Exploring big data opportunities for Online Customer Segmentation

Georgia Fotaki

Marco Spruit

Sjaak Brinkkemper

Dion Meijer

Department of Information and Computing Sciences

Utrecht University

Technical Report UU-CS-2013-021

ISSN: 0924-3275

Department of Information and Computing Sciences

Utrecht University

P.O. Box 80.089

3508 TB Utrecht

The Netherlands

Exploring big data opportunities for Online Customer Segmentation

Georgia Fotaki Marco Spruit Sjaak Brinkkemper Dion Meijer

Abstract

In today's competitive business environment, more and more organizations lead to move or extent their business online. Thus, there is an increasing need for organizations to build concrete online marketing strategies in order to engage with their customers. One basic step towards achieving the objectives related to online marketing is the segmentation of online customers, based on the customer data gathered online. Since there is an onslaught of customer information collected from online sources, new techniques are required for managing and analyzing the huge amount of data; and this is where the concept of "Big Data" can play an essential role.

This research sheds light on three fields: Online Marketing, Customer Segmentation, and Big Data Analytics. The three terms are combined into one framework, which attempts to show how online marketing objectives can be supported by an effective online customer segmentation that can be implemented by techniques and tools applicable to extremely large datasets. For the creation of the framework the following steps are followed: a set of main online marketing objectives is defined; the differences among customer attributes gathered from offline and online channels are discussed and online customer segmentation categories are identified; the concept of Big Data is introduced and relevant techniques and tools suitable for analyzing customer segmentation categories and segmenting customers effectively are described.

Keywords: *Online Marketing, Online Customer Segmentation, Big Data, Data mining, Online Customer Engagement Management.*

PART I: Research Overview

1. Introduction

Today, the Internet has become an essential component of business life and its tremendous impact on organizations' structures is already obvious. During the last decade more and more organizations decide to move or extent their business to an online environment in order to achieve further growth, profits, reputation, as well as to come closer to their customers and meet their needs. Hence, companies today pay even more attention to the establishment of a concrete online marketing strategy. There are already tools in the market facilitating online marketing strategies as well as to strengthen the engagement with online customers such as Web analytics tools, Social Media monitoring software, Web Content Management systems, Audience Targeting tools and Online Customer Engagement Management tools (Fotaki, Gkerpini & Triantou, 2012).

A vital process for Online Marketing is Customer Segmentation. Customer Segmentation constitutes the process of dividing customers into distinct and homogeneous groups and is considered an effective method for managing different customers with different preferences, while developing diverse marketing strategies (Chen et al. 2007, Tsitptis & Chorianopoulos, 2009). Online customers can be segmented according to their characteristics, which are tracked online, with the use of specific techniques and algorithms. Since the amount of customer data that are gathered online grows rapidly, the use of "Big Data" tools and techniques that are able to handle and analyse a huge amount of data on real time has become inevitable. Software tools for Online Marketing and Customer Engagement that already exist in the market can integrate and utilize approaches in order to facilitate and enhance online customer segmentation, as well as other functions they might perform providing a holistic customer view. Therefore, the impact of Big Data techniques and technologies on online customer segmentation needs to be explored.

This research sheds light on three fields: Online Marketing, Customer Segmentation, and Big Data Analytics. These three terms are combined into one framework that attempts to show how online marketing objectives can be supported by an effective online customer segmentation, which can be implemented by techniques and tools applicable to large datasets. In the context of this research the terms "online customer" and "online visitor" both refer to the visitor of a website regardless of their purchasing behaviour.

1.1. Problem Definition and Research Trigger

Online Marketing and Customer Engagement

In today's fiercely competitive business environment, organizations struggle to improve customer experience, achieve customer retention, and expand their customer database. The increasing need to keep the customers satisfied and treating them as individuals by offering right products and services at the right time, has triggered interest in customer engagement. According to Van Doorn et al. (2010), customer engagement is defined as "the behavioral manifestation from a customer towards a brand or a firm which goes beyond purchase behavior". Organizations have already started realizing that traditional Customer Relationship

Management (CRM) approaches that focus on tracking and analyzing customer data should be reconsidered. Furthermore, the tremendous growth of the Internet technologies has lead organizations to pay more attention on building their online marketing strategy and focus on engaging with their customers. Online engagement is defined by Molen & Wilson (2010) as "a cognitive and affective commitment to an active relationship with the brand as personified by the website or other computer-mediated entities designed to communicate the brand value".

In their effort to leapfrog the competition, many organizations have already started employing new Online Customer Engagement Management (OCEM) software tools. Such tools are capable of assisting the whole process of OCE and facilitate the improvement of online campaigns and the engagement of customers with the brand. However, objectives and Key Performance Indicators (KPIs) for online marketing are still, for many organizations, not well defined and not yet aligned with the business strategy (Chaffey et al., 2009). Therefore, marketers still have to put much effort in defining their digital strategy, in order to select the appropriate software tools and effectively utilize them to achieve their business goals.

Big Data

In the meantime, business world is facing the challenge of dealing effectively with the onslaught of data and information. In the past organizations had to handle limited amounts of structured data, which were mostly extracted from traditional business applications, such as Enterprise Resource Planning (ERP) or Customer Relationship Management (CRM). Such data used to be analyzed by traditional data mining techniques. However, as the development of business software application evolves, it is not only the volume of data that grows, but also their complexity. Data and information can be retrieved from both offline and online sources. These data can be structured, semi-structured or unstructured and can be found in several formats, such as video, images or text from social media platforms. This explosion of data in terms of volume, structure and format calls for new approaches capable of processing and analyzing large amounts of data in real time (Forrester, 2012). This kind of approaches of analyzing enormous amount of data is nowadays well known with the term "big data" approaches or analytics.

Big Data is a new term primarily used to describe the data sets that are so large and complex that they require advanced and unique storage, management, analysis and visualization technologies (Chen et al., 2012). Forrester (2012) defines Big Data as *"Techniques and technologies that make handling data at extreme scale affordable*". Big data approaches defer from traditional data mining. Big Data approaches are able to follow the flow of information and analyze data in real time. Since the field of big data analytics is quite new there is not enough scientific literature available yet. However, as mentioned in a recent report by Forrester (2012), organizations are currently using big data sources and integrate new approaches of data analysis in order achieve deeper understanding of their customers and optimization of customer engagement. Hence, it is evident that software tools for OCEM should also utilize big data approaches to enhance the OCE process. The question that arises here is the following: "*How such approaches can assist the processes that should be conducted in order to support online marketing strategies and achieve OCE?"*

Customer Segmentation

One of the basic steps in improving customer's journey and achieving customer engagement is customer segmentation (Fotaki, Gkerpini &Triantou, 2012). Customer Segmentation is a process of dividing the customer base into distinct and internally homogeneous groups in order to develop differentiated marketing strategies according to their characteristics (Tsiptsis & Chorianopoulos, 2009). There are various types of segmentation (thoroughly explained in Chapter 3) based on certain customer criteria or attributes gathered from several sources. Customer segmentation types intent to support different business tasks or activities regarding marketing goals (Tsiptis & Chorianopoulos, 2009). These segmentation types can be analyzed by appropriate analytical techniques or tools, in order to effectively segment a certain customer base.

Although there is a plethora of research conducted for customer segmentation in traditional CRM systems that basically works with data gathered from offline channels, online customer segmentation in real time has not received much attention in research. Another question that arises here is the following: "*How online customer segmentation should be performed in order to serve the goals related to online marketing and OCE?* " Apparently, big data can play a major role in online customer segmentation, since the volume of customer data gathered online rapidly grows. Moreover, there are already available big data tools in the market, which should be able to assist customer segmentation. Therefore, Big Data techniques can be (maybe on their simplest forms) integrated by OCEM tools, in order to perform intelligent customer segmentation resulting in better decision making for assisting online marketing strategy and strengthening customer engagement.

1.2. DEVCORP

Within the spectrum of this research, the OCEM Tool provided by DEVCORP, will be taken as a case study. This product assists most of the business processes of OCEM. The OCEM tool dynamically performs online customer profiling and creates online dialogues, interacting with the visitors on real time. It also performs segmentation of the online visitors. However, the segmentation function is still at an initial point since the product has not integrated any analytics yet. The OCEM Tool is still not able to automatically underline opportunities that might arise from the data. Hence, the product itself constituted the initial trigger of this research. The customer segmentation process should be improved, in order for the product to be able to provide recommendations by analyzing customer data. Moreover, since the OCEM Tool creates dynamic customer profiles by capturing customer attributes on real time, the amount of profiles is growing fast resulting in a huge amount of data. At this point it is worth mentioning that one of DEVCORP clients that implement the OCEM Tool, has created so far more than 50.000.000 customer profiles. Therefore, suitable Big Data approaches should be integrated into The OCEM Tool in order to enhance the intelligent customer segmentation process.

1.3. Research Focus

Based on the specifications described in the paragraphs above, we were triggered to focus this research on the following:

- Identify which are the main business objectives regarding online marketing and customer engagement.
- Point out the differences among customer attributes gathered in offline and online channels in terms of type, volume and structure of data.
- Find out which are the suitable customer segmentation types for each of the online marketing business objectives.
- Find out which Big Data approaches are currently available and what opportunities they offer for analyzing customer data in real-time.
- Explore which techniques, which are also applicable on Big Data, are most appropriate for segmenting online customers according to each customer segmentation type.

1.4. Research Questions

According to the research trigger and research focus stated above, a framework is developed. The framework attempts to conduct an indirect mapping of appropriate Big Data techniques and approaches with goals regarding online marketing that can be assisted by these techniques and approaches, based on online customer segmentation types. The main research question that this research attempts to answer is the following:

MRQ: *"What are the appropriate Big Data techniques that can assist online marketing strategy based on online customer segmentation types***?"**

The framework consists of two parts, while the main research question is divided into three sub-research questions that correspond to each of the parts of the framework.

The first part of the framework includes a set of main business goals that online marketers, who may also make use of OCEM Software tools, aim to achieve. Hence, the first sub-research question that focuses on defining a certain set of online marketing business goals is formulated as follows:

- **RQ1: "***Which are the main business objectives regarding online marketing and online customer engagement?"*

The second sub-research question explores the available customer segmentation types, based on the different sets of customer attributes that can be captured from online channels. The research question also entails the mapping of this segmentation types to the business goals defined in sub-research question 1 that each of the segmentation types can assist. This constitutes the first part of the framework of this research. The second sub-research question is formulated as follows:

- **RQ2:** *"Which are the customer segmentation types that can assist each of the business objectives regarding online marketing?"*

The third sub-research question explores the currently available big data tools, techniques and approaches appropriate for customer segmentation. It also attempts to map the techniques to online customer segmentation types defined in sub-research question 2. The answer to the third sub- research question aims on building the second part of the framework.

- **RQ3: "***Which big data approaches and techniques can be used for each online customer segmentation type?"*

An overview of the main research question of this research and the sub-research questions into which it is divided is given in Listing 1.2:

MRQ: *"What are the appropriate Big Data techniques that can assist online marketing strategy based on online customer segmentation types?"*

- **RQ1:** *Which are the main business objectives regarding online marketing and online customer engagement?*
- **RQ2:** *Which are the customer segmentation types that can assist each of the business goals regarding online marketing?*
- **RQ3:** *Which big data approaches and techniques can be used for each online customer segmentation type?*

Listing 1.2: Research questions

1.5. Scientific and Practical Relevance

Scientific Relevance

Based on the literature review conducted for the purpose of this research, as well as on the interviews and general discussions with the experts of DEVCORP, it can be claimed that most organizations are currently in a transient state; organizations are aiming to put aside traditional marketing techniques and focus more on online marketing and OCE. It is apparent that customer segmentation plays an important role for traditional marketing and the management of relationship with the customer. Therefore, there is enough literature about customer segmentation, describing segmentation types and techniques that are able to support marketing goals. Moreover, there is scientific research on how different customer segmentation types can be analyzed with the use of classic data mining techniques for effective customer segmentation. However, existing research approaches on customer segmentation focuses mostly on the analysis of customer data stemming from offline sources, such as from business applications capable of managing customer data, like CRM. As observed, literature related to the segmentation of online customers is limited and primarily focuses on online buyers' behavior in web shops. It is apparent that there is a need to look deeper into the types of customer attributes that can be gathered from online sources and identify online customer segmentation categories which can support the achievement of online marketing goals.

Moreover, the uprising volume of online customer data calls for new approaches able to handle and analyze the tremendous amount of data. Although, the concept of big data seems to be of great importance for online marketing and online customer segmentation, there is not any scientific literature explaining how big data analytics can be utilized for an effective online customer segmentation, which can support online marketing strategies.

This research raises the issue of the effective online customer segmentation, based on customer attributes gathered from online channels, capable of assisting online marketing objectives with the use of big data tools. The differences among customer data gathered from offline and online sources are highlighted. A framework is proposed, which connects business goals regarding online marketing and OCE with the appropriate online customer segmentation categories, able to support these goals. According to Tsiptsis & Chorianopoulos (2009) there are many different segmentation types, which are based on the specific criteria or customer attributes used for customer segmentation. Each segmentation type can assist certain business goals. Moreover, the framework includes the analytic techniques suitable for analyzing the online customer segmentation types, in order to bring out results for effective online customer segmentation. These analytic techniques can be implemented in a big data environment by big data tools as presented in the framework. Figure 1.1 illustrates a schema of the framework, which is presented extensively in Chapter 6.

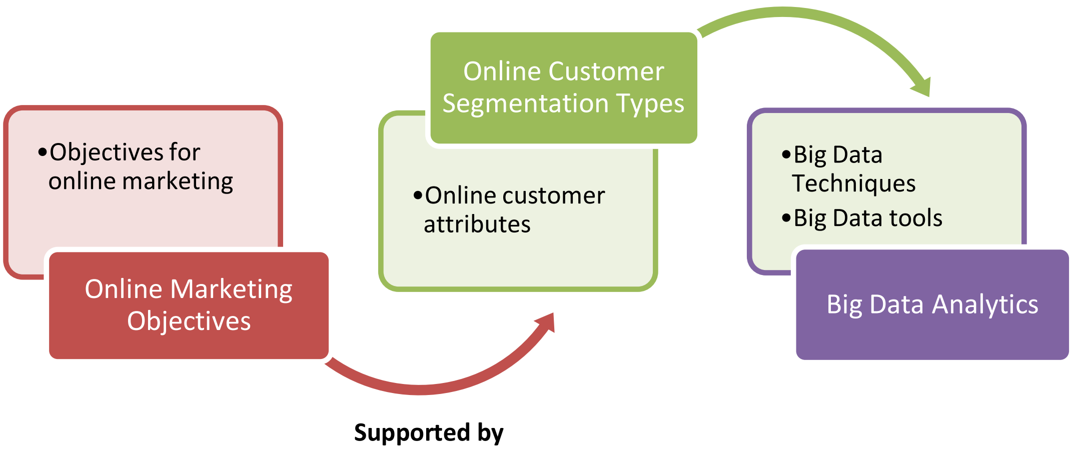

Figure 1.1: Online Visitors Segmentation Framework with Big Data techniques

Practical Relevance

From a practical point of view, primarily this research will be beneficial for DEVCORP and The OCEM Tool. As mentioned above, The OCEM Tool performs segmentation of online visitors but without the use of analytics. Thus, The OCEM Tool's ability of segmenting customers is still at an initial stage. In fact, in order to perform segmentation the user of The OCEM Tool should manually select relevant customer attributes in order to create customer segments. Moreover, the users of The OCEM Tool need the guidance of consultants in order to properly configure the product and decide upon effective segments. For the online marketers-users of The OCEM Tool it's not always easy to define a clear online marketing strategy and decide upon segmentation. Therefore, the framework that this research proposes can be helpful as an overview for online marketers in order to decide upon segmentation when starting with certain business goals. Additionally, it can be also used by business consultants as a guideline when consulting new customers that implement The OCEM Tool on customer segmentation.

As far as big data techniques and technologies are concerned The OCEM Tool uses Apache Cassandra, which is an open source database management system provided by the Apache Software foundation. Apache Cassandra is characterized as a big data technology that is able to handle huge amount of data. Moreover, Apache Solr is used as a search engine, which is highly scalable and able to handle big data. However, so far Apache Solr is used for querying and simple computations. Apart from the aforementioned tools, The OCEM Tool does not integrate any other kind of tool or approach that is related to Big Data. Therefore, the framework that this research proposes will provide DEVCORP with a list of online customer segmentation types suitable for supporting online marketing objectives. The proposed online customer segmentation types can also be inserted into the product afterwards. Moreover, software architects of DEVCORP will have a list of tools or techniques that are suitable for online customer segmentation and can be possibly integrated into The OCEM Tool in order to perform customer segmentation and possibly assist other processes that the product performs.

Concluding, we believe that apart from DEVCORP, the findings of this research can be also beneficial for other vendors providing software tools similar to The OCEM Tool, who wish to integrate big data analytics into their products for effective online customer segmentation capable of assisting online marketing strategies. The proposed frameworks could be used as an initial guideline point in proceeding with effectively managing and analyzing large amounts of customer data, in order to create customer segments that would provide a holistic view of the online customer.

1.6. Report Structure

This thesis report presents the research conducted in order to provide answers to the research question and the sub questions stated in Section 1.4. This report consists of 9 Chapters including the introduction (Chapter 1) and is divided into 5 parts. The first part is called Research Overview and contains Chapter 1 and Chapter 2, which is a detailed description of the research approach followed to answer the research questions. The second part of the research is called "Theoretical Background" and consists of the Chapter 3 and 4. Chapter 3 covers the theoretical background regarding Online Marketing and the objectives related to it. Moreover, it explains in details the task of "Customer Segmentation" and the customer segmentation types that already exist in the literature. Chapter 4 constitutes an introduction to the notion of "Big Data", explains the difference from traditional data mining techniques, and enlists and describes currently available Big Data techniques and tools. The third part is called "Empirical Data" and includes Chapter 5. Chapter 5 presents the empirical data gathered through the case study in DEVCORP. It contains information gathered through the explorative interviews with the experts of DEVCORP regarding the Online Marketing goals. Furthermore, it presents the data gathered from The OCEM Tool, in order to shed light on the differences among the customer attributes gathered from offline and online channels and identify online customer segmentation types. The data presented in Parts II & III are all used as an input for the framework that this research proposes. The fourth part is called "Framework" and contains only Chapter 6, which shows the construction of the framework. Finally, the fifth part of the research contains Chapters 7 and 8. Chapter 7 presents the evaluation of the results of the research. Chapter 8 presents the conclusions and limitations of the research, as well as suggestions for further research. Chapter 9 includes the bibliography used for the purpose of this research. In the end of this document Appendices with detailed tables, images and other information that are not presented in details in the main text, are provided.

2. Research Approach

In the introduction of this report, a comprehensive description of the research domain, the research problem and the research triggers that motivate the research is provided. Moreover, the research questions were formulated and presented and an overview of the main result of the research are illustrated. In this chapter, the approach followed to answer the aforementioned research questions is described.

2.1. Design Science Research

Design Science constitutes a problem-solving research paradigm, which aims in providing innovative solutions in the form of artifacts to related organizational problems. This is achieved through the analysis, management, design, implementation and use of information systems (Hevner et al., 2004).

In order to answer the research questions of this research, and achieve the respective research objectives, a research method appropriate for designing frameworks should be used. According to Vaishnavi, & Kuechler (2012) the Design Science Research (DSR) is appropriate when the solution provided to a specific problem is the design of an abstract artifact, such as a framework, a prototype, a models, a new method etc.

2.1.1. Guidelines to Design Science Research

In order to conduct and evaluate design science research a certain process should be followed. Hevner et al. (2004) suggested the following set of guidelines, which constitutes a clarification of how DSR should be conducted and how is placed into this research.

Guideline 1: Design as an Artifact.
First of all, the DSR bears as a result an artifact such as a model, a method or a framework that can be applicable. This research results in suggesting a framework for Online Customer Segmentation. The framework maps online marketing objectives with relevant online customer segmentation categories, which can assist these objectives. Moreover, it entails the use of analytic techniques that can be implemented by big data tools and technologies.

Guideline 2: *Problem relevance*.
The result of the DSR under investigation should be a solution based on technology and be relevant for a specific business problem. As explained in Chapter 1, the result of this research attempts to highlight the gaps found in the existing literature related to online customer segmentation and its effectiveness to online marketing. Moreover, it raises the issue of big data, which constitutes a current trend in analytics, points out differences and similarities with data mining and explains its relevance to online customer segmentation.

Guideline 3: *Design evaluation*.
The resulted artifact should be evaluated precisely in terms of its utility, quality and efficacy. The framework that this research proposes is high level, and could not be tested on real situation due to time limitation. However, its quality is ensured as it is based on a concrete theoretical background and it is evaluated by experts in terms of utility and effectiveness.

Guideline 4: ***Research contributions***.

The DSR should be clear enough in the areas of design artifact, design foundations and methodologies, in order to be effective. Respectively, this research contains a detailed explanation of the scientific methods and methodologies used for the design of the proposed framework.

Guideline 5: ***Research rigor.***

The construction and the evaluation of the artifact should be conducted with the application of rigorous methods. Therefore, the construction of the framework was based on a combination of literature and empirical data, while its evaluation is based on the opinions of experts on the related domains.

Guideline 6: ***Design as a search process.***

In order to search for an effective artifact it is mandatory to utilize available means to achieve the research goals and comply with the laws of the problem environment. The design of the framework of this research is considered effective, since it is based on existing theory as well as on explorative interviews and data observations which were able to build the base for the final result.

Guideline 7: ***Communication of research.***

The research should be presented and explained effectively to both technology and management oriented people. The research was presented in people related to the fields of online marketing, business, and software engineering data analytics.

2.1.2. Design Science Research Methodology

There are various DSR approaches available in the scientific literature. The approached used for the purpose of this research is the Design Science Research Methodology, as it was proposed and defined by Peffers et al. (2008). DSRM process is considered suitable for this research, since the main objective of a DSRM is to provide a certain artifact, which contains the characteristics of the research outcomes. In this research the artifact is a framework that constitutes an indirect mapping among big data techniques for online customer segmentation and online marketing goals.

The DSRM consists of the six activities as described below:

1. Problem identification and motivation: Define the specific research problem and justify the value of a solution to this problem.
2. Define the objectives for a solution: Identify the objectives for the solution to the problem.
3. Design and Development: Create the artifact.
4. Demonstration: Show how the artifact is capable of solving one or more instances of the problem.
5. Evaluation: Measure and observe how well the artifact supports a solution to the problem.
6. Communication: Communicate the problem and its importance, and indicate its effectiveness to professionals and researchers. Make suggestions for future research.

2.1.3. Research placed into DSRM

Figure 2.1 illustrates how the approach of this research can be placed into DSRM.

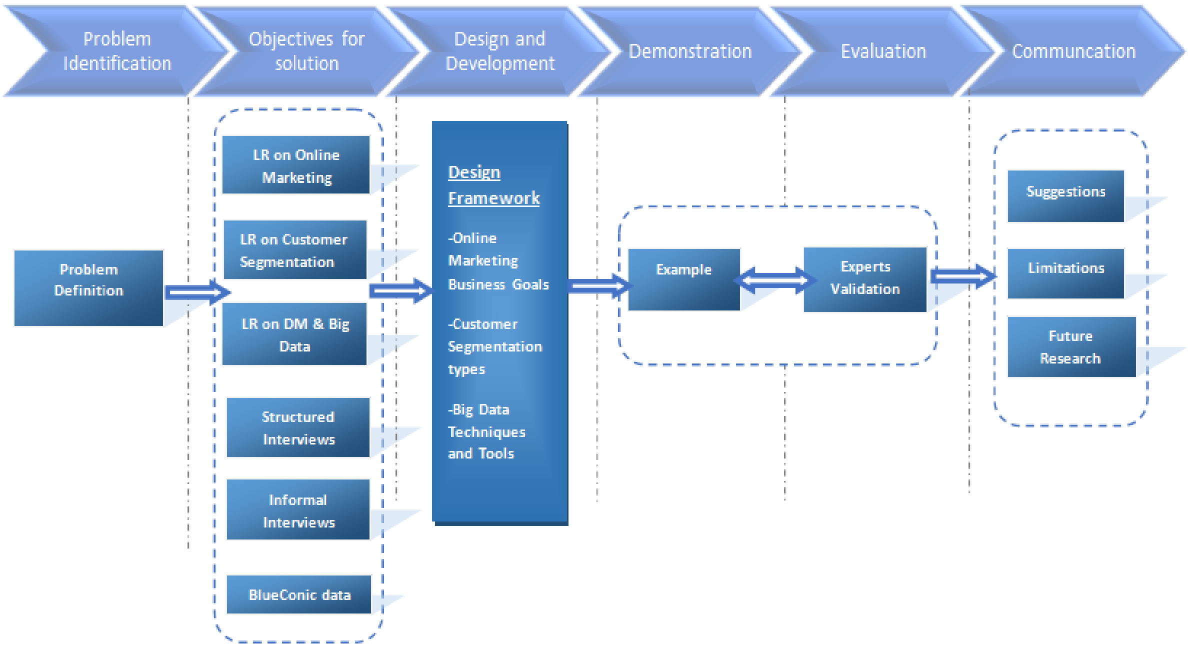

Figure 2.1: Research approach placed into DSRM

Starting with the **first step** of the method the problem is described from both scientific and practical point of view. A literature study is performed in the domains of Big Data, OCE, and Customer Segmentation, so as to discover knowledge gaps. Additionally, several unstructured interviews and discussions with professionals of DEVCORP were conducted prior to the start of the research in order to specify the problem.

As of **second step**, the objectives for the framework are defined. At this point the literature review will focus on gathering data, the analysis of which will be able to answer each of the aforementioned research questions. Additionally, unstructured and semi-structured interviews are conducted.

More specific, in order to design the framework three factors are taken into account; (i) online marketing business objectives, (ii) customer segmentation types, and (iii) big data techniques and tools. Initially, a set of (high level) business objectives regarding online marketing and OCE is going to be determined. The set of goals stems from both the scientific theory and the data gathered during the interviews with the business consultants and marketers. Secondly, several customer segmentation types are thoroughly described. There is already scientific literature on this field, but it is mostly related to CRM and offline channels. A set of different customer segmentation types is presented in the book of Tsiptsis & Chorianopoulos (2009). Therefore, complementary to the literature review on customer segmentation, this research has taken the OCEM Tool as a case study and data are observed and analyzed, in order to identify online customer segmentation types. Finally, a comprehensive literature review is made on current big data techniques and tools, and their differences and similarities to data mining are highlighted. A focused literature study regarding techniques used for customer segmentation is made and approaches that are appropriate to assist customer segmentation are presented and analyzed.

In the **third step**, the above data will be analyzed in order to move on with the design of the framework which will map the Big data techniques and tools with the business goals related to online marketing and OCE , based on different online customer segmentation types.

For the **fourth step**, a hypothetical yet representative example of how the framework could be utilized in the OCEM tool's case is given. Since the proposed framework is high level and theoretical, an implementation on a real situation would require time that would greatly exceed the time limit of this research. Furthermore, for the evaluation, which constitutes the **fifth step** of DSRM the framework is validated by experts.

Finally, recommendations on how to utilize such a framework could be utilized are made.
Furthermore, related gaps and limitations are discussed, while suggestions for future research are made.

Figure 2.2 illustrates the process described above in the form of a Process Deliverable Diagram (PDD) as it is defined by Weerd & Brinkkemper (2008). The processes of the DSRM approach are illustrated in the left part, while the related deliverables are illustrated in the right part. Moreover, in the right part of the diagram, the corresponding parts of the research in which the deliverables are presented are depicted.

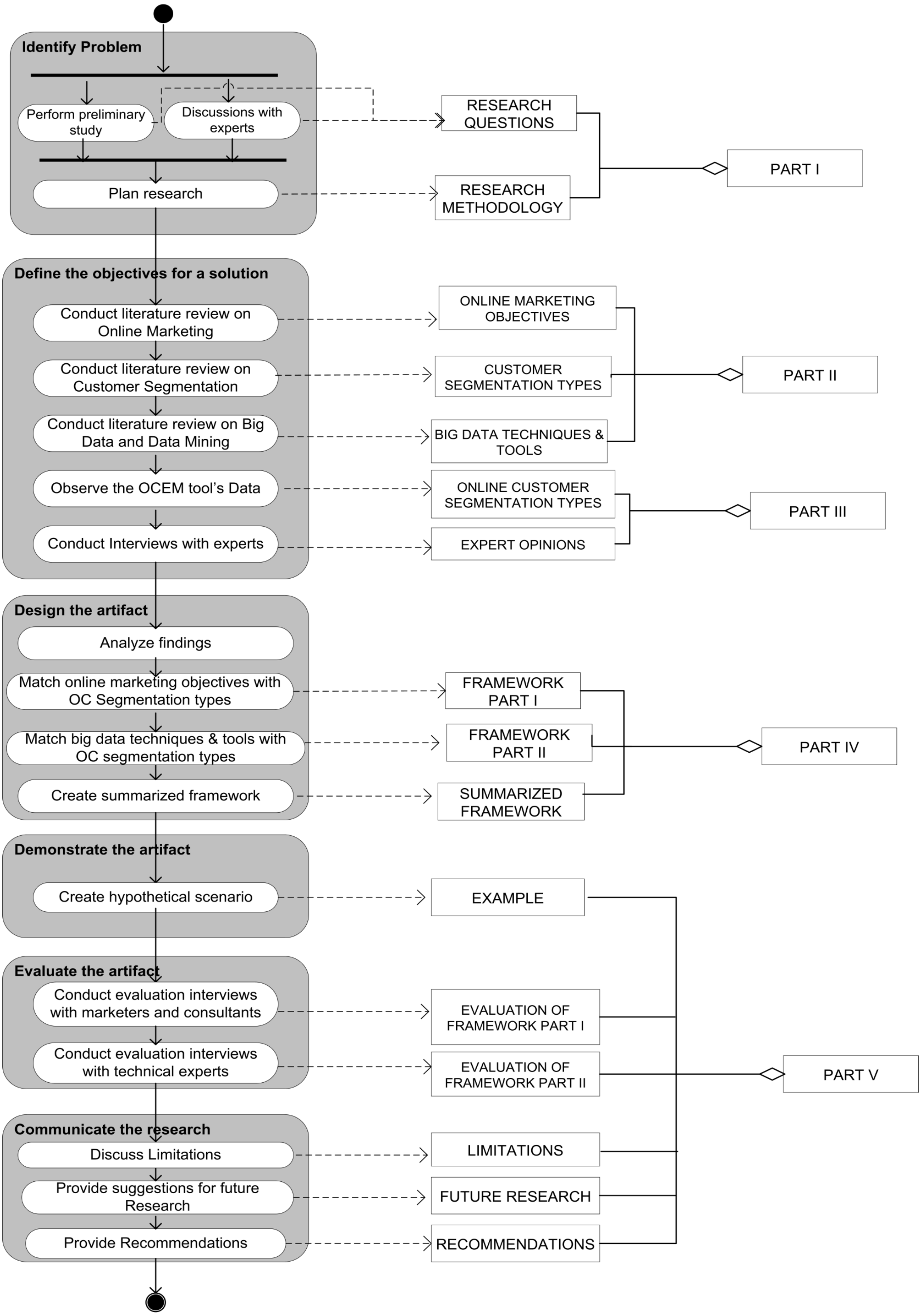

Figure 2.2: PDD of Research Approach

2.1.4. Research questions linked to research activities and deliverables

Table 2.1 shows how the aforementioned research activities are linked to each of the research question, as well as the related deliverables.

Research Questions	Research Activities	Deliverables
RQ1: *"Which are the main business objectives regarding online marketing and online customer engagement?"*	▪ Literature review on online marketing ▪ Unstructured Interviews with DEVCORP experts	▪ Literature review on Marketing ▪ Set of online marketing goals
RQ2: *"Which are the customer segmentation types that can assist each of the business goals regarding online marketing?"*	▪ Literature review on Customer Segmentation. ▪ The OCEM Tool Data ▪ Interviews	▪ Literature Review on Customer Segmentation ▪ Online Customer Segmentation types. ▪ Framework linking online customer segmentation types to objectives
RQ3: *"Which big data approaches and techniques can be used for each online customer segmentation type?"*	▪ Literature review on Big Data ▪ Literature review on data mining for customer segmentation."	▪ List of techniques applied on Big Data ▪ List of main big data tools and technologies ▪ List of main data mining techniques for customer segmentation ▪ Framework suggesting analysis of online customer segmentation types with certain techniques applicable on big data

Table 2.1: Research questions linked to research activities and main deliverables

2.2. Literature Review

In order to answer the research questions that this research poses a literature review was conducted, regarding the following domains: Online Marketing, Customer Segmentation, and Big Data. The literature review was based on academic journals, books, thesis projects, analyst reports, and white papers. Sciencedirect.com, Omega Library and Google Scholar, as well as recommendations from the supervisors of this research were the basic source of literature selection. In order to ensure that the information of this research is up-to-date, we focused on selecting relatively new studies (after 2005). However, valuable content from older studies was also used, when considered necessary.

As explained below, popular and common terms were used in order to create queries for searching papers in the aforementioned databases. Thus, the initial search results of every query were numerous. For this reason, the titles of the publications of the first four pages that each query resulted in were read and the irrelevant studies were excluded. Afterwards, the abstracts of the remaining studies were read. More specific:

- In order to understand Online Marketing and the objectives related to it a literature review was conducted. Hereby, the terms “online marketing”, “internet marketing”, “customer engagement” was searched alone, as well as in combination with the terms “KPIs” and “objectives”, resulting in 9 queries. After reading the titles of the four first pages only scientific publications relevant to the topic were taken into account. In this case, books introducing online marketing and analyzing related KPIs and objectives were preferred. 11 studies were selected, including 5 books.

- In order to explore Big Bata and find out main techniques and tools as well as its implication on customer segmentation. The following queries were conducted: The terms “big data”, “big data tools”, “big data techniques”, “big data analytics” were firstly searched alone and then in combination with the terms “customer segmentation” and “online customer segmentation”, resulting in 12 queries. 12 studies discussing the concept of big data and related techniques and tools were selected. However, when the first four terms were combined with the terms “customer segmentation” and “online customer segmentation”, the occurring research results were irrelevant with the scope of the search. Thus, it was concluded that there is no sufficient related scientific research on the use of big data for customer segmentation.

- In order to explore customer segmentation and find out the basic customer segmentation types, as well as find related literature for online customer segmentation the terms “customer segmentation” & “customer segmentation types” were searched alone, as well as in combination with the term “online”, resulting in 4 queries. From the outcomes of the 18 studies were selected including 2 books, and a master thesis. 3 of the studies were focused on online customer segmentation.

- Since no literature was found in the initial search regarding big data application on customer segmentation, a literature review focusing on data mining techniques for customer segmentation was made. The purpose was to find out, which techniques are preferred for certain customer segmentation types. We searched in the aforementioned databases for academic journals, books or thesis projects, which were published after 2005 and included case studies on customer segmentation with data mining techniques. The keywords used as search terms for this purpose were: “customer segmentation”& “data mining”. 15 studies were selected including a book and master thesis. 3 of the studies discuss customer segmentation based on data gathered from online sources.

  Table 2.3 shows how literature review on each of the fields is linked to each of the sub-research questions:

Research Questions	Literature Study	Number of Studies
RQ1: *"Which are the main business objectives (high level) regarding online marketing and online customer engagement?"*	▪ Literature review on online marketing	11
RQ2: *"Which are the customer segmentation types that can assist each of the business goals regarding online marketing?"*	▪ Literature review on Customer Segmentation.	18
RQ3: *"Which big data approaches and techniques can be used for each online customer segmentation type?"*	▪ Literature review on Big Data	12
	▪ Literature review on data mining techniques for customer segmentation	15

Table 2.3: Literature Review linked to research questions

2.3. Expert Interviews

For the purpose of this research, explorative interviews were conducted with 6 experts of DEVCORP, as well as with an author of one of the books used during the research. Unstructured and semi-structured interviews were organized in two rounds that took place before and after the creation of the framework that this research proposes.

Interview purpose

The aim of the interviews is divided across two interview rounds, both serving a different purpose. Table 2.3 depicts the details of interviewees. The names of the interviewees remain anonymous for privacy purposes.

During the first round, 4 interviews were organized with the experts of DEVCORP. The interviewees were the following: two consultants of DEVCORP, that are responsible for configuring the product and guiding clients on how to use it; the CMO and the product marketer. Hereby, the aims of the interviews were multiple. Firstly, the interviews aimed in gaining a deeper understanding of the online marketing and customer segmentation topic and identify any existing gaps and problems. Secondly, the business goals related with online marketing were discussed in order to decide upon a certain set of objectives that could be used for the purposes of the research. Finally, discussions were made on segmentation categories based on data gathered online and which online marketing goals they can more effectively assist. The data gathered during the first round were used as a complement to the theoretical background of this research and constituted an input for the construction of the framework that this research proposes.

The second round of interviews served the purpose of evaluating the results of the research from the experts' point of view. Therefore, the first part of the framework considering Online Objectives and segmentations techniques was presented to marketers and consultants, and questions concerning the usefulness of each segmentation category for the related objectives were made. Moreover, an interview with an author, who is also specialized in Customer intelligence, was organized in order to evaluate the second part of the proposed framework that suggests which techniques are preferred for certain segmentation categories. Finally, the software architect of DEVCORP also evaluated the second part, while commenting which techniques would be ideal for the case of the OCEM Tool.

Interviewee ID	Interviewee Job Title	Years of Experience	Interview round participated in
M1	Chief Marketing Officer	22	Rounds 1&2
M2	Product Marketer	4	Rounds 1&2
B1	Business Consultant	14	Rounds 1&2
B2	Business Consultant	13	Round 1
B3	Online Marketer /Business Consultant	4	Round 2
A1	Author / Customer Intelligence Expert	13	Round 2
A2	Software Architect	15	Round2

Table 2.3: Anonymized List of Interviewees

2.4. Case Study

For the purpose of this research the case of the OCEM Tool, an OCEM software tool, was examined. The OCEM Tool is a software tool capable of gathering customer information through multiple online channels in order to create dynamic customer profiles. The tool itself provides the process of customer segmentation, but it is still at an initial state, without the integration of analytics.

Within the spectrum of this research, the OCEM Tools' case was examined for multiple purposes. The first purpose was to understand what type of customer attributes can be gathered from online channels, and identify online customer segmentation types. To do so, customer attributes that are gathered from DEVCORP clients who use the OCEM Tool and have created so far more than 1.000.000 customer profiles were observed. The second purpose was to find out which are the goals of online marketers that use a product like the OCEM Tool. Therefore, discussions and interviews with business consultants and marketers of DEVCORP took place to discuss upon the topic from their customers point of view. Finally, the current segmentation process that is basically driven by the consultants of DEVCORP was also observed, in order to clarify how the current clients decide upon customer segmentation and what are the obstacles for effective online customer segmentation.

2.5. Experts Evaluation

The developed framework was validated by experts. Since the framework includes both business and technical aspects, different kind of expertise were needed for the evaluation. Therefore, each of the two parts of the framework were discussed and evaluated separately from people of different expertise

Framework-Part I

The first part of the framework, which relates the online marketing objectives with the appropriate segmentation types, was evaluated during the 4 interviews with the marketers and the consultants of DEVCORP. For this purpose, the framework was presented to the interviewees and questions were addressed regarding the following:

- The relevance of the presented objectives to online marketing
- The relevance of the presented customer segmentation categories. .
- The usefulness of each segmentation category for each of the objectives
- The possibility of utilization of such a framework.
- The problems and the obstacles that would appear in utilizing of such a framework.

Due to the fact that consultants are the ones who configure the OCEM Tool according to their clients' needs, and advise them on how to use the product and perform segmentation, they were also able to answer to the questions expressing also their clients' point of view, giving relevant examples from certain clients' cases when needed.

Framework part II

The second part of the framework, which shows which techniques are more suitable for analyzing certain customer segmentation categories, is based on a literature review made in the field of data mining and big data analytics. At the time that the research was being conducted, it was not possible to test different algorithms on real data in order to further examine their effectiveness and validate the results of the literature review. Hence, two semi-structured interviews with experts were conducted for the evaluation of the second part of the framework. Each one of the interviews aimed to evaluate the final result from a different perspective. The first interview was with the software architect of DEVCORP, who is responsible for the OCEM Tool. He reviewed the results of the literature study and commented on if the algorithms would be applicable for the analysis of the data that the OCEM Tool gathers from online channels and whether they could be effective or not. The second interview was with one of the authors of the book "Data mining in CRM: Inside Customer Segmentation", which is an expert in the field of customer intelligence. He was asked to comment only on the results of the literature review, while giving insights for the effectiveness of each technique. Although he is specialized in CRM systems and customer segmentation and he is not involved with Online Customer Segmentation, he was also able to give his point of view, evaluating the results of the research. To sum up, the questions addressed to the interviewees regarded the following:

- The relevance of the techniques to the segmentation categories.
- The applicability of the techniques to the product.
- The possibility of utilization of such a framework.
- Possible tools that could be used.
- Problems and obstacles that would occur in utilizing such a framework.

PART II: Theoretical Background

3. Online Marketing and Customer Segmentation

The initial point of satisfactory customer segmentation is the determination of business objectives and relevant KPIs related to marketing and the relationship with the customers (Tsiptsis & Chorianopoulos, 2009). Correspondingly, before proceeding to segmentation of online customers, it is essential that online and interactive marketers define a clear online marketing strategy and all the business tactics and processes involved in it. This Chapter introduces the concepts of Online Marketing and Customer Segmentation. Related objectives and KPIs are presented, while customer segmentation types are thoroughly explained.

3.1. Online Marketing

In the following subchapters, the concepts of Online Marketing and online customer engagement are introduced and related objectives and KPIs are presented.

3.1.1. What is Online Marketing?

Since more and more organizations choose to run their business online, online marketing plays a vital role for their success. Online Marketing as a term can be found in many related studies as E-marketing, Digital Marketing or e-marketing. It regards the set of marketing elements, tools and methodologies for advertising and promoting products through the internet. In order to introduce the term Online Marketing, an overview of the different terms that are used for Online Marketing and the related definitions is provided on Table 3.1.

Term used for Online Marketing	Definition
E-marketing	*“The online management process responsible for identifying, anticipating and satisfying customer needs profitably”* (Chaffey & Smith,2008)
	*“Achieving marketing objectives through use of electronic communications technology.”(*Chaffey et al..,2006)
Digital Marketing	*The use of digital technologies to create an integrated, targeted and measurable communication which helps to acquire and retain customers while building deeper relationships with them (Smith,2007)”* (Wymbs, 2011)
	”The management and execution of marketing using electronic media such as the web, e-mail, interactive TV and wireless media in conjunction with digital data about customers’ characteristics and behavior” (Chaffey et al. ,2006)
Internet Marketing	*“The application of the Internet and related digital technologies in conjunction with traditional communications to achieve marketing objectives”(Chaffey et al. 2006)*

Table 3.1: Online Marketing Definitions

3.1.2. Concept of Online Customer Engagement

Customer Engagement is defined as "the behavioral manifestation from a customer towards a brand or a firm, which goes beyond purchase behavior" (Van Doorn et al., 2010). Many researchers support that online customer engagement constitutes a very important notion for Online Marketing, while others do not pay much attention on it (Moley & Wilson, 2010). Online Customer engagement according to Brodie et al. (2011) is an estimation of the degree of visitors' interaction on the site, which is measured against a clearly defined set of goals. This means that online customer engagement indicates customers' behavior towards a brand and reveals the level that the customer is engaged with a brand online. Hence, goals of online marketers are related with the achievement of customer engagement.

There are few tools in the market that are intended for OCE. These tools are different from traditional CRM tools, since they take into account customer behavior from multiple online channels creating a holistic view of the customer. Such tools are Web analytics tools, Social Media monitoring software, Web Content Management systems, Audience Targeting tools and Online Customer Engagement Management tools, such as the OCEM Tool (Fotaki et al., 2012). A summary of the characteristics of such tools is depicted in Figure3.1 below.

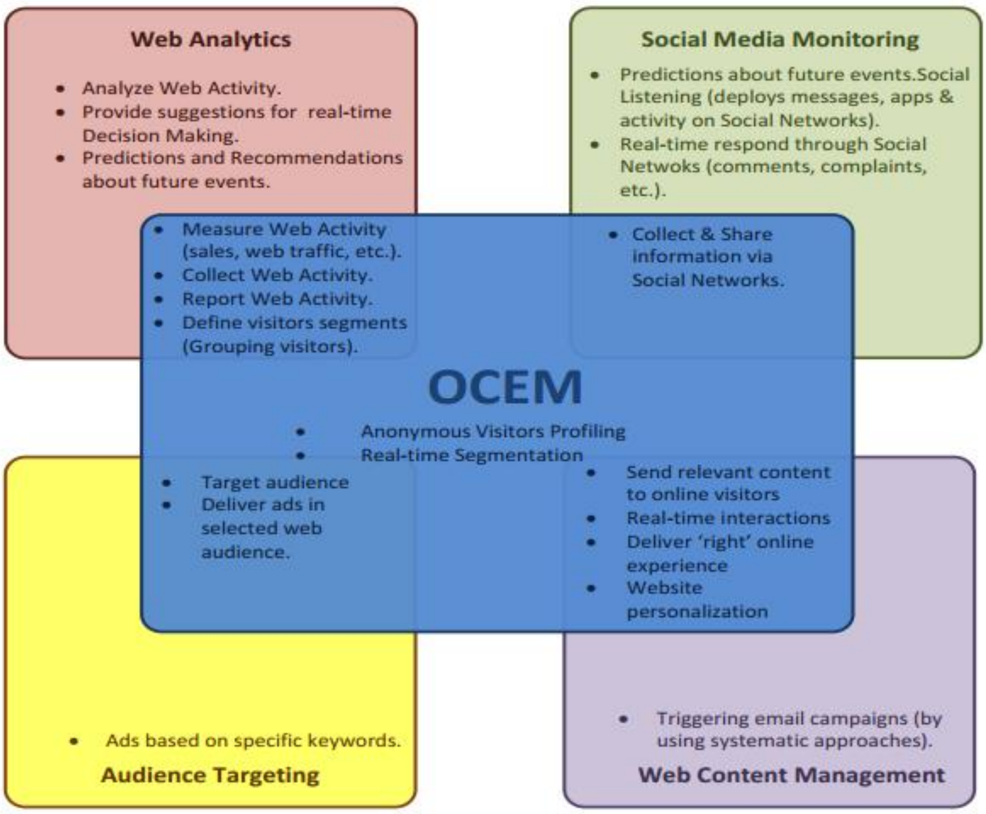

Figure 3.1: Tools for Online Customer Engagement (source: Fotaki et al., 2012)

A framework for Online Customer Engagement is the OCEM (Online Customer Engagement Management) framework, which shows the steps to achieve online customer engagement and the functional requirements of a tool for OCEM (Fotaki et al., 2012). As it is illustrated in Figure 3.2, the OCEM framework consists of three ongoing phases, which result in Online Customer Engagement Optimization: Identification, Evaluation and Reaction. The basic tasks entail the identification of customers' behavior, the analysis of customer characteristics and the development of targeted campaigns according to visitors' online behavior. The software tools mentioned above are able to support, at least some –if not all- of the online customer engagement phases.

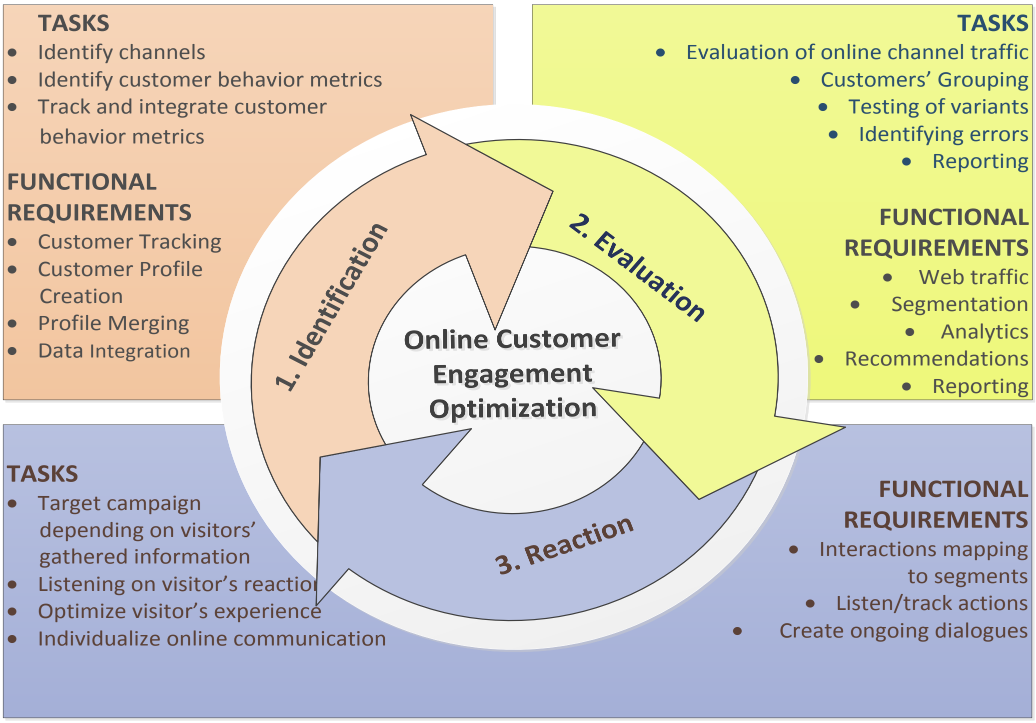

Figure 3.2: Online Customer Engagement Management (OCEM) Framework (Fotaki et al. 2012)

3.1.3. Online Marketing Objectives

As stated by Chaffey & Smith (2008) one of the main difficulties that organizations which choose to go online face, is the lack of clear online marketing objectives that leads them to the random use of e-tools without first agreeing on clearly defined objectives. Furthermore, the unclear responsibilities for many tasks related to online marketing, the absence of measurable objectives, and the treatment of internet as "just another channel to the market", constitute some of the most common challenges in managing an Online marketing Strategy (Chaffey et al., 2009). Organizations are aware of the principles of Online Marketing, but they do not actually act accordingly.

Online marketing itself is clearly a customer centric approach to marketing since the digital media provide marketers with the capability to have a detailed view of their customers and treat them as individuals (Chaffey et al., 2009). This customer-centric approach also entails the notion of OCE, which is still in its infancy. Organizations aim to be engaged with their online customers, but they still do not have a defined strategy to achieve it and thus, their objectives concerning customer engagement are still vague.

For the purpose of this research project a focused literature study was conducted for exploring the main goals related to online marketing. Although there is enough literature regarding Online Marketing and best practices, the related goals are still not well defined. Most of the authors claim that organizations nowadays have started building their online marketing without having defined certain measurable business goals. However, there are general guidelines and frameworks, which organizations can follow in order to build their online marketing strategy and achieve customer engagement.

Chaffey & Smith (2008) in their book summarized online marketing objectives into five broad areas, which are called "the 5Ss; Sell, Serve, Speak, Save, Sizzle". According to these areas specific and measurable business goals can be set. More specific, "Sell" regards the objective for increasing sales, through wider products' and services distribution and promotion. "Serve" concerns the addition of value to the organization and customer engagement, by improving the online experience of customers and encouraging the online dialogue with the customer in order to inspire and inform the development of new products. "Speak" refers to build stronger relationships with the customers, by tracking them, creating online dialogues, or requesting feedback. "Save" regards cost reduction through the increase of online sales and service transactions, online promotions. Finally, "Sizzle" refers to the enhancement of online branding. Table 3.2 shows some examples of online marketing typical goals and KPIs related to the 5Ss.

5S	Examples of Related Objectives
Sell-Grow Sales	• *Achieve 5% increase of online sales in market* • *Increase online sales in a year.*
Serve-Add Value	• *Increase number of customers actively using online services to 30%.* • *Increase time duration on site.* • *Increase interaction with different content on site.*
Speak- Get Closer to Customers	• *Increase visitors to community site section or increase ratings/reviews and discussions.* • *Grow e-mail coverage to 50% of current customer database.*
Save- Reduce Cost	• *Reduce cost of direct marketing by 15% through e-mail* • *Increase web self-service to 40%.*
Sizzle- Extend the brand online	• *Add two new significant enhancements to the customer online experience* • *Improve metrics such as: Brand awareness, purchase intent and Brand favorability.*

Table 3.2: Goals related to the 5Ss (adapted from Chaffey & Smith, 2008)

Chaffey, Chadwick, Johnston & Mayer (2009) in their book refer to Online Marketing as "Internet Marketing" and define Internet marketing strategy, as "the approach by which Internet marketing supports marketing and business objectives". A well-defined online marketing strategy is able to support business processes and marketing objectives.

One of the crucial factors of Online Marketing Strategy is the capturing and understanding of the stage of the customer while their relationship with the organization is created. The stages through which a customer passes during a long term relationship with an organization are called as a whole *Customer Lifecycle*. Those stages are; Acquisition, Retention and Development. *Customer Acquisition* regards the processes to attract new customers. *Customer Retention* includes the activities to maintain relationships with existing customers. Finally, *Customer Development* regards activities aiming to extent customer's involvement with the organization.

Customer development also includes techniques such as cross and up-sell. Figure 3.3 illustrates some online marketing activities that support each of the stages.

Acquisition	Retention	Development
Email Marketing	Email Marketing	Online Content Creation
Online Advertisement	Loyalty Programs	Online Customer Service
Online PR	Online Customer Management	Online proposition development
Pay Per Click Search	Personalization	Site usability & accessibility
Search Engine Optimization	Touch strategy definition	Web Content Management

Figure 3.3: Online Marketing Activities supporting Acquisition, Retention, Development (Adapted from Chaffey et al. 2009)

Online marketing objectives are consistent with the support and facilitation of online marketing activities similar to those shown in Figure 3.1. Acquisition of new customers can be enhanced by effectively guiding them through search engine optimization and exposure to relevant content to the company's website; Retention can be achieved by treating the customer as individual and try to satisfy their needs and expectations; Development can be enhanced by web content management, better service and propositions according to clients' needs. Some examples of objectives that regard the achievement of Customer Acquisition, Retention and Development are illustrated in Figure 3.4 (Chaffey et al. 2009). Obviously, Acquisition is related with the increase of the number of new online customer, which makes use of certain content or services; Retention is related with users' online activity and development with the increase of conversion rates and customers' value.

	Objectives
Acquisition	• Increase online sales of a certain target market. • Migrate a certain percentage of customer to make use of an online billing service. • Increase the number of newly registered users.
Retention	• Increase percentage of active users. • Increase online customer satisfaction rating. • Increase annual repeat new customer conversion rate.
Development	• Increase the average online order value per customer. • Increase website conversion rate. • Increase percentage of online services enquiries fullfiled online

Figure 3.4: Online Marketing Objectives for Acquisition, Retention &Development (Adapted from Chaffey et al. 2009).

	Business types			
KPIS	Online Retailer	Business Site	Content Site	Customer Support
Average Cost per Conversion	✓	✓		
Average Cost per Visit			✓	
Average Order Value	✓			
Average Page Views per Visit			✓	
Average Revenue per Visitor	✓	✓	✓	
Average Product Searches per Visit		✓		
Average Time to Respond to Email Inquiries	✓	✓		✓
Average Visits per Visitor		✓	✓	
Buyer Conversion Rate	✓			
Cart Completion Rate	✓			
Checkout Completion Rate	✓			
Download Completion Rate				✓
Form Completion Rate				✓
Information Find Conversion Rate				✓
Landing Page “Stickiness”	✓	✓	✓	
Lead generation rate per campaign or Campaign Type		✓		
New and Returning Visitor Conversion Rate	✓			
Order Conversion Rate	✓			
Order Conversion Rate per Campaign	✓			
Percent High and Low Satisfaction Visitors and Customers	✓			✓
Percent New and Returning Customers	✓			
Percent New and Returning Visitors		✓	✓	✓
Percent Orders from New & Returning Visitors &Customers	✓			
Percent Revenue from New &Returning Visitors &Customers	✓			
Percent Visitors in a Specific Segment		✓		✓
Percent Visitors Using Search				
Percent Zero Result Searches	✓	✓		✓
Percentage of High, Medium and Low Click Depth Visits (Interest Categories).				✓
Percentage of High, Medium and Low Frequency Visitors			✓	
Percentage of High, Medium and Low Regency Visitors	✓	✓		
Percentage of High, Medium and Low Time Spent Visits (Interest Categories)		✓	✓	
Ratio of New to Returning Visitors		✓	✓	
Search Results to Site Exits Ratio				✓
Search to Purchase Conversion Rate	✓			
Subscription Conversion Rate				✓

Table 3.3 : KPIs per business Type (adapted from Peterson(2006)

Peterson (2006) in his “Big Book of KPI’s” provides a set of Key Performance Indicators for Online Businesses. Key Performance Indicators are types of metrics that are often used to measure performance of several business processes or indicate whether business goals set by an organization are achieved (Chaffey et al. 2009). These KPIs are also used to measure the performance of the objectives related to customer Acquisition, Development and Retention. According to Peterson there are 4 main online business types: Online Retailers, Content Site, Customer Support and Business/ Marketing Sites. **Online Retailers** sell products and services via their website; **Business/Marketing Sites** often provide information about their products and services, without selling them directly through their website; **Content or Advertising Sites** focus on online advertising and publishing content. **Customer Support Sites** provide online customer support. Table 3.2 illustrates the KPIs and the business types for which the KPIS are most suitable. As it can be seen, the majority of KPIs are applicable to the online retailers. Each of the KPIs is related to a specific goal that differs per business model. For instance an online retailer focus on increasing online sales; a business site focus more on creating interest about their product or service; a customer support site focuses more in increasing customer satisfaction and improves; a content site focuses on raising customer loyalty.

To sum up, the objectives of online marketers are strictly related with the phases of customer lifecycle; Acquisition, Retention and Development, as well as with the engagement of the Online Customers. For a successful online marketing strategy it is essential to acquire and attract new online customers, to improve churn rates and customer satisfaction and loyalty in order to achieve customer retention, to increase sales and improve conversion rates in order to reach the stage of customer development (Table 3.4). All these objectives can be measured according to specific KPIs.

Customer Lifecycle	Business goals
Acquisition	▪ Attract new online customers.
Retention	▪ Improve churn rates ▪ Increase satisfaction ▪ Increase loyalty
Development	▪ Increase sales (cross-up sales) ▪ Increase conversion rates
General goal	▪ Increase revenues.

Table 3.4: Summary of main high-level online marketing objectives

For the purpose of this research explorative interviews were conducted, which are presented on Chapter 5. The interviews are complementary to the literature review. The aim was to have the experts' opinion on the topic and to identify specific high level objectives that online marketers who also make use of Online Customer Engagement Management tools have.

3.2. Customer Segmentation

In the following paragraphs the term "Customer Segmentation" is explained and customer segmentation types based on attributes gathered from offline channels and sources are presented and analyzed. The main segmentation types discussed are the following: Attitudinal, Behavioral, Demographics, Loyalty-based and Value based. Moreover, other overlapping segmentation types found in the literature are presented. Additionally, data found during the literature review regarding online customer segmentation are presented.

3.2.1. What is Customer Segmentation?

Customer Segmentation is the process of categorizing customers with heterogeneous characteristics into distinct, homogeneous groups based on common attributes (Hong and Kim, 2012). Customer segmentation is vital for (online) marketing and customer engagement constitutes a part of Customer Relationship Management. It is considered as an effective method for developing differentiated marketing strategies based on customer characteristics. As stated by Chen et al. (2007), effective customer segmentation contributes in raising not only customer satisfaction, but also the expected profits of the organization. Moreover, it assists in maximizing customer value, promotes customer loyalty, and facilitates the CRM processes; customer acquisition, retention and development Hence, customer segmentation is a field of considerable interest for marketers and researchers.

It is often the case that the definition of *Customer Segmentation* is considered the same as that of *Market Segmentation*. However, there is a difference between the two concepts that should be underlined. As defined by Stroud (2006), *market segmentation* "is the process of identifying customers who comprise a homogeneous group of consumers for a specific range of goods and services". On the other hand customer segmentation takes place within a certain and defined market; it is the grouping of the customers within a market, which has already been defined.

Customer segmentation is performed based on customer data, according to which customers can be grouped. Such data can be collected from various sources, such as organizations' databases or invoices, and are most of the times unstructured and continuous. Thus, they should be analyzed and utilized to perform effective customer segmentation. When customer segmentation is conducted with the use of data mining can assist in identifying profitable and loyal customers (Tsiptsis & Chorianopoulos, 2009). There are few empirical researches available, proposing algorithms and approaches for customer segmentation. For very large databases, the application of big data tools and techniques would be advisable.

The main goal of a marketer that performs customer segmentation is to develop effective marketing strategies in order to achieve various marketing objectives. More specific, Online and interactive marketers are interested in knowing their customers better and being engaged with them, by implementing suitable marketing strategies. Therefore, the wise and effective segmentation of their online customers is very important for achieving their online marketing targets and further engage with their customers.

3.2.2. Customer Segmentation types

There is not any one-size-fits-all segmentation scheme that is able to assist effectively all the business needs and goals. There are several segmentation approaches available, each of them suitable for supporting different business needs (Tsiptsis & Chorianopoulos. 2010). Stemming from the results of the literature review on customer segmentation, a set of segmentation types is presented and explained below. Since there is not enough literature on online customer segmentation, these segmentation types are basically based on data that can be collected from traditional CRM systems. Each type regards specific criteria and customer attributes according to which distinct customer segments can be identified. As it is explained below, some segmentation categories overlap each other. This means that either some of the authors refer to the same set of customer attributes- that form a certain segmentation type- with a different naming, or that they describe sub-sets of segmentation types that are included in a wider segmentation category.

Main Segmentation Categories

The segmentation types described below, are the most well-known and are commonly found in papers or other relevant studies.

Value-based segmentation

Value-based is a segmentation process, through which customers are categorized according to their value (Chan 2008; Kim et al. 2006; Hosseni&Tarokh, 2011; Tsiptis & Chorianopoulos, 2009). Customer value has been defined by many researchers and is also known as customer equity or customer profitability (Hosseni & Tarokh, 2011). Moreover, there are several models for estimating customer value. In order to perform value-based segmentation, a certain procedure for determining the value of the customer should be developed. The term "value" does not necessarily entails only the monetary profits derived from the customer. Kumar et al. (2010) proposed 4 value components that constitute the Customer's Engagement Value: Customer Life Time Value (CLV) determined by customer's purchase behavior; Customer Referral value (CRV), that related to the acquisition of new customers; Customers Influencer Value (CIV) regarding customers' behavior to influence other customers; Customer Knowledge Value (CKV) regarding the value added to the firm by receiving customers' feedback.

Behavioral segmentation

Behavioral Segmentation is a process, through which customers are grouped according to usage, attitude and behavior regarding a product or promotion (Stroud, 2006). In a traditional CRM system that integrates data only from offline channels, these customer data can be retrieved from firms' data warehouse (Tsiptsis & Chorianopoulos, 2009). Customer attributes that can be collected for behavioral segmentation contain: product ownership, type of transactions and frequency of transactions, revenue history, payments, and product utilization.

Loyalty or Engagement Segmentation

Loyalty or Engagement Segmentation is used to determine different groupings of customers according to different degrees or loyalty to supplier or brand (Stroud, 2006). The segments for loyalty or engagement segmentation are often created by applying simple business rules and/or cluster models on survey or database information (Tsiptsis & Chorianopoulos, 2009). Such segmentation contributes into identifying the less loyal customers, so as retention actions can be implemented in order to raise customers' loyalty. Attributes related to customer loyalty that can be: Engagement/Loyalty Score, Frequency of purchases, Number of Complains, new or old customer.

Socio-demographics and Life Stage

Through *Socio-demographics and Life Stage* segmentation customers are grouped according to their demographic or social characteristics. This segmentation type is one of the most commonly used, as the characteristics it includes can influence the changes in customers' needs, attitudes preferences and usage behaviors. At this point, it should be underlined though, that it is quite common that customers with the similar demographic and social characteristics have different needs and preferences (Tsiptsis & Chorianopoulos, 2009). Customer attributes that can be tracked and used for socio-demographic segmentation are: Age, gender, income, ethnicity, marital status, education other personal details of the consumer.

Needs or attitudinal-based

Needs or attitudinal based segmentation is performed in order to explore customers' needs that can be fulfilled by the purchase of a product or service, wants, views, attitudes, and preferences (Tsiptis &Chorianopoulos, 2009). This segmentation type can be useful for new product development, since clients preferences that are revealed can be used as an input for the design of new product requirements. Moreover, it can be used for enhancing the effective communication of the brand image and the key features of products or services. In an offline environment the data needed for this kind of segmentation can be basically gathered from external sources like market surveys, through which customers are able to express their opinions and preferences.

Overlapping Categories

The segmentation types enlisted below are also found in relevant studies, but they overlap or complement the aforementioned segmentation categories**.** This means that either the same set of customer attributes - that form a certain segmentation type- is mentioned in some studies with a different naming or that the particular segmentation category includes attributes that constitute a subset of the attributes included in one of the main segmentation categories described above.

RFM segmentation

RFM segmentation is the segmentation of the customers according to Recency, Frequency, and Monetary value of the customer, which actually characterize customers' purchasing behavior (Chou & Wu, 2010; Chan, 2005; Chan, 2008; Tsiptsis & Chorianopoulos, 2009), and are also considered KPIS. RFM segmentation is mostly used by retail enterprises. The Recency measurement indicates how long it has been since the last transaction of the customer. Frequency shows how often the customer purchases as well as the number of transactions during a specific period. Finally, monetary value is used to indicate the total value of purchases during a predefined period (Chan, 2008). Customer data needed for RFM segmentation can be easily collected during the process of purchasing. At this point, it should be underlined that since Recency, Frequency and Monetary indicate customers' purchasing behavior, RFM segmentation overlaps with behavioral segmentation (Recency and Frequency) and also with value-based (Monetary).

Propensity-based segmentation

Propensity-based segmentation (Tsiptsis & Chorianopoulos, 2009) is performed to categorize customers according to propensity scores, which are calculated with the use of certain data mining models. Propensity scores indicate the likelihood of a certain event to happen. Such scores can be churn score, which show the likelihood of a customer to churn, cross-selling and up-selling score or profitability score (Stroud 2006, Tsiptsis & Chorianopoulos 2009). Normally, segmentation based on propensity scores is combined with other types of segmentation, for instance value-based segmentation, in order to assist marketing objectives. For this type of segmentation, customers are grouped according to calculated scores and not on raw data that can be directly withdrawn from databases without prior analysis. Propensity scores are calculated with the application of classification modeling and signify the likelihood of an event to occur. For example churn likelihood or the likelihood to buy a complementary product.

Benefits segmentation

Benefits segmentation is used in order to categorize customers based on their expectations from a product they want to buy or the benefits they are looking for when they want to buy a product. According to Stroud (2006), benefits segmentation is the only segmentation type that takes into account customer's perspective, since it deals with the motivations that lead customers to purchase a product. Such motivations can be product value, usability, availability and brand value. This segmentation type also overlaps with attitudinal/needs segmentation explained above. Data for this kind of segmentation is difficult to be directly tracked. However, through market surveys or by observing and analyzing consumers' behavior, relevant customer information could be gathered.

Interaction Segmentation

Interaction Segmentation is used to group customers according to their preferences regarding channels, payment methods, promotions and communications (Stroud, 2006). Data for this type of segmentation can be gathered easier from online sources.

Lifecycle Segmentation

Lifecycle segmentation is a type of segmentation which aims in categorizing the customers based on the change of their need at several stages of their lives (Stroud, 2006). This type of segmentation overlaps with the socio-demographics segmentation type. Some common attributes that are included in lifestyle segmentation are: marital status, age, employment etc.

Psychographic segmentation

Psychographic Segmentation is performed to group customers according to different degrees of lifestyle, social behavior and personality (Stroud, 2006). This segmentation type involves characteristics lifestyle preferences, interests, values and lifestyles. Ii is also referred as lifestyle segmentation (Miguéis, Camanho & Cunha, 2012). It overlaps with attitudinal/needs segmentation category.

Usage Segmentation

Usage Segmentation is a segmentation type appropriate for categorizing customers according to the type of usage of product and service. It overlaps behavioral segmentation which includes product utilization (Stroud, 2006).

Geographic Segmentation

Geographic Segmentation groups customers based on geographic factors. This category also overlaps with the demographic segmentation (Stroud, 2006).

Occasion segmentation

Occasion Segmentation is performed to group customers according to their consumption of a product or service in certain situations or events or during specific periods (Stroud, 2006). For instance, it has been observed that more beer is purchased during big football tournaments. It overlaps the attitudinal/needs segmentation category.

Table 3.4 summarizes the segmentation types and the relevant papers into which they are described pointing out the customer attributes that are included to each category.

Segmentation type	Overlapping segmentation types	Relevant Customer Attributes	Related Studies
Attitudinal	▪ Benefits ▪ Interaction ▪ Occasion ▪ Psychographic ▪ Usage	Preferences, Interests, Usage occasion, Customer needs, Motivations, Usage situation, lifestyles, personality, necessities, favorite channels, payment methods, preferable promotions, type of communication	Miguéis et al..2012; Stroud ,2006; Tsiptsis & Chorianopoulos, 2009,
Behavioral	▪ RFM	Frequency of transactions, revenue history, payments, product ownership, product utilization, last time of Purchase, Frequency of purchase	Tsiptsis & Chorianopoulos, 2009, Stroud 2006, Leung 2009, Chou & Wu, 2010; Chan, 2005; Chan, 2008;
Demographic	▪ Geographic ▪ Lifecycle	Age, gender, income, ethnicity, marital status, education, country region Marital status, age, employment, number of children	Leung 2009, Stroud ,2006; Tsiptsi s& Chorianopoulos, 2009,
Loyalty-based	▪ Propensity scores	Frequency of purchases, Number of complains, Engagement /Loyalty score, New/old Customer, Propensity scores for loyalty	Tsiptsis & Chorianopoulos, 2009,
Value-Based	▪ RFM ▪ Propensity scores	Customer Value, Monetary Value, Propensity scores for profitability	Chan 2008; Kim et al. 2006; Hosseni & Tarokh, 2011; Tsiptsis & Chorianopoulos, 2009, Kumar et al, 2010.

Table 3.4: Segmentation Types

3.2.3. Online Customer Segmentation

The results of the literature review regarding customer segmentation revealed that while there is theory on Customer Segmentation there is not enough literature that sheds light on the segmentation of online customers and visitors. Although, one would think that there are not any significant differences among “online” and “offline” customers, the augmenting interest in segmenting online customers indicates the opposite. The main difference regards the diversity of characteristics gathered from online channels, indicating online customer’s behavior. As stated by Baranov (2012) customer attributes that can be tracked online allow companies to learn more about individual customers facilitating a direct approach of the customer. Thus online segmentation plays an important role in having a clear view of the needs and desires of the customer.

An interesting insight on how online customer segmentation is related with online marketing goals is given by Peterson, highlighting the importance of an effective segmentation of online customers (2006). According to Peterson, important KPIs (Table 3.3) are based on certain customer attributes that are gathered from online channels. KPIs can be built after a meaningful segmentation of the online customers, which will distill customer behaviors and create focused groups that will enhance the decision making based on the data. Such attributes are cost per conversion, cost per visit, average order value, average revenue per visitor, number of visits, new visitor, returning visitor, subscription, click depth, time spent on each visit. Some of those attributes can be straight captured and added to visitors’ profiles, while others, such as average order value, require calculations.

The available literature on online customer segmentation focuses more on online buyers behavior in retailers website, and do not provide an overview of the customers data that can be tracked from website visits. Therefore, in order to further explore what characteristics can be gathered in order to create informative online customer profiles an explorative case study on the OCEM Tool was conducted and is presented on Chapter 5.

4. Big Data and Data Mining

The tremendous growth of the Internet during the past decades has resulted in the great explosion of the data, influencing the relationship of the organizations with their customers. As a matter of fact, organizations nowadays are provided with the capability to generate huge amounts of customer data, which reveal their characteristics, their needs, their attitudes and their preferences. Hence, there is an increasing need for using new approaches and tools, and implement new strategies in order to handle and utilize the unstructured data and transform them into useful and actionable insights.

This chapter provides an introduction into the notion of Big Data. Moreover, the main available techniques and tools that exist are presented, and it is discussed which of these techniques can be applied to assist customer segmentation.

4.1. What is Big Data?

4.1.1. Big Data Definition

While organizations struggle to attain as many information as possible, the notion of big data is currently becoming a trend and it is a topic of discussion on many scientific forums. Moreover, there are already few techniques and software tools available in the market, able to analyze such a big amount of data. Since, the term itself is quite new currently there are not many scientific researches available on the field.

A number of researchers and analysts have attempted to define "Big Data". However, there is not a standard scientific definition of the term yet. Chen et al. (2012), in their paper, describe big data as *"Data sets and techniques in applications that are so large and complex that they require advanced and unique data storage, management, and analysis and visualization technologies."* Another definition provided by Forrester (2012) refers to big data as "techniques and technologies that make data handling at extreme scale affordable". However, most analysts and researchers that attempt to define big data focus on the most significant attributes or else the so-called V's of big data, which actually characterize the nature of the data (Gartner[1], 2011; Russom, 2011; Forrester, 2012; O'Reilly, 2012; Sathi, 2012). An example of such a definition is the one given by Gartner[1], which describes big data *as "high volume, high velocity, and/or high variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making".* Figure 4.1 illustrates the three most commonly mentioned attributes of big data; Volume, Velocity and Variety.

Volume

Volume refers to the amount of data that are created and can be processed by big data techniques. The volume of data which are created is rapidly growing, since today petabytes of data are created in daily basis (Sathi, 2012) Handling the huge volume of data constitutes a challenge for IT since it increases the need for scalable and dynamic storage and new approaches to querying (O'Reilly, 2012).

Velocity

Velocity regards the frequency at which data are generated as well as the latency of the data (O'Reilly, 2012; Zaslavsky, Perera & Georgakopoulos, 2012; Russom, 2012). The augmenting rate of data generation requires massive parallel process and bigger space in order these data to be analyzed (Sathi, 2012).

Variety

Variety refers to the different types and sources of data. In big data systems, data are generated from various sources and can be found in diverse types. Such data might be images or text from social networks or mobile devices, web logs, streamed video or audio.

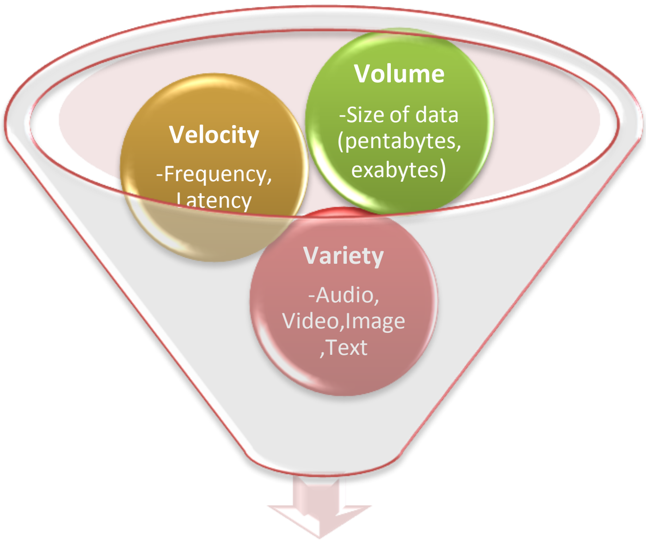

Figure 4.1: Attributes of Big Data (Gartner[1], 2011; Russom, 2011; Forrester, 2012; O'Reilly, 2012; Sathi, 2012)

Based on the above description of Big Data, in this research we adopted the following definition of Big data as :

> *"Data characterized by high-volume, high-velocity, and high-variety, which require advanced and exclusive tools and techniques for information management, processing and storage, analysis and visualization".*

While Big Data are becoming more and more popular, its impact on several aspects of business life and scientific fields has started being obvious. According to Chen et al. (2012) some of the main fields of applications of Big Data analytics are the following (see also Appendix):

- **E-commerce and Market Intelligence:** Due to the explosion of online business and the great use of social media, the amount of information that needs to be handling and analyzed has necessitated the use of big data tools and technologies. Big data can assist

[1] http://www.gartner.com/it-glossary/big-data/

online marketing by applying techniques for highly targeted searches ,personalized product recommendations, text analysis etc. (Chen et al., 2012).

- **Science and Technology**: The fields of science and technology are already benefiting from big data analytics. Big Data applications are already implemented in the fields of Astronomy, biology, genomics and environmental research (Chen et al., 2012; Olleiry, 2012).
- **Health Care System and Wellbeing:** The health sector also faces an onslaught of health related information**.** The fields of genomics and patient electronic health records can reap the advantages that big data analytics offer, in order to extract valuable knowledge and effectively manage the vast amount of patient information (Chen et al., 2012).
- **Security and Public Safety:** The application of big data in crime analysis, criminology, terrorism informatics as well as open source intelligence is considered valuable (Chen et al., 2012).

4.2. Big Data Techniques, Tools and Technologies

4.2.1. Big Data Techniques

As the volume of data increases rapidly, so does the need for using suitable algorithms, techniques or tools capable of analyzing this data and turn them into useful and usable insights for the organizations. Hence, many algorithms, techniques and tools that are able to utilize, analyze and visualize big data have been developed and already adopted by organizations. Such tools and techniques derive from several scientific fields, such as statistics, computer science, economics etc. Many of the Big Data tools and techniques stem from traditional data mining techniques and algorithms that were initially used for limited volume and diversity of data. Those tools and techniques were afterwards adapted and further developed so as to be effectively applied in very large databases which contain wide variety of data (Manyika et al., 2011).

Although there is not enough scientific literature regarding techniques tools and technologies that can be applied on big data there are several reports and some scientific papers that enlist and present big data techniques and tools that are currently available. The most important and well known big data approaches are explained below and are summarized and illustrated in Table 4.1.

Firstly, four basic techniques that are intended for data mining are presented below. Maniyka et al. (2011) and Chen et al. (2012), in their papers refer to these techniques as applicable to Big Data.

Classification

Classification is a very well-known set of techniques used for data mining. It is used in order to classify certain data into distinct groups based on certain characteristics that are defined at the initial state. This means that they are basically used for prediction of future events and behavior. Classification is characterized as a supervised learning technique, as they use a training data set that contains the variable according to which data are categorized into groups, as well as the class attribute. Classification can be used to predict effectively behavior of customers that belong into specific segment where an objective outcome exists or a clear hypothesis can be formulated. For example, it can be used to predict customer churn rate or the buying decisions of the consumers, as well as for optimization of business processes such as scheduling or performance of investments (Mitra et al., 2002; Kim et al., 2006; Turban et al., 2010). Classification techniques can work effectively for very large data sets as well (Maniyka et al., 2011). Some tools commonly used for classification algorithms include: *Decision trees*, which are a flow-charts with a tree structure with nodes that conduct tests on input variables that split the data into smaller subsets, and leaf nodes assigning a class to each of the observations in the resulting segments. They classify data into a certain number of classes according to the value of the dependent variable and they are most appropriate for interval and categorical data (Turban et al., 2010); *Genetic Algorithms*, which are robust and flexible search methods widely used for processes like query optimization or template selections (Mitra et al., 2002); *Neural Networks* that are more effective for large number of variables which involve complex relationships among them. They cannot be trained for very large databases due to the long time needed for training (Turban et al., 2010); *Bayesian networks* which are graphical models showing the relationships between the attributes; *Support Vector Machines (SVM)*, which maps data in a space of high-dimensionality in order to easier separate the data according to certain target groups.

Regression

Regression belongs to the family of traditional statistical techniques, which are also used for data mining. Same as classification, regression techniques are unsupervised learning techniques and are used for prediction and forecasting. Regression is used for mapping each data element to a prediction variable, which is a real number and not a class label as the prediction variable in classification (Mitra et al., 2002). Regression analysis makes visible the change in the typical value of the dependent variable changes when one of the independent variables is varied while the other independent variables remain stable (Raza & Farooqi, 2011). Regression techniques are often used to predict economic values, sales volume, or to identify possible factors that might affect customer loyalty or customer satisfaction. Some popular algorithms for regression are: Linear Regression, in which dependent is continuous; Logistic Regression, in which the response variable is categorical and is mostly used for multi-class classification (Tsiptsis & Chorianopoulos, 2009; Turban et al., 2010).

Cluster analysis

Cluster analysis is a traditional statistical method that was also initially used for simple data mining. This method is used for classifying a collection of things into segments whose members have similar characteristics (Turban et al., 2010). Unlike Classification, Clustering belongs to the unsupervised learning techniques, which entail the modeling of data with the involvement of inputs without having a predefined output. All data inputs are treated similarly in order to obtain information for the determination of groups or associations. Therefore, in clustering the characteristics according to which the objects are categorized into segments or else classes are initially unknown (Tsiptsis & Chorianopoulos, 2009). The clusters are determined when relevant clustering algorithms are applied on the data set under investigation and the similarities in the characteristics of the objects are identified. Some popular algorithms for cluster analysis are: K-means which is a quite fast algorithm applicable in large and wide datasets that requires predetermination of the number of clusters by the user; TwoStep which processes the records in two sets and can automatically determine clusters; Kohonen network/Self Organizing Map (SOM), which is a unique neural network architecture that produces a two-dimensional map of the clusters, and is slower than TwoStep and K-means (Turban et al., 2009; Tsiptsis & Chorianopoulos, 2009). Clustering techniques are often used for customer segmentation. This technique can be applied to huge datasets. There are already software tools in the market able to perform several clustering algorithms (Manyika et al., 2011; Chen et al., 2012).

Association

Association or else Association rule learning constitutes a set of algorithms used for data mining and they belong to unsupervised learning techniques. Such techniques are applied in large databases in order to analyze and discover association rules. This means that they are capable of uncovering interesting relationships among the variables of the database (Mitra et al., 2002, Turban et al., 2010). Two popular kinds of association rule mining are link analysis and sequence mining. With the former the connection among many objects is automatically detected, while with the latter the relationship among the objects is explored according to the order that they occur (identification of relationships over time). It is commonly used in retailing industry for market/basket analysis and description of consumer behavior. More specific is used in finding associations among products that a certain client of a supermarket has bought, which can be recorded by the so called point-of sale or checkout system. An association rule algorithm can be well-fitted in a very large database, since the number of distinct categories present in the data is not required to be specified to move on with the determination of association rules. The most popular algorithm for association rule learning is Apriori, which uses a bottom up approach used for identifying frequent item sets.

Visualization

Visualization techniques are used for creating images, diagrams, or animations, which intend, in combination with other techniques, to provide a better understanding of underlying relationships (Turban et al., 2011). The goal of visualization techniques is to illustrate data and information in a way that they can be more understandable and, thus, decision making can be enhanced (Woo, Bae & Park, 2005). They are used mostly in the field of health and natural sciences. Furthermore, they are often useful for marketers and decision makers, who wish to target profitable and loyal customers (Woo et al., 2005). Currently, there are several techniques and technologies for visualization, which can process huge amounts of numerical or text data. Some well-known techniques that are suitable for big data analysis are: Clustergram, historical flow, TagCloud, and Spatial Information Flow (Manyika et al., 2011). Examples of visualization are illustrated in Figure 4.2.

Clustergram **TagCloud**

Figure 4.2: Visualization techniques examples (source: Manyika et al., 2011).

The rest of the techniques that follow basically stem from statistical theory and can also be associated with data mining. Chen et al. (2012) and Manyika et al. (2012) characterize the following techniques as Big Data techniques, since they can all be applicable in very large and diversified datasets.

A/B testing

As Manyika et al. (2011) mention in their research report for McKinsey institute A/B testing is a technique that stems from traditional statistical theory and although it was initially intended for smaller datasets, it can also be effectively applied to large number of data. A/B testing is also known as split testing or bucket testing. The technique aims in comparing a main control group with a variety of test groups in terms of a certain objective variable or metric, in order to test which group performs better and decide upon actions or changes that will improve that variable or metric. A/B testing is often used for web page optimization and online marketing (Shuckla et al., 2009). For instance, A/B testing could be used in order to decide which text or layout will raise the conversion on a certain Web site. This also means that the end-result of the test should be a measurable variable, such as the click rate or response rate. According, to Manyika et al. (2011) A/B testing can be used for big data as it enables the conduction and analysis of very large number of tests, in order to find significant differences between the control group and the test group. When it comes multiple variables that have to be tested then the method is called multivariate testing.

Network analysis

Network analysis constitutes a set of techniques, which are used to describe the connections and relationships among the nodes in a graph or network. The most well-known technique is social network analysis, where the nodes of the network are individual actors. With social networks analysis the relationships among the individuals of a certain network, -which can be a community or organization- are described and analyzed. It is often used to describe the flow of information and the communication within a company, and thus is useful for knowledge management. Moreover, network analysis approaches the analysis of domains such as online social network and email communications (Abraham, Hassanien & Snâaésel, 2010).

Time series analysis

Time series analysis is a set of techniques which has its origins in statistics and in signal detection and processing theory. It is used for the analysis of sequences of data points, to extract meaningful characteristics from the data (Manyika et al., 2011). It is also suitable for forecasting, by applying a model to predict future values of a certain times series based on past values of the same or other time series. Some common applications of time series analysis and forecasting are: the hourly value of a stock market index; forecasting sales figures in retail sector; in the medical sector can be used for estimating the number of patients diagnosed with a given condition every day or for the prediction of the number of people who will be diagnosed with an infectious disease.

Sentiment analysis

Sentiment Analysis is used for parsing textual sources in order to identify and elicit information**.** It combines the application of natural language processing (NLP), which are computer algorithms that analyze human language, with a bunch of analytic techniques (Manyika et al., 2011). The basic functions of sentiment analysis is the elicitation of the text the sentiment-text classification as positive neutral or negative, and the determination of the strength of the opinion (Pang & Lee, 2008). The big surge in social media activity constitutes sentiment analysis very important since it is often applied for mining and analyzing consumer opinion (Chen et al., 2012). Sentiment-analysis technologies are capable of analyzing large amount of unstructured textual data (Pang & Lee, 2008) and thus they are considered suitable for big data analysis.

	Characteristics of Techniques								
Big Data Techniques	Data illustration	Data Mining	Graph-based	Predictive	Statistics	Supervised learning	Text-based	Time-based	Unsupervised learning
A/B Testing					✓				
Associations		✓	✓						✓
Classification		✓		✓		✓			
Cluster analysis		✓							✓
Network Analysis			✓						
Regression		✓		✓	✓	✓	✓	✓	✓
Sentiment Analysis		✓		✓	✓	✓	✓	✓	✓
Time-series analysis		✓		✓	✓			✓	
Visualization	✓	✓							

✓ Includes Characteristic

Table 4.1: Summary of Big data techniques and related characteristics

4.2.2. Big Data Technologies & Tools

There are an augmenting number of tools and technologies used for big data analysis. Below the most well-known tools and technologies are presented. Such tools and technologies have been also adopted by big organizations, such as Oracle, Microsoft and IBM (Chen et al., 2012). Table 4.2 provides an overview of the tools and their main capabilities.

MapReduce

MapReduce was a revolutionary framework introduced by Google in 2004. It constitutes the first highly-scalable programming model and associated implementation for processing and generating huge amounts of data on large and highly distributed clusters (Dean& Ghemawat, 2004; Gupta, Gupta, & Mohania, 2012). It is used for creating massive scale data applications, and contains two main functions: map and reduce. In the map process, the input data are divided into smaller pieces and sub-problems (key and value), which are afterwards distributed to and processed by a worker node (Gupta et al., 2012). It basically takes an input pair and produces a set of intermediate key/value pairs. In the reduce function takes as an input the intermediate key/value pairs and merges them to create a smaller set of values (Dean & Gehamavat, 2004). A simple prototypical example of Mapreduce can be found in Appendix 10.2. MapReduce constituted the inspiration of the development of several technologies in the big data area. The most popular tool which implements MapReduce is Apache Hadoop (Agneeswaran et al., 2012).

Hadoop

Hadoop or else Apache Hadoop is the most popular open source framework for the distributed processing of huge amounts of data, regardless of its structure (O'Reilly, 2012). Hadoop, which has as its core MapReduce, has been a great incentive for the evolution of big data. It is suitable for the development of applications, which demand the parallel processing of vast amount of data, on large clusters of machines, the number of which can be increased or decreased when it is necessary (Gupta et al., 2012). Figure 4.1 illustrates the components of Apache Hadoop that are also described below:

Hadoop Distributed File System (HDFS): HDFS constitutes the data storage of Hadoop. It is a highly scalable storage, where the data can be automatically replicated across a number of nodes in order to achieve fault tolerance. This also decreases the need for backup. (Gupta et al. 2012, Oreilly 2012,). HDFS is capable of storing data of any type that might also be unstructured.

MapReduce: *Mapreduce* as explained before is also implemented by Hadoop, for distributed processing of massive data.

Apache Mahout: *Apache Mahout* is one of the most important tools provided by Hadoop framework as it constitutes a library of machine learning and data mining algorithms (Oreilly, 2012) which can be implemented in huge sets of data. The core algorithms implemented by Mahout are clustering (K-means/Fuzzy K-means/Mean-Shift), classification (Naïve Bayes, decision trees), and collaborative filtering for recommenders. As part of Hadoop, Mahout is scalable to very large data bases.

Apache Cassandra: Cassandra is an open source database management system capable of storing and managing huge amount of data, while providing high scalability and availability and fault tolerance. Apache Cassandra is a NoSQL distributed database. This means that it is a non-relational decentralized database allowing for unstructured data storage (Agneeswaran, 2012, Manyika et al., 2011).

Other tools included in Apache Hadoop environment are the following (as founded in the official website of Apache Hadoop[2]:

HBASE[2]: HBase is a non-relational dataset, used for fast read and write access.

Hive[2]: is a data warehouse system which allows for managing huge distributed databases. The language that Hive uses for data querying is SQL , a simple language similar to SQL.

Hcatalog[2]: HCatalog is a set of interfaces included in Hadoop, which provides independent access to metadata created by Hive, interoperability among Hadoop tools, and abstraction tables of data, so the users can be at any time aware of where and how their data are stored. Since March 2013, HCatalog merged with Hive.

Pig[2]: Pig is a programming tool provided by Hadoop. It is a high-level data-flow language which allows for parallel executions of complex tasks.

Oozie[2]: Apache Oozie is a scalable workflow scheduler system used to manage the schedule and support the tasks executed by the other Hadoop tools.

[2] http://hadoop.apache.org/

Zookeeper*[2]*: Zookeeper is a distributed centralized coordination service. It provides services to the distributed applications of Apache Hadoop that cannot be implemented by the applications themselves. Such services can be; storing configuration information, fixing bugs, naming etc.

Kafka*[2]*: Kafka is a replicated, scalable log service, which as a part of Apache Hadoop provides the functionality of a messaging system, which is capable of maintaining, reading and writing messages.

R Programming Language

R is an open source programming language used for programming with big data provided by GNU project (Manyika et al, 2011; Sun & Heller 2012). R is a language widely used for creating statistical computing and graphics. Currently R[3] provides "Programming with Big Data in R" (pbdR) which is a project able to utilize huge amount of data, by providing packages for analyzing big data focusing on large scaling computing clusters. R provides a variety of techniques such as clustering, classification, time series analysis, regression etc. It is quite often used for regression analysis on large databases (Manyika et al; 2011)

Apache Solr

Apache Solr[4] is an open source standalone search platform provided by Apache Lucene project. It is written in Java, and its capabilities entail near real time distributing indexing, dynamic clustering, database integration, full-text search, rich document handling and geospatial search. Apache Solr is capable of indexing billion lines of code on nearly real time.

Jaspersoft & Microstategy Visual Insight

Jaspersoft[5] & Microstrategy Visual Insight are two tools for big data visualization (Agneeswaran, 2012). Microstrategy Visual[6] Insight provided by Microstrategy is capable of exploring data contained in spreadsheets, databases as well as in Apache Hadoop and creating multiple visualizations. Similarly, Jaspersoft is a business intelligence tools capable of exploring and visualizing data from both relational databases and Big Data sources.

	Tool Characteristics						
Big Data Tools	**Data Analysis**	**Data Access**	**Data Management**	**Data Processing**	**Data Visualization**	**Data Storage**	**Programming Language**
Apache Hadoop	✓	✓	✓	✓	✓	✓	✓
HDFS						✓	
MapReduce		✓		✓			
Apache Mahout	✓						
Apache Cassandra						✓	
HBase						✓	
Hive(QL)		✓					
Pig					✓		✓

[3] http://www.r-project.org/
[4] http://lucene.apache.org/solr/
[5] https://www.jaspersoft.com
[6] http://www2.microstrategy.com/visual-insight/

Zookeper			✓				
R					✓		✓
Apache Solr		✓		✓			✓
Jaspersoft					✓		✓
Visual Insight					✓		✓
Oozie			✓				

✓ Includes Characteristic

Table 4.2: Big data tools and related characteristics

4.3. Big Data vs. Traditional Data Mining

As Big Data have started becoming a trend, the difference between Big Data techniques and tools and traditional data mining techniques has been a matter of concern. Although in many cases they are considered to be the same, they are too quite distinguishable terms. The key factor of their differentiation is the main characteristics of the data that each technique is able to handle and analyze. A deeper look into the definitions and description of the two terms reveals their main differences.

There are many definitions available for data mining. Fayyad, Piatetsky-Shapiro, & Smyth (1996) simply define data mining as the application of specific methods and algorithms for extracting patterns and information from data. Moreover, many researchers often relate data mining to the term Knowledge Discovery in Databases (KDD). A well-known definition for KDD provided again by Fayyad et al. is; “the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data”. Such data are stored in structured databases, which contain categorical ordinal or continuous variables (Turban, Sharda, Delen & King, 2010). From a technical point of view data mining is a process, which is conducted in order to extract and identify useful information and uncover relevant knowledge or patterns from large sets of data (Shaw, Subramaniam, Tan & Welge, 2001). To do so a set of statistical mathematical and artificial intelligence techniques are used.

It is apparent that traditional data mining techniques and tools are basically intended for structured relational databases of a certain size. On the other hand, as mentioned above big data are characterized by large volume, quick rate of data generation and diversification in terms of types and sources of data. Moreover, these data can be found in any form; structured, semi-structured or unstructured. Therefore, they cannot be easily handled through Relational Database Management Systems (RDBMS). As stated by Chen et al. (2012) unstructured or semi-structured and diverse data require ad-hoc extraction processes, indexing and analytics within scalable and distributed environments, which is quite different than the ones that traditional data mining entails.

Table 4.3 summarizes differences and similarities of traditional data and big data. The key difference between “big data” and “traditional data” lies in the characteristics of the data and the different tools and technologies that these characteristics call for. As it is described in paragraph 4.2., techniques that are used for big data analysis are the same as those for traditional data mining techniques. However, unlike the tools and technologies intended for traditional data, tools and technologies used for big data are able to analyze larger amount of unstructured data, which stem from several sources and are found in various types. As they are also capable of handling and analyzing data that are characterized by high velocity, they can easily follow the flow of information and actually do the analysis in real-time. As big data constitutes a new trend, techniques and technologies for big data analytics that are currently available are continuously improving, while new ones are being developed.

	Traditional Data Mining	Big Data Analytics
Volume	▪ Gigabytes, terabytes	▪ Terabytes, petabytes, exabytes
Variety	▪ Structured data	▪ Semi-structured, unstructured data
Velocity	▪ Centralized databases	▪ Distributed databases
	▪ Data Mining Techniques ▪ Statistics	

Table 4.3: Traditional Data vs. Big Data

4.4. Tools and algorithms suitable for Customer Segmentation

Since the volume of customer data is rapidly growing the issue and customer segments are at a continuous motion, the issue of how effective customer segmentation can be achieved out of the huge amount of data has been raised. However, after the literature research conducted on big data analytics field for the purpose of this research, it is apparent that there is not enough scientific literature that specifies how big data tools can be used in order to assist effective customer segmentation. Nonetheless, there are enough scientific papers, reports and books, which analyze how several data mining algorithms and approaches can be used for customer segmentation. As it was explained in paragraph 4.3, the main data mining techniques for analyzing big data are the same as those for traditional data; only the tools that implement those techniques in big data sets are different. Therefore, techniques and algorithms intended for traditional customer segmentation can also be suitable for large amount of data.

In order to understand, which techniques are most appropriate for customer segmentation a focused literature review was conducted (See Chapter 2 for details). The purpose was to find relevant studies that present and analyze algorithms and techniques appropriate for certain segmentation types. The literature review resulted in 15 studies on customer segmentation using data mining algorithms, as they are illustrated in Table 4.2. Out of 15 studies, 13 are journal articles, one study is a published master thesis research on customer segmentation, and one study is a book. All of the studies present methods for certain customer segmentation types. Moreover, all of the studies contain case studies or implementations of the proposed algorithms, in order to validate and evaluate their result. It should be underline that the book of Tsiptsis & Chorianopoulos (2009) contained several algorithms some of which were also implemented through case studies and thus, they are illustrated separately on Table 4.2. Therefore, in total 19 algorithms or techniques are taken into account. The most of the algorithms found match with the Big Data techniques and approaches, which were presented in Chapter 4.1. This means that such algorithms can be implemented in very large databases with diversified data. A table including more information can be found in Appendix 10.5.

Method	Description	Segmentation types					References
		Attitudinal	Behavioral	Demographic	Loyalty based	Value-based	
Clustering	Soft clustering approach	X	X	X			Wu & Chou, 2011
	K-means		X	X			Chen et al., 2007
	Cluster Analysis	X	X				Miguéis et al.,2012
	ANN/SOM					X	Chan, 2005
	Variety of clustering algorithms		X				Jansen, 2007
	K-means		X	X		X	Ye, Yijun &Zhu, 2013
	K-means &SOM	X					Hong & Kin, 2012
	K-means		X				Cheng & Chen, 2009
	Clustering model	X	X				Tsiptsis & Chorianopoulos, 2009
	Automatic clustering algorithms		X	X			Tsiptsis & Chorianopoulos, 2009
	Clustering model	X					Tsiptsis & Chorianopoulos, 2009
	SOM		X	X			Lee & Park, 2005
Association	Association Rules mining	X	X				Pillai & Vyas, 2012
Classification	Genetic Algorithm				X		Chan, 2008
	Decision tree		X			X	Kim et al., 2006
	Decision tree				X		Tsiptsis & Chorianopoulos, 2009
	Decision tree				X		Han, Lu &Leung, 2012
Visualization	Customer map for customer targeting	X	X			X	Woo et al., 2005
Regression	Binary Logistic Regression				X	X	Hosseni&Tarokh, 2009

Table 4.2: Algorithms for Customer Segmentation

As is can be seen on Table 4.2 the techniques used for Customer Segmentation found in the studies are 5 data mining techniques; Clustering, Classification, Association, Regression and Visualizations. 58% of the studies(11 out of 19) present Clustering algorithms for Customer Segmentation, 21% used Classification, 10% Association, and 5% Visualizations and Regression. Apparently, clustering and classification are the techniques, which are used in most cases for segmenting customers. Each of the techniques described in the papers is used for several types of segmentations. Although, with the first glance someone might think that all techniques can be used for all segmentation types, there are some trends that can be observed

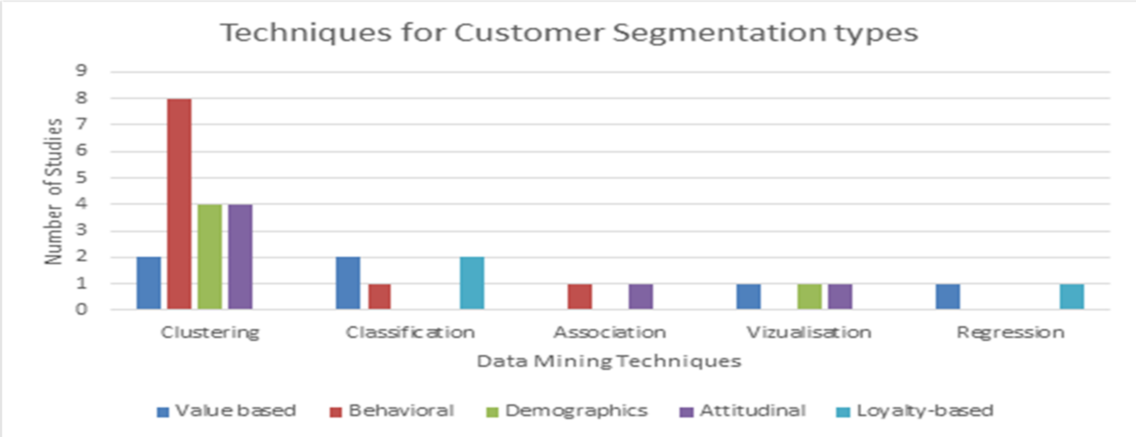

Figure 4.2: Trends in customer segmentation types

Table 4.3 shows the number of studies used a certain kind of technique for segmenting customers according to a specific customer segmentation type or a combination of types. Figure 4.2 illustrates the trends that result from table 4.3. As far as clustering techniques are concerned, 73% of the studies use clustering algorithms for behavioral segmentation, 36 % for demographics, 36% for attitudinal, and 18% for value based segmentation. On the other hand, 3 out of 4 (75%) studies use classification for loyalty-based segmentation, 2 out of 4 (50%) for value-based and only 1 out of 4 for behavioral (25%). These statistics indicate that clustering algorithms are mostly used for segmentation according to behavioral characteristics, demographics as well as their attitudes and needs, while classification algorithms are mostly used for segmenting customer according to characteristics that contain numerical values, like loyalty or customer value. In a few cases it can also be used for value-based segmentation. On the other hand, classification techniques are mostly used for value-based and loyalty segmentation and less for segmenting customers based on demographics or behavior.

As it is observed from the table the other three techniques are used as follows; Association rules are mostly used for behavioral and attitudinal segmentation; Visualization as it was found in a study is used for visualizing customer segments based on their value, demographics and needs and attitudes; Regression algorithms are used for value-based and behavioral segmentation.

Segmentation type	Number of Studies using the method				
	Association	Classification	Clustering	Regression	Visualization
Attitudinal	**1**	**0**	**4**	0	0
Behavioral	**9**	**1**	**1**	**0**	**0**
Demographics	**0**	**0**	**1**	**0**	**1**
Loyalty	**0**	**3**	**1**	**1**	**0**
Value-Based	**0**	**2**	**1**	**1**	**0**

Table 4.3: Number of Studies using methods for customer segmentation

Customer Segmentation with Big Data

All the aforementioned techniques can be implemented by big data tools. As described in Chapter 4.2 two open source tools capable of implementing data mining and statistics techniques in large and diversified databases are Apache Mahout and the programming language R. Both tools can be implemented on the top of Apache Hadoop. Mahout is capable of implementing clustering, classification and association analysis (recommender), while R for big data involves in its libraries algorithms for clustering, classification, association, and regression.

PART III: Empirical Data

5. Online Customer Segmentation – A case study

Based on the literature review regarding Online Marketing and Customer Segmentation, we came up with two basic conclusions. Firstly, although there is enough scientific literature on the field of Online Marketing, it is not easy to capture the main objectives that are related to online marketing and customer engagement. Secondly, the literature that is available for customer segmentation mostly regards customer attributes that can be gathered from offline sources such as organizational databases, CRM or ERP systems. Although there are a few studies on online customer segmentation, they mostly focus on customer segmentation according to transactional data obtained from web shops or online actions (Chan, 2005; Hong & Kim, 2010). Moreover, such studies do not provide a clear overview of the kind of data that can be gathered from online channels and the segmentation categories, which can identified according to those data.

In order to further explore, which are the common basic objectives that regard online marketing and identify which are the segmentation categories based on online customer attributes, the case of The OCEM Tool was examined. Explorative interviews with the marketers and business consultants of DEVCORP were conducted, while the customer data that The OCEM Tool is able to gather online were observed, for the identification of online customer segmentation types. This Chapter presents the results of the data gathered during the interviews and the observations regarding the product.

5.1. The OCEM Tool Case

The OCEM Tool is an Online Customer Engagement software product provided by DEVCORP, which was introduced in the market in 2010. The OCEM Tool was initially built for supporting online marketers in achieving their goals and increasing the engagement of online customers with the organizations. The main users of the OCEM Tool are online marketers. As depicted in Figure 5.1, the core functions of the OCEM Tool are: online customer profiling, creation of online dialogues with the customers and customer segmentation.

The most important function of the product is the creation of dynamic online customer profiles. The OCEM Tool is capable of creating customer profiles, by capturing customer attributes on real time from all online channels owned by an organization. Each time that a new visitor visits a website a profile is created and filled in with visitors' characteristics, behaviors and activities (customer attributes) during his visit. Every time that a visitor goes back to the website their profile is enriched with more information. So far the aggregated number of profiles that have been created by all the OCEM Tool users is larger than 140.000.000.

The OCEM Tool also performs customer segmentation based on customer attributes collected from online channels. The segmentation process is very important, since relevant dialogues and interactions, such as specific offers, content, banners or light boxes, are created according to the segments that reveal customer behaviors, preferences or other characteristics. However, the product does not integrate any analytics for customer segmentation, and thus the segmentation process is quite simple and still at an initial stage. More specific, the OCEM Tool enables its users to create segments of visitors who have common online behaviour or characteristics, which are defined as customer attributes.This means that the users of the product create customer segments themselves or with the guidance of the business consultants of DEVCORP, by selecting among the customer attributes that the OCEM Tool gathers. The user picks certain sets of attributes and sets criteria, in order to create appropriate customer segments that are able to serve their online marketing goals according to each specific company's case. However, the product does not provide any automated recommendation on which online customer segmentation types the user of the OCEM Tool should focus on when starting with a specific goal.

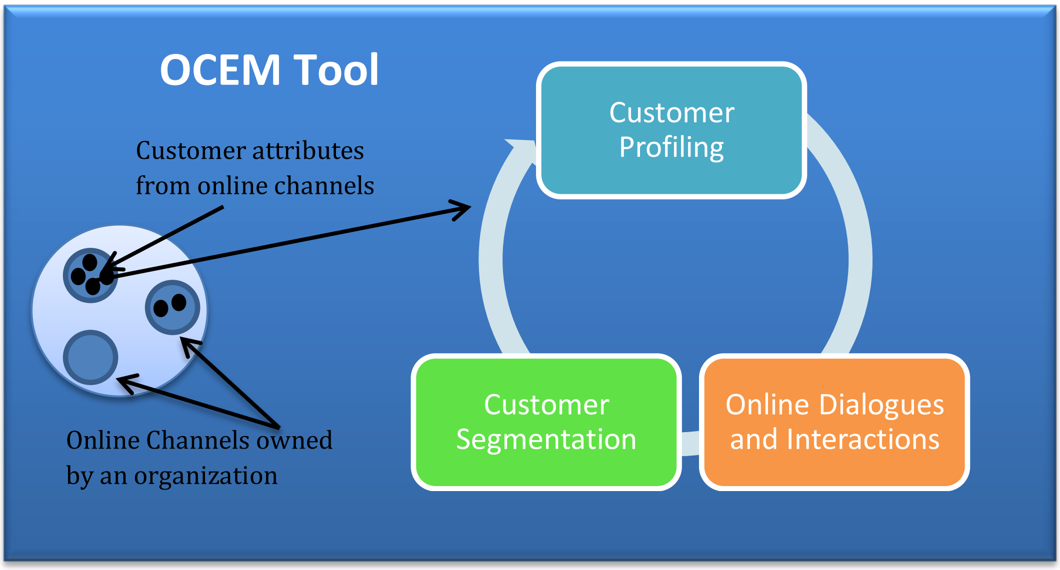

Figure 5.1: Main functions of the OCEM Tool

DEVCORP encounters so far more than 20 clients-companies, who have been signed up with the OCEM Tool and are already active in collecting millions of online customers' profiles. However, since DEVCORP is not only a software provider, but also a service provider, the OCEM Tool clients make use of the product with the guidance of the business consultants. Consultants receive inputs from the users' of the OCEM Tool, who are, in the most cases, online marketers. The inputs are related to the general goals regarding online marketing and customer engagement. However, since most of the businesses that use such products are at a low maturity level regarding their online strategies, consultants often start from focusing on smaller specific projects that serve a broader business objective. After receiving the inputs from their clients the consultants configure the product according to each client's needs and business goals. Moreover, they assist customers with utilizing the outcomes of the OCEM Tool and deciding upon customer segmentation. Customers segments are not automatically created by the product itself. In order to create segments, the user of the product has to decide, which customer attributes should be selected taking always into account the recommendations of the consultants. There is not any specific guideline for the selection of the segments. However, as one of the consultants claimed "*the users of the product would like to verify which attributes are most valuable to look for".*

5.2. Online Marketing objectives

Interviewees

In order to define a certain set of objectives that are important for online marketers, a set of explorative unstructured and semi-structured interviews were organized. The aim was to discuss about the objectives that are typically set by online marketers. For this reason, the basic criteria for the selection of suitable interviewees, was their field of expertise. The experts that were interviewed for the purpose of the research are depicted in the anonymized list provided in table 5.1. The interviewee IDs as shown in the table will be used to refer to each of the interviewee in the rest of the text.

Interviewee ID	Interviewee Job Title
M1	Chief Marketing Officer
M2	Product Marketer
B1	Business Consultant
B2	Business Consultant

Table 5.1: Table of Interviewees

At this point it should be underlined, that contacting online marketers who use the product was not considered effective. After discussing with the consultant, we realized that since the clients do not have well defined online marketing strategies and do not make use of the product themselves, a discussion with them would not add any further value to the research. The consultants were able to respond to the questions expressing their clients' points of view.

Structure of Interviews

There were two kinds of interviews, based on the expertise of the interview. The interviews with M1 and M2 were unstructured and the main question addressed to them was "*What do you consider basic online marketing objectives of an online marketer that would also use The OCEM Tool to achieve them?*" This question was addressed in order to trigger discussion on the topic of online marketing objectives and the interviewees were free to express their views and opinions. During the interviews goals, objectives and KPI's as found in the literature were mentioned and discussed. Moreover, the interviewees could also think and answer from the perspective of an online marketer that collects customer profiles by using a product like The OCEM Tool in order to achieve the objectives discussed.

The interviews with B1 and B2 were semi-structured and consisted again of open questions. The questions aiming in identifying common goals set by online marketers that use a product like the OCEM Tool as well as in discussing the segmentation process followed, and regarded the following:

- The role of the employees in an organization that make use of the OCEM Tool (if not only online marketers).
- The goals of online marketers that use the OCEM Tool.
- The goals that can be achieved by using the product
- The process of consulting the users of the OCEM Tool.
- The identification of common goals among the clients.

- The segmentation process that is followed (how they decide on the selection of the attributes, what can be achieved by segmentation).

Interviews outcome

The results of the interviews came to confirm the findings from the initial literature review made on the field of online marketing. The main goals are provided below. First of all, it is worth mentioning that B1 claimed "*only a few companies act according to the concept of online and digital marketing",* meaning that most of the organizations are currently building their online strategies and are still immature. However, as he also stated "*when there are clear business objectives it's much easier to interact with the customers*".

Almost all interviewees gave answers that lead to the same set of objectives that an online marketer that uses a product like the OCEM Tool wants to achieve. The objectives were high level. Both B1 & B2 said that their clients always start with very small goals they want to achieve according to the specific campaign or situation. However they all lead to the achievement of broader objectives. As it was mentioned by B1 "*the initial goal of the online customer is to publish content from their customer's point of view and not from the company's point of view. On the top of that there are several business goals that are related to it, such as the attraction of new customers, the increase of sales of relevant products in order to generate more profits*. "

The online marketing objectives mentioned during the interviews are enlisted below:

Online marketing goal: "Increase New Customer Acquisition".
All interviewees mentioned during the interviews that customer acquisition and attraction of new customers is among the main goals of online marketers.

Online marketing goal: "Increase Customer Satisfaction".
M1 mentioned that the increase of customer satisfaction is the main purpose and stated "*Customers are satisfied when they feel that they are treated as individuals. By focusing on data that reveal preferences and behaviors, customers can be provided with the right material and guided to the right channel increasing their satisfaction".* M2 and B1 also refer to customer satisfaction as a goal.

Online marketing goal: "Increase Customer Loyalty".
B2 mentioned that the initial goals are to create loyal customer and increase loyalty. M1 mentioned that is "*always important for marketers to measure Customer Loyalty".*

Online marketing goal: "Improve Churn rates".
Both B1 & B2 agreed that the increase of cross and up sales is a very important goal regarding online marketing. M2 referred to the reduction of customer churn as an essential goal set by online marketers.

Online marketing goal: "Increase cross- up sales".
Both B1 & B2 agreed that the increase of cross and up sales is a very important goal regarding online marketing.
Both M1 & M2 mentioned the increase of cross and up sales as an important goal.

Online marketing goal: "Increase conversion rates".
Three of the interviewees agreed that the improvement of conversion rates is the main objective of an online marketer. However, as M1 mentioned it stands at a lower level than the

other objectives, since the marketer defines what a conversion is, according to the business type. For example a conversion can be a purchase, a click on a banner, a contact request or the download of content. So, it can be claimed that the improvement of conversions is connected with the previously mentioned goal. However, as it was referred as an important business goal related to the online marketing it is included on the set of business goals used for the purpose of this research.

Table 5.2 summarizes the above objectives including a short explanation of the objective.

Main Online Marketing Objectives

- **Increase New Customer Acquisition**
  Description: Attract new online visitors
- **Increase Customer Satisfaction**
  Description: Keep online customers satisfied by providing right product and services
- **Increase Customer Loyalty**
  Description: Keep the customer engaged with the online brand.
- **Improve Churn Rates**
  Description: Prevent online visitors from abandoning the brand
- **Increase cross-up sales**
  Description: Sell additional or more expensive products
- **Increase conversion rates**
  Description: Increase visitors online activity

Table 5.2: Set of main Online Marketing Objectives

5.3. Online Customer Segmentation types

Traditionally, customer data can be gathered from several sources, but mostly from offline databases and invoices. This kind of customer data is mostly demographic and transactional data. Additionally, data related to consumer preferences or customer satisfaction can be gathered from market surveys. On the other hand, when it comes to the online environment things are more complex, because of the amount and the variety of customer attributes that can be collected from online channels. One of the key differences of online and offline channels is the great variety and velocity of the customer data that can be gathered. Apart from plain demographics or transactional customer data, data gathered from online channels can also reveal customer behavior and preferences. An OCEM software tool like The OCEM Tool is capable of tracking such customer attributes in real-time, which results in creating dynamic customer profiles. For this reason, the case of the OCEM Tool was examined, in order to find out which customer attributes can be tracked online by such a product, and what segmentation types exist online.

In order to find out which attributes are gathered online, documentation regarding the OCEM Tool was used, and real-customer data from the products' test environment were observed. The OCEM Tool is able to gather a standard set of online customer attributes, which are the same for all the organizations that currently use the product. Table 5.3 shows the list of the standard online attributes that can be gathered online, as it was found in internal documentation. As it can be seen the standard attributes are not categorized according to certain segmentation types.

Attribute Category	Online Customer Attributes
Statistics	• Average Visit , Time on Hosts, Entry Page, Interactions Clicked, Interactions Converted Interactions Viewed Page Views (all visits) Keywords, Page Views (current visit) ,Referrer Hostname Visited Channels Visited Sites Visits
Technical Info	• Flash Version, Java Version, Language, Operating System Name Operating System Version, Screen Resolution, Browser Name
Engagement Loyalty	• Engagement Interests • Engagement Score

Table 5.3: Set of main customer attributes gathered by the OCEM Tool (based on version 2.8)

As it can be seen from the table 5.3 many of the data obtained from online channels by using the OCEM Tool are similar to those gathered from offline channels. For example *visited sites* or *interactions clicked*, indicate preferences and are similar to the attributes of the attitudinal segmentation type, while *visited time* or the *number of visits* are characteristics that belong to behavioral segmentation type. The attributes engagement *interests* and *engagement score* can fall under the loyal-based categories. These attributes are calculated based on business rules set by the organizations that use the OCEM Tool. More specific, the *engagement score* is calculated based on rules that may regard for instance the number of clicks or the number of visits. An index is created and points are scored. An example would be the following: when a visitor comes to the website more than five times then he scores a point. The aggregated number of scores per visitor indicates their level of loyalty and engagement with the brand. Similarly, *engagement interest* indicates how likely is the online customer to be interested on a specific product or content. Apparently, the customer properties shown on Table 5.3 can be categorized into the segmentation types described in Chapter 3.

As we observed there are some certain attributes that cannot be categorized in any of the segmentation categories that already exist from the theory regarding offline systems. These are: Entry Page, Keywords, Referrer Hostname Flash Version, Java Version, Operating System name, Operating System Version. These are basically technical information or referral information. Although, from one point of view they could fall under the categories of preferences or demographics, it is essential that they constitute a separate segmentation category, since they give customer insights that allow the marketer to decide what content is more relevant for a certain website visitor, as well as how the content should be shown to the visitor of a website according to the technical characteristics of the device they use. Therefore, we define two segmentation categories based online customer characteristics that can be gathered only through online channels. These are defined as follows:

Referral Segmentation

Referral Segmentation is the way to segment customers according to referral attributes that are gathered online. Such characteristics show how the visitor ended up on a certain website, revealing their incentives and previous channels they visited. Customers can be segmented according to the source through which they entered a firm's website, or the keyword they typed. Attributes that are related to this segmentation category, which are also gathered by the OCEM Tool are: keywords, referrer hostname, entry page, search engine.

Technical Segmentation

Technical Segmentation is a category that includes technical characteristics according to each online visitor can be segmented. Such characteristics indicate the technical information of the device, operating system, or browser that an online customer uses when entering a website. The visitors can be segmented based on such attributes, so the marketers can decide on the best way that certain content can be illustrated on a visitor's screen. Attributes that are related to this segmentation category, which can also be gathered by the OCEM Tool, are: Flash Version, Java Version, Operating System name, Operating System Version, device, screen resolution, language, and browser name.

Table 5.4 shows how customer attributes that the OCEM Tool gathers from online channels (table 5.3), fall under the basic customer segmentation categories that are described in Chapter 3 (table 3.4) and the online segmentation categories that were previously defined. The segmentation types presented in table 5.4 are defined as Online Customer Segmentation types.

Online Customer Segmentation Types	Online Customer Attributes
Attitudinal	- Interactions Clicked, Interactions Viewed, Interactions Converted, Visited Channels, Visited Sites, Page Views,
Behavioral	- Number of Visits, Number of Page Views, Average visit time, time on host
Demographic	- Location (I.P address)
Loyalty-based	- Engagement Interests, Engagement Score
Referral	- Entry Page, Keywords, Referrer Hostname
Technical	- Flash Version, Java Version, Operating System name, Operating System Version, device, screen resolution, language, browser name

Table 5.4: Online Customer Segmentation types (based on standard attributes gathered by The OCEM Tool version 2.8)

The OCEM Tool already encounters more than 50 companies-clients who make use of it. 15 out of the 50 clients make full use of the product, and each one has created so far more than 2.000.000 customer profiles. As it was previously explained, the OCEM Tool is configured according to each client business type and needs. This means as well, that apart from the aforementioned standard customer attributes that can be collected, there are also several customer attributes being collected which differ according to the needs of the client that owns the OCEM Tool. For the purpose of this research the attributes that are gathered by the clients-companies with the most customer profiles were observed in order to find out which additional attributes can be obtained from online channels. Tables with examples of attributes per client-company can be found in Appendix 10.5.

Table 5.5 provides an overview of both standard and configurable attributes that the OCEM Tool can gather at the moment grouped according to the online customer segmentation they belong to. At this point it should be mentioned, that as it was observed no algorithms are applied in order to calculate values such as Engagement interest or values like purchasing value or order value. As explained before for *Engagement score* certain business rules are applied according to the needs of each client. The value-based attributes are actually based on the aggregation of the purchases or orders of the clients, for which values are applicable. So far the OCEM Tool cannot calculate customer's value based on models like LTV or RFM (as explained on paragraph 3.2.2), which well-known models for the calculation of customers' value are according to a set of monetary and behavioral characteristics. However, according to the opinion of the software engineer of the product, this would be feasible and could be applied on the product, in order to generate extra information for the value of the client.

Online Customer Segmentation Types	Online Customer Attributes
Attitudinal	- Interactions Clicked, Interactions Viewed, Interactions Converted - Visited Channels, Visited Sites, Page Views, - preferred social media, - clicked banners, offers viewed, - product type purchased, product type viewed
Behavioral	- Number of Visits, Number of Page Views, - Average visit time, - membership, returning client, web-shop visits, number of purchases - items purchased, number of orders - last visited time
Demographic	- Age, Gender, City, Job, - Twitter Id, Facebook Id
Loyalty-based	- Engagement Interests, Engagement Score
Referral	- Entry Page, Referrer Hostname - Keywords
Technical	- Flash Version, Java Version, - Operating System name, Operating System Version
Value-based	- Average Purchase value, Average Order Value

Table 5.5: Online Customer segmentation types (based on standard and configurable attributes gathered by the OCEM Tool Version 2.8).

Table 5.6 constitutes an overview of customer segmentation types that can be obtained from both offline and online sources, and the related attributes. The table highlights the differences among the attributes collected from offline and online sources.

	Customer Segmentation types	Attributes from offline sources	Attributes from online sources
Online & Offline	**Attitudinal**	Mostly based on market surveys: -Preferences -Interests, -Customer needs - Motivations -Usage occasion - lifestyles - personality - preferable promotions	-Interactions Clicked, -Interactions Viewed, -Interactions Converted, -Visited Channels, -Visited Sites, -Page Views, -Preferred social media -Clicked banners -Offers viewed -Product type purchased -Product type viewed
	Behavioral	Transactional Data: -Frequency of transactions -Revenue history - Payments -Product ownership - Product utilization -Last time of Purchase -Frequency of purchase	-Number of Visits -Last visit -Number of clicks -Number of Interactions -Subscriptions -Number of purchases
	Demographic	Customer Database: -Age, gender -Income, education, job -Ethnicity, Country region -Marital status, number of children	-Email Address -I.P (indicates location) -Age, Genre -Facebook Id, Twitter Id
	Loyalty-based	-Frequency of purchases, -Number of complaints, -Engagement Loyalty score, - Propensity scores for loyalty	-Engagement score based on business rules regarding customers' behavior
	Value-based	-Customer Value - Monetary Value - Propensity scores for profitability	-Average value of purchases -Average value of orders
Only Online	**Referral**	-----	- Flash Version, Java Version, -Operating System name, Operating System Version
	Technical	------	-Average Purchase value, Average Order Value

Table 5.6: Attributes gathered from offline and online sources.

5.3.1. Meta-data model for Online Customer Segmentation

In order to provide a structural view of the previously presented findings and show how they can be generalized for similar case-studies, hereby, we provide a meta-data model diagram, which is illustrated in Figure 5.2. The diagram is a concept diagram which constitutes the deliverable side of a PDD (Weerd & Brinkkemper, 2008). The model shows that customer segmentation types can be both offline and online; Offline customer segmentation types are based on attributes gathered from offline channels; online customer segmentation types are based on data gathered from online channels by OCEM or similar tools. Online Marketing objectives can be supported when online marketers focus on certain online customer segmentation types in order to segment their customers. As it can be seen the concepts related to customer segmentation types correspond to tables 5.6 and 5.5, which present both offline and online customer segmentation types. Moreover, the concept "Online Marketing Objectives" corresponds in table 5.1.

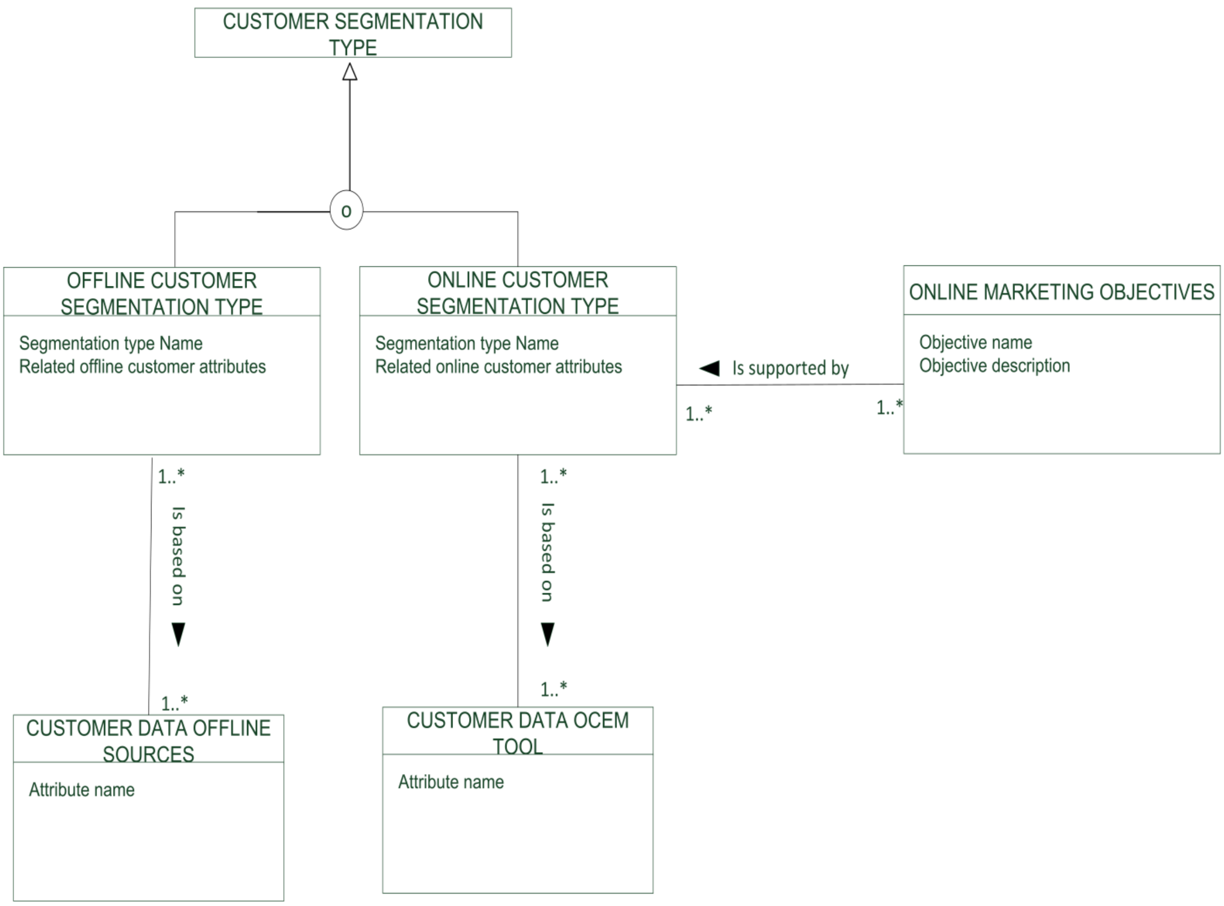

Figure 5.2: Meta-data model of case study findings

5.4. The OCEM Tool and Big Data

As previously mentioned the OCEM Tool is a software tool that does not implement any big data analytics, for the analysis of the online customer profiles it creates. However, this does not mean that the OCEM Tool does not generate big data itself. Data generated by the OCEM Tool involve the three main attributes that characterize big data. More precisely:

- **Volume:** While being two years in the market the OCEM Tool currently encounters so far more than 140.000.000 customer profiles as a total, which are stored in Cassandra database. Although the size of each profile is not big in terms of space (approx. 1 KB per profile) the number of customer profiles and customer properties is becoming in huge. It is worth mention that one of the companies-clients of DEVCORP, has created so far more than 50.000.000 online customer profiles.

- **Velocity** : Customer profiles are created dynamlically and in real-time. This means that the number of profiles is rapidly growing, while the scale on which the profiles are updated when new customer attributes are tracked is extremely large. For example, by a simple visit on a website more than 10 records have to be updated to each customers' profile.

- **Variety:** As it can be easilly understood , the OCEM Tool generates data of different types that are unstructured and semistructured. Web logs and text are the most usual type of data that the OCEM Tool is able to track so far from several online channels including social media, and in several cases from mobile devices.

The OCEM Tool integrates the use of some Big Data tools and technologies. Data generated by the OCEM Tool are stored in Apache Cassandra database, which also constitutes a component of Hadoop and is characterized as a database capable of storing big data. The main reason why Cassandra is used for storing the OCEM Tool data is the fact that is ensures high scalability. So far, Cassandra database is configured in order to fit up to one billion visitor's profile, consisting of a huge number of customer characteristics. Moreover, although the OCEM Tool does not use any tool for analyzing big data, it uses Apache Solr, which is a highly scalable search engine, able to work on big data. Apache Solr is used in the OCEM Tool for indexing and for simple calculations; regarding customer attributes (e.g. calculation of averages).

It is estimated that in the next few years the number of profiles, including a huge number of customer attributes, stored in the OCEM Tool's database will tremendously increase. Although the product is not considered an analytic tool, this increase will sooner or later call for data analysis in order to harness the huge amount of customer data into its full potential and provide a holistic customer approach through customer segmentation. As mentioned during the interviews the improvement of the segmentation process is a matter that is going to be taken into account for the next the OCEM Tool releases. The software architect of the product mentioned that the integration of a big data analytic tool such as Apache Mahout has already been taken into consideration. The integration of such a tool would improve the segmentation process, by providing customer data analysis and revealing relevant customer segments out of the huge amount of data. However, the integration of such a tool would require time, in order to get the tool to work in production scale and configure it according to performance limits. Moreover, the knowledge and the experience of a data analyst or customer intelligence expert might be needed, in order to properly integrate the algorithms that such a tool provides for effective online customer segmentation.

PART IV - Framework

6. Framework for Online Customer Segmentation

In this Chapter a framework for Online Customer Segmentation is proposed. As it is comprehensively described below the framework has as basic inputs the results that occurred from both literature study and empirical data gathering as they are presented in the Chapters 3 4 and 5. The framework consists of two parts. In the first part the set of the objectives related to Online Marketing and Customer Engagement, which was defined in Chapter 3 is matched to the segmentation types described on Chapter 4, which are able to assist each of the business goals. In the second part, the segmentation types are matched with relevant techniques and algorithms, as they were presented in Chapter 5 that can be applied in large data sets. The summary of the two parts is a table, which matches online marketing objectives with segmentation types and appropriate techniques and algorithms.

6.1. Matching Online marketing goals with segmentation types

As depicted by the scheme in Figure 6.1., the first part of the framework for Online Customer Segmentation that this research proposes entails the mapping of business objectives regarding online marketing with Online Customer Segmentation types, which can effectively assist each of those objectives. Online Customer Segmentation types are defined as certain categories of customer attributes that can be gathered from online channels by an OCEM software tool. Based on this segmentation types analysis can be made in order to define underlying online customer segments.

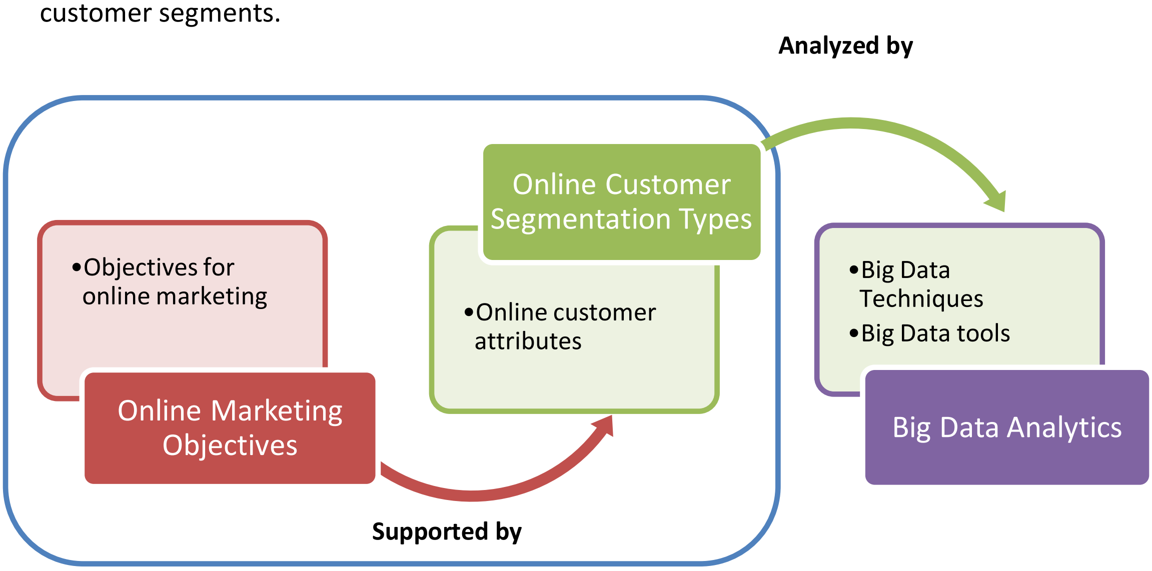

Figure 6.1: Online Customer Segmentation types (1st Step of the Framework)

The first part of the framework for Online Customer Segmentation is illustrated in the form of a matrix in Table 6.2. As it can be seen, the matrix shows which online marketing objectives can be assisted by online customer segmentation types. Since there is not one-size-fits all segmentation for each business objective (Tsiptsis &Chorianopoulos, 2009) none of the customer segmentation types can be excluded during customer segmentation. However, the framework suggests, which segmentation categories are more preferable and useful for assisting each of the related online marketing objectives. In order to indicate the usefulness of each segmentation type for each of the online marketing objectives a High-Medium-Low scale is used. More specific:

- **H:** Indicates that the segmentation type is highly useful for the corresponding objective, and therefore should be preferred.
- **M**: Indicates that the segmentation type is useful in some cases and complementary to the highly useful segmentation types.
- **L:** Indicates that the segmentation type is not that useful for the corresponding business goal, especially if it's used alone.

Table 6.2 shows 6 high level business goals that online marketers usually have. The objectives shown stem from the literature review as well as from the interviews conducted, as it was described in Chapters 3 and 5. Each of the goals can be assisted by certain segmentation types. Based again on the literature and on the discussions with the experts the relation between the business goals and the segmentation types are defined. More specific, in order to identify the level of usefulness of the segmentation types for each of the business goals the following factors were taken into account:

- Definition of the objective
- Measurements of the objective
- Usual actions related to the achievement of the objective
- Opinions of the interviewees regarding the objectives in relation with the customer attributes, on which more attention should be paid for the achievement of the objectives.

The analysis of the Table 6.2 is provided below:

Increase New Customer Acquisition

Online marketers aim to form relations with new customers and achieve customer acquisition. New Customer Acquisition is achieved when new visitors enter a website, and move on with interacting with it (Peterson, 2004). Especially, for business sites and content sites it is essential to have an increasing number of new visitors in their websites (Peterson, 2006). Therefore, it is important to start with reaching new visitors and then triggering visitors to start an activity on the website. As it stated by B3 "*it is very important that online customers are directed to the right online channels and be exposed to content similar to their preferences*." In order to achieve this, online marketers should focus on Referral and Technical segmentation. Referral and technical attributes are the first attributes that are gathered when a new customer profile is created. With referral segmentation customers can be segmented according to referral characteristics which include: whether a customer is new or returning by capturing the I.P address**;** how the user reached a certain website; how they returned to this site; which was the initial source or which keywords they searched for in order to find the site. Apparently, not only can it be identified if the customer is new or not, but also which are the incentives that lead them to the website. Therefore, based on referral segmentation, marketers can show targeted content (banners, offers etc.) to each visitor so as to facilitate customer acquisition. Moreover, segmenting the customers according to the technical characteristics, such as the device they use, screen resolution, the browser, and the language of the browser, is also important for new customer acquisition. By technical segmentation, marketers can adapt the content showed to the customer based on the technical perspectives of the device or browser that each customer uses, in order to create a friendlier interface for the new customer. After the first visit, in order to achieve acquisition preferences should also be taken into account. Therefore, attitudinal category would be also considered quite useful in combination with referral and technical info categories. Other segmentation categories can be used, but are not that useful when used alone for attracting new customers to the website.

Improve Churn Rates

The improvement of churn rates is related to customer retention, since customer retention is achieved when the likelihood of a customer to churn is decreased. As it has been already mentioned, it is very important for online marketers to prevent their customers from leaving their brand so as to improve their churn rates and achieve customer retention. Customer retention entails targeting the most valuable and profitable customers and keeping them buying or visiting the website (Chaffey et al., 2009). Therefore, in order to improve churn rates value-based segmentation and behavioral segmentation are the segmentation types that an online marketer should first focus on. Value-based segmentation allows for the identification and targeting of profitable and valuable customers. Behavioral segmentation since it reveals whether a client is likely to churn or not according to its transactional behavior and the number of website visits. By focusing on Behavioral and Value-based segmentation types the most profitable customers can be identified, so as further marketing activities can be triggered to achieve customer retention and improve churn rates. For example, online marketers can provide most valuable customers with tailored propositions or promotions or award them with offers. At this point it should be underlined that the improvement of churn rates is a goal of high priority for online retailers. As it was mentioned by B3 during the interview" *the improvement of churn rates is also related with having your clients exposed to content according to their preferences*." Hence, after finding whether customers are likely to churn then marketers should target them based on their preferences. Therefore, segmenting them according to attitudinal segmentation type constitutes a complement to value-based and behavioral segmentation types and therefore rated as of medium usefulness. Other segmentation categories can be used, but do not add extra value when they used alone for improving churn rates.

Increase Customer Satisfaction

Customer satisfaction is a very important goal that online marketers set. As M1 mentioned during the interviews "*it is very important to give to the online customer the feeling that they are recognized and treated as individuals*." Therefore, it is crucial, that online customers or visitors get relevant information from the website and are exposed to certain interactions, such as banners, light boxes, or emails. The most effective type of segmentation in this case is attitudinal segmentation type, which includes characteristics that indicate customer preferences and interests. By using this segmentation type, online marketers can provide their online customers with relevant content and relevant offers based on their needs, avoiding having them exposed to irrelevant content that in most cases becomes annoying. Customers' satisfaction is difficult to be captured and measured online, since most of the times it can be obvious by a phone call or by filling in a satisfaction form. However, segmenting the customers according to behavioral characteristics or their engagement/loyalty score reveals the level of satisfaction, since the frequency of visits and interactions is an indication the level of visitors' satisfaction.

Increase Customer Loyalty

It is essential for online marketers to gain their customers' trust, increasing loyalty with the brand and further engaged with their online customers. According to Peterson (2004) "loyalty can be measured as the number of visits any visitor is likely to make over lifetime as a visitor", and can be typically measured by the number of visits per visitor. As explained before engagement or loyalty score can be calculated by setting certain business rules (Tsiptsis & Chorianopoulos, 2009). For example, a simple business rule could be based on the number of visits of a visitor: if a customer visits more than 2 times per day a website then he scores a point. The aggregated number of points results in a score that indicates customers' loyalty. Such score shows how often the visitor visits the website or their level of interest for certain products or content. As stated by M1 "*a loyalty or engagement score is a way to check customers' loyalty and engagement with the brand*". Loyalty-based segmentation includes attributes that indicate the loyalty and the interest of the customer for the brand. Hence, it is very important to be used as a segmentation type for identifying loyal or disloyal customers and support the increase of customer loyalty. According to Peterson (2006) the frequency of visits, or the time that a visitor spends on a website also indicates customers' loyalty. Therefore, behavioral segmentation can also be used for distinguishing loyal and disloyal customers, and proceed to corresponding marketing tactics to increase customer loyalty. Other customer attributes are of low importance for supporting the increase of customer loyalty.

Increase Cross-Up Sales

The increase of cross and up sales belongs to customer development stage of customer lifecycle. It is a very important objective for online retailers and entails the selling of additional and more expensive products to customers (Chaffey et al, 2009; Chaffey & Smith, 2008; Peterson, 2006). In order to achieve the increase of cross up sales, the online marketer has to understand the needs and preferences of the customers their buying behavior. Moreover, most valuable and loyal customers should be targeted for the sale of products of higher price. Therefore, it is suggested that online marketers should focus on the following segmentation types: Attitudinal, Value-based and Behavioral segmentation. Firstly, attitudinal segmentation type should be used by online marketers in order to provide customers with relevant products and promotions based on their needs and lifestyles. Behavioral segmentation is also important, since it shows the most active clients, who will be more willing to move to a purchase. Demographics, like the age, genre or the marital status, are also valuable for delivering relevant content to the customer and increase cross-up sales. However, as it as stated by B3 "*demographics category contains attributes that are not easy to be captured online, and thus are only useful in some specific business cases".* Therefore, since it can be used complementary to the attitudinal segmentation type is considered of medium usefulness. Tailored propositions can be offered to them according to their needs, so as cross-up sales can be increased. Additionally, decisions about the design of new products or the improvement of existing ones can be made. However, it is also important to take into account how profitable the customers are and how willing they are to proceed into further purchases. Therefore, value based segmentation type should be also used for identifying and targeting the most profitable customers for the increase of up sales.

Improve Conversion Rates

The improvement of conversion rates is as stated by B3 and M2 the most important objective for online marketers. B3 stated that "*The performance of conversion rates is actually what online marketers are judged upon.* A conversion entails the successful completion of certain activities by online customers that have a positive effect for the online organization (Peterson, 2004). There are different ways to measure conversions, according to each activity. However, the high visiting frequency of the website is the basic measurement, since conversion rates are calculated by the proportion of visitors that move on to a certain activity on a website, such as click on interactions, proceed on a purchase or on a subscription and obtain information (Winer, 2001).

As far as the meaning of conversion is concerned, "it always depends on the type of business" as stated by B3. For example for an insurance company, a conversion could be an online application for insurance. For other organizations can be a purchase, a contact request, a download of a product or just a click on a banner. Therefore, for the improvement of conversion rates all segmentation types can be important, depending always on the marketers' point of view. Attitudinal and referral segmentation types can play an important role for improving conversion rates, since the first indicates preferences and the latter the incentives that lead the visitor to the website, both allowing for distributing relevant content to the visitor. B3 stated that *"past behavior is also essential for the improvement of conversions"* . This is attributed to the fact that it shows how frequent the client visits, the number of actions they already did on the websites, indicating their willingness to move on with further actions. Therefore, behavioral segmentation is also important for online marketers, in order to identify customers with lower activity level and decide upon further marketing actions to trigger those customers to become more active on the website. Moreover, the adjustment of the content according to technical characteristics, can also lead the online customer to stay on the website and complete several activities, thus, technical segmentation could be also useful in some cases.

Online Marketing Objectives with appropriate segmentation types							
	Online Customer segmentation types						
	Attitudinal	Behavioral	Demographic	Loyalty-based	Referral	Technical	Value-based
Objectives							
Increase New Customer Acquisition	M	L	L	L	H	H	L
Improve Churn Rates	M	H	L	L	L	L	H
Increase Customer Satisfaction	H	H	L	M	L	L	L
Increase Customer loyalty	L	H	L	H	L	L	L
Increase Cross –up sales	H	H	M	L	L	L	H
Improve conversion rates	H	H	L	L	M	M	M

Level of usefulness: H: High M: Medium L: Low

Table 6.1: Online Marketing Objectives with appropriate segmentation types

6.2. Techniques for Online Customer Segmentation

Figure 6.2 shows the second step of the framework, which entails the techniques that are appropriate for each of the previously presented Online Customer Segmentation types. As shown on the scheme the analysis of the Online Customer Segmentation types requires certain techniques which can be implemented by Big Data tools. As described in Chapter 4, from the literature research on Big Data field, it became obvious that the most used and well known tool that can implement algorithms for data analysis in large databases is Apache Mahout. Apache Mahout can implement various data mining techniques that can be used for segmentation in very large and complex databases.

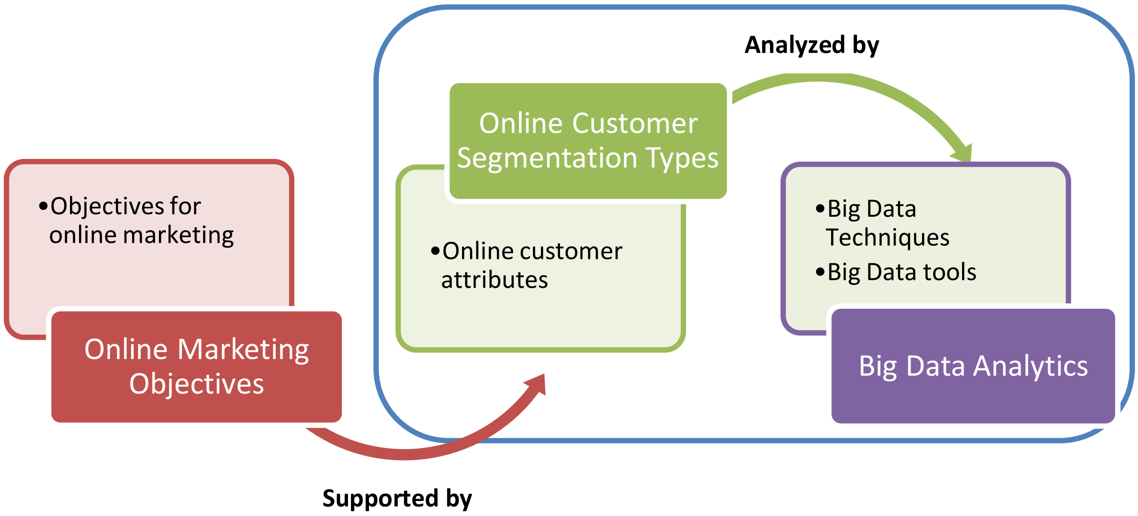

Figure 6.2: Techniques for Online Customer Segmentation (2nd Step of the Framework)

The link between Online Customer Segmentation types and Big Data tools and techniques is mostly based on the literature review made on Customer Segmentation using data mining techniques. The results from the literature review (Chapter 4) are used as an input for the framework. Moreover, experts' opinions during the interviews were taken into account for the creation of the framework. Table 6.2 presents the second part of the framework that this research proposes. As it can be seen the table consists of three columns, and it suggests techniques that are more suitable for analyzing each of the online customer segmentation types, as well as appropriate tools that can implement these techniques in large data bases.

The observations described below regard the big data techniques used for customer segmentation and constitute the explanation of the framework illustrated in Table 5.2.

Clustering

As it was observed from the literature review presented on Chapter 4.3, the prevailing technique that suits in most cases for customer segmentation is *clustering*. Cluster analysis belongs to the unsupervised modeling techniques. In that case all customer attributes from a selected set can be simultaneously analyzed, in order to create customer segments with similar characteristics. A cluster model is able to manage a large number of attributes and reveal data-driven segments which are not known in advance. As it was also mentioned by A2 clustering is used when it is difficult to know the segments in advance, and thus a general algorithm is needed. Therefore, clustering is preferred for behavioral segmentation. An example would be: If a marketer wants to create segments of customers in order to identify which of the customers of the website are more active and visit the website frequently, then a cluster algorithm can be used in order to analyze behavioral data and create customer segments using behavioral data. The clustering algorithm could take as inputs a combination of behavioral attributes such as number of clicks, number of visits, time of last visit etc. The algorithm would create homogeneous groups of customers, which would show which of them are more active or less active on the website. One well-known big data tool that can be used for the implementation of clustering techniques is Apache Mahout.

Classification

On the other hand, a very popular technique for customer segmentation is *classification*, which is used for supervised modeling. Classification is used when it is required for the segments to end in a specific result. This means that there is always a dependent variable according to which segments are created. According to the results of the literature review regarding techniques for customer segmentation, (see Chapter 4.3) it is preferably used for loyalty-based segmentation or value-based segmentation. The amount of money spent a customer spend or the degree of loyalty could constitute the dependent variable. However, it can also be used in some cases for behavior segmentation or for attitudinal segmentation, but still is preferred for loyalty-based or value-based segmentation. An example would be: When a marketer wants to create customer segments in order to identify the most profitable customers that are probably willing to purchase more products, based on the average amount of money they spend, then a classification technique is possible to be used. The dependent variable in this case would be “the average amount of money spent”, while the independent variables would be attributes like “amount of orders”, “average order value”, “number of purchases” etc. A classification algorithm would reveal groups of customers that are more profitable and thus more likely to move on with more purchases. Classifications techniques can be implemented as well by the big data tool Apache Mahout.

The techniques that are described below are not straightforward segmentation approaches. However, they can assist online customer segmentation and enhance decision making, by providing holistic approach of the online customer.

Association

Association is a technique, which also belongs to unsupervised modeling and is used for recommendation engine for web page personalization (Cakir & Aras, 2012). Association is not a straightforward segmentation technique, since it is basically used for uncovering relations among the data. However, some characterize it as segmentation technique since it is capable of “predicting” customers’ preferences according to previous behavior so segments with recommended products for certain customers can be created. Therefore, it is highly useful for segmentation based on customers’ preferences, in order to find out what type of content the visitor might prefer to see or what type of product they might prefer to buy. Apache Mahout is a big data tool that provides recommendation engines able of predicting patterns of user preferences out of very large amount of data (Owen et al. 2012). Thus, it is very effective when used for products recommendations and web page personalization based on the user preferences.

Regression

Regression is a traditional statistical technique which is basically used for prediction and forecasting. Although regression analysis cannot be considered a straightforward segmentation technique, it is actually able to create customer segments by predicting the segment in which a customer belongs. It is very useful for loyalty segmentation, where the customers are segmented according to their degree of loyalty (loyal or disloyal). As it was also mentioned by A2 regression analysis is also useful for value-based segmentation where customers are grouped according to their profitability, predicting if a client is profitable or not.

As mentioned before, R is an open source programming language which contains libraries for statistical analysis that can also be implemented in huge amount of data. Currently, R is a known open source tool that used for regression analysis of big data.

Visualization

Visualization is not a segmentation technique itself, but is used complementary to segmentation in order to illustrate segments and make them more understandable. Especially, in the case of a software tool that is addressed to people with limited knowledge of analytics, visualization of data becomes extremely useful for simplifying complex data patterns and thus, facilitating decision making. For online customer segmentation, it is very important to visualize huge amounts of online customer data into segments with the use of visualization tools for big data, indicating preferences, behaviors or consumer characteristics. There are plenty of tools for visualization of Big Data such as Visual Insight & Jaspersoft.

As explained before, during the research two new online customer segmentation types were identified based on online customer attributes; technical and referral. These segmentation types are not found in the literature, and thus it is not possible to make an assumption for the big data techniques that are suitable for analyzing these data, without implementing and comparing the techniques on real data. However, as it was also discussed during the interviews with A1 and A2, it is obvious that technical and referral segmentation can be assisted by both clustering and classification techniques, depending always if a dependent variable is used or not. Some examples could be: Create clusters of customers according to the device they use or the keyword they searched for; Use classification in order to find which clients came from a certain affiliate website or from a certain social media channel.

	Segmentation types					Big Data tools				
Method	Clustering	Association	Classification	Visualization	Regression	Recommender Apache Mahout	Clustering Apache Mahout	Classification Apache Mahout	R for Logistic Regression	Visual Insight &Jasper soft
Attitudinal	X	X		X		X	X			X
Behavioral	X			X						
Demographic	X			X						
Loyalty -based			X		X			X		
Technical	X		X	X			X	X		X
Value Based	X		X		X				X	
Referral	X		X	X			X	X		X

Table 6.2: Framework Part II -Techniques for Online Customer Segmentation

6.3. Summary

Table 6.3 constitutes an overview of the two frameworks, showing which online customer segmentation types should be preferably analyzed by certain techniques in order to assist the achievement of main online marketing objectives. In the table shows the level of usefulness of the online customer segmentation types based on online customer attributes gathered from online channels is indicated by different colors. In the colored cells the big data techniques suitable for analyzing each online customer segmentation type, are shown.

	Online Customer segmentation types						
	Attitudinal	Behavioral	Demographic	Loyalty-based	Referral	Technical	Value-based
Objectives	Big Data Techniques (Implemented by big data tools)						
Increase New Customer Acquisition	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Improve Churn Rates	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Increase Customer Satisfaction	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Increase Customer loyalty	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Increase Cross –up sales	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression
Improve conversion rates	Association Clustering Visualization	Clustering Visualization	Clustering Visualization	Classification Regression	Classification Clustering Visualization	Classification Clustering Visualization	Classification Regression

Level of usefulness of online customer segmentation types: High Medium Low

Table 6.3: Summary of Online Customer Segmentation framework

Part V: Evaluation and Conclusions

7. Evaluation

The following chapter presents the results of the evaluation of the two frameworks presented in Chapter 6. The two frameworks were separately evaluated by experts of different expertise. The first part of the framework, which matches the online customer segmentation types with online marketing goals, was evaluated by two marketers and two business consultants. The second part of the framework, which matches the online customer segmentation types with the big data techniques and tools, was evaluated by a software architect and an author of one of the books used for the purpose of this research, specialized in customer intelligence. Moreover, in order to illustrate how the two parts of the framework would work in practice, a walk-through of the framework presenting an ideal scenario of how the frameworks would be used in a hypothetical situation is provided.

7.1. Walk-through of the framework

Since it was not possible to test the framework on a real situation, a hypothetical yet representative example of how the framework could work in the case of a customer that uses the OCEM Tool is provided below. This example serves in better understanding how the frameworks that this research proposes could be used in practice. For the creation of this example, a sample of customer profiles created by the OCEM Tool was used. During the time this research was conducted it was not possible to make use of an analytic tool applicable to large and diversified data set. Therefore, for the purpose of the example the dataset provided was simplified, and analyzed with the use of SPSS. The hypothetical business scenario presented below is based on a real business case of one organization-client of DEVCORP who has employed the OCEM Tool.

Business Scenario

Company X is a TV and radio broadcaster that uses an OCEM tool provided by DEVCORP, called the OCEM Tool. X owns a bunch of online channels, since apart from the main website it owns a different website for every TV or radio program, Facebook pages and Twitter streams. Company X has employed the OCEM Tool in order to create customer profiles by collecting online customer attributes in real time from all owned channels, understand individual customers' journey, understand customers' behavior and provide more relevant content and information to the individual customer. Through the websites of company X a visitor can purchase a membership.

In general company X focuses on increasing customer satisfaction, improving churn rates and conversion rates and attracting new customers and. The objective that online marketers of company X should reach this year is to improve conversion rates by increasing the number of visitors that choose to move on with purchasing a membership. In order to reach the main goal, company X focuses on promoting the membership by showing to users relevant banners according to their preferences, interests and past activities, and instead showing additional promotions, in order to avoid churn and decrease churn rates. Moreover, they want to avoid showing banners for memberships to those who are already members. Table 7.1 shows the online customer attributes that company X gathers by using the OCEM Tool. Table 7.2 shows how these attributes fall under the online customer segmentation types, as they are defined in this research.

Categories	Online Customer Attributes gathered by Company X
Person	Age, Gender, City, Job, Member, Email , phone number, Social, keywords
Statistics	- Number of Visits, - Interactions Clicked, Interactions Viewed - Average visit time - last visited time - Entry Page, Referrer Hostname - Keywords on website - Click Count - engagement score
Other	- Gift preference - Program preference - Visitor type - Interests - Visited pages
Technical Info	- Flash Version, Java Version, - Operating System name, Operating System Version

Table 7.1: Customer Attributes gathered by the OCEM Tool for Company X

Online Customer Segmentation Types	Online Customer Attributes gathered by company X
Attitudinal	- Interactions Clicked, Interactions Viewed - Visited Channels, Visited Sites, - Program preference - Program fan - Visitor type - Interests - Entry page
Behavioral	- Number of Visits, - Number of clicks per visit - Average visit time, - member - last visited time - Click count
Demographic	- Age, Gender, City, Job, - Social Media Id - Email , phone number
Loyalty-based	- Engagement Score
Referral	- Referrer Hostname - Keywords - Keywords on website
Technical	- Flash Version, Java Version, - Operating System name, Operating System Version, browser name
Value-based	- N/A

Table 7.2: Online Customer Segmentation types based on data gathered by company X

In order to improve the conversion rates of visitors becoming subscribers and also improve the churn rates company X can to do the following:

- Distinguish the non-subscribers from subscribers.
- Identify how often customers visit company X websites.
- Identify online visitors' preferences
- Identify online visitors 'interests.
- Promote banners based on visitor interests.
- Promote banners based on visitors' preferences.
- Show more relevant content based on visitors preferences and interests.
- Show particular banners for members.
- Do not show banners for membership to members.

An effective segmentation of the visitors of the online channels that company X owns, would help in conducting the aforementioned tasks. According to the proposed framework in order to increase conversion rates one can start segmenting their online visitors based on the behavioral and attitudinal segmentation types. Tables 7.3 and 7.4 show behavioral and attitudinal attributes with a short description of each attribute. By using a clustering technique clusters of customers can be identified according to attitudinal and behavioral characteristics.

Behavioral attributes	Description
Number of Visits	Total number of times that the customer has visited X's website
Clicks per visit	Number of clicks per visit.
Average visit time	Average time spent on time per visit
Click count	Total amount of clicks
member	Visitor owns a membership or not
last visited time	The last time that an online customer visited one of X's websites.

Table 7.3: Behavioral Attributes

Attitudinal attributes	Description
Interactions Clicked	Interactions on website clicked by the customer
Interactions Viewed	Interactions on website clicked by the customer
Visited page	Number of clicks per visit
Entry X page	The first X's page that the customer visited
Interests	Visitors interests in categories: music, reality, nature etc.
Program preference	Preference to a specific program (TV or radio)
Program fan	Fan of a specific program (TV or radio)

Table 7.4: Attitudinal Attributes

Behavioral segmentation type

For the purpose of the example a sample of 1200 customer profiles were obtained, and the IBM SPSS statistic data editor was used for analyzing the data. In order to create clusters of customers, the TwoStep algorithm, which belongs to the clustering techniques, was used. The results of the segments are visualized in Figures 7.1 and 7.2:

Input (Predictor) Importance

1,0 0,8 0,6 0,4 0,2 0,0

Cluster	1	4	5	3	2
Label	Moderate active visitor (uknown ...	Not a Member	Moderate active member	Higly active member	Higly active visitor (uknown ...
Description	Uknown Membership, lower visited time and clicks	Visitor is not a member	Visitor is a member with lower visiting time and visits	Visitor is a member, with high visiting time and clicks	Uknown Membership, but highly number of clicks and visits
Size	66,5% (799)	16,3% (196)	11,9% (143)	3,6% (43)	1,7% (20)
Inputs	ismember 0 (100,0%)	ismember 2 (100,0%)	ismember 1 (100,0%)	ismember 1 (81,4%)	ismember 0 (55,0%)
	visits 291,99	visits 249,40	visits 383,76	visits 1.414,07	visits 728,50
	averagetime 0,71	averagetime 1,03	averagetime 0,50	averagetime 1,19	averagetime 41,80
	clickcount 762,60	clickcount 773,44	clickcount 1.003,36	clickcount 4.087,53	clickcount 7.556,20
	visitclicks 2,76	visitclicks 2,96	visitclicks 2,59	visitclicks 4,33	visitclicks 26,65

Figure 7.1: Online customer clusters created based on the behavioral customer segmentation type

As it can be seen in Figures 7.2 and 7.1 the TwoStep algorithm techniques revealed 5 basic online customer segments with similar behavioral characteristics. The pie chart in Figure 7.2 constitutes a simple visualization of the clusters and their size. Figure 7.1 illustrates the clusters and the average values of the specific behavioral attributes. As it can be seen the attributes are sorted and colored according to their input importance, as it was decided by the algorithm. The quality of the clusters is shown in Appendix 10.5, which indicates that the five attributes are a good selection for creating customer clusters. The largest group labeled as "moderate active visitor (unknown membership)" contains 66.5% of the online customers. The online customers that belong in this group for whom it is not known yet if they own a membership, and the level of activity according to their behavioral attributes (visiting frequency, number of clicks, average time spent). The second group labeled as "Not a member" includes 16.3% of the online customers. This group consists only of online customers for who is known that they are not members. The third group (11.9%), labeled as "moderate active members", shows the members who are less active, while the fourth (3.6%), labeled as "highly active members", the members who are more active during their visits on X websites., the smallest group (1.7% of the visitors), labeled as "highly active visitor (unknown membership)"includes visitors for who it is not known if they own a membership or not, but they seem to be highly active based on their behavioral attributes.

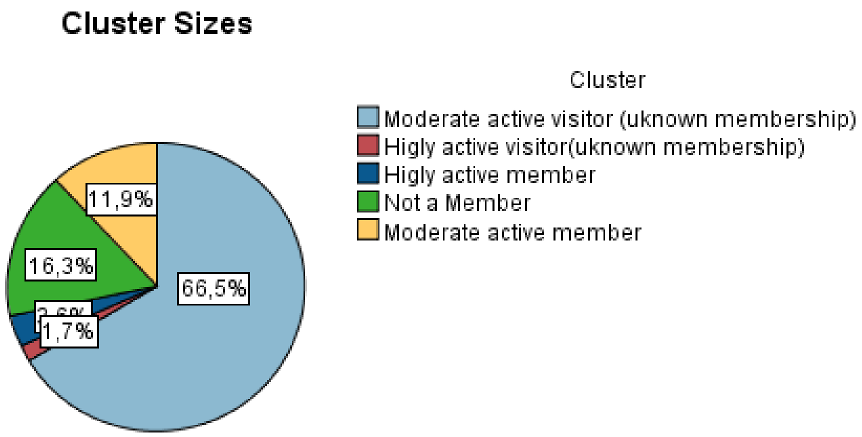

Figure 7.2: Size of Online customer segments

For each of these groups online marketers can create different campaigns. For example, they can target online customers that belong to “No Members” group, with relevant banners for membership. Moreover, they can target “highly active visitors” with relevant banners for memberships, since due to their highly visiting frequency they would be more likely to buy a membership. This strategy could help in leading more visitors to buy memberships, increasing the related conversion rates. Moreover, online marketers could avoid showing to visitors who belong to the categories “Highly active member” and “Moderate active member” banners which promote the purchase of a membership, in order to avoid deregistration due to exposure to irrelevant content, improving churn rates.

Similarly, using a clustering technique for attitudinal attributes, segments of customers with certain preferences and interests could be identified. Banners, email campaigns or other interactions could be based on the resulted clusters, in order to deliver more relevant content to the client.

7.2. Experts Evaluation Part I

In regard to the evaluation of the first part of the framework four interviews were conducted. Table 7.3 shows the anonymized list of the experts that were interviewed for the purpose of the evaluation of the first part of the framework. In the text that follows, the interviewees are mentioned with their Interviewee ID as depicted on the table 7.3.

Interviewee ID	Interviewee Job Title
M1	Chief Marketing Officer
M2	Product Marketer
B1	Business Consultant
B3	Business Consultant/Online Marketer

Table 7.3: List of Interviewees part I

The 4 interviewees were asked questions related both to the results of the framework as well as to its usefulness and utilization. The interviewers were provided with a preliminary version of the framework and they were asked for their opinions on the relevance of the objectives and customer segmentation categories, and the level of usefulness of its segmentation category.

Relevance of online marketing objectives

As far as the relevance of the set of the objectives presented on the framework is concerned, all interviewees agreed that the objectives are all relevant to online marketing and can summarize the basic high-level objectives of an online marketer. Thus, they agreed that they were relevant for the proposed framework. However, M1 mentioned that the improvement of conversion rates is at a level lower than the other objectives of the framework. M2, B1 and B3 agreed that the improvement of conversion rates is a very basic goal for online marketers, but it can be measured based on what a marketer considers a conversion. M2 and B3 also mentioned that although the improvement of churn rates and the improvement of loyalty are two distinct objectives, there might be occasions that these two objectives are not easily distinguishable, and thus, they have to be well defined by marketers and measured by certain KPIs.

Relevance of the online customer segmentation types

When asked for the online customer segmentation types all four interviewees mentioned that all segmentations types are appropriate for online customer segmentation and the online customer attributes fit perfectly to each of the categories. B1 and B3 mentioned that such a categorization of attributes could also be integrated in a software product capable of capturing online customer properties and creating customer profiles, to make it easier for the marketers to define their segments. When they were asked if they would segment their clients according to referral and technical attributes all of the respondents answered positively, while M1 and B3 of them commented that especially the technical info are very important to decide upon for real-time interactions and in general for real-time trigger generation.

Usefulness of the Online Customer segmentation types

In order to evaluate the framework in terms of the usefulness of each of the online customer segmentation types for each of the objectives, the interviewees were asked on which segmentation types they would focus more when they think of starting with a specific objective. The answers they provided confirmed the framework, while some useful remarks were made. More specific:

When asked for the "***Increase of new customer acquisition"*** objective, M1, M2 and B3 commented that technical and referral segmentation types are very important for the first visit of the customer, when actually the optimization of customers' journey in the website starts. As mentioned by one of the interviewees "for *the first contact is very important to categorize new visitors according to their referral and technical characteristics*". As mentioned by M2 and B1 it is important after segmenting visitors according to the aforementioned attributes, it would also be useful to take into account their preferences and show them more relevant content in order to complete customer acquisition.

When asked for the ***"Improvement of churn rates"*** objective**,** M2 and B3 mentioned that value-based and behavioral characteristics are very important, since they are useful for categorizing customers according to their profitability and past behavior. As mentioned, value-based is more useful for the online retailers, since the calculated values are results of customers purchasing behavior, while for business sites it's easier to identify customers that are more likely to churn by segmenting them according to their behavioral characteristics. M1 mentioned that characteristics that indicate needs and preferences have to be considered as well, after identifying customer's profitability and the risk of churn according to behavioral and value-based characteristics. An important indication made by M2 is that the keywords entered by the visitor indicate in a few cases the intention of the visitor to abandon the brand. "*It has been observed that words "cancel" or "cancel subscription" followed by the name of the brand, is used by people who want to terminate subscription or abandon the brand. In that case segments according to the keywords customers used are also important, to find out if they are about to churn or not".*

When asked for the ***"Increase of Satisfaction"*** objective, all four interviewees responded that attitudinal are the first characteristics that they have to be taken into account in order to create segments showing the preferences of the customer. As was commented by one of the interviewees "*Customers should be guided correctly through the channels and get what they need, in order to be satisfied. Therefore characteristics that indicate preferences are very important*." M2 and B1 also considered behavioral characteristics important for the specific objective, as the frequency of the visits or the subscriptions constitute an indication for their satisfaction. B3 mentioned that also the loyalty-based segmentation might give an indication of customers' satisfaction, but still when a customer is loyal does not mean that they are necessarily satisfied. However, at this point it should be underlined that M2 and M1 mentioned that satisfaction is not always easily measured only from the online behavior of the customer. "*It is very important to have feedback from the clients to increase satisfaction. When a client fills in a form, or calls at the customer support center to complain or to solve a problem, then the degree of satisfaction can be determined. Therefore, it is important that information from offline channels should be integrated into the product and taken into account for customer segmentation*", was commented by M2. M1 also mentioned, that it worth the case to conduct surveys for satisfaction, in order to be able categorize customer according to their degree of satisfaction, and move on with keeping the customers satisfied.

When asked for the ***"Increase of Loyalty"*** objective, all four interviewees agreed that the t loyalty-based segmentation type is very useful for the specific objective. As M1 stated "*a score that indicates the engagement, loyalty or the interest of the client in a certain brand should be always taken into account, so segments according to degree of loyalty can be created*". An engagement or loyalty score can be calculated by setting specific business rules that regard behavioral characteristics. Thus, the interviewees agreed that behavioral segmentation is also important to start with, in order to enhance the increase of loyalty. M2 mentioned that in his opinion loyalty, satisfaction, and the churn rate are related, and thus, a certain segmentation type that should be taken into account for the three related objectives is the behavioral segmentation.

When asked for the ***"Increase of Cross – up Sales"*** objective, M2 and B3 mentioned that the first step is to see which customers are most likely to buy products. Therefore, value-based segmentation and even more behavioral based segmentation is very essential in order to see whether a customer is likely to spend or whether they are willing to move on with purchases. M1 mentioned that behavior segmentation is necessary in order to see what types of products a customer has already purchased, in order to avoid suggesting the same products or be able to decide upon additional offers. All four interviewees mentioned that they would use attitudinal based on segmentation in order to segment customers according to their preferences and suggest new products for purchase. It is worth saying that 3 of the interviewees considered demographic segmentation important, but they believe that online customer characteristics such as age or genre are sometimes difficult or it takes time until they are captured online and added in a visitor's profile. As specifically B3 mentioned *"I would use demographic segmentation* *when demographics are available and for specific type of products, such as insurances, for which demographic characteristics are very important."*

When asked for the **"Increase of conversion rates**", all the interviewees mentioned that it depends on the online marketers' point of view and what they consider a conversion in each case. However, all agreed the attitudinal segmentation type and the behavior segmentation are important in most cases. B3 provided examples : " *Preferences and past behavioral is always the first thing you would see, when you consider as conversion visitors clicking on a specific banner, or moving on with purchasing a product or downloading content from your website. However, if somebody new enters referral or technical attributes would also be important to look into. Segmentation types like loyalty or a value-based segmentation, wouldn't add any value*". The low usefulness of value-based and loyalty-based segmentation is basically attributed to the fact that conversion rates are related with the activities that an online visitor completes while being on a website. This means that attributes that reveal visitors' preferences, intentions and behaviors are more important in order to lift conversion rates than the ones that reveal loyalty or profitability.

Utilization of the first framework

The interviewees were also asked whether they would utilize the first part of the proposed framework for online customer segmentation and how this could be done according to their opinions.

B3 mentioned that such a framework would be useful as an overview for business consultants in order to advise their clients who make use of OCEM products upon segmentation. Moreover, he mentioned that the framework would constitute a guideline for use-cases considering campaign creations, in order to identify customer segments, to which the campaigns should target.B1 stated that the framework "*would be useful for a quick overview when an online marketer start with specific objectives in order to know where they should focus or not, what they should do or not."* He also added that *"it could be given to the clients that use an OCEM software product in order to help them with segmentation. It would raise a lot of discussion and trigger them to think more on their objectives".* M1 mentioned that it would be a first guideline to show how different segmentation categories or a combination of them are needed in order to help in the achievement of the objectives of online marketers. M2 mentioned that it would help the clients to think of their objectives based on their needs, and possibly it would fit for specific use cases including segmentation that would automatically be embedded in the product.

Moreover, all interviewees mentioned that the categorization of the online attributes according to the segmentation categories would possibly fit in most cases in an OCEM product that creates customer profiles by capturing online customer attributes. In such a case, the online customer attributes that are gathered could be also categorized in the product according to the segmentation categories.

Finally, the interviewees were asked on what, in their opinion, would be the problems or the obstacles that would appear in the utilization of such a framework. M2 and B3 mentioned that although the approach is interesting, it would require much time in order to see actual results on certain business goals. M1 mentioned that this framework would be the start on achieving a goal by effectively segmenting the customers, but still there is much more that needs to be done in achieving those objectives. B1, it was mentioned that a standardization of segmentation types according to the proposed framework , could be integrated to a product for OCEM, but still there would be cases that would not fit exactly and, thus the product would have to be configured for the specific case.

Summary

To sum up, after finalizing the round of interviews, the first framework was confirmed by the experts and it was considered useful as a first step for an effective online customer segmentation that would enhance the achievement of objectives related to online marketing. Such a framework could be utilized as a guideline for effective online customer segmentation that assists business goals. However, due to the nature of online marketing it would take time to see the actual results of online customer segmentation on a specific objective.

7.3. Experts Evaluation Part II

With regards to the evaluation of the second framework, which shows which techniques -also applicable in big data- are appropriate for segmenting online customers according to certain segmentation types, two interviews were conducted. Table 7.4 shows the anonymized list of the experts that were interviewed for the purpose of the evaluation of the second part of the framework. In the text that follows, the interviewees are mentioned with their Interviewee ID as depicted on the Table 7.4.

Interviewee ID	Interviewee Job Title
A1	Author / Customer Intelligence Expert
A2	Software Architect

Table 7.4: List of Interviewees part II

The first interview was with the author of one of the books used for the purpose of the research. The author was asked to confirm the results of the literature review on customer segmentation on which the framework is based. Moreover, being an expert as well on the field of Customer Intelligence he was able to judge and confirm the results of the literature review and to give his point of view on which techniques would bear better results in each case. The second interview was with the software architect of DEVCORP who is also responsible for the OCEM Tool. He reviewed the second framework and commented if the techniques would be applicable for each of the online customer segmentation types and whether they could be effective or not. The effectiveness of big data tools and the possibility of implementing big data tools were also discussed. Moreover, the usefulness of such a framework and the way it could be utilized was discussed.

Relevance of techniques for Online Customer Segmentation

First of all, both A1 and A2 agreed with the online customer segmentation types and the techniques used for segmentation. More specific A2 commented that such a categorization would fit in an OCEM product, which creates online visitors profiles by collecting online attributes, in order to facilitate the segmentation process. From his point of view and his experience he mentioned that all classification, clustering, association, and regression are techniques relevant for segmenting customers according to the characteristics that are gathered online.

When asked for **clustering technique**, A1 mentioned that cluster analysis is the prevailing and most commonly used technique for segmenting customers. It is used when the natural groups that constitute the segments are not known in advance and a clustering algorithm is needed in order to define those groups and categorize the customers. He confirmed that clustering is the most commonly used for behavioral segmentation type and demographics. An example from banking sector would be: If it is required to identify the most profitable clients that are retailers, then a clustering technique should be used. Although A1 is an expert on customer segmentation in CRM and does not consider himself much experienced on the online customer segmentation, he commented that clustering technique is very useful for an online environment where the visitors of a site are basically unknown. "*In an online environment, you would always check how long the customer stayed in the website and how did he behave during his visit site."* In that case clustering could be used to create customer segments that reveal their behavior in terms of visiting time, clicks, page views etc. Furthermore, from experience he commented that technical and referral segmentation types could be both analyzed by clustering as well as classification techniques. A2 also commented that clustering technique would fit in most cases. Specifically for technical and referral segmentation, clustering in his opinion would be the most useful technique. While bearing some examples he stated "*Clusters can be made out of search engines and referring sites or the keywords that are mostly used*". However, he mentioned that also classification would fit; depending also on the segments that somebody wants to define.

When asked for **classification**, A1 mentioned "*We use classification when we want the segments to end in a specific result*". A1 confirmed that classification technique is very effective for classifying customers according to their loyalty score. In the case of value-based segmentation category he does not consider the classification very effective, because normally the marketers want to identify the most profitable clients, without knowing the customer value. However, it can be used in cases for behavioral segmentation as well. An example where classification would best fit would be when through segmentation there is a need to identify clients that use more a certain product. Another example A2 provided regarding the banking sector was: if it is required to identify clients that make high use of a certain banking product, then a classification tree should be used. Similarly, A2 mentioned that he considers classification always useful when there is a dependent variable according to which segments can be created.

When asked for **association**, A1 mentioned that association is not a technique for segmentation, but there are several analysts and scientists who might treat it as such. Association is used for recommendation engines, and especially for basket analysis as it happens in the case of Amazon.com. Association is always useful to find out what visitors prefer to buy. As A1 commented, in an offline environment is very difficult to understand the preferences and attitudes of the clients, and therefore data from surveys or market analysis are used. In an online environment it is much easier to track some of those characteristics. In such cases association analysis is useful for characteristics that indicate customer preferences (attitudinal segmentation type). Respectively, A1 mentioned that he considers association analysis very helpful for creating recommendation and for micro-segmentation according to the user preferences.

When asked for **regression** analysis A1replied that it cannot be used alone for segmentation. However, it estimates the probability that a customer belongs in a certain segment. It is often used to estimate churn probability and, thus, segment customers to churners or non-churners, based on value-based characteristics.

When asked about **visualization** A1 confirmed that visualization is very important for illustrating segments providing clear view, and is similar to cluster analysis. It actually, illustrates the results of a clustering.

Big Data tools

When A2 was asked whether an analytic tool for big data could be implemented in an OCEM software product to facilitate online customer segmentation he stated: "*From a theoretical point of view a big data tool can be implemented in the product for analysis and customer segmentation. However, from a technical point of view it always takes time to integrate such a product, since technical perspectives such as performance or multi-tenancy, should be taken into account.*" Moreover, A2 mentioned that big data analytic tools such as Apache Mahout or language R could be used for analyzing online customer data and perform customer segmentation. These tools could also be used in an OCEM tool. However, operational problems might appear that would regard to set a tool to work in the production scale and configure it according to the performance limits and the settings they have set for the product.

Utilization of the second framework

A2 commented that the second framework that this research proposes, would be a starting point to look into segmentation types and implement tools , like apache mahout, that are able to implement segmentation techniques. It shows which techniques are preferred for each segmentation category. Then it could be easily check what works or not, according to each specific situation.

Summary

To sum up, after finalizing the interviews, the second framework was confirmed by the experts. Both interviewees believe that there are not one-size-fits-all segmentation techniques. However, clustering and classification are the prevailing segmentation techniques, which can be used according to the required result. The second framework was considered useful as a starting point to look into online customer segmentation types and the techniques and tools through which they can be analyzed. It constitutes an overview that one can use in order to check what can be used in each case. As A1 mentioned, normally, the marketers should be able to ask for specific segmentation types, according to their needs and then data analysts are called to do the analysis and provide them with the best solution. This opinion indicates that the second framework in combination with the first framework proposed in this research, could be used to help marketers decide on specific segmentation types which would assist them in achieving their goals, and data analysts to have an overview techniques and big data tools that would me more effective in analyzing the specific segmentation types in each case.

8. Conclusions and Discussions

8.1. Main Findings

The rapid growth of the Internet and the continuously raising number of businesses that choose to go online has resulted in the generation of tremendous amount of customer data. As the organizations struggle to handle and utilize effectively all the information available in order to provide better products and services and gain competitive advantage, online customer segmentation and the so called "big data" analytics are two fields that currently constitute matter of concern and discussion.

This research sheds light on the two fields, highlighting the differences between an "online" and "offline" customer, as well as the differences in handling and analyzing "big data" and "traditional data". Moreover, it points out the need for building a concrete online marketing strategy in order to be able to get the appropriate information about the customer and create effective segments that would assist the achievement of specific online marketing goals.

During the research a comprehensive literature review in the fields of online marketing, customer engagement, big data and data mining was conducted. Moreover, an online customer engagement management tool provided by DEVCORP, called the OCEM Tool, was used to find out how customer data are generated from online channels, what type of data are generated, and how online marketers currently handle this data to target their customers, since the product does not yet provide data analysis.

Answers to the research questions of the research

The main objective of the research was to answer the main research question "What are the appropriate Big Data techniques that can assist an online marketing strategy based on online customer segmentation types?" The main deliverable of the research, which constitutes an answer to the main research question was a framework consisting of two parts; the first part showing online customer segmentation types able to assist online marketing objectives, while the second part shows which techniques that can be used for big data analysis are suitable for segmenting online customers according to each of, the online customer segmentation types. The main research question is answered by 3 sub-research questions which are answered as follows:

- ***RQ1:*** *Which are the main business objectives regarding online marketing and online customer engagement?*

Research question 1 was examined in chapters 3 and 5. Chapter 3 introduces the concepts of online marketing and online customer engagement. It presents the findings from a literature review regarding Online Marketing, focusing on finding main online marketing objectives. The literature review showed that organizations are currently building their online strategies and thus, their objectives are not clear and well defined in most cases. However, they are related to the steps of customer lifecycle: Acquisition, Retention and Development and there are plenty of KPI's, which can also be calculated by including online customer attributes that are gathered from online channels. In order to identify a set of high level business goals, and further discuss the matter a set of interviews with marketers and consultants of DEVCORP. After the interviews, the following set of high-level business goals was defined:

- Increase New Customer Acquisition

- Increase Customer Satisfaction
- Increase Customer Loyalty
- Improve Churn Rates
- Increase cross-up sales
- Increase conversion rates

❖ ***RQ2:*** *Which are the customer segmentation types that can assist each of the business goals regarding online marketing?*

Research question 2 was examined in chapters 3 and 5. Since not enough scientific literature on online customer segmentation was available, a literature review on traditional customer segmentation was made. In chapter 3 the notion and the importance of an effective customer segmentation for achieving is explained, while customer segmentation types, which are based on attributes gathered from offline channels and sources are presented. In Chapter 5, the OCEM tool is taken as a case study, in order to figure out what kind of attributes can be gathered online and point out the differences from offline attributes. As it was found out the most of the attributes gathered online match with the attributes gathered offline and fall under the main customer segmentation categories, as they are defined from traditional customer segmentation theory. Two new online customer segmentation types were defined that are based on referral attributes such as keywords, or host site and on technical info such as screen resolution, operation system, browser, language. In total 7 online customers segmentation types were identified;

- Attitudinal (also in offline)
- Behavioral (also in offline)
- Demographics (also in offline)
- Loyalty based(also in offline)
- Value-based; (also in offline)
- Referral (only online)
- Technical (only online)

As it was understood from the literature and from the discussions with the experts on the topic, effective online customer segmentation plays an essential role for achieving online marketing objectives. However, as it was concluded there is no one-size-fits-all segmentation. The answer to research question 2 is provided in Chapter 6 in table 6.2, where the first part of the framework is presented. The main inputs of the framework were the data collected from the literature review and empirical data as demonstrated in Chapters 3 and 5 correspondingly. The framework shows the level of usefulness of each online customer segmentation type for each of the aforementioned online marketing objectives. As it was observed behavioral and attitudinal segmentation types are of high usefulness for most of the objectives.

❖ ***RQ3:*** *Which big data approaches and techniques can be used for each online customer segmentation type?*

Research question 3 was examined in chapter 4. Chapter 4 introduces the notion of Big Data, while the basic techniques that can be applied on Big Data and basic tools able to handle and analyze big data are presented. Moreover, differences and similarities between big data and traditional data are explained. As it was found the techniques that can be applied on big data, stem from data mining and statistics. The difference is related to the characteristics of the data, while the analytic techniques that are used for big data analysis do not differ from those used for traditional data mining. The difference exists in the tools that are able to handle and analyze the large volume and the diversification of the data. As far as the tools are concerned, the most well-known open source framework for big data is Apache Hadoop, which includes tools for big data management, processing, storing and analysis. Apache Hadoop includes Apache Mahout for big data analysis using clustering, classification and recommendation engines (associations). Moreover, programming language R

Since the main techniques used for customer segmentation can also be applied in big data, a literature review on customer segmentation and data mining techniques was made. A set of 15 studies relevant to the topic was found, and trends of techniques used for certain types of customer segmentation were pointed out. The main techniques found were: Clustering, Classification, Visualization, Regression, Associations. Clustering and Classification are the techniques that are used more often and can be applied to the most of the customer segmentation types**,** while the rest of the techniques are not straight segmentation techniques, but their results can be used for customer segmentations. All techniques can be implemented on big data by Apache Mahout and the programming language R.

The answer to the 3d research question is provided in Chapter 6 in Table 6.3, where the second part of the framework for Online Customer Segmentation is presented. The framework suggests which techniques can preferably be used for online customer segmentation according to each of the online customer segmentation types.

8.2. Research Limitations

The results of this research should be interpreted in regard to the following limitations:

Lack of prior research studies on the topic: Although there is prior research on the field of customer segmentation and data mining, there was no scientific research found on the opportunities that big data offer for the online customer segmentation. Furthermore, the related literature found focusing on online customer segmentation was limited. Moreover, there was no framework found that connects online customer segmentation to online marketing objectives. Therefore, the results of the research are based on a combination of literature data stemming from online marketing, traditional customer segmentation with data mining, and currently available big data theory, as well as on empirical data gathering.

Lack of available data: During the research it was not possible to interview the clients that use the OCEM tool, in order to find out their objectives and needs in terms of segmentation. Since, the product is relatively new the clients still work with it under the guidance of consultants, and thus interviewing them it was not advisable. Moreover, it was not possible to observe customer data that are gathered by tools similar to the OCEM tool, in order to search for other possible online customer data that can be gathered online.

Time limitation: The research was conducted in DEVCORP and lasted 6 months. During the research discussions and interviews with the experts of DEVCORP were conducted, internal documentation was reviewed, while data from the OCEM tool were observed. Due to time limitation, it was not possible to test the proposed frameworks on a real-situation. First of all, the application of big data tool capable of implementing techniques for online customer segmentation, such as Apache Mahout on the OCEM tool, was considered time-consuming , due to the complexity of the tool and the configuration it requires. Moreover, as it was also stated by the experts testing the big data techniques in order to see the actual results in online marketing strategies , is a process that would require time that would much exceed the time limits of this research.

Moreover, at this point it should be underlined that this research focuses on techniques used for customer segmentation and not on specific algorithms that this techniques involve. This means that there are algorithms that belong to a specific technique that might perform better thatn the others. Moreover there are cases where algorithms that belong to a specific technique, cannot be applicable to very larga datasets , basically due the fact that they are not as fast as required.

8.3. Discussion and Future Research

This research raised the following issues:

- Online customer segmentation is a core process for assisting an online marketing strategy. However, there is limited scientific research related to the field.
- Huge amount of online customer data are continuously generated. However, there was no scientific research found for the use of big data tools in online customer segmentation.
- In the world of business a gap between online marketers and data analysts emerges. Normally, online marketers should be able to select combinations of online customer segmentation types, that would better serve their needs and goals, and then data analysts are called to do the analysis and provide them with the best solution. However, online businesses are currently building their online strategies and thus they do not have well defined online marketing goals. Therefore, marketers do not provide data analysts with the appropriate information, while proceeding with valuable and effective online customer segmentation becomes difficult.

Bearing the above in mind, it is apparent that a general guideline for effective online customer segmentation is needed, while the opportunities that big data offer for online customer segmentation should be further explored.

The two frameworks that this research proposes could constitute a first step towards an effective online customer segmentation approach capable of assisting an online marketing strategy. In feature research, the frameworks could be tested as a whole on more than one real situation. Firstly, the actual usefulness of online customer segmentation types for each of the objectives, according to the first framework can be tested on a real situation. Starting with a certain online marketing objective certain online customer segmentation types can be selected to be analyzed for creating actionable customer segments. After the online customer segmentation types that should be analyzed have been chosen, the implementation of big data techniques and tools as shown in the second framework should be tested. Big Data tools, as those proposed in the second framework, can be used in order to implement techniques to analyze the online customer segmentation types and create effective customer segments. Each of the techniques should be tested in order to find out what bears the best result in each case. This would, of course, require a longitude research since it would require time to see the actual effectiveness of the online customer segmentation in online marketing objectives.

Moreover, in a future research that would test the framework, more OCEM software tools, which are capable of gathering online customer attributes should be observed, in order to explore if there are additional online customer attributes that could form new online customer segmentation types. Additionally, the effects of the online customer segmentation online marketing objectives of lower level can be tested.

Furthermore, a future research could focus on comparative evaluation of specific algorithms that are suitable for analyzing each of the online customer segmentation types, in order to explore which algorithms bears the best result. Moreover, the comparative evaluation should also focus on whether the algorithms can perform well in large data sets.

Finally, it is apparent that the huge amount of customer data gathered online calls for a beneficial harnessing of those data in order to create a holistic view of the customer. Obviously, Big Data can play a major role in the field of customer segmentation. Therefore, further research should focus more on how big data tools and approaches can assist online customer segmentation by using tools for the implementation of the techniques for customer segmentation as it is also suggested by the second of framework of this research.

9. Bibliography

Abraham, A., Hassanien, A. E.,& Snâaésel, V. (2010). Computational social network analysis: trends, tools and research advances. Springer.

Agneeswaran, V. S. (2012). Big-Data : Theoretical, Engineering and Analytics Perspective. In Big Data Analytics (pp. 8-15). Berlin Heidelberg: Springer.

Anil, R., Dunning, T., & Friedman, E. (2012). *Mahout in action*. Shelter Island: Manning.

Baranov, A. 2012. Building Online Customer Relationship. *Bulletin of the Transylvania University of Brasov, Economic Sciences.* 5(54), 15-18.

Brodie, R. J., Hollebeek, L. D., Juric, B., & Ilic, A. (2011). Customer Engagement: Conceptual Domain, Fundamental Propositions, and Implications for Research. *Journal of Service Research*, *14*(3), 252–271.

Cakir, O., & Aras, M. E. (2012). A Recommendation Engine by Using Association Rules. *Procedia-Social and Behavioral Sciences*, *62*, 452-456.

Chaffey, D. Chadwick-Ellis, F., Mayer, R., Johnston, K. (2009), Internet Marketing: Strategy, Implementation and Practice, Essex: Prentice Hall.

Chaffey, D., & Smith P.R. (2008). eMarketing eXcellence: Planning and optimizing your digital marketing. Oxford: Elsevier

Chan, C. C. H. (2005). Online auction customer segmentation using a neural network model. *International Journal of Applied Science and Engineering*, *3*(2), 101-109.

Chan, C., C., H. (2008). Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer. *Expert Systems with Applications* 34(4): 2754–2762.

Chen, H., Chiang, R., H., L., & Storey, V., C. (2012). Business Intelligence and Analytics: From Big Data to big impact, *MIS Quaterly* 36 (4), pp.1165-1188

Cheng, C.H., & Chen, Y.S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, *36*(3), 4176–4184.

Chen, Y., Guozheng, Z., Dengfeng, H. & Chua, F. (2007). Customer segmentation based on survival character. *Journal of Intelligent Manufacturing* 18(4): 513–517.

Chu, B., H., Tsai, M., S. & Ho, C., S. 2007. Toward a hybrid data mining model for customer retention. *Knowledge-Based Systems* 20(8): 703–718.

Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107-113.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). “From Data Mining to Knowledge Discovery in Databases.” *Al Magazine,* Vol. 17(3), pp. 37-54

Farooqi, R., & Raza, K. (2011). A Comprehensive Study of CRM through Data Mining Techniques. *Business*, *2*(i).

Fotaki, G., Gkerpini, N., Triantou, A, I., Brinkkemper, S. (2012). *Online Customer Engagement Management.* Utrecht University.

Gogia, S. (2012). The Big Deal about Big Data For Customer Engagement. Forrester Research. Retrieved from: http://www.forrester.com/The+Big+Deal+About+Big+Data+For+Customer+Engagement/fulltext/-/E-RES72241

Gupta, R., Gupta, H., & Mohania, M. (2012). Cloud Computing and Big Data Analytics: What Is New from Databases Perspective?. In *Big Data Analytics*(pp. 42-61). Springer Berlin Heidelberg.

Hevner, A., & Chatterjee, S. (2010). Design Research in Information Systems. *22*. Springer

Hevner, A., R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information System Research. *MIS Quarterly. 28*(1), 75–105.

Hong, T. & Kim, E. (2010). Segmenting customers in online stores from factors that affect the customer's intention to purchase. *International Conference on Information Society*, *39*(2), 383–388.

Hosseni, M, .B, Tarokh M. J. (2011). Customer Segmentation Using CLV Elements. *Journal of Service Science and Management* 04(03): 284–290.

Jansen, S. M. H. (2007). Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior. *A Vodafone Case Study.*

Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, *31*(1), 101-107.

Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S. (2010). Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. *Journal of Service Research*, *13*(3), 297–310.

Lee, J., & Park, S. (2005). Intelligent profitable customers' segmentation system based on business intelligence tools. *Expert Systems with Applications*, *29*(1), 145–152.

Leung, H., C. (2009). An Inductive Learning Approach to Market Segmentation based on Customer Profile Attributes. *Asian Journal of Marketing.*

Mitra, S., Pal, S. K., & Mitra, P. (2002). Data mining in soft computing framework: a survey. *IEEE transactions on neural networks*, *13*(1), 3–14.

Manyika, J., Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. "Big data : The next frontier for innovation, competition, and productivity." *McKinsey Global Institute* 364(May): 156.

Miguéis, V. L., Camanho, A. S., & Falcão e Cunha, J. (2012). Customer data mining for lifestyle segmentation. *Expert Systems with Applications*, *39*(10), 9359–9366.

Mollen, A., & Hugh W. (2010). Engagement, telepresence and interactivity in online consumer experience: Reconciling scholastic and managerial perspectives. *Journal of Business Research* 63(9-10): 919–925.

Ngai, E.W.T., Xiu, L., Chau, D.C.K., (2009) Application of data mining techniques in Customer Relationship Management: A literature review and classification.

O'Reilly (2012). Big Data Now: 2012 Edition.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Peterson, E. T. (2004). *Web analytics demystified: A marketer's guide to understanding how your web site affects your business*. Celilo Group Media & CafePress.

Peterson, E.T. (2006). The Big Book of Key Performance Indicators. *Web Analytics Demystified.*

Pillai, J., & Vyas, O. P. (2012). CSHURI – Modified HURI algorithm for Customer Segmentation and Transaction Profitability, *2*(2), 79–89.

Russom, P. (2011). Big Data Analytics. *TDWI Best Practices Report,* Fourth Quarter

Sathi, A. (2012). Big Data Analytics: Disruptive Technologies for Changing the Game. USA: MC Press Online.

Sengamedu, S. H. (2012). Scalable Analytics–Algorithms and Systems. In *Big Data Analytics* (pp. 1-7). Springer Berlin Heidelberg.

Shukla, A., Wexler, M., Singh, V., Kan, E., Joshi, D., & Lestiyo, I. (2012). Micro-bucket testing for page optimization. *U.S. Patent No. 8,126,930*. Washington, DC: U.S. Patent and Trademark Office.

Srinivasa, S., Bhatnagar, V. (2012) Big Data Analytics. Berlin: Springer

Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, *31*(1), 127–137.

Stroud, Dick. (2006). Customer Intelligence. *Journal of Direct Data and Digital Marketing Practice* 7(3): 286–288.

Sun, H., & Heller, P.(2012). Oracle Information Architecture : An Architect ' s Guide to Big Data. An Oracle White Paper in Enterprise Architecture

Tsiptsis, K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. Wiley.

Turban E., Sharda, R., Delen D., & King, D. (2010). Business Intelligence: A Managerial Approach.(2nd ed.) New Jersey: Prentice Hall

Van Doorn, J.,Lemon, K.,N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P., C. (2010). Customer Engagement Behavior: Theoretical Foundations and Research Directions. *Journal of Service Research* 13(3): 253–266.

Weerd, I. van de, Brinkkemper, S. (2008). Meta-modeling for situational analysis and design methods. In M.R. Syed and S.N. Syed (Eds.), *Handbook of Research on Modern Systems Analysis and Design Technologies and Applications (pp. 38-58)*. Hershey: Idea Group Publishing.

Winer, R. S. (2001). A framework for customer relationship management. *California management review*, *43*(4), 89-105.

Woo, J., Bae, S., & Park, S. (2005). Visualization method for customer targeting using customer map. *Expert Systems with Applications*, *28*(4), 763–772.

Wu, R.S & Chou P.H. (2011). Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. *Electronic Commerce Research and Applications* 10(3): 331–341

Wymbs, C. (2011). Digital Marketing: The Time for a New “Academic Major” Has Arrived. *Journal of Marketing Education*, *33*(1), 93–106.

Ye, L., Qiuru, C., Haixu, X., Yijun, L., & Guangping, Z. (2013). Customer Segmentation for Telecom with the k-means Clustering Method. *Information Technology Journal*, *12*(3), 409-413.

Zaslavsky, A., Perera, C., & Georgakopoulos, D. (2012). Sensing as a Service and Big Data. *Proceedings of the International Conference of Advances in Cloud Computing. 8: 21-29*

Zhou, Y., Wilkinson, D., Schreiber, R., & Pan, R. (2008). Large-scale parallel collaborative filtering for the 86etflix prize. In *Algorithmic Aspects in Information and Management* (pp. 337-348). Springer Berlin Heidelberg.

10. Appendices

10.1. Basic Definitions

Big Data	*Data sets and techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis and visualization technologies.* ***(Chen et al., 2012).*** *Big Data is the frontier of a firm's ability to store, process, and access (SPA) all the data it needs to operate effectively, make decisions, reduce risks, and serve customers.* ***(Forrester, 2012)*** ***Big data*** *is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making****.(Gartner, 2010)***
Data Mining	*Data mining is the process of searching and analyzing data in order to find implicit, but potentially useful, information. It involves selecting, exploring and modeling large amounts of data to uncover previously unknown patterns, and ultimately comprehensible information, from large databases* ***(Shaw et al., 2001)***
Online Customer Engagement	*Online engagement is a cognitive and affective commitment to an active relationship with the brand as personified by the website or other computer-mediated entities designed to communicate brand value. It is characterized by the dimensions of dynamic and sustained cognitive processing and the satisfying of instrumental value (utility and relevance) and experiential value (emotional congruence with the narrative schema encountered in computer-mediated entities).*
Customer Engagement	*Customer engagement aims to go beyond transactions and can be defined as the customer's behavioural manifestations that have a brand or firm focus resulting from motivational characters* ***(Doorn et al., 2010).*** *Three stages:* *a)Identify Customer Engagement Behavior and Customers* *b)Evaluate Engagement Manifestations* *c)Reacting on Customer Behavior*
Customer Segmentation	*Research area that intrigued many researchers in data mining. Effective method for managing different customers with different preferences. It increases not only customer satisfaction but also the expected profits of the company. Diverse marketing strategies used in customer segmentation also increase customer value. (****Chen et al. 2007****)*
Market segmentation	*The process of identifying customers who comprise a homogeneous group of consumers for a specific range of goods and services.(****Stroud, 2006****)*
Conversion Rate	*A conversion rates is the number of "completers" divided by the number of 'starters' for any online activity that is more than one logical step in length. (****Peterson, 2004****)*
Loyalty	*Loyalty can be measured as the number of visits any visitor is likely to make over lifetime as a visitor. It should be measure as the raw number of visits all visitors have made since measurement was initiated, and the number of visits should be de-duplicated. (****Peterson,2004****)*

Table 10.1: Basic Definitions

10.2. Big Data: Example of MapReduce

Figure 2 shows an example of the MapReduce implementation for a scenario where one wants to find the list of customers having total transaction value more than $1000.

```
void map(String rowId, String row):
 // rowId: row name
// row: a transaction recode
customerId= extract customer-id from row
transactionValue=    extract    transaction    value    from    row
EmitIntermediate(customerId, transactionValue);

void reduce(String customerId, Iterator partialValues): // customerId:
Id to identify a customer // partialValues: a list of transaction values int
sum = 0;
for each pv in partialValues: sum += pv; if(pv > 1000)
Emit(cutsomerId, sum);
```

Figure 10.2: MapReduce Example 1(Gupta et al., 2012)

10.3. Segmentation types with related tasks and algorithms

The Table 10.2 shows the business tasks that can be achieved by each of the customer segmentation types, and the algorithms with which they were analyzed, as found in the relevant studies used durint the literature review.

Algorithms	Segmentation type	Business Task Assisted	References
Soft clustering approach	Behavioral, demographic, attitude(website usage)	Customer retention, Customer Satisfaction	Wu & Chou, 2011
K-means	Demographic, behavioral	Customer Retention	Chen et al., 2011
Cluster Analysis	Behavioral, attitudinal	Customer Satisfaction (meet the needs of customers)	Miguéis et al.,2012
ANN/SOM	Value based-RFM	Understand buyers behavior to offer reasonable prices	Chan, 2005
Variety of clustering algorithms	Usage-behavior		Jansen, 2007
K-means	Behavioral , demographics Value based,	Improve target marketing, make recommendations	Ye, Yijun &Zhu, 2013
K-means &SOM	Psychographic data	Customer Identification differentiated marketing programs	Hong & Kin, 2012
K-means	Behavioral, usage-behavior	New product development. Tailored offers and incentives	Tsiptsis & Chorianopoulos, 2009

Clustering model	Behavioral Demographic	Improve Customer handling	Tsiptsis & Chorianopoulos, 2009
Automatic clustering algorithms	Attitudinal	Customize loyalty programs(Increase Loyalty	Tsiptsis & Chorianopoulos, 2009
Clustering model	Demographic Behavioral	Increase revenues.	Lee & Park, 2005
SOM	Behavioral , Attitudinal(focus on rare expensive items)	Maximize customer retention, loyalty and profitability	Pillai & Vyas, 2012
HURI	Attitudinal, Behavioral	Recommendation Engine. Cross& Upsales	Tsiptsis & Chorianopoulos, 2009
Genetic Algorithm	Value-based	Identify high value customers for campaigns. Increase customer loyalty and customer LTV	Chan, 2008
Decision tree	Value based(LTV), Loyalty	Increase Cross & Upsales, provide better services (increase satisfaction), increase loyalty.	Kim et al., 2006
Decision tree	Behavior	Customer Identification(Acquisition)	Tsiptsis & Chorianopoulos, 2009
Decision tree	Value –based, loyalty	Retain High value customers	Han, Lu &Leung, 2012
Customer map for customer targeting	Attitudinal, demographic, customer value	Customer retention, customer targeting	Woo et al., 2005
Binary Logistic Regression	Value-based, behavioral	Calculates customer loyalty& Churn Risk. Useful for cross-up sales	Hosseni&Tarokh, 2009

Table 10.2: Segmentation types with related tasks and algorithms

10.4. Example Data gathered from the OCEM Tool

The tables that follow illustrate online customer attributes that are collected by different type of organizations that use the OCEM Tool.

10.4.1. Example of customer data gathered from business- online retailer's website:

Per Product	Personal Info
Interested in Subscription	Age
Average Older Value	Gender
Ecards sent	Has email
Ecards received	URL
Favorite Payment method	Mobile
Purchase hours	
Hours since last order	
Offers viewd	

Previous magazines/books selected	
Product types purchased	
Number of purchases	
Purchased	
Shopping Interest	
Webshop visits	
Nr of Webshop baskets	
Preferences (favourite author,genre,ebook etc)	
Average Older Value	
Webshop order	
Ipad user	

Table 10.3 Segmentation types with related tasks and algorithms

10.4.2. Example of online customer data gathered by the OCEM Tool for business site:

Personal Info	Social	Preferences
Age	Community member	Fav Beer
City	Facebook Fan	Brand
Country	Fb ID	Occasion
Email	Facebook User	Festival
Name	Twitter ID	Football Club
Gender	Twitter User	
Street		
Job		
Zipcode		
Username		
Date of Birth		

Table 10.4: Online customer data gathered by The OCEM Tool for business site

10.4.3. Example of online customer data gathered by the OCEM Tool for retailer site:

Visitors Info	Engagement	Products	Membership Properties
City	Favorite Offer Category	Number of Coupons	Cellphone
Days Since Join	Car Care Engagement Index	Products in Shopcard	Location
Is a member	Dining &Food Index	Products in Waller	Newsletter
Keywords	Shopping Engaement index	Products Viewed	
Restaurants index	Travel Engagement Index	Coupons Print	
Name	Favorite Subcategory		
User Identifier	Favorite Homepage		
Zipcode	Store Clickthrough		
Entertainment Index	Stores Viewed		
Visited Subcategories	Customer Loyalty Index		

Table 10.5: Segmentation types with related tasks and algorithms

10.5. Quality of clusters

Figure 10.2 shows the quality of the clusters created after using 5 behavioral online customer attributes in order to create customer segments, using the TwoStep algorithm. As it can be seen the cohesion and separation of the clusters approaches 1, indicating that the 5 inputs were quite good for creating clusters.

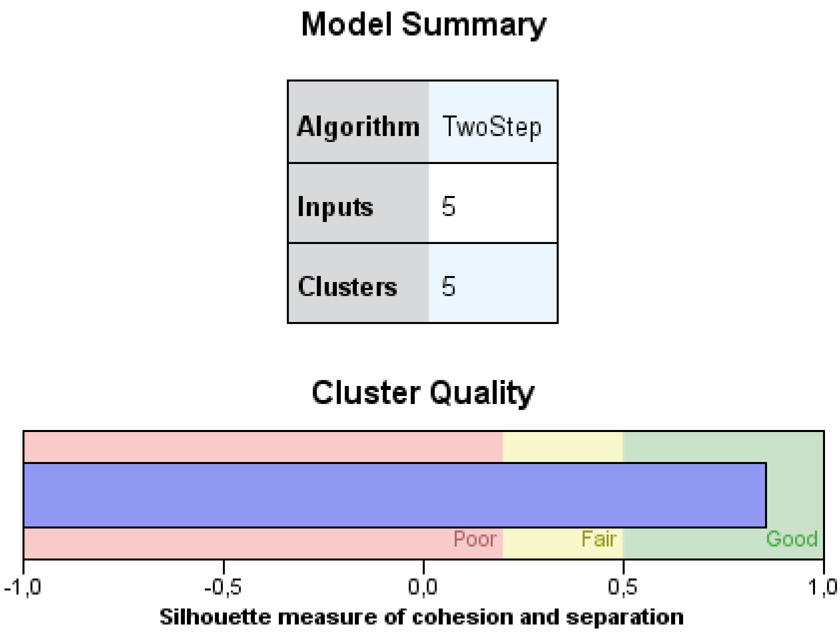

Figure 10.3: Cluster quality