INTERNATIONAL
HELLENIC
UNIVERSITY

Recommendations for personalized marketing and efficient budget allocation

Thomai Fytili

SID: 3305170006

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Digital Marketing & e-Business

December 2019

THESSALONIKI – GREECE

Recommendations for personalized marketing and efficient budget allocation

Thomai Fytili

SID: 3305170006

Supervisor:	Ioannis Magnisalis
Supervising Committee Members:	……………………
	……………………

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Digital Marketing & e-Business

December 2019

THESSALONIKI – GREECE

I hereby declare that the work submitted is mine and that where I have made use of another's work, I have attributed the source(s) according to the Regulations set in Student's Handbook.

Abstract

Traditional marketing has been replaced by digital marketing, online disruptive technologies and a big variety of performance monitor tools which generate big data. Access in an enormous variety of generated data, created the necessity in marketing to improve data analysis skills, in order to interpret the data, create insights, convert it to knowledge and integrate this knowledge into the decision-making process. Optimization of online campaigns in terms of personalization and efficient budget allocation, are two critical problems that marketers need to handle. An efficient budget allocation in combination with personalized content based on customers' interests, can lead to better performance and return of investment. Advertisers' inability to analyze data from previous or current campaigns and interpret it, is most of the time a serious impediment to this procedure. In our days, marketing decisions should be derived from data analysis and past performances. Marketers not only should be able to build creative campaigns, but also to monitor them, gauge their performance and take actions based on the results.

Through this work, we analyze data from two different business examples, using processes and visualization in RapidMiner. The main goal is to give marketers easy solutions, in order to make data interpretations or hypothesis and use them in the decision-making process. Also, main goal of this analysis is to give recommendations about which type of data to focus, how to handle it and why personalized content and efficient budget distribution can optimize campaign performance in a great extent.

Keywords: Digital Marketing, Marketing Campaigns, Decision making, Web Analytics, Prescriptive Analytics, Budget Allocation, Budget Optimization

Thomai Fytili

Date: 02.12.2012

Acknowledgements

I would like to express the deepest appreciation to my supervisor, Ioannis Magnisalis, for his assistance and guidance throughout my studies. In addition, I would like to thank all the professors that I had the chance to meet and work with, during this journey of my studies. I would especially like to thank professor Christos Moridis who helped me realize why marketing is important and what a good marketer and person should be.

I would also like to say thank you to my boss Hermann Semeloglou, for his understanding, support and trust in the last two years. Additionally, many thanks to my family and my friends whose love and guidance are always with me. Last but not least, I wish to thank also Frank Blanning for his unconditionally support and help through all the difficult moments.

List of Figures

List of Tables

1 Introduction

The emergence and evolvement of the Internet has had a huge impact on multiple aspects. We live in a new digital era, where the Internet forms the biggest medium to transact goods and services. [1] Online platforms, channels, e-shops and social media are becoming an integral part of our daily life. Magazines, newspapers and other traditional media have lost their power, while online communication, e-commerce, digital advertising and share of information in online platforms are increasing continually (Marketing news, May 2013 p.16) [2].

During the last 10 years online advertising and the number of internet users have grown exponentially. All the types of advertising that Internet users are exposed to, are combined and consist of digital marketing campaigns. A digital marketing campaign usually consists of sub-campaigns and different ad-sets. Each ad-set may have different goal, target audience, design and content. The main goals usually a digital marketing campaign has, is to attract users' attention, enhance engagement between brands and customers, generate traffic, conversions or revenue. Different campaigns presenting different products or services, have different goals and use different distribution channels. Planning is the first stage of a campaign and usually consists of the outline of campaigns' goals, the decision of the target audience, the time that the campaign will be launched and its duration. The next step is to develop a strategy plan which includes the main content of the campaign and how and when it will be distributed to the target audience. Additionally, the budget that should be allocated in different campaigns is a major decision to be made in this stage. At the completion of the campaign, it is essential to decide if a campaign was successful and how much contributed to the achievement of the primary goal.

In our modern information and technology society, every day users receive multiple advertising messages, but only a few hold their attention. Digital marketing contrary to traditional marketing can be measured and collect data about the campaign performance. We can have access to information such as how many users were exposed to the campaign, how many clicked and how many finally converted. Also, we can discover when they are willing to receive the message, what content interests them and how they engage with it. Reaching the right customer, in the right moment with the right message is very important. This fact gives marketers the advantage to target specific groups and offer them personalized and enhanced experiences better than ever before. [3] Social media is a great tool of collecting such data, as people use social media in order to read the news, interact with their friends, search about others' opinion and share their own. All this information can be exploited and contribute to the optimization of campaign performance.

Digital transformation and exponential technology disruption created many chances, opportunities but also challenges regarding to marketing. The problem is that most of the times the collection of the data is a difficult procedure, while the analysis and interpretation require skills and great effort. Data could be derived also from the behavior that users have when they visit a website, watch a promotion video or write a review on social media. The huge amount of data and the multiple different sources create the problem of information overload. Marketers are responsible not only to advertise a product but also to monitor the campaigns' performance and decide which data is important for further analysis. In this way they can create a strategy plan and make decisions based on real data and previous results.

Today advertisers have more than 7.000 tools and marketing technologies to collect and analyze data. [3] Most of these tools can be used in an easy way and don't require programming skills. The main goal of this research is to use one of these tools in order to handle, visualize, interpret the data and finally decide on which actions should be made in order to optimize the campaign performance. The purpose of the analysis of two different datasets is to highlight how important is to have the whole picture of customer's online journey and integrate all the useful available information to the decision process, in order to enhance the performance of digital marketing campaigns.

This work is organized as follows:

The first section is a search that has been conducted based on two literature questions. The first is about the advantages that digital marketing has over traditional marketing, while the second is about the contribution of big data analytics in the decision-making process. The next section focuses on the methodology and tools that have been used. The third section consists of the main analysis. Then, general recommendations, achievements and limitations of this analysis are discussed. At the end, the conclusion and future work of our research is presented.

2 Literature Review

2.1 Digital Marketing vs Traditional Marketing

Digital marketing is the marketing of products and services using mostly online digital channels. [4] It is also famous as "online marketing", "internet marketing" and "web marketing". The term first appeared in the 1990s through advertising for consumers on the web. In 1993, the first web-ad banner went live. [5] In the 2000s, online advertisement evolved because of the advancement of mobile applications and new technologies. In 2010, digital marketing became bigger due to the introduction of social media platforms. [6]

Today, people use all the more digital platforms and devices instead of physical shops. The internet is the place with the biggest audience to promote, buy and sell goods and services. According to CISCO's Visual Network Index, until 2022 the 60% of the population will be Internet users and more than 20 billion devices will be connected to the Internet. [7] Online platforms, mobile applications, eshops and other digital media are all parts of our everyday life. [8] The last years, new online organizations, such as Amazon, eBay, Lulu, YouTube and Netflix have been emerged and now dominate the market. *"As of 2018, 57 percent of global internet users had purchased fashion-related products through the internet, making apparel the most popular online shopping category worldwide. Footwear was ranked second with a 47 percent online purchase reach.*" (Statista 2019) [9] Digital marketing also brought to life innovative platforms such as Kickstarter. Kickstarter is a crowdfunding page where customers have the possibility to support the seller in the early development stage and launch the product successfully.

The combination of technology and marketing gave a big advancement to product development [10] New smart, automated and connected products have been developed the last years. The IoT (Internet of things) brought new innovative, automated and data-centric products. Those products are connected to the Internet and constitute a great source of real time data. [11]

Digital marketing is more easily measured compared to traditional marketing. [12] In the past, the procedure of collecting and analyzing data was an extremely difficult task. [13] Additionally, in digital marketing with the use of tools and techniques, we can easily anticipate future risks and even prevent them before they occur. [10] Having statistics and analytics in real time, gives managers the flexibility to make instant changes regarding their campaigns, because they can easily notice malfunctions. [14] "Performance marketing" is a trial and error procedure which tries to continually make improvements to bring better results by monitoring the customers. Knowing information such as clicks, page views, interactions, etc., help us develop a clear view about users' online journey and interests. [15]

In digital era, people are bombarded with ads, messages and promotion emails every day. In reality people seek for an authentic and unique experience, and not for more advertisements. The competition is extremely high, the customers have more options available contrary to the past and as a result, the purchase procedure is more complicated than ever before.

The path to purchase journey

Awareness: When a person has the first interaction with the brand. It is the stage where the user first sees a product and yet doesn't have the desire to buy it. In this stage, the interactions with the users serve to give information, highlight the strengths of the product and not try to sell it.

Consideration: At this stage the person has the need and considers buying a specific product. It is the moment when starts to compare products and tries to find more information.

Conversion: In this stage the potential customer is ready to buy. It is the right moment to promote more and tempt him with discounts or other tactics.

Loyalty: After the purchase, businesses should continue to show to their customers that they care about them. In this stage they should offer their help and opinion. In that way, customers will feel satisfied and they will have the desire to buy again. Additionally, they will become loyal and brand advocates.

Advocacy: Satisfied customers tend to recommend their favorite brand to others, give supportive feedback and a lot of times share their experience through social media. Having a hole community with people that supports your brand, is the best advertisement. [16]

Other differences between digital marketing and traditional marketing

- Traditional marketing used to communicate a message to the customer in one-way dissemination. On the other hand, the use of the internet gives the opportunity to companies to spread their message in a two-way communication. The online users can interact with their favorite brand and even contribute to it. [17]
- In the past, it was difficult for a small business to face the competition, but now with digital marketing even small companies with low budget can attract the correct audience. [14]
- Traditional marketing required high budget and many resources for commercials, (e.g TV ads, radio, print media, etc.) contrary to digital marketing which is far more cost-effective.
- Digital marketing can be displayed to people in every corner in the world without limits as far as people have access to the internet. [14]
- Today the user has the ability to compare easily products from different suppliers and make the best decision. [14]
- The customer cannot touch the product, so may hesitate to buy it contrary to the purchase in a physical store, where customer can interact with the product [18]
- Now with big data analytics, managers can have insights about customers' social behavior, an information that couldn't be acquired with classical questionnaires. [10] The effective analysis of customers' behavior and interaction is a key requirement to optimize advertising campaigns and budget allocation. [2]
- In the new digital era, customers have to be in the center of everything. Also, it is a fact that users trust the opinion of other users more than the advertisements. [19] Social media, which didn't exist before, is an easy way to let your customers share their opinion with other users without spend much. [2]
- Branding is different in the digital era. The brand represents the whole experience of the user and not only a logo. [20] There are communities where businesses and customers come really close, they share experiences and they enhance their relationship creating better results for both sides. [21]

In a few words, Digital marketing is more effective and efficient, as it is more user-focused, interactive, ubiquitous, agile, personalized and it can be easily measured. Digital marketing is more complex than traditional marketing. That is the reason why techniques such as search engine optimization (SEO), content marketing and e-commerce marketing are necessities to marketing strategies. [21]

2.2 Digital Marketing Techniques

Many digital marketing practices and techniques have been developed in order to enhance marketing, expand brand awareness and increase conversions. It is difficult for a company to find only one effective technique, as the best results usually come from a combination of these methods. Using the right combination, will help the company to develop a competitive advantage and create a great impact on the audience. Each method aims to a different goal, so before the adoption of any technique, the managers should define their goals and their digital marketing strategy. [22]

2.2.1 Search Engine Optimization (SEO)

Search engine optimization is the process that tries to increase the site's volume and quality traffic through the quality of the site's content. The first step is to define a list of keywords and phrases which is possible to be typed by the user in a search engine. The next step is to get the page into the search engine index. The third step includes the optimization of SEO (e.g the use of metatags). According to Weideman 2009, 67% of users who search to find an information on the internet, read only the sources which are placed in the first SERP (Search Engine Result Page). This fact proves how important is the SEO in a marketing campaign. In general, SEO is the best way to bring organic visits to the page and also one of the most-effective strategies. [14] [23]

2.2.2 Search Engine Marketing (SEM)

Search engine marketing is the strategy to bring traffic to the page through paid work. This method is also known as Paid Search Marketing (SEM), Pay-pair-click (PPC) and cost-per-click (CPC). The strategy includes display advertisement, Search retargeting, Paid social advertising etc. The most famous platform for this job is Google Ad Words. *"63% of people said they'd click on a Google ad.*" (Search Engine Land, 2019) [24] SEM is very important especially for new companies which don't have yet organic advantage and loyal audience.

The procedure of this technique is quite easy. The user types a keyword or a phrase in a search engine and the results are displayed in a specific order. The order of the results depends on the amount of money that the companies offered for that specific term. If many companies have paid for this specific term, an action takes place and the results on the top are from the companies which offered the most. The paid results usually are above the organic results and mention that they are paid advertisements. PPC could become a costly strategy as the keywords could be used from many advertisers and the competition could be extremely high. [14], [23]

2.2.3 Content Creation

Content can be displayed in many forms such as videos, photos, articles in blogs etc. The content is the best way to brand your business and create trust among the customers. It is very important to use different content for different devices. For example, the content for mobile and desktop should be different, because mobile screens have limited space. Also, content in different platforms should be different because each platform is addressed to different targeting groups. The content on Facebook could be more casual, contrary to the content on LinkedIn which should be more formal. [14] Last but not least, content addressed to audiences with diverse demographic characteristics and interests, should be different for each separate group, even in the same platform.

2.2.4 Personalization

Over the last decades, new companies have been emerged like Alibaba, eBay and Netflix which focused their strategy in creating a special relationship with the customer and became the drivers of the modern economy. [25] Personalization is one of the most significant tactics to retain your users and keep them satisfied. Intimacy is very important for the customers. They love to receive recommendations based on their interests, be remembered on their birthdays and feel special. [4] *"Mapping the customer experience and use of personalized content are deemed the most effective tactics for optimizing marketing automation."* (Ascend2, 2018) (Source: https://www.hubspot.com/marketing-statistics) [24]

Personalization in display advertisement increases the click through rate. [26] Today, brands have access to data regarding to customers' behavior and preferences. The growth of social media created the opportunity to exploit information about users' interests and use this information in order to profile users and create personas. [1] A persona is a representation of a group of customers who share same characteristics and interests. Using personas can help marketers to deliver personalized messages to the right audience at the right moment.

2.2.5 Social media

Social media is an extremely useful tool which can lead to brand construction, effective communications, better performance, crowdsourcing and word-of mouth. Through social media a brand can be created, maintained and expanded. People use social media in order to communicate with their friends, search about others' opinion or share their own. (Baily et all) [8], [27] On the other hand, for a marketer social media is a tool to understand how people feel and think about a brand. Social media and other external sources can be used as a source to organize the marketing budget and mitigate the risk to address resources to users who aren't interested. [28] Other contributions of social media includes: Opinion mining, targeted advertisement, competitive advantage, personalized responses and customer relations [17]

The problem with social media is that we cannot gauge their extract contribution on the final purchase. The most common mistake is when we analyze a purchase, considering only the last click and ignoring that the customer has an online journey before the stage of the purchase. Additionally, bad criticism and recommendation written on social media can be seen by everyone. According to Chen, Wang and Xie 2011, an observational experiment was conducted in Amazon.com. During the years 2005 and 2006 researchers were collecting information about the impact of eWOM and the result revealed that the negative eWOM affects more users than the positive eWOM. [2], [29] [25]

2.2.6 Mobile Marketing

Mobile Marketing takes place in mobile devices. The strategy's aim is to communicate the right message to the right people in the right location. [14]

2.2.7 Viral Marketing

Viral Marketing is another way to make your brand famous or bring traffic to your site. Viral marketing strategy is to spread a unique and creative content which will be liked and shared from many people through social media in a short period of time. [14]

2.2.8 Email Marketing

Email marketing is the way to approach customers through sending them a commercial content via emails. It is important to display to the users personalized content, which fits their interests, and not irrelevant ads. [14]

2.2.9 Display Marketing

Display marketing use display advertising in many formats in order to target potential customers. Those formats could be banners, text and video commercials etc. According to Match Craft, 2018, *"The most popular display ad types are banners, native ads, and social media* *ads"*. (Source: https://www.hubspot.com/marketing-statistics) More than 5 trillion banner ads are created each year. [24]

Display advertisement is a complex procedure, as it involves many intermediate parties, end users, publishers, ad platforms etc., all of which generate a large number of data. [30] Usually, publishers offer space on their websites and advertisers give money in order to place their advertisement on this space. Advertisers could be entrepreneurs, who try to develop and expand an innovative idea, or big multinational companies which target a bid audience in many geolocations. [31] [Figure 2.1]

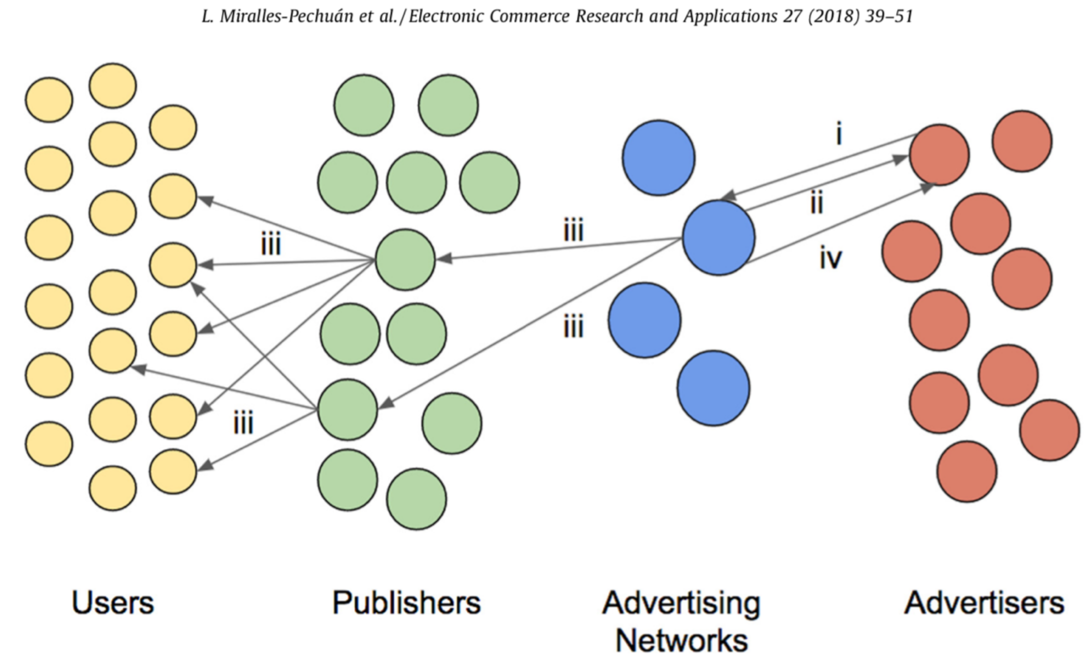


Figure 2.1 Main steps of the process of launching an online campaign (Source: 81)

Advertisements are displayed every day in websites and social media. Facebook display advertisement let us track impressions, click through rates and conversion rates [3]

2.2.10 Retargeting

The visitors not always have the intend to buy something when they visit a site. Also, even if they want to buy something, they don't usually purchase it in their first visit. Usually, they need time to think and revisit the site a couple of times before make their finally decision. [32]

Retargeting is focused on people who already know the brand. Commercial and promotion coupons will be delivered to already visitors, in order to remind and encourage them to finally finish their purchase.

2.2.11 Affiliate Marketing

In affiliate marketing publishers offer space in their website, in order to advertise another company. This type of marketing is quite useful for small and new companies with low traffic. The reference to visit a start-up company from another well-known company will create trust and more effective results. [14]

2.2.12 Blog – Influence marketing

A blog is created by a person who shares his thoughts, reactions and ideas in public. The last years, it is very common for brands to be advertised by bloggers and influencers. People don't feel that this is an advertisement, but a recommendation from a person that they like and trust. [33]

2.2.13 Permission marketing

Permission marketing was introduced by Seth Godin. The idea of this concept is that the customer has to give first permission in order to receive commercials. In that way brands give the perception that care about the desire of the customers and not just try to allure them. [20]

2.3 Big data analytics and their contribution in business decision making

Nowadays, Internet users can express their opinion through online platforms, websites, forums or blogs without any constraint. Analytics had their first appear back to 1940s and started to attract more attention during 1960s, when computers started to participate in the decision support systems. Data science appeared in 2001, when Analytics 2.0 era begun. [34], [35] Big data analytics, which is used in different domains, is a useful set of methods, technologies and other tools that contribute to the generation, collection, analysis and interpretation of a big amount of data. [36], [13], [34] The advancement of technologies and the big variety of available tools give businesses the opportunity to grow very fast, exploit new technologies, have access to many sources of useful data and create insights based on them. What makes data such a valuable asset is its hidden information. [13] Web analytics is under the umbrella of Big data analytics. The

WA industry begun in the middle of 90s from the companies Webtrends, Omniture and NetGenesis (Web Analytics Association, 2010)

Web Analytics is used in order to understand and optimize the web usage. According to Waisberg and Kaushik 2009 [37], the term can also be defined as the tool that is used for the improvement of a site's persuasion and relevancy in order to achieve better conversion rates. *"60% of professionals support that analytics increases the sales' velocity".* [38] Although WA is an important and useful tool for performance measurement, often tends to be useless without the proper analysis. [37] Google analytics is one of the most used WA tools, as it can be used free of charge. WA analytics have enhanced the digital marketing performance and increased the revenue coming from sales. [39]

2.3.1 Benefits of web analytics

- Analyze information coming from a digital visitor such as pages viewed, interactions, number of downloaded files etc. This information gives to managers the opportunity to profile their customers and have a clear view about their activities and preferences. Buzz monitoring helps to explore someone's opinion and have a better understanding about his preferences. [40] Then, different customer segments can be generated based on geo-location, user profiling or other users' specific characteristics.
- Calculate site's traffic and get insights about trends, for example how often a user purchase and why.
- Gauge the performance of online campaigns and track customers' online journey. This gives to the managers the possibility to detect their possible errors, site malfunctions or other bottlenecks on navigation. For example, according to [40], the analysis of a Brazilian shoe e-shop revealed that visitors who don't speak Latin languages gave up extremely early in the first page. The problem finally was that the language button wasn't in a visible position and the visitors couldn't continue easily their navigation. [40]
- Contributes to cyber security and complements web fraud detection.
- WA collects other information about the transaction such as the browser type, endpoint type, the user's language, country etc. Also, they can recognize if a visitor is a bot or a human. For that purpose, CAPTCHA is used in order to decrease bot interactions. [41]
- The collection of the data can be standardized and automated without a lot of effort. [42]

- Understanding your customers and their behavior assist managers to set the correct price for their products and services.
- It can cooperate and be integrated in other enterprise tools and software such as CRM and social software (Digital Marketing Depot, 2014)

Tag manager

Another web analytics strategy is to implement the functionality of tag manager into the page. A tag is a JavaScript code implemented in page's html file, in order to collect data and revert it to information about visitor's behavior. [43] Tag manager doesn't collect data about the page performance, as Google Analytics does, but can track specific moves a user makes, for example clicks to download a file from the site and clicks to play a video. So, we can have multiple tags tracking different actions on a webpage, which can be activated only when the user makes an action or has a specific characteristic.

Having access to such information, helps marketers to know exactly what problems exist in each webpage and make improvements accordingly. Also, managers who understand the exact visitor's journey can create focused advertisements, personalized content and make better decisions on where to spend the marketing budget. [44]

2.3.2 Web Analytics Challenges

Web analytics is a useful tool which also has its limitations and disadvantages. First of all, it cannot calculate the exact return of investments. Digital marketing provokes purchases not only online but also offline. The problem is that there isn't a way to match offline purchases with a specific digital marketing campaign. Also, data scientists although they have quantitative and analytical skills, they usually lack marketing background. [2] Others, even if they have marketing background, they lack training and knowledge in web tools and technologies. [45] Both fields, data science and digital marketing, are very important and most of the time is difficult to find someone expert in both.

2.3.3 Big Data Analytics techniques

Analytics explore past and current performance, in order to get insights from a big amount of data and contribute to decision making and planning. [46], [13] Big data analytics can be applied in almost any area and have many business applications. [13] Many offline marketing activities have digital elements that can be analyzed by analytics. Big data analytics consists of methodologies such as cluster analysis, decision trees, Bayesian analysis, filtering, regression analysis etc. [13], [47]

Big data is described by 4 characteristics, immense Volume, high Velocity, Veracity and big Variety. [28] [48] The emergence of big data provokes management changes as the managers found in a position that they had to collect, process and analyze a big amount of data in order to get the knowledge. [28] Usually, data is complex, difficult to be processed and demands adequate technology - tools. According to Davenport et al., 2014 *"You cannot be analytical without data, and you cannot be really good at analytics without really good data"*. Redman, 2013 writes that data is qualitative when it is accurate, relevant, complete precise and valid. [49] [28] [2]

Big data derives from traditional transactions or other unstructured sources such as social media, sensors, videos, emails, loyalty cards, coupons etc. [50] More than a billion of people all around the world use every day social media platforms and create a huge amount of unstructured data. [51] In order to manage all this unstructured data, many data mining and machine learning techniques have been developed the last years.

Descriptive analytics

Descriptive analytics is the category of analytics that is used by most of the businesses. (IBM, 2012) [52] [46] Using descriptive analytics, managers can have answers about what happened in the past. The analysis of historically data lead to have a deeper understanding of what caused failure or was successful in the past. This can be extremely useful for ongoing activities. Descriptive analytics mostly is used for reporting, sales monitoring and identifying correlations among the data [51], [53]. The results are more often displayed and interpreted via graphics and charts [13]

Diagnostic analytics

Diagnostic analytics explore data in order to find out what caused an event. In that way, organizations acquire important information about complex problems which may help them later to find solutions.

Predictive analytics

Predictive analytics, after historically data analysis, make assumptions about future actions. It is actually the next step of descriptive analytics. This category uses statistical analysis, machine learning, data mining, game theory, forecasts and other predictive models to detect possible relationships between elements and predict future events. [28], [51] The common methods are regression, decision tree, Bayesian statistics etc. The data in predictive models is used to detect risks and opportunities. Also, a lot of times it is used to make recommendations [54]. Another use of predictive analytics is in the optimization of customer relationships. The procedure is to analyze customers' data and predict their future behavior based on their preferences. [51] , [55] Predictive analytics can for example analyze users' purchase history and other information and create personalized promotions, customer segments etc. [56] For example, a small coffee shop in New York saved 38% of its marketing costs because after the analysis of customers' data, made predictions and methodically approached the customers with personalized discounts. In that way, made them convert easily and retain them as loyal customers. [38] In Gartner's 2017-2018 CMO Survey, most of the leaders who were surveyed, said that the 2/3 of their budget is dedicated to customer retention. [57]

Prescriptive analytics

Prescriptive analytics is the last stage in data analysis. After the analysis of descriptive and predictive analytics, prescriptive analytics give recommendations about what should be done in order to optimize the predictions that have been made. Prescriptive analytics could give accurate predictions and accurate decision choices. [28] Based on prescriptive analytics, managers can exploit early future opportunities. Some of the techniques that are used in prescriptive analysis are mathematical programming, game theory and other simulation techniques. [51]

Natural Language Processing

Natural language processing, sentiment analysis and social network analysis are the big data analytics which were used in social media and enhanced the organizational decision making. Sentiment analysis is implied in order to understand what people feel about a certain topic through classification, polarity, targeting collection and aggregation process by automated tools. [58], [45], [59] Although, it is easy for humans to learn a language, it isn't so easy for a machine. The ambiguity of a language and the different slang words or meanings people use in their expressions, make difficult for a computer to process natural language. [51]

Segmentation

Analytics is used also in order to create customer segments, in other words create groups of users that appear to have similarities in their behavior (e.g Clicks, length of video watched etc.) or personal characteristics such as users' location, education, career, and income. The procedure to divide the customers in segments, is a prerequisite of a marketing campaigns. In that way, managers can have a clear image of users' intents and offer betters services. According to [43], social media could be a useful resource to collect information about the users and create different segments. Segmentation is the previous step of personalization. In general, different people need different treatment. Recommendations should be made based on users' experiences and characteristics. It is very important to create different and appropriate content for each different group. In that way the bond between the brand and the customer become stronger. [43] Also, each segment should be analyzed differently. This happens because the page's visitors have different behavior and different needs. So, an average metric representing the behavior of all these visitors is not realistic. [32] Tesco's data collection system which integrated information from market research, websites, loyalty cards etc., in order to create segments, was the key to make the company number one retailer company in UK. (Humby, Hunt & Philips, 2008) [60], [2]

2.3.4 Big Data Analytics Challenges

An important challenge is that data sources usually contain noise and inconsistencies. Entry duplications, security and data irregularities are some of the problems. A lot of users don't have specific instructions and enter incorrect data or make mistakes. Also, the transfer of data from one system to another is possible to provoke loss of significant information and have bad impact on the results. [28], [51]

Complex data and data quality

Many users create unstructured and fuzzy content through social media, which includes usually many forms of data such as text, images and videos. All these data form user's personal point of view, makes the extraction of high-quality information significantly important. The effort to analyze complex data with low quality is always bigger and time consuming. [51]

A data quality problem is the endogeneity problem. This problem happens when data is treated as random while it isn't not. Datasets suffering from endogeneity, cannot generate trustworthy results. In that case, the addition of extra entries isn't a solution. Data of greater variety should be added in the analysis. [56] Data quality is a prerequisite for good estimations.

Data availability

When machine learning techniques are applied, the whole dataset has to be accessible because the algorithm examines the training set and then create a predictive model. In case of having data streaming, this procedure isn't possible.

Data locality

The data is usually located in many files which are placed in different locations. The integration of data from many locations usually makes the analysis quite difficult. [51]

2.3.5 Better models

In order to implement predictive analytics, we need to take into consideration 5 major steps. First of all, define the primary goals and objectives of the organization. Then collect only the relevant data from all the possible sources and databases. The third and more important step is to enhance the quality of the data using data extraction and cleaning techniques. In this stage managers should fix duplicate records and normalize the data in order to ensure consistency. The next step is to decide how the data analysis will take place. Managers have the options to undertake the analysis and create their own models. This procedure requires expertise in data science and good knowledge of programming. Another solution could be to outsource the work to another company which has the specialty in the data science field. Nowadays, there are available many online software tools which may not be as effective as a professional data analyst would be but are easy to use and can be a good start for a new small business. Last but not least, validating and evaluating the produced model is extremely important at the end, because this process errors and mistakes could be detected and fixed. [38]

Large databases demand greater effort, are time consuming and more costly in their analysis. Nowadays, there are a couple of methods which can lead to compression of the size of a dataset without information loss. A smaller dataset provides the same information and allows the creation of better models or methods to be developed. In general, the reduction of the dimensionality of the data is a hot topic in marketing literature and has been proposed from many authors. Bayesian methods supports the creation of smaller datasets with no losses. They have the ability to allow the users to update parameters at any point without the necessity to rerun the algorithm [56]

The role of data scientists includes data preparation and analysis. From this exploration they should be able to draw conclusions from the hypothesis that are produced during the analysis. As the data increases exponentially, it is difficult for a person to examine all the potential scenarios. According to Gartner, 2018 By 2020, *"More than 40% of data science tasks will be automated"*. Automated algorithms will be created with the capacity to explore all the hypothesis, spot easily the hidden patterns and make conclusions without personal bias. Given that, the productivity will be increased and the data insights won't suffer from bias.

2.3.6 Data mining

Data mining is known as "knowledge discovery in databases" and is the procedure of discovering interesting patterns that could be analyzed and be useful in decision making. [61], [62] Data mining integrates machine learning, statistical analysis and other visualization techniques. Machine learning is a method that automates the procedure of knowledge from examples [63], [61]. Machine learning techniques which can handle many data types, such as ordinal, nominal etc., can be extremely useful for data mining approaches. [61] In the field of marketing those techniques are used in the following applications:

Market segmentation: Divide the target audience into smaller groups of customers which have similar characteristics or preferences.

Lifestyle Behavior analysis: Understand consumers' behavior and their intends. Investigate their activities and opinions.

Market basket analysis: Analyze the combinations of products which bought together and use them in future promotions to attract more consumers.

Customer reaction to promotions: Find out if a promotion had the results you thought it would bring. [61]

Problem types

- **Classification** is the identification of attributes that are related with classes or clusters. As soon as the training finishes and the classes have been identified, new unknown entries can be analyzed and categorized as well. This procedure can mitigate the risk in many procedures and marketing activities. [61]

- **Prediction** is the procedure to find future values based on data that already have been processed. For example, predict product performance and forecast mistakes and risks of future activities [61]
- **Association** is the identification of relationships that exist in groups. An example is the market basket analysis. Association rules usually supports product bundling or decisions about combination of products for placement or promotion. [59]
- **Detection** is the detection of outlier, irregularities and elements with a deviated behavior.
- **Cluster analysis** is a data mining procedure, which tries to detect groups of elements that have similar characteristics or behavior [59]

2.3.7 Decision making based on data

According to "The Future of Jobs Report 2018" by the World Economic Forum, problem-solving is one of the most significant skills that someone has to possess. [64] Also, the ability to identify key business decisions and to be able to give recommendations on complex analytical problems are skills with high demand in our days. [59] For that reason, during the last years, many Decision Support Systems have been developed. Those systems are computer – information systems which use specific decision rule models and decision maker's insights [65].

Data driven decisions help managers to detect problems faster and more fact-based. [64] Big Data Analytics support marketing decisions and improve processes. Managers should also understand the cause of the events and current trends. Also, not only they should be able to know if a strategy works, but also the reason why it works in order to be capable to recognize why this specific solution will work or not in another different situation. [56] In other words, they should be able to interpret the analytical results but also develop their critical thinking. [56] Classical marketing analytics lack of the ability to gain insights useful that could be used in order to solve complicated problems. [66] On the other hand, Big Data Analytics exist for this purpose. Decision making is supported by big data analytics and business intelligence. Those tools have the ability to reveal customers' needs, motives and understand their lifecycle. Additionally, they can foresee future problems and difficulties and prevent them before they occur. [10]

The biggest barrier for managers is to interpret the results from analytics and integrate them in the decision making. [67] Although web analytics offer objective, standardized and quantitative metrics, many marketing managers continue to rely on their experience and intuition, ignoring available information when they make decisions. [68], [34] Also, another critical point is the allocation of tasks related to Analytics. Managers should be able to share properly the information and the insights coming from analytics and cooperate with their colleagues. [67] The target is to make intelligent decisions, observe hidden values, remove limitations and manage the big data properly. Visualizations of data is a way to spot meaningful patterns, outliers and hypothesis that cannot easily be detected in standard statistical methods as they include colors, texture, brightness and have specific data and information highlighted [69], [61] Visualizations and predictive models, cooperate in order to improve the quality of knowledge that comes from the data analysis. [59]

First thing that companies should do before the collection of data, is to organize strategical methods for the data management. [13], [62] According to Chaffey & Patron, 2012, managers should start from the identification of key performance indicators before start to analyze the Web or Big data metrics. [70] Selecting the appropriate KPIs is very important because each company has each own goals and tactics, thus there aren't any standards to fit the needs of all organizations. [67] After the selection of the most appropriate KPIs, the metrics can be evaluated based on their relationship with the selected KPIs. By prioritizing the metrics, managers are able to focus their attention on the most significant objectives and avoid the information overloaded. [67] KPIs in order to be successful should be simple, map to key business activities, actionable, reliable and timely (SMART). [59]

Decision making should be based on the analysis of big data. The three main steps required for the decision making are: 1st, be able to explain the current situation (descriptive analysis), 2nd based on the current situation, predict future results (predictive analysis) and 3rd suggest improvements for better decision making (prescriptive analysis). [65] These steps aren't only applied in Digital marketing but in all sectors.

Budget Allocation

Budget distribution is one of the most difficult decision problems that managers have to decide. The difficulty to allocate effectively the budget is greater in case of limited budgets. It is very important to allocate the budget wisely in different campaigns and ad-sets, in different time periods and track the performance in real time.

Campaigns are at the top and they contain the total budget. This budget should be divided and distributed to different campaigns, ad-sets or ads which target specific audiences. [Figure 2.2] [1]

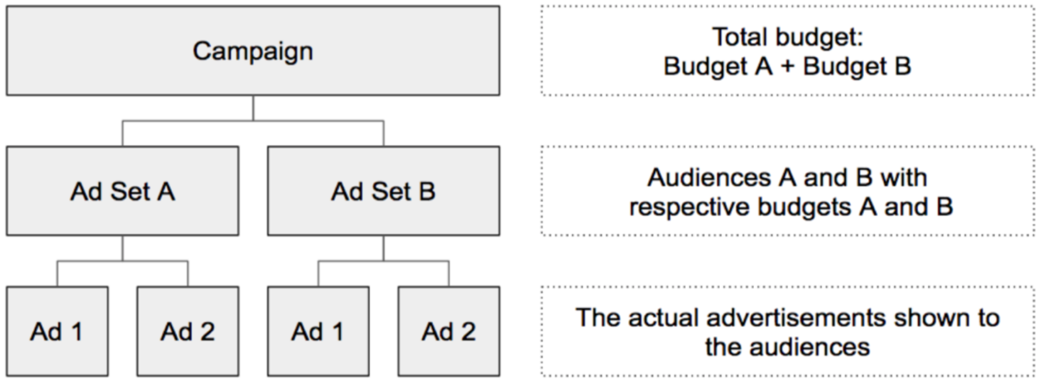


Figure 2.2 Structure of display an advertising campaign (Source: 1)

The problems that may distract the effective budget allocation could be budget constraints and lack of expertise and time. Additionally, the underlying mechanism of the bidding process is quite complex. Also, marketers usually have to deal with many campaigns, different markets and distribution channels. [71] Last but not least, the solution isn't always to spend more money on advertisement, in order to make profit.

The advertiser set the campaigns' goal. In most cases this goal is the generation of revenue. Facebook data assist in the analysis of how much money spent in each stage of the conversion funnel. In [Figure 2.3], we see an example of an ad displayed in Facebook. This ad displayed to 100 users, clicked 10 times and finally lead to two conversions.

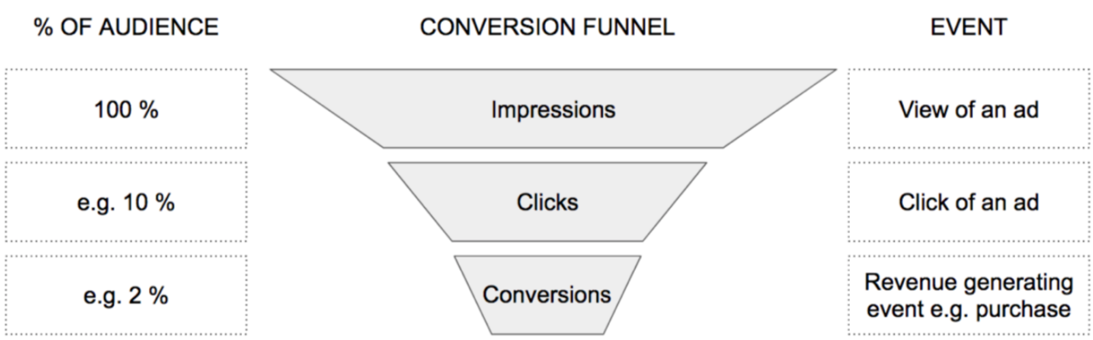


Figure 2.3 Conversion funnel in display advertising (Source: 1)

According to Tull et al.1986 and Fiscer et al. 2011, instead of increasing the total budget, allocating effectively the budget among different products and regions can lead to better results. The budget allocation strategy should be based on the average click through rates, the target audience, competitors' strategies and the effectiveness of the website. Additionally, the budget should be allocated based on the business size, the results of the marketing activities and the potential growth. [71] Allocate the budget among different keywords, choose a bidding strategy, decide on a daily or monthly budget and differentiate the budget on different target groups are also good strategies. [72]

Business Intelligence

BI enables the analysis of information to enhance processes and produce better outcomes. It is used more often in the decision making of issues about supply chain, sales, marketing and finance [73,74,75] The advancement of BI is that offers good visualizations which supports the produced insights and make it easier to understand them. BI requires skills, technologies and good data. The main task of BI tools is to reverse data into useful information. During the last years, BI has been affected a lot from faster analytics and Big data. However, the research shows that there are many opportunities yet undiscovered. [76] BI systems are complex and usually consist of many components which have to be integrated all together.

Examples found according to the literature investigation

In [77] they try to develop a regressed-based model that makes predictions about customers lifetime value. The current customers' value may not be useful. Instead potential customer value can help managers to make better decisions and manage the budget for each segment accordingly.

Two company activity examples are referred to [67] "Steel company" and "Machinery and Paper". The first one after implementing analytics was able to measure Digital marketing outcomes, be aware of how effective social media are in relation with the campaign and last but not least, they finally had a better understanding of what their customers want. On the other hand, the 2nd company used analytics as a supplementary tool for decision making when they wanted to know the performance of specific actions, for example if a video on the welcome page attracts the audience. [67]

3 Problem Definition, Materials and Methods

3.1 Introduction

Finding the right dataset and the most appropriate tool was a difficult process in our analysis. This chapter provides information about the method that was used to conclude to the right materials and tools. It gives a detail description of the steps that were followed until the step of experiment implementation, which will be discussed in the 4th chapter. This part was the most time-consuming part of this study. First, the research questions of the literature review are discussed. Then we present the datasets that we tried to use and which of them we finally included to our research. At the end of this chapter we focus on the tools that we examined and the one that we concluded to use.

3.2 Literature Review and Research questions

At the begging, a research about digital marketing and big data analytics has been conducted. The libraries that we used are: IEEE Xplore, ScienceDirect, and Google Scholar. The research is based on two research questions. The first one is: "*What are the differences between traditional and digital marketing*". Keywords such as: "digital marketing" and traditional marketing" were used. Analyzing this question, gave us a great chance to find the differences and the alterations that happened in marketing during the last decades. Through this process we present the new tools that digital marketing uses today, the online user's journey and the general benefits of new technologies.

The second research question is about "*Big data and their contribution in business decisions*". We used keywords such as "web analytics", "prescriptive analytics", "marketing campaigns", "data mining" and "decision making". First, through this question, we had the chance to explain the web analytics benefits and challenges. Then, to discuss about big data analytics techniques, data mining and other more complicated procedures. Last but not least, to focus on the advantage of taking into consideration the data analysis in business decision making and problem solving.

With this research we wanted to make an introduction about how useful digital marketing is today and the relationship that it has with other technologies. Today, marketers have to be eligible to use many tools and ground their decisions on data and previous behaviors.

3.3 The selection of the datasets

The second step was to find a proper dataset for the analysis. The ideal dataset could be a dataset that includes information about customer characteristics such as their purchase location, purchase history, demographics and other personal information such as interests. Also, we wanted to have data about the cost that different campaigns had.

3.3.1 Google Merchandise Store Dataset

At the begging, we tried to use the “Google Merchandise Store Dataset” because it includes Google Analytics data from a real online store. More specific, it includes information about users’ behavior when they use the site. For example, what pages the user visited before he finally purchases a product and how he interacted with the content. Also, it includes organic, paid and display traffic. Last but not least, it includes information about conversion and transactional data. A limitation is that the dataset can only be accessed via BigQuery platform. In Figure 3.1, we can see an example of running a query in Google BigQuery platform. This query gives us the number of the average revenue by user per visit. In general, we can run many queries and extract information about the dataset, but this procedure requires a good knowledge of databases and coding.

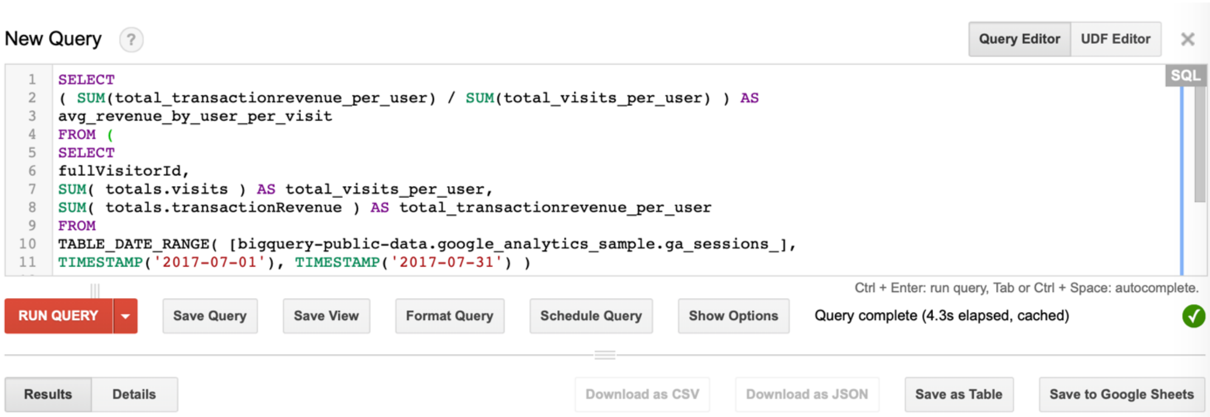


Figure 3.1 Query in Google BigQuery

Another limitation regarding to this dataset is that all users do not have administrative rights while using it. This means that the users can run queries, have results based on real data, create reports but they cannot process the dataset as administrators. Furthermore, changes such as removal of columns or normalization of the data cannot be made. Because of these limitations, the “Google Merchandise Store” dataset wasn’t used in this research.

3.3.2 Bombas’ dataset

The third solution was to use a dataset that was found in one of the papers that analyzed in first place. The dataset owner is the company Bombas, a startup company which manufactures socks. The novelty is that the company donates one pair of socks for each one purchased. Bombas founded in 2013 and is a crowdfunding project first launched in page “Indiegogo”. The dataset consists of 1099 registers and 13 columns.

The information about the columns:

- **Customer id**: The unique id of each user who made a purchase to the site.
- **State**: The State that the user made the purchase.
- **Postcode**: The postcode of the neighborhood that the user made the purchase.
- **Registration date**: The day that the user registered to the site for first time.
- **Last site visit**: The last time that the user visited the site.
- **Number of site visits**: How many times each user visited the site
- **Number of product views**: How many products each user checked.
- **Number of add to cart events**: How many times the user added products to the cart.
- **Number of checkouts**: How many times the user made a purchase.
- **Number of orders placed**: How many times each customer ordered.
- **Items purchased**: How many products each user bought.
- **Order 1 revenue**: The price that each customer paid for the order.
- **Order 1 purchase date**: The date that the purchase made.

The goal of the paper is to offer instructions to students and help them analyze an e-commerce organization to increase revenues by analyzing performance data. This example could help many businesses to make strategic growth decisions based on real data. The dataset was analyzed in order to extract information about users’ interaction with the site and give recommendations for better performance.

Information such as the location that most buyers have, the time, the day and the month that most purchases took place, was used in order to decide for strategic future plans. This information categories could be very useful sources for successful future campaigns and decisions that should be made.

The first part of the analysis is based on this dataset. The dataset as it mentioned above contains only information about users' interaction with the site. As we wanted to take into consideration also other information related to previous paid campaigns, we tried to find another dataset which will cover this part.

3.3.3 Sales Conversion Optimization dataset

After finding many datasets with insufficient information, we ended up with the "Sales Conversion Optimization" dataset, found in Kaggle website. This csv dataset contains information about some social media ad campaigns, conducted by an anonymous organization. It consists of 1143 observations and 11 variables. The exact description of the variables is below.

ad_id: the unique id for each different ad.

xyz_campaign_id: an id which is related to each different ad campaign of XYZ company.

fb_campaign_id: another id which indicates how actually FB tracks each different campaign.

Age: the age of the person who saw the ad

Gender: the gender of the person who saw the ad

Interest: a code which indicates the category of person's interests (interests are mentioned in user's personal profile and are separated based to their context in different categories)

Impressions: the actual number of times the ad was displayed.

Clicks: the number of clicks on this specific ad

Spent: the amount the company xyz paid to FB platform in order to display this ad

Total conversion: the number of people who wanted to learn more about the product after they saw the ad.

Approved conversion: the number of people who bought the product after they saw the ad.

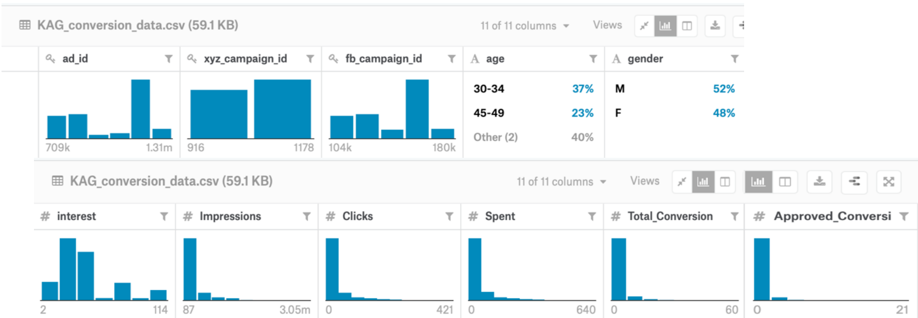


Figure 3.2 The columns of the Dataset "Sales Conversions Optimization"

(Source: https://www.kaggle.com/loveall/clicks-conversion-tracking)

3.3.4 Conclusion

At the begging, the goals of our work were to prove that marketers should also be good analysts. Also, to create solutions for predictions and give automatically recommendations about personalization and budget allocation improvements. Unfortunately, we couldn't find the ideal dataset in order to make predictions and automatically recommendations, so we decided to continue the analysis and focus on how important is for marketers to analyze information from previous campaigns and take it into consideration for future ones, using online tools that do not require programming skills.

After the research that we made on datasets we chose to analyze 2 of them, as it mentioned above. Those two are "*Bombas*" dataset and the "*Sales Conversion Optimization*" dataset. The first one helped us to extract some information regarding to users' interaction with the site and use it for further analysis and decision making. On the other hand, the second dataset, gave us the opportunity to analyze FB campaign's information for the same reason. Data coming from social media campaigns is a very significant part of customer's online journey and should be taken into consideration in decision making related to campaign strategy. After selecting the datasets, we made a research for data mining tools in order to decide which one we will be used in our research.

3.4 Public dataset Sources

In order to find a dataset, we searched many pages which offer public datasets for research and analysis. The pages that we used are the following:

kaggle.com: Kaggle is a webpage where anyone can have access to public datasets. Also, Kaggle is an environment where data scientists and machine learning engineers can collaborate, make research, create new machine learning models and take part in world-wide competitions in order to solve critical problems. On March 2017, Google acquired Kaggle.

data.gov: Data.gov was found in May 2009 and is a U.S Government website. The purpose of data.gov is to enhance the access to machine learning datasets related to federal, state and government information.

data.world: Data.world is another webpage which is open and free of charge. It is an online environment where anyone can find public datasets, make analysis and share it with the world.

archive.ics.uci.edu: The UCI Repository is a machine learning community that offers an online collection of datasets. It was founded in 1987 and from then it is used by students and researchers for machine learning analysis.

All the sources mentioned above, are online communities where people can use real data examples, make experiments and finally create new algorithms and patterns. Other sources that were used in order to find an appropriate dataset were github and online forums. Github is an environment which is used to host software development version control. Github also offers an online community forum where users can make questions and give answers to others' people problems. Multiple other forums were used to find an appropriate dataset, because were appeared after Google search. Searching to solve a problem in discussion forums, could be the best solution.

3.5 Data mining tools

The big amount of data requires techniques to be used for better analysis and conclusions. It is impossible for an individual to analyze the data without any help and make the right decisions. For this reason, during the last years, have been developed many open and free data mining tools on the web. According to [78] there isn't a data mining tool better that all the other. The performance and the results of an algorithm depends on the type of the dataset, the type of the algorithm and the reasons of the analysis.

3.5.1 Weka

Weka developed at the University of Waikato in New Zealand and first released in 1997. Weka is used for data mining and machine learning operations. For this purpose, it uses many algorithms written in Java. Many professors and universities all around the world use Weka in order to teach techniques such as data processing, classification, regression, clustering and visualization. [78], [79] Weka has three user interfaces, Explorer, Experimenter and Knowledge Flow. Explorer supports preprocessing, learning and visualization. In [Figure 3.3], we can see the user interface.

Disadvantages in Weka

- The files' format should be ARFF.
- The documentation is poor and not up to date.

- It has scaling problems, as it takes much time to run processes when the dataset is big.

[80

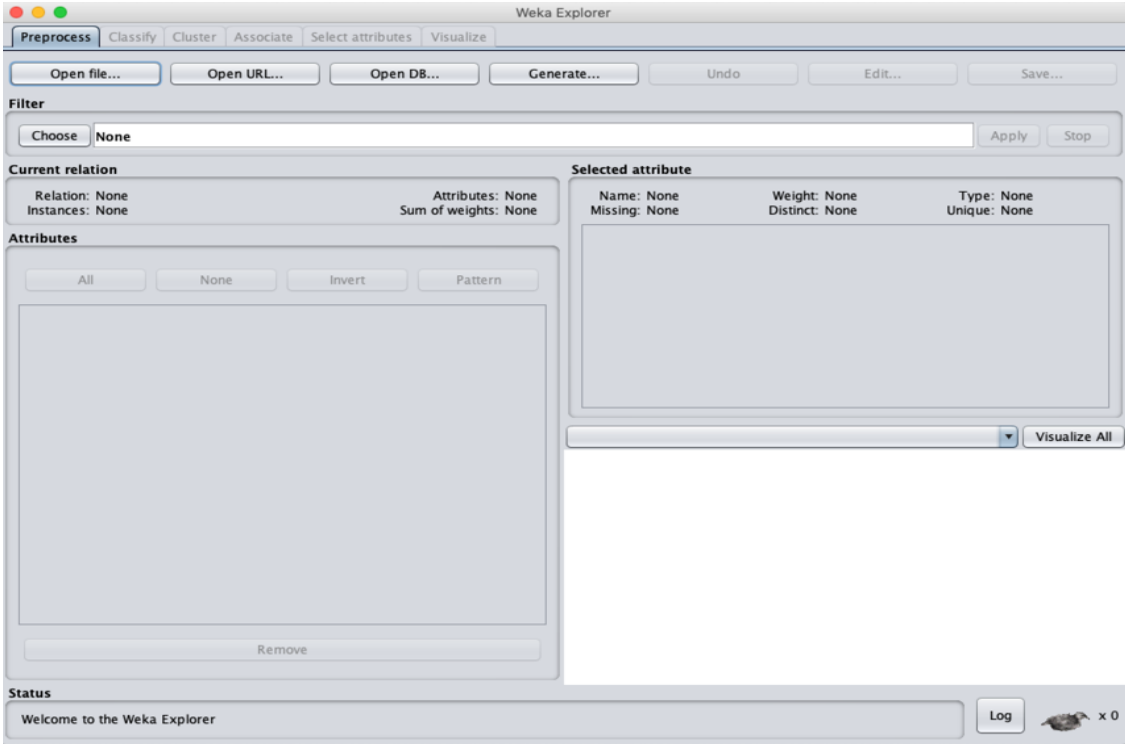


Figure 3.3 Weka – Explorer Interface

3.5.2 Orange

Orange is a data mining and machine learning open source package, which also can be used as a Python library. It developed by Faculty of Computer and Information Science at University of Ljubljana in 2009. It has many algorithms which are used for data processing and manipulation. Decision trees, bagging and boosting are a couple of tasks that can be executed using the toolkit Orange. In general, is an easy tool for users and researchers who want to test their hypothesis and make conclusions. [78], [79] Orange as we can see in [Figure 3.4], has a friendly interface where the user can design easily a model.

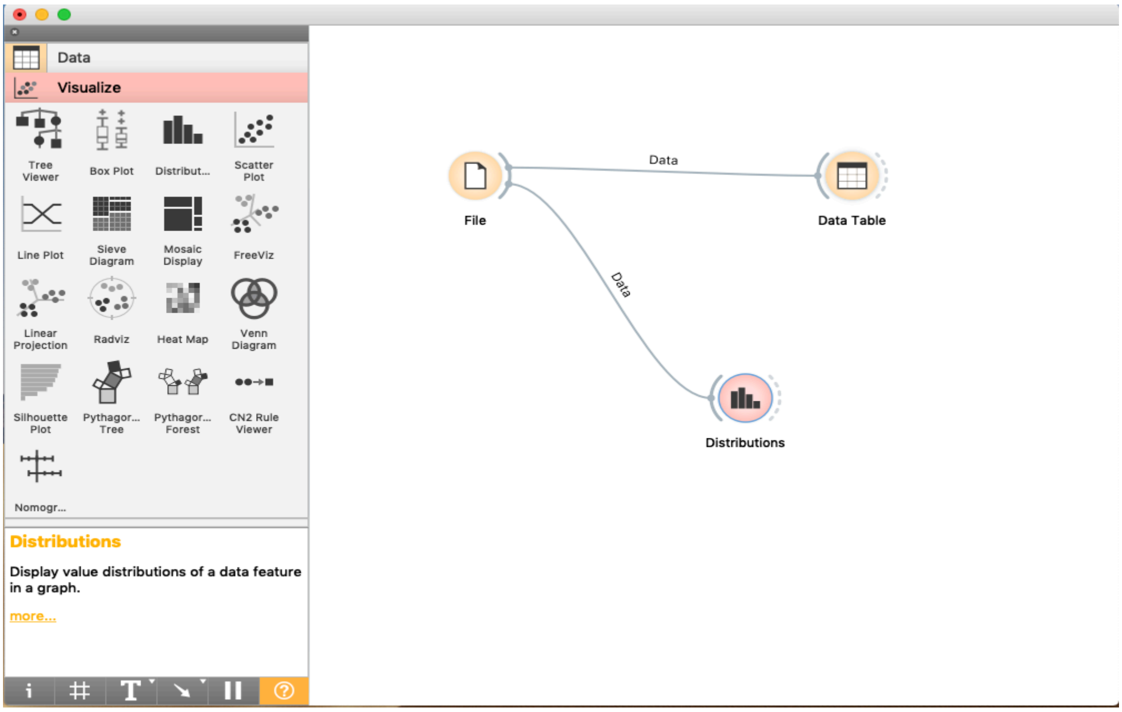


Figure 3.4 Orange Interface

3.5.3 RapidMiner

RapidMiner is also a data science platform with more than 1.500 native algorithms and data science functions. The development started back in 2001 and was known as YALE (Yet Another Learning Environment), in the University of Dortmund. RapidMiner is a free and open source platform which supports data mining and machine learning tasks such as regression, classification and clustering. It has the capability to design and operate complex machine learning problems.[79]

RapidMiner offers a friendly environment to the user. Additionally, it has a proper and adequate documentation with many tutorials. Last but not least, it helps the user as it displays recommendations for the next steps and suggestions what parameters to change, in case of mistake. [80]

After comparing the three machine learning tools, we concluded to use RapidMiner because it can be used easily from a Marketer and it hasn't got many limitations.

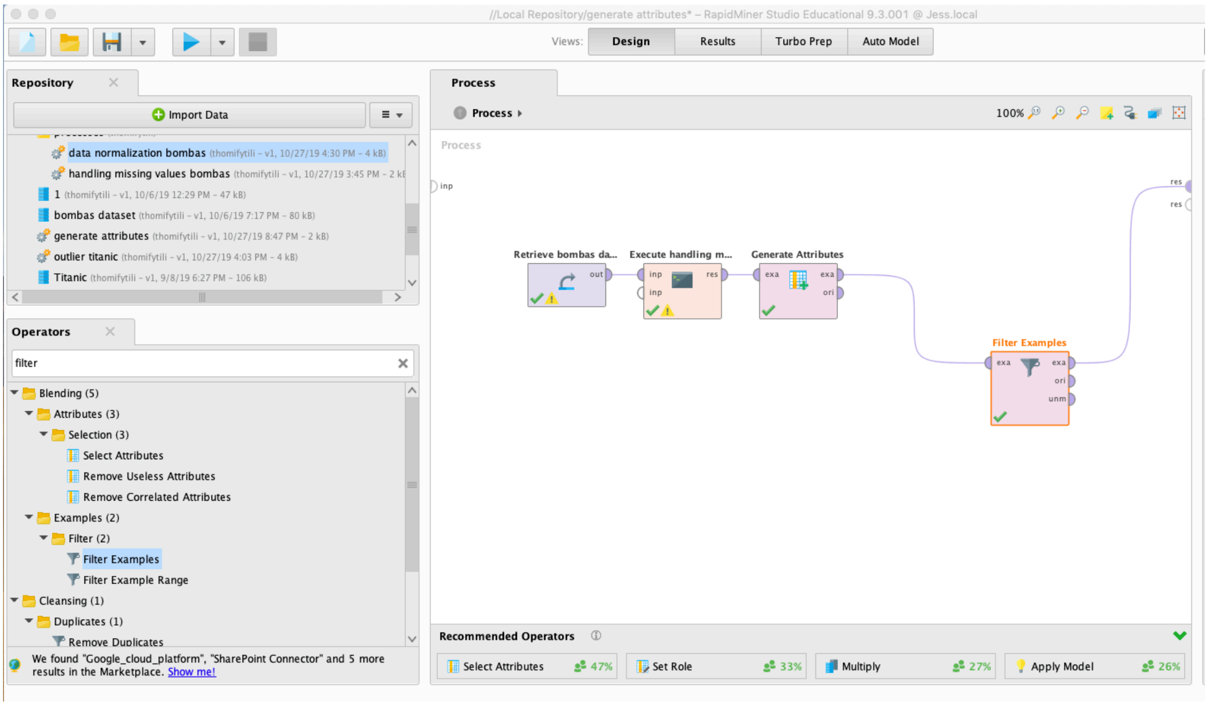


Figure 3.5 Rapid Miner Interface

4 Implementation & Results

4.1 Introduction

In this chapter the implementations and the results of the study will be discussed. We are going to see step by step the analysis of the two datasets, “Bombas” and “Sales Conversion Optimization”.

4.2 Bombas Dataset Analysis

Data analysis starts with collecting the right data. In Bombas dataset, the data was collected by Google Analytics, while the second dataset was created by the Business manager of Facebook. The first thing that an analyst should do is to understand the data. Trying to interpret the data and combine the information that is available is a very important process.

4.2.1 Data Import

The first step of our analysis is to upload the data to the Repository of RapidMiner. We open the RapidMiner interface and in the window that pops up, we click “**Blank**” in order to start a new process. [Figure 4.1]

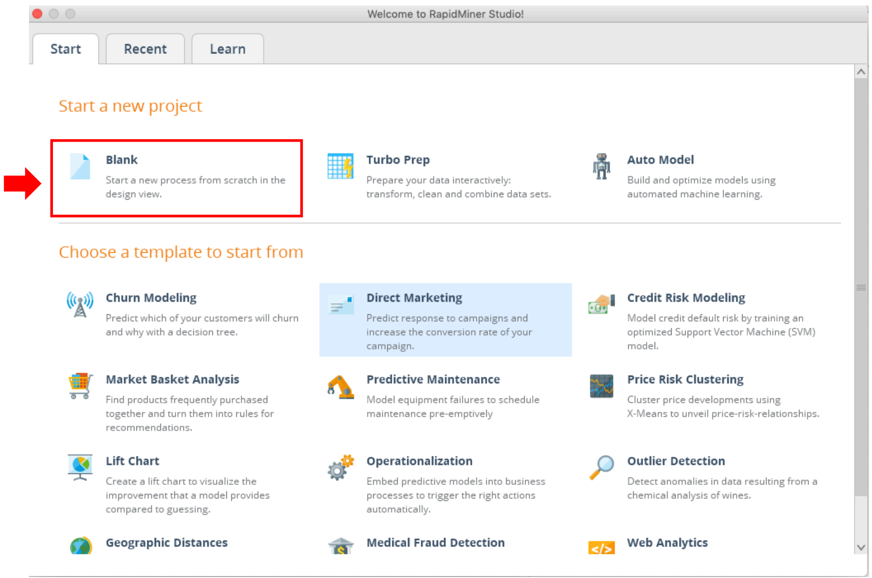


Figure 4.1 New Process in Rapid Miner

Then we click the button “**Import Data**” that we see in the left side of the screen. After clicking it, we select the location that the dataset has in the computer. Then, we select the dataset that we want to analyze and we click: “**Next**”.

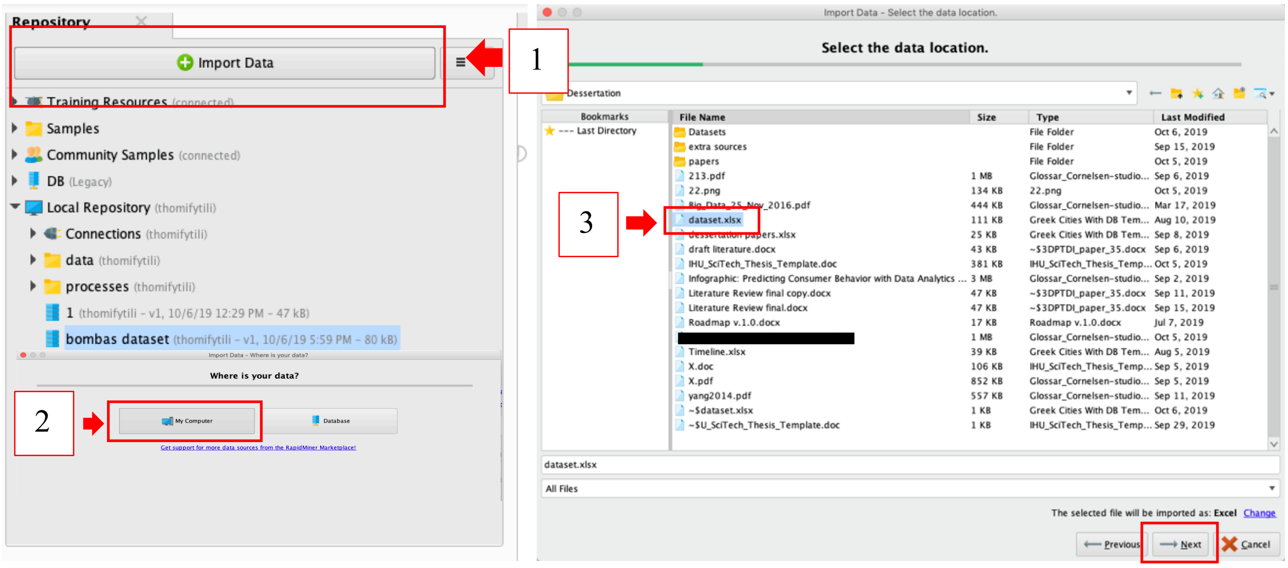


Figure 4.2 Import & Selection of the dataset

The file is uploaded, and we can have a quick view of the data. In case there is data to other sheets of the file, we can include it on the left top corner. In the ***cell range***, we can choose which columns we want to include to our analysis. In our case, we select all the columns, so we leave

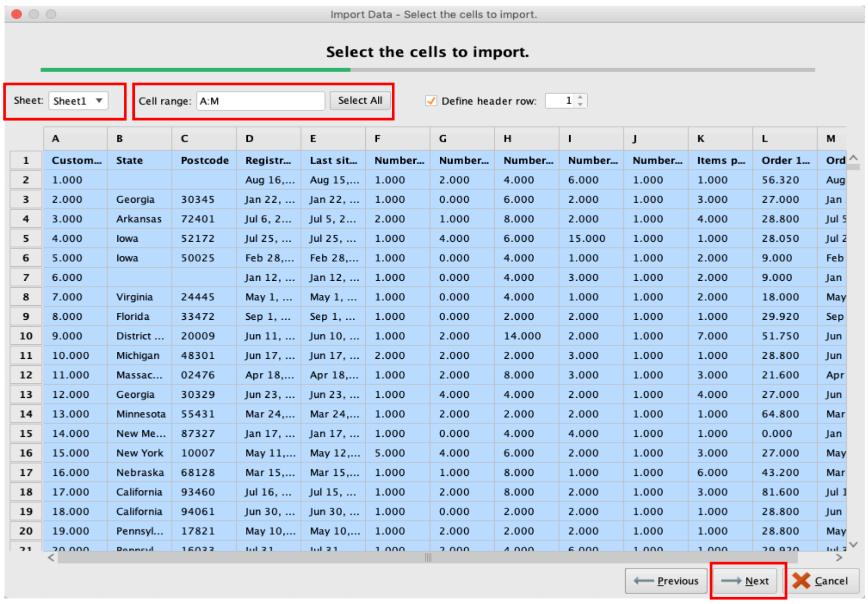


Figure 4.3 Select the cells to import

it as it is. Automatically all the columns are selected. In order to continue, we click “**Next**”. [Figure 4.3]

In the following step we check the button “**Replace errors with missing values**”. There are cells in the dataset without information or others which contain mistakes. A mistake for example could be to have a non-numerical value in a column which is supposed to have numerical values. If we try to import the dataset with spaces or mistakes, a reading error will occur. In that case, the data cannot be uploaded. If this box is checked, spaces and mistakes will be converted to missing values and the dataset will be uploaded successfully. We are going to handle the missing values later in the data preparation section. The gear gives the possibility to change the type of each column, change the role, rename the column and exclude it. All these options can also be accomplished in following steps.

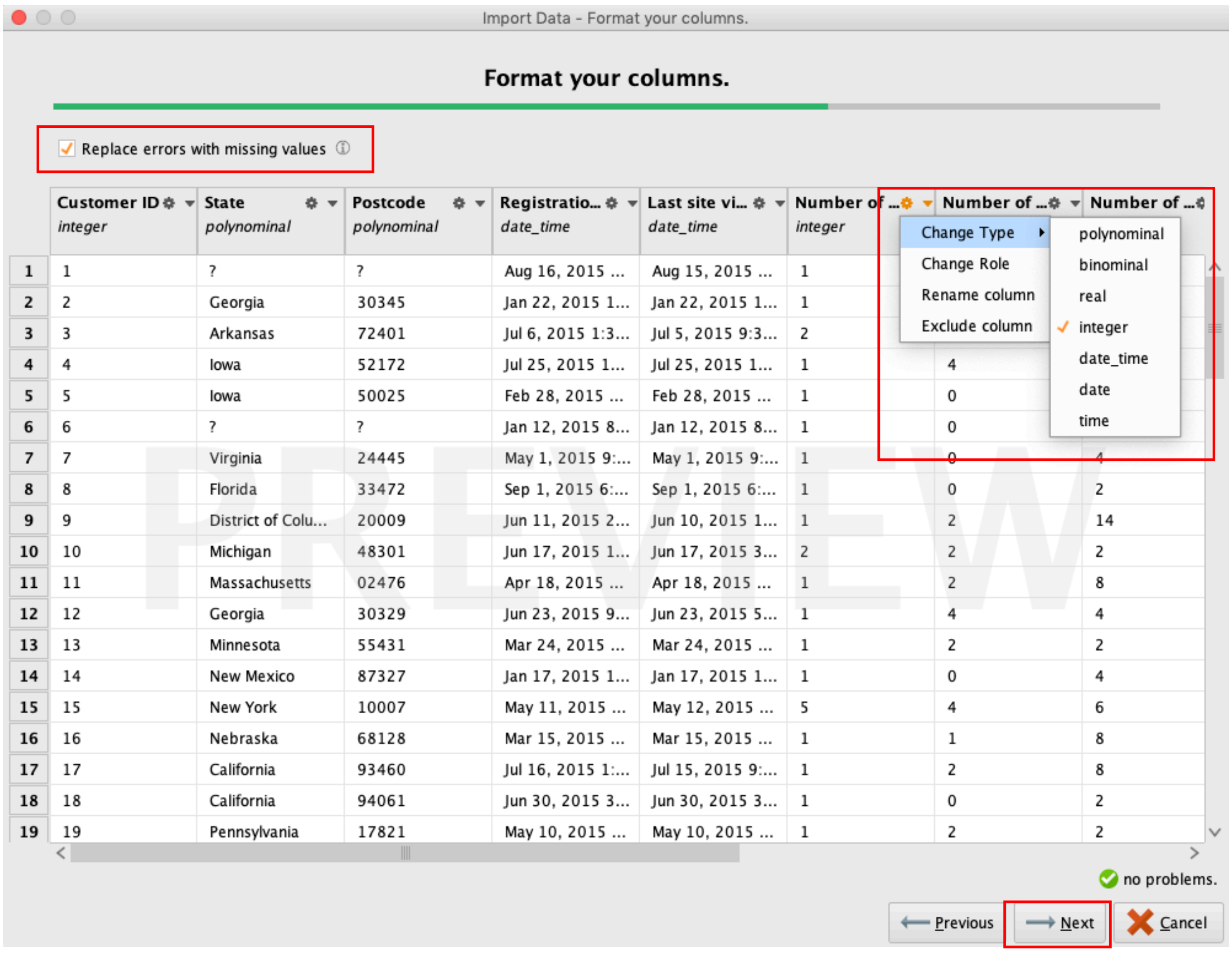


Figure 4.4 Format the columns

The dataset is automatically saved in the "**Repository panel**". RapidMiner's repository describes the metadata together with the data in a way that makes easier the design of the process. We hit the "**Finish**" button and the dataset is successfully uploaded. [Figure 4.5]

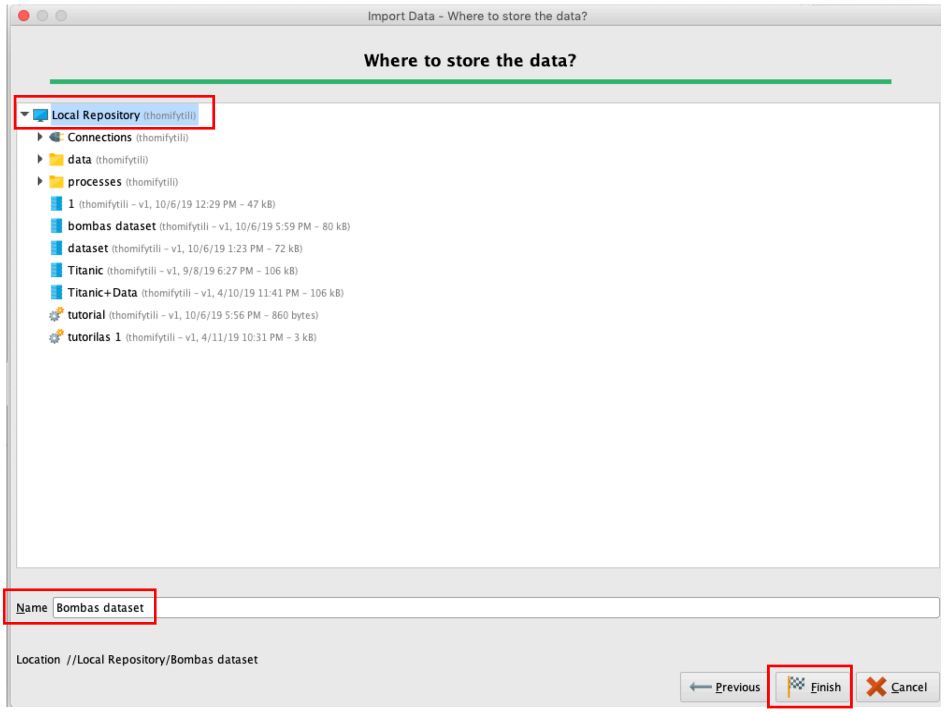


Figure 4.5 Repository Panel

4.2.2 Data processing

After uploading the data, we can find it in the Local Repository on the left top corner of the interface. We drag and drop the file into the process panel. In order to run the data, we should connect the output port of Retrieve bombas dataset (out) with the result port (res) by dragging a line. At the end, we need to hit the run process button at the top left in order to start the process. [Figure 4.6]

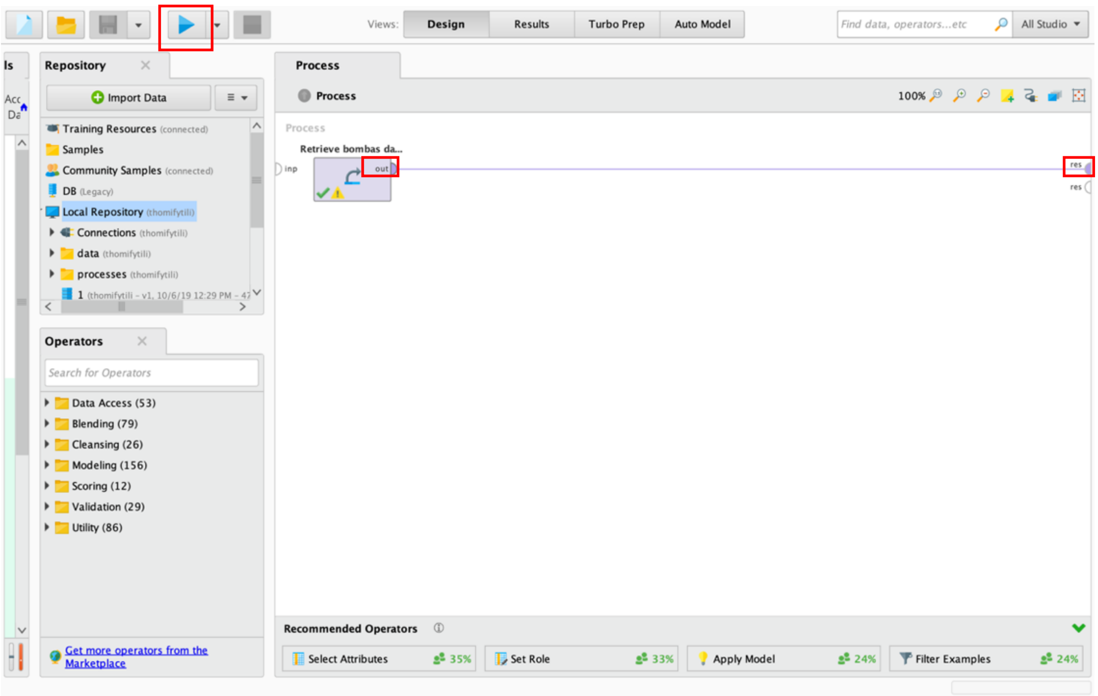


Figure 4.6 Retrieve the dataset

Results

In the Results tab we can see the table of the data. In RapidMiner, the dataset is called Example Set, as we can see in [Figure 4.7]. The rows are called examples and the columns have the name of attributes. Special attribute are the values that we want the model to learn to predict. In our example, we haven't set any attribute as special. On the top right corner, we can display all the examples or only those which contain missing values. We can also change the sequence in display, as we can change the position of the attributes by just dragging and dropping them. Last but not least, we can change the sort order by clicking on the attribute's name, one for ascending, twice for descending and three times to remove this order.

Statistics

In the Statistics tab we can have access to dataset's information. On the left side, all the attributes' names are displayed. The next column is about the type of each attribute. In our example the majority of the attributes have an integer type. With the next column we can find out how many missing values exist. Additionally, basic statistics are displayed as minimum, maximum and average values of each attribute. If we click the arrows on the left side, we can notice that a mini chart is displayed. We can click "**Open visualizations**" and a new window will be opened with a graph representation. [Figure 4.8]

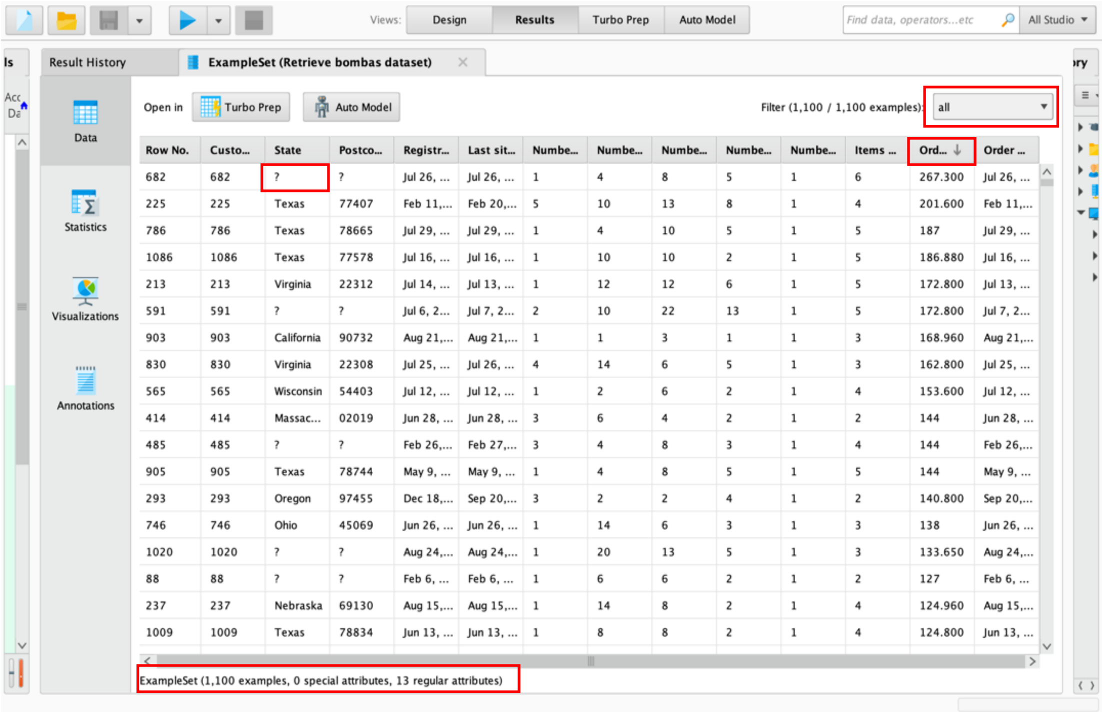


Figure 4.7 Dataset Table Overview

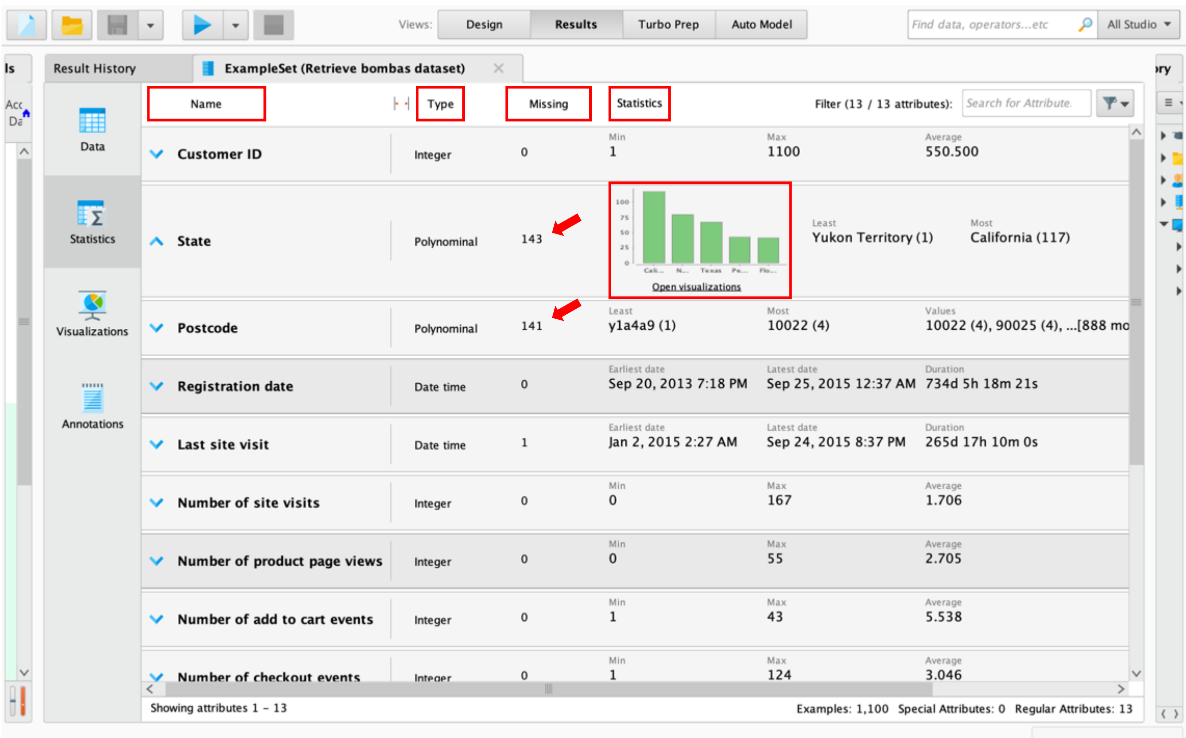


Figure 4.8 Statistics Overview

Visualizations

In the visualizations section, we can find all the possible visualizations of the dataset. In RapidMiner, more than 30 different visualization methods are available. Displaying the data is a way to interpret it. Graphs are actually all the representations which display the connection and the relationship between the data. We present below the most famous visualization methods in RapidMiner. (manual rapidminer)

Bar chart

Bar column or horizontal is a chart which displays how often a value appears. Bar representations are very important and help us understand how an attribute is distributed.

Pie chart

Pie chart is another method of data visualization. It is displayed by a circle divided to sectors. Each sector represents a specific quantity of a value.

Scatter plot

In scatter plots the data is displayed as a bunch of points. Each point has a value which is indicated by the position of the point.

Boxplot

Boxplot is a graph which displays data through their quartiles. In this method, outliers usually are indicated by the white dots. By using this graph, we can have a clear view about the outliers of the dataset.

Heatmaps

A heat map shows how dense a region is. The values are presented by colors. The region which accumulates the most data usually has the color red.

4.2.3 Data Preparation

Data preparation is usually the most time-consuming step of the process, but also the most critical and important. In this part of the analysis, the data is cleaned or is replaced with another data in order to give useful information. According to Gartner, only 20% of data is in a structured form and the rest 80% is unstructured. This means that this data has to be preprocessed before being able to make a model for analysis. Getting the data in a good shape for the analysis is very important. According to Forbes, most of the data scientists say that data preparation is the worst part of their jobs.

Handling Missing Values

Usually the datasets have examples with missing values. Missing data is a tough problem that we should take early into consideration and try to solve it.

According to bombas dataset, in the Statistics panel, we notice that three attributes contain missing values. Those three attributes are "*State*" with 143 missing values, "*Postcode"* with 141 missing values and "*Last site visit*" with only one missing value. We can handle the problem of missing values with the following three ways.

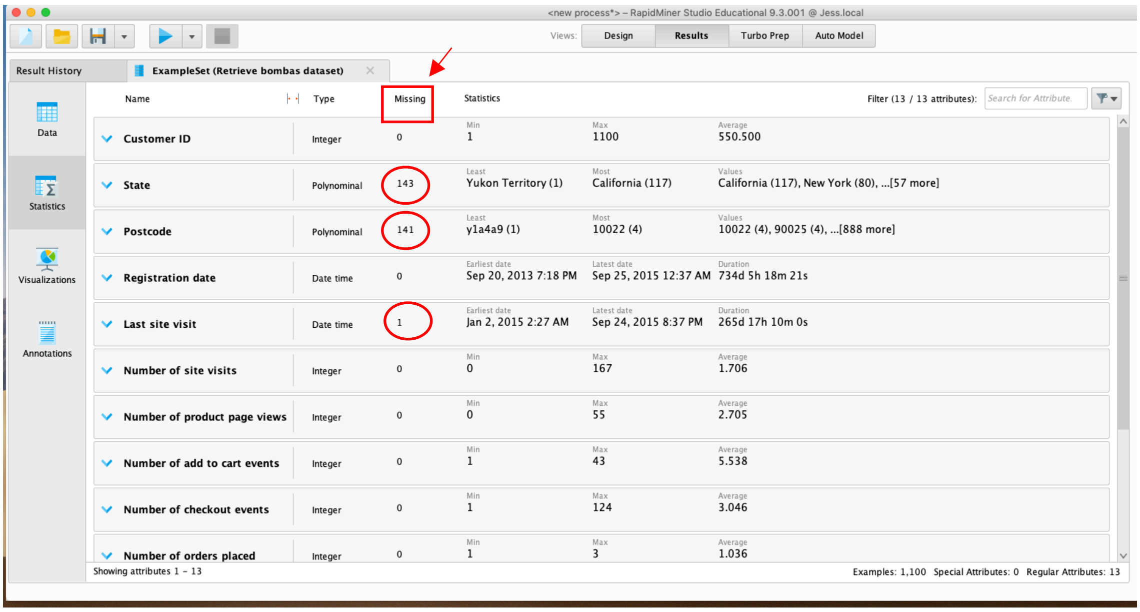


Figure 4.9 Statistics panel

The first way is to use the operator "***Select Attributes***" and exclude from the list of the attributes those that don't have important information for our analysis. In our case the attribute "S*tate*" contains useful information, as we can later use it for further analysis. The attribute "Postcode" could also be quite useful if we had a large dataset. In our analysis we separate the customers coming from different States based on their location, but those customers most of the times need further analysis. Even though people come from the same State, they could have many differences. One difference could be their economic background. Sometimes the place that they live and more specific their neighborhood could be an indicator of their financial condition. Consequently, postcodes could be a source for this kind of analysis. Such information gives us the advantage to locate people and place advertisements with more expensive products to some neighborhoods and more economic solutions to others, with effective results.

The size of Bombas dataset isn't appropriate for this kind of analysis. For that reason, we can use the first way in order to remove this attribute and all the missing values that contains. We select the operator "***Select Attributes***". Then, from the parameters setting we set the ***attribute filter type*** to ***subset*** and we click ***select attributes*** in order to select all the attributes that we want to include to our analysis. We select all the attributes, but the postcode and we click ***apply***. [Figure 4.10]

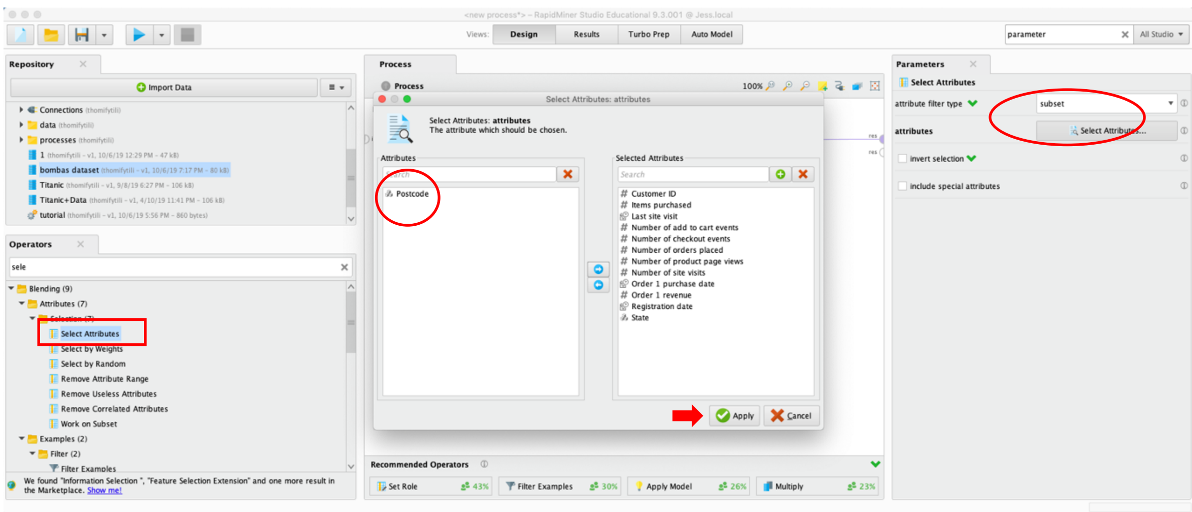


Figure 4.10 Excluding the attribute "postcode"

If we visit now the statistics tab, we notice that the attribute postcode doesn't exist in the list with attributes, as we can see in the [Figure 4.11].

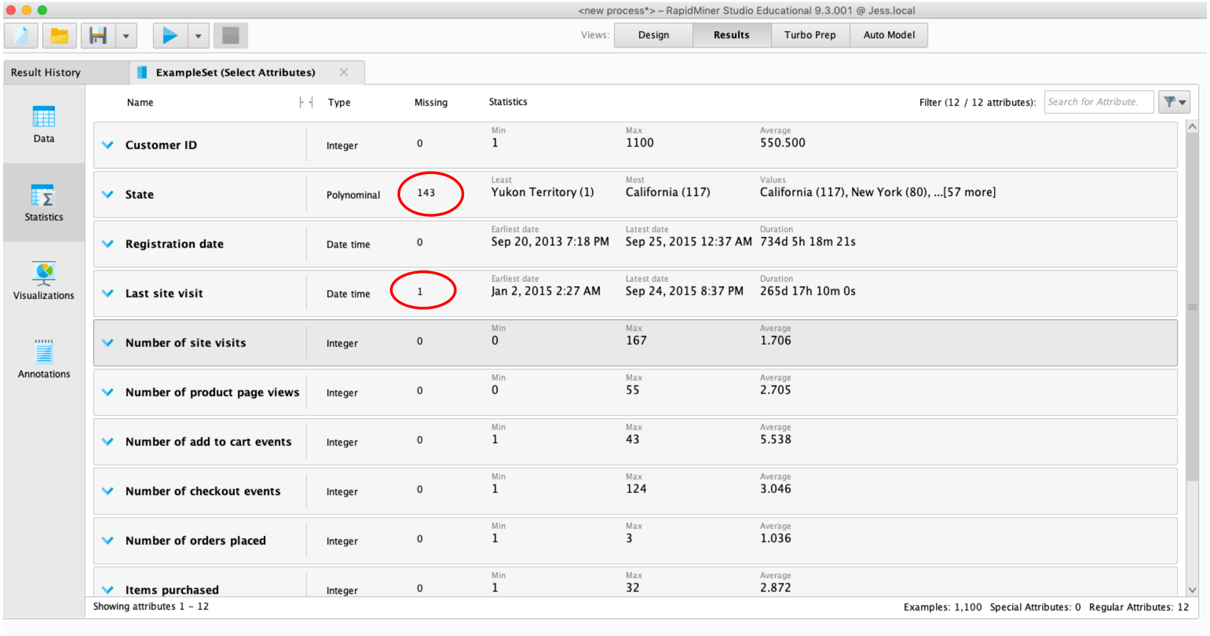


Figure 4.11 Statistics panel after excluding "postcode" attribute

The second way to handle the missing values is to replace them with other values. That could be useful if we had for example an age attribute. In that case, if there are missing values related to age, it would be useful to replace those missing values with an average value of age, instead of removing the whole attribute from the dataset.

The third way is to filter specific examples. In that way, we can remove examples with missing values and keep the rest values of the attribute. We choose the operator ***Filter Examples*** as we can see in the [Figure 4.12]. In the parameters settings we set the ***condition class*** to ***no missing attributes***. We run the process and if we go to statistics tab, we notice that there aren't missing values. Our dataset has now 956 examples. [Figure 4.12]

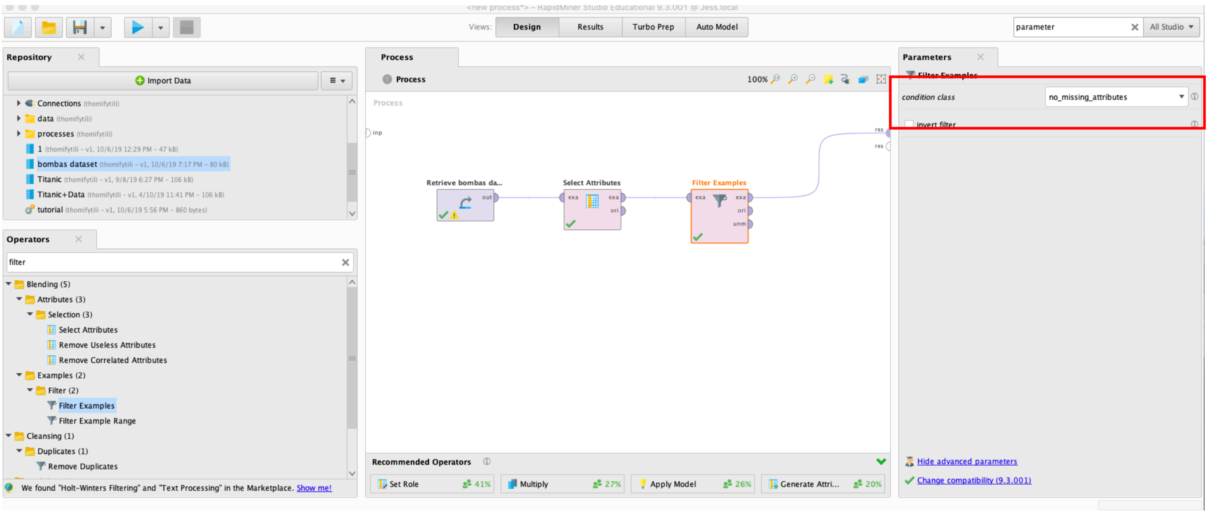


Figure 4.13 Filter Examples

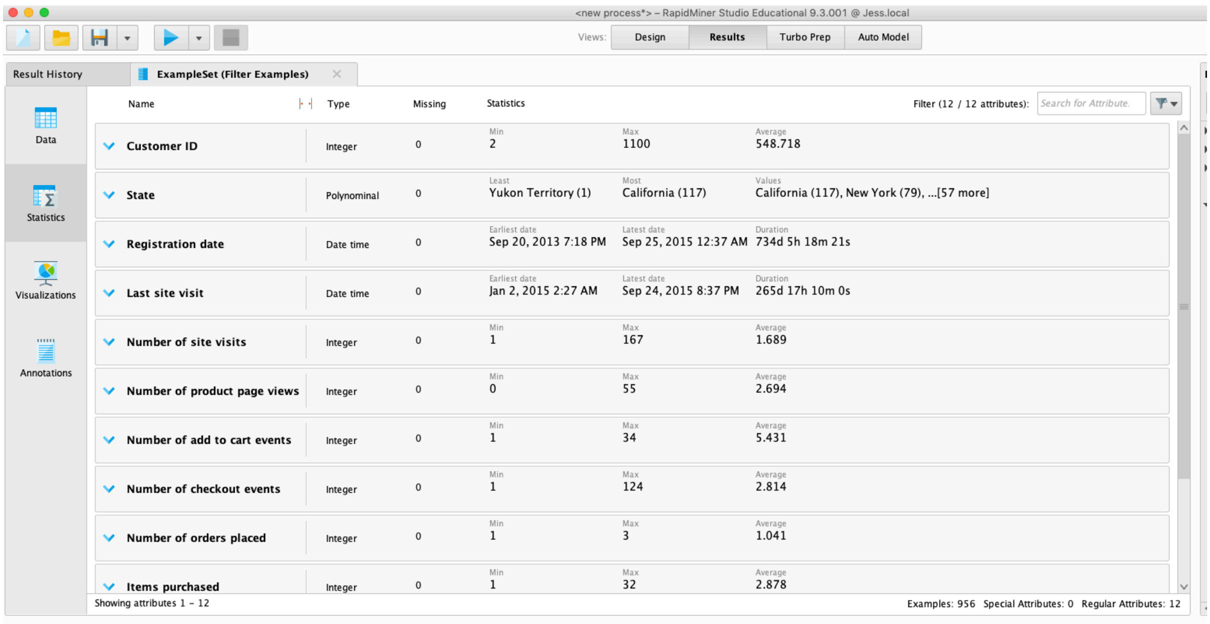


Figure 4.12 Statics panel after handling missing values

After handling the missing values, we can continue with the analysis of the dataset.

4.2.4 Data Analysis

After the data preparation, we continue with the main analysis of the dataset. The goal is to analyze the available data, create graphs, extract information, improve personalization and budget allocation and be able to make decisions about future campaigns based on this information.

Based on the orders placed in each State

In first place, an analysis was made based on the State that the customers come from, in order to study local performance and try to identify patterns. As we can see in the left corner the graph results are ***grouped by the State*** and the value is the **sum of the items purchased in each State**. The graph indicates that the top 3 States are California with 326 items purchased, New York 226 and Texas 193. On the other hand, other States such as Victoria and NSW have a few items purchased in total.

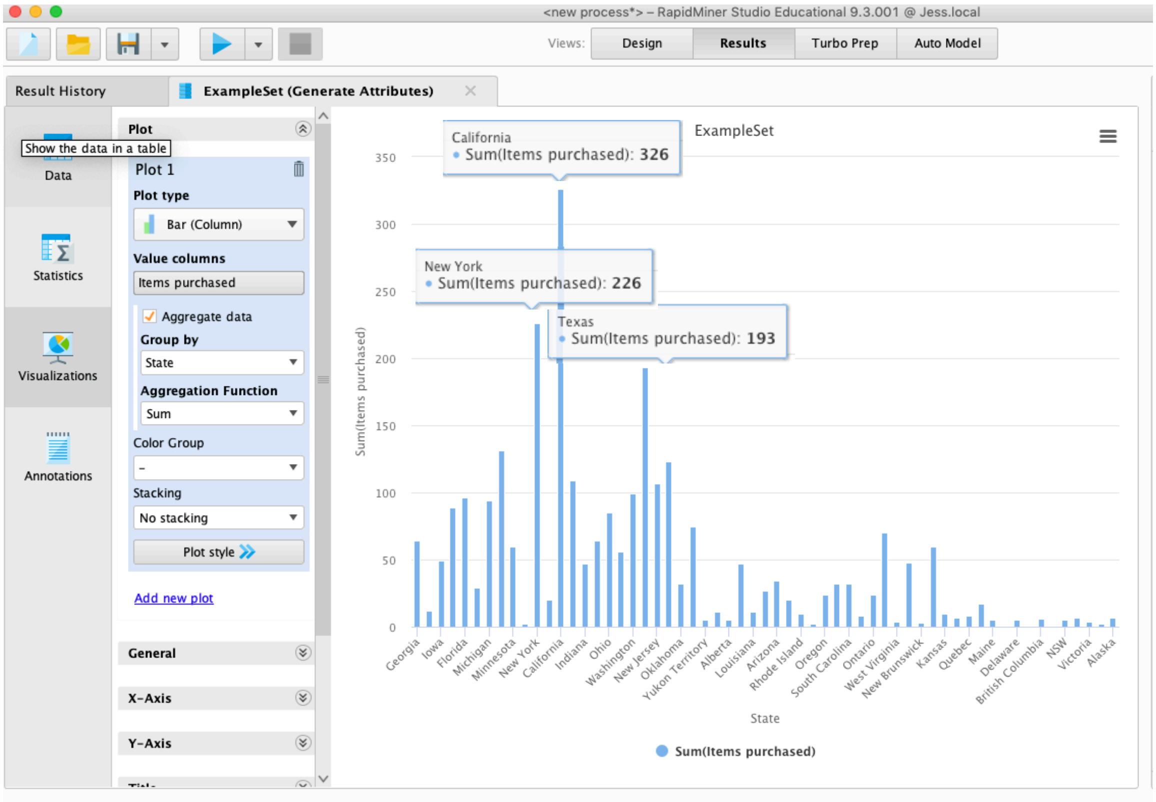


Figure 4.14 The total purchases for each State

The [Figure 4.15] illustrates the total revenue created from each State during this year. We notice that California, New York and Texas are again the top 3 countries but although customers from Texas purchased less items than the customers from New York, they created higher revenue. That probably means that they bought more expensive products.

The heatmap [Figure 4.16] is also a good representation of the amount of revenue each State produces. We notice that California is the only State in the range 4k-6k. In the next level is New York and Texas with yellow color. Follows states such as Florida, Michigan and Minnesota because as we see their color is a light yellow. Those States in blue have the less amount of revenue.

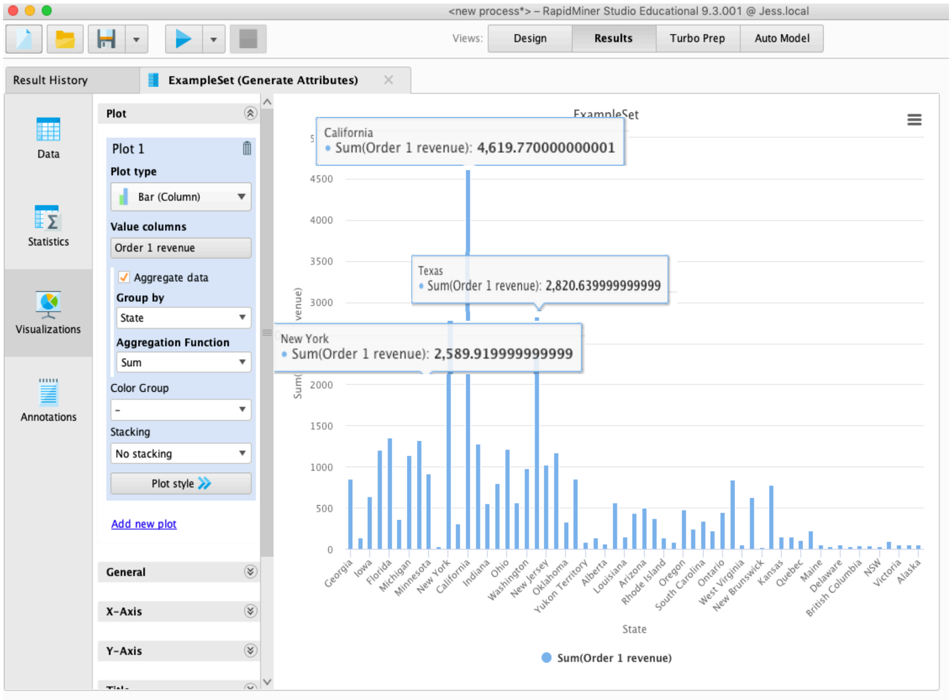


Figure 4.15 The total revenue for each State (1)

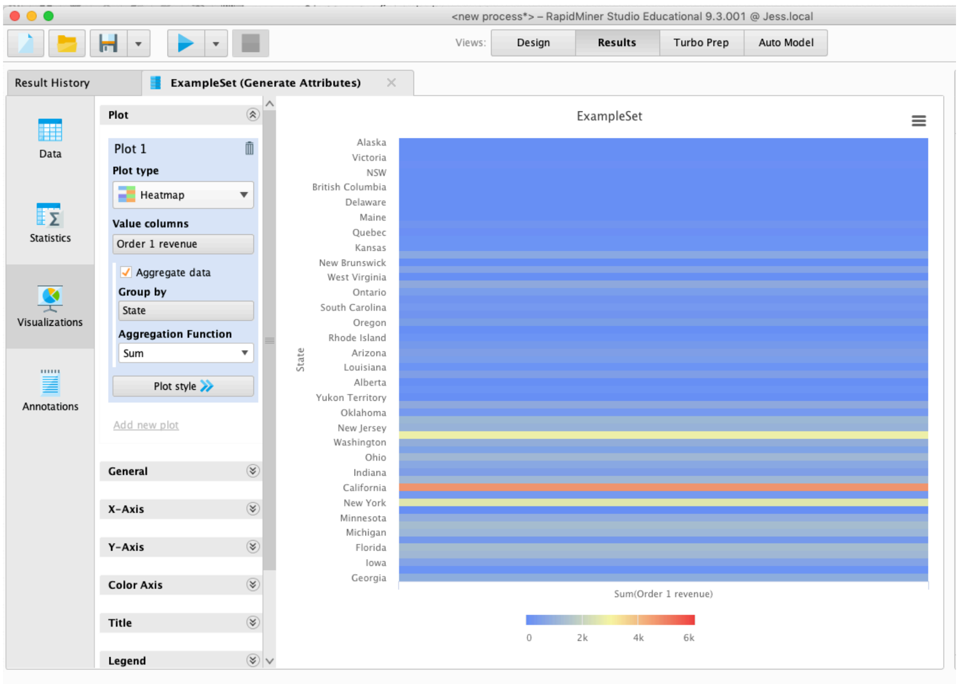


Figure 4.16 The total revenue for each State (2)

Recommendations:

Geography location plays a crucial role in creating customized and personalized campaigns. A company which offers its products to many countries should analyze constantly the traffic from each country in order to have a clear view and adjust future activities. A set of recommendations should be given based on the location criteria for future marketing campaigns:

- **Make research for different Countries - States:** It is very important to target the right people with the right message at the right moment. People's wants and needs are different from place to place. That means marketers should understand audiences' local habits, holidays and what they usually do in their free time. Then they should find a connection with their product - service and each of these occasion in order to promote it successfully. It is a good idea to take advantage of local events and offer discounts or gifts to your customers. People when celebrate tend to be more eager to buy for themselves or for loved ones.

- **Adapt the prices to States or even cities with different income brackets:** People's income most of the times plays a critical role in their desire to buy a product or not. A solution here could be to make a small decrease in price for audiences with lower income in order to tempt them buy the product. Another solution is to promote different products to different places based on their price. In Bombas case for example, a campaign with more expensive products could be launched in Texas because based on the previous graph, customers from Texas bought less products but more expensive, contrary to customers from New York.

- **Adjust the content of the campaign to each different culture:** The campaign's content should be personalized in order to attract the audience. People from different countries have different cultures. Something funny or enjoyable for some people may be disturbing for others. Also, people in different places use different language or different terminology to describe same things. Words most used in each place can be used as keywords to the marketing campaigns.

- **Take advantage of the weather forecast:** Make relevant content for each geography locations and connect it with the weather forecast such as cold, rainy and snow days. In that case different campaigns could lead to different landing pages for different geolocations. For example, Bombas can promote warm socks for cold days in New York with a relevant landing page, while promoting stylish socks for sunny walks in California beaches with another landing page.

- **Exclude areas that don't bring profits:** If a company implements all the above and notices that the profit coming from a specific area isn't adequate to the effort and budget spent, then should take the decision and leave this market and focus on others which are more profitable.

Based on the statistics coming from the State, a recommendation that could be made is to try to eliminate the budget spent on the top States such as California, New York and Texas. The company has a numerous audience in these countries so it could attract them with different ways and not with high cost campaigns. For example, try to attract and keep the existing customers with personalized emails, offers and discounts. Also, it could create content which will provoke the engagement of present customers. Make them post their Bombas socks on social media, make hashtags and take part in competitions. Those activities don't require high budget and they could be quite effective in order to keep those States in top of the profitable states. On the other hand, in the begging it could dedicate more budget to the States which have some profits but not so high. Those States are depicted with a light yellow in the heatmap. Some of them are Florida, Pennsylvania, Ohio and Virginia. Trying to preserve the top profitable States, while boosting little by little the other States and implementing the recommendations that referred previously, could be a good growth plan.

Based on month, day and time

We continue the analysis based on the day month and time that most purchases took place. The attribute ***Order 1 purchase date*** contains information about the month, the date, the year and time that the purchase happened. We would like to analyze those metrics separately. For that reason, we are going to use the operator **"Date to Numerical**" and isolate each time the attribute that interests us. [Figure 4.17]

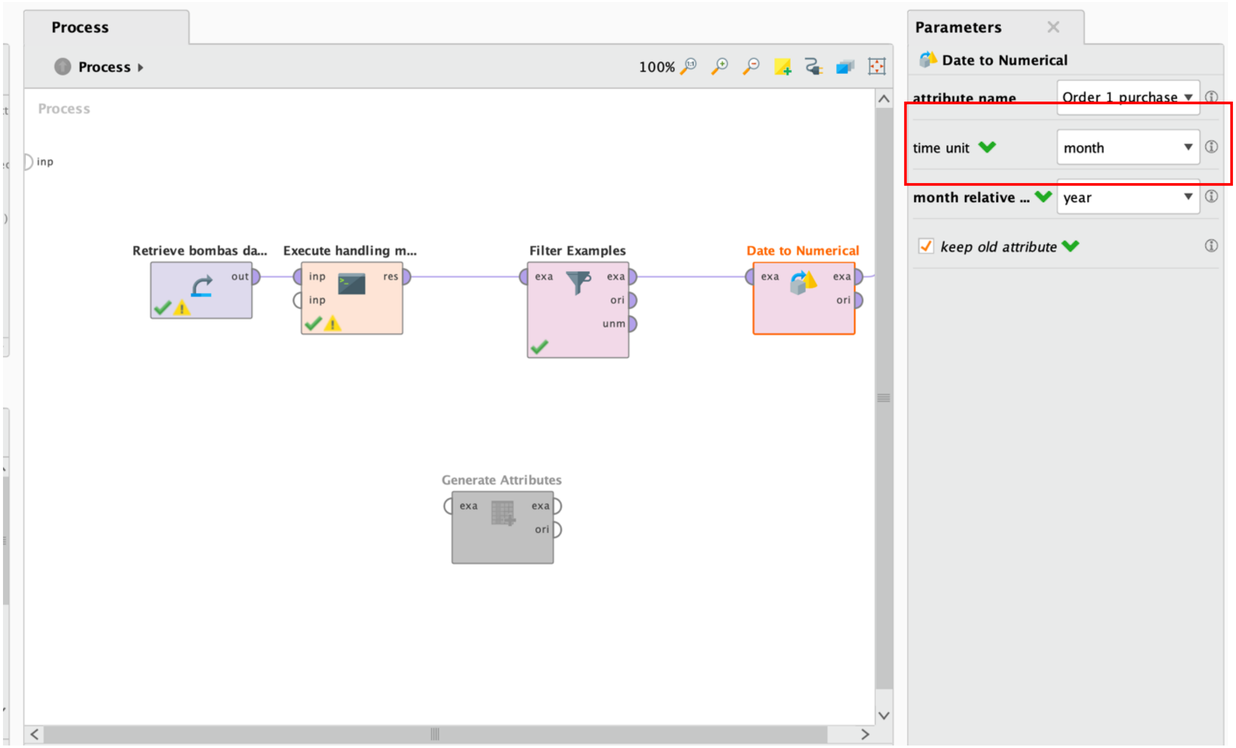


Figure 4.17 Date to Numerical Attribute

Analysis based on the Month

Now it is very important to understand in which months the customers make the most purchases. This information will help us organize better the campaigns and invest more in those months that customers tend to buy more. In [Figure 4.18] we notice that June (6), January (1) and February (2) are the 3 top months that people tend to buy more. On the other hand, April (4) and September (9) are those with the lowest performance.

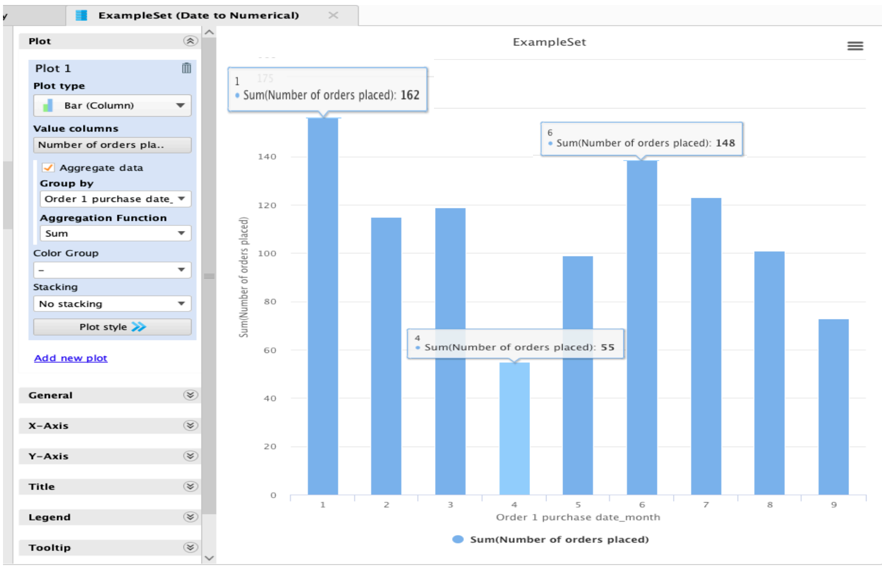


Figure 4.18 The total orders that took place in each Month

Now we are going to examine the same thing but for each State separately. We add to the process the operator ***Filter Examples.*** In the parameters section we define the filter as **State equals** to **California** and we hit the button **Ok.** We run the process and we check the same graph but this time only about the State California.

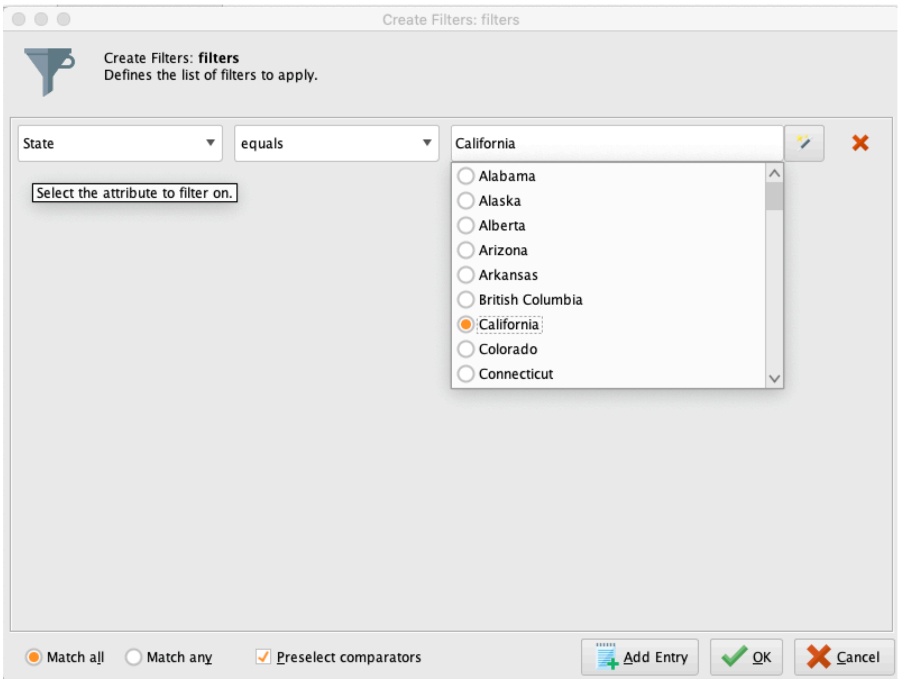


Figure 4.19 Filter and gives results separately for each State

As we can see in [Figure 4.20], again June (6) and January (1) are the months with the most purchases, but contrary to the [Figure 4.18] February (2) in California has a few purchases. In that case, more budget should be spent in campaigns for the months June (6), January (1), March (3) and May (4).

If we try to do the same analysis for the State of New York [Figure 4.21], we will notice that the month June (6) is again the month with the most purchases. Contrary to California analysis, July (7) and February (2) are the months with the most purchases while March (3) didn't have such a good performance.

If we create graphs for the rest of the States, we will observe that the purchases in each state are different in each month. This happens because people in different locations behave differently. We should follow these statistics and adjust our campaigns and the budget accordingly and not spent the same amount of money for all the States, the same period of time.

We chose to examine New York (79 examples) and California (117) because are the highest traffic States according to this dataset.

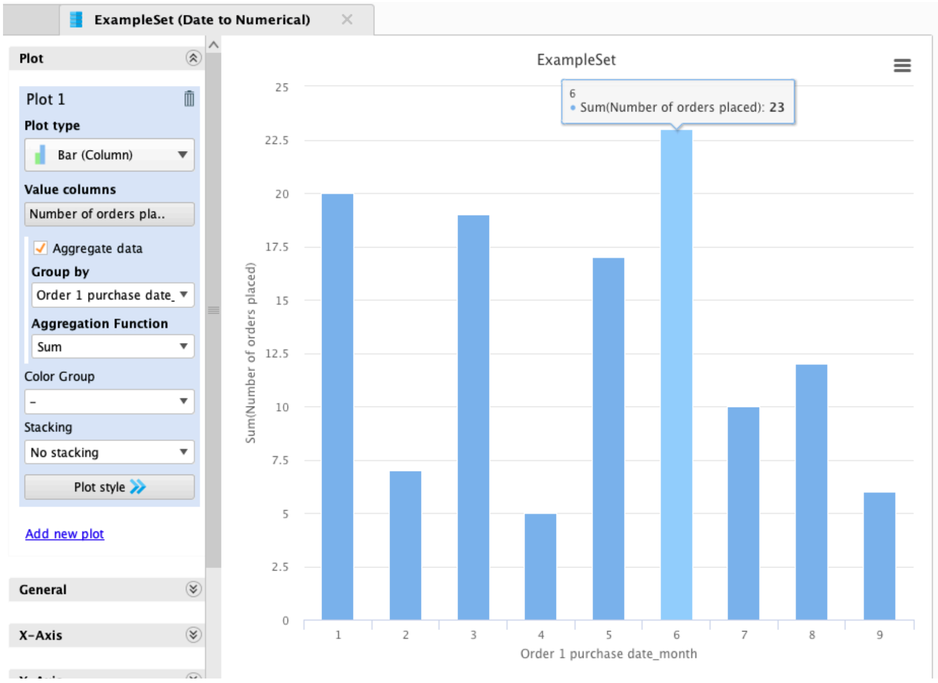


Figure 4.20 The total number of orders placed in **California** based in each **Month**

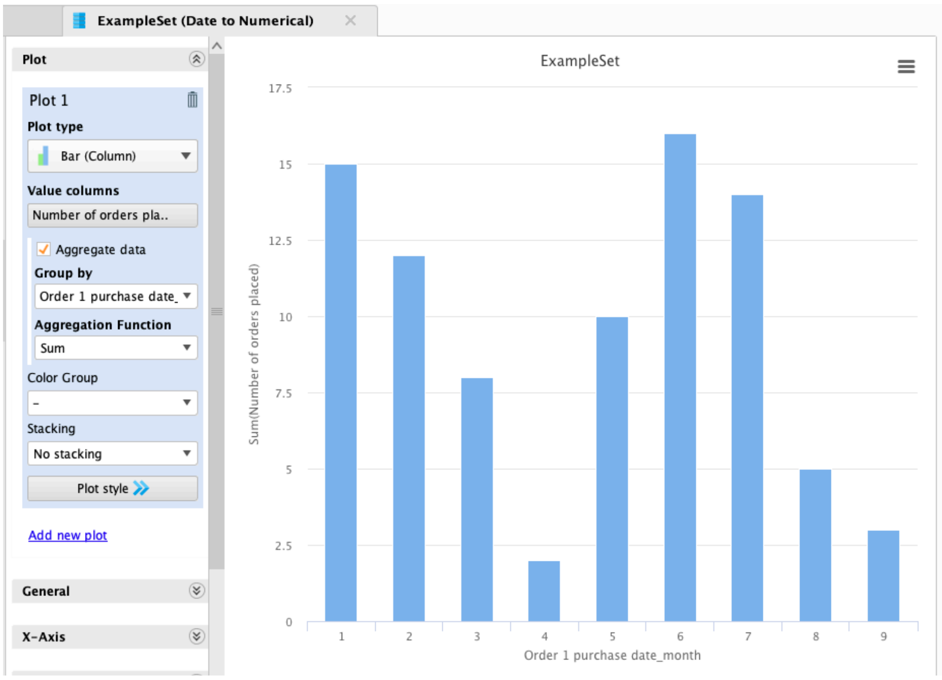


Figure 4.21 The total number of orders placed in **New York** based in each **Month**

Analysis based on the Day

We follow the same tactic for the different days of the week. In general, in all States the most purchases took place during the weekdays with the most famous the day Monday (2), while Friday (6), Saturday (7) and Sunday (1) had the less purchases.

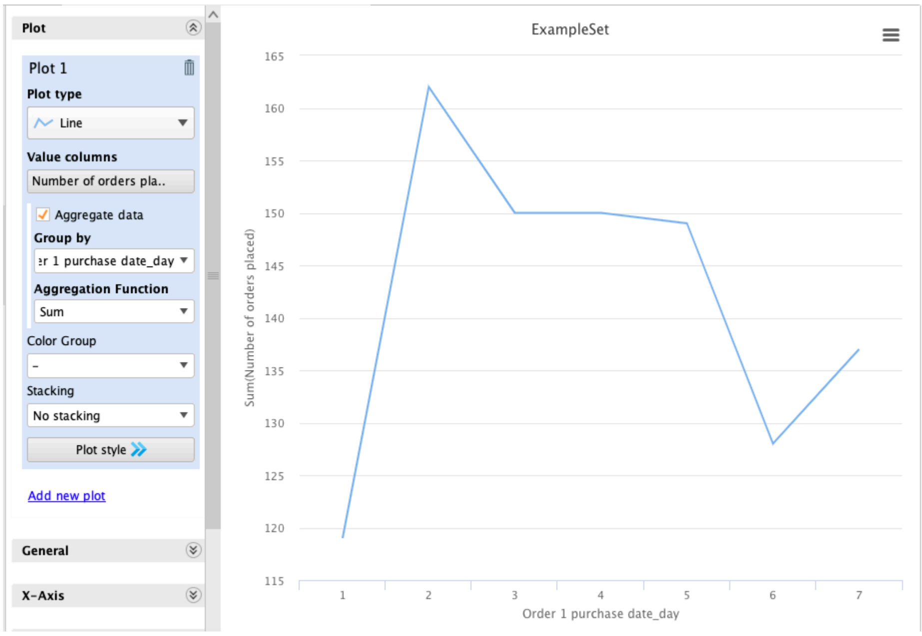


Figure 4.22 The total number of orders placed based on the **Day**

We implement the same Filter to check if there is different performance in each different State.

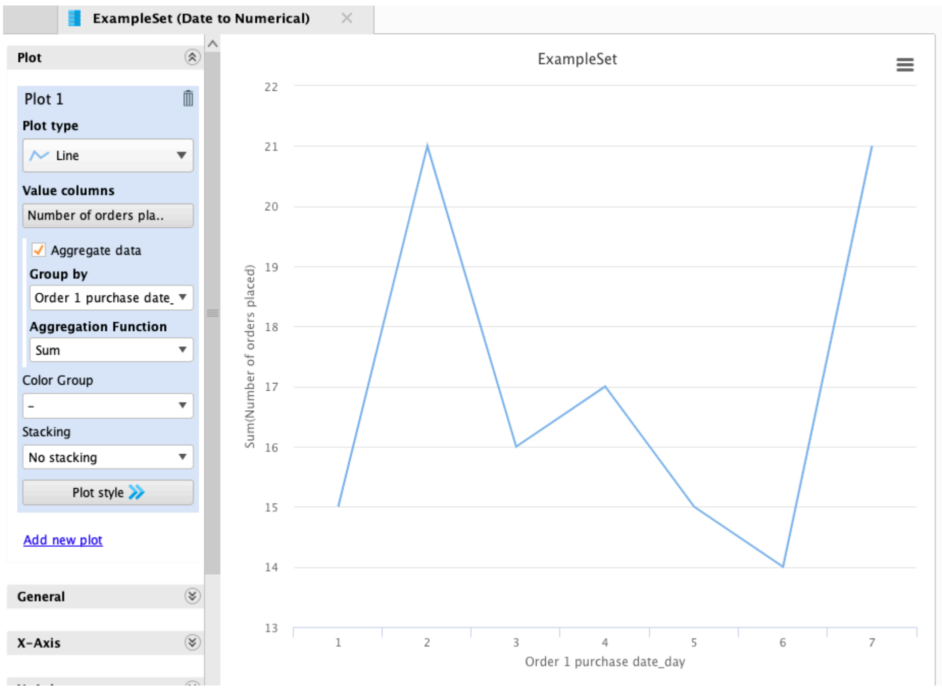


Figure 4.23 The total number of orders placed in **California** in each **Day**

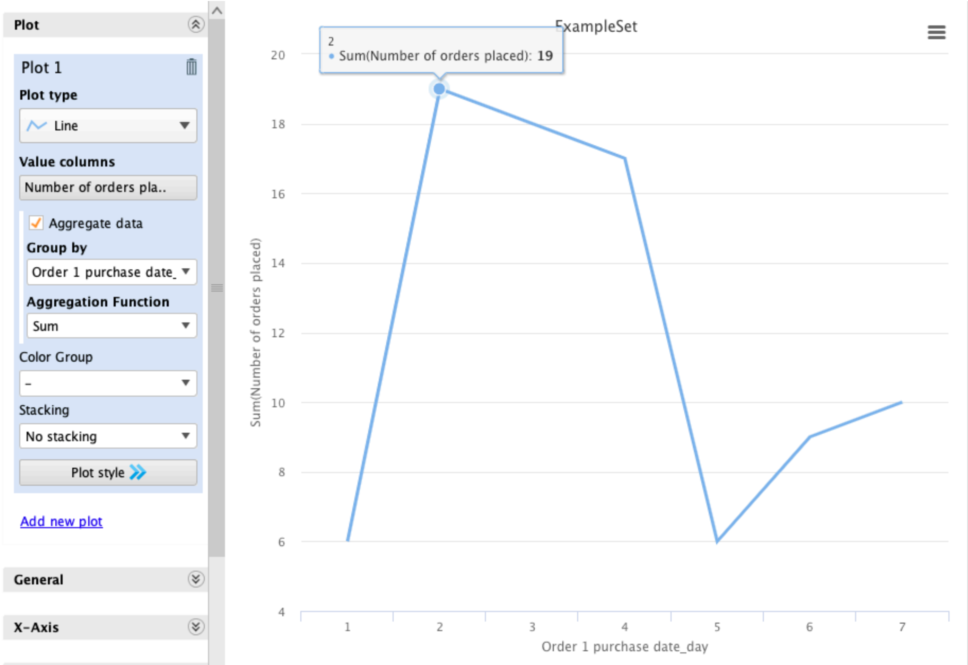


Figure 4.24 The total number of orders placed in **New York** in each **Day**

We notice from the graphs above that in New York, the most purchases took place during the first 3 days of the week (2, 3, 4), while in California the most purchases took place on Monday (2) and Saturday (7). Having this in mind advertisements should be displayed in New York more often during the first days of the week and less on the weekend, while in California should be more often during Mondays and Saturdays.

Analysis based on the Time

Now we check the hours of the day that most of purchases took place. We notice a few purchases during the time period [1-7 AM], which is normal because most people sleep this time of the day. Many purchases are observed during [11-16 PM] and become less during the period [17-23 PM]. We notice a pick at 20 PM and then as the time pass, the number of purchases tend to decrease.

Comparing Figure 4.26 and Figure 4.27 we notice differences regarding the time in each different State. In California many purchases made at 21PM while in New York only a few. In California many people purchase at 17PM while in New York not. Display an ad in both States at 21PM or at 17PM probably not be the same effective. This information should be taken into consideration because the advertisements should be placed to different audiences in different times.

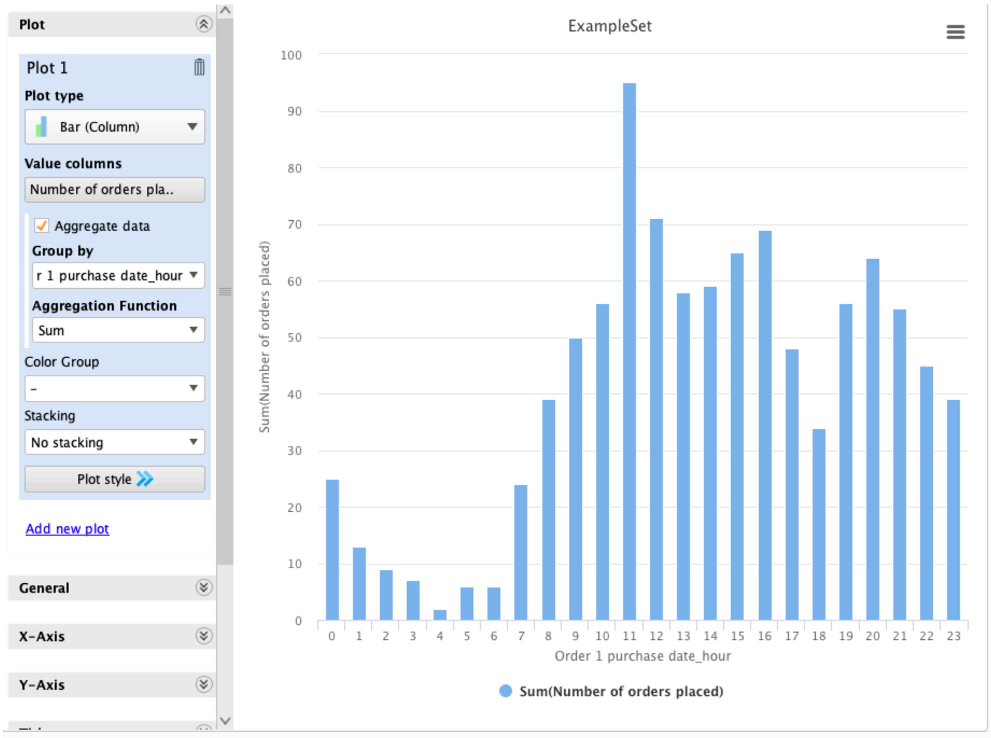


Figure 4.25 The total number of orders placed in different **Time** of the day

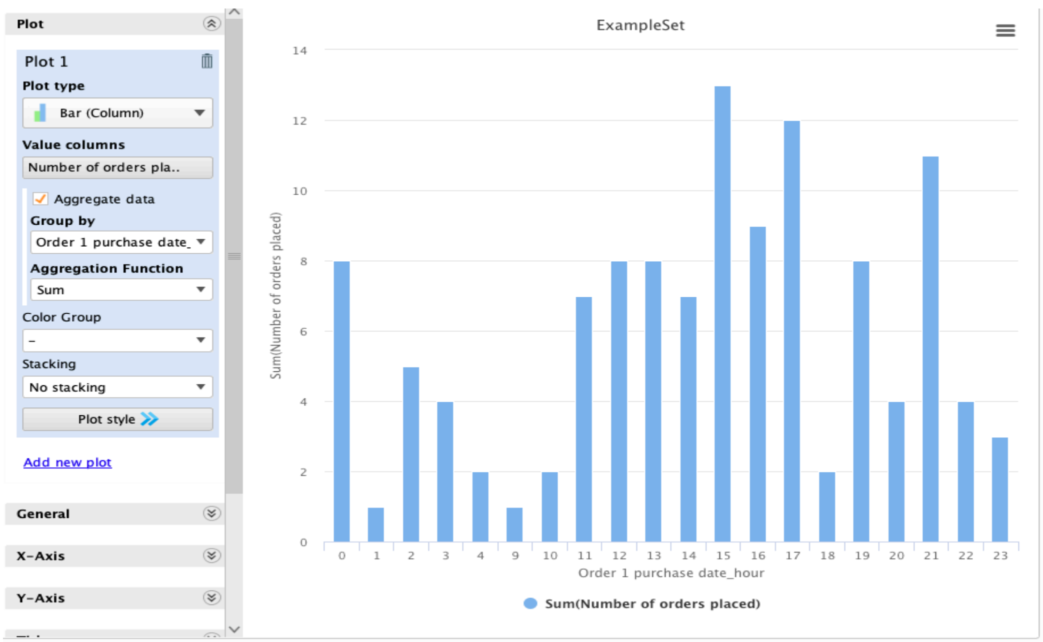


Figure 4.26 The total number of orders placed in **California** in different **Time** of the Day

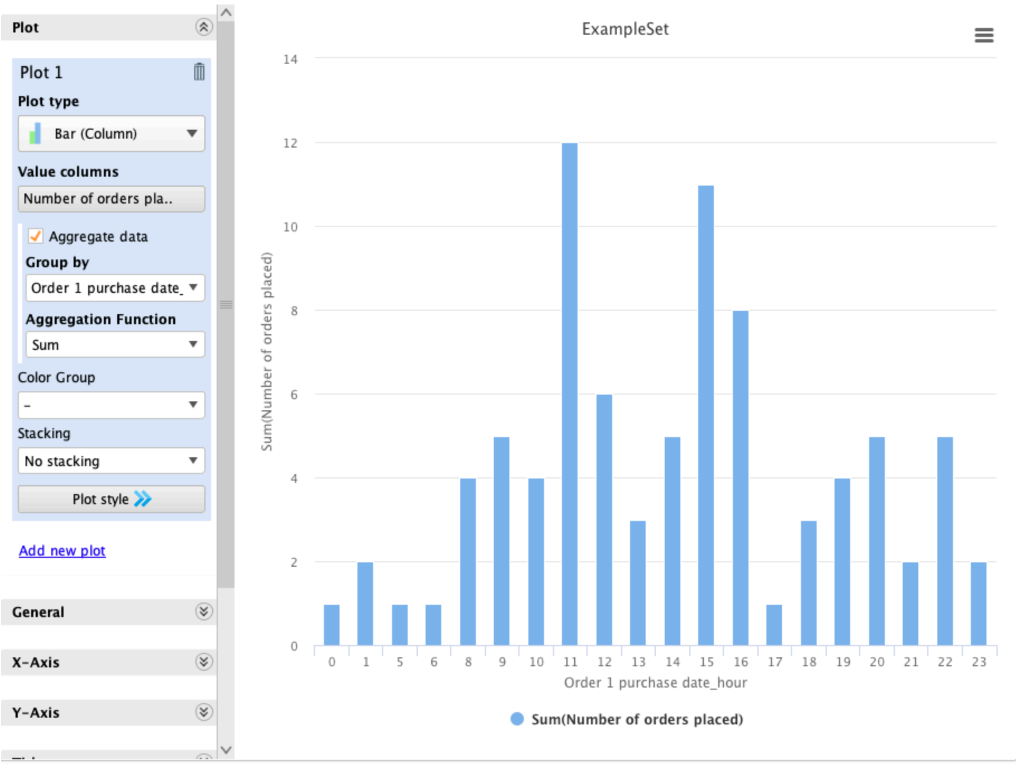
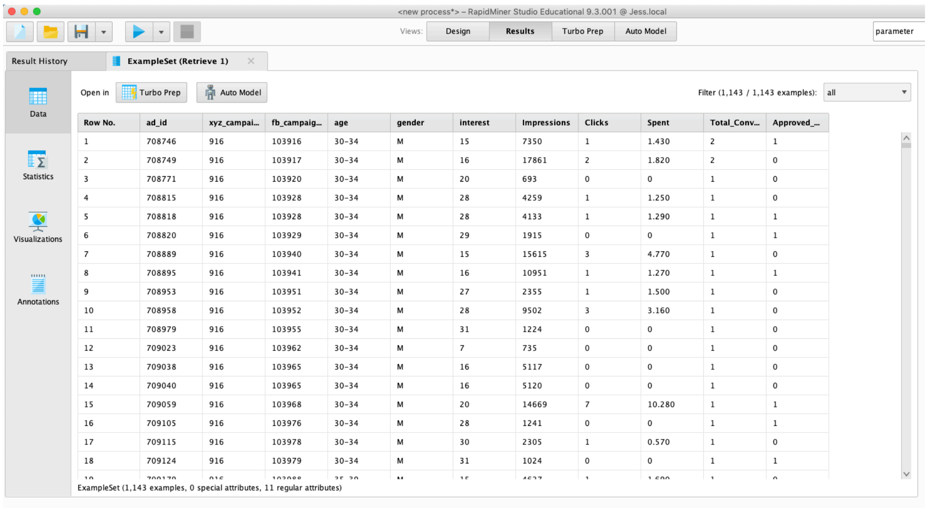


Figure 4.27 The total number of orders placed in **New York** in different **Time** of the Day

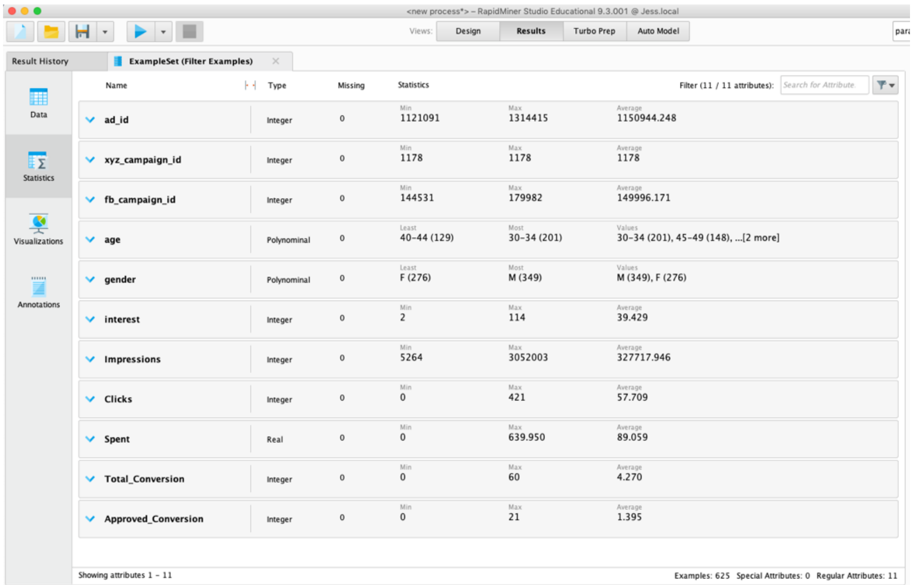
4.3 Sales Conversion Optimization dataset Analysis

We upload the dataset to RapidMiner and the procedure is the same with the Bombas dataset.



Row No.	ad_id	xyz_campai...	fb_campaig...	age	gender	interest	Impressions	Clicks	Spent	Total_Conv...	Approved_...
1	708746	916	103916	30-34	M	15	7350	1	1.430	2	1
2	708749	916	103917	30-34	M	16	17861	2	1.820	2	0
3	708771	916	103920	30-34	M	20	693	0	0	1	0
4	708815	916	103928	30-34	M	28	4259	1	1.250	1	0
5	708818	916	103928	30-34	M	28	4133	1	1.290	1	1
6	708820	916	103929	30-34	M	29	1915	0	0	1	1
7	708889	916	103940	30-34	M	15	15615	3	4.770	1	0
8	708895	916	103941	30-34	M	16	10951	1	1.270	1	1
9	708953	916	103951	30-34	M	27	2355	1	1.500	1	0
10	708958	916	103952	30-34	M	28	9502	3	3.160	1	0
11	708979	916	103955	30-34	M	31	1224	0	0	1	0
12	709023	916	103962	30-34	M	7	735	0	0	1	0
13	709038	916	103965	30-34	M	16	5117	0	0	1	0
14	709040	916	103965	30-34	M	16	5120	0	0	1	0
15	709059	916	103968	30-34	M	20	14669	7	10.280	1	1
16	709105	916	103976	30-34	M	28	1241	0	0	1	1
17	709115	916	103978	30-34	M	30	2305	1	0.570	1	0
18	709124	916	103979	30-34	M	31	1024	0	0	1	1

Figure 4.29 Dataset Table Overview



Name	Type	Missing	Statistics		
ad_id	Integer	0	Min 1121091	Max 1314415	Average 1150944.248
xyz_campaign_id	Integer	0	Min 1178	Max 1178	Average 1178
fb_campaign_id	Integer	0	Min 144531	Max 179982	Average 149996.171
age	Polynominal	0	Least 40-44 (129)	Most 30-34 (201)	Values 30-34 (201), 45-49 (148), ...[2 more]
gender	Polynominal	0	Least F (276)	Most M (349)	Values M (349), F (276)
interest	Integer	0	Min 2	Max 114	Average 39.429
Impressions	Integer	0	Min 5264	Max 3052003	Average 327717.946
Clicks	Integer	0	Min 0	Max 421	Average 57.709
Spent	Real	0	Min 0	Max 639.950	Average 89.059
Total_Conversion	Integer	0	Min 0	Max 60	Average 4.270
Approved_Conversion	Integer	0	Min 0	Max 21	Average 1.395

Figure 4.28 Statistics Tab Overview

As can be seen from the [Figure 4.29 & Figure 4.28], the exampleset contains of 1.143 examples and 11 regular attributes. Also, all the attributes contain numerical values except from the attributes ***gender*** and ***age*** which contain ***Polynominal values.*** Contrary to Bombas dataset, this dataset doesn't have missing values. So, there is no need to handle missing values. For the analysis of this dataset we consulted the notebook: Measuring Facebook Advertising ROI from Chris Bow. (Source: https://www.kaggle.com/chrisbow/an-introduction-to-facebook-ad-analysis-using-r) [82]

The dataset contains information about 3 campaigns. Each campaign consists of multiple set of ads with the similar campaign goal. We continue with filtering the dataset and analyzing the campaign with the id: **1178,** because we have more examples regarding to this campaign. Unfortunately, the observations aren't so many but compering the 3 campaigns that the dataset has, this one has the biggest number of observations. The bigger the number of the observations, the more precise would be our assumptions. More specific there are 625 examples and 11 attributes for the campaign 1178.

In first place, additional metrics such as **CTR** and **CPC** could be calculated based on the information that already exist. Those metrics help us to measure the results of a campaign and conclude if a campaign is successful or not. We can see some metrics below:

Click through rate (CTR): This metric counts how many impressions converted to clicks. A high ratio of CTR means that a lot of people who see an ad want to interact with it and click it. In other words, the campaign has relevant content with the characteristics and interests that the target audience has.

Conversion rate (CR): Conversion rate is the rate which measures the percentage of how many clicks become a conversion. The conversion could be whatever action is the goal of the campaign. For example, a direction to a landing page, a download of a file, a subscription to a newsletter, a video view, a purchase completion etc.

Cost per click (CPC): This metric counts on average how much each click costs. This cost depends on several factors such as the platform that is used for the advertisement, the industry and the competition. Companies with more expensive products or services tend to spend more in order to attract the audience to click on the link.

Cost per conversion: Is the metric which counts how much each conversion costs. In other words, it calculates the effectiveness of a campaign.

In first place, we are going to calculate the metrics **CPC** and **CTR** for our dataset and create two new columns presenting this information. We add the operator ***Generate Attributes,*** we complete the two function expressions and the names for the two new attributes. The attributes are generated successfully [Figure 4.31]. Now we can also remove the attributes that contain information that we aren't going to use such as the ***ad id***, ***xyz_campaign_id*** and ***fb_campaign_id.***

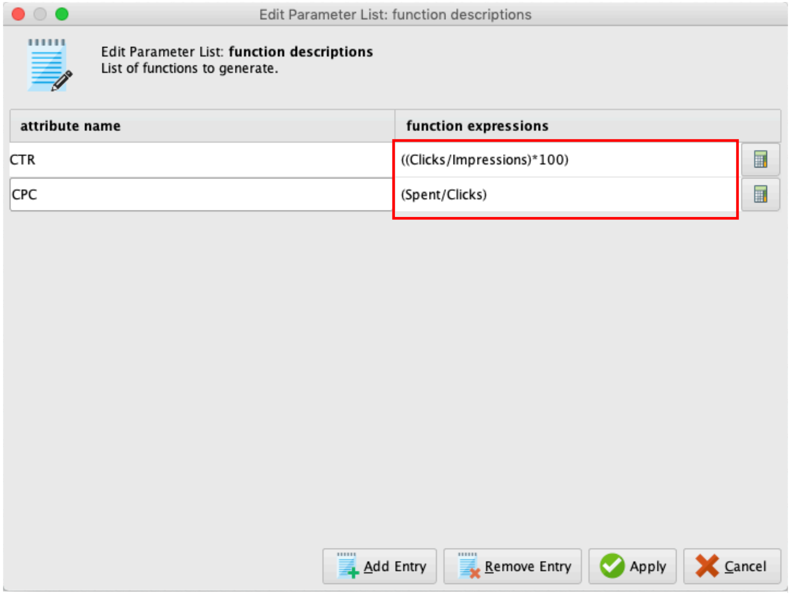
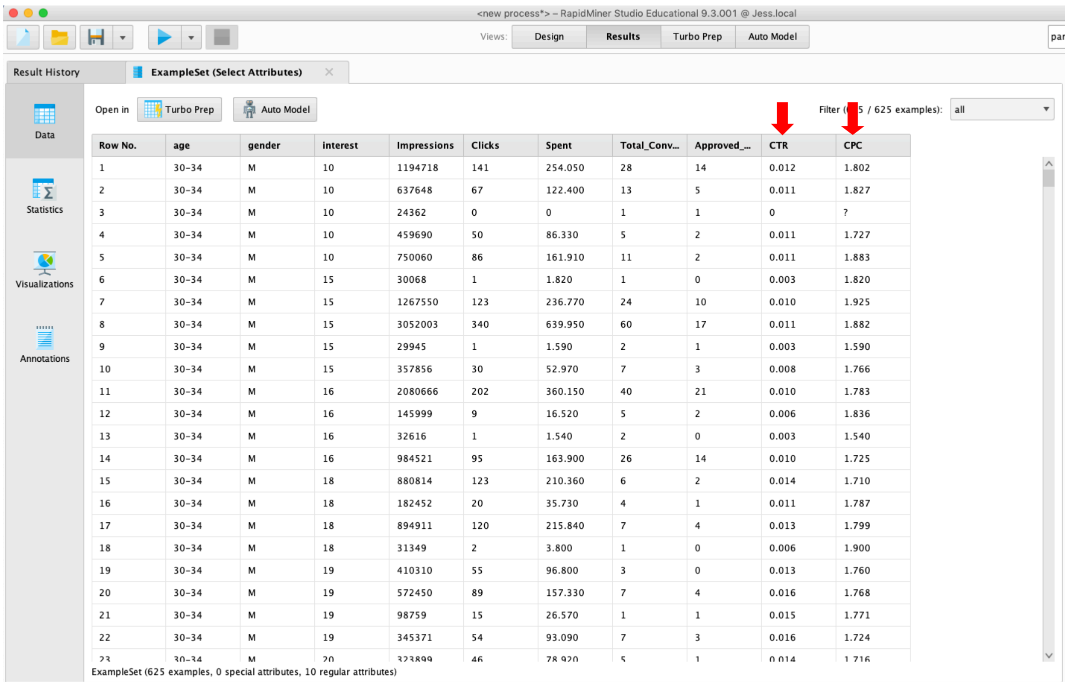


Figure 4.30 Generate attributes CPC & CTR



Row No.	age	gender	interest	Impressions	Clicks	Spent	Total_Conv...	Approved_...	CTR	CPC
1	30-34	M	10	1194718	141	254.050	28	14	0.012	1.802
2	30-34	M	10	637648	67	122.400	13	5	0.011	1.827
3	30-34	M	10	24362	0	0	1	1	0	?
4	30-34	M	10	459690	50	86.330	5	2	0.011	1.727
5	30-34	M	10	750060	86	161.910	11	2	0.011	1.883
6	30-34	M	15	30068	1	1.820	1	0	0.003	1.820
7	30-34	M	15	1267550	123	236.770	24	10	0.010	1.925
8	30-34	M	15	3052003	340	639.950	60	17	0.011	1.882
9	30-34	M	15	29945	1	1.590	2	1	0.003	1.590
10	30-34	M	15	357856	30	52.970	7	3	0.008	1.766
11	30-34	M	16	2080666	202	360.150	40	21	0.010	1.783
12	30-34	M	16	145999	9	16.520	5	2	0.006	1.836
13	30-34	M	16	32616	1	1.540	2	0	0.003	1.540
14	30-34	M	16	984521	95	163.900	26	14	0.010	1.725
15	30-34	M	18	880814	123	210.360	6	2	0.014	1.710
16	30-34	M	18	182452	20	35.730	4	1	0.011	1.787
17	30-34	M	18	894911	120	215.840	7	4	0.013	1.799
18	30-34	M	18	31349	2	3.800	1	0	0.006	1.900
19	30-34	M	19	410310	55	96.800	3	0	0.013	1.760
20	30-34	M	19	572450	89	157.330	7	4	0.016	1.768
21	30-34	M	19	98759	15	26.570	1	1	0.015	1.771
22	30-34	M	19	345371	54	93.090	7	3	0.016	1.724
23	30-34	M	20	323899	46	78.920	5	1	0.014	1.716

Figure 4.31 Data Table Overview of the campaign: 1178

We are going to investigate the potential correlation between the attribute ***spent*** and the rest of the attributes, using the operator ***Correlation Matrix***. Correlation is a statistical technique which reveals what attributes are related. Correlation matrix gives us very important information which we can use later in order to create patterns.

In order to produce the correlation Matrix table, we have to include to the design process the operator ***Correlation Matrix.*** We connect the ***mat*** port of the operator to the **res** of the output and we click ***play***.

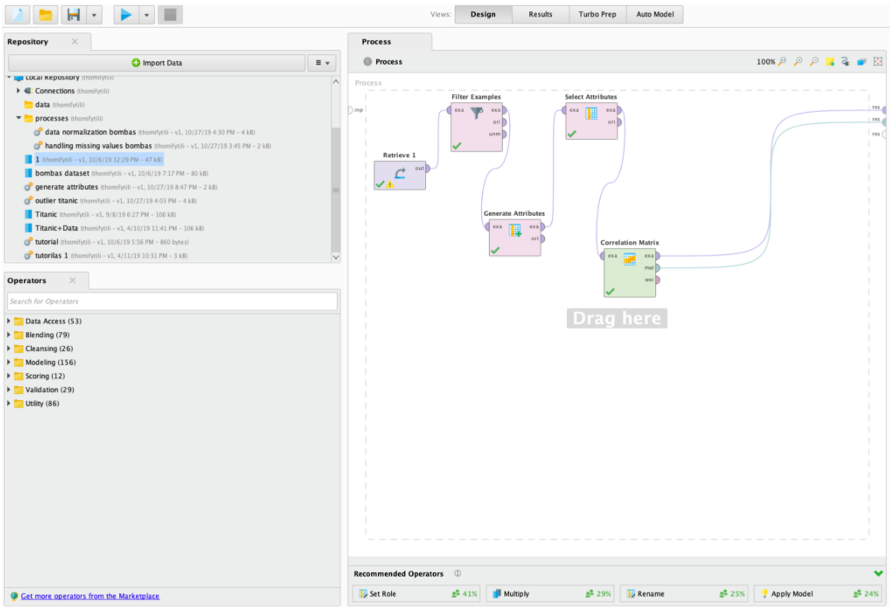


Figure 4.32 Correlation Matrix Operator

Attributes	age	gender	interest	Impressions	Clicks	Spent	Total_Conversion	Approved_Conversion	CTR	CPC
age	1	?	?	?	?	?	?	?	?	?
gender	?	1	0.086	0.158	0.300	0.235	0.082	0.009	0.511	-0.579
interest	?	0.086	1	-0.039	-0.046	-0.073	0.034	-0.023	-0.134	-0.286
Impressions	?	0.158	-0.039	1	0.933	0.962	0.790	0.671	0.203	0.013
Clicks	?	0.300	-0.046	0.933	1	0.991	0.646	0.520	0.400	-0.151
Spent	?	0.235	-0.073	0.962	0.991	1	0.684	0.562	0.354	-0.064
Total_Conversion	?	0.082	0.034	0.790	0.646	0.684	1	0.871	0.023	0.051
Approved_Conversion	?	0.009	-0.023	0.671	0.520	0.562	0.871	1	-0.015	0.100
CTR	?	0.511	-0.134	0.203	0.400	0.354	0.023	-0.015	1	-0.473
CPC	?	-0.579	-0.286	0.013	-0.151	-0.064	0.051	0.100	-0.473	1

Figure 4.33 Correlation Matrix Table

Each cell in the correlation matrix table shows if there is any correlation between two attributes of the dataset. From the table, it is clear that there is a strong correlation between the ***money spent*** for the campaign and the ***impressions***. There is also a strong correlation between the ***spent*** attribute and the total ***clicks*** that the campaign had. Correlation also exists in the relationship of ***clicks*** and ***impressions***. On the other hand, the correlation isn't so strong between the total spent and the two attributes ***total_conversion*** and ***approved_conversion.***

4.3.1 Analysis based on demographic characteristics

In this chapter we are going to segment and analyze the dataset based on the demographic characteristics gender and age.

Analyze by Gender

We are going to add one more attribute regarding the Conversion Rate. In other words, the percentage of how many clicks became conversions. We use the **Generate Operator** and the **Expression: ((Approved_Conversion/Clicks)).** We observe in the data overview table that there are 3 examples which have 0 clicks and 1 approved_conversion. This happens probably because either the click wasn't tracked, or because attributed elsewhere. We use the **filter operator** in order to keep only the values where the **CR is ≠ +-Infinity** in order to exclude these 3 examples from the dataset, because may cause problems to the analysis later.

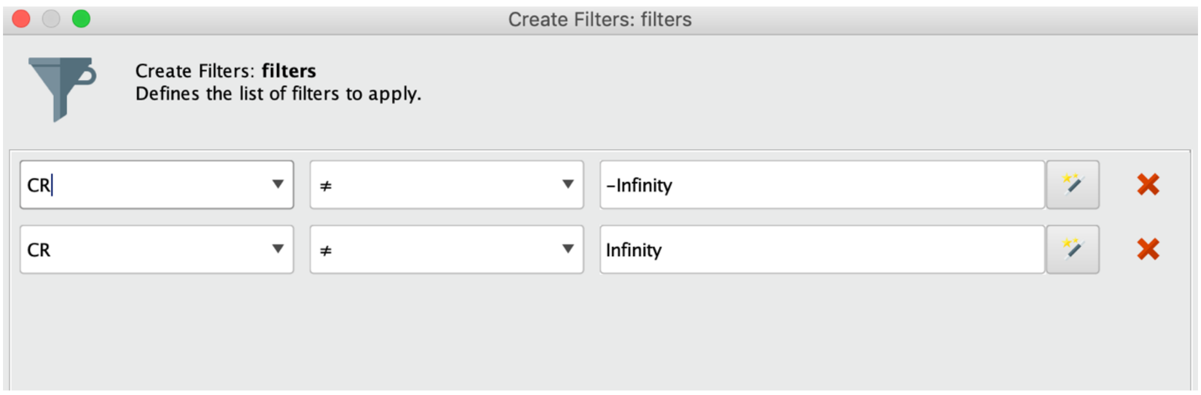


Figure 4.34 Filter the attribute CR

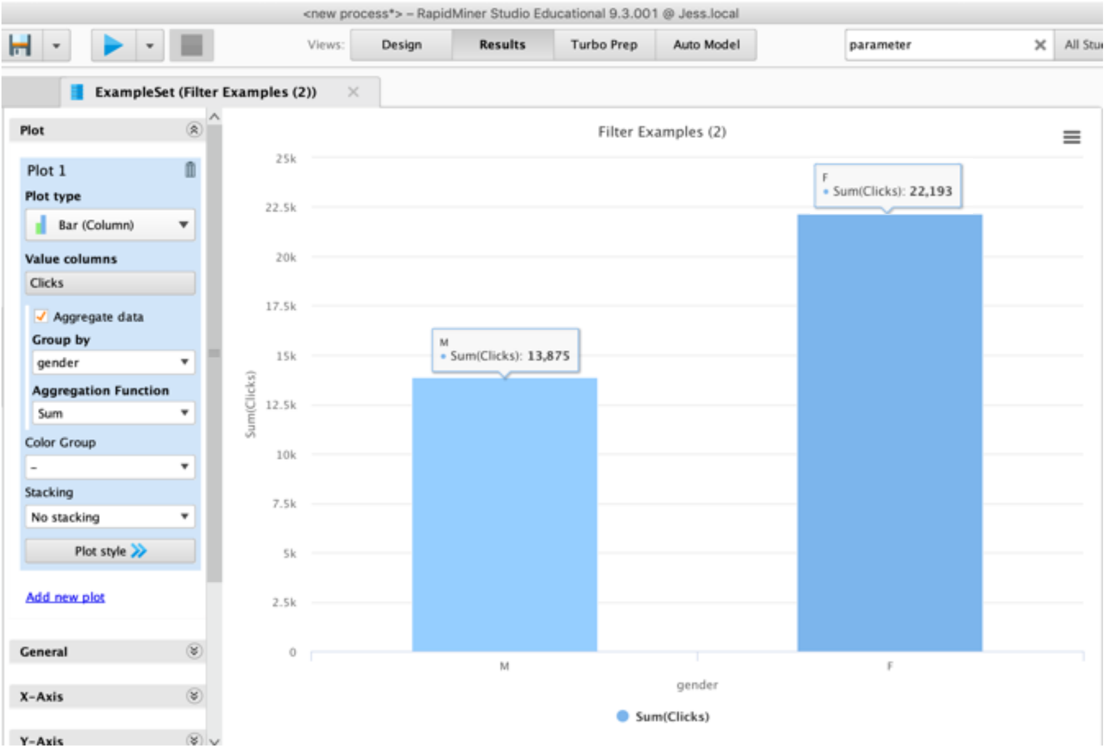


Figure 4.35 Sum of Clicks based on Gender

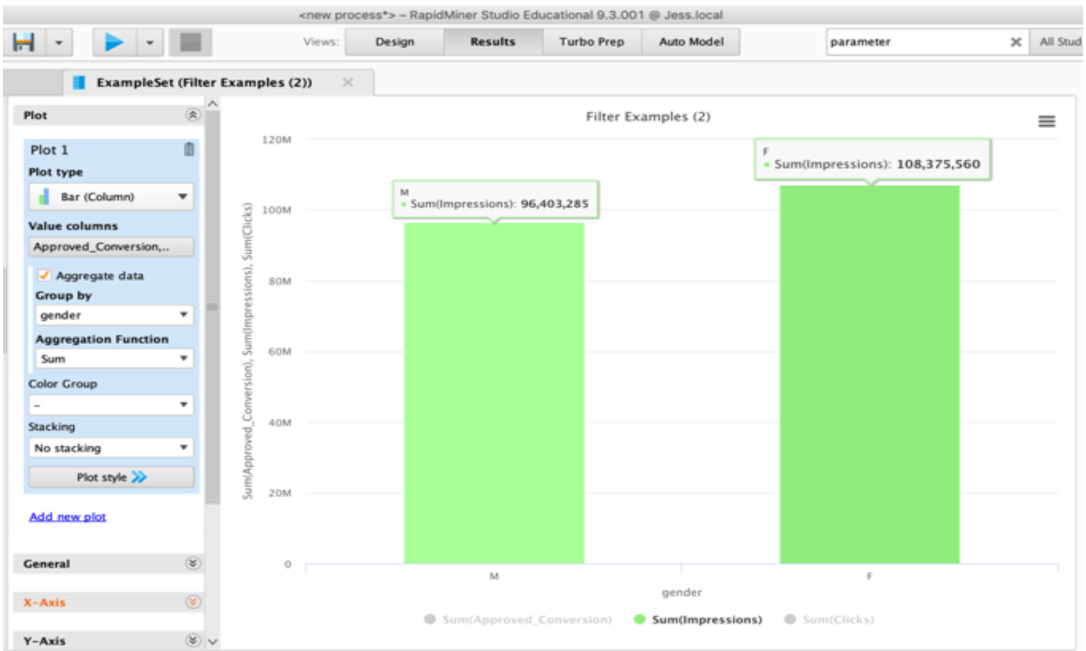


Figure 4.36 Sum of Impressions based on Gender

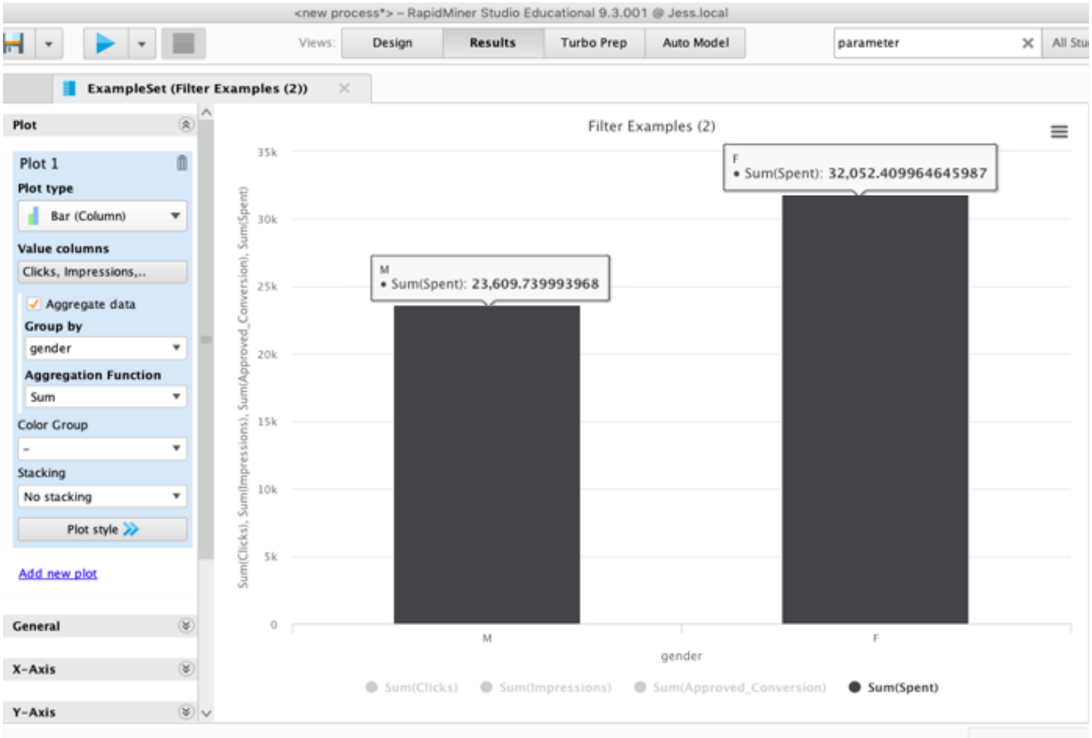


Figure 4.37 Sum of Spent based on Gender

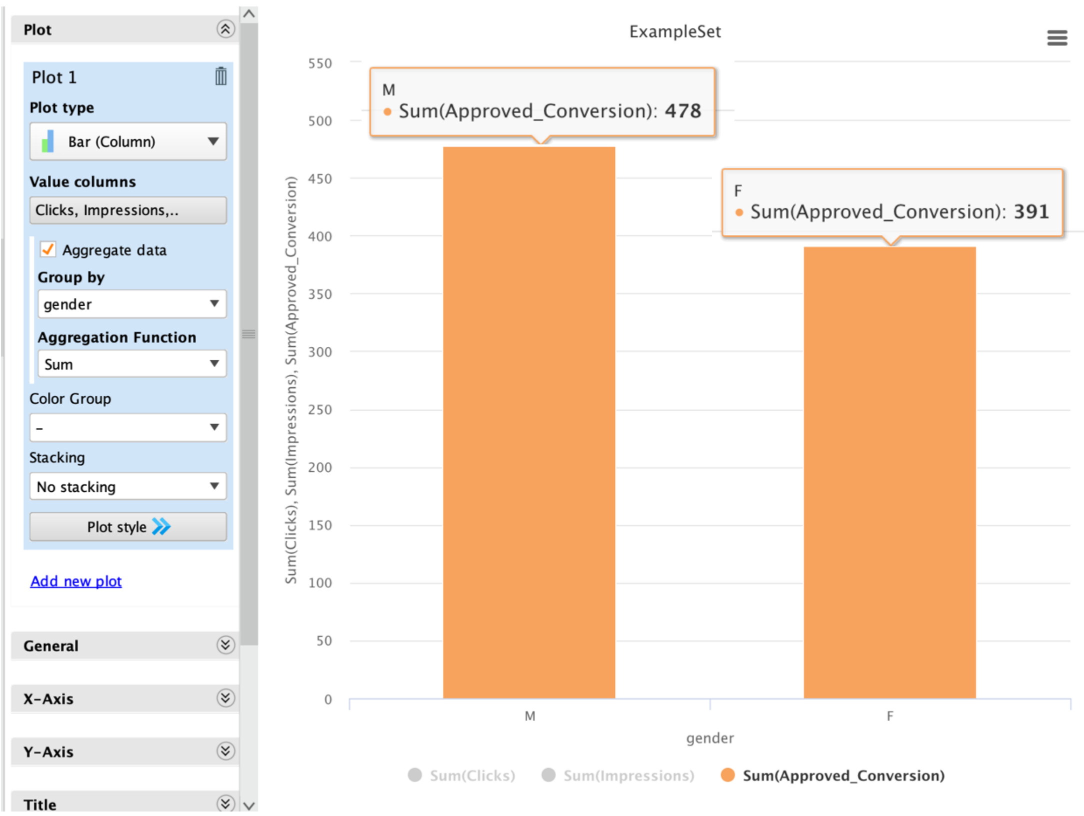


Figure 4.38 Total Approved Conversions based on Gender

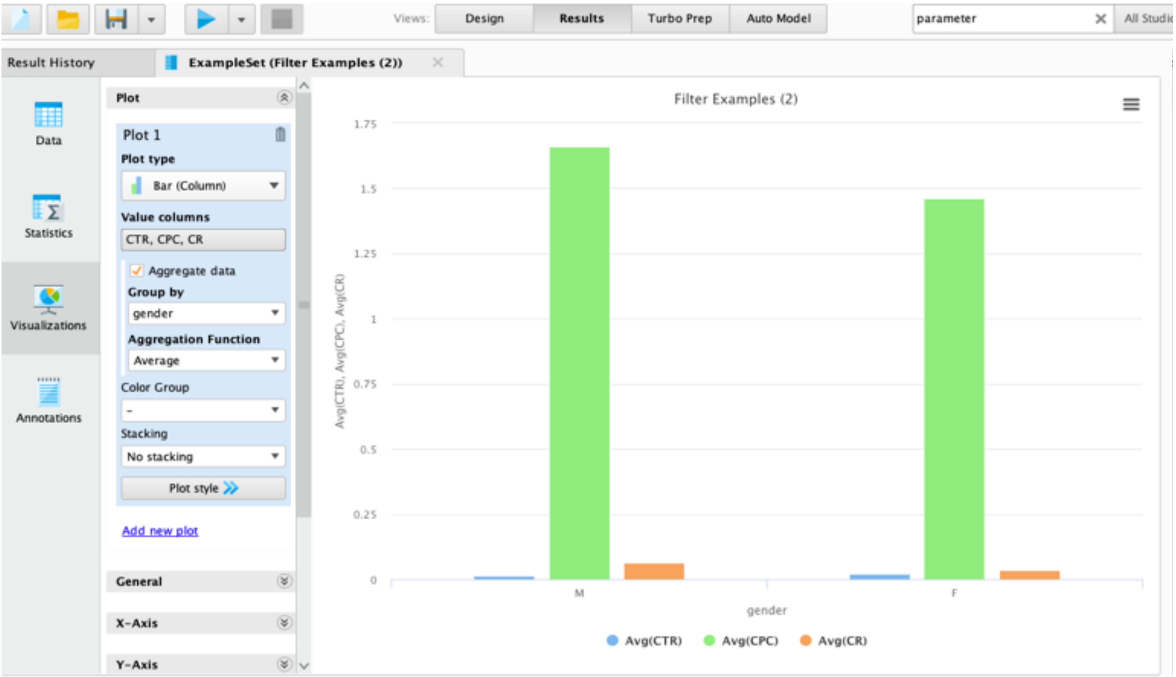


Figure 4.39 Average CTR, CPC & CR

Table 1: Performance Results based on the Gender

Gender	Impressions	Clicks	Spent	Approved conversions	Avg. CTR	Avg. CPC	Avg. CR
Male	96,403,285	13,875	23,609.74	478	0.0135	1.66	0.065
Female	108,375,360	22,193	32,052.41	391	0.0198	1.47	0.0365

Comparison

The [Table 1] compares the outcomes of the campaign performance and the impact that it had separately to women and men. It shows the performance based on values such as impressions, clicks, approved conversions etc. With green color are highlighted the largest values of each attribute. It is clear from the table that the impressions, clicks and spent values regarding the women are higher than the same metrics for men. More specific:

- The campaign was displayed to 11,972,075 more women than men.
- Women clicked 8,318 more times than men.
- The total money spent for women was 8,442.67£ more than money spent for men.

We saw before that the values for impressions, clicks and spent are correlated. Because we had more impressions and clicks for women it is normal that the spent is higher in this category. Regardless the popularity of the campaign to the women, the number of men conversions exceeds the conversions of women. It seems that the campaign was quite attractive for women before the first interaction and more attractive for men at the conversion stage.

We check now the avg. CTR, CPC and CR. The CTR is a little better for the women, the CPC metric is more or less the same for both segments and the CR for men is almost twice than that of women. We reach the conclusion that the campaign had set the same cost for both categories, as the cost per click is very close to both audiences. Last but not least, it had better performance and content for the audience of men.

Recommendations

As we don’t know the context and the goal of the campaign, we cannot be sure about what action could be recommendable. In case the goal of the campaign is the increasement of the conversions, a recommendation could be to decrease budget for women audience and increase the budget for men audience, because men convert more. Also, the campaign should be differentiated between the two groups. Women after clicking the advertisements probably should be directed to another page with a more relative content with their interests in order to convert.

Analyze by Age

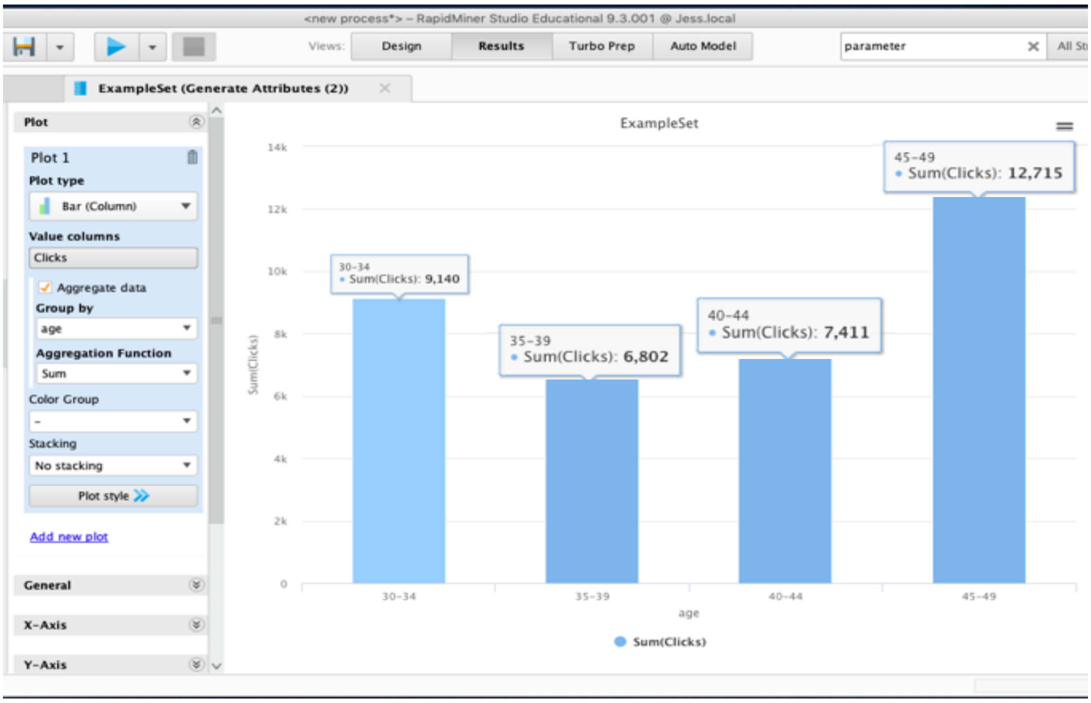


Figure 4.40 Sum of Clicks based on the Age

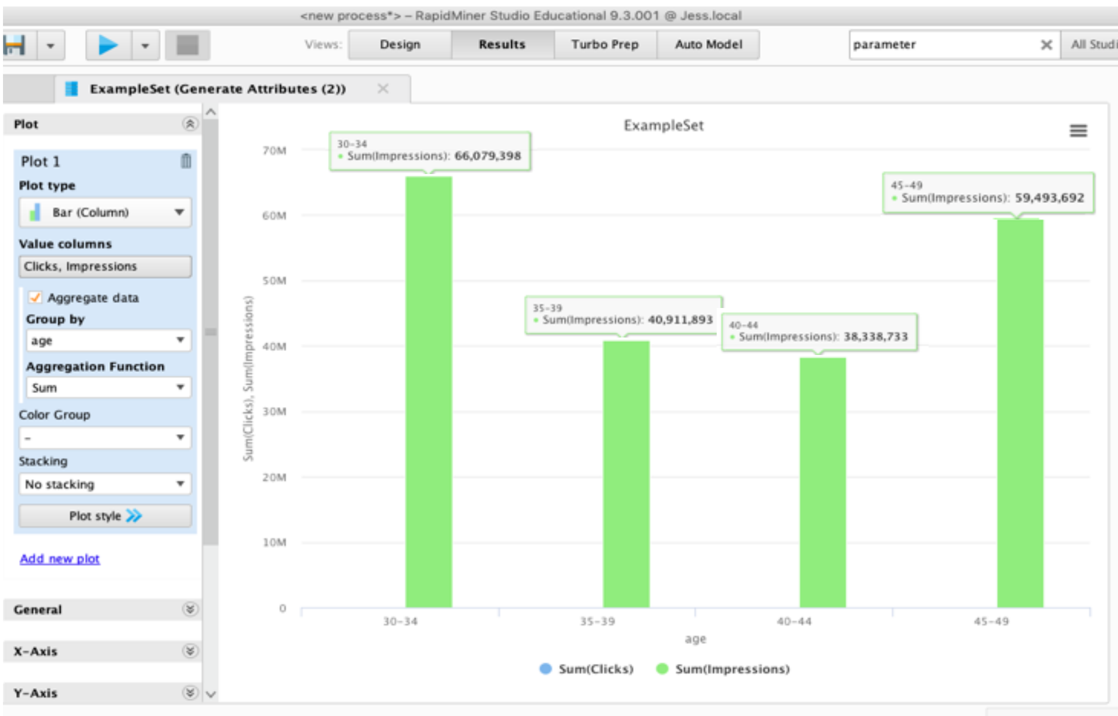


Figure 4.41 Sum of Impressions based on the Age

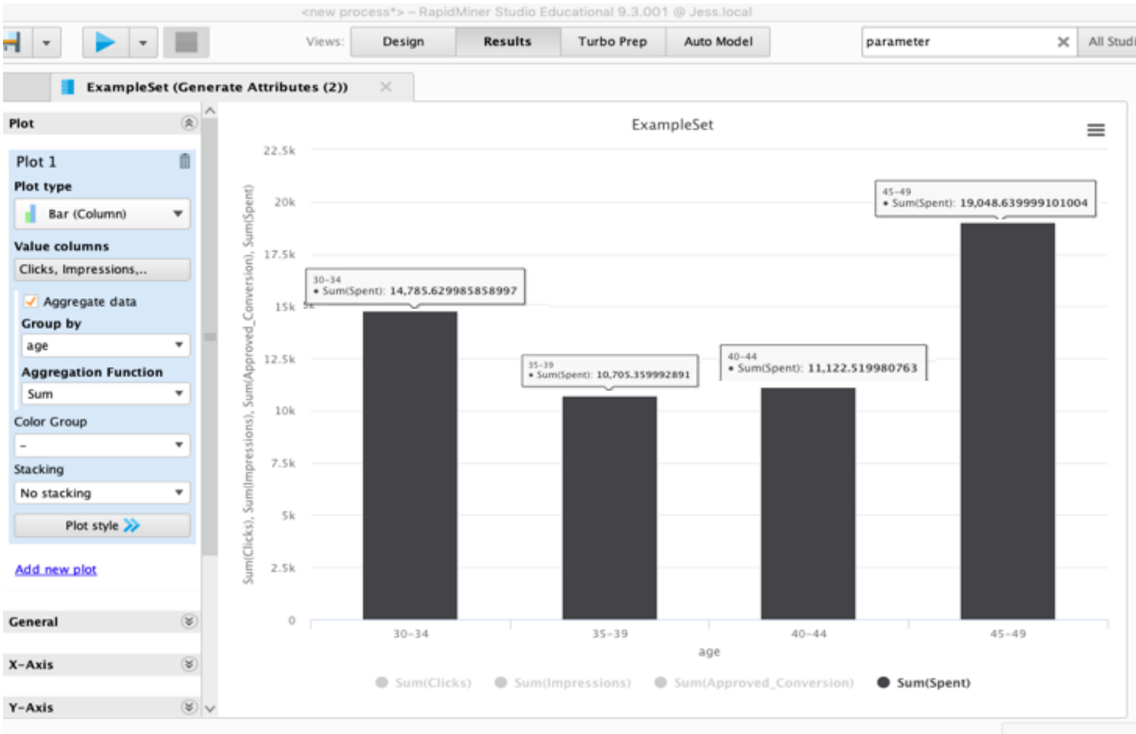


Figure 4.42 Money Spent based on Age

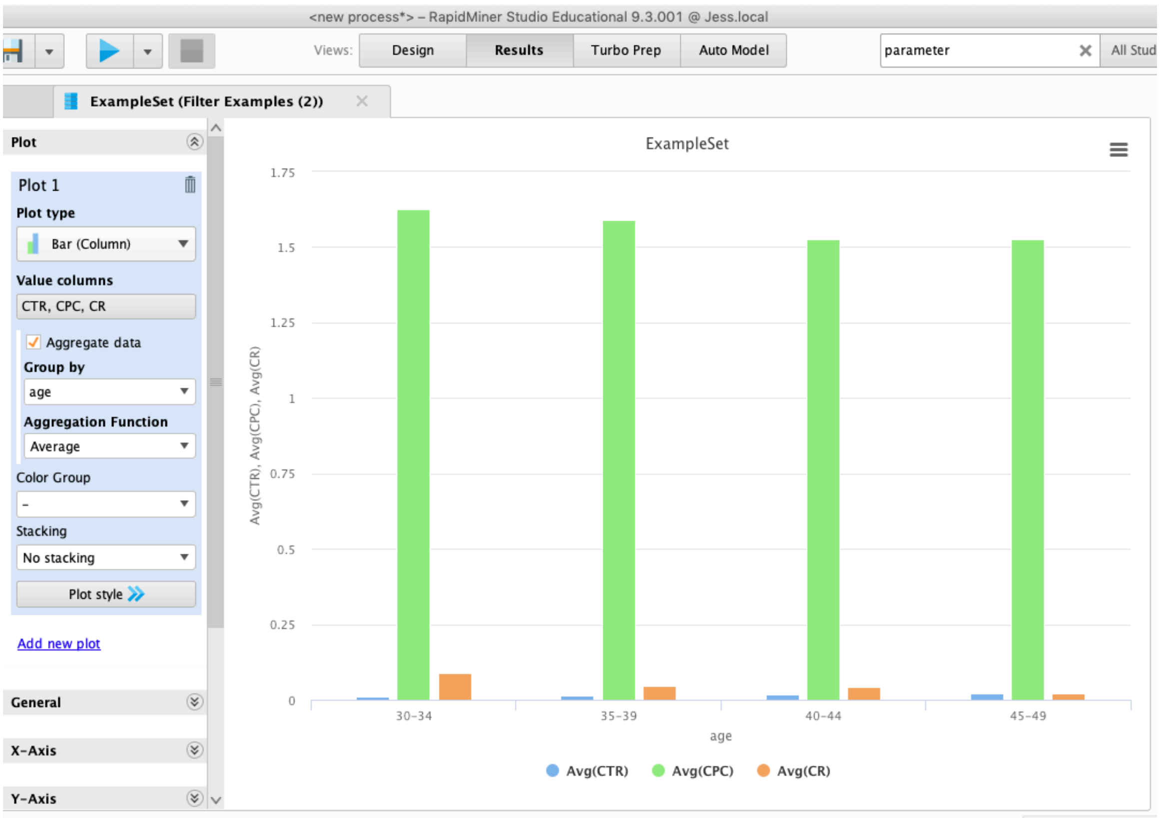


Figure 4.43 Average CTR, CPC & CR

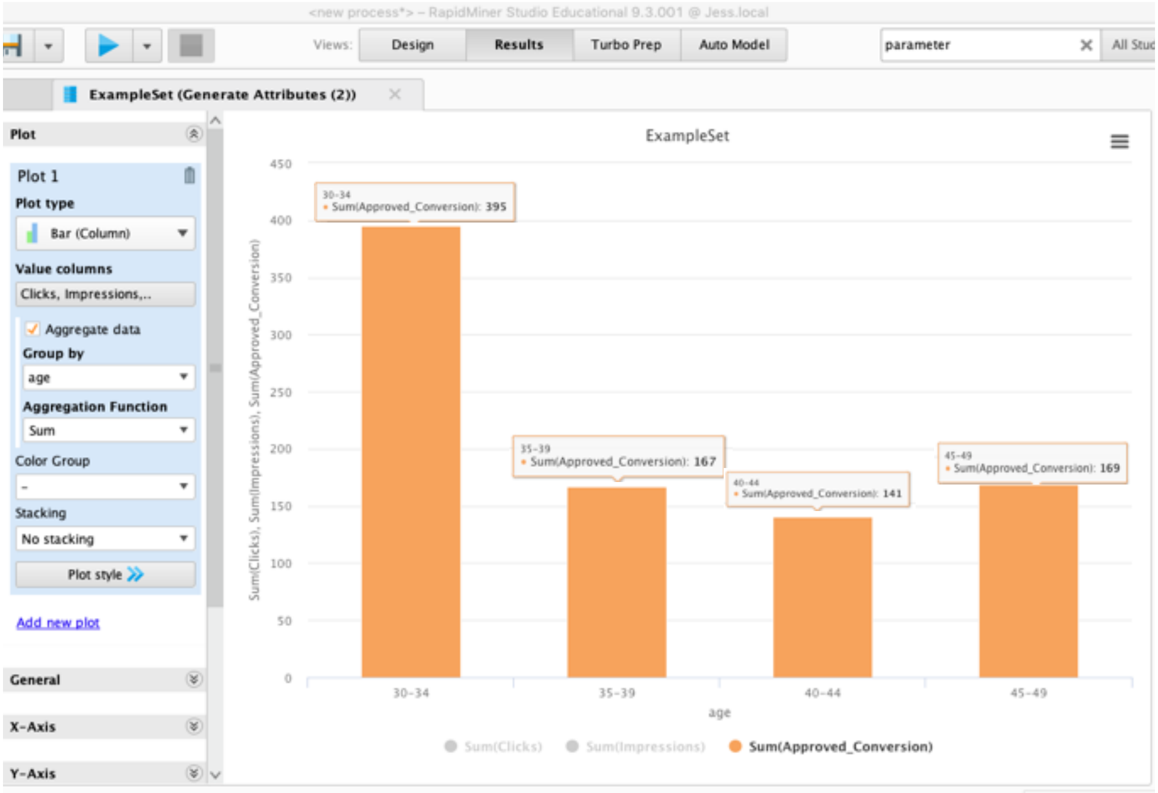


Figure 4.44 Sum of Approved Conversions based on Age

Table 2: Performance Results based on the Age

Age	Impressions	Clicks	Spent	Approved conversions	Avg. CTR	Avg. CPC	Avg. CR
30-34	66,079,398	9,140	14,785.63	392	0.012	1.625	0.088
35-39	40,911,893	6,802	10,705.36	167	0.015	1.592	0.045
40-44	38,338,733	7,411	11,122.52	141	0.019	1.525	0.041
45-49	59,493,692	12,715	19,048.64	169	0.02	1.527	0.022

Comparison

The [**Error! Reference source not found.**] compares the results of the campaign performance and the impact that separately had to each age group. It shows the performance based on values such as impressions, clicks, approved conversions etc. With green color are highlighted the largest values of each attribute, while with red the smallest. We observe that although the campaign displayed mostly to the group 30-34, it was more attractive for the group 45-49 as more users clicked on the adsets. Finally, 392 people by the age of 30-34 converted while only 169 by the age 45-49. Although, in this campaign ads were displayed mostly to people by the age of 30-34 and more visitors converted the cost for the group 45-49 was bigger because contrary to the group age 30-34, more users clicked.

Comparing the avg CTR column, we notice that the group 45-49 has the best CTR of all, as it generated the biggest number of clicks in relation with the impressions that it had, among all age groups. In first place, the campaign was very attractive for this group, but it didn't make them convert. On the other hand, the group 30-34 has the smallest CTR and the biggest CR. That means that the campaign wasn't as attractive as to the group 45-49, but finally more people converted. The results of these metrics confirm our previous observation.

Recommendations

The biggest amount of money was spent for the group 45-59, which at the end had less conversions than the conversions of the age-group 30-34. Also, the last group had the largest CR among the 4 age-groups. Therefore, a recommendation could be to invest more money in this group and at the same time create ads more attractive, in order to make more users at this age to click. On the other hand, the budget for group 44-49 could be reduced and at the same time the campaign's content at the conversion stage could become more attractive in order to make the audience at this age convert.

4.3.2 Analyze by Interest

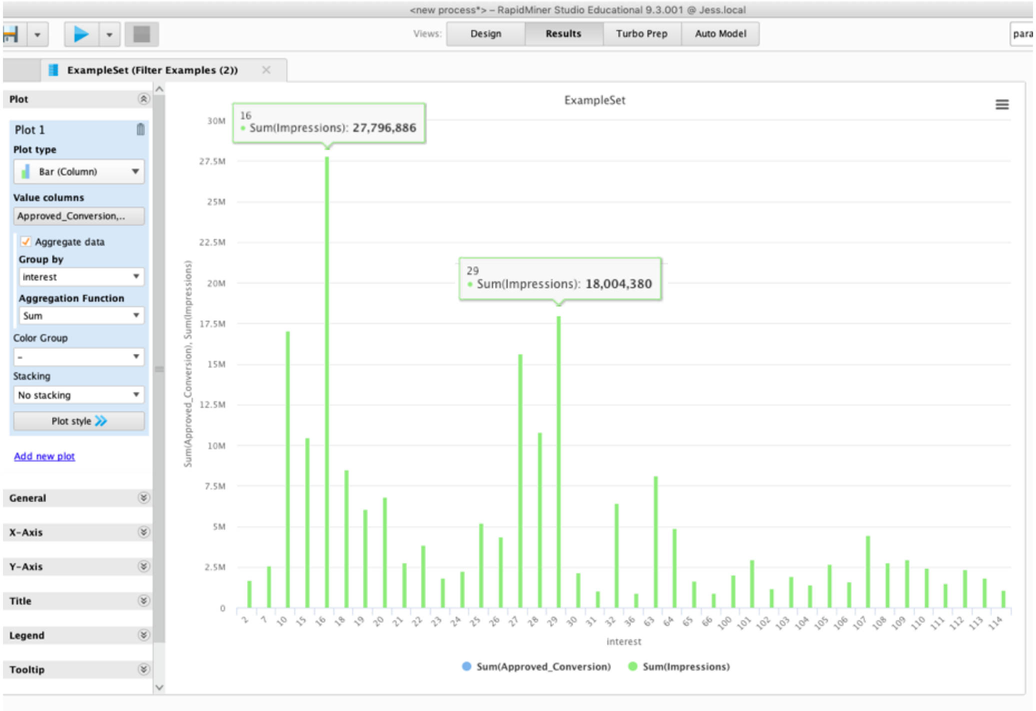


Figure 4.46 Sum of Impressions based on Interest

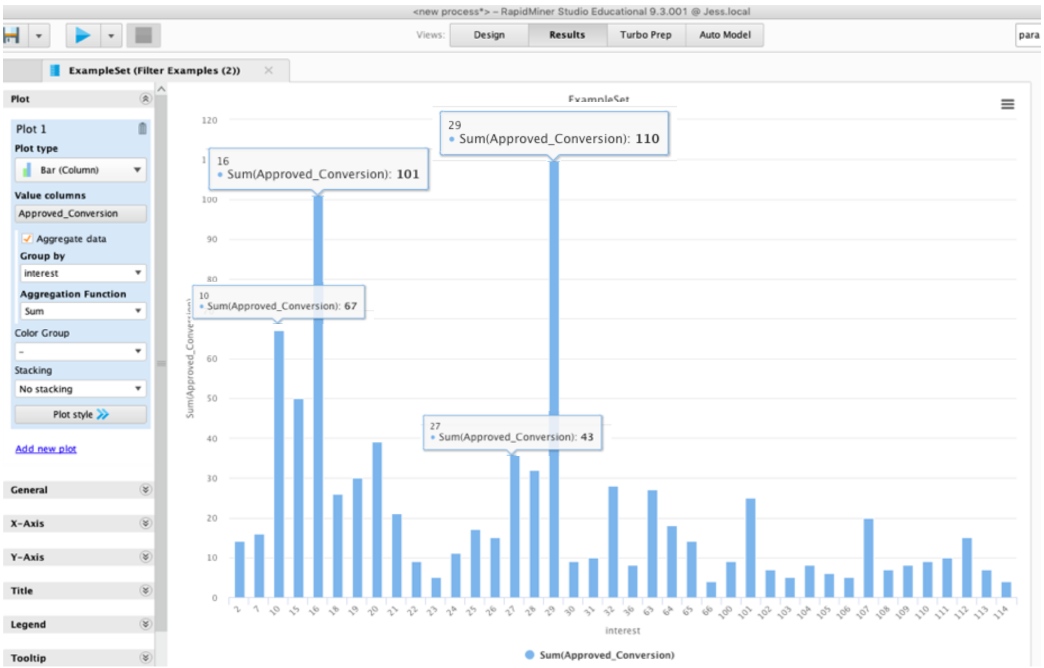


Figure 4.45 Sum of Approved Conversions based on Interest

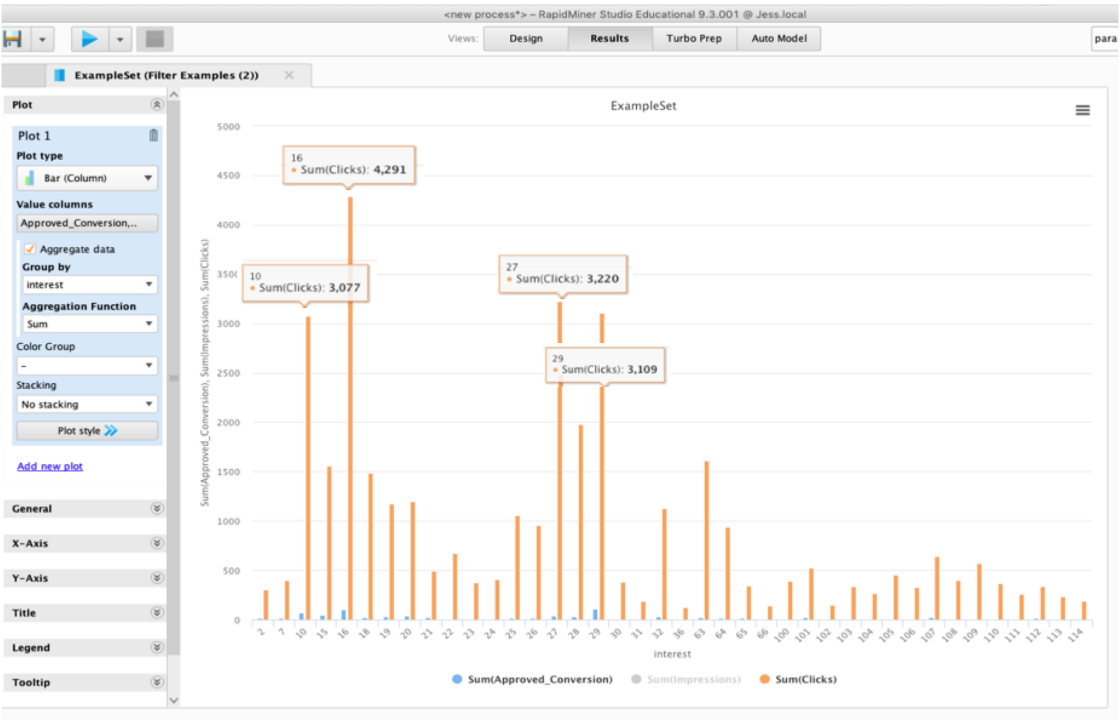


Figure 4.48 Sum of Clicks based on Interest

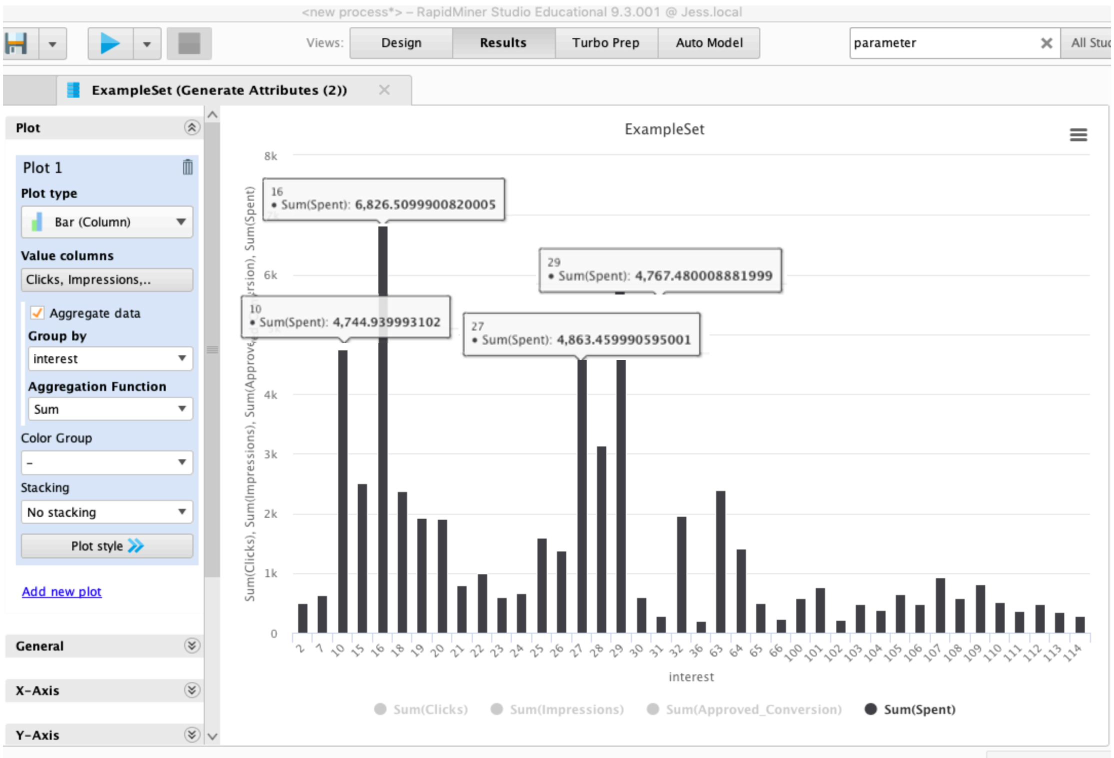


Figure 4.47 Sum of Spent based on Interest

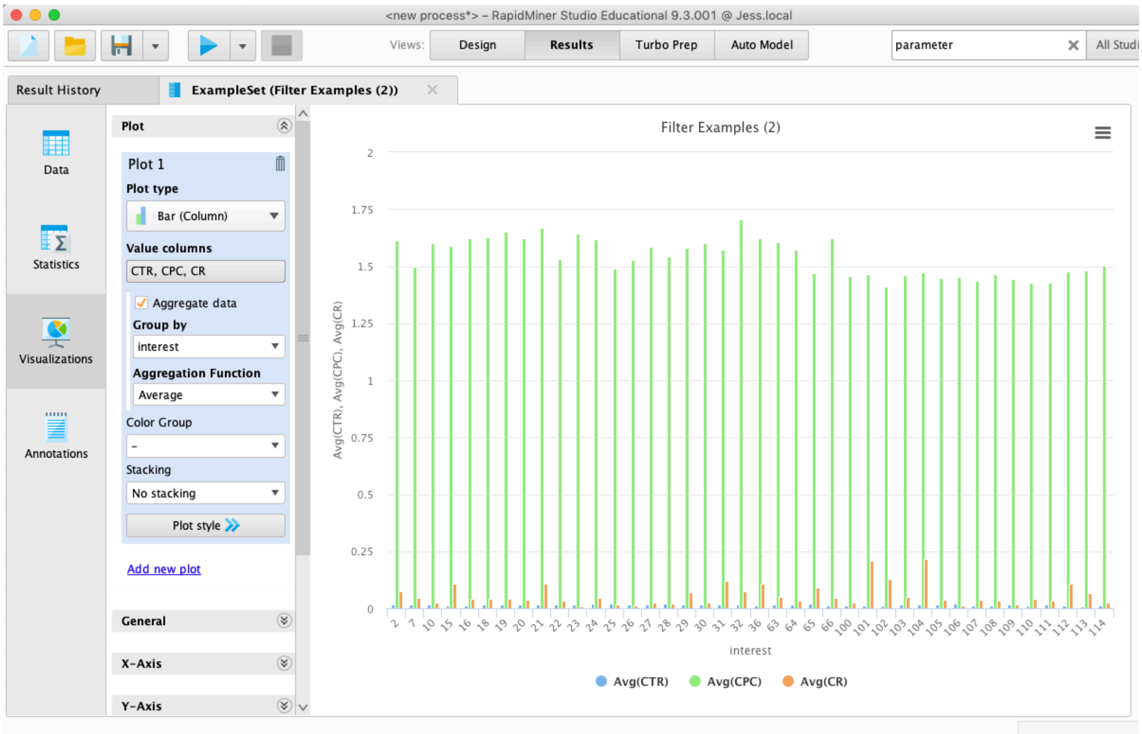


Figure 4.49 Average CTR, CPC & CR

Table 3: Performance Results based on the Interest

Interest	Impressions	Clicks	Spent	Approved conversions	Avg. CTR	Avg. CPC	Avg. CR
10	17.052,327	3.077	4.744,94	67	0.017	1.6	0.026
15	10.458,266	1.554	2.514,3	50	0.0144	1.59	0.109
16	27.796,886	4.291	6.826,51	101	0.0145	1.62	0.0395
27	15.615,162	3.220	4.863,46	43	0.0185	1.585	0.0245
29	18.004,380	3.109	4.767,48	110	0.0155	1.58	0.069

Comparison

We grouped the dataset by interest categories and then we observed the results that each interest has in relation with each different attribute (impressions, clicks, approved conversions, spent, average CR, CPC and CTR). We analyzed the top 5 interests with the most conversions. With green color are highlighted the largest values of each attribute, while with red the smallest. Unfortunately, we don't know the context of each interest, so it is difficult to interpret the results because some of them could refer to interests with similar context, while others could have totally different context. In first case, it is better to handle the similar interest categories in a similar way and group them, while in the second case it is better to treat them differently in order to have better results. In our study we suppose that the interest categories have differences among them and should be treated differently.

According to the [Table 3], the campaign displayed mostly to people with the interest number 16. Then follows the interest 29 and 10 with small difference between them. We notice that the number of impressions that the audience 16 has, is quite bigger than the rest of the interests. The group, which was exposed less to the campaign, is the group having the interest 15. The interest 16 has the biggest number of impressions, clicks and spent, while the interest 15 has the smallest. This is reasonably as previously the correlation matrix table revealed that the values impressions, clicks and spent have a strong correlation among them.

The next column shows the conversions that each different interest group made. We observe that there is different behavior among the different interest categories. The 29-interest group had the most conversions and surpassed the number of conversions that made from the 16-interest group, which was exposed mostly in this campaign. Additionally, was spent around the same amount of money for the groups 27 and 29 because had around the same number of clicks, but the 29 generated more than double conversions than 27. Although the 15-interest group had the lowest values regarding to the impressions, clicks and spent, it made more conversions than the 27-interest group.

The CPC column has more or less the same percentage for each interest. That happens because the budget as we might assume, has been allocated the same for each interest category. We notice that the 27-interest group has the highest CTR average value and the lowest CR. This means that a lot of people were attracted from the campaign and clicked it, but a few actually converted.

The 15 group has the lowest average CTR value, while it has the highest CR. This fact means that not so many people clicked the ads of the campaign, but from those many converted.

Recommendations

Like it mentioned above, we don't know the context of each interest and we assume that each interest is unique and different from the others. Even though the interest 27 had almost the double clicks and money spent from the interest 15, it had less conversions. Therefore, a recommendation could be to increase the budget for targeting users with interest 15, as they convert more and decrease the budget for users with interest 27. If we still believe that the campaign could also be successful to group with interest 27, we should change the content at the conversion stage in order to convince the group to convert. At the same time, ads should be created which will attract more the group with interest 15 and make it click.

Comparing also the clicks, money spent and conversions between the audiences with interests 27, 10 and 29 we observe that the clicks and money spent for the three interests are almost the same, but the conversion number for the interest 29 is almost the double in comparison with each one of the previous two. A recommendation could be to reduce the budget for the first two, while increase the budget for the interest group 29. Additionally, the campaign's content should be differentiated for different groups in different interaction moments.

4.3.3 Group Specific Personas

After analyzing the dataset based on demographics and interests, we are going to divide more the audience and create different personas. In the first table we analyze the ad performance separately for women and men having different ages and the interest 10. In the second table we do the same, but this time each group has the interest 15.

Table 4: Grouped by interest 10

	Sum. Impressions	Sum. Clicks	Sum. Appr. Conversions	Spent	Avg. CTR	Avg. CPC	Avg. CR
Female (30-34)	531,058	76	6	102.63	0.0143	1.44	0.077
Female (35-39)	1,842,153	406	3	565.03	0.0215	1.42	0.011
Female (40-44)	2,256,544	529	7	732.13	0.023	1.389	0.0078
Female (45-49)	3,014,393	759	7	1,037.81	0.025	1.379	0.0100
Total:	**7,644,148**	**1,770**	**23**	**2,437.6**			
Male (30-34)	3,024,116	344	23	624.690	0.0112	1.809	0.06
Male (35-39)	3,042,116	203	7	378.48	0.0108	1.813	0.0254
Male (40-44)	1,679,307	249	6	428.05	0.0149	1.6929	0.0304
Male (45-49)	2,975,363	511	8	870.120	0.0167	1.724	0.0168
Total:	**10,720,902**	**1,307**	**44**	**2,301.34**			

Comparison

Female (30-34)

This group has the less impressions and number of clicks. Therefore, it has the smallest Avg. CTR. On the other hand, it appears to have the best Avg. CR of all the audiences of the table. As we understand, in first place, the ad isn't so attractive for women at this age with this specific interest, but when they open it is quite possible to convert.

Female (35-39)

This group have way more impressions and clicks contrary to the previous group, but it has only 3 conversions. This observation means that the users don't have the appropriate motivation to convert when it is time to convert or that after clicking the ad, they realize that the content isn't in their interest.

Female (40-44) vs Female (40-45)

Comparing the two groups, we observe a difference between them according to the number of impressions, clicks and spent, but they both have the same number of conversions. Even though the number of conversions of women (40-44) is more or less the same with the number of conversions of women (30-34), the cost for the first group is almost 10 times higher.

Male (30-34) vs Male (35-39)

These 2 audiences have the most impressions among the 4 men groups, but not the most clicks. The first group has more clicks contrary to the second. Their biggest difference comes to conversions as male (30-34) seems to convert more.

Male (40-44) vs Male (45-49)

These two groups have very same number of conversions while the number of impressions, clicks and the cost for audience male (45-49) is more than the double.

In general, we observe that even women and men having the same age behave differently. The same happens for the same gender having different ages.

Table 5: Grouped by interest 15

	Sum. Impressions	Sum. Clicks	Sum. Appr. Conversions	Spent	Avg. CTR	Avg. CPC	Avg. CR
Female (30-34)	1,632,981	256	7	383.82	0.0154	1.474	0.028
Female (35-39)	157,467	27	2	39.95	0.0135	1.404	0.27
Female (40-44)	582,725	142	2	194.81	0.0244	1.372	0.0141
Female (45-49)	1,221,803	302	0	430.02	0.0247	1.422	0
Total:	**3,594,976**	**727**	**11**	**1048.60**			
Male (30-34)	4,737,422	495	31	933.01	0.007	1.797	0.246
Male (35-39)	1,150,574	168	4	283.22	0.0140	1.745	0.033
Male (40-44)	163,181	26	1	40.02	0.0159	1.54	0.038
Male (45-49)	812,113	138	3	209.36	0.0157	1.53	0.055
Total:	**6,863,290**	**827**	**39**	**1,465.61**			

Female (30-34)

Women (30-34) have the biggest number of impressions among women. However, the Avg. CR is quiet low.

Female (35-39) vs Female (40-44)

Comparing these two groups, we notice that they have the same number of conversions but there is a big difference between impressions, clicks and cost spent for each group.

Female (40-45)

Although this audience has the biggest number of clicks and the highest cost, it didn't convert at all. That means that ad's content is quite attractive, but after the users click the ad they don't be motivated to convert or they realize that the real content is irrelevant with the advertisement or their expectations.

Male (30-34) vs Male (40-44)

The first audience has the better Avg. CR among all the groups having the interest 15. It also has the biggest number of conversions and at the same time the highest cost. The group male (40-44) has only one conversion and total spent 40.02£ Having in mind only the conversions, it is like 1 conversion cost 40.02, while in first example if we divided the spent with the conversions, each conversion costs around 30.1.

Male (35-39) vs Male (45-49)

These two groups have very similar number of conversions, but quite different number of impressions, clicks and cost.

Recommendations

In the [Table 4: Grouped by interest 10], we observe that more money spent for the women audience rather than men audience. However, out of the 67 conversions, the 44 are made by men. Although the fact that men by the age of 45-49 made 167 more clicks than men by the age of 30-34, the second group made 15 more conversions with less budget. In [Table 5: Grouped by interest 15], we observe that also men by the age of 30-34 has the better performance among all groups.

Since we don't know the context and the goal of the campaign, we cannot understand why the campaign had better conversion outcomes for men audience and especially for men by the age 30-34. If we do the same analysis for the rest of the interest groups and we observe the same pattern, an interpretation could be that the product or the service which be promoted is more attractive for men at this age. In that case we could try to target and give more budget to attract men 30-34. On the other hand, the fact that the women made more clicks to the total adsets of the campaign than men, let us assume that the object of the campaign is also attractive for women. In this case the problem is that more money is spent for women, because when they see for the first time the ad they are attracted, but the majority of them do not finally convert. If this is the case, a recommendation could be to differentiate the links after clicking between men and women. Create content and attract women based on their interests and characteristics and do the same for men. If we continue to observe differences between same gender, same age and different interest we should differentiate the landing pages of the ads based on the persona that we created and not just only based on individual characteristics.

5 Discussion

In this sector, general recommendations about marketing campaign optimization will be given, based on the results of the conducted analysis. Next, the achievements and the limitations of this study will be discussed.

5.1 General Recommendations

A marketing campaign is a complex process which requires many steps of creativity, consideration and data driven decisions. Before the launch of the campaign, marketers should make decisions about the following:

- Set specific and smart goals. What we want to achieve with this campaign and what are the expected results, are good questions that marketers have to answer first. The possible categories of campaign advertising goals are:
  - **Awareness:** Make a big number of people aware that the product exists, without the expectation of them to convert directly. Campaigns having this objective, do not intend a specific action from the user, just to increase awareness. It is recommended for large companies.
  - **Consideration:** In this category we wish to engage more the audience in order to increase the websites' traffic or post's engagement, video views, page likes etc. Again, advertisers try to build awareness, but this time with users' engagement. This engagement could be a click to the site, a like or just a video play.
  - **Conversion:** Conversion category intend to direct the users to another environment and make them complete a procedure, for example download a pdf file, subscribe to a newsletter or buy a product.
- Segment the audience, create different personas and target each different audience with different campaigns.
- Don't waste too quickly the budget. Run A/B tests and decide on alterations and improvements based on the results, before running the campaign. Those alterations could be:
  - Set different budget for different target groups
  - Exclude target audiences which had very low performance and didn't convert
  - Try to alter and improve the content
  - Set different content and landing pages for different target audiences

- Continue to monitor the campaign and make improvements while the campaign is active. In this case, be careful because when a lot of changes made, it takes time to the algorithm to adjust to new requirements and may fail.
- Monitor CTR, CR and CPC, which most of the time are good indicators to understand the progress of the campaign.
- Monitor the whole journey of the users and integrate all the information that you get in your digital campaign strategy. For example:
  - When they visit the webpage
  - Which days of the week or which part of the day they usually convert
  - How long they stay on the page
  - What is their next move after a conversion
  - After how much time they return to the website after a conversion etc.
- Boost the budget on specific days and times within a day.
- Make research and use online tools in order to integrate data from different sources, combine it, create useful insights and use it in the decision-making process.

5.2 Achievements

Our work outlines the advantages of new digital marketing techniques in combination with other data science tools such as RapidMiner. In traditional marketing there wasn't the possibility to gauge the performance of marketing campaigns. With our analysis we prove that in digital marketing not only we have the chance to measure the outcomes, but also to exploit this information in order to make improvements and make better decisions in real time.

Nowadays, marketers and advertisers can use online tools and analyze data without the requirement of programming skills. They can handle the data, find correlations, detect errors etc. In this way, they can extract useful insights about previous performances and make better decisions for future actions. Through our analysis, we achieved to prove how easy is to make a performance data analysis, using an online tool such as RapidMiner.

This work also highlights the importance of visualizations. With only a few clicks we managed to create informative graphs, which helped us to interpret the data and discover useful information about campaigns' performance and users' online behavior. The exploitation of this information can lead to performance improvement. Also, this supports previous findings in the literature review about visualizations. [59, 61, 69]

As proposed in [43], we divided the audience in different personas, using social media. This fact helped us to understand better our customers, their needs and their preferences and be able to give them better future experiences. The creation of personas enhances the delivery of personalized messages to the right target audiences in the right moment.

According to Tull et al.1986 and Fiscer et al. 2011, instead of increasing the total budget, the better allocation based on the different products and regions can be more effective. Today, the advancement of technology gives us the chance to have access to more detailed information about our customers. Through this analysis, we achieved to demonstrate that the unbalanced budget distribution in different audiences based on their performance results, could improve the performance of digital marketing campaigns. Also, our analysis shows a clear advantage of personalization based on a combination of users' habits, demographic characteristics and personal interests. Last but not least, our findings support the importance of the combination of efficient budget allocation and content personalization for better campaign optimization.

5.3 Limitations

Our work clearly has some limitations. Both datasets that were used for this analysis have a small number of observations and aren't a representation of big data. Additionally, when we made an analysis based on a specific attribute or a combination of characteristics, the sample became even smaller and sparse. In this case, we couldn't be sure if the outcomes had a real meaning or came up by chance. Another problem that we faced, was that we didn't know the context and the objectives of the Facebook campaign. This limitation prevented us of having a clear view about the campaign's content and giving recommendations to specific alterations and adjustments regarding the content improvement, budget allocation and target audience splitting. Last but not least, the size and the information of both datasets didn't encourage us to apply machine learning methods and make predictions.

6 Conclusion & Future Work

In this sector, firstly the summary and the conclusion will be presented. Lastly, suggestions about future work and analysis will be recommended.

6.1 Summary and Conclusion

Advancement in technology has created a whole new world, where the big amount of data and the access in technological tools give marketers many opportunities. According to our work, the ability of data selection, analysis and interpretation is the basis to effective decision-making. Marketers should ground their decisions about many aspects on historical performance of previous campaigns. Information about customers' demographic characteristics, could be a great start to analyze, in order to detect patterns, but aren't enough. Today, brands can exploit information about customers' online behavior and interests. This fact can help organizations to understand better the needs and wants of their customers. Deliver relevant messages and content to customers based on their preferences can increase the CTR in a great extent. But is this enough? As we observed in our analysis, audiences with high CTR not always convert. Most of the times, display advertisement campaigns give personalized information to the users, but when the users click the ad, it directs them to pages with generic content. When users realize that the information in the landing page is irrelevant with their interests or it isn't what they were expecting, they probably abandon the page and never convert. In these cases, there is no need to allocate the budget differently or dedicate more budget to this campaign, as the target to attract customers' attention was completed. Here, advertisers should focus on the content that follows the ad.

Personalization isn't only useful in order to attract the attention of the user, but also to keep him engaged during all the steps of his online journey. Also, the placement and the timing of the ad display can lead to different performance outcomes, even though the ad has the same content and appearance. Therefore, even if the budget is allocated in the best way, improvements to content, design and display should also be taken into consideration, in order to improve the performance of the campaign. Knowing what the customers want, what matters to them and what motivates them can help marketers to create messages with great impact.

In general, there are easy solutions that can help marketers to analyze data and make decisions based on the analysis. Also, efficient budget distribution in different campaigns and ad-sets is a critical process that can have a great impact on the campaign performance but not without the contribution of personalized content strategy. The better targeting and customer profiling, the monitoring of past performances, the realization of how, where and when the target audiences react to the display messages, the efficient budget allocation in different target audience based on their responses and the continuous personalization in different stages, is a combination of actions that can optimize the campaign performance and return of investments.

6.2 Future work

The small dataset and the fact that we didn’t have control and knowledge about the content of the campaign, were significant limitations for our analysis. We hope that further study and experiments with a larger dataset will confirm our findings. So, in future analysis, is very important to have access and control of the campaign, in order to practice A/B testing and record the results of different alterations in budget or content. Additionally, further work in the future will help to create an application which will check the campaigns’ content text or image before the launch of the campaign, compare it with the performance of previous campaigns and predict if this content will perform well and to which target audience. In that way, people will be exposed to personalized content and the budget distribution will be distributed in a more efficient way. This analysis requires thousands, or even millions records of campaigns’ information, in order to give good predictions.

Bibliography

1 Suvi Mäkinen (2017). Profit maximization through budget allocation in display advertising. (Master's thesis). Aalto University School of Business Finance

2 Leeflang, P., Verhoef, P., Dahlström, P. and Freundt, T. (2019). Challenges and solutions for marketing in a digital era.

3 Instapage.com. (2019). Changes in Advertising Over Time: How Personalization Keeps Getting Better. [online] Available at: https://instapage.com/blog/changes-in-advertising [Accessed 30 Nov. 2019].

4 Behera, R., Gunasekaran, A., Gupta, S., Kamboj, S. and Bala, P. (2019). Personalized digital marketing recommender engine. Journal of Retailing and Consumer Services, p.101799.

5 Monnappa, A. (2019). The History and Evolution of Digital Marketing. [online] Simplilearn.com. Available at: https://www.simplilearn.com/history-and-evolution-of-digital-marketing-article [Accessed 30 Nov. 2019].

6 Fierro, I., Cardona Arbelaez, D.A., Gavilanez, J., 2017. Digital marketing: a new tool for international education. Pensam. Gest. (43), 241–260.

7 Prassaki, E. (2018). Έως το 2022 το 60% του παγκόσμιου πληθυσμού θα είναι χρήστες του διαδικτύου. [online] Available at: https://gblogs.cisco.com/gr/2018/12/14/vni2019/ [Accessed 30 Nov. 2019].

8 Albuquerque, P., Pavlidis, P., Chatow, U., Chen, K. Y., & Jamal, Z. (2012). Evaluating promotional activities in an online two-sided market of user-generated content. Marketing Science, 31(3), 406–432.

9 Share of internet users who have purchased selected products online in the past 12 months as of 2018. (2019). [online] Statista. Available at: https://www.statista.com/statistics/276846/reach-of-top-online-retail-categories-worldwide/ [Accessed 30 Nov. 2019].

10 Saidali, J., Rahich, H., Tabaa, Y. and Medouri, A. (2019). The combination between Big Data and Marketing Strategies to gain valuable Business Insights for better Production Success. Procedia Manufacturing, 32, pp.1017-1023.

11 Cline, G. (2017). PRODUCT DEVELOPMENT IN THE ERA OF IOT: TYING THE DIGITAL THREAD. [ebook] Available at: https://www.ibm.com/downloads/cas/YPR650OA [Accessed 30 Nov. 2019].

12 Hennig-Thurau, T., Malthouse, E., Friege, C., Gensler, S., Lobschat, L., Rangaswamy, A., et al. (2010). The impact of new media on customer relationships. Journal of Service Research, 13(3), 311–330

13 Sanjay, M. and Alamma, B. (2016). An insight into big data analytics — Methods and application. *2016 International Conference on Inventive Computation Technologies (ICICT)*.

14 Verma, Deepak. (2018). A Critical Review of Digital Marketing. 8. 321-339.

15 Tonkin, Sebastian, Caleb Whitmore, & Justin, Cutroni. (2010). Performance market- ing with Google analytics: Strategies and techniques for maximizing online ROI.Hoboken, NJ: Wiley Pub.

16 Shuteyev, P. (2018). Awareness, Consideration & Decision: Conversion Goals & Means At Each Stage. [Blog] *SnovioLabs*. Available at: https://snov.io/blog/awareness-consideration-decision-what-to-convert-with-at-each-stage/ [Accessed 30 Nov. 2019].

17 Jacobson, J., Gruzd, A. and Hernández-García, Á. (2019). Social media marketing: Who is watching the watchers?. *Journal of Retailing and Consumer Services*.

18 Bhayani, Sanjay. (2018). Internet Marketing vs Traditional Marketing: A Comparative Analysis. 3. 53-63.

19 Jayaram, D., Manrai, A. and Manrai, L. (2015). *Effective use of marketing technology in Eastern Europe: Web analytics, social media, customer analytics, digital campaigns and mobile applications*.

20 Kotler, P., Kartajaya, H. and Setiawan, I. (2017). *Marketing 4.0*. John Wiley & Sons, Inc., Hoboken, New Jersey, p.48.

21 Yang, Yanwu & Wang, Fei-Yue. (2014). Budget Allocation in Competitive Search Advertisements, Part II: In Dynamic Environments. 10.1016/B978-0-12-411457-9.00008-2.

22 HARRIS, D. (2017). 10 Cutting-Edge Digital Marketing Techniques. [Blog] *GetResponse*. Available at: https://www.getresponse.com/blog/digital-marketing-techniques [Accessed 30 Nov. 2019].

23 Kritzinger, W. T., & Weideman, M. (2013). Search Engine Optimization and Pay-per-Click Marketing Strategies. *Journal of Organizational Computing and Electronic Commerce*, *23*(3), 273–286. https://doi.org/10.1080/10919392.2013.808124

24 HubSpot. (2019). 2019 Marketing Statistics, Trends & Data — The Ultimate List of Digital Marketing Stats. Retrieved November 30, 2019, from https://www.hubspot.com/marketing-statistics

25 Kannan, P. K. & Li, Alice. (2017). Digital marketing: A framework, review and research agenda. International Journal of Research in Marketing. 34. 22-45. 10.1016/j.ijresmar.2016.11.006.

26 Bleier, Alexander & Eisenbeiss, Maik. (2015). Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where. Marketing Science. forthcoming. 10.1287/mksc.2015.0930.

27 Tiago, M. T. P. M. B., & Veríssimo, J. M. C. (2014). Digital marketing and social media: Why bother? *Business Horizons*, *57*(6), 703–708. https://doi.org/10.1016/j.bushor.2014.07.002

28 Appelbaum, D., Kogan, A., Vasarhelyi, M., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. *International Journal of Accounting Information Systems*, *25*, 29–44. https://doi.org/10.1016/j.accinf.2017.03.003

29 Chen, Y., Wang, Q., & Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. Journal of Marketing Research,

48(2), 238–254.

30 Ramesh, A., Teredesai, A., Bindra, A., Pokuri, S., & Uppala, K. (2013). Audience segment expansion using distributed in-database k-means clustering. Proceedings of the Seventh International Workshop on Data Mining for Online Advertising - ADKDD ’13. https://doi.org/10.1145/2501040.2501982

31 Miralles-Pechuán, L., Ponce, H., & Martínez-Villaseñor, L. (2018). A novel methodology for optimizing display advertising campaigns using genetic algorithms. Electronic Commerce Research and Applications, 27, 39–51. https://doi.org/10.1016/j.elerap.2017.11.004

32 Sekiguchi, A., & Tsuda, K. (2014). Study on Web Analytics Utilizing Segmentation Knowledge in Business to Business Manufacturer Site. Procedia Computer Science, 35, 902–909. https://doi.org/10.1016/j.procs.2014.08.164

33 Cinstantin Brancoveanu, 2009, Evolution of Digital Marketing. University Pitesti / Faculty of Administrative and Communication Sciences Braˇila, Constantin Braˆncoveanu University Pitesti / Faculty of Management Marketing in Business Affairs Braˇila

34 Delen, D., Moscato, G., & Toma, I. L. (2018). The impact of real-time business intelligence and advanced analytics on the behaviour of business decision makers. 2018 International Conference on Information Management and Processing (ICIMP). https://doi.org/10.1109/icimp1.2018.8325840

35 Delen, D. (2014). Real-World Data Mining: Applied Business Analytics and Decision Making. New Jersey: Pearson Education, Inc.

36 J. Burby and A. Brown (2007), Web Analytics Association – Web Analytics Definitions, Version 4.0 (Accessed March 2016) from http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAn alyticsDefinitionsVol1.pdf

37 Waisberg, D., Kaushik, A., 2009. Web Analytics 2.0: Empowering Customer Centricity, SEMJ.org, V. 2 I. 1, 7p.

38 Maria, G. (2018, September 18). How to use predictive analytics: 5 steps to get started. Retrieved November 30, 2019, from https://lab.getapp.com/how-to-use-predictive-analytics/

39 Phippen, A., Sheppard, L., & Furnell, S. (2004). A practical evaluation of Web analytics. Internet Research, 14(4), 284–293.

40 Laurindo, Fernando. (2011). A Framework of Web Analytics - Deploying the Emergent Knowledge of Customers to Leverage Competitive Advantage.. ICE-B 2011 - Proceedings of the International Conference on e-Business. 125-130.

41 Onwubiko, Cyril. (2016). Exploring web analytics to enhance cyber situational awareness for the protection of online web services. 1-8. 10.1109/CyberSecPODS.2016.7502355.

42 Russell, M. (2010). A call for creativity in new metrics for liquid media. Journal of Interactive Advertising, 9(2), 44–61.

43 Jansen, B. J., Jung, S.-G., Salminen, J., An, J., & Kwa, H. (2017). Leveraging Social Analytics Data for Identifying Customer Segments for Online News Media. 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA). https://doi.org/10.1109/aiccsa.2017.64

44 Bengel, Andy & Shawki, Amin & Aggarwal, Dippy. (2015). Simplifying web analytics for digital marketing. 1917-1918. 10.1109/BigData.2015.7363968.

45 Sivarajah, U., Irani, Z., Gupta, S., & Mahroof, K. (2019). Role of big data and social media analytics for business to business sustainability: A participatory web context. *Industrial Marketing Management*. https://doi.org/10.1016/j.indmarman.2019.04.005

46 Song, S.-K., Kim, D. J., Hwang, M., Kim, J., Jeong, D.-H., Lee, S., … Sung, W. (2013). Prescriptive Analytics System for Improving Research Power. *2013 IEEE 16th International Conference on Computational Science and Engineering*. https://doi.org/10.1109/cse.2013.169

47 Davis, C. K., "Beyond data and analytics", Communication of the ACM, 57, 39 –41, 2014.

48 Laney, D., 2001. 3D data management: controlling data volume, velocity and variety. In: META Group Research Note. 6. pp. 70.

49 Redman, Thomas C., 2013. Data Driven: Profiting From Your Most Important Business Asset. Harvard Business Press (ISBN 978-1-4221-6364-1).

50 Zhang, J., Yang, X., Appelbaum, D., 2015. Toward effective big data analysis in continuous auditing. Account. Horiz. 29 (2), 469–476.

51 Ghani, N. A., Hamid, S., Targio Hashem, I. A., & Ahmed, E. (2018). Social media big data analytics: A survey. *Computers in Human Behavior*, *101*, 417–428. https://doi.org/10.1016/j.chb.2018.08.039

52 IBM, 2013. Descriptive, predictive, prescriptive: transforming asset and facilities management with analytics. In: Thought Leadership White Paper, (October 2013).

53 Simpao, A. F., Ahumada, L. M., Gálvez, J. A., & Rehman, M. A. (2014). A review of analytics and clinical informatics in health care. Journal of Medical Systems, 38(4), 45.

54 T. H. Davenport and J. G. Harris, Competing on Analytics: The New Science of Winning. Harvard Business School Press, 2007, vol. 8, no. 5.

55 Delen, D., & Demirkan, H. (2013). Data, information and analytics as services. Elsevier.

56 Bradlow, Eric & Gangwar, Manish & Kopalle, Praveen & Voleti, Sudhir. (2017). The Role of Big Data and Predictive Analytics in Retailing. Journal of Retailing. 93. 79-95. 10.1016/j.jretai.2016.12.004.

57 Kune, L. F., & Meyers, J. (2018). *A Marketer's Guide to What Is — and Isn't — a Customer Data Platform*. Retrieved from https://www.gartner.com/en/documents/3865417

58 Shanthi, C., & Pappa, N. (2017). An artificial intelligence based improved classification of two-phase flow patterns with feature extracted from acquired images. ISA Transactions, 68, 425–432.

59 Kokina, J., Pachamanova, D., & Corbett, A. (2017). The role of data visualization and analytics in performance management: Guiding entrepreneurial growth decisions. *Journal of Accounting Education*, *38*, 50–62. https://doi.org/10.1016/j.jaccedu.2016.12.005

60 Humby, C., Hunt, T., & Philips, T. (2008). Scoring points: How Tesco continues to win customer loyalty. London: Kogan Page Limited.

61 Bose, Indranil & Mahapatra, Radha. (2001). Business data mining - A machine learning perspective. Information & Management. 39. 211-225. 10.1016/S0378-7206(01)00091-X.

62 SAS, "Big data meets big data analytics: Three key technologies for extracting real-time business value from the big data that threatens to overwhelm traditional computing architectures", http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/big-data- meetsbig-data-analytics-105777.pdf. , 2014.

63 P. Langley, H.A. Simon, Applications of machine learning and rule induction, Communications of the ACM 38 □11), 1995, pp. 55±64.

64 Meister, Maximilian & Beßle, Julia & Cviko, Amir & Böing, Tobias & Metternich, Joachim. (2019). Manufacturing Analytics for problem-solving processes in production. Procedia CIRP. 81. 10.1016/j.procir.2019.03.001.

65 E.Turban,J.E.Aronson,T.-P.Liang,DecisionSupportSystemsandIntelligentSystems, seventh ed., Prentice-Hall, Upper Saddle River, NJ, 2004.

66 TechAmerica Foundation, Demystifying big data: A practical guide to transforming the business of Government. 2012 http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica- bigdatareport-final.pdf VanBoskirk, TechAmerica Foundation's Federal Big Data Commission.

67 Järvinen, J., & Karjaluoto, H. (2015). The use of Web analytics for digital marketing performance measurement. *Industrial Marketing Management*, *50*, 117–127. https://doi.org/10.1016/j.indmarman.2015.04.009

68 Germann, F., Lilien, G., & Rangaswamy, A. (2013). Performance implications of deploying marketing analytics. International Journal of Research in Marketing, 30(2), 114–128.

69 Lurie, Nicholas & Mason, Charlotte. (2007). Visual Representation: Implications for Decision Making. Journal of Marketing. 71. 160-177. 10.1509/jmkg.71.1.160.

70 Chaffey, D., & Patron, M. (2012). From web analytics to digital marketing optimization:Increasing the commercial value of digital analytics. Journal of Direct, Data and Digital Marketing Practice, 14(1), 30–45.

71 Yang, Yanwu & Zeng, Daniel Dajun & Yang, Yinghui & Zhang, Jie. (2015). Optimal Budget Allocation Across Search Advertising Markets. INFORMS Journal on Computing. 27. 285-300. 10.1287/ijoc.2014.0626.

72 Archak, Nikolay & Mirrokni, Vahab & Muthukrishnan, Senthilmurugan. (2019). Budget Optimization for Online Advertising Campaigns with Carryover Effects.

73 Negash, S., & Gray, P. (2008). Business intelligence. In F. Burstein, & C. W. Holsapple (Eds.), Handbook on decision support systems 2 (pp. 175–193). Berlin, Heidelberg: Springer.

74 Gartner IT Glossary. (2013). Definition of business intelligence.. Retrieved from: http://www.gartner.com/it-glossary/business-intelligence-bi/ (accessed on 15.04.16.)

75 Halpern, F. (2015). Next-generation analytics and platforms for business success: tDWI research report. Available on: www.tdwi.org

76 Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. International Journal of Information Management, 36(5), 700–710. https://doi.org/10.1016/j.ijinfomgt.2016.04.013

77 Ekinci, Y., Ülengin, F., Uray, N., & Ülengin, B. (2014). Analysis of customer lifetime value and marketing expenditure decisions through a Markovian-based model. European Journal of Operational Research, 237(1), 278–288. https://doi.org/10.1016/j.ejor.2014.01.014

78 Wahbeh, Abdullah & Al-Radaideh, Qasem & Al-Kabi, Mohammed & Al-Shawakfa, Emad. (2011). A Comparison Study between Data Mining Tools over some Classification Methods. International Journal of Advanced Computer Science and Applications - IJACSA. 2. 10.14569/SpecialIssue.2011.010304.

79 Naik, A., & Samant, L. (2016). Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science, 85, 662–668. https://doi.org/10.1016/j.procs.2016.05.251

80 Rangra, Kalpana and Kishori Lal Bansal. “Comparative Study of Data Mining Tools.” (2014).

81 Miralles-Pechuán, L., Rosso, D., Jiménez, F., Garc ́a, J.M., 2017. A methodology based on deep learning for advert value calculation in cpm, cpc and cpa networks. Softw. Comput. 21 (3), 651–665.

82 Bow, C. (2018, February 26). An introduction to Facebook ad analysis using R. Retrieved December 1, 2019, from https://www.kaggle.com/chrisbow/an-introduction-to-facebook-ad-analysis-using-r