DS227: Business Analytics for Data Science

Final Project

Student: Viktoria Melkumyan

Instructor: Yevgenya Bazinyan

Customer Personality Analysis

Problem Statement

Business Question:

How can the company understand customer purchasing patterns to improve marketing ROI?

Goal:

Segment customers, personalize campaigns, and identify high-value customers.

Dataset Overview

29 variables

Key features:
Demographics,
Spending,
Campaign
Engagement,
Purchase Channels,
Recency

2,240 customers

Each customer appearing only once with his/her whole shopping history

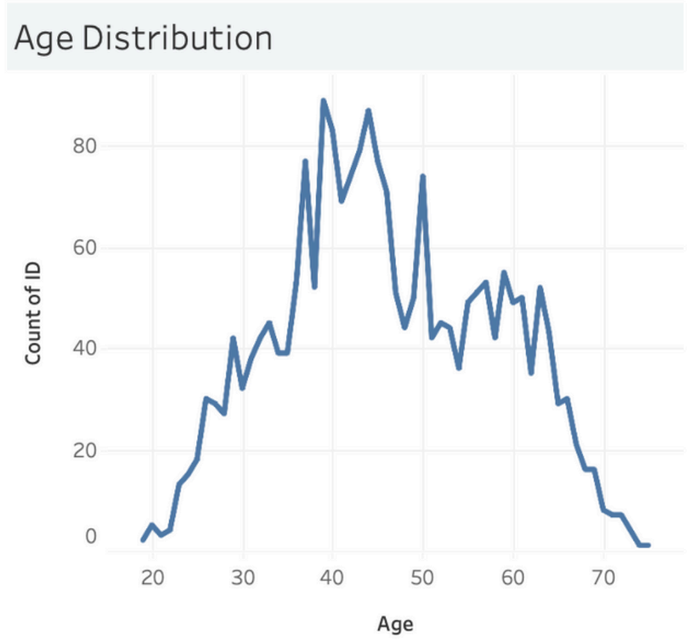
Data Preparation & Cleaning

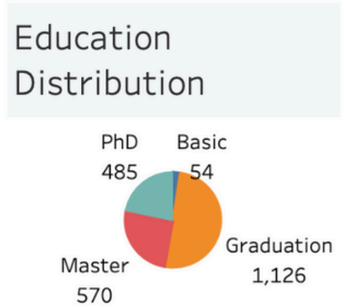
Handled missing values and outliers

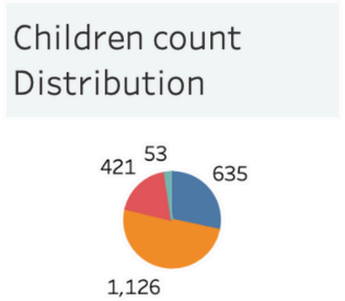
Feature engineering: Age, Total Spend, Total Children

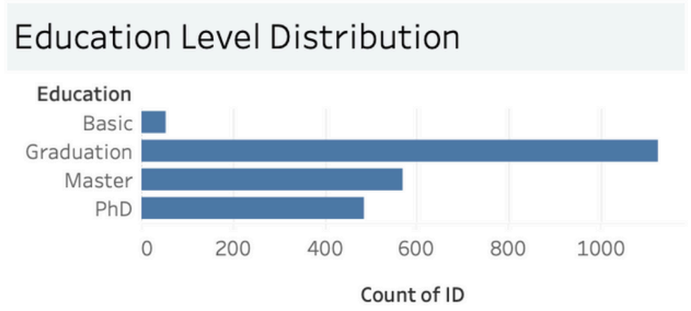
Expressed categorical variables as numerical: Education, Marital Status

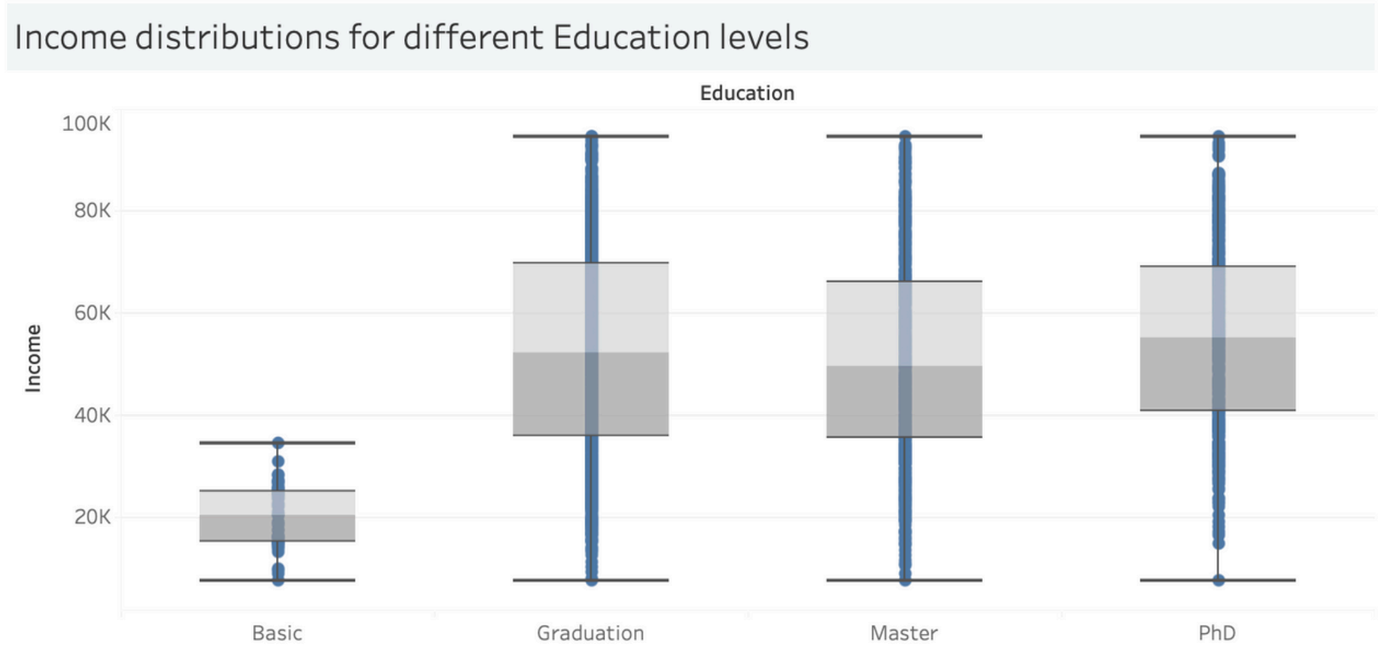
Descriptive Analysis: Demographics





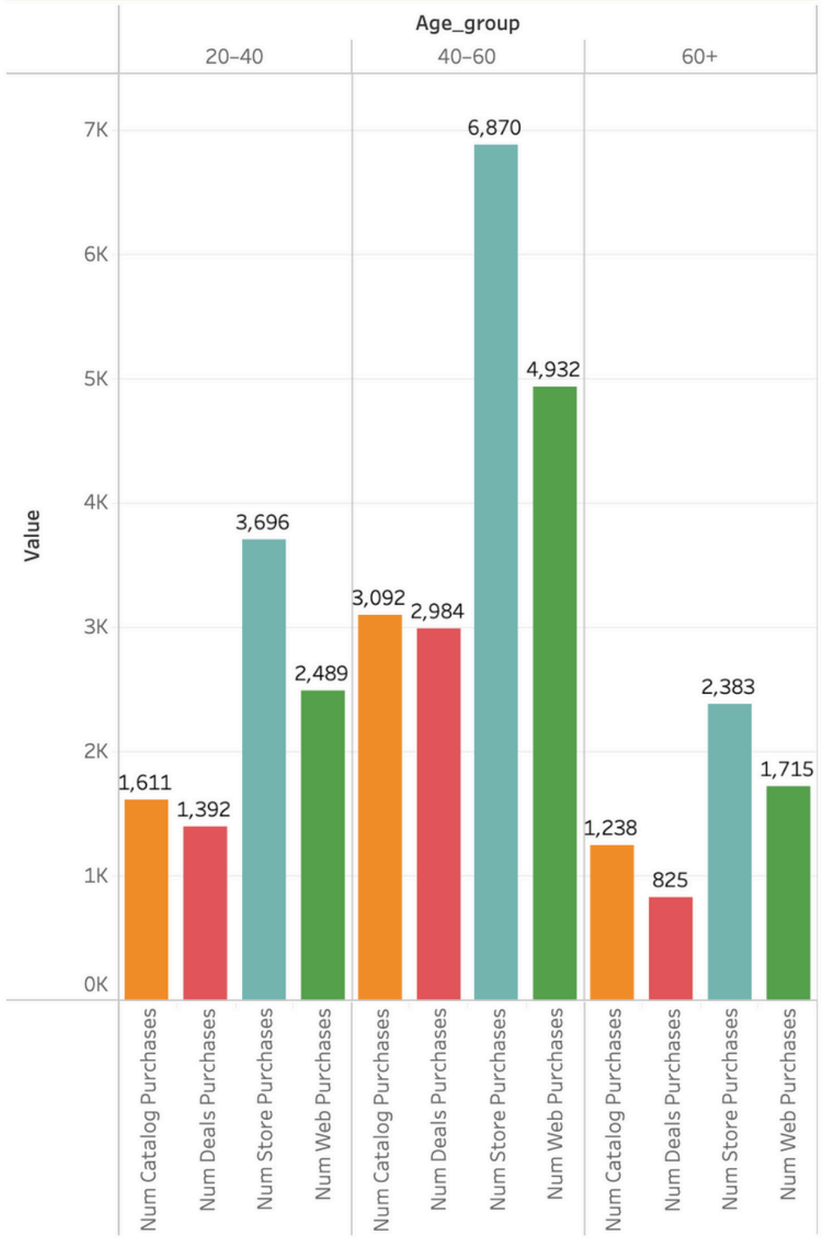




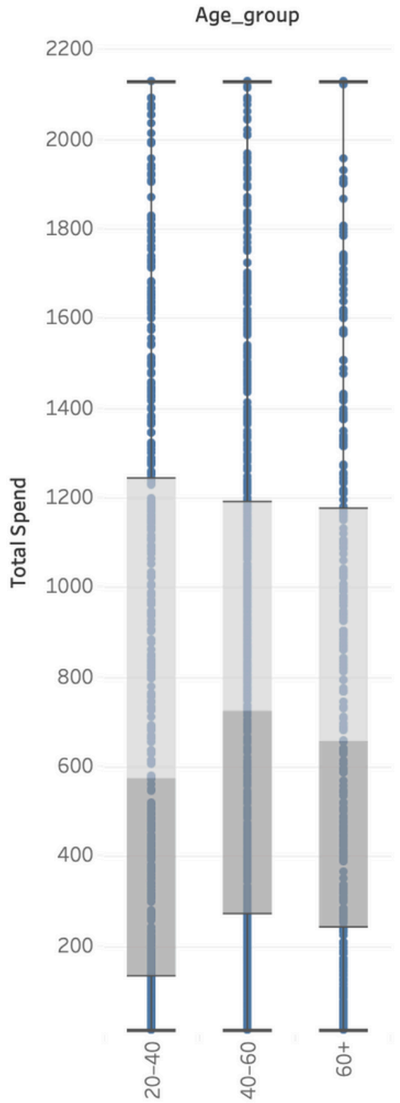


Descriptive Analysis: Spendings

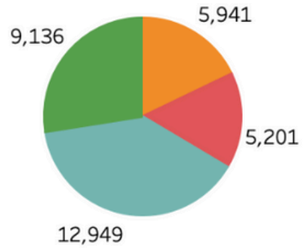
Channel Preference vs Age group

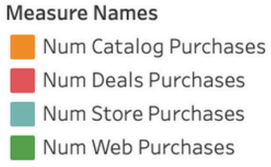


Total Spendings Distribution per Age Group

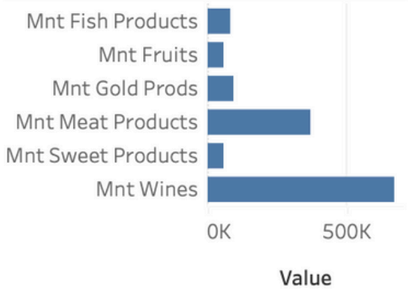


Channel Usage Distribution





Total Spendings per Category

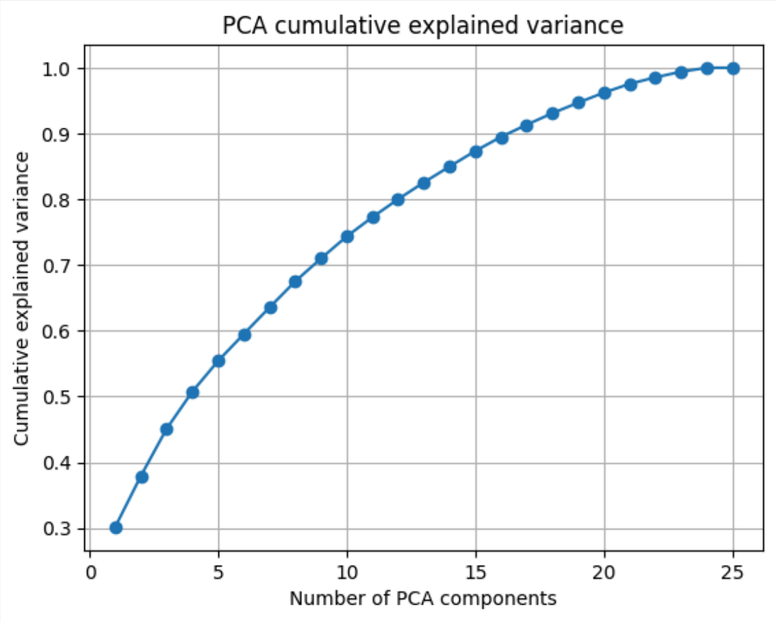


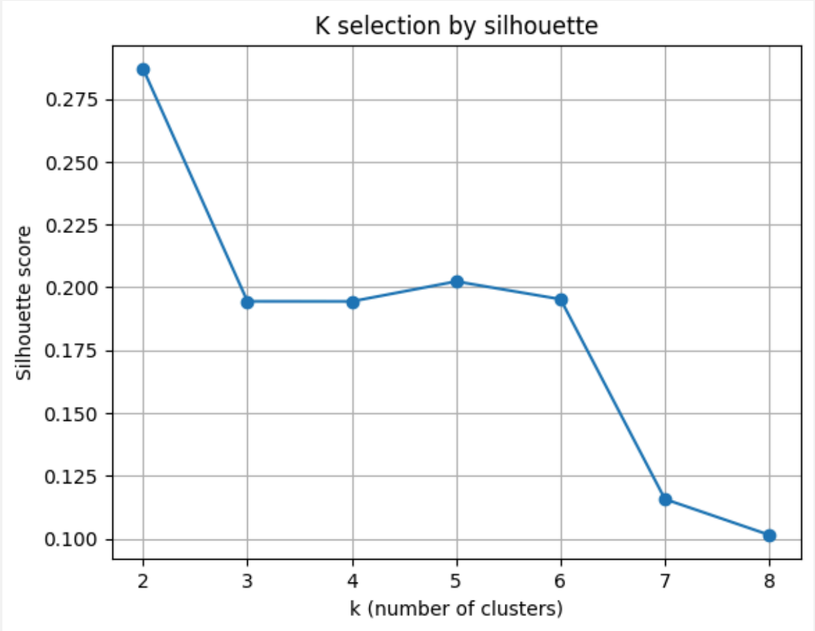
PCA & Clustering

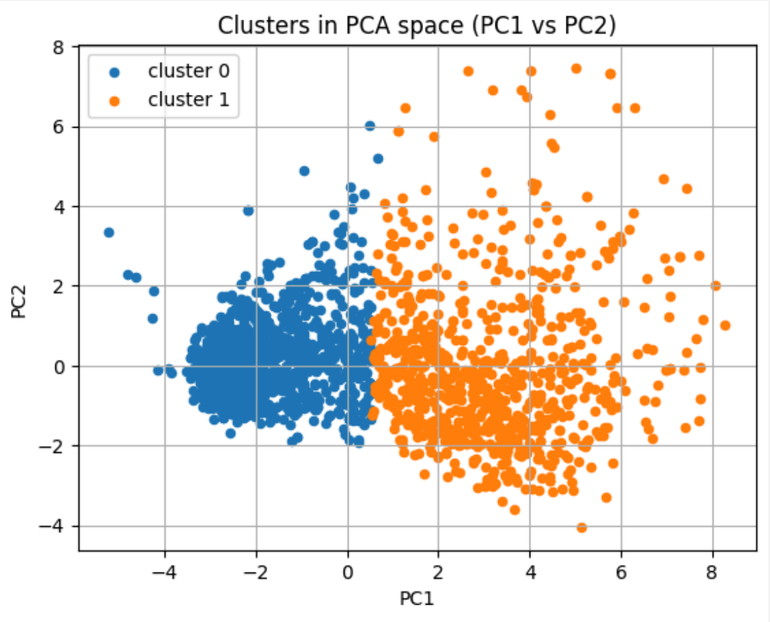
Reduced dimensionality using PCA

Optimal number of clusters

Clustering Results plotted on 2 principal components



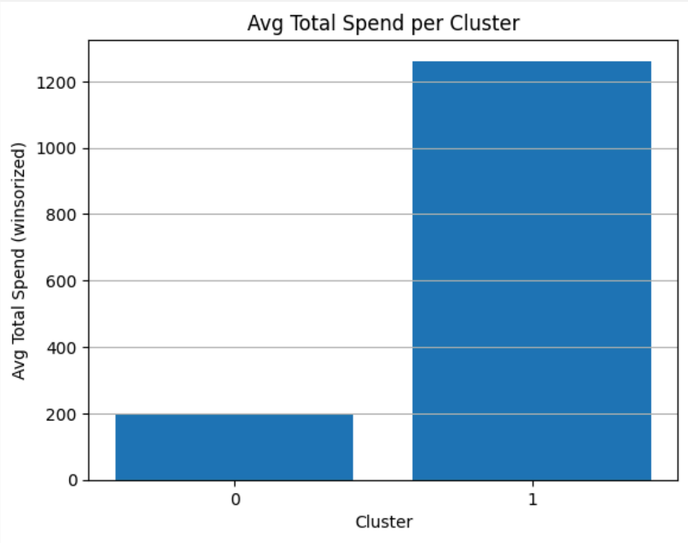




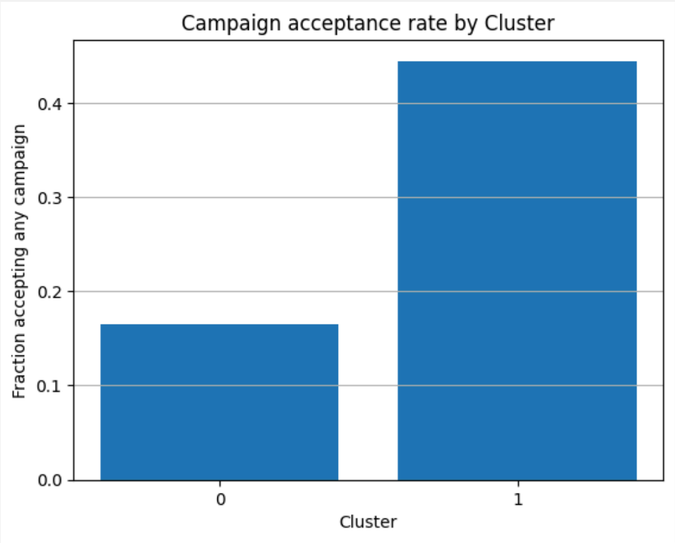
used 17 components

Cluster Insights

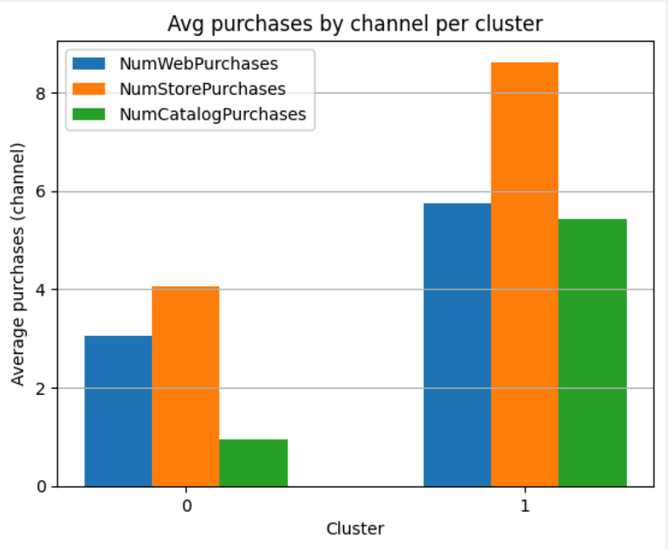
Spendings



Campaign acceptance



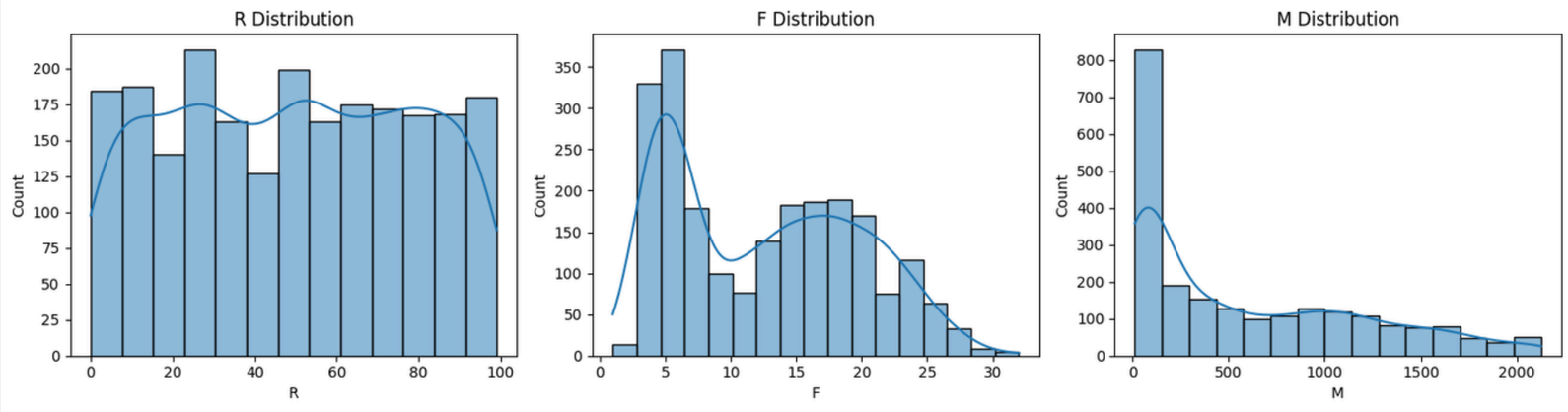
Channels

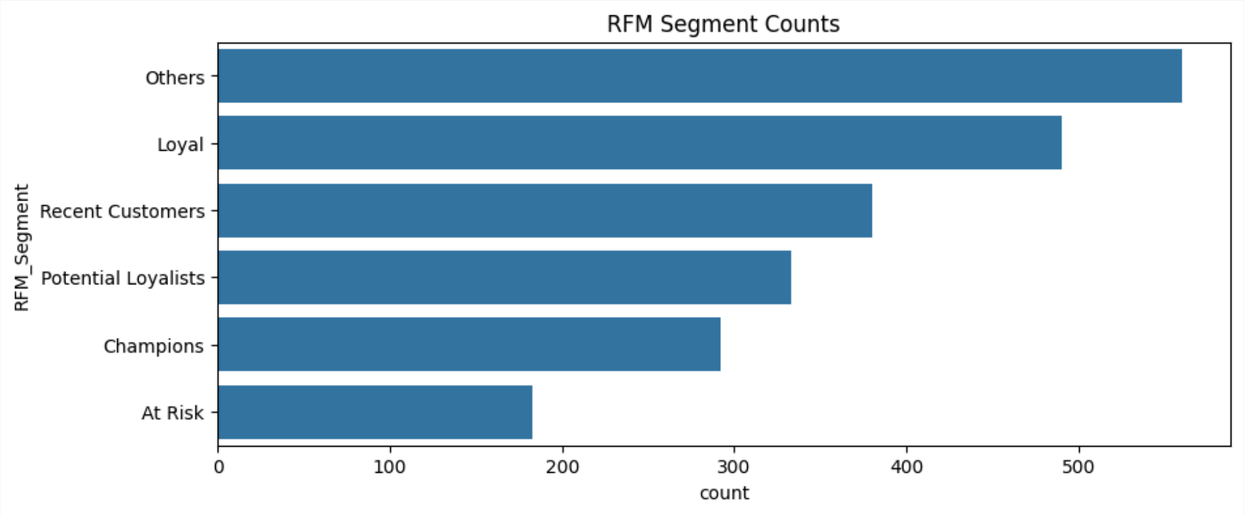


RFM

Recency, Frequency, Monetary

Identified loyal, frequent, and high-spend customers

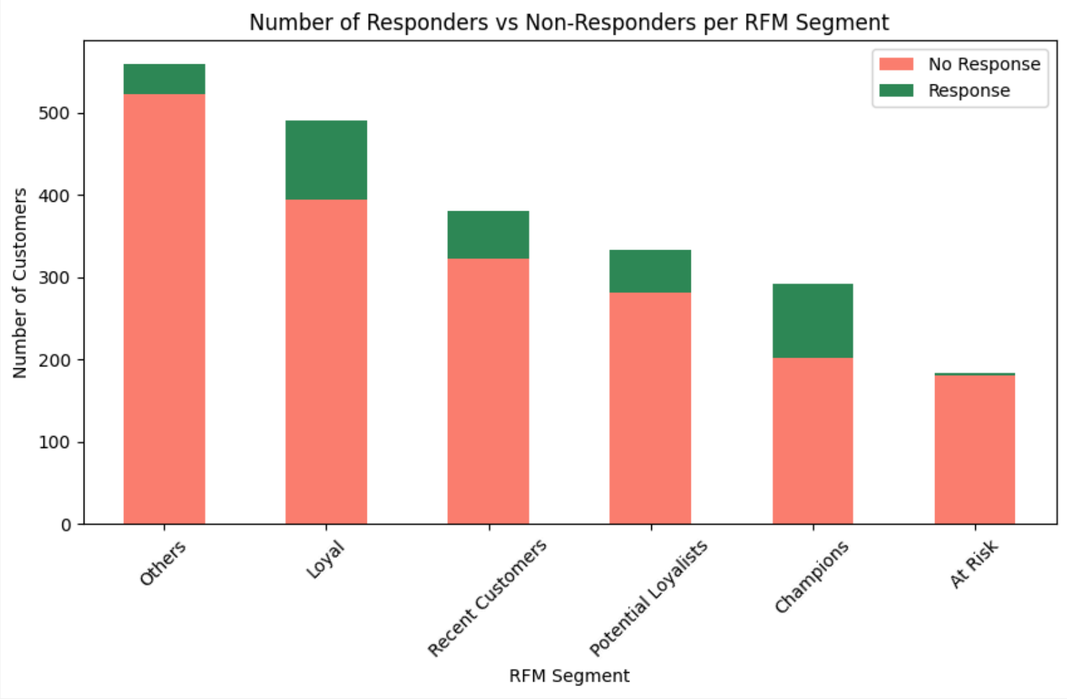


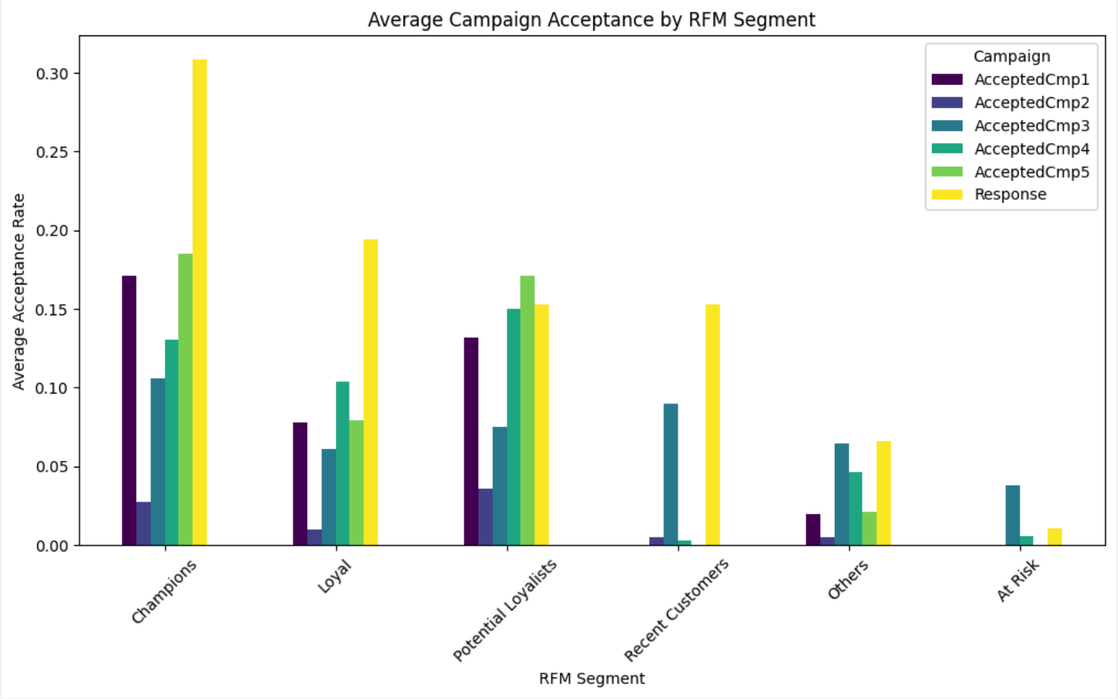


RFM : Campaign Responses

Recency, Frequency, Monetary

Identified loyal, frequent, and high-spend customers





Predictive Modeling

Logistic Regression

- **Goal**: Predict customer response to marketing campaigns
- **Features used**: Age, Income, Recency, Total Spend, Children, Purchases, Channel Activity, Campaign Acceptance
- **Method**: Logistic Regression (train/test split, standardized features)
- **Output**: Accuracy, confusion matrix

```
Accuracy: 0.9159212880143113

Confusion Matrix:
 [[459  17]
 [ 30  53]]

Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.96      0.95       476
           1       0.76      0.64      0.69        83

    accuracy                           0.92       559
   macro avg       0.85      0.80      0.82       559
weighted avg       0.91      0.92      0.91       559
```

Thank you !!

References

Romero-Hernandez, O. (n.d.). Customer Personality Analysis [Dataset]. Kaggle.
Monitor revenue, track conversion rates, and analyze the sales pipeline regularly.