Proceedings of the International Conference on Industrial Engineering and Operations Management
Pilsen, Czech Republic, July 23-26, 2019

Applicable Models of Customer Analytics for a Retail Company in Mexico

Daniela Garza Gutiérrez
Engineering Management Academic Program
Universidad de Monterrey
Nuevo León, México
daniela.garzag@udem.edu

Juan Ignacio González Espinosa
Engineering Department
Universidad de Monterrey
Nuevo León, México
juan.gonzaleze@udem.edu

Luz María Valdez de la Rosa
Engineering Department
Universidad de Monterrey
Nuevo León, México
luz.valdez@udem.edu

Abstract

Big Data has become a worldwide tendency, having strong presence in the technological sphere as well as an increased growth in all market sectors. In this final evaluation project, extensive exploration on Big Data literature, predictive analytics, client analytics, as well as that of tools and technologies associated with the compilation and processing of mass data was undertaken. Furthermore, the elaboration of a data analysis model is being performed for a company within the ambit of the retail sector in Mexico. The abovementioned, through the development of a pilot test, which consists of three general stages, identification of the client, validation and refinement of the model and a forecast of new clients. The pilot test implies the formulation, refinement and reading of data of mass scale. This is undertaken by making use of three analyses: discriminant analysis, hierarchical cluster and k – media cluster. The methodology employed is DMADV, which is used where there is a need of designing or a re-designing of products and/or processes or, such as this case, the process of data analysis. As a result, a model capable of identifying 5 different segments, which the potential of providing analytical capacities in order to know, grow, monitor and maintain e-commerce clients within the retail business of Mexico.

Keywords
Big data analytics, customer analytics, descriptive analytics, predictive analytics.

1. Problem

Companies, in order to survive the competitive environment, must transform their management practices and focus on generating a competitive advantage that differentiates them from their competition. To achieve such differentiation, companies must create value in their services, that now, giving to the age of the customer (Forrester, 2013), is to get to know their customers, to understand their needs, it is essential to establish personalized relationships that allow companies to obtain such knowledge.

Currently, the retail company in Mexico seeks to attract customers and keep them through the quality and freshness of its products, complemented with an excellent service. However, they have no knowledge of how customers are buying, since they do not have them identified in clusters or groups with profiles. They only monitor their customer's transactions. Within the company, the customer's information is centralized, however, there's no vision of how to make an exploitation of this information within the different areas of the business.

For purposes of the project and based on the transactions made by the customers is necessary to identify where are those transactions carried out, how often, what kind of transactions, and identify buying patterns to provide a more efficient service and define better marketing strategies. Likewise, it's important to know how profitable our client is at this time, how profitable can be in the future, and how to make the segmentation and profiling of customers.

2. Objectives

2.1 General objectives

Provide the analytical capacities in order to know, grow, monitor and maintain e-commerce clients within the retail business in Mexico.

2.2 Particular objectives

- Identify applicable customer analytical models through research and analysis of relevant literature, industry practices and activities of the same retail in USA.
- Design and develop a data analytics model for e-commerce clients.
- Develop a pilot test model

3. Literature Review

According to Cuzzocrea, Song, and Davis, (2011), "Big Data" refers to enormous amounts of unstructured data produced by high-performance applications falling in a wide and heterogeneous family of application scenarios: from scientific computing applications to social networks, from e-government applications to medical information systems, and so forth. In this case, the application in the retail company was of an e-commerce application scenario. Chen, Chiang and Storey, (2012), mentions that the excitement surrounding business intelligence and analytics (BI&A), and big data has arguably been generated primarily from the web and e-commerce communities.

Data stored in the underlying layer of all these application scenarios have some specific characteristics in common, among them: *large-scale data*, which refers to the size and the distribution of data repositories; *scalability issues*, which refers to the capabilities of applications running on large-scale, enormous repositories to scale over growing-in size inputs rapidly; supporting advance *Extraction-Transformation-Loading* (ETL) *processes* from low-level, raw data

to somewhat structured information; designing and developing *easy and interpretable analytics* over big data repositories in order to derive intelligence and extract useful knowledge from them. (Cuzzocrea, Song, and Davis, 2011).

Specifically, Akter and Wamba, (2016), mentions that in the e-commerce context, big data enables merchants to track each user's behavior and connect the dots to determine the most effective ways to convert onetime customers into repeat buyers. Big data analytics enables e-commerce firms to use data more efficiently, drive a higher conversion rate, improve decision making and empower customers. Gandomi and Haider, (2014), says that to enable such evidence-based decision making, organizations need efficient processes to turn high volumes of fast-moving and diverse data into meaningful insights. The overall process of extracting insights from big data can be broken down into five stages shown in Figure 1. These five stages from two main sub-processes: data management and analytics. Data management involves processes and supporting technologies to acquire and store data and prepare and retrieve it for analysis. Analytics, on the other hand, refers to techniques used to analyze and acquire intelligence from big data.

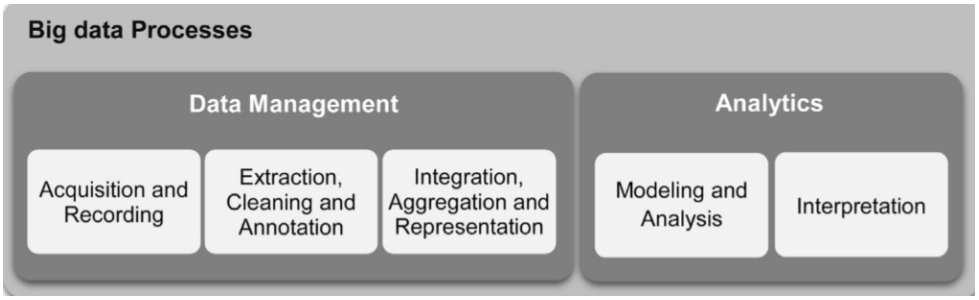


Figure 1: Processes for extracting insight from big data.

As well, Fan, Lau and Zhao, (2015), mention that with big data analytic technologies, key factors for strategic marketing decisions, such as customer opinions toward a product, service, or company, can be automatically monitored by mining social media data, creating marketing intelligence.

Big data analytics have been embraced as a disruptive technology that will reshape business intelligence, which is a domain that relies on data analytics to gain business insights for better decision-making. (Fan, Lau, Zhao, 2015). Other authors as, Sun, Zou, and Strang, (2015), agrees with the above mentioned, saying that big data and big data analytics has become one of the important research frontiers, and that these emerging technologies are and will make big changes in the way e-commerce and e-services operate.

According to a study of Gartner, worldwide business intelligence and analytics software, consisting of BI platforms, analytic applications and advance analytics totaled $14.4 billion in 2013, an 8% increase from 2012 revenue.

The principal tools for BI include software for database query and reporting tools for multidimensional data analysis, data mining and data warehousing (Sun, Zou and Strang, 2015). Big data analytics is mainly concerned with three types of challenges: storage, management, and processing. (Fan, Lau, Zhao, 2015).

There´s an innovation term known as "Database as a Service" (DaaS), this term defines a set of tools that provide final users unified mechanisms for creating, storing, accessing and managing their proper databases on remote servers, DaaS, according to Cuzzocrea, Song, and Davis, (2011), is the most appropriate computational data framework to implement big data repositories. As a data-centric approach, business intelligence and analytics (BI&A), has its roots in the enduring database management field. It relies heavily on various data collection, extraction and analysis technologies. (Chen, Chiang, Storey, 2012).

As data analytics capabilities become more accessible, data scientists need a methodology capable of providing a strategy, regardless of the technologies, data volumes or approaches involved. This methodology support data mining and several new practices in data science such as the use of large data volumes, the incorporation of text analytics into predictive modeling and the automation of some processes. (Rollins, 2015). The predictive modeling or predictive

analytics comprise a variety of techniques that predict future outcomes based on historical and current data. In practice, predictive analytics can be applied to almost all disciplines, among them, predicting customer´s next moves based on what they buy, when they buy and even what they say on social media. (Gandomi, Haider, 2014).

The methodology shown in Figure 2 consists of 10 stages: (1) Business understanding, (2) analytical approach, (3) Data requirements, (4) Data collection, (5) Data understanding, (6) Data preparation, (7) Modeling, (8) Evaluation, (9) Deployment, (10) Feedback. (Rollins, 2015).

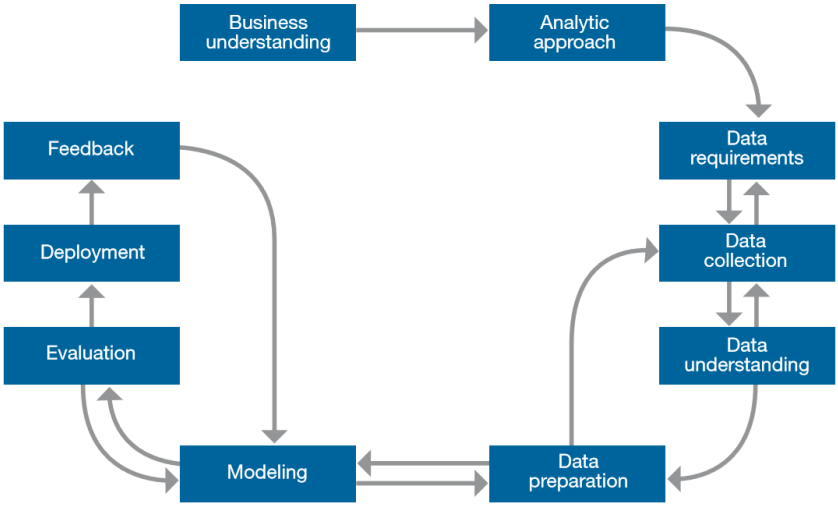


Figure 2: Foundational Methodology for Data Science

This methodology illustrates the process of problem solving, the more the data scientist learn about the data, they frequently return to a previous stage to adjust. Models are not created once, instead, through feedback, refinement and redeployments, models are continually improved and adapted to evolving conditions. This way the model will be providing continuous value to the organization for as long as the solution is needed.

However, in the marketing intelligence purpose, Fan, Lau, and Zhao, (2015), propose the following framework for different applications.

	People	Product	Promotion	Price	Place
Data	• Demographics • Social Networks • Customer Review • Click Stream • Survey Data	• Product Characteristics • Product Category • Customer Review • Survey Data	• Promotional Data • Survey Data	• Transactional Data • Survey Data	• Location-based social networks • Survey Data
Method	• Clustering • Classification	• Association • Clustering • Topic Modeling	• Regression • Association • Collaborative Filtering	• Regression • Association	• Regression • Classification
Application	• Customer Segmentation • Customer Profiling	• Product Ontology • Product Reputation	• Promotional Marketing Analysis • Recommender Systems	• Pricing Strategy Analysis • Competitor Analysis	• Location-based Advertising • Community Dynamic Analysis

Figure 3: A marketing mix framework for big data management

The 4P model has been the most relevant for consumer marketing. However, it´s production-oriented, researchers proposed a fifth P (people). Fan, Lau and Zhao, (2015), adopted the 5P model of the marketing mix because these perspectives play critical roles in developing successful marketing strategies. Figure 3 is an overview of the big data management framework for marketing intelligence. First, data from various sources are retrieved and utilized to generate vital marketing intelligence. Second, a variety of analytics methods are applied to convert raw big data to actionable marketing knowledge. Finally, both data and methods are combined to support marketing applications with respect to each perspective of the marketing mix model.

E-commerce refers to the online transactions: selling goods and services on the internet. In e-commerce, data are the key to track consumer shopping behavior to personalize offers, which are collected over time using consumer browsing and transactional points. The ultimate challenge of big data analytics (BDA) is to generate business value from this explosion of big data. "Value" referred as the generation of economically worthy insights and/or benefits by analyzing big data through extraction and transformation. For example, by injecting analytics into e-commerce, managers could derive overall business value by serving customer needs (79%); creating new products and services (70%); expanding into new markets (72%); and increasing sales and revenue (76%). Amazon is a classic example of enhancing business value and firm performance using big data. The first application of big data for e-commerce firms is the provision of personalized service or customized products. (Akter, Wamba, 2016).

The e-commerce firms deal with both structured and unstructured data. Structured data focuses on demographic data including name, age, gender, date of birth, address, and preferences, unstructured data includes clicks, likes, links, tweets, voices, etc. The challenge is to deal with both types of data in order to generate meaningful insights to increase conversions.

Researchers often come across situations best resolved by defining group of homogeneous objects, whether they are individuals, firms, or even behaviors. The most commonly used technique for this purpose is cluster analysis. According to Hair, Black, Babin and Anderson, (2010), cluster analysis is a group of multivariate techniques whose primary purpose is to group objects based on the characteristics they possess. The concept of the variate is important, it represents a mathematical representation of the selected set of variables which compare the object´s similarities. Cluster analysis work performing a task innate to all individuals, pattern recognition and grouping. (Hair, Black, Babin and Anderson, 2010).

4. Methodology

The methodology used to solve the problem was the CRISP-DM (Cross Industry Standard Process for Data Minning) methodology. According to Wirth and Hipp, (n.d.), the methodology defines a process model which provides a framework for carrying out data mining projects. The CRISP-DM process model aims to make large data mining projects, less costly, more reliable, more repeatable, more manageable, and faster.

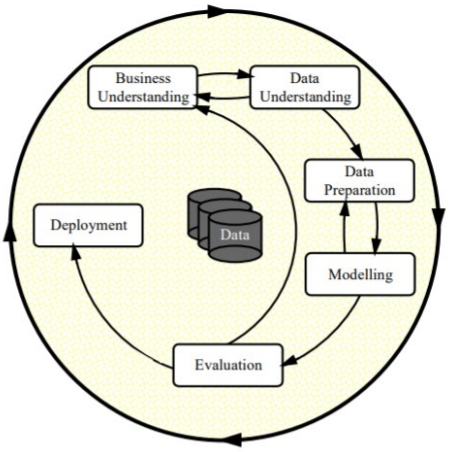


Figure 4: Phases of the CRISP-DM Process Model for Data Mining

This model provides an overview of the life cycle of a data mining project. It contains the phases of a project, their respective tasks, and their outputs.

In the following, there's the outline of each phase:

- Business Understanding

The initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary project plan designed to achieve objectives.

The main objective of the company was to increase loyalty among their customers. They wanted to identify groups of customers according to their buying patterns and profiles and have the capacity to make analysis in order to make relationships among them.

- Data Understanding

This phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

The company's data base was at a transaction level. The information available was: order ID, client ID, branch office, SKU, family and department, price, quantity, type of delivery, date and sale promotion. About demographic data about the customer: gender, first and last name, birth date, and marital status.

- Data Preparation

The data preparation phase covers all activities to construct the final dataset. Tasks include table, record and attribute selection, data cleaning, construction of new attributes, and transformation of data for modeling tools.

In Figure 5 there's the process of how data was collected.

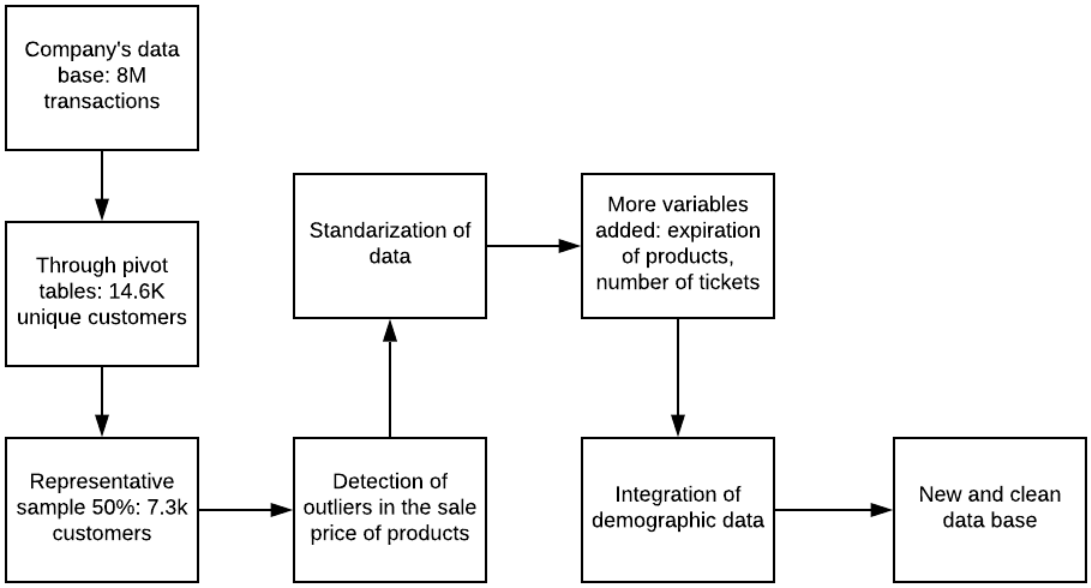


Figure 5: Data preparation

- Modeling

Modeling techniques are selected and applied, and their parameters are calibrated to optimal values.

The cluster analysis was selected as the main technique, then it proceeded to selecting the clustering variables:

- x_1 - number of products;
- x_2 - price;
- x_3 - number of tickets;
- x_4 - discounts;
- x_5 - expiration of products.

Pooled Within Groups Matrices						
		Average # of Products	Average Price	Average Expiration Date of Products	Sum of $ of discounts	Number of tickets
Correlation	Average # of Products	1.000	.108	-.250	.023	.103
	Average Price	.108	1.000	-.078	-.010	.011
	Average Expiration Date of Products	-.250	-.078	1.000	.014	-.110
	Sum of $ of discounts	.023	-.010	.014	1.000	.041
	Number of tickets	.103	.011	-.110	.041	1.000

Table 6: Variable Correlation

After selecting the clustering variables, the next step is measuring similarity. According to Hair, Black, Babin, and Anderson, (2010), the concept of similarity is fundamental to cluster analysis. The variables can be measured in a variety of ways, but the best fit for the project was the correlational measure shown in Table 6.

Then there was the selection of the portioning procedure used for forming clusters. From the hierarchical cluster procedures, it was selected the agglomerative methods and from the clustering algorithms, the centroid method.

The agglomerative methods start with all observations as their own cluster, so that the number of clusters equals the number of observations. Similarity measure combine the two most similar clusters into a new cluster, reducing the number of clusters by one. The process repeats using the similarity measure at each step, combining the most similar clusters into a new cluster a total of n-1 times until all observations are contained in a single cluster. In Table 7 the partial agglomeration schedule is shown.

Partial Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
7300	1	43	29.177	7291	7292	7301
7301	1	37	29.585	7300	7297	7304
7302	4014	5555	30.077	7284	0	7311
7303	859	2499	40.515	0	0	7314
7304	1	77	40.635	7301	7295	7307
7305	30	270	52.378	7264	7299	7310
7306	194	2036	60.115	0	7298	7309
7307	1	3114	75.552	7304	7296	7308
7308	1	27	109.944	7307	7294	7309
7309	1	194	227.736	7308	7306	7311
7310	30	2230	350.568	7305	0	7312
7311	1	4014	421.124	7309	7302	7312
7312	1	30	553.608	7311	7310	7313
7313	1	7088	980.296	7312	0	7314
7314	1	859	1050.877	7313	7303	7315
7315	1	7126	3750.106	7314	0	0

Table 7: Partial Agglomeration Schedule

Similarity between two clusters is the distance between the cluster centroids. Cluster centroids are the mean values of the observations on the variables in the cluster variate.

After making the portioning procedure the decision on the number of clusters was made using the stopping rule and making a comparison of cluster solutions (see figure 8).

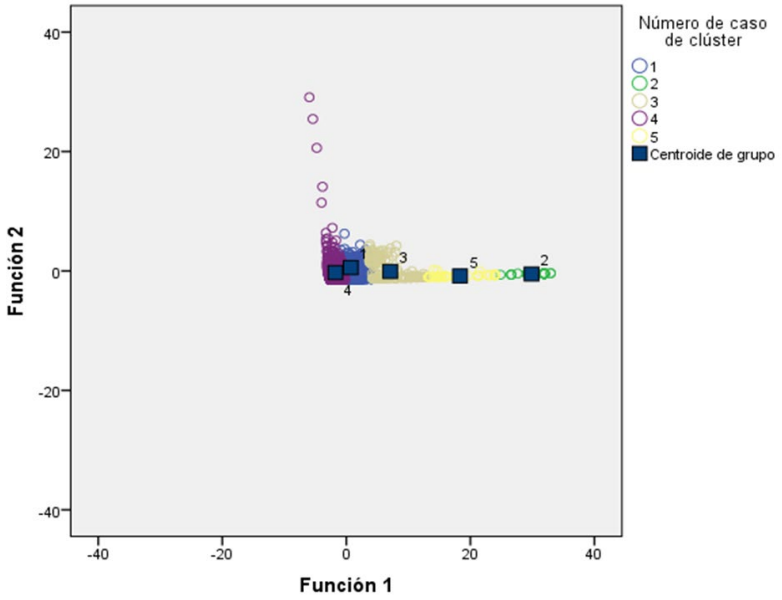


Figure 8: Cluster Centroids

Finally, a discriminant analysis was made in order to prove that the assigned observations really belong to each cluster.

Clasification Results

Cluster			Predicted Group Membership					Total
			1	2	3	4	5	
Original	Count	1	2078	0	6	322	0	2406
		2	0	93	0	0	0	93
		3	30	0	179	0	1	210
		4	10	0	0	4488	0	4498
		5	0	0	0	0	104	104
	%	1	86.4	0.0	.2	13.4	0.0	100.0
		2	0.0	100.0	0.0	0.0	0.0	100.0
		3	14.3	0.0	85.2	0.0	.5	100.0
		4	.2	0.0	0.0	99.8	0.0	100.0
		5	0.0	0.0	0.0	0.0	100.0	100.0

95.0% of original grouped cases correctly classified.

Table 9: Discriminant Analysis

- Evaluation

Before proceeding to a final deployment of the model, it is important to evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives.

The model assigned the observations to clusters in the following way (see figure 10).

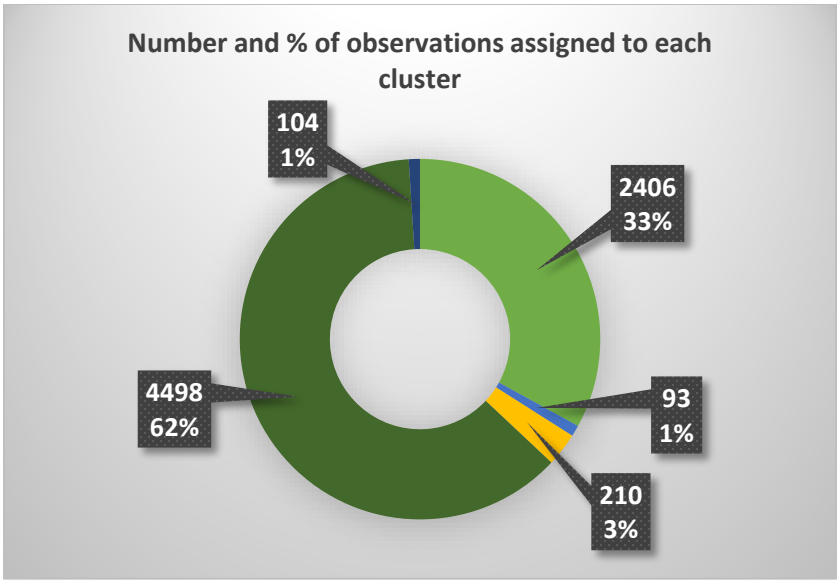


Figure 10: Assignation of observations into clusters

However, its relevant knowing the detail of each cluster. The variables are the following:

- x_1 - average number of products;
- x_2 - average price;
- x_3 - number of tickets;
- x_4 - number of $ on discounts;
- x_5 - average expiration of products.

Cluster 1 has 2,406 observations and is most distinguished by a **high mean** for number of products (x_1), and price (x_2). As well, its important to observe the gap range in the number of tickets, meaning the minimum and maximum number of visits in the period.

Statistics Cluster 1

	x_1	x_2	x_3	x_4	x_5
N	2406	2406	2406	2406	2406
Mean	37.02	1405.94	4.44	1.21	5.05
Std. Deviation	20.34	392.07	5.725	15.02	1.88
Minimum	1.00	923.73	1	0.00	1.00
Maximum	216.30	2567.70	62	350.00	10.00

Table 11: Statistics Cluster 1

Cluster 2 has 93 observations and is most distinguished by a **low mean of products** bought (x_1), however, a **high price** (x_2). The number of visits in the period as well, are low.

Statistics Cluster 2

	x_1	x_2	x_3	x_4	x_5
N	93	93	93	93	93
Mean	1.30	12236.03	1.03	0.00	10.00
Std. Deviation	1.04	561.69	0.23	0.00	0.00
Minimum	1.00	10348.00	1.00	0.00	10.00
Maximum	7.33	13398.00	3.00	0.00	10.00

Table 12: Statistics Cluster 2

Cluster 3 has 210 observations and is most distinguished by a **relatively high mean of products** and a **high mean in price** (x_2). The maximum number of visits in the period is relatively high.

Statistics Cluster 3

	x_1	x_2	x_3	x_4	x_5
N	210	210	210	210	210
Mean	27.67	3729.62	1.65	0.53	8.22
Std. Deviation	41.24	848.18	1.78	4.11	2.56
Minimum	1.00	2576.90	1.00	0.00	2.00
Maximum	156.00	5799.00	16.00	48.70	10.00

Table 13: Statistics Cluster 3

Cluster 4 has 4,498 observations and is most distinguished by a **relatively high mean of number of products** (x_1). This cluster has the highest number of visits in the period being analyzed, so its interesting for marketing matter, to take advantage of this high frequency of visits and start doing efforts to increase the number of products bought in each visit for this type of customers.

Statistics Cluster 4

	x_1	x_2	x_3	x_4	x_5
N	4498	4498	4498	4498	4498
Mean	16.84	441.37	3.66	1.10	5.90
Std. Deviation	28.94	260.08	9.21	14.26	2.58
Minimum	0.60	1.80	1.00	0.00	0.00
Maximum	972.00	923.55	284.00	551.05	10.00

Table 14: Statistics Cluster 4

Cluster 5 has 104 observations and is most distinguished by a **high mean of price** (x_2). However, the range in the number of visits is low.

Statistics Cluster 5

	x_1	x_2	x_3	x_4	x_5
N	104	104	104	104	104
Mean	1.73	7898.88	1.21	0.00	9.93
Std. Deviation	3.19	1363.47	0.50	0.00	0.41
Minimum	1.00	5999.00	1.00	0.00	7.10
Maximum	24.74	9999.00	4.00	0.00	10.00

Table 15: Statistics Cluster 5

After the statistics analysis, a deeper analysis was made in order to observe what type of products each cluster was buying and that way, a validation and profiling of clusters could be made.

The analysis was completed, and the profiling resulted the following way:

Cluster 1: Shopping Moms

73% of the customers assigned to the cluster were women with an average age of 36.8 years old. In the following table, the detail of the products being bought by this cluster is shown. The top 9 products are basic groceries.

Department	Number of Products
Fruits and vegetables	92,464
Inedible groceries	25,804
Dairy products	22,980
Delicatessen	22,020
Fast food	21,749
Breakfast and desserts	21,322
Basics	16,866
Meats	15,247
Drinks and snacks	13,360
Frozen food	4,894
Babies	4,868
Live healthy	4,247
Health	4,211
Personal care and hygiene	3,529
Bakery	2,880
Fish and seafood	2,246
Tortilleria	2,053
OTC	1,264
Beauty products	1,181
Sweets and stationery	748
Wine, beer and liquors	662
Kitchen	523
Home	502
Pharmacy	298
Prepared foods	283
Coffee	202
Do it yourself	170
Electronics	139
Toys and entertainment	82
Flower shop	74
Specialized meals	16
Total	**286,884**

Table 16: "Shopping Moms" Buying Patterns

Cluster 2: Technology Guys

66% of the customers assigned to the cluster were men with an average age of 36.4 years old. This cluster is unique by the type of products they're buying, exclusively electronics and home products (see Table 17).

Department/Family	Number of Products
Electronics	**100**
Televisions	91
Videogames	5
Computers	4
Home	**2**
Mattresses	2
Total	**102**

Table 17: "Technology Guys" Buying Patterns

Cluster 3: Equals/Same Role

44% and 45% of the customers assigned to the cluster were women and men respectively, with an average age of 37.2 years old. This cluster doesn't hold characteristics that make them unique. The type of products being bought are basic groceries.

Department	Number of Products
Fruits and vegetables	1,500
Inedible groceries	564
Fast food	392
Breakfast and desserts	369
Dairy products	369
Meats	362
Delicatessen	360
Basics	347
Drinks and snacks	263
Babies	119
Live healthy	105
Home	97
Health	96
Frozen food	86
Electronics	85
Personal care and hygiene	71
Fish and seafood	65
Bakery	61
Wine, beer and liquors	38
Do it yourself	26
Beauty products	26
OTC	25
Tortilleria	25
Kitchen	19
Sweets and stationery	16
Pharmacy	12
Coffee	8
Prepared foods	2
Flower shop	1
Toys and entertainment	1
Total	**5,510**

Table 18: "Equals/Same Role" Buying Patterns

Cluster 4: Future Shoppers

65% of the customers assigned to the cluster were women with an average age of 35.3 years old. This cluster have similar buying patterns as the "Shopping Moms", the main difference are in the number of products bought (see Table 19).

Department	Number of Products
Fruits and vegetables	66,501
Inedible groceries	22,160
Breakfast and desserts	17,980
Dairy products	17,972
Delicatessen	16,320
Fast food	15,753
Drinks and snacks	13,722
Basics	13,376
Meats	9,141
Bakery	4,082
Frozen food	4,001
Health	3,596
Live healthy	3,178
Babies	2,885
Personal care and hygiene	2,870
Tortilleria	1,647
Sweets and stationery	1,461
OTC	1,455
Wine, beer and liquors	1,407
Fish and seafood	1,298
Beauty products	1,105
Kitchen	916
Home	588
Do it yourself	435
Prepared foods	404
Pharmacy	371
Electronics	280
Toys and entertainment	236
Flower shop	152
Coffee	137
Specialized meals	22
Total	**225,451**

Table 19: “Future Shoppers” Buying Patterns

Cluster 5: Dads

58% of the customers assigned to the cluster were men with an average age of 36 years old. This cluster, as well, is unique, the buying patterns are a mix of electronics with groceries. This is an interesting case for the marketing department (see Table 20).

Department	Number of Products
Electronics	118
Dairy products	14
Fruits and vegetables	13
Basics	12
Inedible groceries	9
Drinks and snacks	8
Meats	8
Fast food	7
Do it yourself	6
Live healthy	5
Delicatessen	5
Health	5
Breakfast and desserts	3
Frozen food	2
Home	2
Babies	2
Personal care and hygiene	2
Kitchen	1
Toys and entertainment	1
Tortilleria	1
Bakery	1
Total	**225**

Table 20: "Dads" Buying Patterns

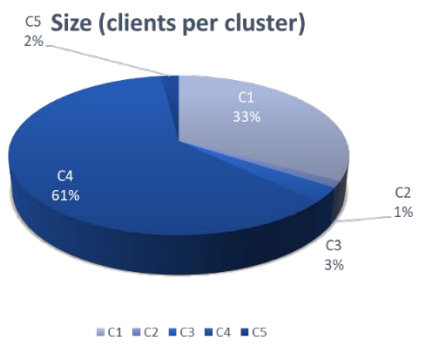
- Deployment

Depending on the requirements the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process.

Customer's preferences and demand change. For the model to work correctly, constantly there must be adjustments, and getting new input into the model so the output is accurate.

As well, the team made important recommendations for further modeling: (1) obtain more variables, specially from unstructured data. (2) do a further segmentation for e-customers that use the modality of "pick and go". (3) use discriminant analysis with 2018 data to validate clusters and compare with 2017. (4) develop causal models to forecast sales for each cluster (5) identify the perfect marketing for each cluster.

5. Results



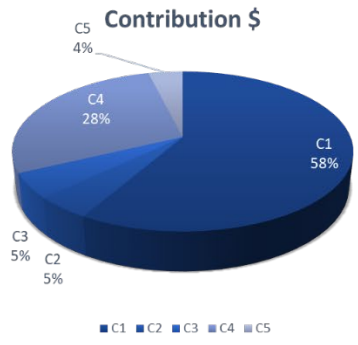


Figure 16: Cluster Results by Size and Economic Contribution

After analyzing the results, this were the findings:

The biggest cluster in size (Cluster 4) doesn´t represent the biggest economic contribution to the company. The team made a proposal to the company in order to make this group of customers more profitable through targeted marketing, encouraging sales.

Another important outcome is that Cluster 1, generates more than half of the income of the e-retailer. So, it's important to keep close attention to this group of customers. They're valuable customers.

Finally, Cluster 2, represents 5% of sales, even though in size it only represents 1%. This cluster has the potential to be a very representative group of customers. The team made a proposal to create new strategies to grow this cluster that has the potential to contribute in a bigger scale to the company.

6. Conclusions

The execution conclusions were that the IT team got the first efforts and insights related to analytics applied on the ecommerce platform. Its important to mention that six months later the company created a data analytics department so that the company can make in collaboration with undergraduate students from the University of Monterrey, data analytics projects.

This project was followed up by two more teams to transition from prescriptive analytics to predictive analytics.

The customer analytics model was presented to the CEO at the end of the project and the results were beyond expectations.

As well, general conclusions, big data analytics (BDA) enables e-commerce firms to use data more efficiently, improve decision making and empower customers (Akter, Wamba, 2016).

BDA support business needs such as:

- Identifying loyal and profitable customers,
- Determining the optimal price,
- Detecting quality problems,
- Or deciding the lowest possible level of inventory.

Data analytics and big data aren't just part of a new trend, it's part of a new approach to customers and their needs, and how the company need this information in order to stay competitive and be able to respond to market changes and demand.

References

Cuzzocrea, A., Song, I.Y., and Davis, K.C. *Analytics over Large-Scale Multidimensional Data: The Big Data Revolution!* Glasgow, Scotland, UK, 2011.
Fan, S., Lau, R.Y.K., and Zhao, J.L. *Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix*, Elsevier Inc., 2015.
Chen, H., Chiang, R.H.L., and Storey, V.V. *Business Intelligence and Analytics: From Big Data to Big Impact*, Mis Quarterly, vol. 36, no. 4, pp. 1165-1188, 2012.
Sun, Z., Zou, H., and Strang, K. *Big Data as a Service for Business Inteligence*, ResearchGate, 2015.
Akter, S., and Wamba, S.F. Big data analytics in E-commerce: a systematic review and agenda for future research, Springer, 2016.
Wirth, R., and Hipp, J. *CRISP-DM: Towards a Standard Process Model for Data Mining*, Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf, (n.d.).
Rollins, J.B. *Foundational Methodology for Data Science*, IBM Corporation, United States of America, 2015.

Gandomi, A., and Haider, M. *Beyond the hype: Big Data concepts, methods, and analytics*, Elsevier, vol. 35, pp. 137-144, 2015.
Hair, J.F., Black, W.C., Babin, B.J., and Anderson, R.E. *Cluster Analysis*, Pearson Prentice Hall, Chapter 9 of Multivariate Data Analysis, 2010.

Biographies

Daniela Garza Gutiérrez earned B.Sc. in Engineering Management at the University of Monterrey, Mexico. She is certified in customer and operations analytics by Wharton University, USA. She has experience in the elaboration and communication of key performance indicator reports, in the planning, execution and closing of projects focused on training, and improvement projects while as an in Oxxo's corporate in the HR department. She also worked more than a year and a half in the credits department at Value Arrendadora Financial Group.

Juan Ignacio González Espinosa earned Bsc. in Marketing graduate by the ITESM Campus Monterrey with mention of excellence. He earned a MBA from EGADE Business School and PhD in administration, with a major in international business and strategy. Trainee exchange as a doctoral student at The Ohio State University, at the Fisher College of Business, and specialization in Peking University. He´s full time professor at Engineering Department in the Engineering management academic program, at University of Monterrey.

Luz María Valdez de la Rosa is Academic Chair of B.S. in Engineering Management for the University of Monterrey, in the state of Nuevo Leon, Mexico. She earned B.S. in Industrial Engineering and Systems and Masters in Quality Management from University of Monterrey, Mexico, and she is currently studying the Ph. D. in Administration Sciences from the Autonomous University of the State of Nuevo Leon, Mexico. She has participated as consultant for the manufacturing and services in the quality field, and participated as ASQ member, IISE member.