

A  
PROJECT REPORT  
ON  
**A Study of customer personality Analysis using machine  
learning algorithms**

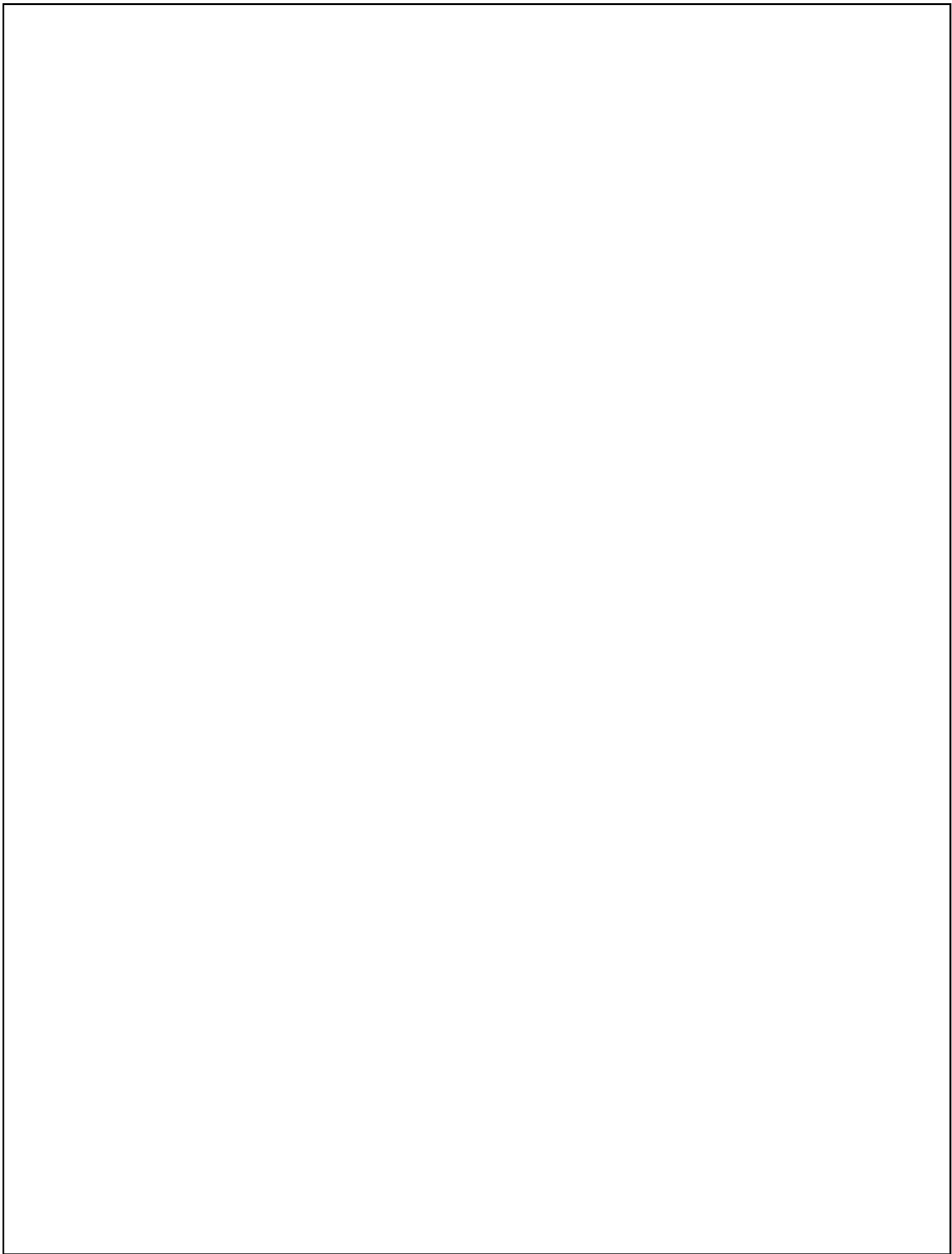
SUBMITTED BY  
**SAURABH SHANKAR PANSKAR**  
UNDER GUIDANCE OF  
**PROF. ROHINI NIKAM**  
SUBMITTED TO



**SAVITRIBAI PHULE PUNE UNIVERSITY**  
IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF  
THE DEGREE OF MASTER OF BUSINESS ADMINISTRATION (MBA)  
THROUGH

SAI BALAJI INTERNATIONAL INSTITUTE OF MANAGEMENT SCIENCE,  
PUNE 411033

ACADEMIC YEAR 2022-24



Date: 14/11/2023

3  
**THE PATAN URBAN  
CO-OP .BANK LTD  
PATAN**

Phone No.:(02372) 282921

Office Address: 886 A, MAINROAD PATAN.

fax: 283350

**TO WHOMSOEVER IT MAY CONCERN**

This is to certify that Mr. SAURABH PANASKAR a student of MBA (Major in Business Analytics - summer semester), Sai Balaji international institute of management, Pune has successfully completed 2 (Two) months (From 4th August, 2023 to 6th October, 2023) long internship programme at this branch.

During his internship he displays enthusiasm, punctuality, hardworking nature, and an inquisitive mindset. He actively engaged in the assigned tasks, contributing positively to the workplace environment.

We appreciate his dedication and efforts during his internship, and we believe that he has gained valuable insights and practical experience that will benefit him in his academic and professional journey.

We wish him every success in his future life.

Place - Patan

FOR THE PATAN URBAN CO-OP. BANK LTD., PATAN



*Prav...Panaskar*  
**BR. MANAGER**  
*BR. Mh...harpeth*  
Pravin Panaskar  
(Branch Manager)

## **ACKNOWLEDGEMENT**

I would want to take this opportunity to thank all of the people that helped me with my MBA summer internship project. Throughout the endeavor, I am grateful for their aspirational leadership, priceless constructive criticism, and kind counsel. I truly appreciate them offering their insightful and honest opinions on several project-related topics.

For his assistance and direction, I would like to extend my sincere gratitude to Prof. (Dr.) L. K. Tripathi, Director of the Saibalaji International Institute of Management Sciences in Pune.

I also want to express my gratitude to Prof. Rohini Nikam, my project internal guide, and my project external guide, Manager, as well as everyone else who helped me with my MBA project by providing the necessary resources and conducive environment.

Thank You,

Saurabh Shankar Panaskar

## INDEX

<b>Serial No</b>	<b>Topic</b>	<b>Page No</b>
<b>1</b>	Introduction	5-6
<b>2</b>	Objectives	7
<b>3</b>	Company Profile	8-10
<b>4</b>	Review of Literature	11
<b>5</b>	Research Methodology	12-13
<b>6</b>	Statistical Tools and Techniques	14
<b>7</b>	Data Preprocessing	15-20
<b>8</b>	EDA	21-27
<b>9</b>	Findings	29
<b>10</b>	Conclusion	30
<b>11</b>	References	31

## INTRODUCTION

Nowadays there is cut-throat competition in the market between the various companies to maintain their traditional customer base and serve them best so that they can grow their business by using various marketing strategies. For this purpose, the approach of the company looking towards the customers is most important. But attracting customers and customer satisfaction is a very big problem for the company. With a variety of marketing techniques, the corporation is attempting to promote its product to consumers. The business promoted its items and gave discounts to consumers. The business organizes a number of marketing initiatives to draw clients. Instead of accomplishing all of this, a lot of the businesses failed to keep their customer base and were unable to draw in new ones since they were ignorant of what kind of consumers would buy. Customer personality analysis enters the picture for this reason. According to Dion et al. (1995) and Kassarjian (1971), a personality is only a psychological construct that may be described as a "set of responses that individuals perform in their environments."

Analysis of the personalities of customers is now a crucial component in developing marketing strategies. An in-depth examination of a business's ideal clients is known as customer personality analysis. This analysis delves deeply into consumer purchasing patterns to assist businesses better understand their clientele, which enables them to tailor their offerings to meet the wants and desires of their target market and increase customer happiness. A number of factors influence a customer's purchase behavior, including their habits, health concerns, and whether or not they are a parent.

Unsupervised machine learning models are crucial in this situation. The unsupervised machine learning techniques encompass a variety of models, including DBSCAN clustering and hierarchical k-means. A data exploration approach called clustering enables objects with comparable properties to be grouped together for easier processing down the road. Large datasets are necessary for these models to automatically identify patterns and similarities between the various client segments. The company's working staff has always performed manual client personality analysis; however, with the advent of machine learning, this procedure can now be completed more quickly and automatically. In the fiercely competitive landscape of today's business environment, companies are relentlessly pursuing strategies to not only retain their

existing customer base but also to attract new customers and fuel business growth. Central to this pursuit is the ability to understand and cater to the unique needs, preferences, and behaviors of customers. Despite employing various marketing techniques, many businesses find it challenging to retain customers and attract new ones due to a lack of insight into the diverse consumer base they serve. Recognizing this challenge, a paradigm shift has emerged in the form of customer personality analysis, a process aimed at comprehensively understanding the psychological constructs that influence consumer behavior.

Customer personality analysis involves a deep examination of a business's ideal clients, exploring their purchasing patterns, habits, and concerns. This analysis enables businesses to tailor their products and services to meet the specific wants and desires of their target market, ultimately enhancing customer satisfaction. In the contemporary landscape, the integration of unsupervised machine learning models has revolutionized customer personality analysis, offering a more efficient and automated approach

## OBJECTIVE

To help a business by extracting insights based on its target customers from various customer segments. This helps them better understand their customers and makes it easier for them to adapt products according to the unique wants, behaviors, and concerns of different sorts of customers.

- To check the relationship between customers' Income and their spending.
- To check the response of the customers to the promotions done by the company.
- To check the response of the customers to the campaigns

# Banking Sector Overview

## Introduction:

The banking sector plays a pivotal role in the global economy, serving as the backbone of financial systems across the world. This dynamic industry encompasses a wide array of financial institutions, each contributing uniquely to the economic landscape. In this comprehensive analysis, we will delve into the various aspects of the banking sector, exploring its structure, functions, regulatory environment, and emerging trends.

## Structure of the Banking Sector:

The banking sector is typically divided into two main categories: commercial banks and central banks. Commercial banks are the primary interface between customers and financial services, offering a range of products such as savings accounts, loans, and investment options. Central banks, on the other hand, are responsible for regulating and overseeing the entire financial system, controlling the money supply, and implementing monetary policy.

## Commercial Banks:

Commercial banks are the most visible face of the banking sector, providing essential financial services to individuals, businesses, and governments. These institutions accept deposits, facilitate withdrawals, and grant loans, forming the core of the banking experience for the general public. Commercial banks are categorized based on their size and the scope of services they offer, ranging from small community banks to large multinational institutions.

## Regulatory Environment:

The banking sector is subject to strict regulatory oversight to ensure financial stability and protect the interests of depositors and investors. Regulatory bodies, such as the Federal Reserve in the United States or the European Central Bank in the Eurozone, establish guidelines and conduct regular assessments to monitor compliance with regulations.

## Emerging Trends in the Banking Sector:

### Digital Transformation:

The banking sector is experiencing a rapid shift towards digitalization. Online banking, mobile apps, and digital wallets have become integral parts of the customer experience, offering

convenience and accessibility.

#### Fintech Disruption:

The rise of fintech companies is challenging traditional banking models. Fintechs leverage technology to provide innovative financial services, ranging from peer-to-peer lending to robo-advisors, often offering more streamlined and cost-effective solutions.

#### Blockchain and Cryptocurrencies:

The adoption of blockchain technology and cryptocurrencies is gaining momentum in the banking sector. These technologies offer enhanced security, transparency, and efficiency in financial transactions.

#### Data Analytics and Artificial Intelligence:

Banks are increasingly using data analytics and artificial intelligence to analyze customer behavior, assess credit risk, and enhance overall decision-making processes. This allows for more personalized services and improved risk management.

#### Regulatory Compliance and Cybersecurity:

With the increasing complexity of financial transactions, banks are focusing on robust regulatory compliance and cybersecurity measures to protect customer data and ensure the integrity of financial systems.

#### Conclusion:

The banking sector is a dynamic and integral component of the global economy, providing essential financial services that facilitate economic growth and development. With ongoing technological advancements and evolving consumer preferences, the industry is undergoing a transformative phase. Adapting to these changes while maintaining regulatory compliance will be crucial for the sustained success and stability of the banking sector in the years to come.

## **Company profile – PATAN URABAN CO –OP BANK LMT.**

**Perspective:** Our goal is to rank among India's top urban cooperative banks in terms of client satisfaction, profitability, and superior service delivery within the banking industry

**Mission:** At Patan Co-Operative Bank, we are dedicated to providing our shareholders with the highest dividend returns possible while maintaining the quality of our services. We do this by managing our business effectively and efficiently.

**Background** -The banking industry was split into nationalized banks, urban cooperative banks, and rural cooperative banks in 1980, when the Indian economy had not yet opened up. Additionally, credit societies had a strong hold on the realm of financial transactions.

Four gentlemen from Gujarat's Patan Jain Community founded "Patan Co-operative Society Ltd.", a credit cooperative organization, with the goal of assisting other traders and businesspeople.

### **Successes:**

In 1983, the bank operated out of a single branch and had total deposits of Rs. 49.16 lacs and total advances of Rs. 25.00 lacs. The humble starting has allowed for incredible advancement throughout time. The bank now has over Rs. 29.00 crores in advances and over Rs. 83.00 crores in deposits. Patan Cooperative Bank Ltd., with just three branches, obtained the numbers that are typically attained by banks with several branches. This might be the first instance of its kind in the Indian banking sector.

The Bank has a strong track record of consistently and yearly delivering dividends to its shareholders.

The bank's gross non-performing assets (NPA) as

## Review of Literature

D.T. Pham(2004), et al. [1] proposed a measure to determine the clusters for the k means algorithm for the different datasets.

- Unsupervised classification can be called as clustering if there is no virtual for labeled data. Clustering is used to classify the classes and assign objects to each class. By using Clustering techniques we put similar types of objects in the same class or same group or they are different or unrelated from other classes or other groups [2].
- Shreyash S. Mane, Shweta M. Barhate, Satish J. Sharma (June 2022), et al.[3] From this research paper We got that The cluster representation of the k-means algorithm is based on the mean and to initialize the cluster, select randomly k-means.
- Savi Gupta and Roopal Mamtora (2014), et al. [4] Use data mining methods to identify relationships between large datasets. It makes sense to look at a customer's buying behavior. This paper describes existing data mining algorithms for shopping cart analysis.
- Pradeep Rai, Shubha Singh(2010) et al. [5] demonstrated the types of hierarchical clustering and how it works.

## RESEARCH METHODOLOGY

- Data Source

For this Project the customer personality analysis dataset which is the secondary dataset has been collected from the [ineuron.ai](http://ineuron.ai) website I neuron is a reputed institution in the data sciencetraining field from India.

The Dataset contains 2240 rows and 30 columns. This dataset includes three categorical columns such as ‘Marital Status’, ‘Education’ and ‘Dt Customer’. 27 columns are numerical columns such as ‘ID’, ‘Income’, ‘Teenhome’, ‘kidhome’, ‘Recency’ and so on.

Website - [www.ineuron.ai](http://www.ineuron.ai)

- Variables Description:

1. ID- Customer's unique identifier
2. Year Birth - Customer's birth year
3. Education: Customer's education level
4. Marital\_Status - Customer's marital status
5. Income- Customer's yearly household income
6. Kidhome - Number of children in customer's household
7. Teenhome- Number of teenagers in customer's household
8. Dt\_Customer - Date of customer's enrollment with the company
9. Recency- Number of days since customer's last purchase
10. Complain- 1 if the customer complained in the last 2 years, 0 otherwise
11. MntWines - Amount spent on wine in last 2 years
12. MntFruits - Amount spent on fruits in last 2 years
13. MntMeatProducts -Amount spent on meat in last 2 years
14. MntFishProducts - Amount spent on fish in last 2 years
15. MntSweetProducts - Amount spent on sweets in last 2 years
16. MntGoldProducts - Amount spent on gold in last 2 years
17. NumDealsPurchases - Number of purchases made with a discount
18. AcceptedCmp1 - 1 if customer accepted the offer in the 1st campaign, 0 otherwise
19. AcceptedCmp2 - 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
20. AcceptedCmp3 - 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
21. AcceptedCmp4 - 1 if customer accepted the offer in the 4th campaign, 0 otherwise
22. AcceptedCmp5 - 1 if customer accepted the offer in the 5th campaign, 0 otherwise

23. Response -1 if customer accepted the offer in the last campaign, 0 otherwise
24. NumWebPurchases - Number of purchases made through the company's website
25. NumCatalogPurchases - Number of purchases made using a catalogue
26. NumStorePurchases -Number of purchases made directly in stores
27. NumWebVisitsMonth -Number of visits to the company's website in the last month

## **Statistical Tools and Techniques**

Techniques:

Unsupervised machine learning algorithms

- Hierarchical clustering
- Kmeans clustering

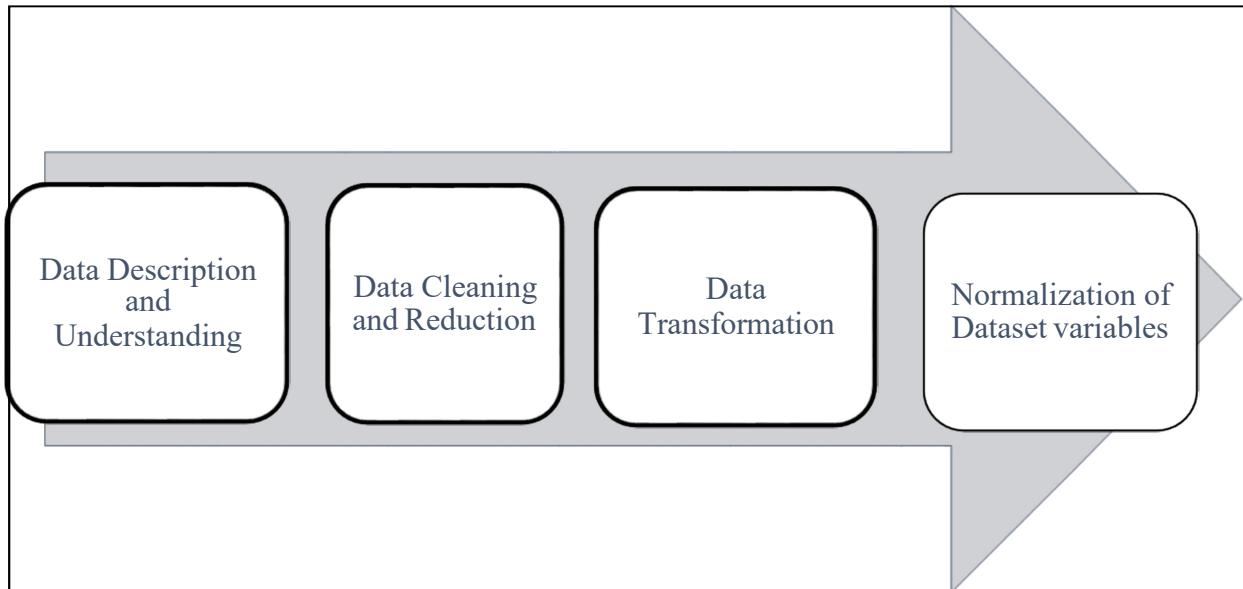
EDA

Tools:

- Python programming in jupyter notebook

# Data Preprocessing

The data preprocessing is undertaken in the following steps,

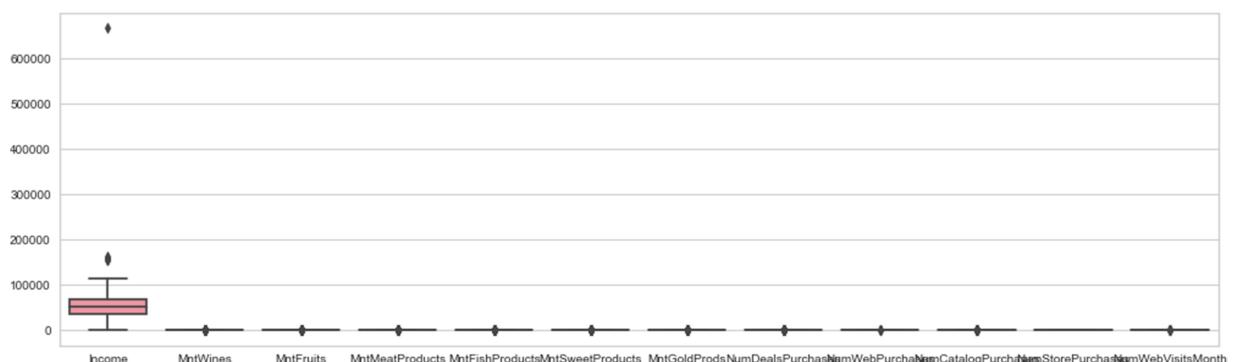


- Data Understanding and Description :

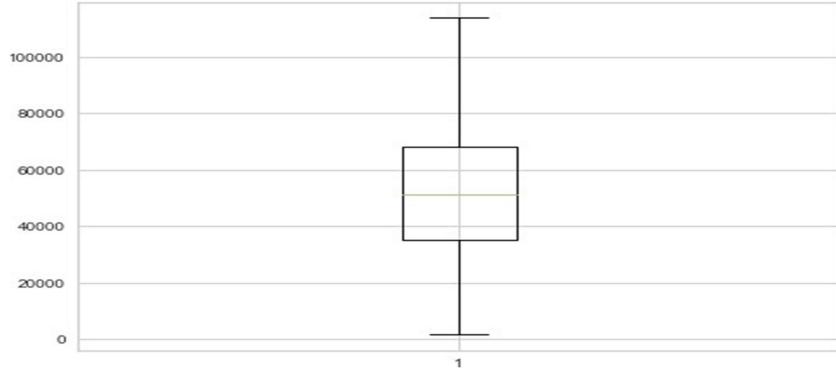
To predict the attrition of an employee based on the variables, describe the variables before analyzing them so that have a clear understanding. Information about the number of observations, and the data type of each variable.

- Data Cleaning and Reduction :

In Data Cleaning, to check for the missing values and observed that in this dataset income column has 24 missing values, these missing have been filled by using the KNN method. After filling in the missing values boxplot has been drawn to check the outliers in the data, and outliers have been detected in the income column.



Here I by using IQR method these outliers have been removed.



- Data Transformation:

- Attribute Construction:

- The first ‘Current year’ column has been constructed, by using the ‘Current year’ column and ‘year\_birth’ column we created the ‘Age’ column. ‘Total\_spend’ is created by using some existing columns. The ‘Registered\_days’ column has been created by using the ‘Dt\_Customer’ column. There were outliers in ‘Age’ Column, so by using cap method these outliers have been removed.

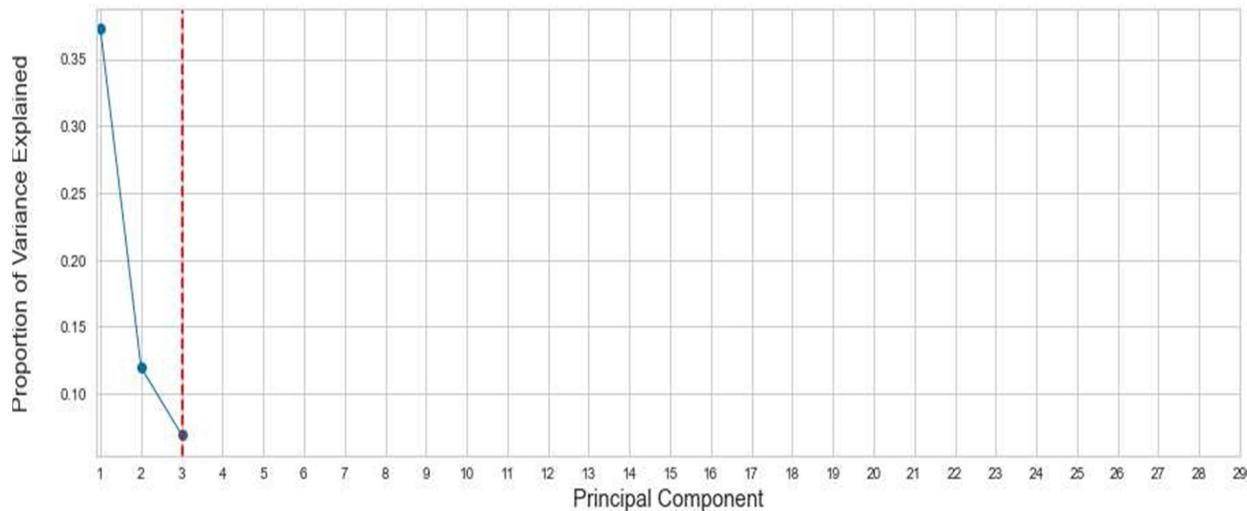
- Normalization of Dataset

- The data has been normalized by using standard scaler

PCA for dimensionality reduction:

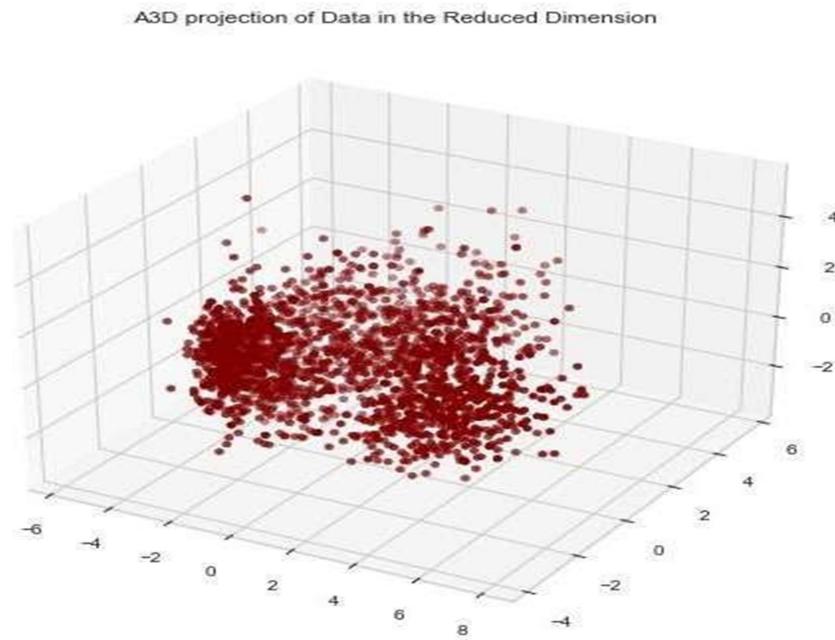
PCA or Principal Component Analysis is the unsupervised machine learning tool for the analysis of huge datasets containing the highest number of variables or features or dimensions, increasing the interpretability of data with preserving the maximum amount of information. PCA reduces the problem of overfitting in the phase of training the model by decreasing the number of features or variables so for this particular reason, before running features or attributes through a classifier we have to reduce the dimensions of the data. PCA is the statistical procedure that uses an orthogonal transformation to convert a set of correlated variables into uncorrelated variables which means the principal components which we obtain should be independent of each other. PCA is the tool which is most widely used in exploratory data analysis and in machine learning for the predictive model. PCA is also called as black box technique.

Scree Plot



The number of components we choose ultimately depends on our preference. Typically, it is recommended that the total variance explained by all components should be between 70% to 80%, which would equate to roughly 6 components in this case. Despite this, for the present example, only 3 components will be selected, which should preserve a little over 55% of the variance. While this may not be ideal in social sciences, it is still a reasonable level of variance retention. PC1 is responsible for around 36% of the overall variation in the dataset and has the greatest impact on it.

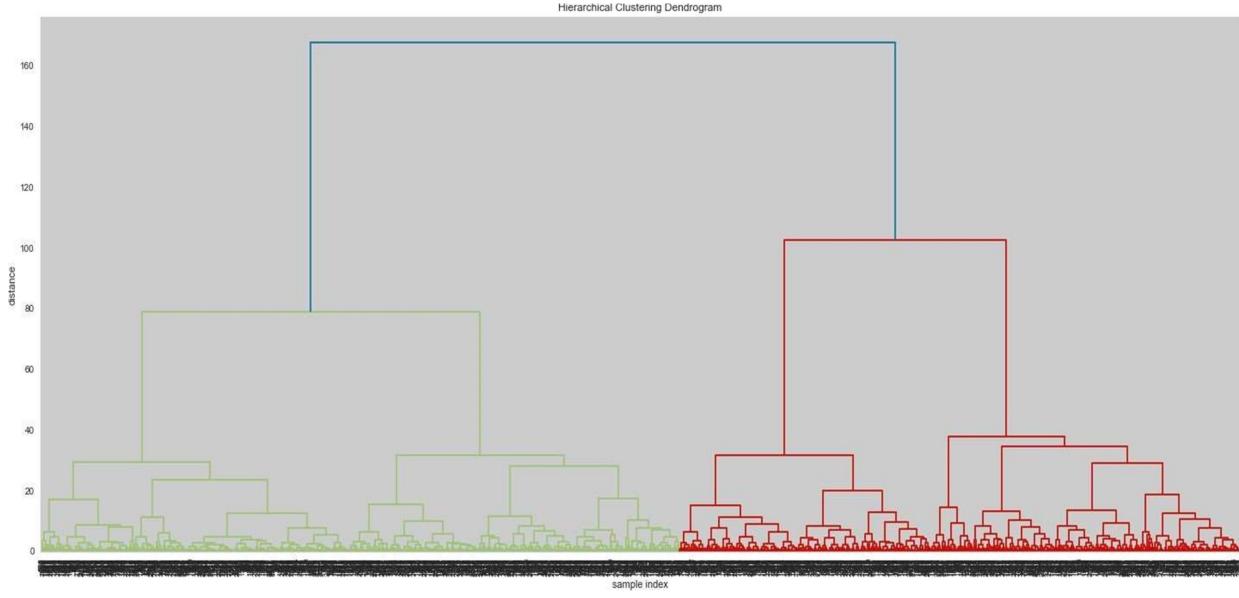
3D Representation of the principal components:



Above is the 3D representation of the Principle Components.

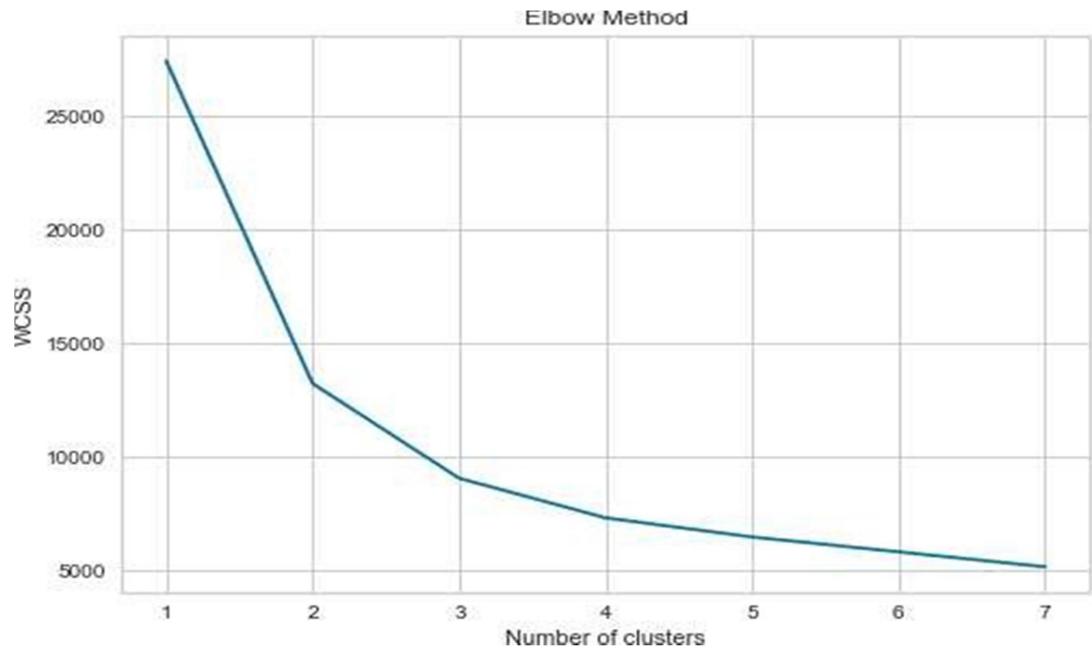
Hierarchical Clustering:

Hierarchical clustering is what we used in this research study. The unsupervised machine learning approach used to cluster the data is called Hierarchical Clustering. It is a cluster analysis technique that aims to create a tree-like hierarchy among the clusters. Agglomerative clustering and divisive clustering are the two hierarchical clustering techniques that group objects together based on shared attributes. The most widely used technique, however, employs a bottom-to-top approach and treats each object as a singleton cluster: agglomerative clustering. Pairs of clusters are gradually combined until all clusters are integrated into one big cluster with all components. The dendrogram, a tree-based visualization, is the result of hierarchical clustering.



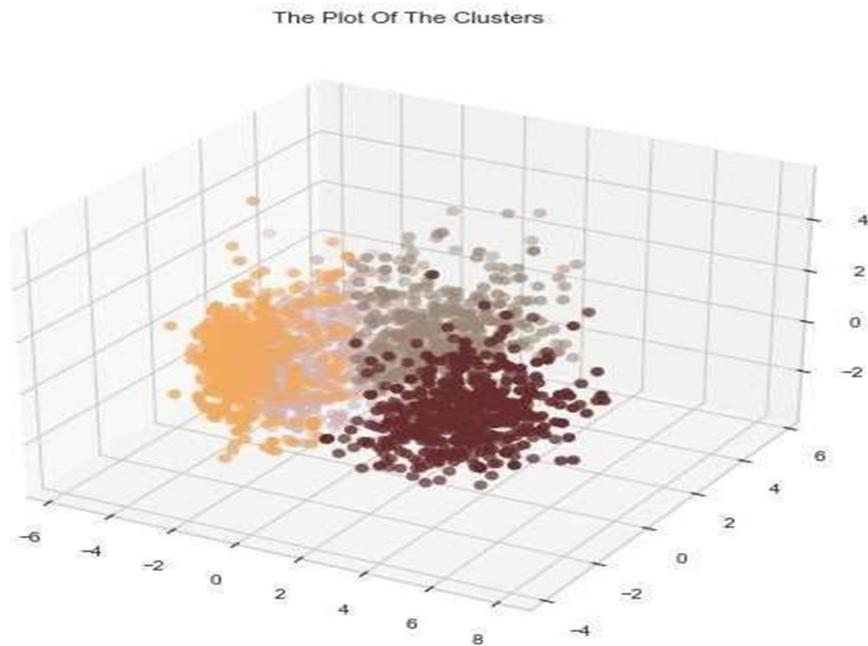
### K-means Clustering:

K-means clustering is a popular unsupervised machine-learning technique for grouping purposes. The vector quantization technique known as K-means clustering originated in signal processing. Creating K clusters with somewhat similar properties for each observation within each cluster and distinct observations within each cluster is the ultimate aim of this strategy. K-means is a centroid-based clustering technique in which the distance between a data point and a centroid—which may be calculated using the Euclidean distance formula—determines which cluster the point belongs to. We must choose how many clusters to incorporate observations in when using this strategy. In order to establish a predetermined number of clusters, we employed the elbow



From the above elbow plot, we decide that our data is going to be clustered into 4 clusters.

3D representation of clusters:



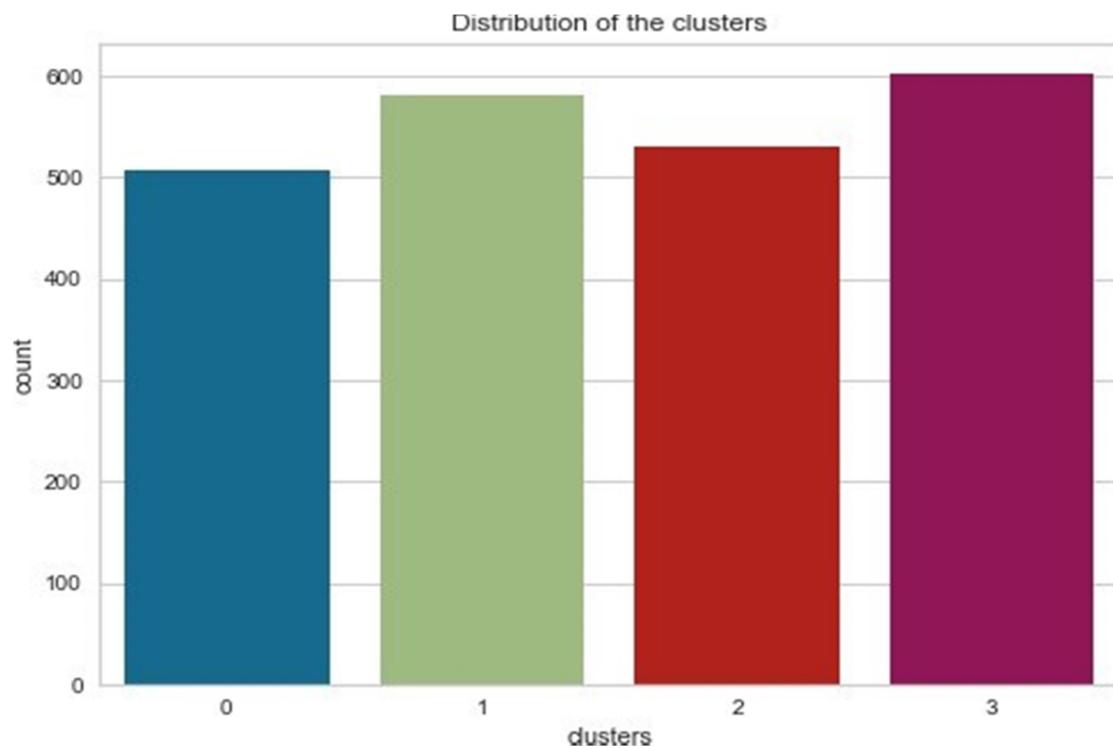
The above plot represents the result of K-means clustering which is in the 3D plot.

## EXPLORATORY DATA ANALYSIS

Based on clusters which are formed by using Kmeans method

Bar chart:

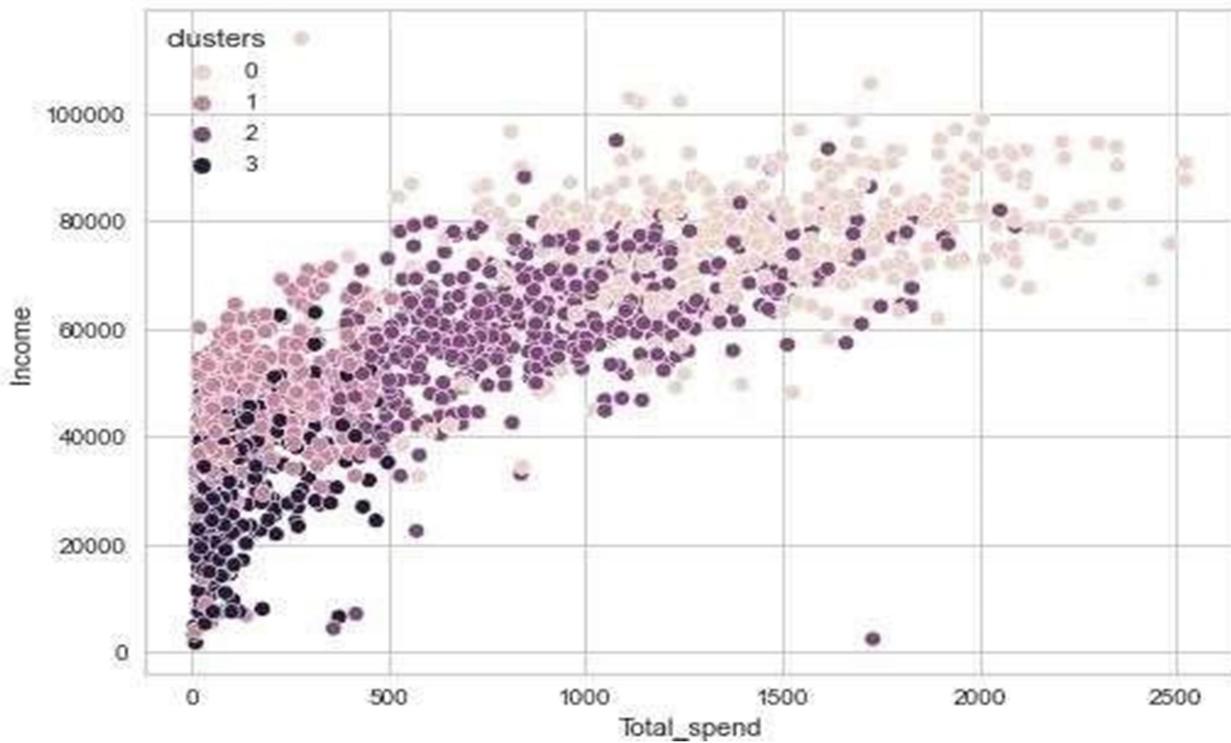
To see the count of each cluster



The count plot of the clusters created by the K-means clustering technique is shown below. There are 508 observations in cluster 0, 581 observations in cluster 1, 531 observations in cluster 2, and 602 observations in cluster 3.

scatter plot -

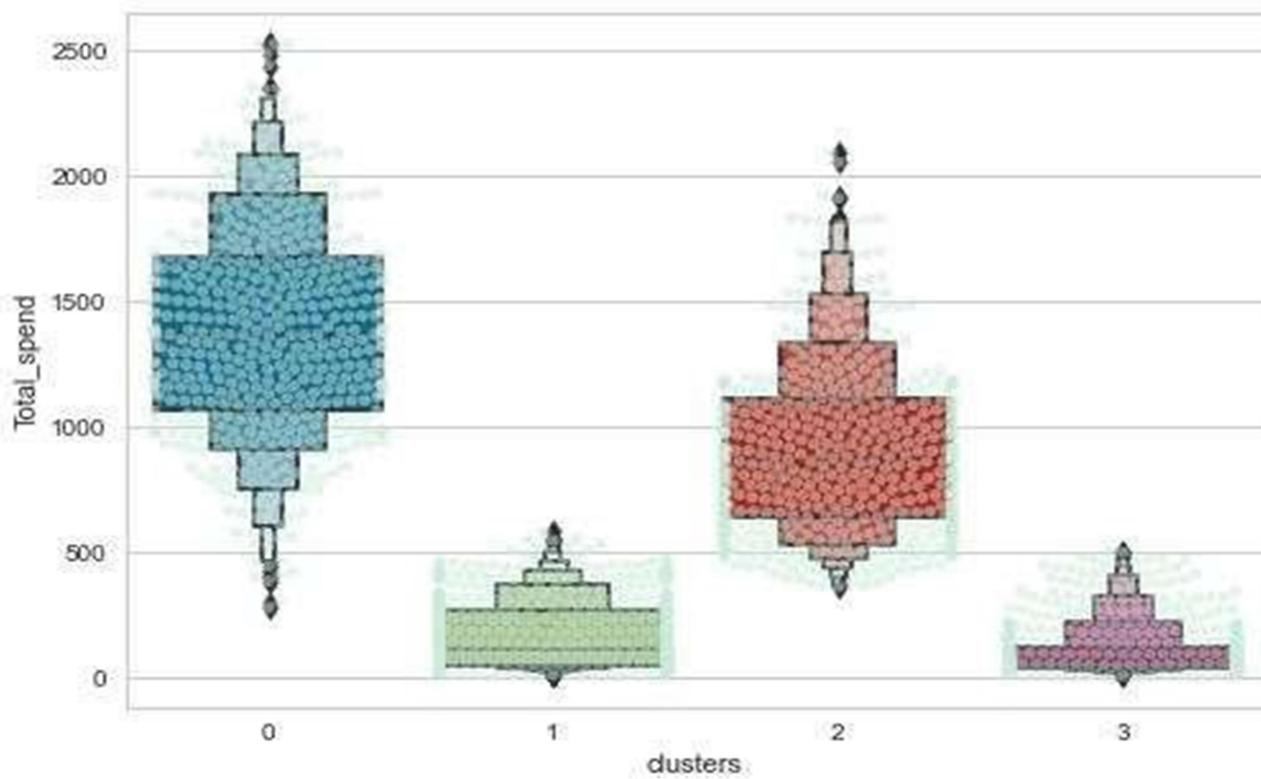
To check the relation between customer's income and spend



The relationship between income and total spending is depicted in this scatter plot, and it is clear that as income rises, so does the total amount spent. Customers in Cluster 3 are low earners and low spenders; customers in Cluster 2 are average earners and average spenders; customers in Cluster 1 are high earners but low spenders; and customers in Cluster 0 are the highest earners and highest spenders.

Swarmplot -

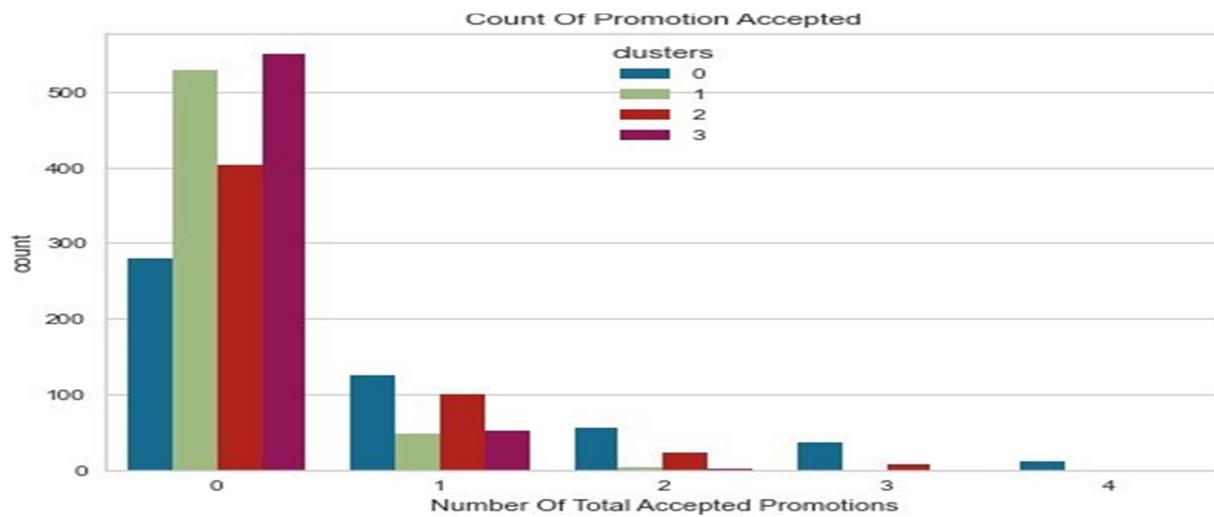
To show the group of customers according to the total spend



We can see from this map that cluster 0 has the largest client base, followed by cluster 2, cluster 1, and cluster 3, which has the smallest customer base. This allows us to examine each cluster's spending patterns and determine which one to concentrate on while developing a marketing plan to reach the highest spending clusters.

Bar plot -

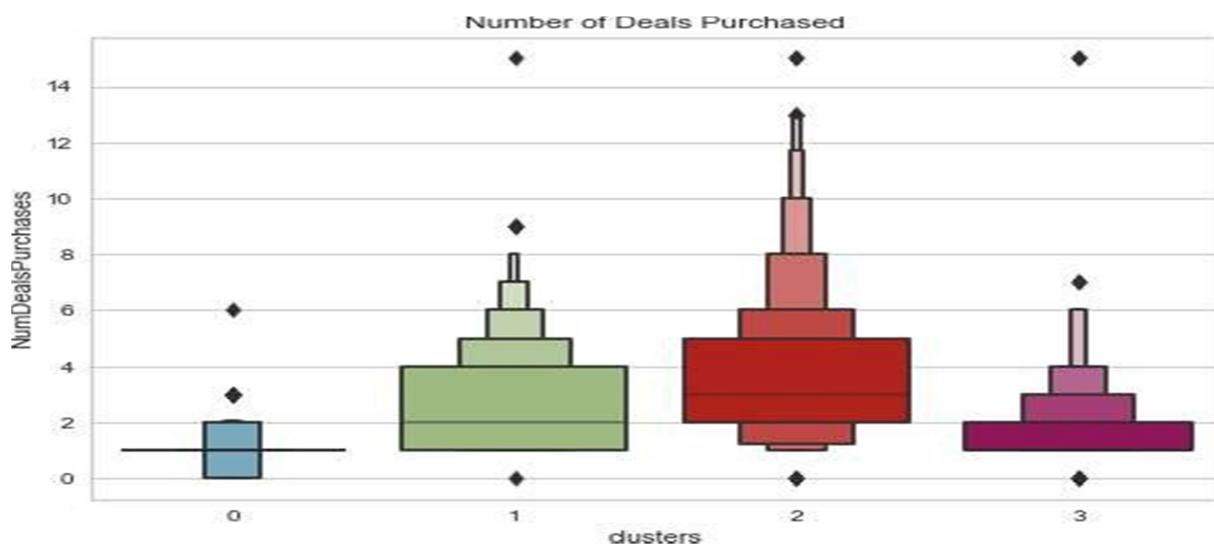
To check the response of the customers to the promotions done by the company



We can see that response to the marketing campaign by the customers is very low, most of the customers did not participate in the marketing campaign. Few customers participated in the marketing campaign. This suggests that more targeted and well-planned campaigns are needed so that more customers would participate in the campaigns and sales could be boosted.

Boxenplot :

To check the response of the customers to the deals offered by the company

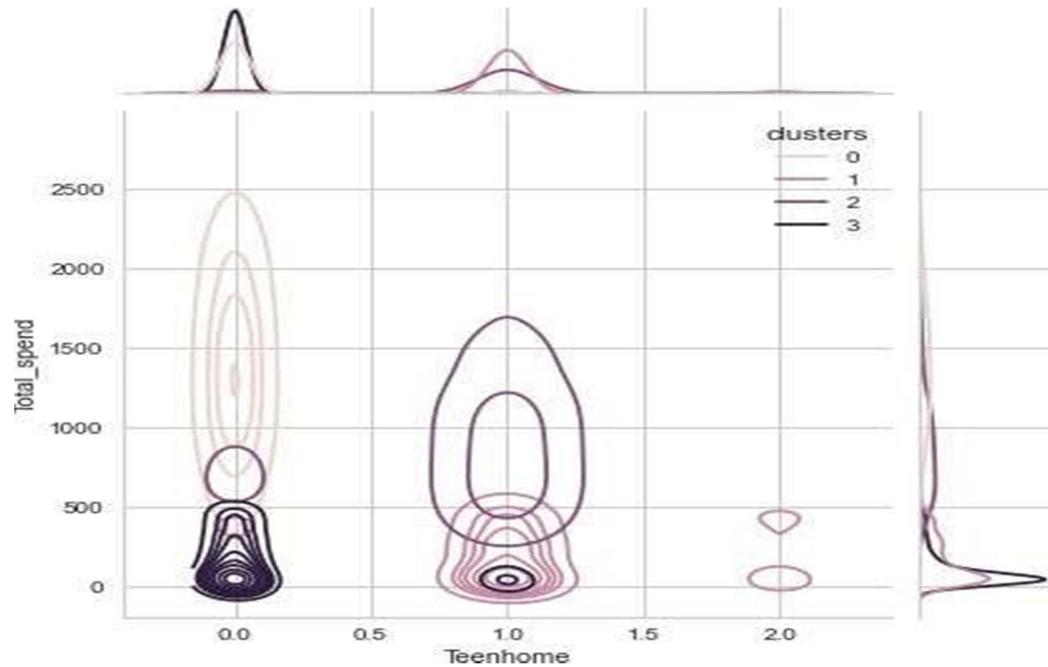


In contrast to campaigns, customers have given better responses to the deals offered by the company. The company had the best results in cluster 1 and cluster 2. Customers from cluster 0 and 3 seems to be not interested in the deals.

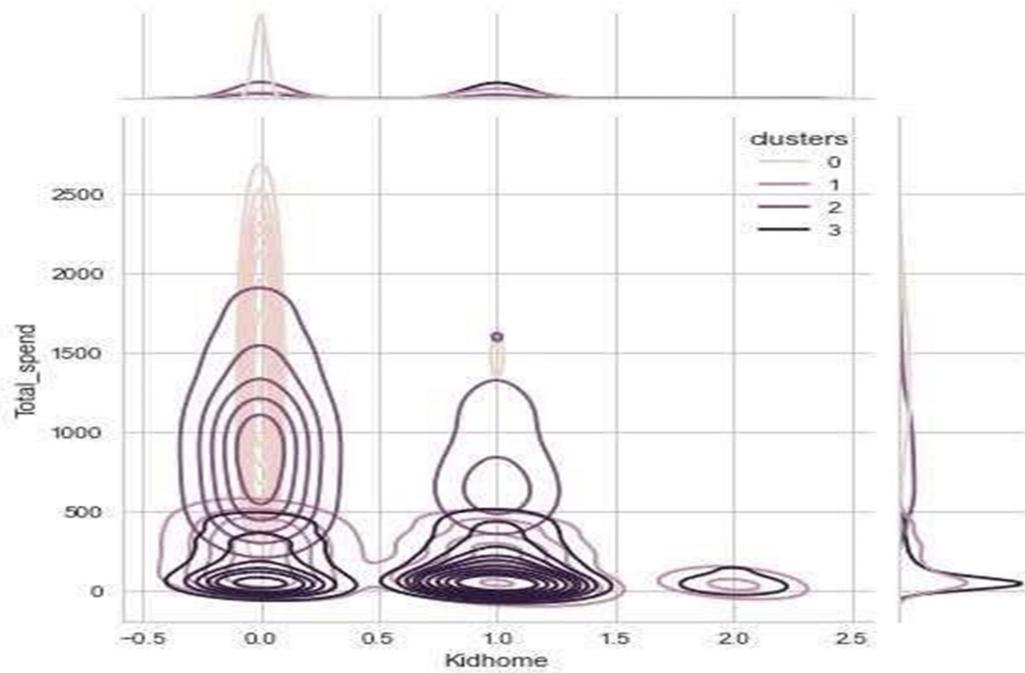
Joint plot:

For the profiling of the customers

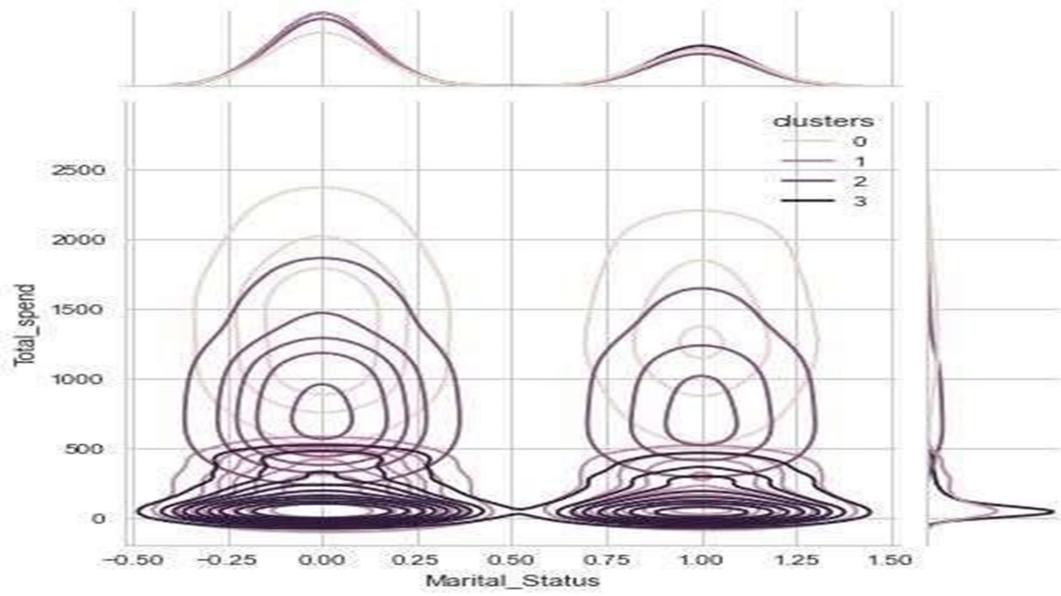
Total\_spend Vs Teenhome -



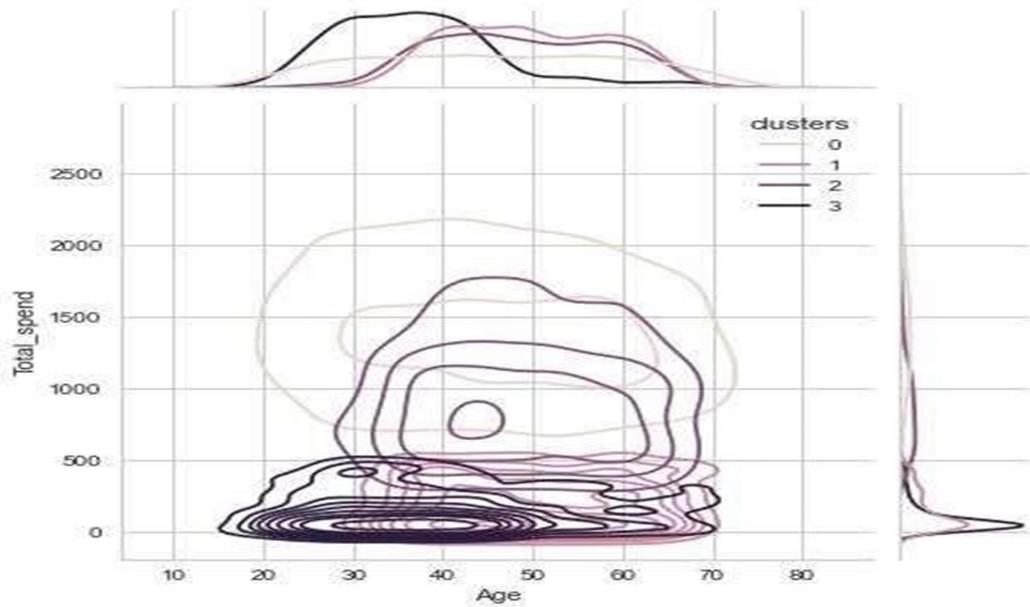
Total\_spend Vs Kidhome



Total\_spend Vs Marital\_Status:



Total\_spend Vs Age



About cluster 0:

As shown in the above figures this is the group of such customers there are no teenagers and kids in their homes but the majority of them are married. It has a wide range of age, from teenagers to older and it is high spending group so it should be a high-income group.

About Cluster 1:

As shown in the figure this group of such customers there are teenagers and kids in their homes but a minimum of them are married. This cluster includes most of the adults and older customers. It is a relatively low-income group so it should be a low-income group.

About Cluster 2:

As shown in the above figure there is more number of customers who don't have kids but have teenagers in their homes. There are a minimum number of customers who are married. This group contains only adults and it is an average spending group so it should be an average income group.

About Cluster 3:

As shown in the above figure there are half of the customers have kids and half of the customers don't have kids in their homes but most of the customers don't have teenagers in their homes. Half of the customers are married and half of the customers are single. It has wide a range of ages from teenagers to older. It is very low spending so it should be a low-income group.

## **Findings**

### **Income-Spending Relationship:**

One of the objectives is to check the relationship between customers' income and their spending. The analysis may reveal insights into how different income groups allocate their spending across various product categories. For instance, it could identify whether higher-income customers tend to spend more on luxury items or experiences, while lower-income customers focus on essential goods.

- **Customer Response to Promotions:**

The objective of checking the response of customers to promotions is crucial for understanding the effectiveness of marketing strategies. The analysis may uncover patterns related to which customer segments are more responsive to promotional activities. It could reveal whether discounts, special offers, or loyalty programs have a significant impact on specific customer groups.

- **Customer Segmentation:**

Unsupervised machine learning models, such as DBSCAN clustering and hierarchical k-means, are employed to identify customer segments with similar behaviors and preferences. The major finding here would be the identification of distinct customer segments based on various factors like purchasing patterns, preferences, and responses to promotions. This segmentation can then guide targeted marketing strategies for each segment.

- **Product Adaptation and Customer Satisfaction:**

The overarching objective is to help businesses adapt their products according to the unique wants, behaviors, and concerns of different customer segments. The analysis may provide insights into specific product features or modifications that could enhance customer satisfaction within each identified segment. This could include tailoring marketing messages, product design, or service

offerings to better align with customer preferences.

- Data-driven Decision Making:

The shift from manual customer personality analysis to leveraging machine learning models suggests an emphasis on data-driven decision-making. The major finding in this context would be the efficiency and speed gained by automating the analysis process. It allows businesses to extract valuable insights from large datasets in a more timely and accurate manner, enabling proactive decision-making.

- Enhanced Customer Understanding:

The overall objective of the analysis is to help businesses better understand their customers. Major findings would include a more nuanced understanding of customer behaviors, preferences, and needs. This understanding can empower businesses to tailor their strategies, leading to improved customer satisfaction, loyalty, and ultimately, business growth.

It's important to note that specific findings would depend on the nature of the data, the effectiveness of the machine learning models, and the depth of the analysis conducted by the business.

## **Conclusion**

In this article for ‘Customer personality analysis’ unsupervised machine learning can be an effective method. By using customer data we find the relations within customer data and can predict customers’ tendencies and loyalty based on their purchasing behaviors and characteristics. This can provide the company with important insight into customer data so that the company decide marketing and customer service strategy to give the best service to them according to their preferences.

Cluster 0 is the highest spending group of the company closely followed by cluster 2, the company should keep on updating their marketing strategy to maintain this loyal customer base, cluster 1 is the average income group but they didn’t purchase enough from the company so the company should decide different marketing strategy other than cluster 0 and 2 so that cluster 1 customers would purchase more goods from the company. Cluster 3 is low-income as well as low-spending groups, company should also think about them, the company can provide affordable goods with good quality. Customers didn’t give the response for the campaigns but customers gave better responses for deals offered by companies, so the company should focus on the deals than the campaigns.

## References

- [1] D T Pham, S S Dimov, and C D Nguyen (2004) Selection of K in K-means clustering.  
DOI: 10.1243/095440605X8298
- [2] Data Mining Techniques <http://www.zentut.com/data-mining/datamining-techniques>.
- [3] Pradeep Rai, Shubha Singh. A Survey of Clustering Techniques. International Journal of Computer Applications (0975 – 8887) Volume 7– No.12, October 2010
- [4] Shreyash S. Mane, Shweta M. Barhate, Satish J. Sharma. CUSTOMER BEHAVIOUR ANALYSIS USING DATA MINING.ISSN 2349-5162.2014
- [5] Savi Gupta and Roopal Mamtora. A Survey on Association Rule Mining in Market Basket Analysis, (2014), ISSN 0974-2239