



Customer Personality Analysis

Submitted by (TEAM #2)

**Prasad Thamada
Ankita Manjarekar
Pallavi Bongu
Swetha Kancharla**

Submitted for DSCI-6007-01
DR. ARDIANA SULA

CONTENTS

Executive Summary..... 2

Abstract..... 2

Highlights of Project..... 3

Introduction..... 4

Process Flow..... 5

Methodology..... 6

Results..... 10

Conclusion..... 20

Reference..... 21

Executive Summary

Customer Personality Analysis is a comprehensive examination of a company's diverse customer base. This analysis enables businesses to gain a deeper understanding of customer behavior, preferences, and concerns, facilitating the customization of products and services to cater to specific customer segments. By leveraging this insight, companies can optimize their marketing strategies by targeting the most receptive customer segments, thereby enhancing the efficiency of their promotional efforts, and ultimately driving sales. Customer Personality Analysis is a powerful tool that empowers businesses to align their offerings with the distinct needs of various customer groups, leading to improved customer satisfaction and loyalty.

This summary encapsulates the essence of Customer Personality Analysis, emphasizing its pivotal role in enabling businesses to tailor their products and marketing initiatives to different customer segments, thereby enhancing customer engagement and driving business growth.

Abstract

This project aims to leverage unsupervised learning techniques, including Principal Component Analysis (PCA) and K-Means clustering, to gain valuable insights into the characteristics, preferences, and behavior of a company's ideal customers. By segmenting customers into distinct groups based on their personalities and behavioral patterns, the project seeks to enable the business to better understand its customer base and tailor its products, services, and marketing strategies to meet the specific needs of each segment. The goal is to drive improved customer satisfaction, loyalty, and business performance by delivering more personalized and impactful customer experiences. The project has the potential to provide significant value to the business by enhancing its ability to adapt to the evolving needs of its diverse customer base and make data-driven decisions to inform strategic choices and drive business growth.

Highlights of the Project

The Customer Personality Analysis model is a strategic project that enables businesses to tailor their products and marketing strategies based on the specific characteristics and behavioral patterns of their target customers. This in-depth analysis involves segmenting customers into distinct groups based on their personalities and behaviors, allowing businesses to gain valuable insights into their diverse customer base.

To achieve this, the model leverages advanced techniques such as Principal Component Analysis (PCA) to reduce the dimensionality of the dataset and remove redundant data, thereby facilitating feature engineering. Subsequently, KMeans clustering is applied to the refined features, leading to the segmentation of customers into different clusters. This segmentation enables businesses to identify the types of customers most likely to purchase a particular product at a given time, thereby optimizing their marketing efforts, and enhancing customer satisfaction.

Customer Personality Analysis is a pivotal tool that empowers businesses to better understand their customers and adapt to their evolving needs. By leveraging this model, businesses can create more personalized and impactful customer experiences, ultimately driving customer satisfaction, loyalty, and business success. The analysis also facilitates the identification of new opportunities for product development, innovation, and market expansion by uncovering unmet needs or underserved customer segments. Through continuous monitoring and analysis of customer data, businesses can ensure that their products and services remain relevant and competitive in the dynamic market landscape.

In conclusion, Customer Personality Analysis is a critical component of data-driven decision-making, enabling businesses to gain a comprehensive understanding of their customer base and make informed strategic choices. By harnessing the power of advanced analytical techniques, businesses can effectively tailor their offerings and marketing strategies to different customer segments, ultimately driving customer engagement, loyalty, and long-term business growth.

Introduction

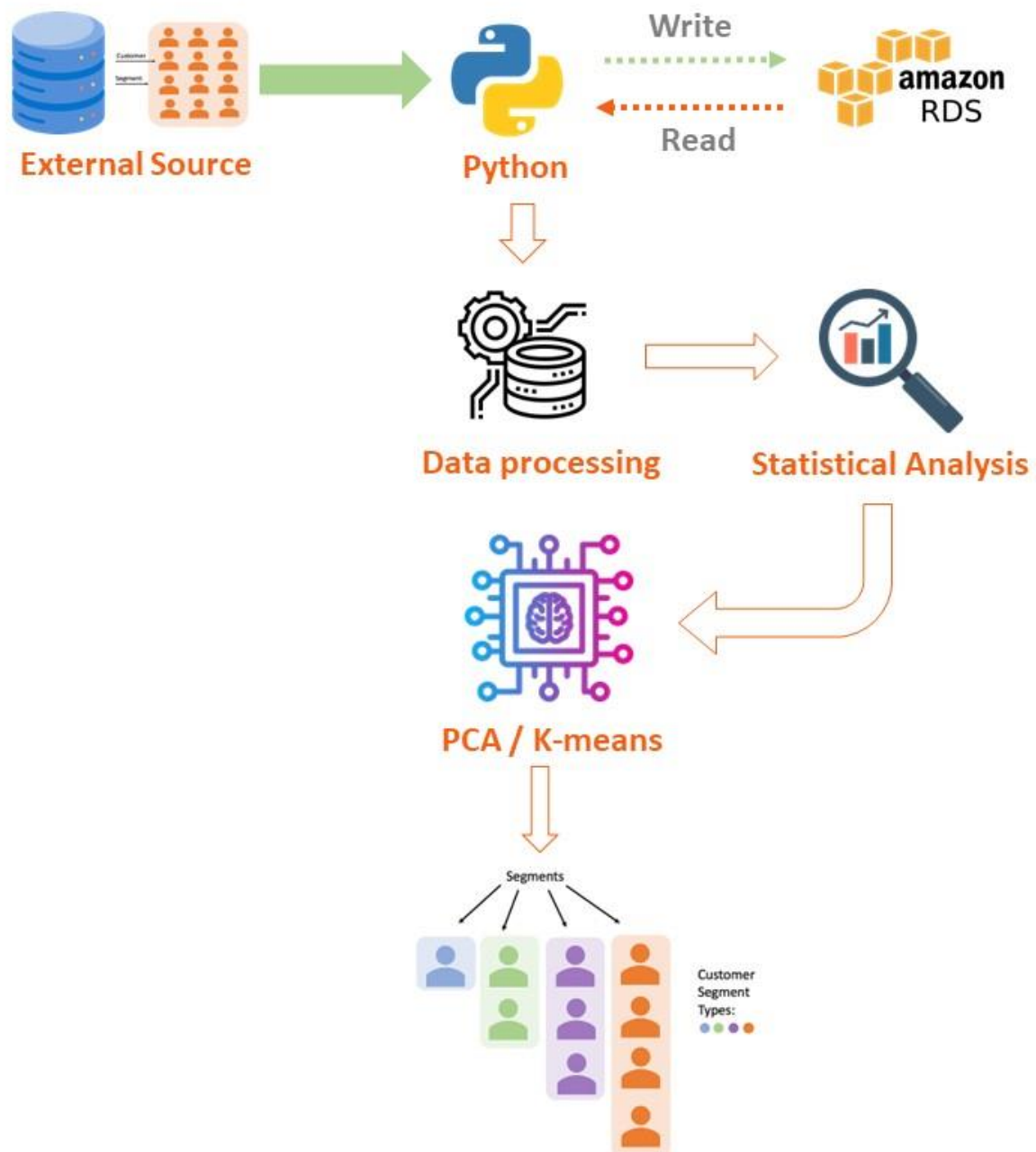
Customer Personality Analysis is a comprehensive process that involves the detailed examination of a company's ideal customers to gain valuable insights into their characteristics, preferences, and behavior. This analysis plays a crucial role in helping businesses better understand their customers, thereby enabling them to tailor their products, services, and marketing strategies to meet the specific needs of different customer segments. By leveraging various data sources and analytical techniques, businesses can create customer profiles or personas that represent distinct segments of their customer base, ultimately leading to more effective decision-making and resource allocation.

The primary objective of Customer Personality Analysis is to empower businesses to enhance customer satisfaction, drive business growth, and improve overall business outcomes. By gaining a deeper understanding of their customers, businesses can create more personalized and impactful customer experiences, leading to improved customer engagement and loyalty. This analysis also enables businesses to identify new opportunities for product development, innovation, and market expansion by uncovering unmet needs or underserved customer segments, ultimately contributing to the long-term success and sustainability of the business.

In summary, Customer Personality Analysis is a powerful tool that provides businesses with the insights needed to better understand their customers and adapt to their evolving needs. By leveraging this analysis, businesses can create more meaningful and impactful customer experiences, ultimately driving customer satisfaction, loyalty, and business success.

GitHub link : https://github.com/dsci6007-team2/customer_personality_analysis

Process Flow



Methodology

1. Data Collection

The dataset used for this project has four tables which are ingested in csv format. The four tables are mainly Customers, Products, Promotion and Purchasing Place. The dataset link is

<https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis/>

The columns for each of the tables are as follows:

a. Customers

- ID: Customer's unique identifier
- Year_Birth: Customer's birth year
- Education: Customer's education level
- Marital_Status: Customer's marital status
- Income: Customer's yearly household income
- Kidhome: Number of children in customer's household
- Teenhome: Number of teenagers in customer's household
- Dt_Customer: Date of customer's enrollment with the company
- Recency: Number of days since customer's last purchase
- Complain: 1 if the customer complained in the last 2 years, 0 otherwise

b. Products

- MntWines: Amount spent on wine in last 2 years
- MntFruits: Amount spent on fruits in last 2 years
- MntMeatProducts: Amount spent on meat in last 2 years
- MntFishProducts: Amount spent on fish in last 2 years
- MntSweetProducts: Amount spent on sweets in last 2 years
- MntGoldProds: Amount spent on gold in last 2 years

c. Promotion

- a. NumDealsPurchases: Number of purchases made with a discount

- b. AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise
- c. AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
- d. AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
- e. AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise
- f. AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
- g. Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

d. Purchasing Place

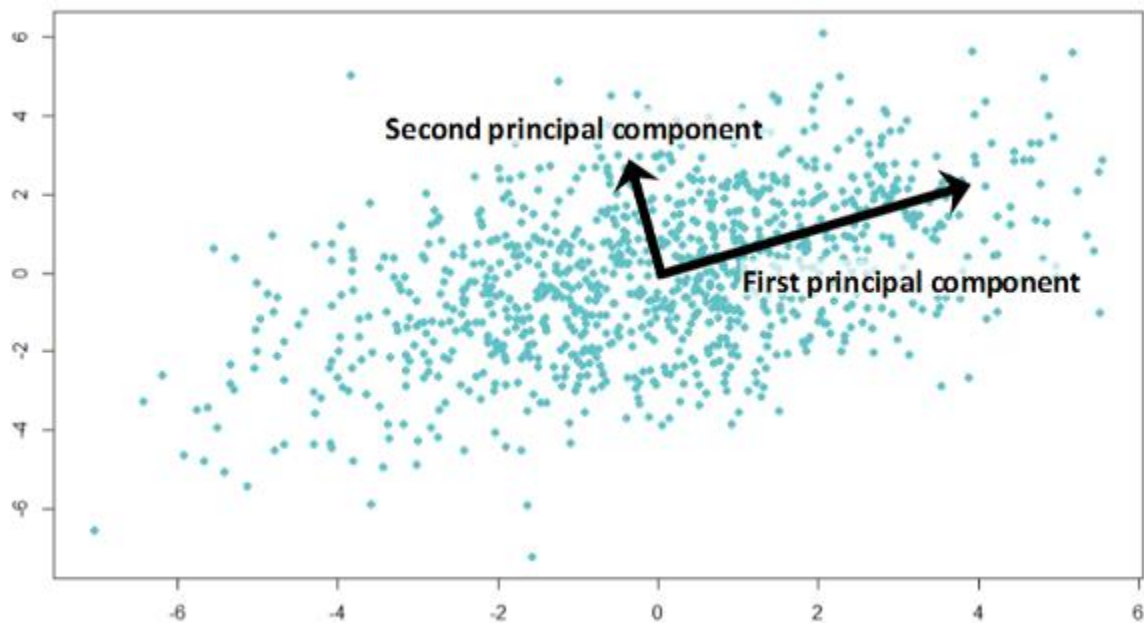
- a. NumWebPurchases: Number of purchases made through the company's website
- b. NumCatalogPurchases: Number of purchases made using a catalogue
- c. NumStorePurchases: Number of purchases made directly in stores
- d. NumWebVisitsMonth: Number of visits to company's website in the last month

2. Machine Learning Algorithms (Code: https://github.com/dsci6007-team2/customer_personality_analysis)

a. Principal Component Analysis (PCA):

It is a statistical technique used for dimensionality reduction, noise reduction, collinearity mitigation, feature selection, data compression, and preprocessing for clustering. It works by identifying correlations between data points and finding the most important lines, planes, or hyper-planes that approximate the data, allowing for the data to be represented by a smaller set of "summary indices." PCA is widely used in various fields, including data analysis, data compression, and signal processing, due to its ability to extract relevant information from complex data sets and facilitate data visualization and trend analysis.

In the context of the Customer Personality Analysis project, Principal Component Analysis (PCA) serves as a crucial technique for data preprocessing and dimensionality reduction. By applying PCA, the project aims to reduce the dimensionality of the customer data, thereby mitigating noise, reducing redundancy, and facilitating feature selection. This process allows for the extraction of essential information from the data while preserving its accuracy, ultimately enabling more effective clustering and segmentation of customer groups. Additionally, PCA supports data compression, making it easier to visualize and analyze the customer data, and provides a solid foundation for subsequent analysis and modeling.

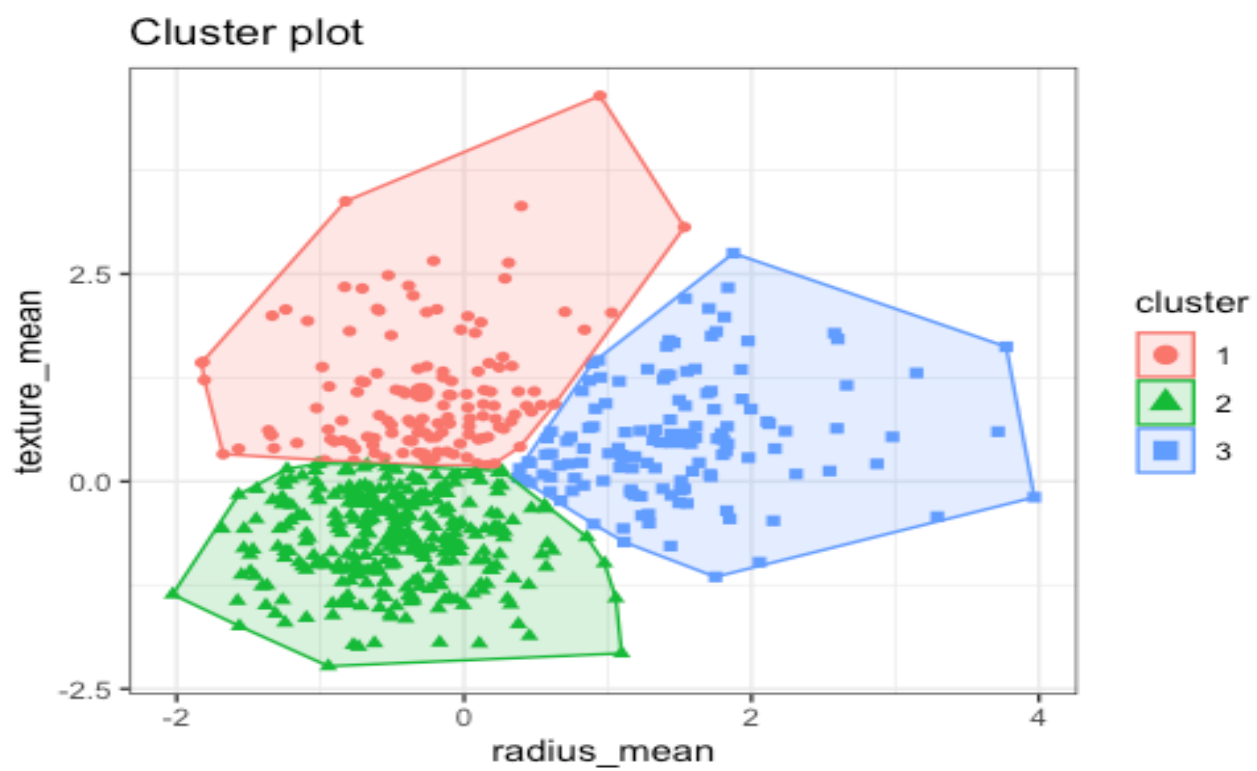


Fig(i) Principal Component Analysis

b. KMeans Clustering:

K-Means clustering is a popular unsupervised machine learning algorithm used for clustering data into distinct groups. The algorithm works by iteratively assigning data points to a specified number of clusters and then adjusting the cluster centroids to minimize the within-cluster variance. This process continues until the cluster assignments stabilize. K-Means clustering is widely used for customer segmentation, image compression, and anomaly detection, among other applications.

In the context of the Customer Personality Analysis project, K-Means clustering is employed to segment customers into various categories based on their characteristics and behavior. By applying K-Means clustering to the customer data, the project aims to identify distinct customer segments with unique personality traits and preferences. This segmentation enables the business to better understand its customer base and tailor its products, services, and marketing strategies to meet the specific needs of each segment. K-Means clustering facilitates more targeted and effective customer engagement, leading to improved customer satisfaction and business performance.



Fig(ii) KMeans Clustering

Results

i. Data Preprocessing

```
jupyter team2_solution_customer_personality_analysis Last Checkpoint: 6 hours ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (pykernel)

host = 'restaurant.cmjra71bg93.us-east-1.rds.amazonaws.com'
port = 3306
database = 'sample'

# Connect to the database
connection = pymysql.connect(
    host=host,
    port=port,
    user=user,
    password=password,
    database=database
)

# Create a cursor object
cursor = connection.cursor()

# Execute a SQL query
cursor.execute('SELECT * FROM marketing_campaign')

# Fetch the results
results = cursor.fetchall()

# Print the results
for result in results:
    print(result)

(5524, 1957, 'Graduation', 'Single', 58138, '0', '0', '04-09-2012', 58, 635, 88, 546, 172, 88, 88, '3', 8, '10', 4, 7, '0',
'0', '0', '0', '0', '0', 3, 11, '1')
(2174, 1954, 'Graduation', 'Single', 46344, '1', '1', '08-03-2014', 38, 11, 1, 6, 2, 1, 6, '2', 1, '1', 2, 5, '0', '0', '0',
'0', '0', '0', 3, 11, '0')
(4141, 1965, 'Graduation', 'Together', 71613, '0', '0', '21-08-2013', 26, 426, 49, 127, 111, 21, 42, '1', 8, '2', 10, 4, '0',
'0', '0', '0', '0', '0', 3, 11, '0')
(6182, 1984, 'Graduation', 'Together', 26646, '1', '0', '10-02-2014', 26, 11, 4, 20, 10, 3, 5, '2', 2, '0', 4, 6, '0', '0',
'0', '0', '0', '0', '0', 3, 11, '0')
(5324, 1981, 'PhD', 'Married', 58293, '1', '0', '19-01-2014', 94, 173, 43, 118, 46, 27, 15, '5', 5, '3', 6, 5, '0', '0', '0',
'0', '0', '0', 3, 11, '0')
(7446, 1967, 'Master', 'Together', 62513, '0', '1', '09-09-2013', 16, 520, 42, 98, 0, 42, 14, '2', 6, '4', 10, 6, '0', '0',
'0', '0', '0', '0', '0', 3, 11, '0')
(965, 1971, 'Graduation', 'Divorced', 55635, '0', '1', '13-11-2012', 34, 235, 65, 164, 50, 49, 27, '4', 7, '3', 7, 6, '0',
'0', '0', '0', '0', '0', 3, 11, '0')
(6177, 1985, 'PhD', 'Married', 33454, '1', '0', '08-05-2013', 32, 76, 10, 56, 3, 1, 23, '2', 4, '0', 4, 8, '0', '0', '0',
'0', '0', '0', 3, 11, '0')
```

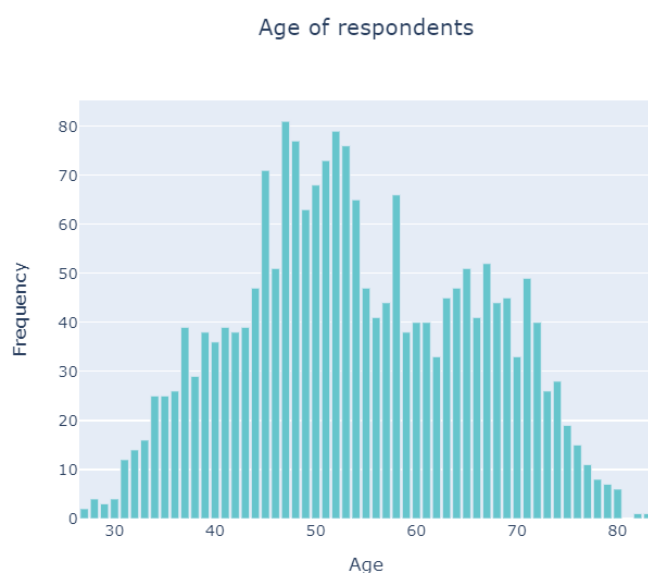
```
In [5]: column_names = [col[0] for col in cursor.description]
mark_camp = pd.DataFrame(results, columns=column_names)
mark_camp.tail()
```

```
Out[5]:
```

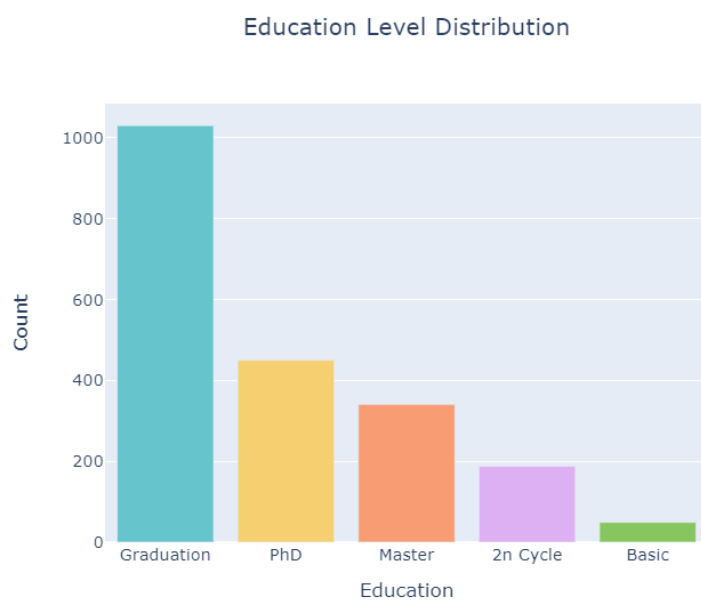
	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	AcceptedCmp3	
2235	10870	1967	Graduation	Married	61223	0	1	13-06-2013	46	709	...	5	0	
2236	4001	1946	PhD	Together	64014	2	1	10-06-2014	56	406	...	7	0	
2237	7270	1981	Graduation	Divorced	56981	0	0	25-01-2014	91	908	...	6	0	
2238	8235	1956	Master	Together	69245	0	1	24-01-2014	8	428	...	3	0	
2239	9405	1954	PhD	Married	52869	1	1	15-10-2012	40	84	...	7	0	

5 rows × 29 columns

Age Distribution Graph

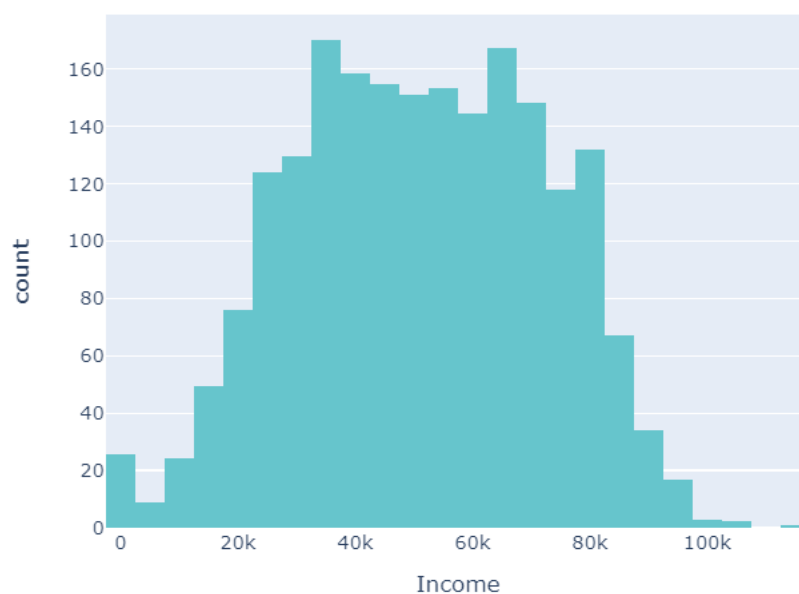


Education Level among Customers



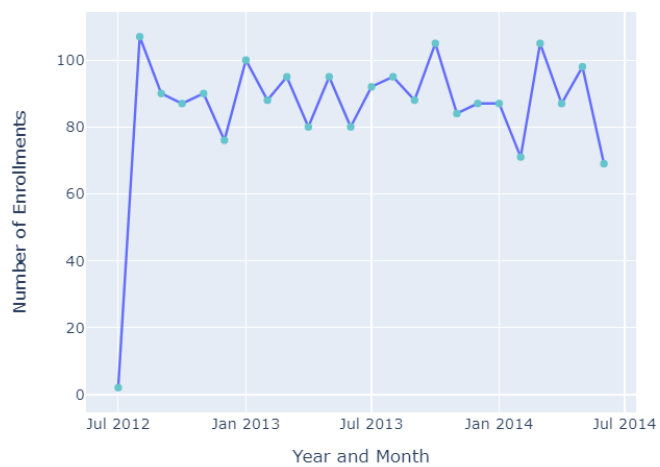
Income distribution

Income Distribution

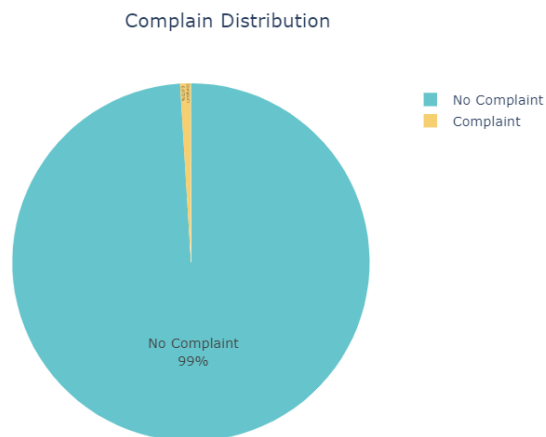


Company Growth based on number of Customers enrolled

Company Growth Over Time



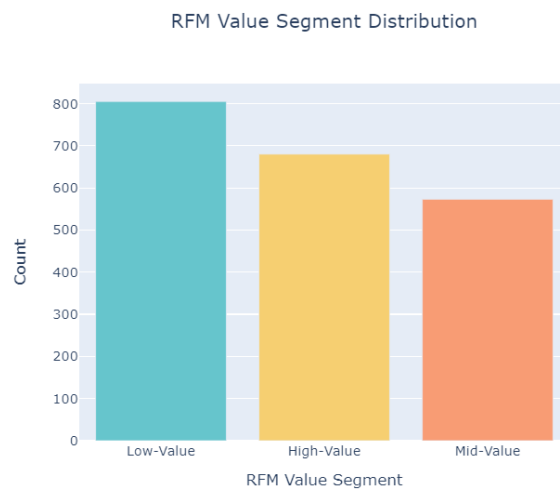
Complain Distribution

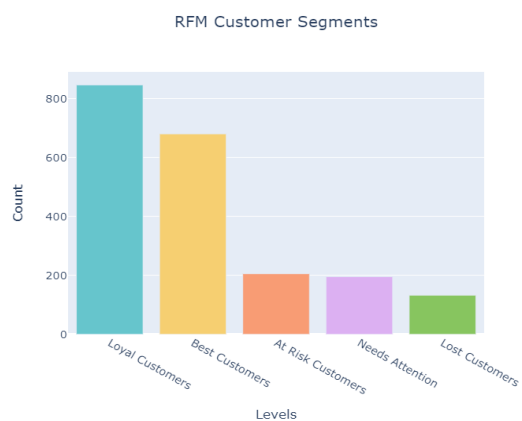


RFM (Recency, Frequency and Monetary)

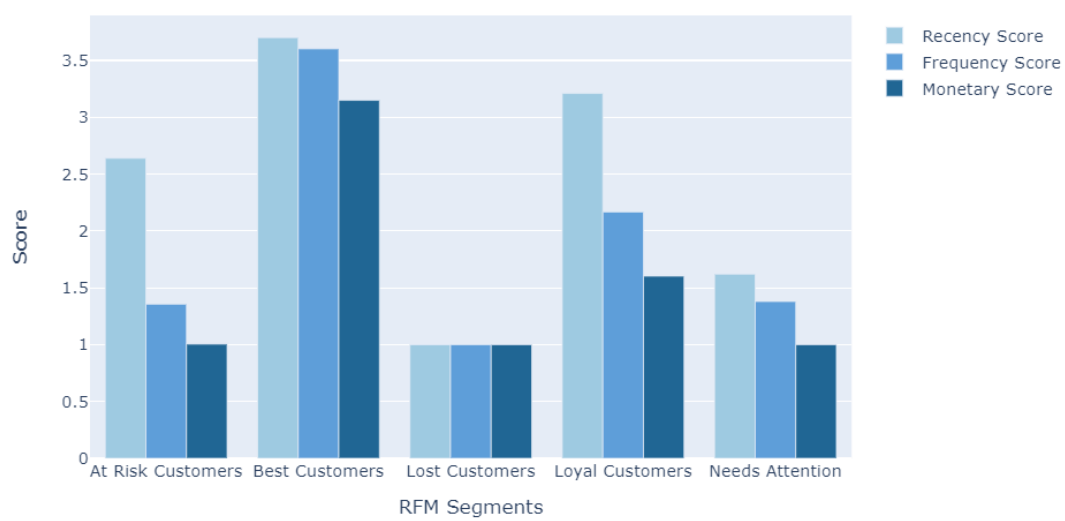
RFM is a method used for segmenting customers based on their past behavior. It looks at three key dimensions:

- **Recency** (how recently a customer made a purchase)
- **Frequency** (how often a customer makes a purchase)
- **Monetary Value** (how much money a customer spends)

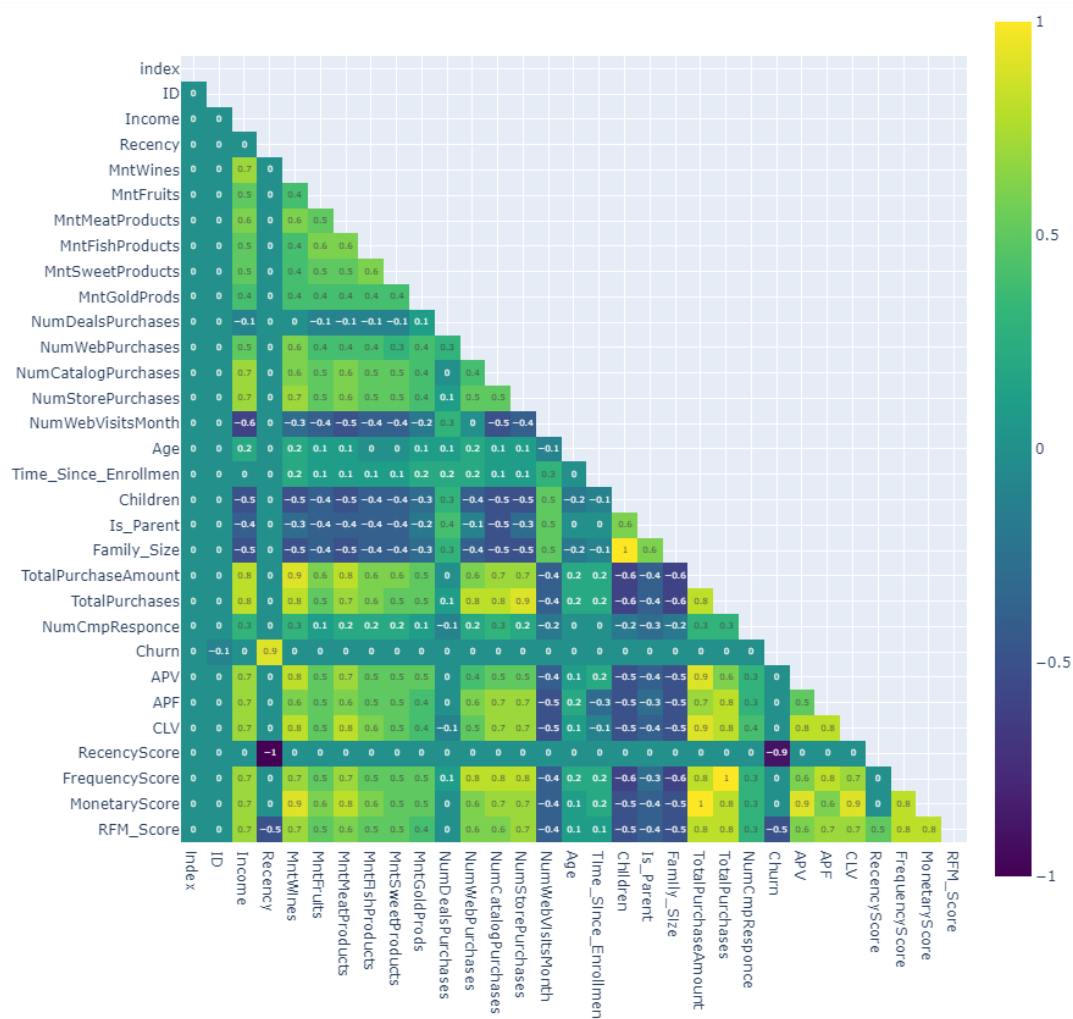




Comparison between different Customer's Recency, Frequency and Monetary Score



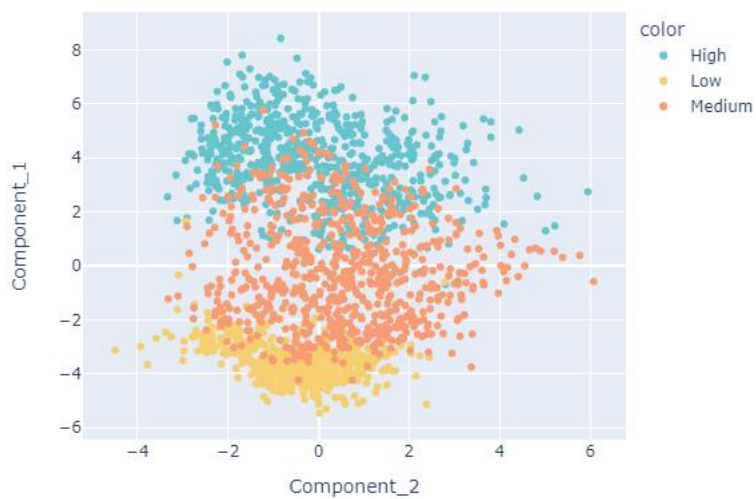
Correlation between distinctive features in the dataset



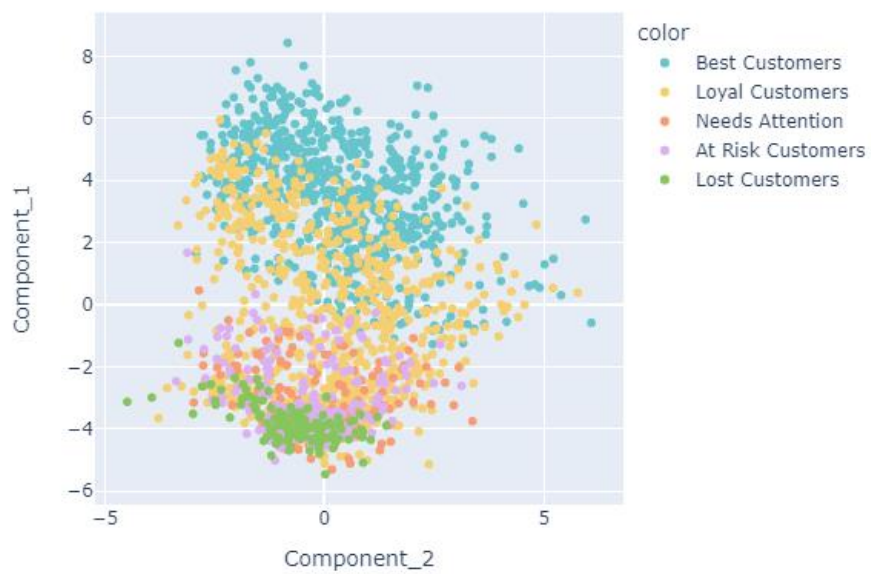
ii. Principal Component Analysis

Principal Component Analysis and Cumulative Explained Variance

PCA Components and CLV Analysis



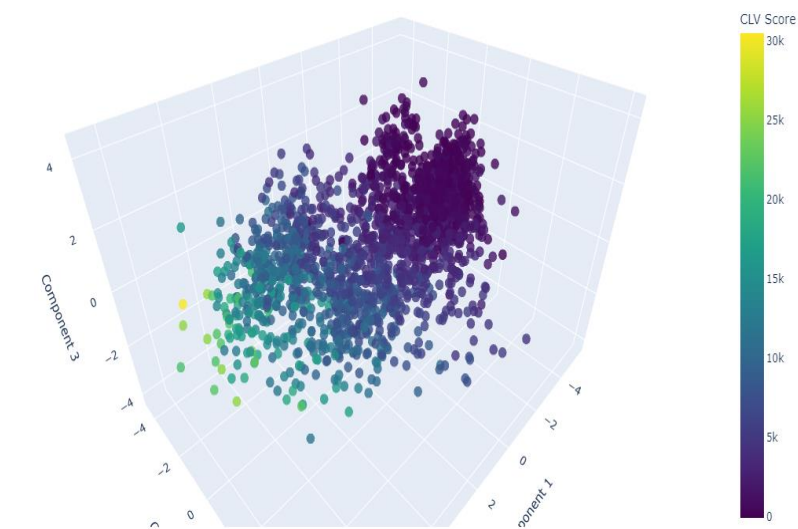
PCA Components and RFM Analysis



PCA in 3D Scatter Plot using RFM Value

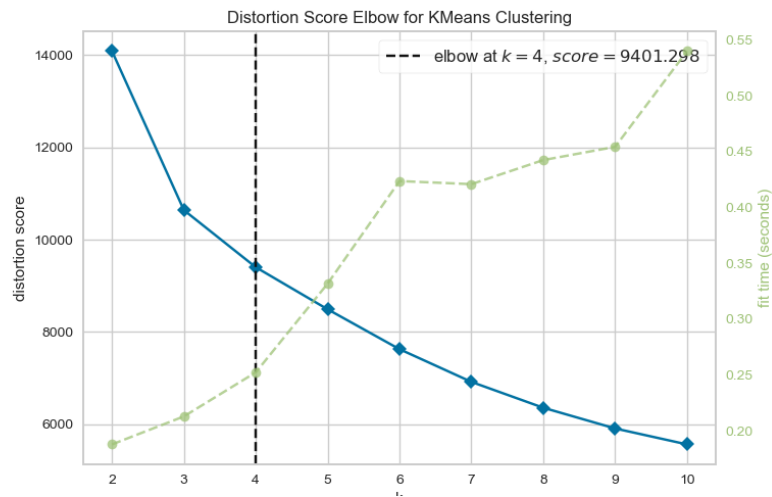


PCA in 3D Scatter Plot using CLV Score

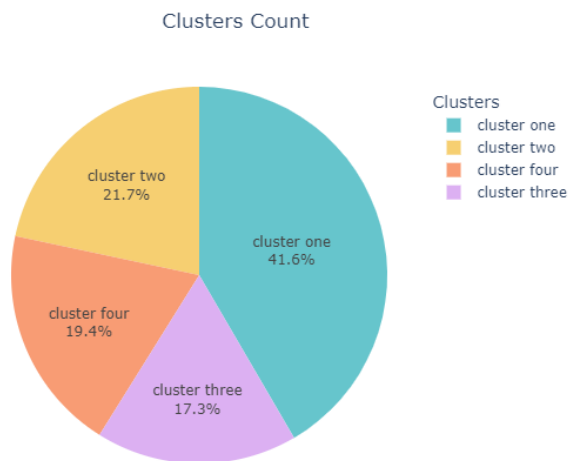


iii. KMeans Clustering

```
In [106]: # Create an instance of KElbowVisualizer with KMeans model (n_init="auto") and range of clusters (k=10)
plt_elb = KElbowVisualizer(KMeans(init = 'k-means++', random_state = 42, n_init=10), k=10)
# Fit the KMeans model to the cluster_data and perform the elbow method to find the optimal number of clusters
plt_elb.fit(cluster_data)
# Display the elbow plot which helps to determine the optimal number of clusters
plt_elb.show();
```

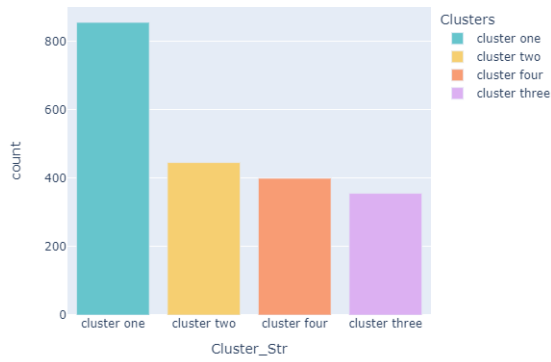


Cluster Analysis

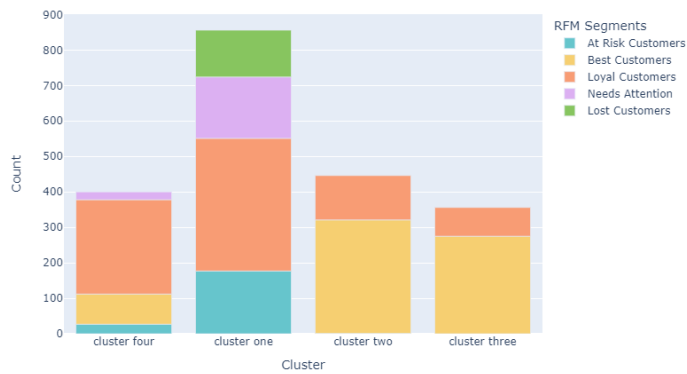


Clusters Count

Clusters Count



Age VS Income by Clusters



Conclusion

The Customer Personality Analysis project has provided valuable insights into the characteristics, preferences, and behavior of the company's ideal customers. Through the application of advanced techniques such as Principal Component Analysis (PCA) for dimensionality reduction and K-Means clustering for customer segmentation, the project has successfully identified distinct customer segments with unique personality traits and preferences. This segmentation has enabled the business to better understand its customer base and tailor its products, services, and marketing strategies to meet the specific needs of each segment. The project has laid a solid foundation for more targeted and effective customer engagement, leading to improved customer satisfaction and business performance.

In conclusion, the Customer Personality Analysis project has demonstrated the power of data-driven insights in understanding and adapting to the needs of a diverse customer base. By leveraging advanced analytical techniques, the project has paved the way for more personalized and impactful customer experiences, driving customer satisfaction, loyalty, and business success. As the project continues to evolve, there is significant potential for further exploration of additional customer attributes and behaviors to enhance the depth and accuracy of the customer segmentation, contributing to the long-term growth and sustainability of the business.

References

1. Uzila, A. (2022). K-means Clustering and Principal Component Analysis in 10 Minutes. Towards Data Science. [Online]. Available: <https://towardsdatascience.com/k-means-clustering-and-principal-component-analysis-in-10-minutes-2c5b69c36b6b>
2. Ding, C., & He, X. (2023). K-means Clustering via Principal Component Analysis. [PDF]. Available: <https://ranger.uta.edu/~chqding/papers/KmeansPCA1.pdf>
3. 365 Data Science. (2021). How to Combine PCA and K-means Clustering in Python? [Online]. Available: <https://365datascience.com/tutorials/python-tutorials/pca-k-means/>
4. implilearn. (2023). Data Visualization in Python: Overview, Libraries & Graphs. [Online]. Available: <https://www.simplilearn.com/tutorials/python-tutorial/data-visualization-in-python>
5. Towards Data Science. (2020). Data Visualization with Python. [Online]. Available: <https://towardsdatascience.com/data-visualization-with-python-8bc988e44f22>
6. MachineLearningMastery.com. (2022). Data Visualization in Python with matplotlib, Seaborn, and Bokeh. [Online]. Available: <https://machinelearningmastery.com/data-visualization-in-python-with-matplotlib-seaborn-and-bokeh/>