

Customer Segmentation

Based on Kaggle Dataset: Customer Personality Analysis

MSiA 421 Final Project

Group 9: Ziqiao Ao, Xin Shu, Ke Xu, Yuwen Meng

Literature Review

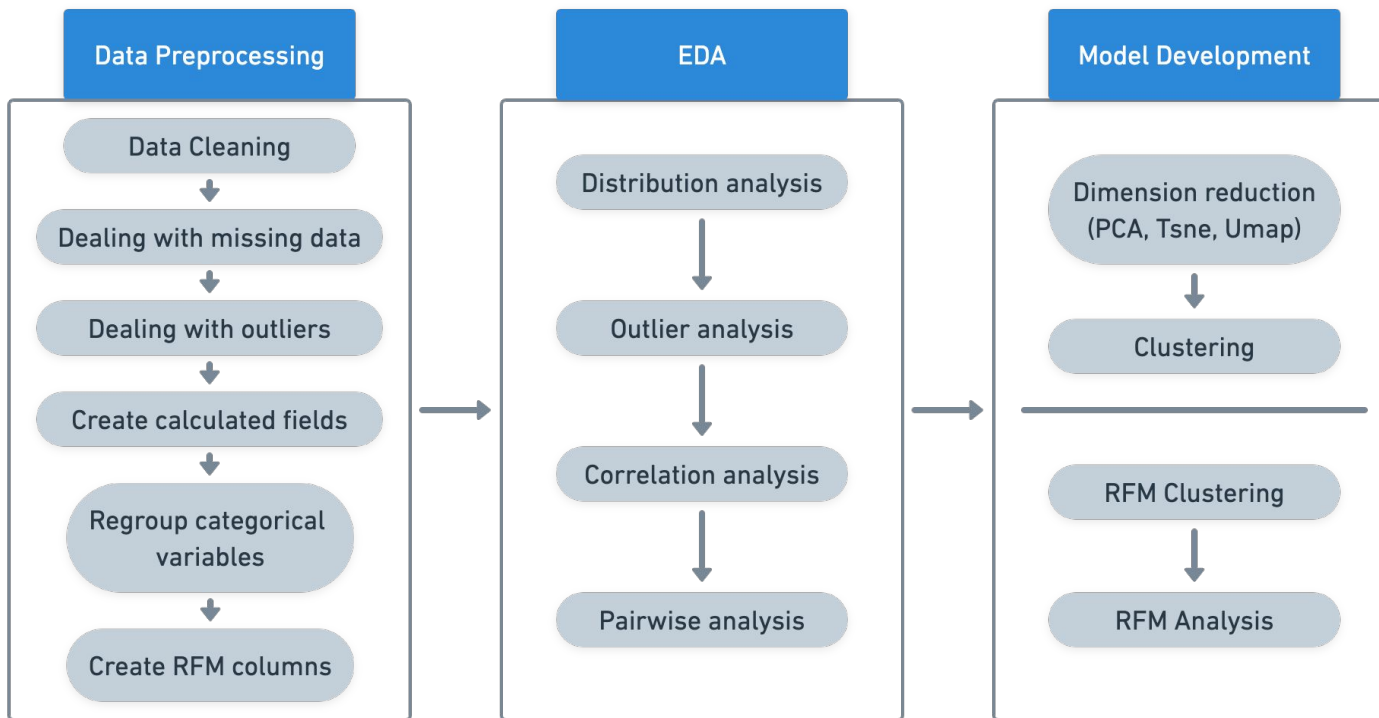
Customer Values

- RFM on sports industry (Dogan, Onur, et al. 2018)
- RFM on online retail with decision tree analysis (Chen, Daqing, et al. 2012)
- Customer LTV model on wireless communication (Kim, Su-Yeon, et al. 2006)

All Features

- Testing clustering algorithms on Banking (Zakrzewska, D., and J. Murlewski 2005)
- PCA techniques with K-means and hierarchical clustering (Azad Abdulhafedh 2021)

Model Approach



Data Preprocessing

01

Demographic Features

Age, Income, Education, Marital Status, Parent

RFM

Recency

Monetary: MntWines, MntFruits, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProds

Frequency: NumWebPurchases, NumCatalogPurchases, NumStorePurchases

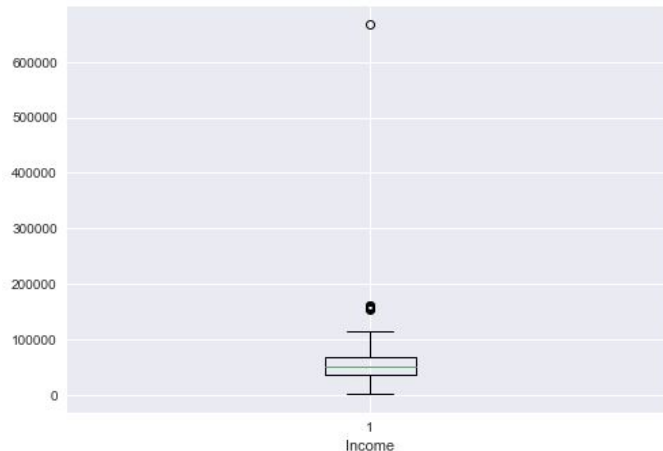
Others

Complains, Response to campaigns, Dt_Customer

02

Remove Outliers

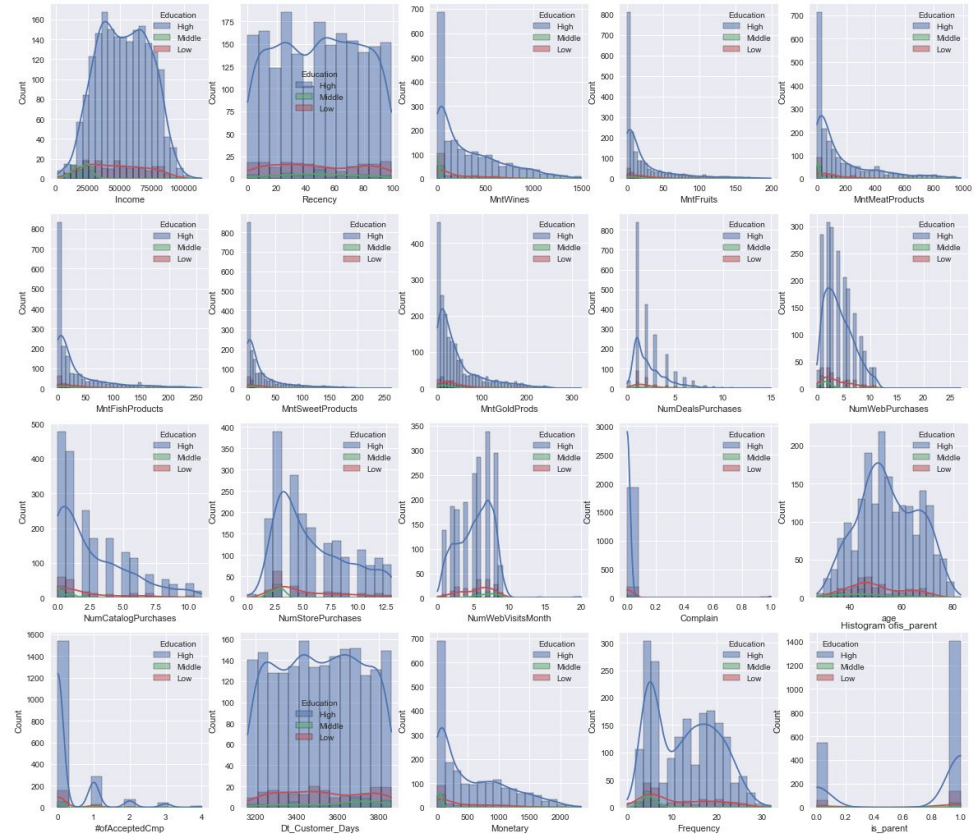
Income, Age, and MntMeatProducts



Exploratory Data Analysis

Distribution Analysis

- The distributions are largely right-skewed in the purchase monetary.
- The income and age have normal distribution while recency gets a relatively uniform distribution.
- Customers with higher education levels tend to have higher monetary, frequency, tenure, and income.



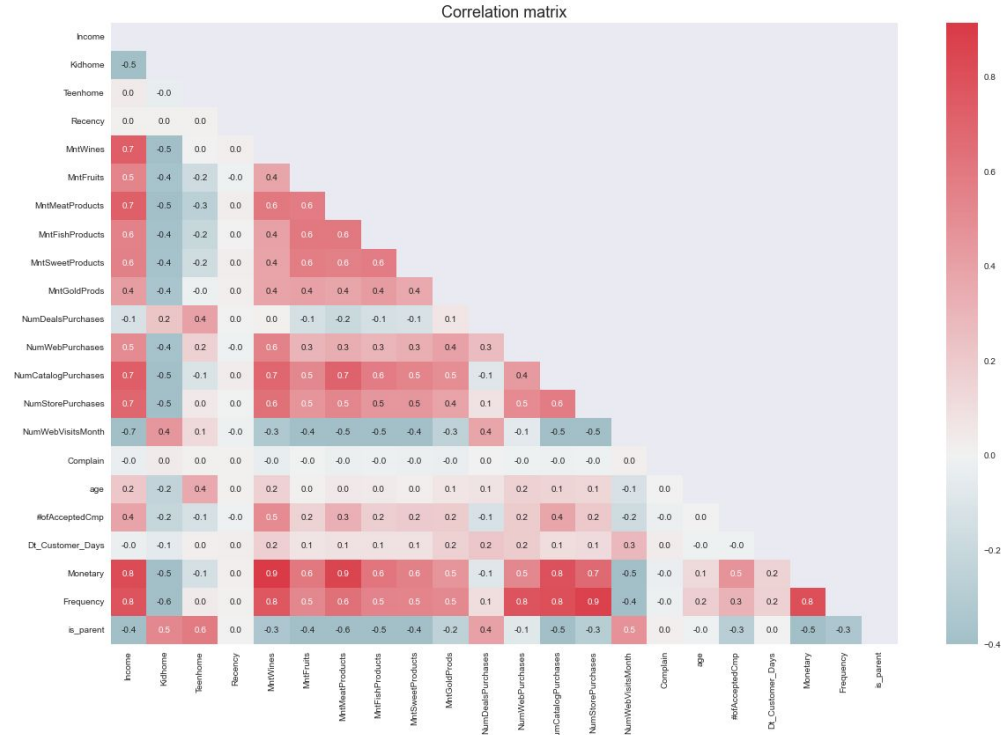
Exploratory Data Analysis

Correlation Analysis

- Most of the variables are positively related.
- The number of kids and teenagers at home, and whether the customer is a parent is negatively correlated with most of the other variables.

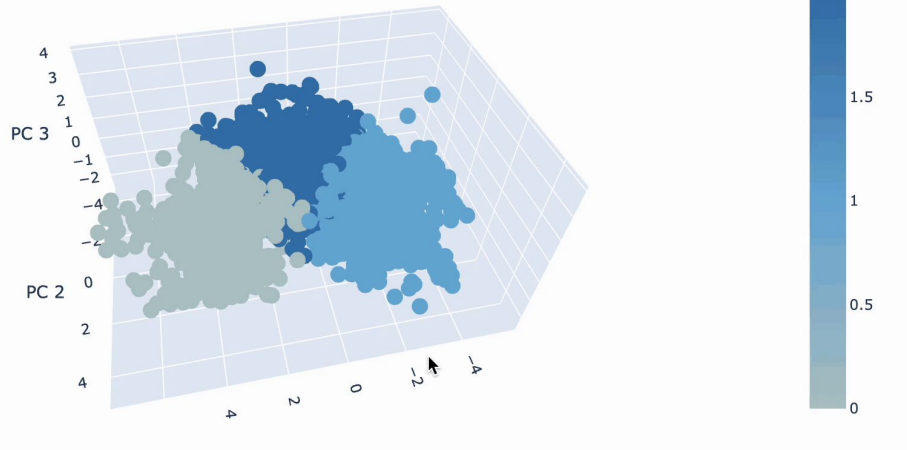
Pairwise Analysis

- As income increases, the monetary and frequency increase in general, this clear positive relationship also exists between monetary and frequency.
- Customers with higher education levels have relatively higher incomes and money spent.
- Single customers' money spent varies largely (a larger range compared with married customers).



Clustering Based on Dimensionality Reduction

The KMeans clustering algorithm with $K=3$ after PCA with 18 components produced the best clustering model for this analysis. According to their demographic features and purchasing behaviors, customers are grouped into 3 cluster.



Age

An average age of 55

Income

High income

Complaint Rate

Lowest complaint rate in the last year

Education

Middle education level (90% of people have high education)

Campaign Acceptance

The highest rate of accepting the company's campaign

Is parent

Most people did not have children.

Spending Habits

Spent the most in each product category in the last 2 years

Cluster 0



Marketing Strategy

Targeted with higher-end products and customized marketing campaigns that cater to their spending habits.



The infographic features a central teal circle with a white silhouette of a person's head and shoulders. The text 'Cluster 1' is written in bold black font across the silhouette. Six colored lines radiate from the circle to six surrounding text boxes: blue (top-left), orange (top-right), red (middle-right), pink (bottom-right), light blue (bottom-center), and yellow (bottom-left).

Cluster 1

Age

An average age of 50
(Youngest group)

Education Level

Lowest education level among
the three clusters (Only 83% of
people have high education)

Income

Lowest income among
three clusters

Campaign Acceptance

The lowest rate of accepting
the company's campaign

Is parent

Most people in this cluster
had kids at home

Spending Habits

Spend the least in each
product category in the last
2 years.

Marketing Strategies

It is recommended to focus on affordability and family-friendly products, along with incentives to encourage campaign acceptance.

01

Age

An average age of 58
(Oldest group)

02

Income

Middle Income

03

Education Level

Highest education level



Cluster 2

Is Parent

Most people have
teenagers at home

04

Single Rate

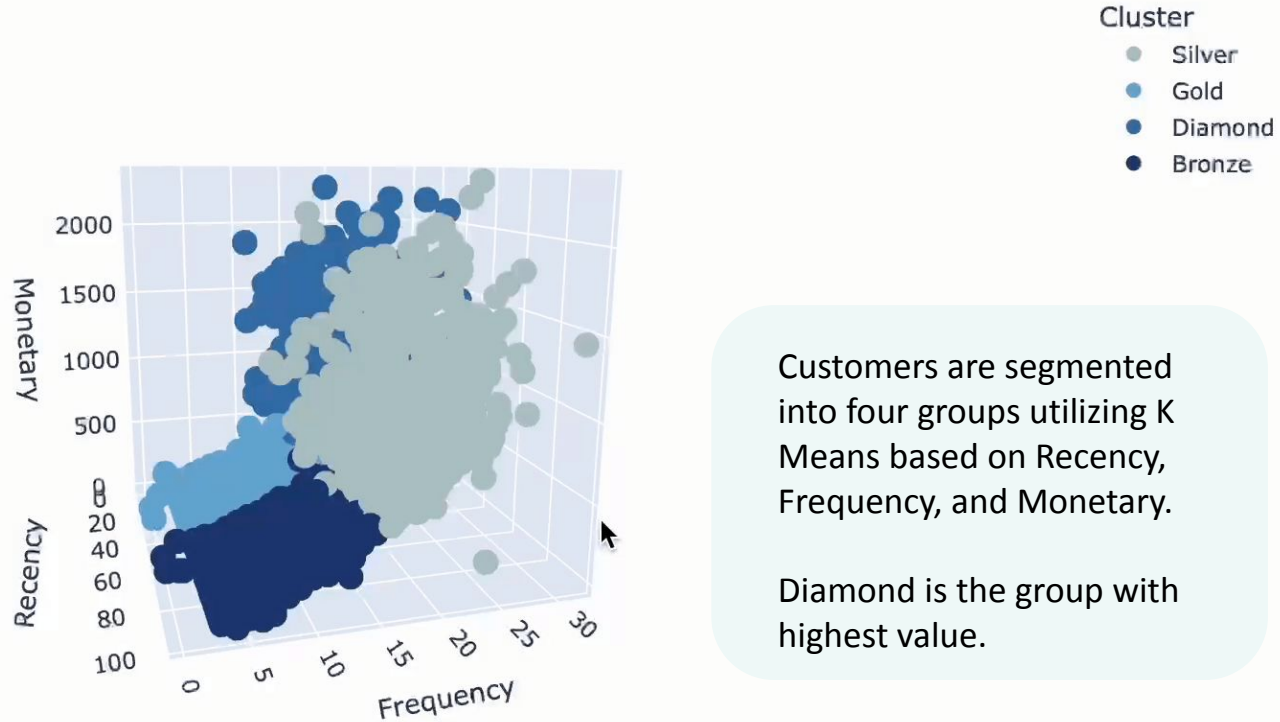
Highest single rate among
the three clusters.

05

Marketing Strategies

This cluster could be targeted with tech-savvy or cultural products that could increase their spending.

RFM Analysis



Customer infographics



Diamond

Low Recency,
High Frequency,
High Monetary,
High Income,
High Age,
Less likely to be parent,
Least complaints

Loyalty rewards or exclusive perks such as early access to new products or services;
Send personalized emails or messages to show appreciation



Gold

High Recency,
High Frequency,
High Monetary,
High Income,
High Age,
Highest Alone Rate,
No Middle Level Education

Targeted marketing campaigns to get them back; Tailor marketing messages appealing to customer alone



Silver

High Recency,
Low Frequency,
Low Monetary,
Low Income,
Low Age,
More likely to be parent

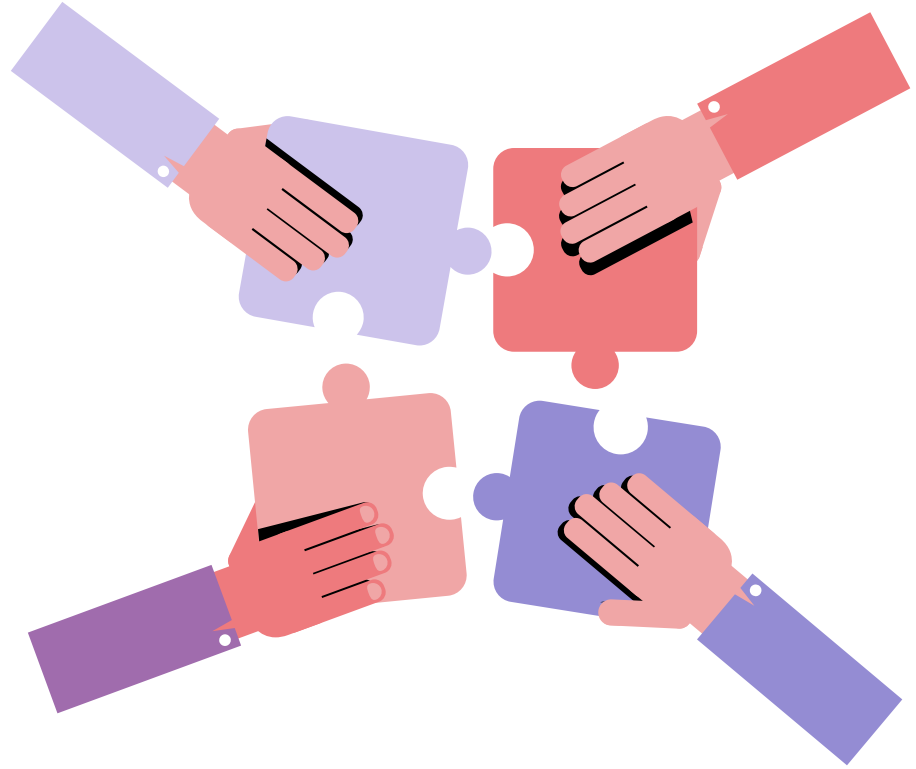
Personalized email or social media campaigns especially for parent; Personalized discounts or promotions to encourage them to make repeat purchases; Target middle level education customer only in these 2 groups with relevant promotions align with their interests or needs to motivate them to be higher value customers



Bronze

Low Recency,
Low Frequency,
Low Monetary,
Low Income,
Low Age,
More likely to be parent

Thank you!



Q&A