CUSTOMER

segmentation by

RFM & K-Means Clustering

analysis methods.

www.ahlam93.github

Python Project

Table Of Contents

www.ahlam93.github

Executive Summary :

The Capstone Project on Consumer Segmentation aims to explore customer purchasing behavior and its implications for marketing strategies. By analyzing factors influencing purchasing frequency, the project seeks to underscore the importance of targeted marketing campaigns. Through thorough research and analysis, actionable insights are uncovered to drive effective marketing strategies, ultimately enhancing customer engagement and maximizing business profitability. This project serves as a critical exploration of customer behavior dynamics, offering valuable insights to inform strategic decision-making in marketing and customer segmentation.

Key Points:

1. **Customer Segmentation using RFM Analysis:**

   Utilizing RFM (Recency, Frequency, Monetary) analysis, customers are segmented into distinct groups based on transactional behavior. Analysis of RFM scores identifies various customer segments such as "hibernating," "at-risk," and "loyal customers." This segmentation aids in understanding diverse customer needs and behaviors, enabling targeted marketing efforts.

2. **Predictive Modeling with K-Means Clustering:**

   Employing K-Means clustering predicts customer clusters based on demographic and transactional attributes. This predictive modeling identifies two primary customer clusters with distinct spending patterns and income levels, facilitating tailored marketing strategies and personalized customer experiences.

Executive Summary :

3. Insights and Recommendations:

- **Targeting Potential Loyalists:** Emphasize converting potential loyalists and loyal customers into champions through tailored campaigns and incentives, leveraging their high engagement for brand advocacy.
- **Understanding Spending Behavior:** Analyzing spending behavior reveals insights based on income, total expenses, and number of kids. Higher-income customers tend to spend more while spending patterns differ among customers with varying numbers of children.
- **Campaign Effectiveness:** Analyze campaign acceptance rates to refine and tailor promotions, enhancing effectiveness and driving higher participation rates among specific customer segments.

In summary, this project underscores the significance of data-driven insights in optimizing marketing strategies and fostering customer relationships. By utilizing segmentation techniques and predictive modeling, businesses can gain deeper insights into their customer base, identify growth opportunities, and implement targeted initiatives to drive engagement and loyalty, ultimately leading to sustainable business growth.

Market Search Questions

To be answered by analysis/

1. Demographic Analysis:
   - How do demographic factors like age, education, and income influence purchasing behavior?
2. Customer Behavior and Preferences:
   - How do customers' household composition (e.g., number of children at home, marital status) affect their purchasing preferences?
3. Customer Engagement and Response to Marketing Campaigns:
   - How effective have previous marketing campaigns been in terms of customer response (e.g., acceptance of marketing offers)?
4. Customer Loyalty and Retention:
   - Are there any segments of customers that exhibit higher levels of loyalty compared to others?
   - What strategies can be implemented to improve customer retention and foster long-term relationships?

Introduction

Understanding customer behavior and preferences is crucial for businesses to tailor their marketing strategies effectively and maximize their return on investment. In today's competitive market landscape, companies rely heavily on data-driven insights to gain a competitive edge and optimize their marketing efforts. The provided dataset offers a wealth of information regarding demographic attributes, purchasing behavior, and responses to marketing campaigns, presenting a valuable opportunity for comprehensive analysis.

The primary objective of this research is to delve into the factors influencing customer purchasing behavior and their responses to marketing campaigns. By employing advanced analytical techniques such as RFM (Recency, Frequency, Monetary) analysis and K-means clustering, we aim to gain deeper insights into consumer preferences, engagement levels, and loyalty. RFM analysis allows us to segment customers based on their recency of purchase, frequency of purchases, and monetary value, providing a holistic view of their purchasing behavior. On the other hand, K-means clustering enables us to identify distinct customer segments with similar purchasing patterns, facilitating targeted marketing strategies and personalized communication.

The decision to focus on this research question stems from its significance in the contemporary business landscape. As competition intensifies and customer expectations evolve, businesses must continually adapt their marketing approaches to remain relevant and competitive. Leveraging data analytics, including RFM analysis and K-means clustering, allows organizations to gain deeper insights into customer behavior, enabling them to tailor their marketing efforts more effectively and allocate resources efficiently.

Furthermore, the availability of a rich dataset presents a unique opportunity to apply advanced analytical techniques to uncover hidden patterns and correlations within the data. By adopting a data-driven approach that incorporates RFM analysis and K-means clustering, we can extract actionable insights that go beyond surface-level observations, enabling businesses to make informed decisions and drive strategic growth.

In summary, this research question holds immense potential for leveraging sophisticated analysis techniques to address real-world marketing challenges. By analyzing the provided dataset comprehensively and employing RFM analysis and K-means clustering, we aim to uncover valuable insights that can guide businesses in optimizing their marketing strategies, enhancing customer engagement, and driving sustainable growth in today's dynamic marketplace.

METHODS

About the Dataset

The data source is from the Kaggle website and the author acknowledged that it’s provided by Dr. Omar Romero-Hernandez. And it consists of 2240 rows*29 columns before doing some cleaning. The distribution of columns is as follows:

Customers

- ID
- Year_Birth
- Education
- Marital_Status
- Income
- Kidhome
- Teenhome
- Dt_Customer
- Recency
- Complain

Places

- NumWebPurchases
- NumCatalogPurchases
- NumStorePurchas
- NumWebVisitsMonth

Products

- MntWines
- MntFruits
- MntMeatProducts
- MntFishProducts
- MntSweetProducts
- MntGoldProds

Promotions

- NumDealsPurchases
- AcceptedCmp1
- AcceptedCmp2
- AcceptedCmp3
- AcceptedCmp4
- AcceptedCmp5
- Response

Data cleaning summary :

- After Checking for missing data we found 24 missing values in the ‘Income’ column so we replaced it with its median.
- We calculated customer age from ‘Year_Birth’ and renamed the column to ‘Age’.
- We have some outliers in the ‘Age’ & ‘Income’ columns, so we set caps on ‘Age ’and ‘Income ’for outlier removal.
- We grouped the education column into two groups (postgraduate/undergraduate) to reduce the features and encoded them to numerical values (0=postgraduate & 1=undergraduate).
- We grouped the 'Marital_Status' column into two groups (single/relationship) to reduce the features and encoded them to numerical values (0=relationship & 1=single).
- The columns 'Z_CostContact' and 'Z_Revenue' have a single value, meaning they are meaningless, so We dropped them.
- To analyze the total number of products each customer has bought (Total_Expenses) we combined the 'MntWines', 'MntFruits', 'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts', 'MntGoldProds' columns into a single column.
- To analyze the total number of kids each customer has we combined the (Kidhome & Teenhome) into a single column. (Total_ kids_Number.).
- To analyze the total number of accepted Campaigns by each customer we combined 'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5', 'AcceptedCmp1', 'AcceptedCmp2' into a single column ('Total_Accepted_Campaigns').
- To analyze the total number of purchases by each customer (Total_Purchases_Number) we combined the 'NumDealsPurchases', 'NumWebPurchases',' NumCatalogPurchases', 'NumStorePurchases', 'NumWebVisitsMonth' into a single column.
- We convert the datatype of the 'Dt_Customer' column to Datetime which shows the day that a specific customer signed up with the business. To See which dates are the oldest and the newest on file.

DESCRIPTION AND JUSTIFICATION OF SELECTED METHOD FOR ANALYSIS

1. RFM segmentation Method

Recency, Frequency & Monetary

In the first method of analysis, we choose to make the analysis based on the most three significant metrics or features (Recency, Frequency & Monetary) ,Here's a description and justification of using RFM analysis for the provided dataset:

Description of RFM Analysis:

1. **Recency:** Refers to how recently a customer has made a purchase. In the context of the dataset, we can calculate recency by determining the number of days since each customer's last purchase. Customers who have made purchases more recently are likely to be more engaged and responsive to marketing efforts. Analyzing recency can help identify active and inactive customers and tailor marketing strategies accordingly.
2. **Frequency:** Measures how often a customer makes purchases. For this dataset, we can calculate frequency by determining the total number of purchases made by each customer over a specific period. Customers who make frequent purchases are often more loyal and valuable to the business. Analyzing frequency can help identify high-value customers and target them with personalized marketing campaigns to encourage repeat purchases.
3. **Monetary:** Monetary value represents the total amount of money spent by each customer on purchases. In the dataset, we can calculate monetary value by summing the total expenses of each customer. Customers who spend more money are typically more valuable to the business and warrant special attention. Analyzing monetary value can help identify high-spending customers and tailor marketing strategies to maximize revenue.

1. RFM segmentation Method

Recency, Frequency & Monetary

Justification for Using RFM Analysis:

1. **Segmentation:** RFM analysis allows for the segmentation of customers into distinct groups based on their purchasing behavior. This segmentation enables businesses to better understand their customers' needs and preferences and tailor marketing strategies accordingly.
2. **Personalization:** By segmenting customers based on RFM scores, businesses can personalize their marketing efforts to target specific customer segments more effectively. This personalized approach can lead to higher engagement and conversion rates.
3. **Retention and Loyalty:** RFM analysis helps identify loyal and high-value customers who are more likely to make repeat purchases. By focusing on retaining these customers through targeted marketing campaigns and loyalty programs, businesses can improve customer retention and foster long-term relationships.
4. **Optimization**: RFM analysis provides actionable insights that businesses can use to optimize their marketing efforts and allocate resources more efficiently. By focusing on high-value customer segments, businesses can maximize their return on investment and drive sustainable growth.

In summary, RFM analysis is a valuable method for analyzing customer behavior and optimizing marketing strategies. By leveraging the RFM framework, businesses can gain deeper insights into their customers' purchasing habits and preferences, leading to improved customer engagement, loyalty, and profitability.

2. K-Means Model (clustering):

In the second method of analysis, we choose to make the analysis based on clustering the whole features used In the dataset, Here's a description and justification of using the K-Means model for the provided dataset:

Description of K-Means Model:

The K-Means clustering algorithm is an unsupervised learning technique used to partition a dataset into K distinct, non-overlapping clusters. The algorithm works iteratively to assign each data point to the nearest cluster centroid and then updates the centroids based on the mean of all data points assigned to each cluster. This process continues until the centroids no longer change significantly, or a specified number of iterations is reached.

2. K-Means Model (clustering):

Justification for Using K-Means Model:

1. **Cluster Analysis:** K-Means clustering is effective for identifying inherent patterns and structures within the dataset. By grouping similar data points into clusters, businesses can gain insights into different customer segments based on their shared characteristics.
2. **Segmentation:** The K-Means model facilitates customer segmentation by dividing the dataset into distinct groups based on relevant features such as income, age, total expenses, and other demographic variables. This segmentation enables businesses to tailor marketing strategies to different customer segments, leading to more targeted and personalized campaigns.
3. **Identifying Customer Profiles:** K-Means clustering helps identify distinct customer profiles or personas within the dataset. By analyzing the characteristics of each cluster, businesses can gain a deeper understanding of their customer base and tailor products, services, and marketing messages to better meet their needs and preferences.
4. **Predictive Modeling:** Once clusters are identified, businesses can use the insights gained from K-Means clustering to develop predictive models for customer behavior. These models can help forecast future purchasing patterns, identify potential high-value customers, and optimize marketing strategies for better customer engagement and retention.
5. **Optimizing Resource Allocation:** K-Means clustering assists in optimizing resource allocation by identifying customer segments with the highest growth potential or value. By focusing resources and marketing efforts on these high-potential segments, businesses can maximize their return on investment and drive revenue growth.

2. K-Means Model (clustering):

In conclusion, the K-Means clustering algorithm is a useful tool for examining customer data and locating pertinent dataset segments and trends. Businesses can improve overall consumer satisfaction and profitability, make well-informed decisions, and create focused marketing campaigns by utilizing the insights offered by K-Means clustering.

RESULTS

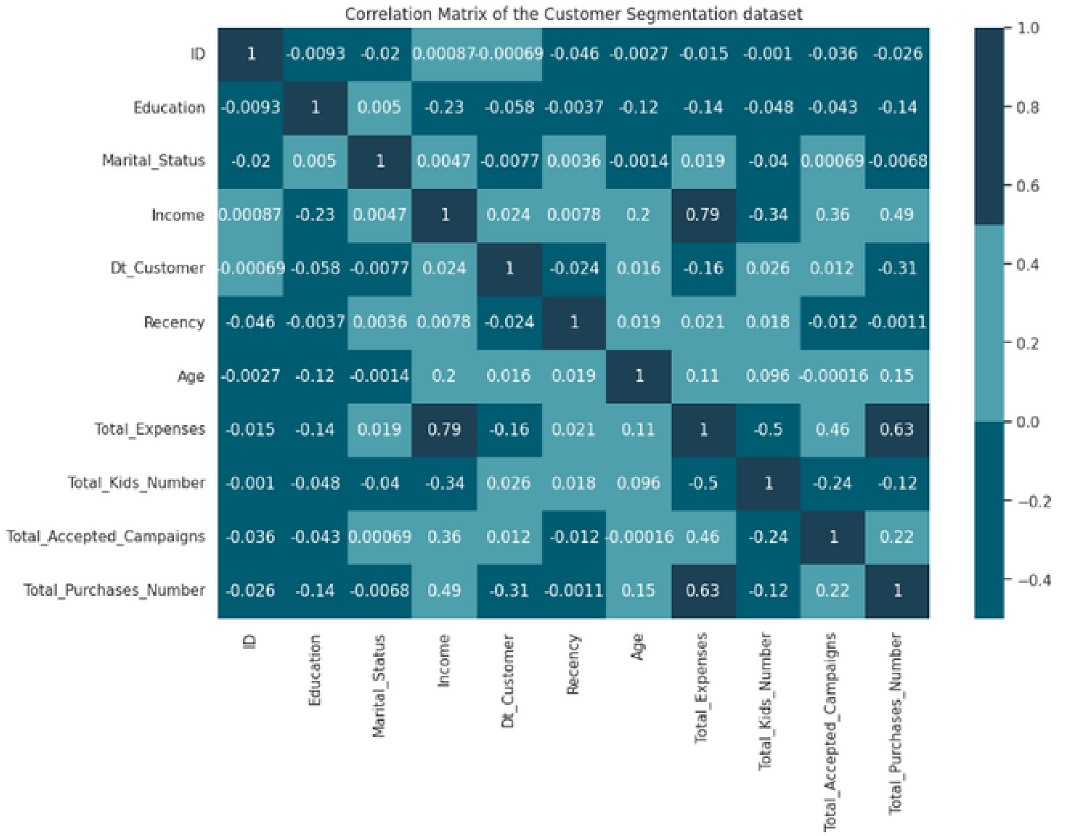
Descriptive Analysis & the most significant Graphs:

In this section, we present the results of our analysis on customer segmentation and predictive modeling. The analysis encompasses a comprehensive examination of customer purchasing behavior and its implications for marketing strategies. Through descriptive statistics, visualizations, and interpretation of results, we aim to provide insights into the diverse characteristics of customer segments and their corresponding behaviors. By leveraging segmentation techniques and predictive modeling, we uncover actionable insights that can inform strategic decision-making in marketing and customer engagement initiatives. The results presented herein offer valuable insights into customer preferences, spending patterns, and campaign effectiveness, facilitating a deeper understanding of the dynamics driving consumer behavior in the marketplace.

Descriptive Analysis:

The correlation Matrix provides insights into the strength and direction of relationships between variables. so in our provided dataset, we note that the most positively correlated features with 'Total_Expenses' in descending order are:

- Income.
- Total_Purchases_Number.
- Total _Accepted_Campaigns.

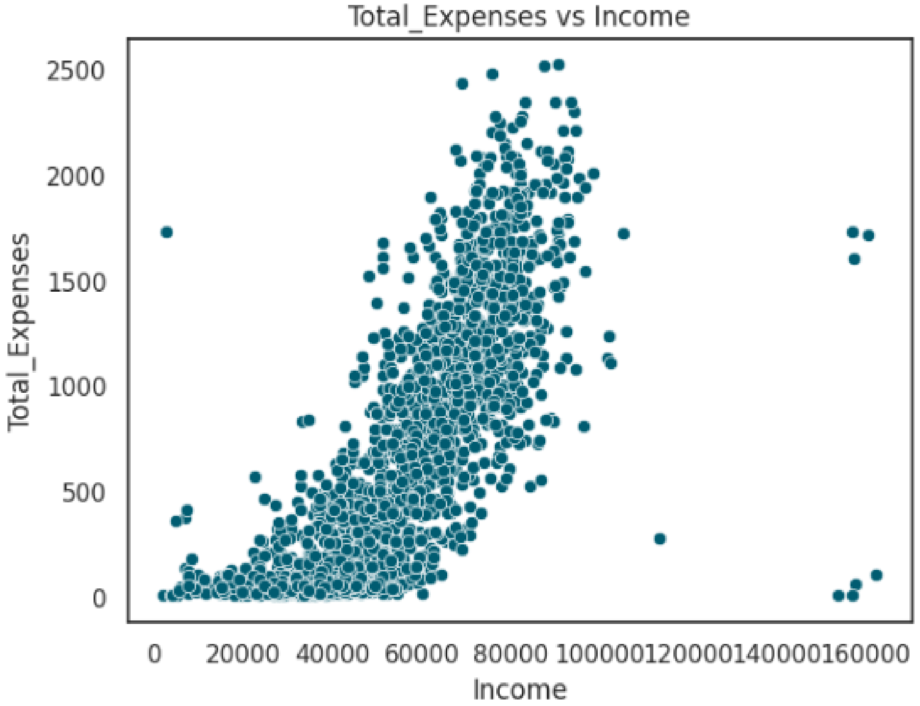


We note that the most positive correlated feature is Income with Total_Expenses and Total_Purchases_Number.

www.ahlam93.github

Descriptive Analysis:

This scatter plot illustrates the relationship between Income and Total _Expenses in our provided dataset.

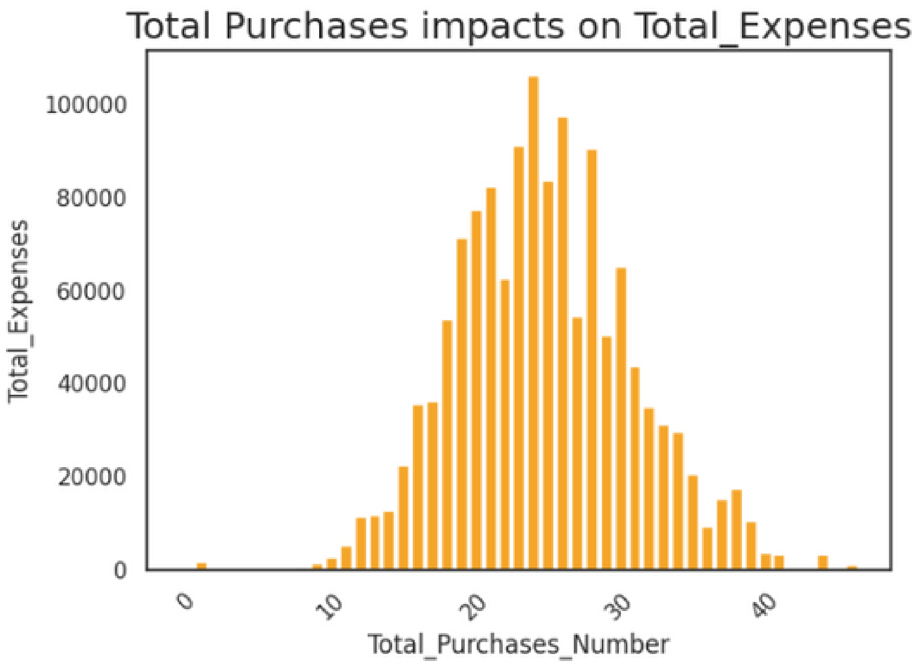



```
Correlation between Income and Total_Expenses: 0.7893749146446831
```

From the chart we that as income increases, Total_Expenses also tend to increase, so the strength of the correlation indicates that changes in income are closely associated with changes in Total_Expenses.

Descriptive Analysis:

This Bar charts provide a quantitative representation of the relationship between Total_Purchases_Number and Total_Expenses. The length of each bar accurately represents the magnitude of Total Expenses associated with each category of Total Purchases No.

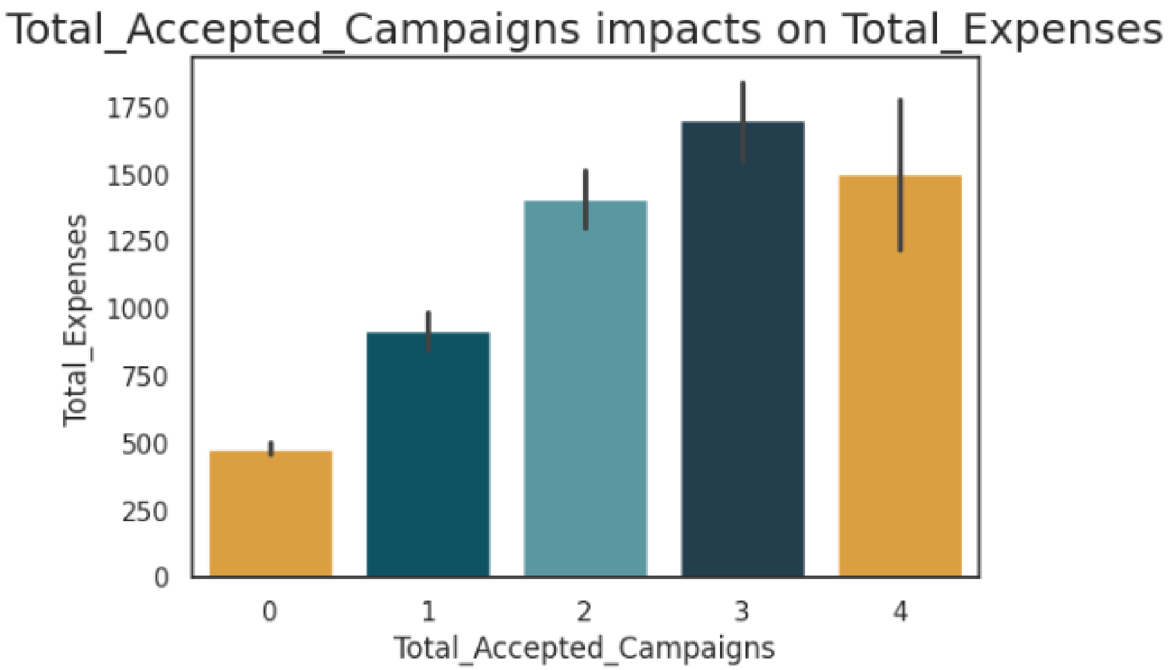


Correlation between Total_Purchases_Number and Total_Expenses: 0.6279254476829537

We note that, a correlation coefficient of approximately 0.63 suggests a moderately strong positive correlation between the total number of purchases and Total_Expenses. This means that as the total number of purchases made by customers increases, their Total_Expenses also tend to increase, indicating a relationship where customers who make more purchases also tend to spend more money overall.

Descriptive Analysis:

This bar plots can reveal patterns or trends in the relationship between Total_Accepted_Campaigns and Total_Expenses.



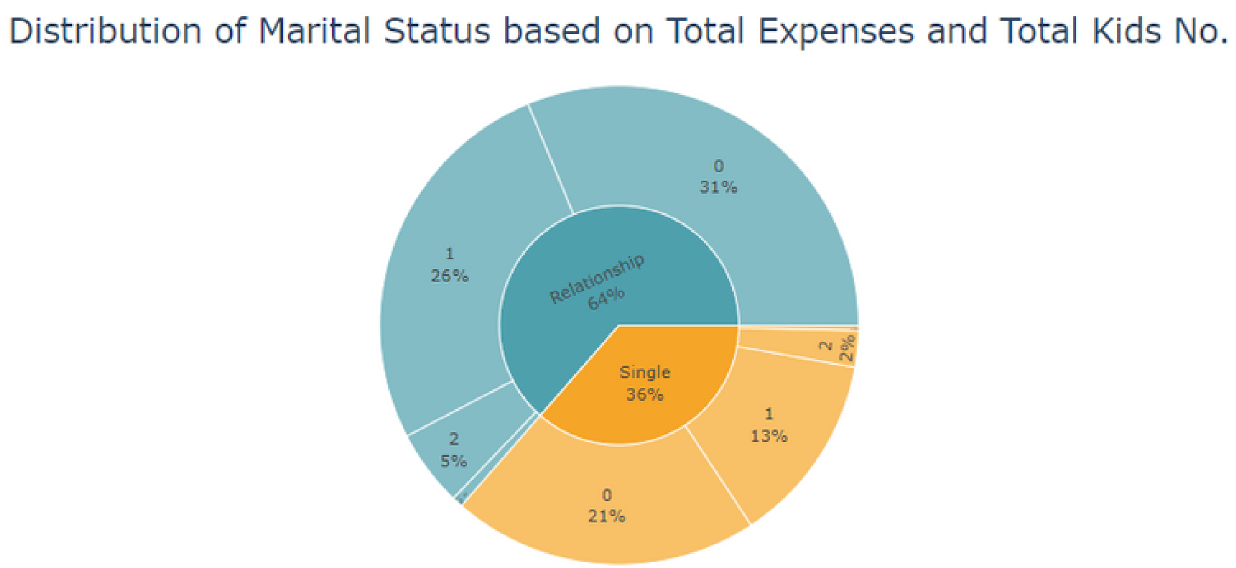
Correlation between Total_Accepted_Campaigns and Total_Expenses: 0.45892496164956675

So we note that Customers who accepted 3 campaigns have more Total expense rate.

www.ahlam93.github

Descriptive Analysis:

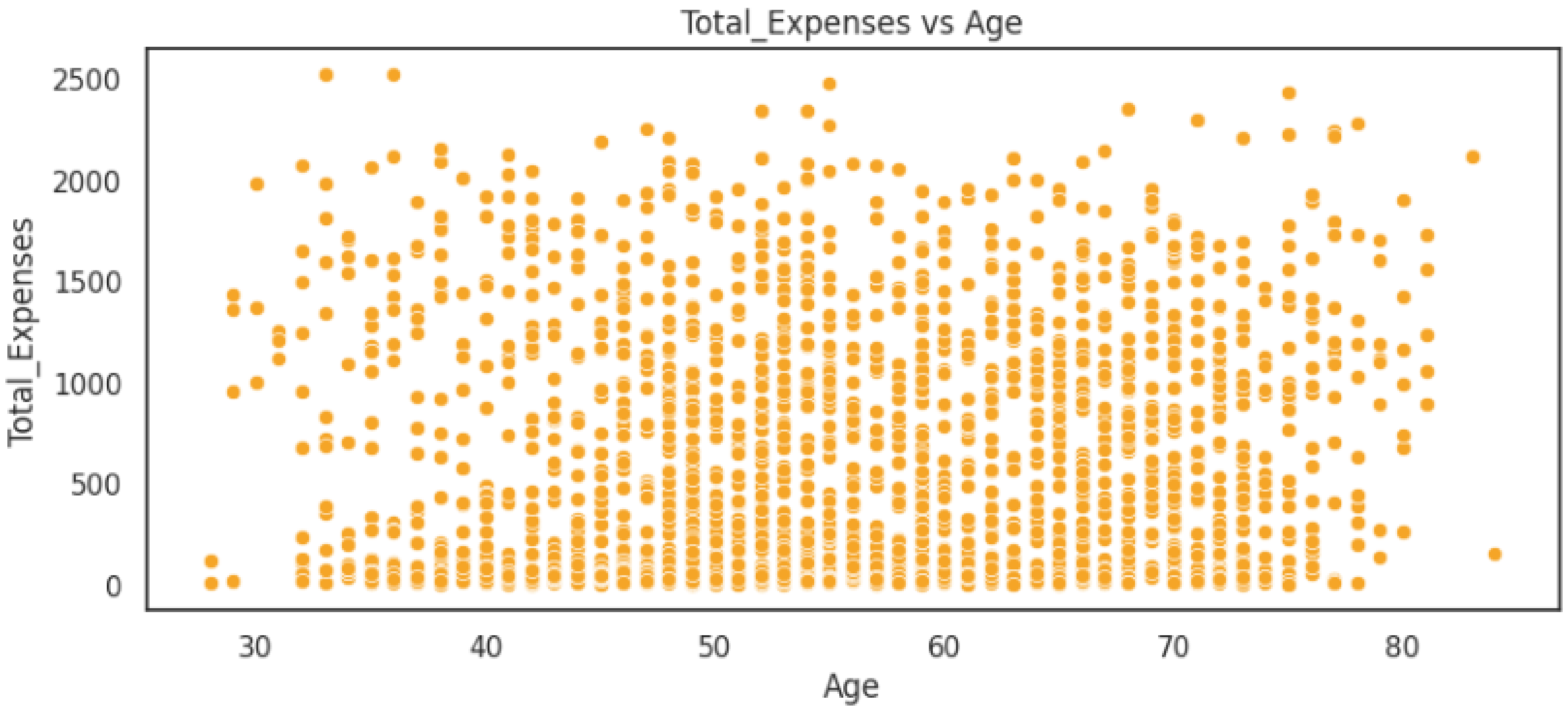
Features that might have little impact on total expenses for each customer such as Material_statuse & number of kids:



From the graph We note that customers in relationships with 0 number of kid have more Total Expense rate.

Descriptive Analysis:

Features that might have little impact on total expenses for each customer such as Age:




```
Correlation between Age and Total_Expenses: 0.11361767213727979
```

We note that there is a slight tendency for Total_Expenses to increase slightly as age increases, or vice versa. However, the correlation is weak, meaning that age and Total_Expenses are not strongly related.

www.ahlam93.github

RFM EXPLINATION

RFM Calculation:

To apply RFM analysis to our provided dataset we group the data by 'ID' of each customer and aggregate RFM metrics as follows:

1. **Recency:** Calculate the number of days since each customer's last purchase (using the "Recency" column) already given in the dataset so no calculation is needed.
2. **Frequency:** Count the total number of purchases made by each customer within the given timeframe (using the "Total_Purchases_Number" column).
3. **Monetary:** Determine the total amount spent by each customer on purchases (using the "Total_Expenses" column).

- We Check if there is a customer with 0 expenses to remove them because it will not add any value to our analysis.

```
rfm = rfm[rfm['Monetary(Expenses)']>0]
```

	Monetary(Expenses)	Recency	Frequency(Purchases_Num)
ID			
0	1198	66	18
1	577	0	23
9	120	86	19
13	32	57	12
17	1028	81	33

RFM Scoring:

To apply RFM scoring to the dataset provided we do the following steps:

1. **Recency:** Calculate the Recency score based on quantiles (5 bins) of Recency values. Assign a score based on this value, with higher scores indicating more recent purchases.
2. **Frequency:** Calculate Frequency score based on quantiles (5 bins) of rank of Frequency value. Assign a score based on this frequency, with a higher score indicating more frequent purchases.
3. **Monetary:** Calculate the Monetary score based on quantiles (5 bins) of Monetary value. Assign a score based on this monetary value, with higher scores indicating higher spending.

ID	Monetary(Expenses)	Recency	Frequency(Purchases_Num)	recency_score	frequency_score	monetary_score
8082	34	75	12	2	1	1
1951	839	92	36	1	5	4
9733	80	52	11	3	1	2
255	21	31	12	4	1	1
3524	167	9	19	5	3	2
10648	1289	78	22	2	4	5
6173	969	46	27	3	5	4
4931	1730	13	29	5	5	5
5221	879	93	34	1	5	4
3870	1006	28	28	4	5	4

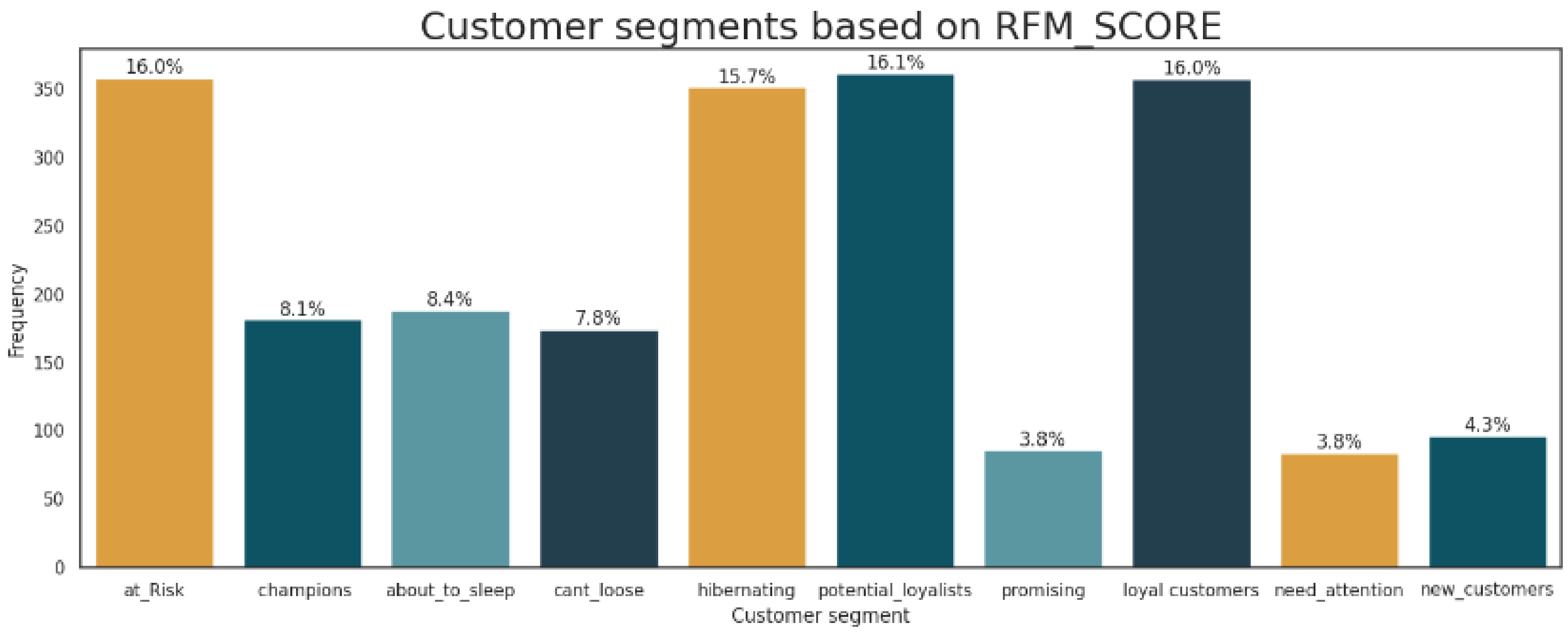
RFM Scoring:

Now we'll focus on 2 features Recency and Frequency to evaluate customer loyalty by concatenating their scores So, a single score will be created for each customer that captures both the recency and frequency of their purchases to be segmented into different groups based on their RFM scores. For example, customers with a high RFM_Score would be considered 'champions' whereas customers with a low RFM_Score would be considered as 'hibernating'.
We choose to ignore Monetary because Customers may purchase items at different price points, which can skew their monetary value without necessarily indicating loyalty. For example, a customer who makes a few high-value purchases may have a higher monetary value than a customer who makes frequent but lower-value purchases. Ignoring monetary value in such cases allows for a more balanced assessment of loyalty based on recency and frequency of purchases.

ID	Monetary(Expenses)	Recency	Frequency(Purchases_Num)	recency_score	frequency_score	monetary_score	RFM_SCORE	segment
0	1198	66	18	2	3	5	23	at_Risk
1	577	0	23	5	4	3	54	champions
9	120	86	19	1	3	2	13	at_Risk
13	32	57	12	3	1	1	31	about_to_sleep
17	1028	81	33	1	5	4	15	cant_loose
20	183	91	12	1	1	2	11	hibernating
22	309	99	22	1	3	3	13	at_Risk
24	47	96	19	1	3	1	13	at_Risk
25	1115	9	25	5	4	4	54	champions
35	210	35	16	4	2	3	42	potential_loyalists

RFM Segments:

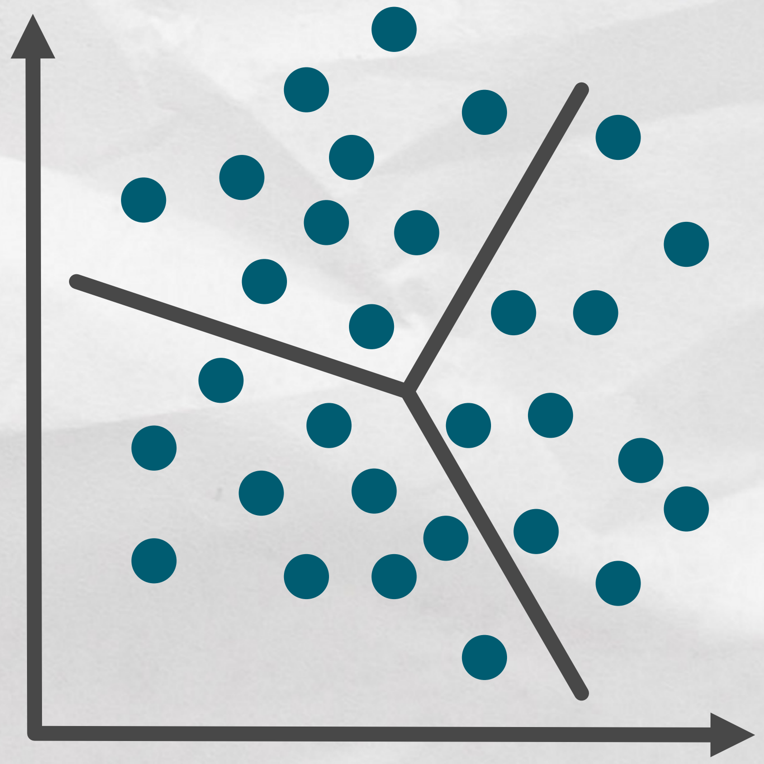
The countplot provides a clear visual representation of the distribution of customer segments, making it easy to understand the composition of the customer base according to RFM_SCORE.



Based on the chart we conclude that we have to pay big attention for potential_loyalists & loyal customers to convert them to champions by tailed special campaigns and offers for them.

www.ahlam93.github

K-MEANS EXPLINATION



K-means Explination:

The K-means clustering algorithm is a popular unsupervised machine learning technique used for clustering similar data points together based on their features. Here's how the K-means model can be applied to the provided data:

- **Data Preprocessing:**

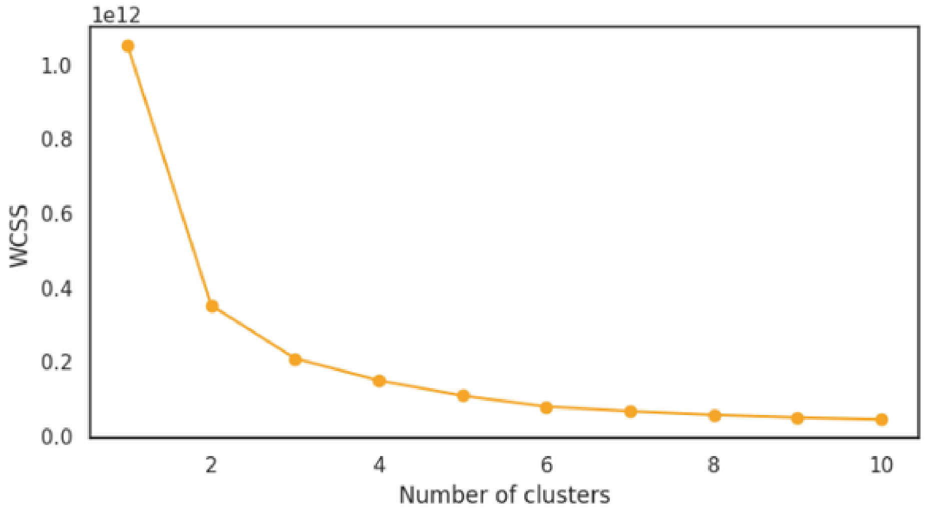
In the cleaning data step we applied essential preprocessing of the data such as (Check for missing values & encoding categorical variables such as Education).

- **Feature Selection:**

Identify the features that we want to include in the clustering analysis. In this dataset, features like Education, Marital_Status, Income, Recency, Age, Total _Expenses, Total_Kids_Number, Total _Accepted_Campaigns, and Total_Purchases_Number can be considered for clustering.

- **Choosing the Number of Clusters (K):**

The ***K*** parameter represents the number of clusters to form. we used the elbow method to determine the optimal number of clusters for the dataset.



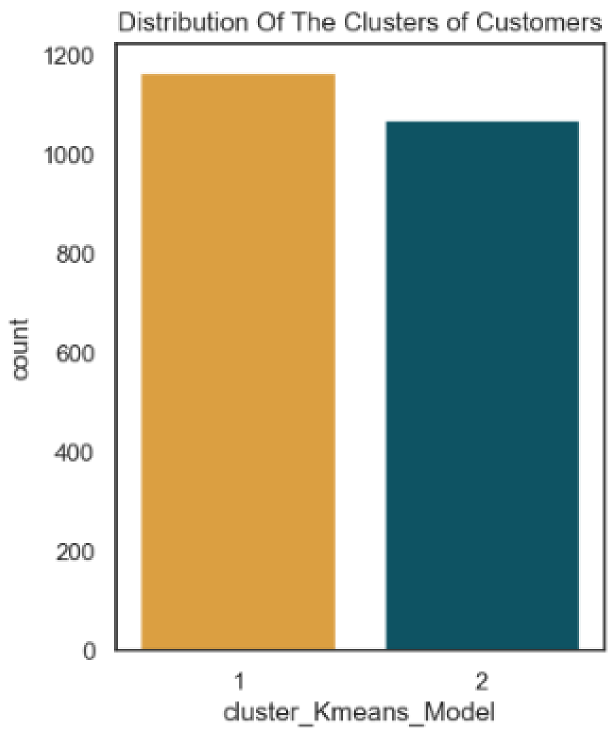
From the chart we conclude that the best number of *K* is 2

Applying K-means:

Steps for applying K-means:

- Training a K-Means Algorithm.
- Initializing and fitting a K-Means model with 2 clusters to the dataset.
- Predicting cluster labels for data points using the trained KMeans model.
- Appending the predicted cluster labels to the main data frame .

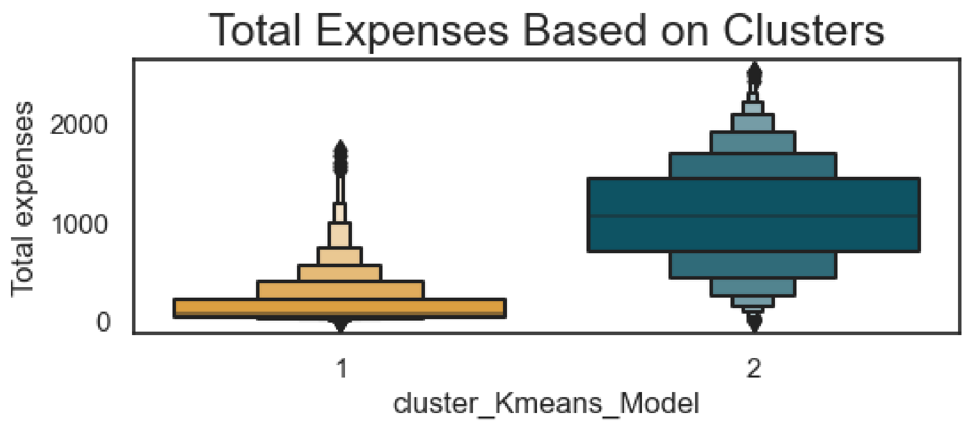
	ID	Education	Marital_Status	Income	Recency	Age	Total_Expenses	Total_Kids_Number	Total_Accepted_Campaigns	Total_Purchases_Number	cluster_Kmeans_Model
0	5524	0	1	58138.0	58	67	1617	0	0	32	2
1	2174	0	1	46344.0	38	70	27	2	0	11	1
2	4141	0	0	71613.0	26	59	776	0	0	25	2
3	6182	0	0	26646.0	26	40	53	1	0	14	1
4	5324	0	0	58293.0	94	43	422	1	0	24	2



www.ahlam93.github

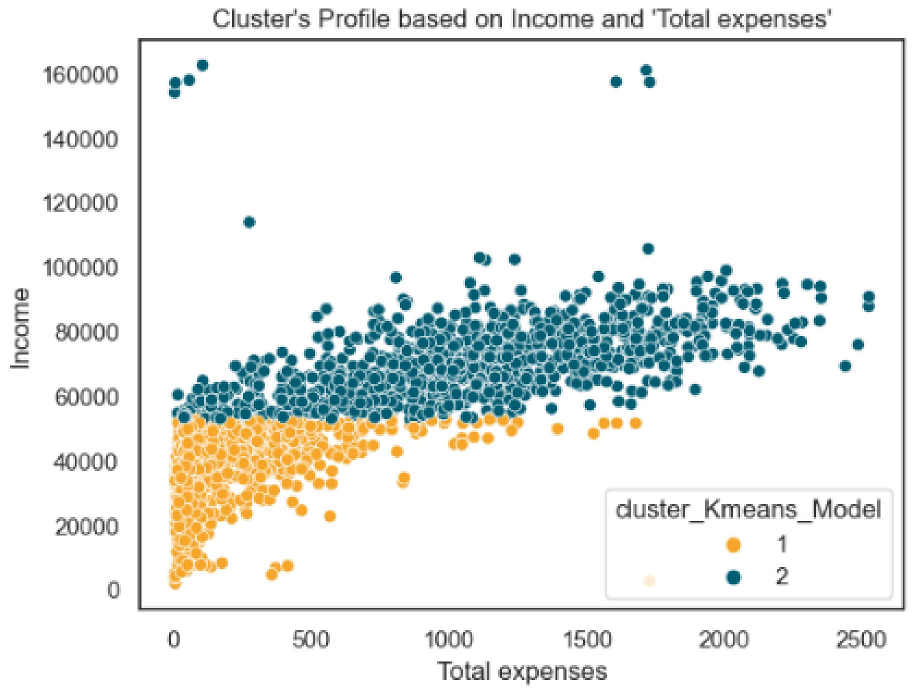
Applying K-means:

Considering that this clustering is unsupervised. We lack a labeled feature with which to assess or grade our model. This section aims to analyze the patterns in the clusters that were created and identify the types of patterns in the clusters.
In order to do that, we will use exploratory data analysis to examine the data in light of clusters and make inferences as follows:



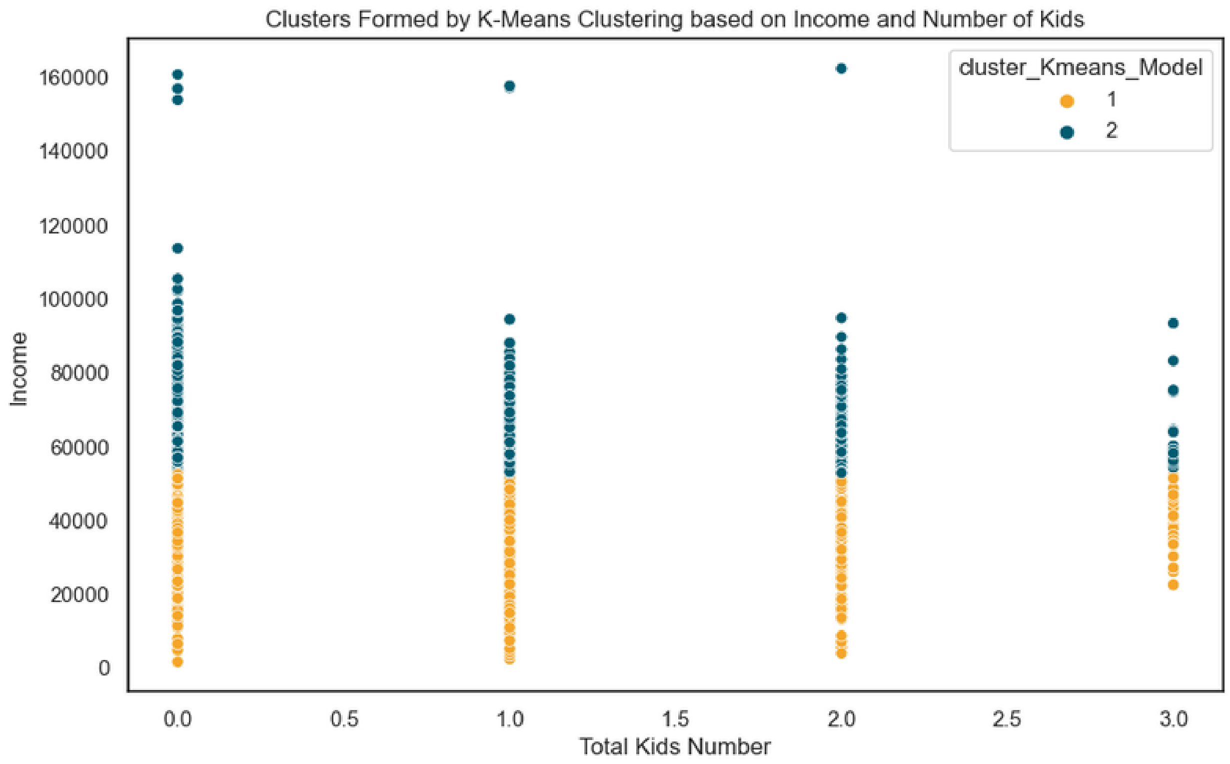
We conclude that Cluster 2 has High Expenses whereas Cluster 1 has low Expenses.

Applying K-means:



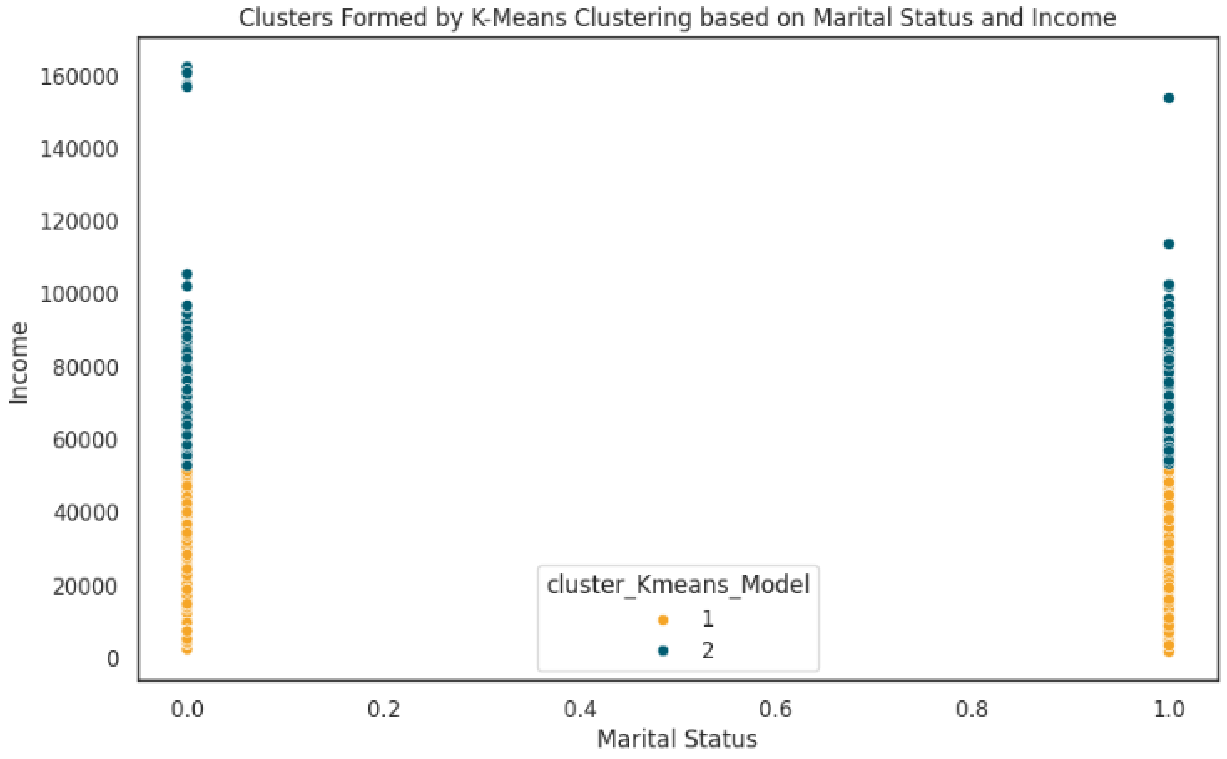
From the chart we conclude Cluster 2 has high-income rate and expends more than cluster 1.

Applying K-means:



People who have more kids fall into cluster 1 and they expend less than other and the People who having less number of kids falling in cluster 2 which they spnend more.

Applying K-means:



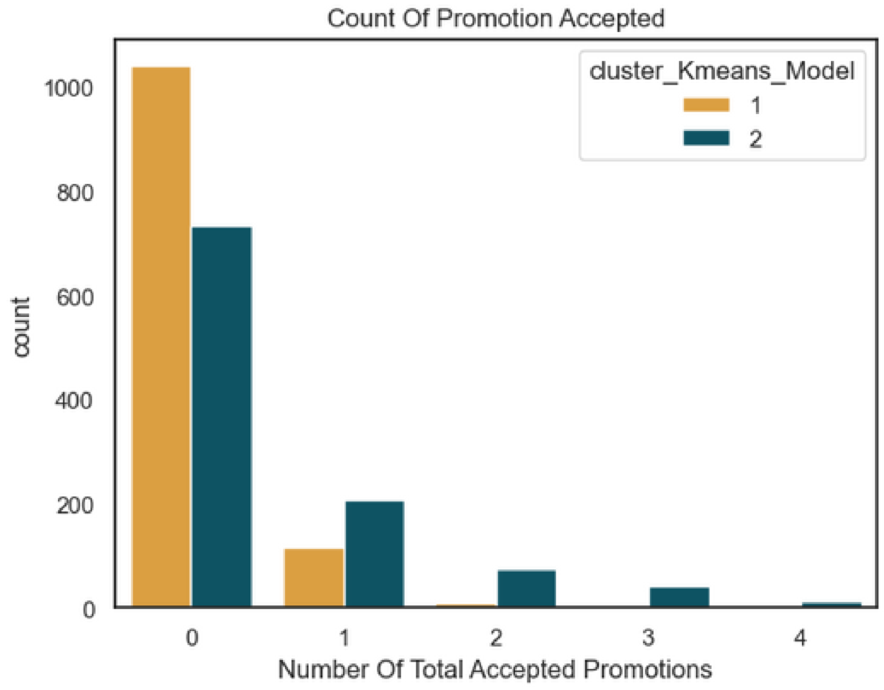
Both clusters have almost the same number of different relationships which means no impact on the result.

Applying K-means:

We conclude that people having less income falling in cluster 1 whrease people with high income falling in cluster 2.

www.ahlam93.github

Applying K-means:



We observe that the campaigns have not yet received a resounding response. Overall, very few people took part. Furthermore, nobody participates in all five of them. Perhaps to increase sales, more carefully thought out and targeted promotions are needed.

Conclusion of K-Means model:

Cluster 1 :

- Customers with less Total _Expenses.
- Customers who are married and parents of more than three children.
- Customers with modest income.

Cluster 2 :

- Customers with more Total_Expenses.
- Customers who are single or parents who have less than three kids.
- Customers with high income.
- Although age is not a criterion, it is somewhat recognized that older individuals belong to this category.

As a result, customers in cluster 2 tend to spend more money, thus businesses might target this group to sell their products.

RFM vs. K-means

Both RFM (Recency, Frequency, Monetary) analysis and K-means clustering are widely used methods for segmentation, but they differ in their approaches and applications.

RFM:

- **Advantages:**
  - Simple and intuitive method.
  - Focuses on specific behavioral aspects related to purchasing.
  - Provides actionable insights for targeted marketing campaigns and retention strategies.
- **Limitations:**
  - May not capture all dimensions of customer behavior.
  - Assumes equal importance of recency, frequency, and monetary value.
  - Does not consider interactions between variables.

K-MENAS:

- **Advantages:**
  - Flexible and adaptable to various types of data.
  - Captures complex relationships and interactions between variables.
  - Automatically identifies natural groupings in the data.
- **Limitations:**
  - Requires a predefined number of clusters (K).
  - Sensitive to initial centroid selection and outliers
  - Interpretation of clusters may be challenging without domain knowledge.

RFM vs. K-means

- **Complexity:** RFM analysis is more straightforward compared to K-means clustering, making it suitable for quick insights and actionable strategies. In contrast, K-means clustering is more complex but offers greater flexibility and sophistication in capturing nuanced customer segments.
- **Interpretability:** RFM analysis provides clear and interpretable segments based on recency, frequency, and monetary value scores. K-means clustering, on the other hand, may produce clusters that are not immediately interpretable without further analysis or domain knowledge.
- **Data Requirements:** RFM analysis requires data on customer transactions and interactions, focusing primarily on purchasing behavior. K-means clustering can accommodate a broader range of data types and variables, including demographic, behavioral, and psychographic attributes.
- **Application:** RFM analysis is commonly used for customer retention, loyalty programs, and targeted marketing campaigns. K-means clustering is suitable for exploratory analysis, customer segmentation, and identifying hidden patterns or trends in the data.

In summary, both RFM analysis and K-means clustering are valuable methods for customer segmentation, each with its strengths and limitations. The choice between the two depends on the specific objectives, data availability, and desired level of sophistication in segmentation analysis.

The Comparison Conclusion

To sum up, RFM answers a significant market search question related to customer loyalty and retention.
However, K-means is a more comprehensive method for customer segmentation due to it could use the whole features in the dataset to examine customer behaviors so we chose it to build future decisions that related to the oriented market campaigns.

www.ahlam93.github

CONCLUSION

Summary of the work we did on the dataset:

Data Cleaning and EDA (Exploratory Data Analysis):

- Loaded the dataset and checked its basic information.
- Conducted data cleaning by handling missing values.
- Examined the distribution of various features like ‘Age’, ‘Income’, ‘Total_Expenses’, etc., using visualizations such as scatter plots and barplot.
- Grouped and encoded categorical variables like ‘Education’ level and ‘Marital_Status’.
- Derived new features like‘Total_Expenses’, ‘Total_Kids_Number’, ‘Total_Purchases_Number’, and ‘Total_Accepted_Campaigns’.
- Explored the distribution of ‘Total_Purchases_Number’ and its impact on ‘Total_Expenses’.

Bivariate & Descriptive Analysis:

- Using the correlation matrix to provide insights into the strength and direction of relationships between variables.
- Explored correlations between pairs of features using scatter plots and calculated correlation coefficients such as ‘Income’ vs. ‘Total_Expenses’, ‘Age’ vs. ‘Total_Expenses’, etc
- Examined the impact of the Total_Kids_Number’, and ‘Marital_Status’ on ‘Total_Expenses’.
- Visualize the impact of the ‘Total_Accepted_Campaigns’ on ‘Total_Expenses’.

Summary of the work we did on the dataset:

RFM Analysis (Recency, Frequency, Monetary):

- Grouped customers by their purchasing behavior and calculated RFM metrics.
- Assigned RFM scores and segments to each customer based on their RFM scores.
- Visualized the distribution of customer segments based on RFM scores.

Predictive Analysis using K-Means Clustering:

- Applied K-Means clustering algorithm to perform customer segmentation.
- Determined the optimal number of clusters using the Elbow method.
- Predicted cluster labels for the customers and visualized the distribution of clusters.

www.ahlam93.github

Recommendations:

Some recommendations for the next steps based on the result of analysis:

1. **Tailored Marketing Campaigns:** Design targeted campaigns based on customer segments, focusing on needs and behaviors.
2. **Customer Acquisition:** Targeting promising segments with promotions to convert them into loyal customers.
3. **Customer Experience Enhancement:** Improve overall experience to increase satisfaction and loyalty.
4. **Data Collection and Analysis:** Gathering and analyzing customer data for timely strategy adjustments.
5. **Feedback and Monitoring:** Collecting feedback to address issues promptly and maintain satisfaction.
6. **Customer Lifecycle Management:** Tailoring communication strategies to each stage of the customer journey.
7. **Alignment with Business Goals:** Ensuring initiatives align with business objectives and KPIs for effective evaluation.

THANK YOU

www.ahlam93.github