

# MARKETING AND CUSTOMER ANALYTICS

Advanced Programming in Python (DSE 309)



**SUBMITTED TO:**

Prof. Parthiban Srinivasan

Adjunct Faculty

Data Science and Engineering

**SUBMITTED BY:**

Shraddha Agarwal

19294

Data Science and Engineering

# **1. ACKNOWLEDGEMENT**

I would like to express my gratitude towards Dr. Parthiban Srinivasan, Professor of Advanced Programming in Python for providing us such a wonderful opportunity to work on this project and guiding us throughout this journey.

# **2. ABSTRACT**

Marketing analytics is the analysis of data collected via marketing campaigns in order to find trends in things like how a campaign influenced conversions, customer behavior, geographical preferences, creative preferences, and so on. Customer analytics plays a vital role in generating insights from big data to improve service innovation, product development, personalization, and managerial decision making. It is the process of systematically examining a company's customer data and behavior in order to find, attract, and keep the most lucrative customers. In this project, the marketing data of a company is analyzed thoroughly to draw the reasons about the ineffectiveness of its marketing campaign and providing the company with the data-driven solutions.

# **3. INTRODUCTION**

In the last few years, "Big Data" has posed a constant challenge to businesses, presenting an exciting new frontier of opportunity. Modern businesses are putting up efforts to use analytics in order to acquire a competitive edge in the quickly changing data-rich business environment. By 2020, the global market for business intelligence and analytics is expected to be worth \$200 billion, with many industry analysts forecasting that a company's customer analytics capabilities would improve the overall performance of the company in the big data environment. Large and small businesses are inundated with information on customer preferences. However, having more knowledge does not automatically imply making better judgments. The problem is learning how to understand data, and marketers in particular are increasingly expected to use analytics to inform and defend their actions.

Marketers may use marketing analytics to monitor, control, and evaluate their performance in order to increase their efficacy and maximize their return on investment (ROI). Marketing analytics, in addition to the obvious sales and lead generating uses, may provide deep insights into consumer preferences and patterns, which can be used to inform future marketing and business choices. Customer analytics refers to a cutting-edge technology that analyses huge amounts of marketing data to address customer-centric concerns. A vast body of research focuses on the benefits of customer analytics, with little or no consideration paid to the firm's ability to create customer analytics capacity in a data-rich environment.

## **3.1 MARKETING ANALYTICS**

Marketing analytics comprises the processes and technologies that enable marketers to evaluate the success of their marketing initiatives. This is accomplished by measuring performance. Marketing analytics uses important business metrics, such as ROI, marketing attribution and overall marketing effectiveness. Through all of these analyses, the strengths, weaknesses, opportunities and threats of a company can be identified. Marketing analytics as a technique aims to leverage these trends and discoveries to improve future campaigns based on what worked in the past. Marketers and customers both benefit from marketing analytics. This analysis enables marketers to get a better return on their marketing spending by determining

what works best in terms of conversions, brand recognition, or both. In addition, analytics guarantees that customers receive a higher number of targeted, customized campaigns that appeal to their unique needs and interests, rather than the annoying mass communications.

### **3.2 CUSTOMER ANALYTICS**

Marketing analytics has a sub-dimension called customer analytics. Customer analytics is a process in which market segmentation and predictive analytics are used to assist make crucial business choices based on data from customer behavior. Businesses use this information for things like direct marketing, location selection, and client relationship management. Marketing gives services to clients in order to please them. With this in mind, the productive system is examined from the point of origin at the production level to the point of consumption at the conclusion of the cycle. Customer analytics plays a critical role in predicting customer behavior. In marketing, the customer is king, and a company's long-term viability is determined by the lifetime value of its customers.

Customers have access to information about where to shop, what to purchase, how much to pay and so on, at any time and from any location. As a result, using predictive analytics and data to forecast how customers will react while dealing with companies is becoming increasingly vital. It aims to establish a single, accurate perspective of a customer so that businesses may make informed decisions about how to effectively attract and keep consumers, as well as discover and communicate with high-value customers. The more a client's purchase patterns and lifestyle preferences are understood, the more accurate predicted behaviors emerge, and the better the customer journey gets.

## **4. AIM OF THE PROJECT**

The primary goal of this project is analyzing the dataset of a company to understand the ineffectiveness of the marketing campaigns and proposing the data-driven solutions.

The recent marketing campaigns the company were not much effective. To build the understanding of the reasons behind it, the marketing and customer data of the company have been analyzed using exploratory and statistical data analysis, data visualization and customer segmentation. Various classification models have also been implemented to predict the success of the latest campaign. At the end, recommendations were proposed to improve the success rate of the campaigns based on the conclusions made with the data analysis.

## **5. INFORMATION OF THE DATASET**

The dataset which is used in this project includes the following:-

- The unique identification of 2240 customers
- Birth Year of the customers
- Education level of the customers
- Marital Status of the customers
- Yearly household income of the customers
- Number of children in the household of the customer
- Number of teenagers in the household of the customer
- Date of customer's enrolment with the company

- Number of days since customer's last purchase
- Amount spent by the customers on wine in the last 2 years
- Amount spent by the customers on fruits in the last 2 years
- Amount spent by the customers on meat in the last 2 years
- Amount spent by the customers on fish in the last 2 years
- Amount spent by the customers on sweets in the last 2 years
- Amount spent by the customers on gold in the last 2 years
- Number of purchases made by the customers with a discount
- Number of purchases made by the customers through the company's web site
- Number of purchases made by the customers using a catalogue
- Number of purchases made by the customers directly in stores
- Number of visits to company's web site in the last month
- Performance of the 1st campaign
- Performance of the 2nd campaign
- Performance of the 3rd campaign
- Performance of the 4th campaign
- Performance of the 5th campaign
- Performance of the latest campaign
- Customer complains in the last 2 years
- Location of the customer

The dataset is available at the link: <https://www.kaggle.com/jackdaoud/marketing-data>

## 6. METHODOLOGY

The project has been divided into 5 sections:-

### 6.1 Exploratory Data Analysis:

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, or spot anomalies, or test hypothesis and to check assumptions with the help of summary statistics and graphical representations. EDA is different from initial data analysis (IDA) which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed.

The feature containing null values are identified and filled with the median (mean is not filled to avoid skewness of the data due to the presence of outlier), features containing outliers are identified using the outliers plot and dropped some of the outliers which are impossible to occur in real, feature that need transformation is identified and transformed from 'object' dtype to 'datetime' dtype. Features are being engineered by introducing some new features using the initial features followed by plotting the correlation maps for identifying feature correlation to determine the patterns and anomalies present in the data. All the observations drawn has been described at the bottom of the plots.

```
In [33]: # View the customer ID with new features.
df[['ID', 'Age', 'Children', 'TotalEnrollDays', 'TotalMnt', 'TotalPurchase', 'TotalCampaignsAcc']]
```

Out[33]:

	ID	Age	Children	TotalEnrollDays	TotalMnt	TotalPurchase	TotalCampaignsAcc
0	1826	51	0	2728	1190	15	1
1	1	60	0	2729	577	18	2
2	10476	63	1	2762	251	11	0
3	1386	54	2	2764	11	4	0
4	5371	32	1	2797	91	8	2
...	...	...	...	...	...	...	...
2235	10142	45	1	3194	689	20	0
2236	5263	44	1	3238	55	5	0
2237	22	45	1	3288	309	14	0
2238	528	43	0	3292	1383	20	0
2239	4070	52	2	3381	1078	18	3

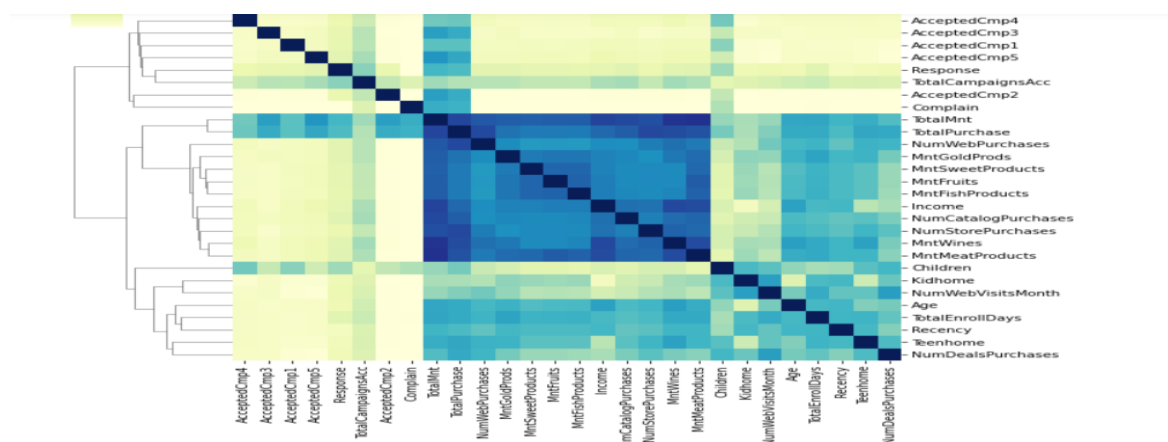
2229 rows x 7 columns

## Feature Engineering Dataframe

### 6.2 Statistical Analysis:

Statistical analysis is the collection and interpretation of data in order to uncover patterns and trends. It is a component of data analytics. Statistical analysis can be used in situations like gathering research interpretations, statistical modeling or designing surveys and studies. It can also be useful for business intelligence organizations that have to work with large data volumes. In the context of business intelligence (BI), statistical analysis involves collecting and scrutinizing every data sample in a set of items from which samples can be drawn. In statistics, a sample is a random sample obtained from a larger population. The goal of statistical analysis is to identify trends. Statistical analysis is used by a retail company to uncover trends in unstructured and semi-structured consumer data that can be used to improve the customer experience and boost revenue.

The factors affecting the features like ‘number of store purchases’ can be determined by visualizing its positive and negative correlations. In terms of total purchases, the performance of India than rest of the world is evaluated by employing the null hypothesis test and t-test. The correlation between amount spend on gold in the last 2 years with the number of store purchases has been verified using Kendall Correlation Analysis and the correlation between married PhD candidates with the amount spent on fish in the last 2 years has been verified using visualization along with determining the factors affecting ‘amount spend on fish’. The relationship between the geographical region and the success of 4<sup>th</sup> campaign is verified using chi-square test. The results are described at the end of the test.

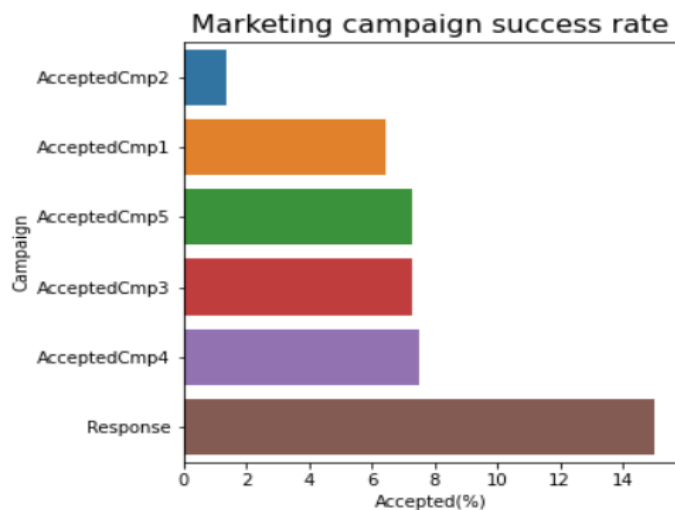


Clustermap depicting correlation between different features

## 6.3 Data Visualization:

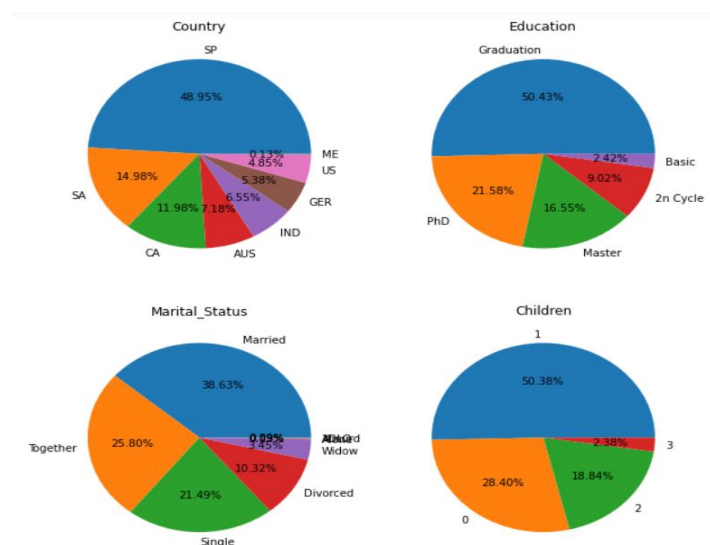
Data visualization is the process of translating large data sets and metrics into charts, graphs and other visuals. The resulting visual representation of data makes it easier to identify and share real-time trends, outliers, and new insights about the information represented in the data. As the amount of big data increases, more people are using data visualization tools to access insights on their computer and on mobile devices. Dashboards are used by business people, data analysts, and data scientists to make data-driven business decisions. Data visualization positively affects an organization's decision-making process with interactive visual representations of data. Businesses can now recognize patterns more quickly because they can interpret data in graphical or pictorial forms.

In the project, data is visualized using bar plots and pie plots to determine the answers to the basic questions of the company like the most successful marketing campaign, average customer look like to the company, best performing products and underperforming channels. The answers are described at the end of the plots.



Bar plot depicting success rate of marketing campaigns

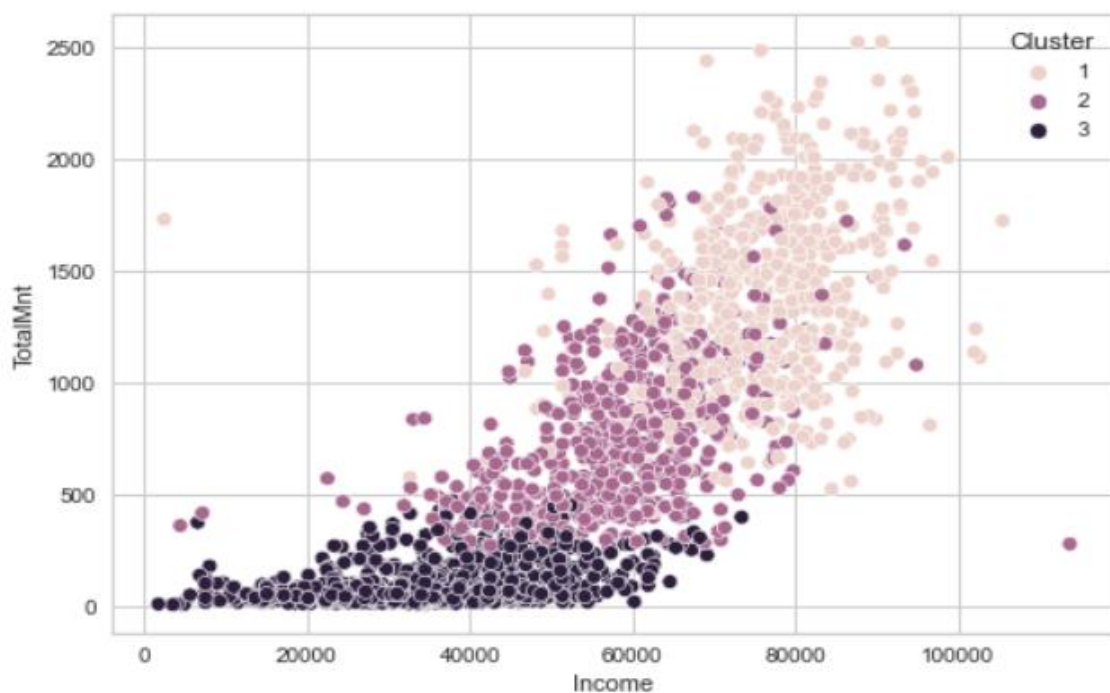
Pie plot depicting how average customer looks to the company



## 6.4 Segmentation of the customers (Clustering):

Customer segmentation is the practice of dividing a company's customers into groups that reflect similarity among customers in each group. The goal of segmenting customers is to decide how to relate to customers in each segment in order to maximize the value of each customer to the business. Customer segmentation has the potential to allow marketers to address each customer in the most effective way. Using the large amount of data available on customers (and potential customers), a customer segmentation analysis allows marketers to identify discrete groups of customers with a high degree of accuracy based on demographic, behavioural and other indicators.

The customers are segmented by employing K-Means clustering model and optimum number of clusters chosen are 3. The KElbowVisualizer is imported from yellowbrick library which implements the 'elbow' method to select the optimal number of clusters by fitting the model with a range of values for KK. If the line chart resembles an arm, then the "elbow" (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. In the visualizer "elbow" will be annotated with a dashed line. Model has been built and data gets clustered into 3 clusters with similarities and differences among them based on the features. These clusters can be visualized using Facet.Grid plots. The observations drawn after visualizing the graphs are written below the graph.



Clustering of Customers

## 6.5 Classification Modelling:

Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. It's not always binary, but always discrete. In mathematical term, classification predictive modelling has the task of approximating a mapping function ( $f$ ) from input variables ( $X$ ) to discrete output variables ( $y$ ).



In this project, the classification models are employed to predict whether the customer accepts or rejects the offer in the recent campaign. Eight classification models have been implemented that are: Logistic Regression, Decision Tree Classifier, Gradient Boosting Classifier, Extra Trees Classifier, MLP Classifier, SVC, random Forest Classifier and k-Neighbors Classifier. The accuracy score and performance measures of these models are compared and visualized using bar plots. The ROC Curves for these models are also drawn followed by classification reports and confusion matrices.

## **7. RESULTS AND ANALYSIS**

Each section has produced their own results which are described in details at the end of each section in the Jupyter notebook. Following are the section-wise results of the project:-

### **7.1 Exploratory Data Analysis:**

- 'Income' contains 24 null values.
- Multiple features contain outliers but the outliers of 'Birth\_Year' is unacceptable as it's not possible for a customer to have birth year less than 1900 and also some of entries in 'Income' column is higher than '200,000'.
- 'Dt\_Customer' needs transformation since it contains date.

### **7.2 Statistical Analysis:**

- 'Amount spent on wine', 'Income', 'Total purchase', 'spent on meat', 'number of catalog purchases' and 'total amount spent' are positively correlated while 'Number of children' and 'Number of websites visit per month' are negatively correlated with 'number of store purchase'.
- Total purchase made by India is equal to the average of total purchase made by rest of the world in terms of total purchases.
- There is no correlation between people who spent an above average amount on gold in the last 2 years and in store purchases.
- Married PhD customers does not have a significant relationship with the amount spent on fish.
- There is no significant relationship between geographical regional and success of the 4th campaign.

### **7.3 Data Visualization:**

- The most successful campaign is the most recent campaign ('Response')
- The average customer belongs to Spain, is graduated and married and has age 52, income of about 50,000 and at least a child.
- Wine is the best performing product in terms of revenue and revenue per customer followed by meat.
- Deals purchases is underperforming channel followed by catalog purchases.



## 7.4 Segmentation of Customers:

There are a number of similarities across certain attributes but the key ones which help distinguish the clusters of customers from one another have been summarised as follows:-

Cluster 1 includes those customers who have high income.

- Majority of them have no kids.
- This group has higher expenses and it is evident from the fact that they earn the most.
- As far as the products are concerned, there is nothing that stands out which tells whether the group as a whole prefer a certain product more or less.
- The cluster prefers to make catalog purchases and has the least deal purchases.
- Based on the previous 5 marketing campaigns, the highest success rate has come from this group.

Cluster 2 includes those customers who have moderate income.

- Majority have at least 1 child. (Child = Kid(s) + Teen(s))
- The group has a significant affinity to wine products.
- This cluster has the highest number of people.

Cluster 3 includes those customers who have lower income.

- Customers with Basic education are found in this group only.
- Due to lower income, the expenses in absolute terms is also the least here.
- The group prefers to spend more on fruits, meat and fish and less on wine.
- The preferred channel to purchases products here is through stores.
- Majority of campaigns have failed to attract this group. The success rate is the lowest.

## 7.5 Classification Modelling:

- Gradient Boosting Classifier has highest accuracy among all the 8 different classification models.

## 8. CONCLUSION

The project is concluded after proposing the data-driven solutions through the data analysis of the marketing dataset. Following recommendations can be proposed to the company to make their marketing campaigns more effective:-

- Conduct future advertising campaigns using the same model recently implemented.
- Focus advertising campaigns on boosting sales of the less popular items.
- Focus advertising campaigns on the more successful channels to reach more customers.
- Marketing fish products to married people who have an education level at or above graduation is recommended.
- Making wine more accessible will increase sales in store.
- Focus advertising campaigns to target lower and moderate-income customers by providing them with special discounts as these clusters have the greatest number of people.

## 9. REFERENCES

- [1] Dawn Iacobucci, Maria Petrescu, Anjala Krishen, Michael Bendixen. The state of marketing analytics in research and practice. *Journal of Marketing Analytics*
- [2] Md Afnan Hossain, Shahriar Akter, Venkata K. Yanamandram. Customer Analytics Capabilities in the Big Data Spectrum: A Systematic Approach to Achieve Sustainable Firm Performance. *Sydney Business Schools- Papers*
- [3] Omer Minkara, CUSTOMER ANALYTICS: HOW TO MAKE BEST USE OF CUSTOMER DATA. Aberdeen Group
- [4] Aggarwal, P., R. Vaidyanathan, and A. Venkatesh. 2009. Using lexical semantic analysis to derive online brand positions: An application to retail marketing research. *Journal of Retailing*
- [5] Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*
- [6] <https://www.scikit-yb.org/en/latest/api/cluster/elbow.html>
- [7] <https://seaborn.pydata.org/generated/seaborn.FacetGrid.html>
- [8] <https://www.optimove.com/resources/learning-center/customer-segmentation>
- [9] <https://searchbusinessanalytics.techtarget.com/definition/customer-analytics>
- [10] <https://www.syntelli.com/solutions/marketing-consumer-analytics#:~:text=Marketing%20analytics%20provides%20a%20holistic,data%2C%20and%20track%20customer%20acquisition.>