CUSTOMER PERSONALITY ANALYSIS

Snigda Gedela

CHAPTERS

BUSINESS SUMMARY

Stakeholders | Background

ANALYTICAL PROCESS

Goal | Plan | Dataset

UNSUPERVISED MODEL

PCA | Clustering

SUPERVISED MODEL

Random Forest |
Multinomial Logistic

ANALYSIS REPORT

Result summaries

BUSINESS SUMMARY

01

Company

Specialty grocery store

02

Stakeholder

CMO

03

Background

Company A is a specialty grocery chain that has a national footprint. It offers both online and offline service. The company was expanding quickly in the past 2 years. The CMO is looking for solutions to better target different customer segments and improve campaign effectiveness

ANALYTICAL PROCESS

GOAL

Use the company's customer dataset to create customer segments, identify demographic and behavioral characteristics, and develop targeted campaigns based on prediction models

PLAN

Data Collection
↓
EDA
↓
Feature Engineering
↓
Segmentation
↓
Prediction

OUTPUT

- ❑ Defined and characterized customer segments
- ❑ Prediction for customers' responsiveness to campaign offers
- ❑ Recommendation on campaign strategy

RAW DATA OVERVIEW

*before cleaning

Dataset:	Marketing Campaign (https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis)
Number of Samples:	2240*
Number of Variables:	29 (25 int, 3 str, 1 id)*

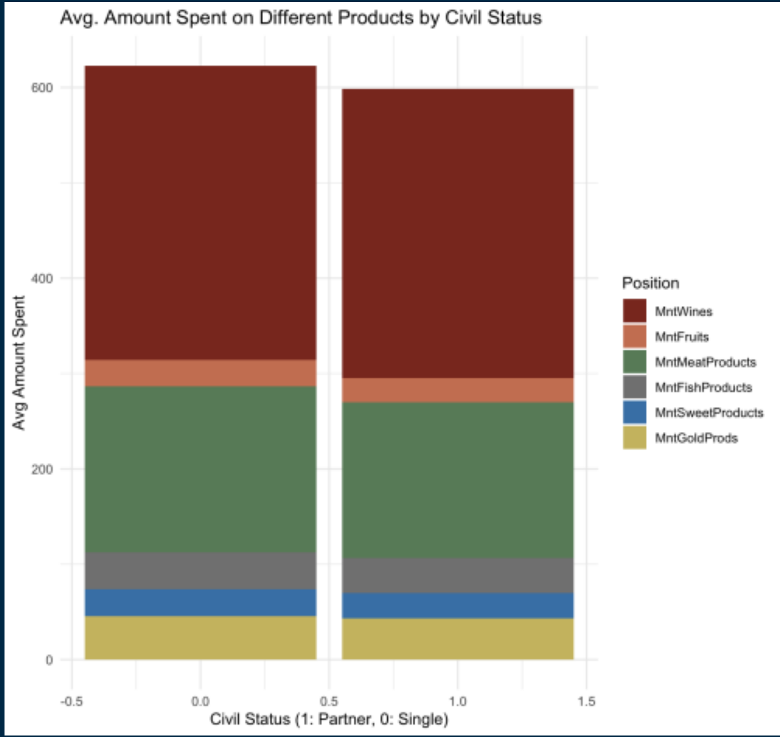
Overview Information

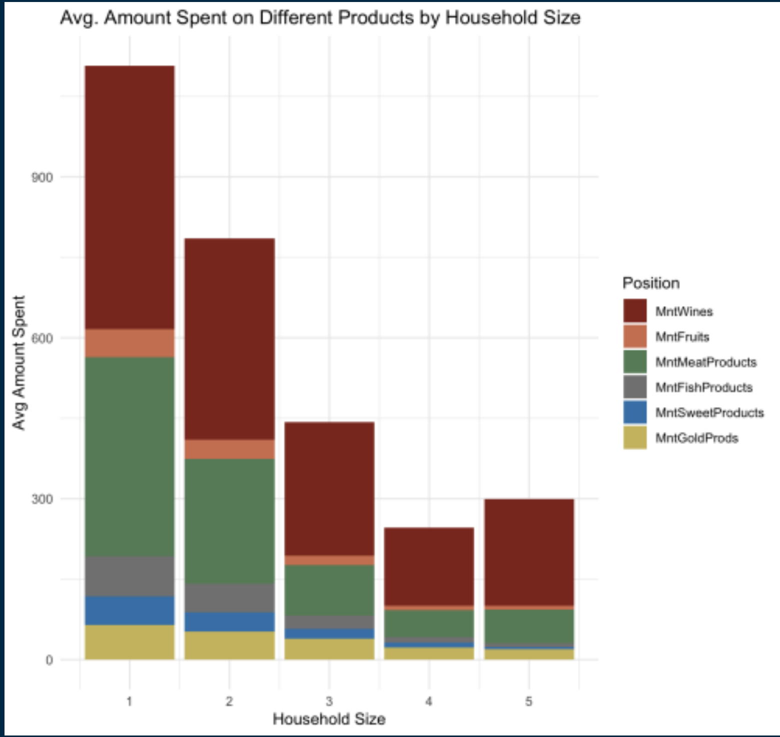
- Data was collected in the last 2 years (i.e. amount spent on meat in the last 2 years)
- Data collected is not consistent (i.e. Education level ->Graduate Degree = 2nCycle, Master, Grad)

About the Dataset

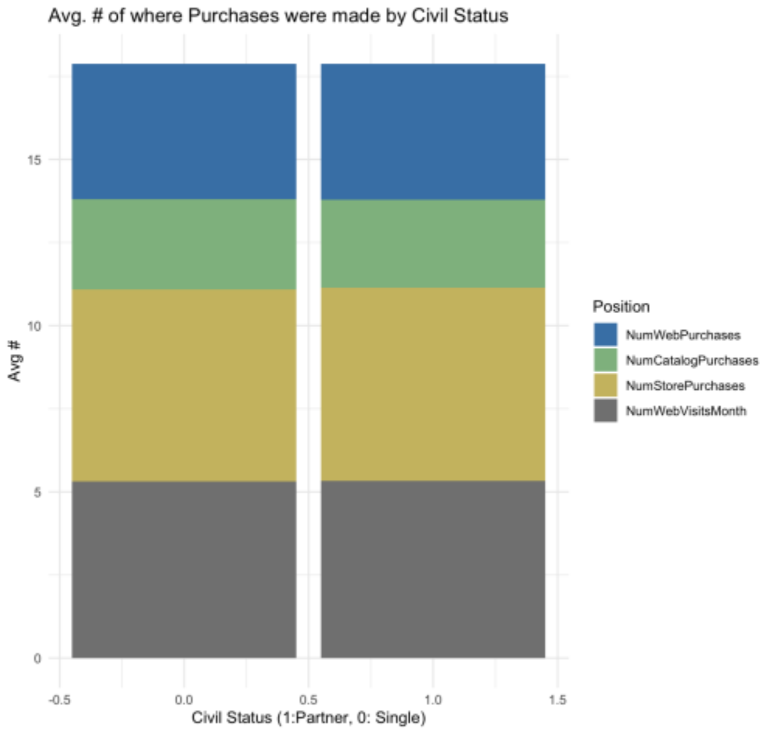
- Information on customer (education, year birth, etc)
- Information on customer's spending habits
- Information on customer's reaction towards promotion/campaigns
- Where the purchases were made

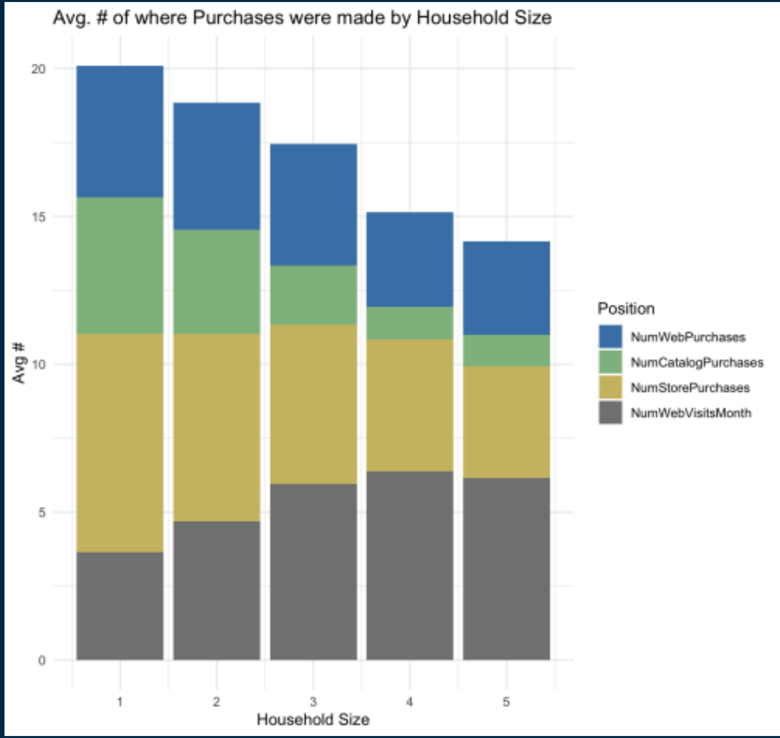
EDA: Avg Amount Spent



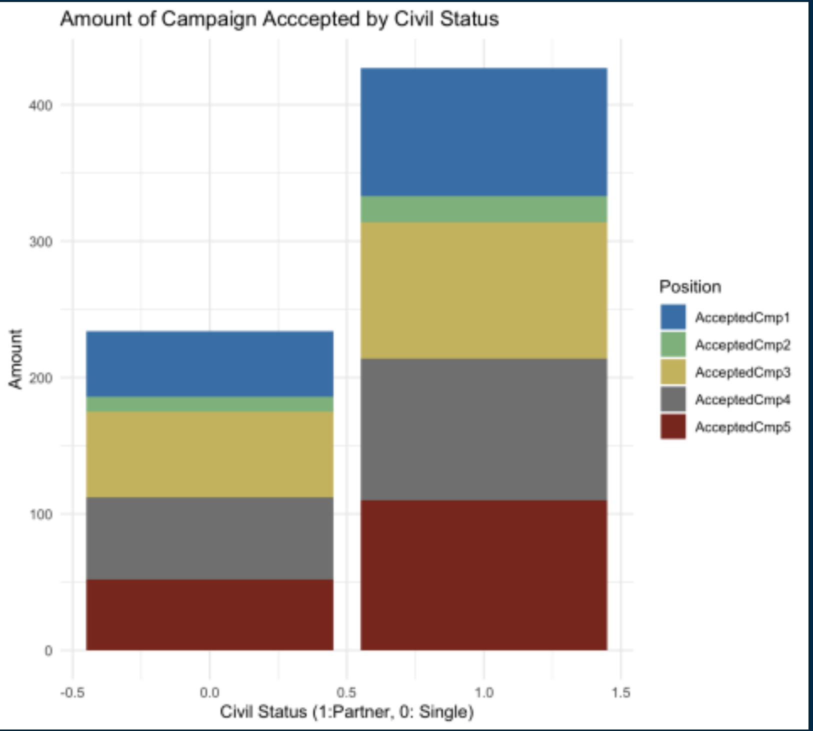


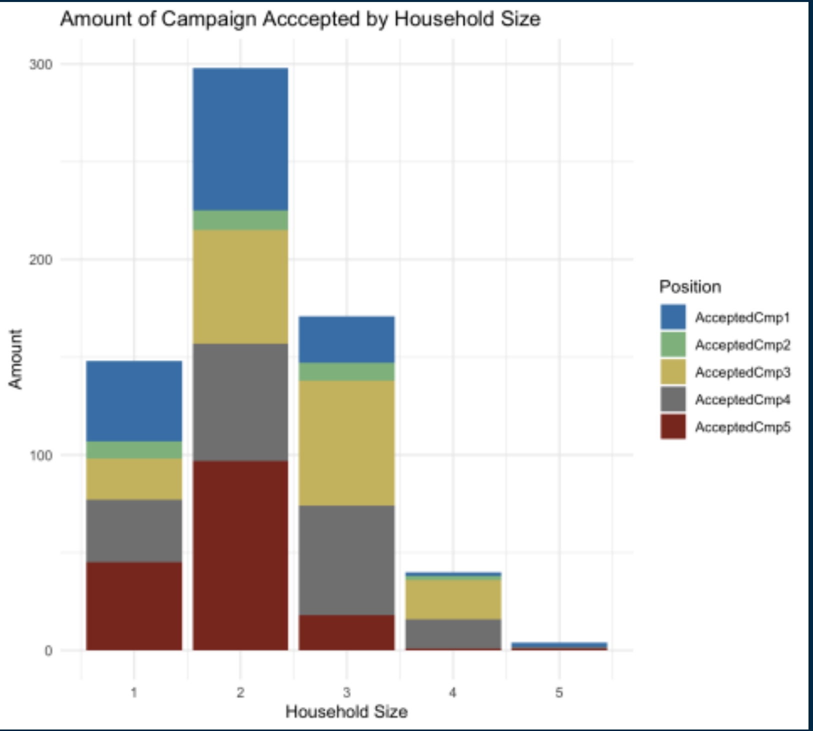
EDA: Where Purchases were Made





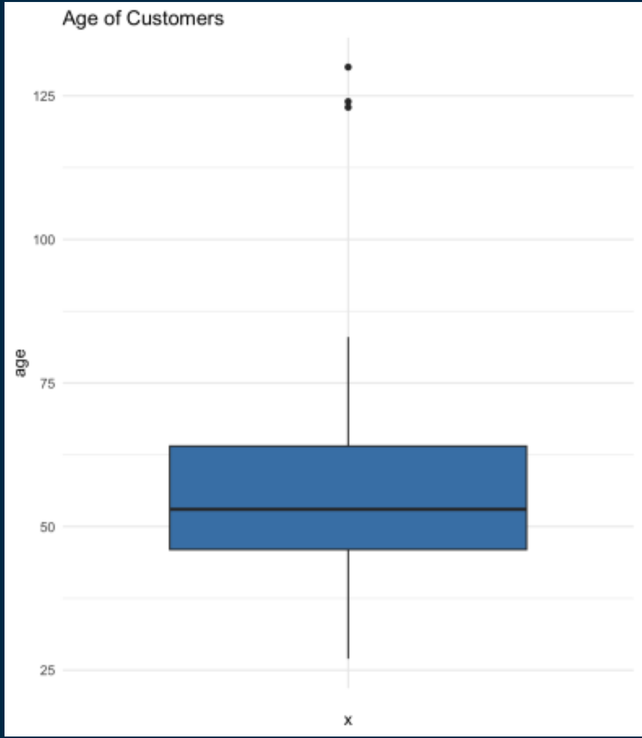
EDA: Amount of Campaign Accepted





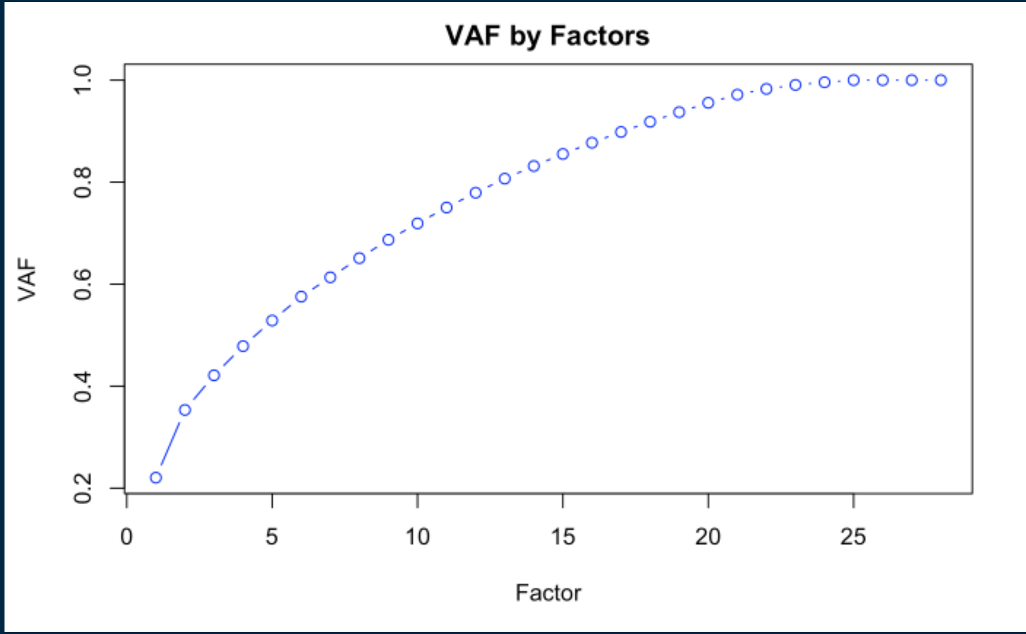
DATA CLEANING

- ❑ Dropped missing data
- ❑ Standardized Education field
- ❑ Standardized Marital Status field into Binary
- ❑ Combined the number of kid and teen into 1 column of children
- ❑ Converted Year_Birth into Age (2023-Year_Birth)
- ❑ Added total_spent and create a % of amount spent on wine, fruits, meats, etc
- ❑ Added total_place and create a % of purchases made through web, catalog, store
- ❑ Removed outlier; age > 90 and income > $600,000
- ❑ Removed columns that are not needed (ID, Year_Birth, Num, etc)



PRINCIPAL COMPONENT ANALYSIS

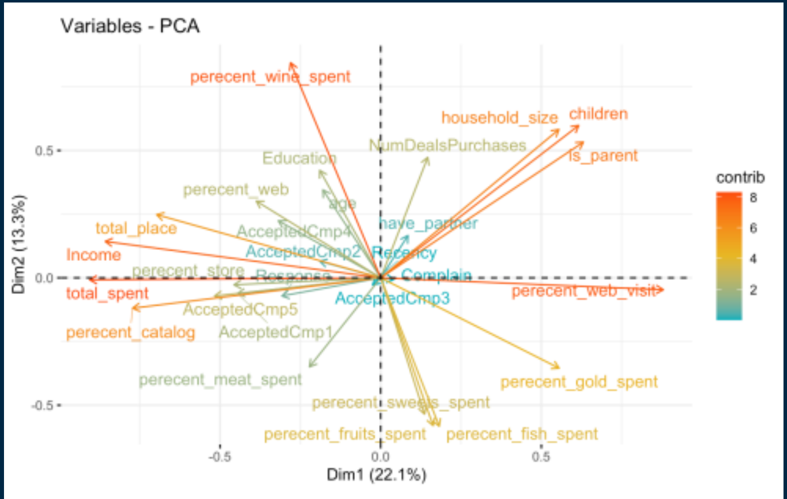
Many variables contribute to the total variance

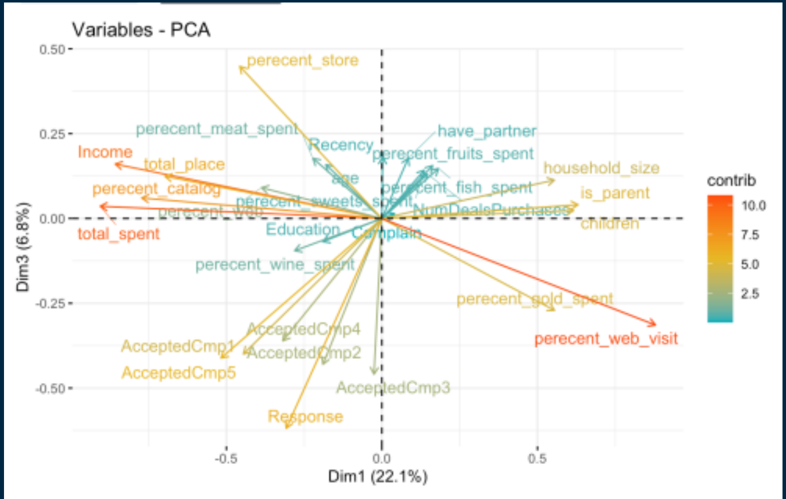


PRINCIPAL COMPONENT ANALYSIS

The Principal Component Analysis has identified three principal components that capture different components of the customer data.

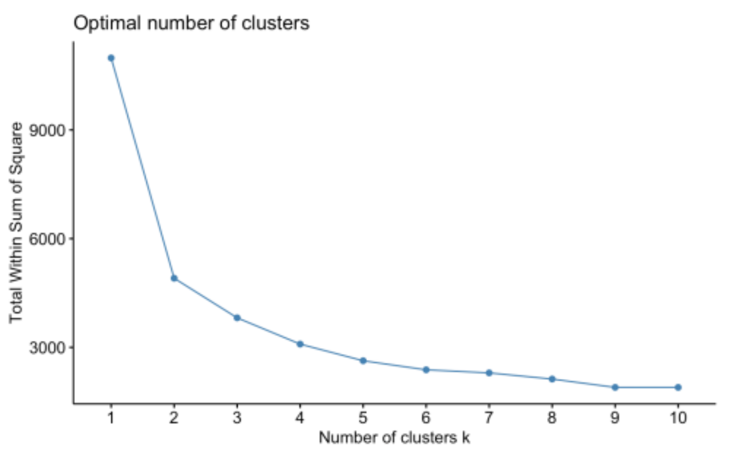
- PC1: The general spending habits and website usage of customers
- PC2: The purchasing habits of customers with respect to certain products
- PC3: The tendency of customers to accept campaign offers

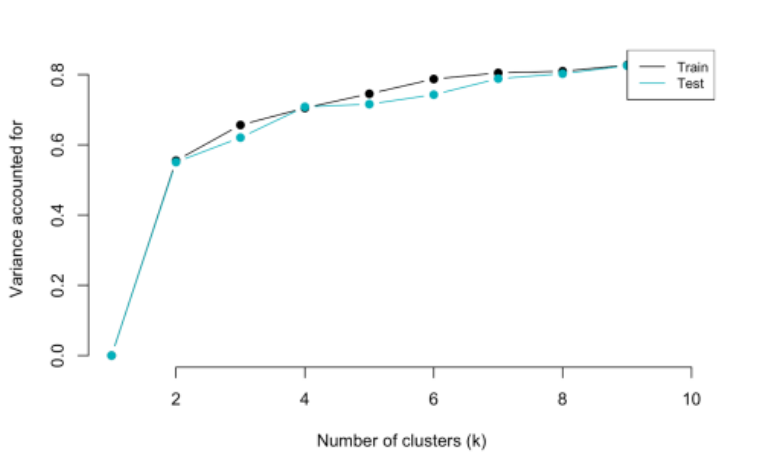




CLUSTERING

- Used variables based on the Principal Component Analysis results and generated 1-10 k-means clusters
- Generated scree plots of percentage of Variance-accounted-for (VAF) and Total Within Sum of Square





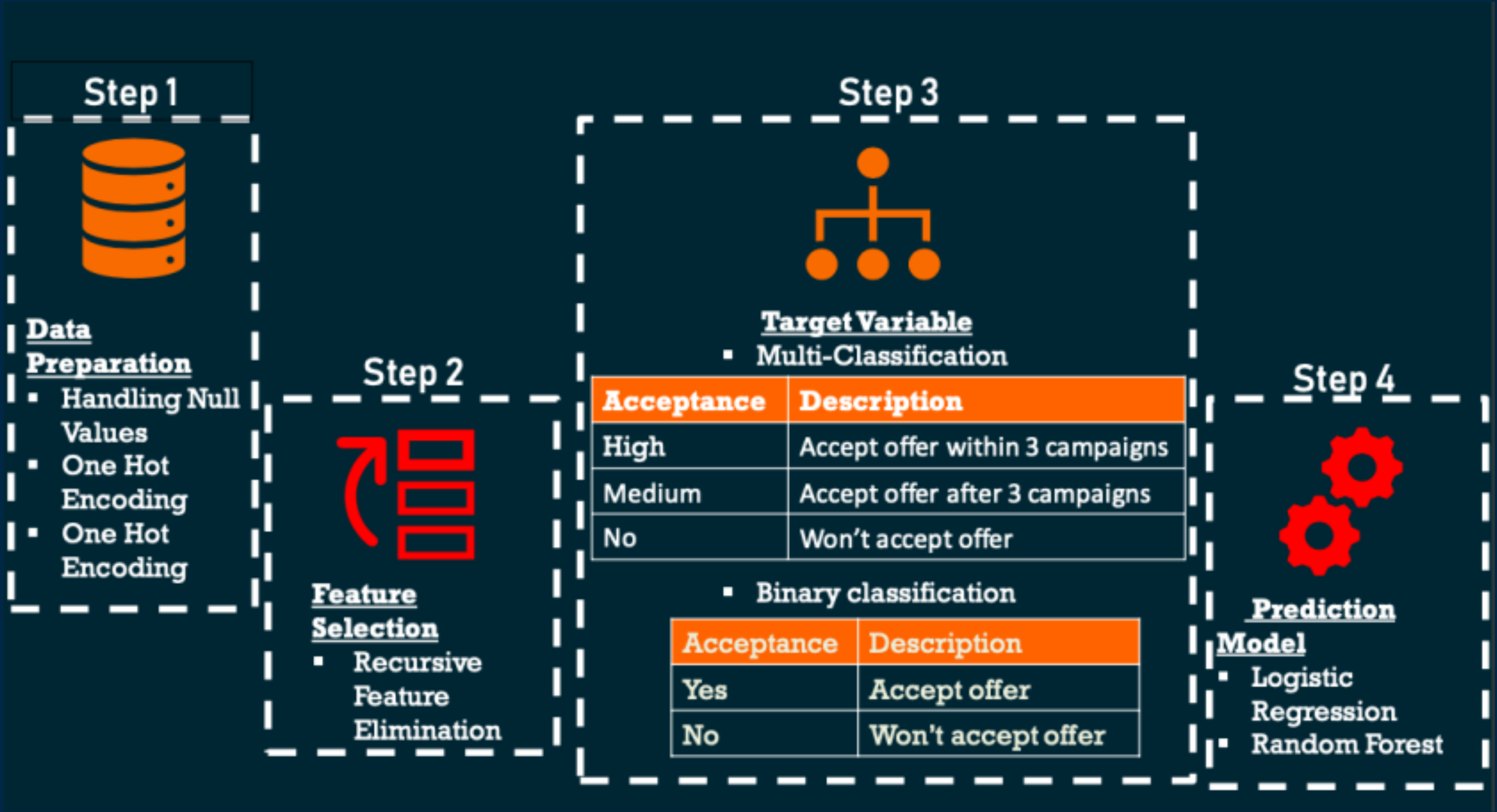
CUSTOMER SEGMENTS

- The Scree plots suggest that our customers can be separated into 2 groups:

- Low income
- More children
- Spend less in total
- Less frequent purchases
- Mostly purchase online

- Greater income
- Fewer children
- Spend more in total
- More frequent purchases
- Don’t purchase online as often

Customer Responsiveness Prediction



Acceptance	Description
High	Accept offer within 3 campaigns
Medium	Accept offer after 3 campaigns
No	Won't accept offer

Acceptance	Description
Yes	Accept offer
No	Won't accept offer

RANDOM FOREST - Multiclass Classification

Acceptance	F1 Score	Recall	Precision
High	54%	49%	59%
Medium	46%	40%	54%
No	88%	92%	85%

Observations:

- ❏ The model is able to predict customer that do not accept offers.
- ❏ But the model is not able to distinguish between High and Medium Acceptance customers.

RANDOM FOREST - Binary Classification

Acceptance	F1 Score	Recall	Precision
Yes	89%	92%	86%
No	68%	61%	75%

Observations:

- ❑ Binary Classification works better than multiclass classification for the data that we used.
- ❑ Might require more features to be able to distinguish different levels of acceptance of the customers

LOGISTIC REGRESSION / MULTINOMIAL LOGIT

Acceptance	Accuracy	Sensitivity	Specificity
High	66%	44%	88%
Medium	64%	40%	87%
No	72%	76%	67%

Acceptance	Accuracy	Sensitivity	Specificity
Yes	75%	71%	78%
No	74%	78%	70%

Observations:

- ❑ Similar results as random forest in classification methods
- ❑ Generally random forest delivers a better result in prediction

ANALYSIS REPORT

Executive Summary

The analysis used the customer demographic and behavioral data in the past 2 years to identify customer segmentations and build a prediction model on responsiveness to campaign offers. The methodology and result of the analysis will demonstrate an effective and economical way to the CMO of a national specialty grocery brand to develop target marketing plans, improve conversion rate, and increase average customer spend.

Analysis & Output

<u>Two customer segments:</u>

- ❏ Online family shoppers
- ❏ In store frequent shoppers

<u>Predict responsiveness to campaign offers</u>

- ❏ Final Model: Random Forest

ANALYSIS REPORT

Recommendation

Campaign target

- Segment 1 : Online family shoppers
  Focus on online campaigns and SEO, improve online shopping user experience, promote kids relevant product categories and products in bulk
- Segment 2: In store frequent shoppers
  Promote new and high value products

Campaign frequency

- Group 0 (won't accept offers): decrease campaign frequency, only send out major promotional information with targeted product categories based on relevance
- Group 1(accept offers): create a silent period once offers are accepted, rotate campaign contents

Next step

- Deeper analysis into product categories and sales channels
- Need more data in campaign contents, channels and frequency to better evaluate campaign effectiveness
- Test new campaign methods and validate results in the next 2 months

THANK YOU!