MMAI 831 – AI in Marketing

Executive Marketing Challenge

Superstore Campaign Analysis using AI & ML

Team Bay

June 11, 2024

AGENDA

Exciting new Offer:

20%

Discount on All Purchases for Gold Membership Customers

Challenge:

Year-End Sale Planning: *The superstore is preparing for a year-end sale*

New Offer: *Launching a gold membership that gives a 20% discount on all purchases*

Pricing: *Membership offered at $499 during the campaign, compared to $999 on other days*

Campaign Strategy: *Promotion through phone calls to existing customers*

Cost Reduction Goal: *Aim to reduce campaign costs by using a predictive model*

Why is it Important?

Cost Efficiency

Reducing unnecessary marketing expenses through precise targeting saves significant resources

Increased Sales

By identifying and targeting the right customers, the superstore can maximize the sale of gold memberships

Enhanced Experience

The additional benefits and discounts can enhance the overall shopping experience, leading to higher customer satisfaction and retention

Superstore Data Overview

Independent Variables (Features)

21 Features

**Demographics:** *Birth year, Education, Marital, Kids, Teens, Income*

Customer Engagement: *Complain, Date of customer's enrollment , Website Visits, Recency*

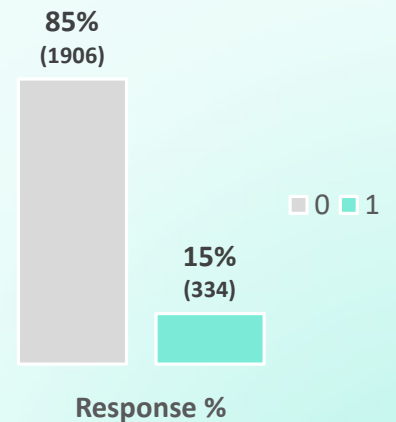**Purchase Behavior:** *Deals Purchases, Catalog Purchases, Store Purchases, Web Purchases*

Product Spending: *Fish, Meat , Fruits, Wines, Gold Products*

Dependent Variable (Target)

Response: Whether customer accepted the offer in the last campaign

0 — No

1 — Yes
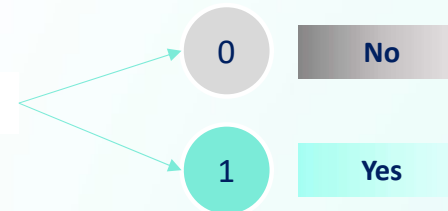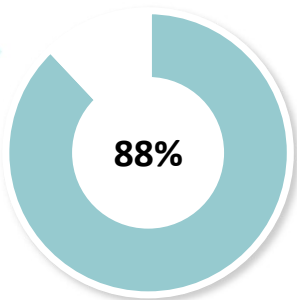
Classification Problem!

85% (1906)

15% (334)

■ 0 ■ 1

Response %

Data is heavily imbalanced

Descriptive Data Analysis

Customer Demographics

88%

Well-educated individuals

(Graduates, Masters, PhDs)

63%
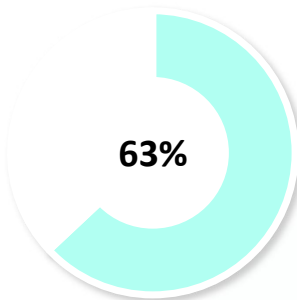
Living Together

(Married, Together)

Avg Income

PhDs + Married	**$56K**
Master + Married	$53K
Graduates + Married	$51K
Basic Edu + Other Marital Status	**$20K**

Higher Positive Response Rate for Membership

Ages 40 - 55

92%
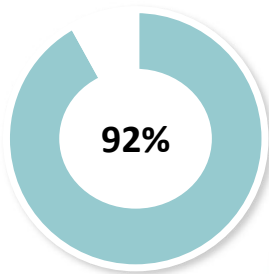
Well-educated individuals

Others, 21%
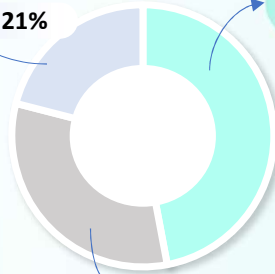
Living Together, 47%

Single, 32%

Avg Income

$60K

Spending Habits

Average Amount Spent on Different Products by Response

● Fish Products ● Fruits ● Sweet Products ● Gold ● Meat ● Wines

1	52	38	38	61	294	503	986

For positive responses, an average of $986 is spent

Average high spending on Wine, Meat and Gold products

Meat, Gold, Wines

BY RESPONSE, TOTAL_CHILDREN, KIDHOME, TEENHOME

● Meat ● Gold ● Wines

540

600   800   1000

unt Spent on Different Products (in dollars)
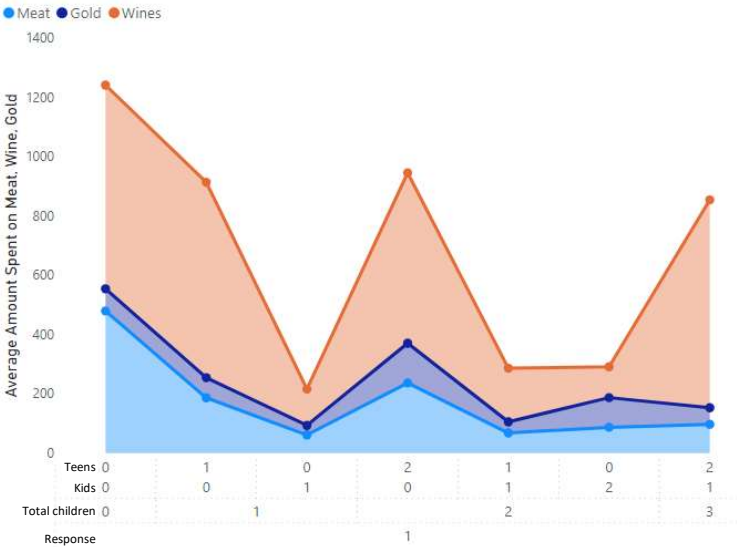
People spent more on wines and gold when they have less dependents or teens

Teens	0	1	0	2	1	0	2
Kids	0	0	1	0	1	2	1
Total children	0		1		2		3
Response				1			

Catalog Purchases, 21%

Store Purchases, 46%
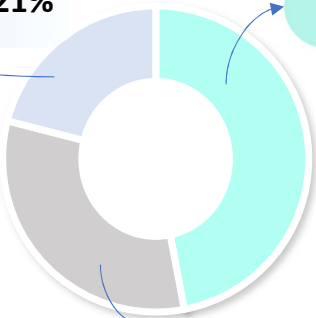
 # Purchases via Different Channels

Web and Store purchases are more preferred

Web Purchases, 32%

3

1
0
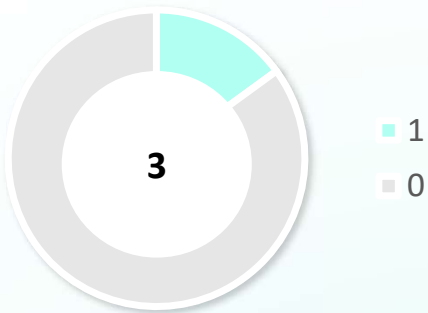
 # Complaints for positive response

High customer satisfaction among positive responses

Response vs Recency

Boxplot of Recency by Response

Recency

50

45

40

35
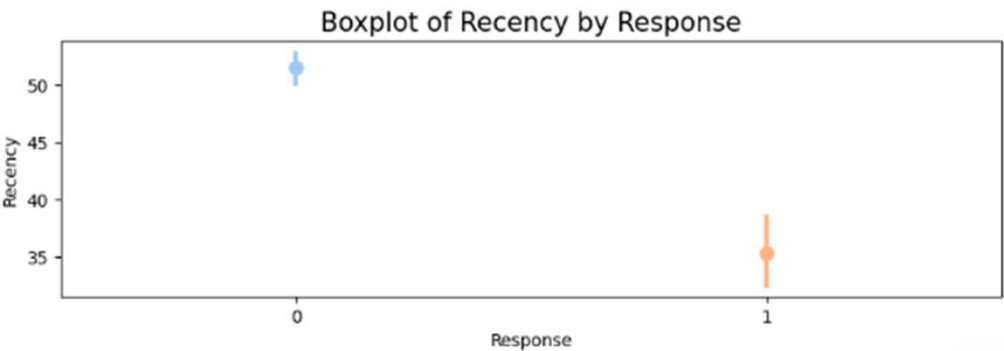
0   Response   1
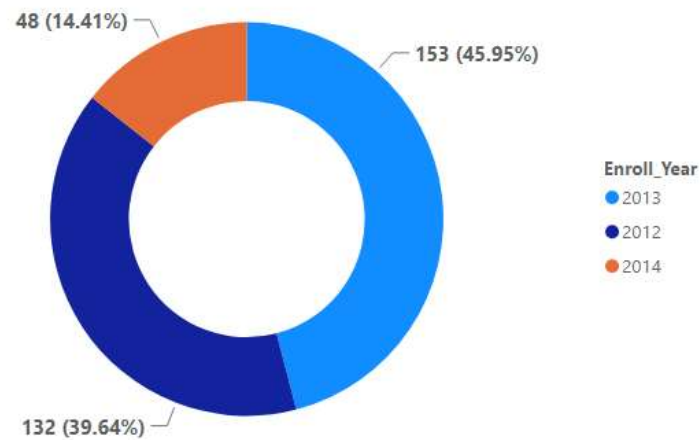
Recent interactions generate more positive responses

Average Recency of 35

Customer Tenure and Loyalty

Sum of Response
BY ENROLL_YEAR

48 (14.41%)
153 (45.95%)

Enroll_Year
● 2013
● 2012
● 2014

132 (39.64%)

Sum of Response
BY ENROLL_QUARTER, ENROLL_YEAR

Sum of Response

Enroll Quarter and Year

Higher interest in membership among customers who enrolled in 2012 and 2013
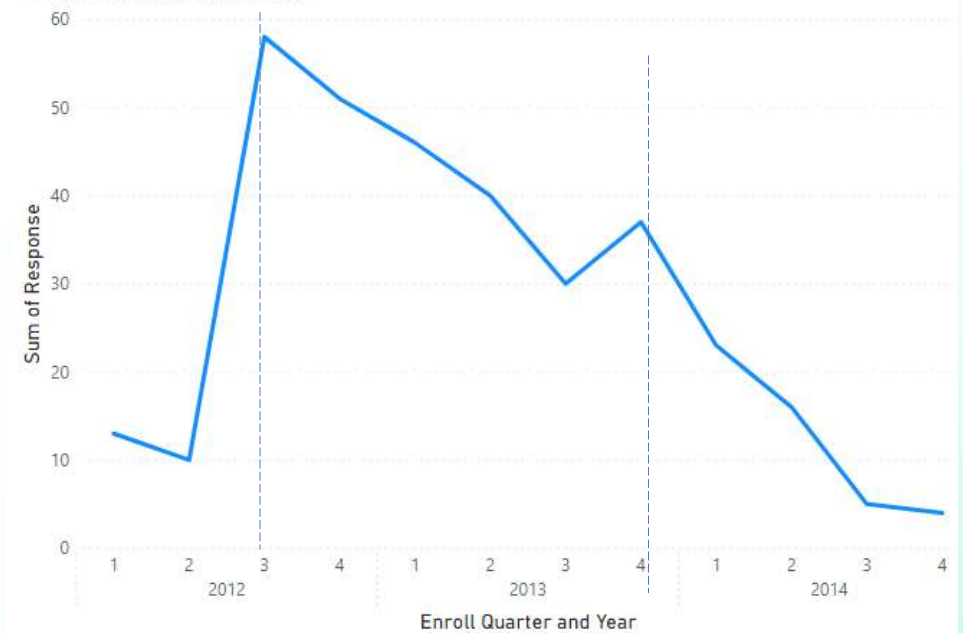
Downward trend in the Response over the years people enrolled

Data Modelling Analysis
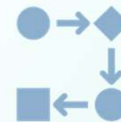
Data Modelling Analysis

1

Exploratory Data Analysis

Train – Test Split

Handling Data Imbalance
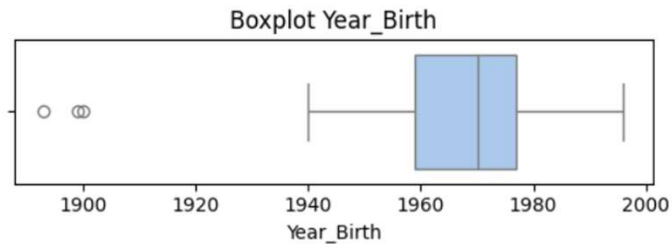
Feature Engineering + Handling Outliers

Cross Validation to Various ML Models

Feature Selection & Hyperparameter Tuning

Evaluation

Outlier Analysis

Boxplot Year_Birth

Boxplot Income

Age > 100 present

24 null income values

Income Outlier > 4 M

Data Imbalance

Count of Response

Data heavily imbalanced, more bias towards negative response
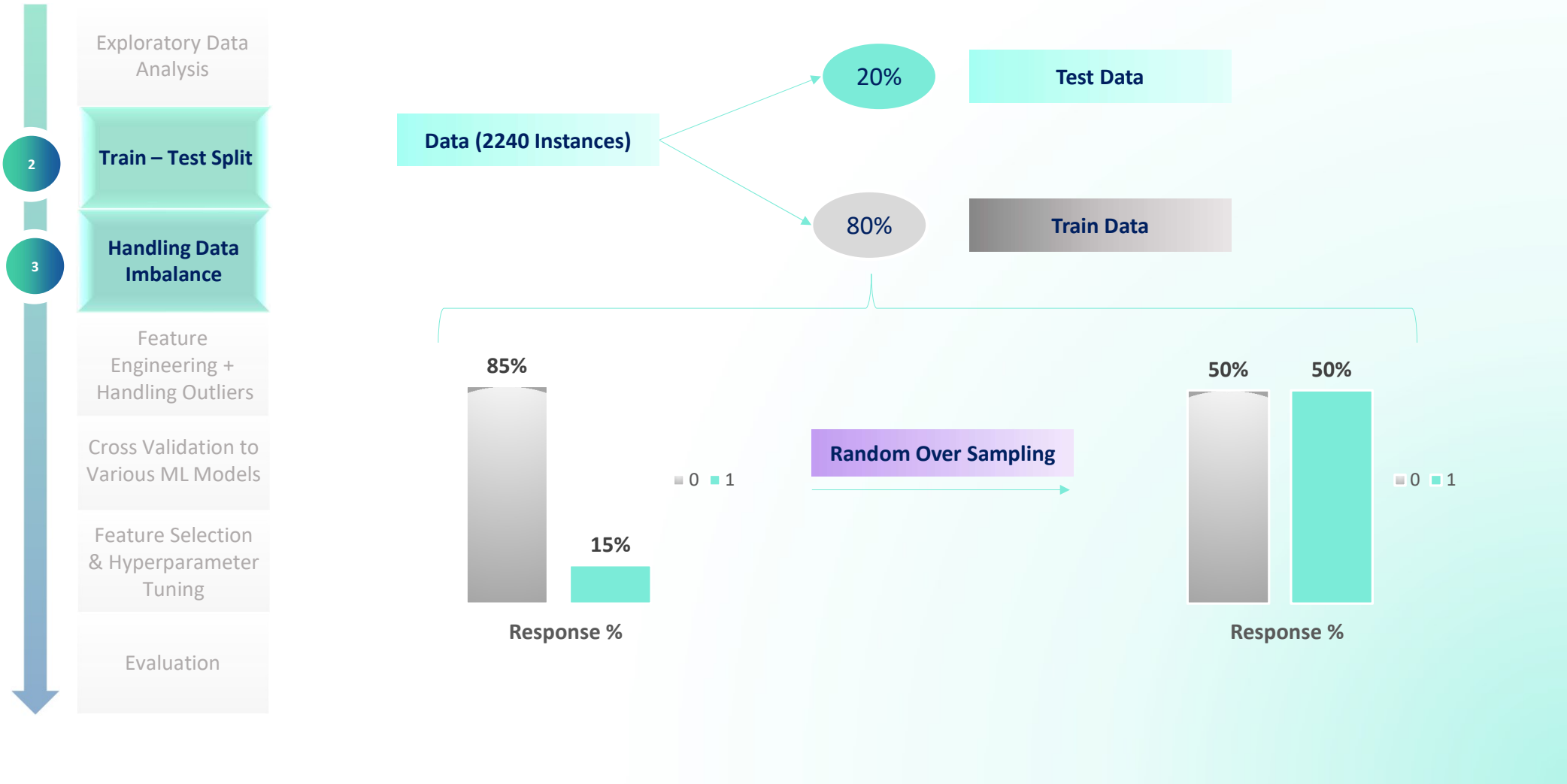
Data Modelling Analysis

Exploratory Data Analysis

2 Train – Test Split

3 Handling Data Imbalance

Feature Engineering + Handling Outliers

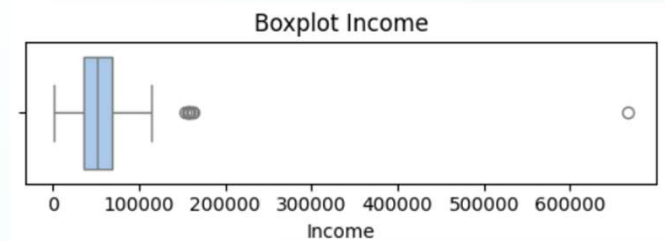Cross Validation to Various ML Models
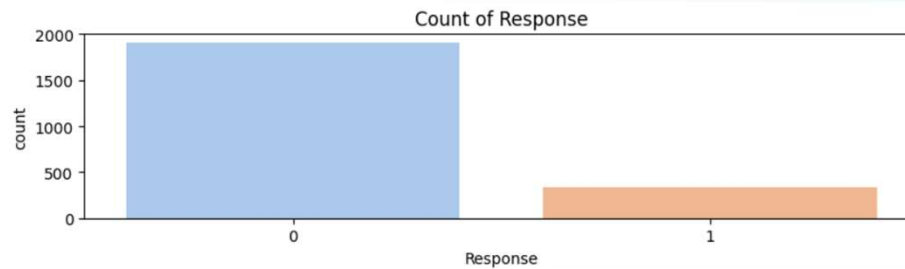
Feature Selection & Hyperparameter Tuning

Evaluation

Data (2240 Instances)

20% → **Test Data**

80% → **Train Data**

85%

15%

■ 0 ■ 1

Response %

Random Over Sampling

50%   50%

■ 0 ■ 1

Response %

Data Modelling Analysis

4

- Exploratory Data Analysis
- Train – Test Split
- Handling Data Imbalance
- **Feature Engineering (FE)**
- Cross Validation to Various ML Models
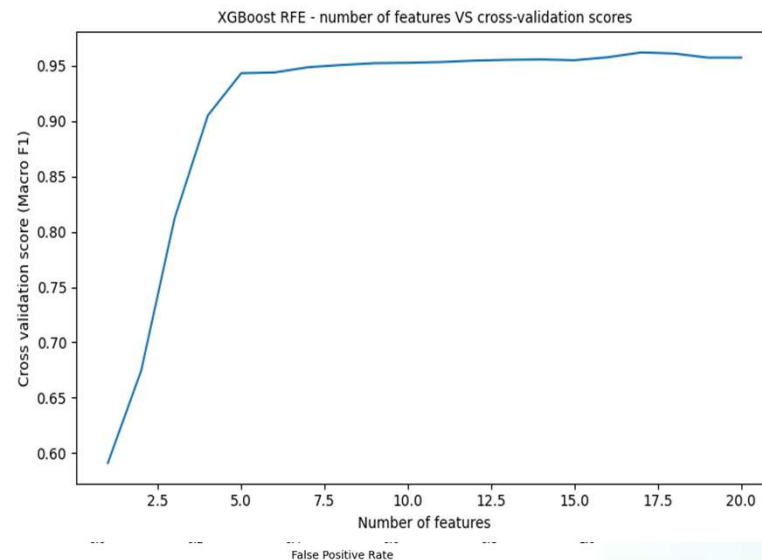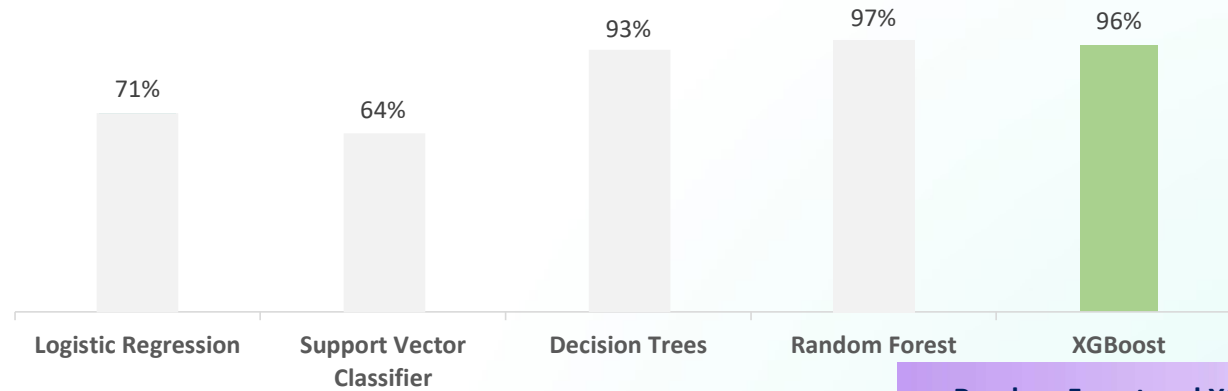- Feature Selection & Hyperparameter Tuning
- Evaluation

Handling Outliers

- ✓ Filtered rows with Year Birth > 1920
- ✓ Filtered Rows with Income < 4M

Rare Label and One Hot Encoding

- ✓ Rare Label: Pick top 4 labels, and classify rest as 'rare'
- ✓ Education of customers:
  - Categories: Graduate, Master, PhD...
- ✓ Marital status of customers:
  - Categories: Married, Single, Together, …

Other Features

- ✓ Total Children at Home: Kids + Teens
- ✓ Total Spending
- ✓ Total Purchase
- ✓ Deal Purchase Ratio
- ✓ Web Purchase Ratio
- ✓ Average spending per purchase
- ✓ Income to Spending Ratio

Handling Missing Income

- ✓ Replaced by average income per marital status and education

DateTime Features

- ✓ Age
- ✓ Customer Enrollment Date --> Enrollment Year, Month, Quarter, Day

Scaling

- ✓ Standard Normal Scaler

~ 38 Features after FE

Data Modelling Analysis

Cross Validation Scores (Macro F1 Score) of Different Models

Logistic Regression — 71%	
Support Vector Classifier — 64%	
Decision Trees — 93%	
Random Forest — 97%	
XGBoost — 96%	

Pipeline steps (left navigation):
- Exploratory Data Analysis
- Train – Test Split
- Handling Data Imbalance
- Feature Engineering + Handling Outliers
- **5 Cross Validation on Various ML Models**
- **6 Feature Selection & Hyperparameter Tuning**
- Evaluation

XGBoost RFE - number of features VS cross-validation scores
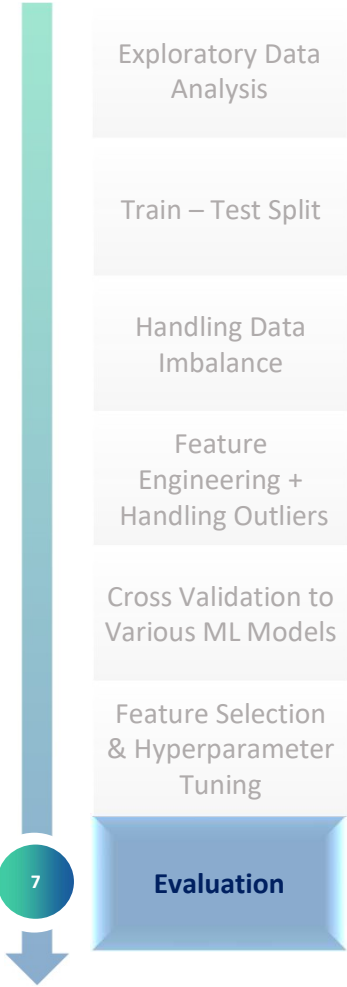
- **Random Forest and XGBoost perform better**
- **Model Picked: XGBoost** (*Better Performance, higher recall, precision*)
- **Feature Selection using Recursive Feature Elimination + CV method**
- **32 optimal features selected**
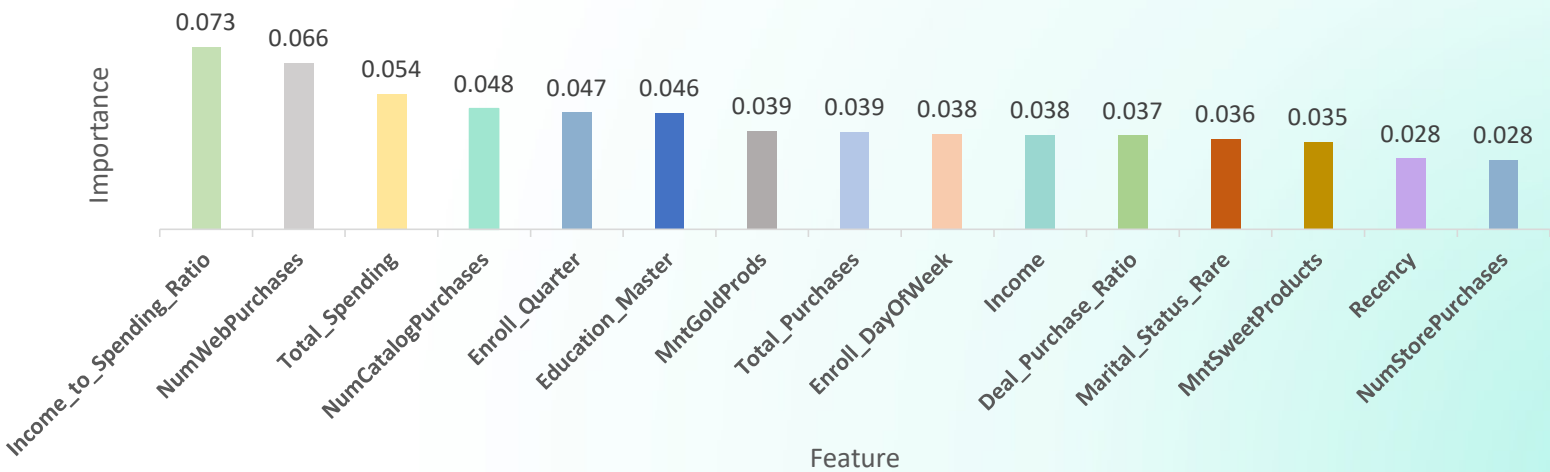- **Hyperparameter Tuning using GridSearchCV**

Evaluation

XGBoost Model Performance Evaluation

Confusion Matrix

		Predicted		
		0	1	
Actual	0	358	18	376
	1	28	39	67
		386	57	443

Classification Matrix Report

	Precision	Recall	F1-Score	Support
0	0.93	0.95	0.94	376
1	0.68	0.58	0.63	67
Accuracy			0.90	443
Macro Avg	0.81	0.77	0.78	443
Weighted Avg	0.89	0.90	0.89	443

XGBoost Feature Importance

Feature importance values: Income_to_Spending_Ratio 0.073, NumWebPurchases 0.066, Total_Spending 0.054, NumCatalogPurchases 0.048, Enroll_Quarter 0.047, Education_Master 0.046, MntGoldProds 0.039, Total_Purchases 0.039, Enroll_DayOfWeek 0.038, Income 0.038, Deal_Purchase_Ratio 0.037, Marital_Status_Rare 0.036, MntSweetProducts 0.035, Recency 0.028, NumStorePurchases 0.028

Results + ROI

Pre-Implementation

Cost of Marketing Campaign [1]	443 * $35	**$15,505**
Total Revenue	67 * $499	$33,433
Net Profit	$33,433 − $15,505	**$17,928**

- $ROI = \frac{Net\ Profit}{Total\ Investment} * 100 = \frac{17,928}{15,505} * 100 = \mathbf{115.62}\% $

Post-Implementation

Cost of Marketing Campaign [1]	57 * $35	$1,995
Cost of Development [2]	$60/hour * 40hr	$2,400
Deployment and Maintenance [3]	$0.20 * 24 * 365	$1,752
Total Cost		**$6,147**
Total Revenue	39 * $499	$19,461
Net Profit	$19,461 − $5,867	**$13,314**

- $ROI = \frac{Net\ Profit}{Total\ Investment} * 100 = \frac{13,314}{6,147} * 100 = \mathbf{216.59}\% $

		Predicted		
		0	1	
Actual	0	358	18	376
	1	28	39	67
		386	57	443

1.87x

Increase in **ROI**

60.35%

Decrease in **Campaign Cost**

1. https://www.magellan-solutions.com/blog/cost-of-telemarketing/
2. https://www.ziprecruiter.com/Salaries/Ai-Ml-Developer-Salary
3. https://aws.amazon.com/sagemaker/pricing/

Recommendations

Campaign Strategy

- **Pre-Campaign Awareness**
  Conduct surveys to gauge customer preferences and refine campaign strategies.

- **Ongoing Feedback**
  Continuously gather feedback during the campaign to adjust tactics.

- **Referral Incentives**
  Implement referral perks to encourage existing customers to promote the gold membership.

Targeted Segments

- **Low Recency**
  Focus on customers with recent interactions (recency < 35).

- **Educational Background**
  Prioritize customers with advanced degrees (Masters and PhDs).

- **Marital Status**
  Target married or cohabiting individuals.

- **Enrollment Period**
  Engage customers enrolled between Q3 2012 and Q4 2013.

- **Spending Behavior**
  Focus on customers with high average spending per purchase, particularly those who spend more on wines.

Optimizing Spend

- **Effective Resource Allocation**
  Focus on high-impact channels and strategies to maximize conversions. Prioritize phone calls and explore enhancing targeted website ads.

- **Elasticity Focus**
  Target customers likely to respond to the offer, especially those with higher potential (TP + FN).

Conclusion

Key Benefits of AI/ML Implementation

Reduced campaign costs by 60%

Accurate identification of potential customers to boost sales

Actionable insights on customer demographics, spending habits, engagement, and loyalty that can enhance overall customer experience

Thank You

Appendix

Statistics

- Takeaways from the CMO Intentions 2024 study by the CMO Council and Zeta Global Holdings Corporation. - based on a survey of nearly 200 chief marketing executives throughout North America and Europe.
  - AI And GenAI Leading MarTech Investment, 60% of marketers view this initiative as providing the most value and return on investment (ROI)
  - **1 in 3 CMOs say their campaign performance needs improvement or has underperformed.**
- The global market revenues of AI usage in marketing expected to reach 36 billion U.S. dollars in 2024. – Ref
- $107 billion is the global market revenue forecast for AI in marketing in 2028. – Ref
- In 2022, 90% of marketers working with AI say that the technology has helped them personalize the customer journey across different channels. - Ref
- More than 80% of marketers worldwide integrate some form of AI into their online marketing activities. - Ref

Achievable
Leverage historical customer data and machine learning algorithms to identify and target the top 20% of customers most likely to purchase the gold membership.

Measurable
Aim for a 15% increase in gold membership purchases among targeted customers compared to last year's campaign

Relevance
Enhance customer loyalty and boost year-end sales revenue through targeted marketing efforts
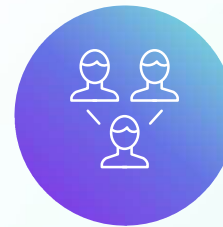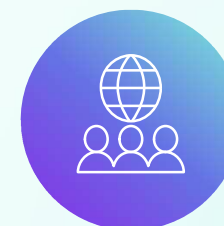
Specific
Increase the conversion rate of existing customers purchasing the gold membership during the year-end sale

Time-Bound
Achieve the objective by the end of the year-end sale period, which spans from December 1st to December 31st

SMART Objectives

EDA

Univariate Analysis: Categorical Variable Analysis
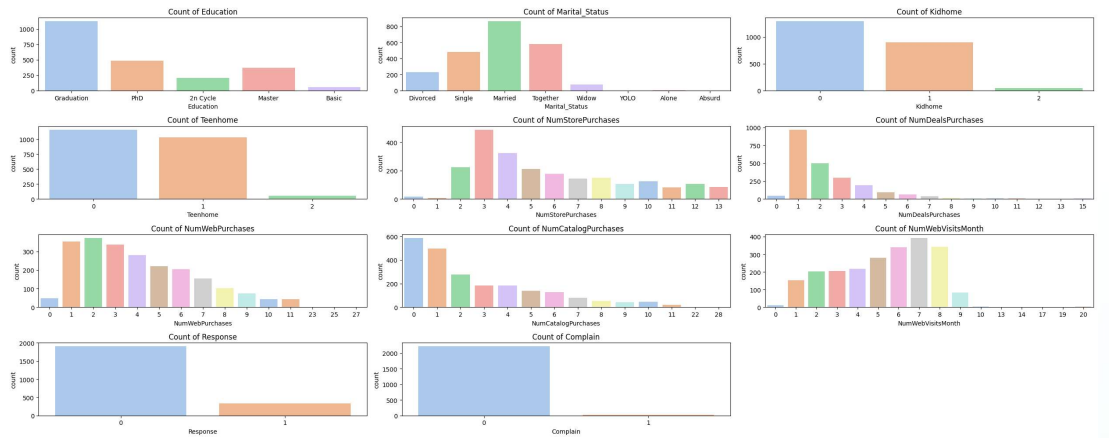
## Bivariate Analysis wrt Response
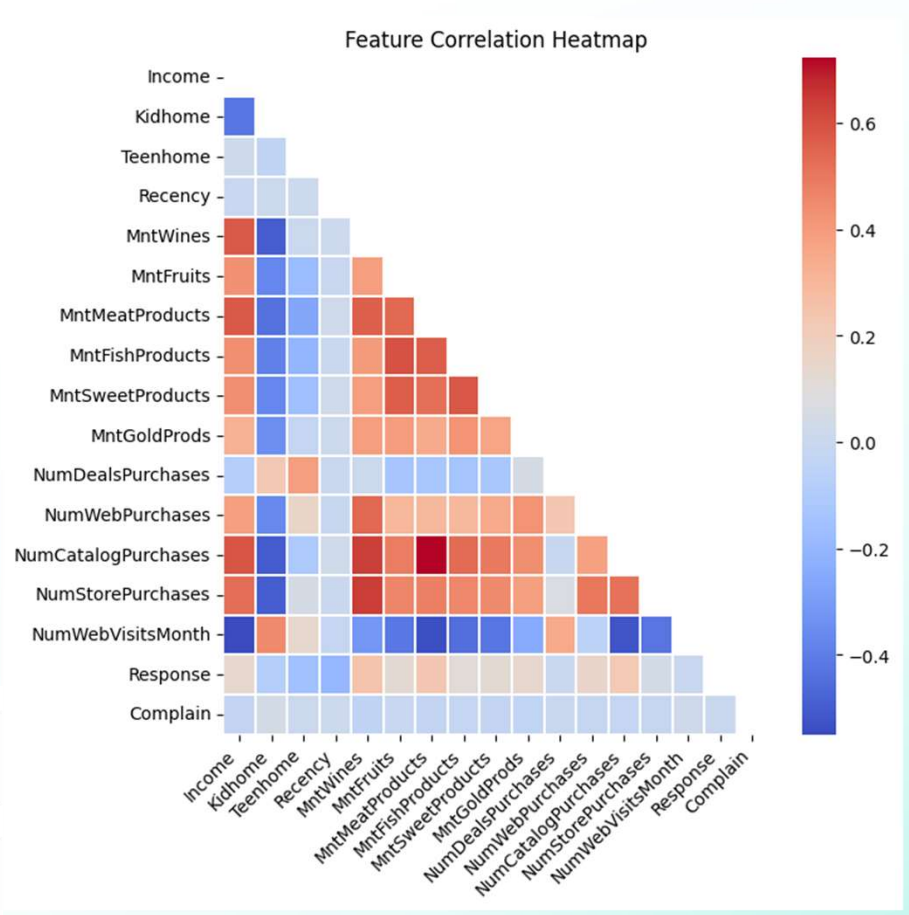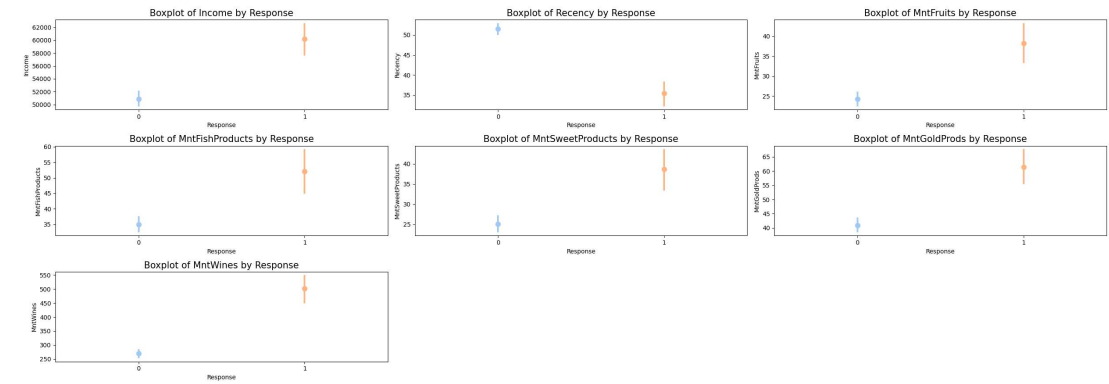
A high positive correlation between meat products and catalog purchases. A strong negative correlation between web visits and income, web visits and meat products, web visits and catalog purchases, wines and kids, income and kids.

Count of Deals Purchases and Total Purchases by Res...

For positive responses, almost all bought products with deals

Number of Purchases via Different Channels and Sum of Response by Year and Quarter

● Catalog Purchases ● Store Purchases ● Web Purchases ● Deals Purchases ● Sum of Response
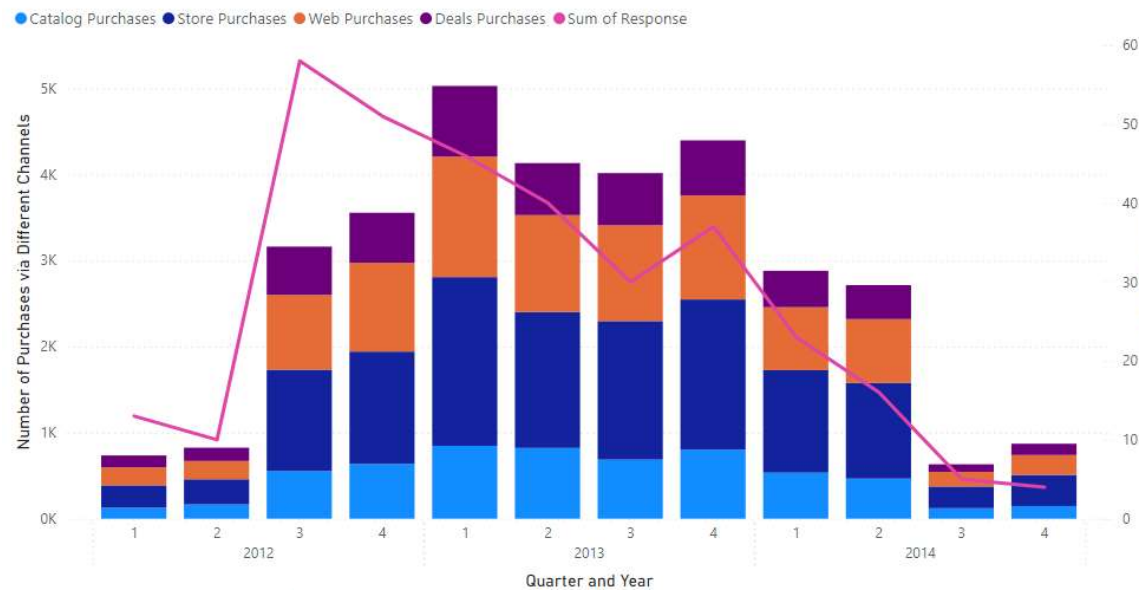
Number of Purchases via Different Channels by Year

● Catalog Purchases ● Store Purchases ● Web Purchases
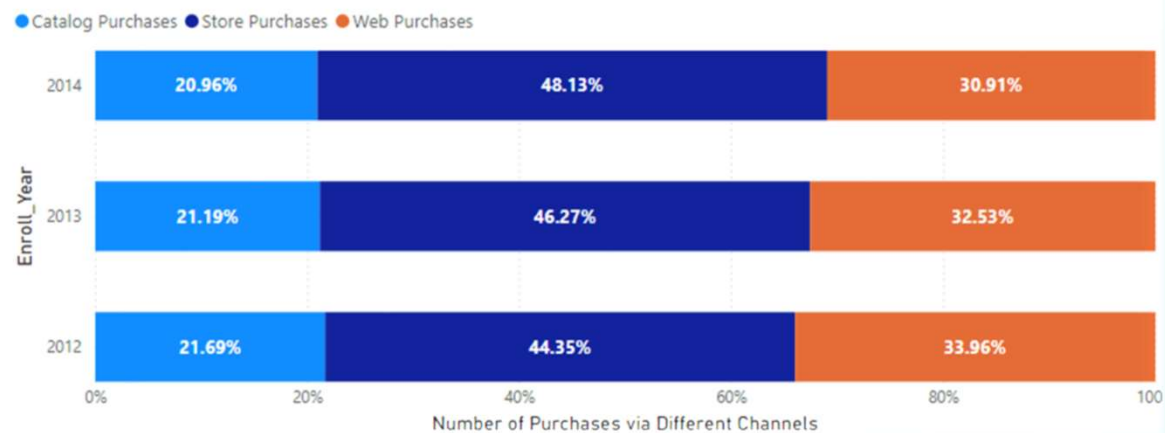
XGBoost Model

Hyperparameter grid for GridSearchCV for XGBoost model

```
param_grid_xgb = {
    'classifier_xgb__learning_rate': [0.01, 0.1, 0.2],
    'classifier_xgb__max_depth': [3, 5, 7, 10],
    'classifier_xgb__min_child_weight': [1, 3, 5],
    'classifier_xgb__gamma': [0, 0.1, 0.2],
    'classifier_xgb__subsample': [0.5, 0.7, 1],
    'classifier_xgb__colsample_bytree': [0.5, 0.7, 1]
}
```

Best hyperparameters for XGBoost model

```
XGB best params: {'classifier_xgb__colsample_bytree': 1,
'classifier_xgb__gamma': 0,
'classifier_xgb__learning_rate': 0.2,
'classifier_xgb__max_depth': 7,
'classifier_xgb__min_child_weight': 1,
'classifier_xgb__subsample': 1}
```

XGBoost Pipeline
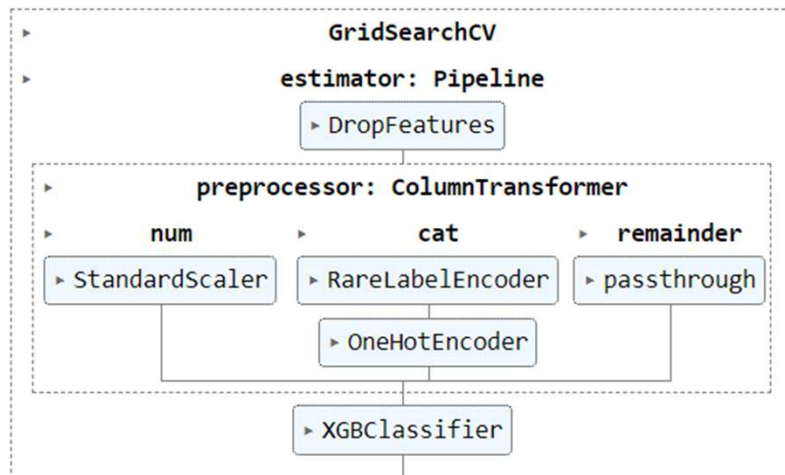
*XGBoost Performance after feature selection*

```
Best Model with Feature Selection Classification Report

              precision    recall  f1-score   support

          0       0.93      0.95      0.94       376
          1       0.68      0.58      0.63        67

   accuracy                           0.90       443
  macro avg       0.81      0.77      0.78       443
weighted avg       0.89      0.90      0.89       443
```

Clustering

1. *Using clustering we could identify the patterns – this chart is for all the features including new ones created – could identify a clear pattern*

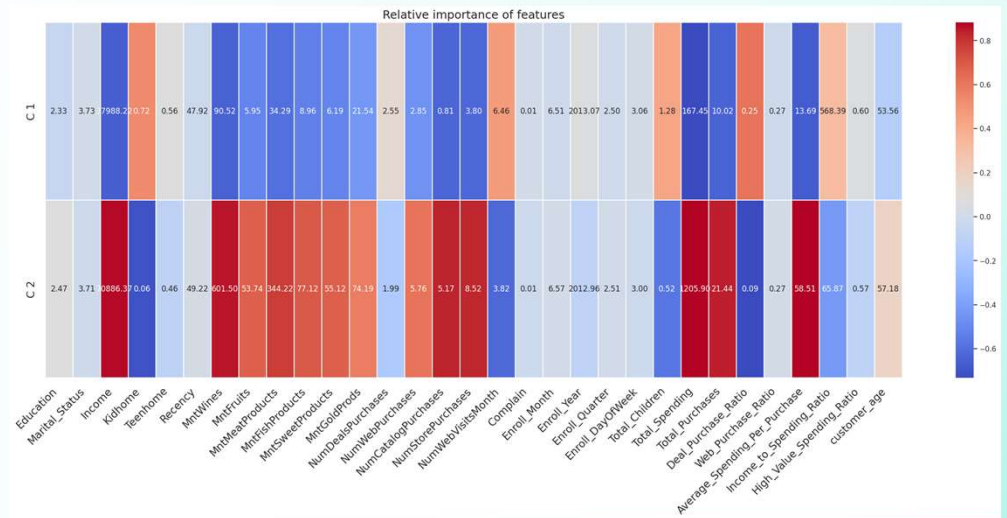Snake Plot, Hierarchical Clustering, K=6
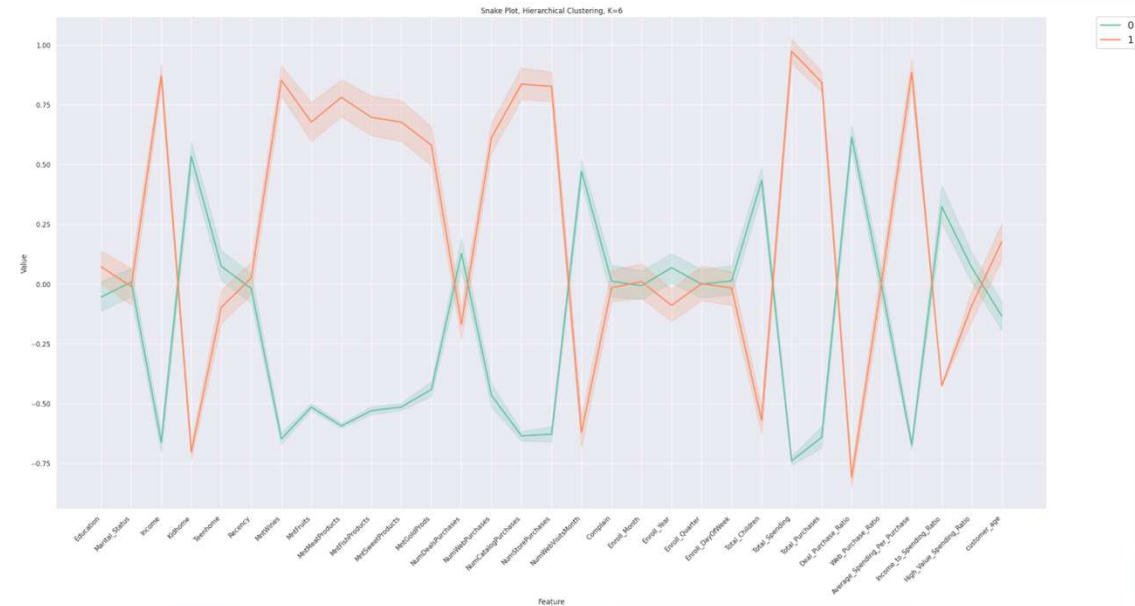
```
Cluster 0:
Response 0 count: 901, Percentage: 89.65%
Response 1 count: 104, Percentage: 10.35%

Cluster 1:
Response 1 count: 162, Percentage: 21.20%
Response 0 count: 602, Percentage: 78.80%
```
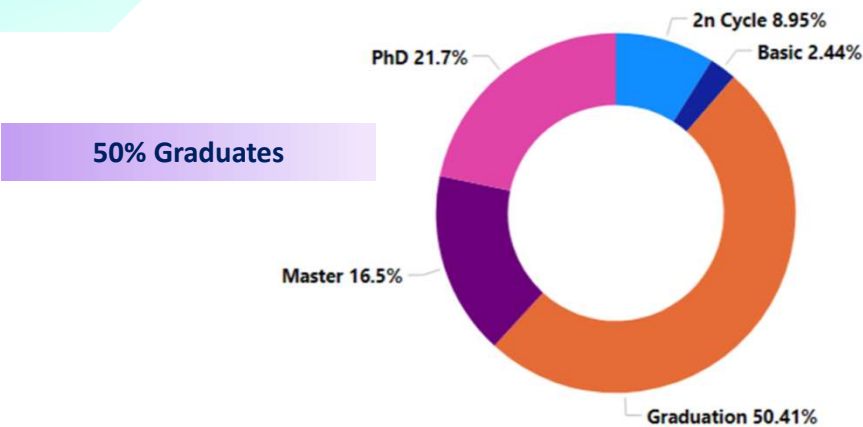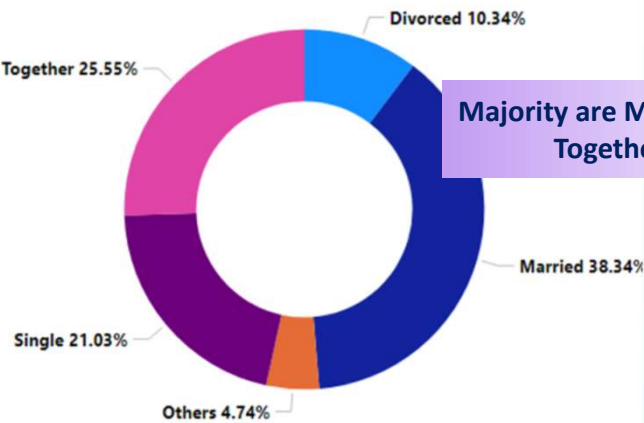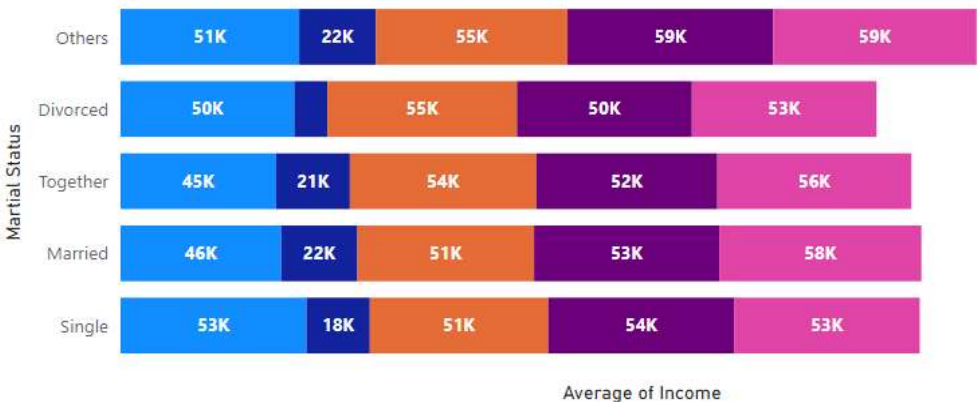

Relative importance of features
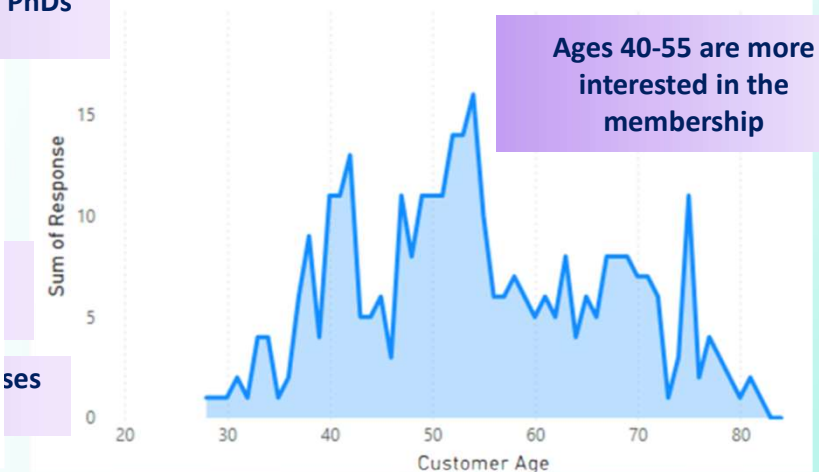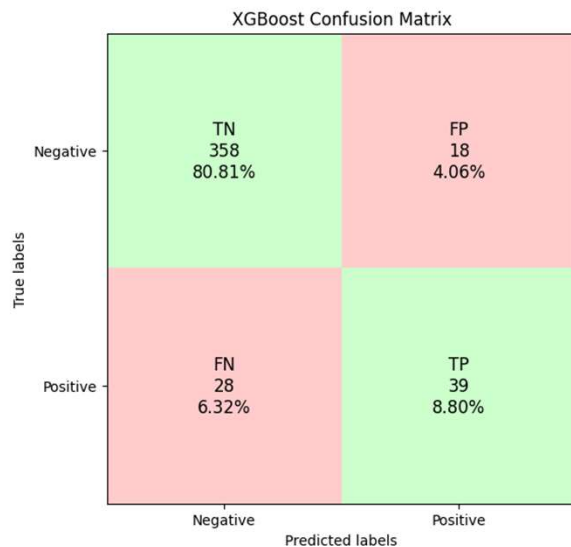
XGBoost Confusion Matrix

How Much Money Does a Model Make/Save?

- Current situation:
  - 2,000 customers
  - 5% churn per month
  - Each churn costs an average of $400
- Cost of current situation: 100*$400 = $40,000
- Churn prediction model:
  - If model predicts Yes, will offer 20% discount offer
    - Average cost of $50
    - Assume everyone accepts offer
  - Current model has below confusion matrix
- How much does the model save them per month?

Confusion Matrix

	Predicted Yes	Predicted No	
Truth Yes	90	10	100
Truth No	285	1615	1900
	375	1625	

⊙

Cost Matrix
Cost per transaction

	Predicted Yes	Predicted No
Truth Yes	$50	$400
Truth No	$50	-

=

Total Cost
$22,750

	Predicted Yes	Predicted No
Truth Yes	$4,500	$4,000
Truth No	$14,250	

Extracted from MMAI 869 – AI and ML