Business Intelligence Capstone Project

Matthew Chapman

The Problem

- Company:
  - Well-established retail food company
  - Serves 100ks of registered customers
  - Sells Products from 5 major categories
  - 3 sales Channels
- Forecasts show profit growth is not promising
- Want to improve performance of marketing activities
  - Emphasis on marketing campaigns
- Low marketing budget

The Solution

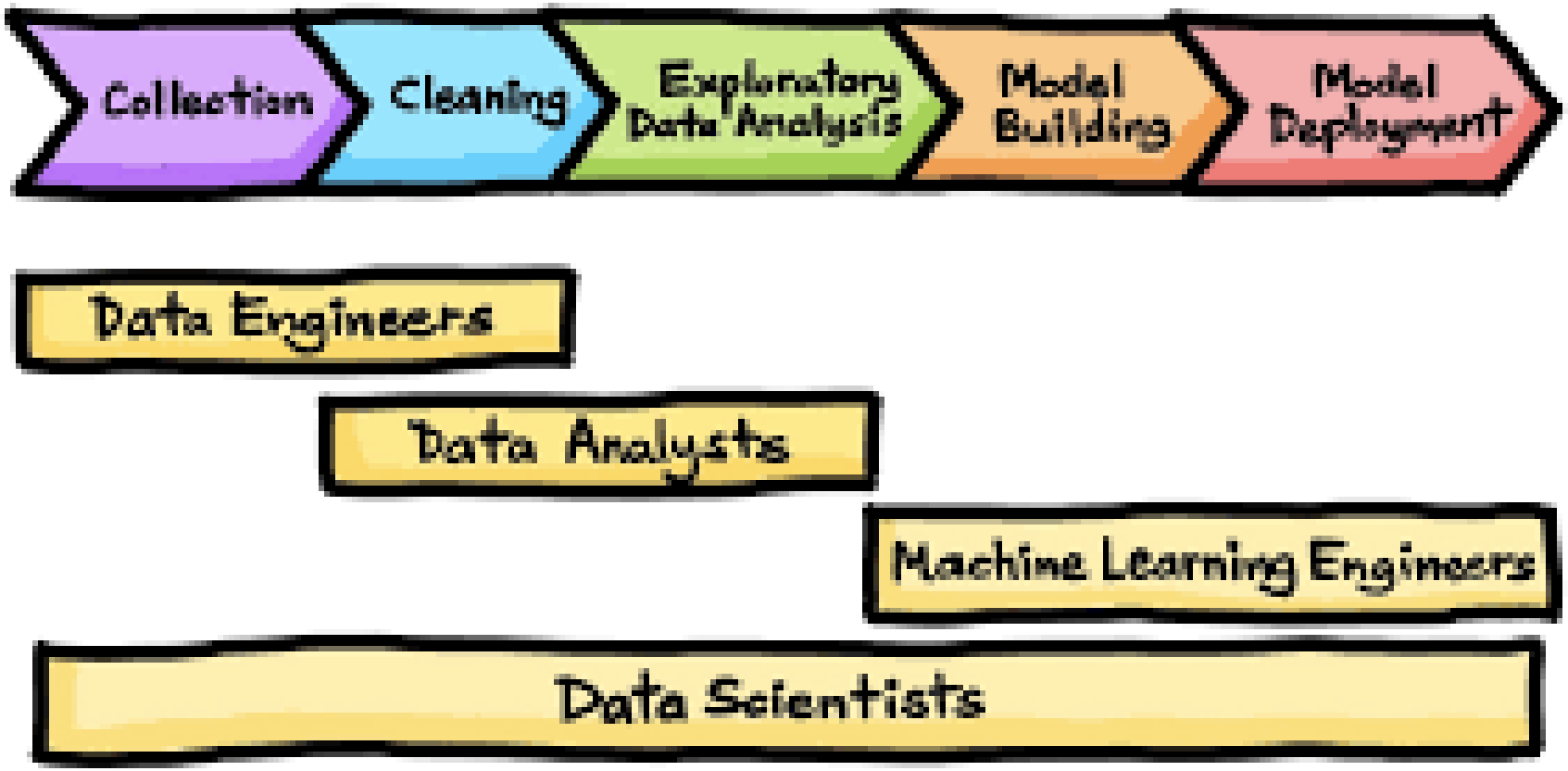
- Accurately predict if customer will accept current campaign offer

-Logistic Regression Model

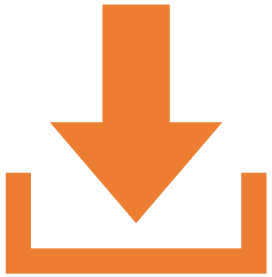
  - Increase accepted offers, minimize marketing costs

- Data:
  - Customer Traits
  - Customer History

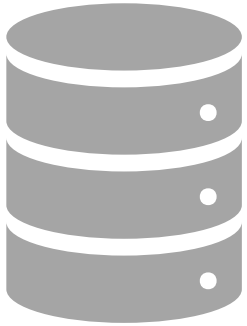
The Process


Collection
Cleaning
Exploratory Data Analysis
Model Building
Model Deployment
Data Engineers
Data Analysts
Machine Learning Engineers
Data Scientists

Data Collection



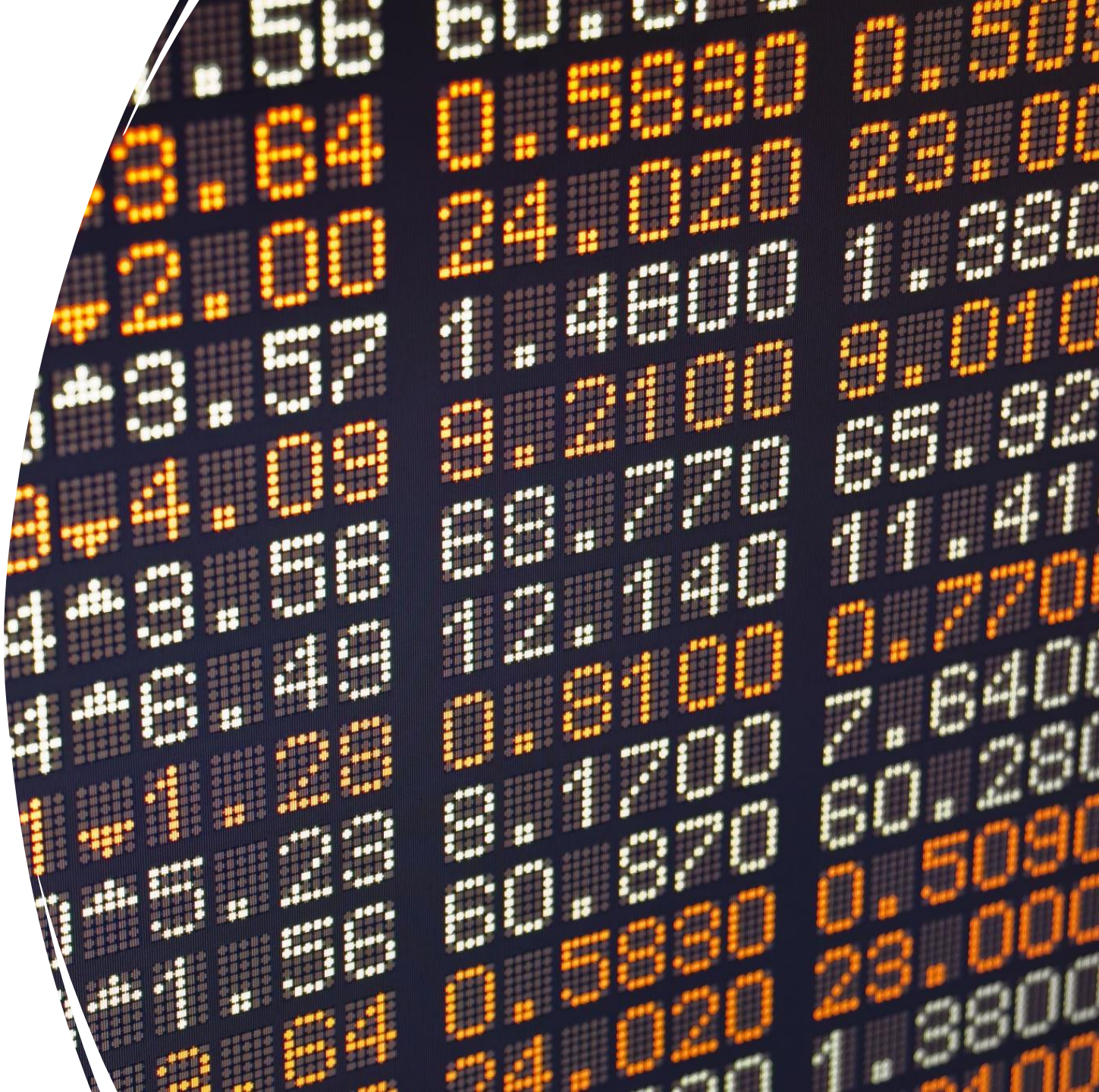
Download Data



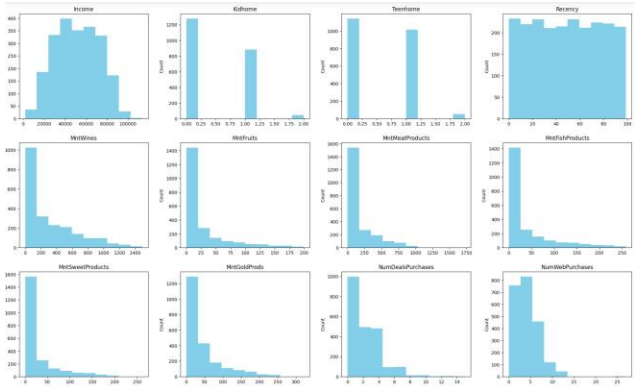
Import Data

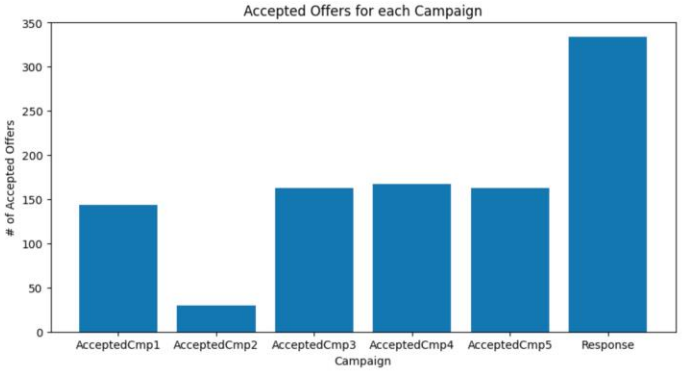
Cleaning/Manipulation

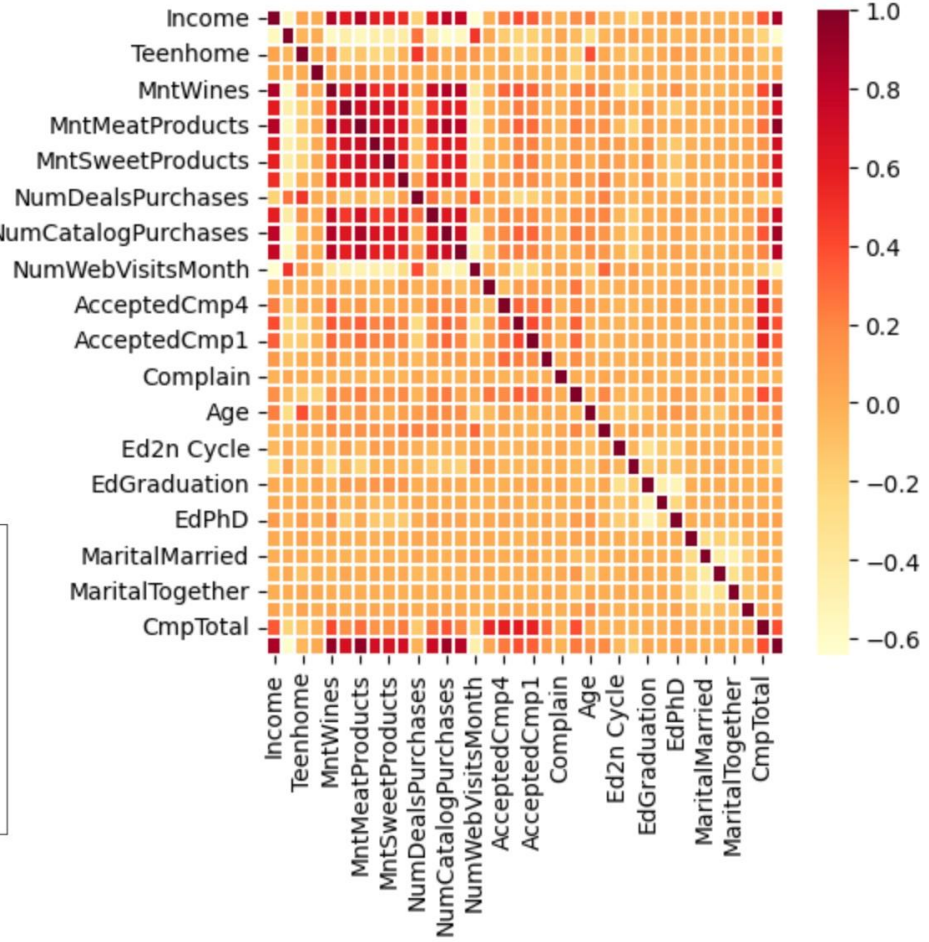
- Convert Categorical Variables to Binary Variables (Dummy Variables)
- Convert Date Variables to Numeric
- Create Summary Variables (Total Revenue, Total Accepted Offers)
- Delete Data with:
  - Null Values
  - Outliers

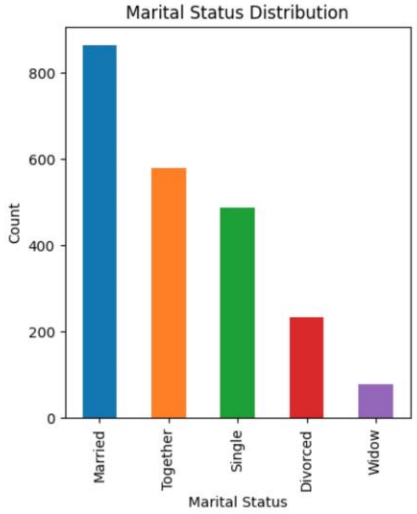


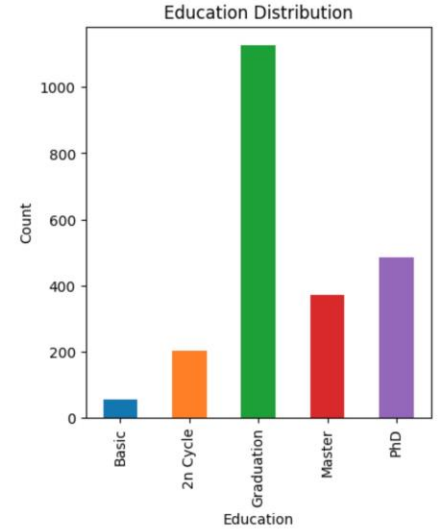
Analysis

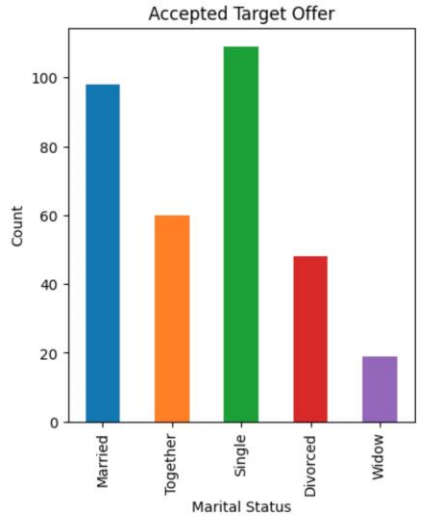

Income
Kidhome
Teenhome
Recency
MntWines
MntFruits
MntMeatProducts
MntFishProducts
MntSweetProducts
MntGoldProds
NumDealsPurchases
NumWebPurchases

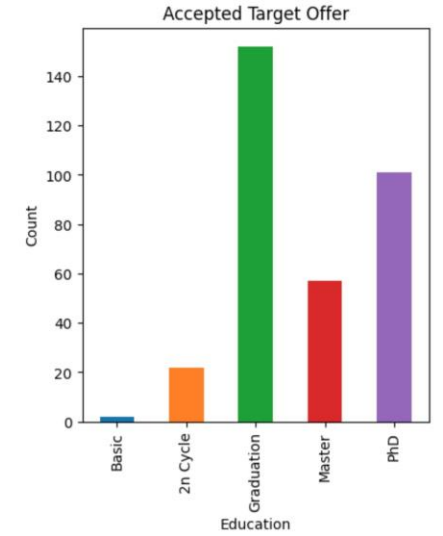

Accepted Offers for each Campaign
# of Accepted Offers
0
50
100
150
200
250
300
350
AcceptedCmp1
AcceptedCmp2
AcceptedCmp3
AcceptedCmp4
AcceptedCmp5
Response
Campaign


Income
Teenhome
MntWines
MntMeatProducts
MntSweetProducts
NumDealsPurchases
NumCatalogPurchases
NumWebVisitsMonth
AcceptedCmp4
AcceptedCmp1
Complain
Age
Ed2n Cycle
EdGraduation
EdPhD
MaritalMarried
MaritalTogether
CmpTotal
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


Marital Status Distribution
Count
0
200
400
600
800
Married
Together
Single
Divorced
Widow
Marital Status


Education Distribution
Count
0
200
400
600
800
1000
Basic
2n Cycle
Graduation
Master
PhD
Education

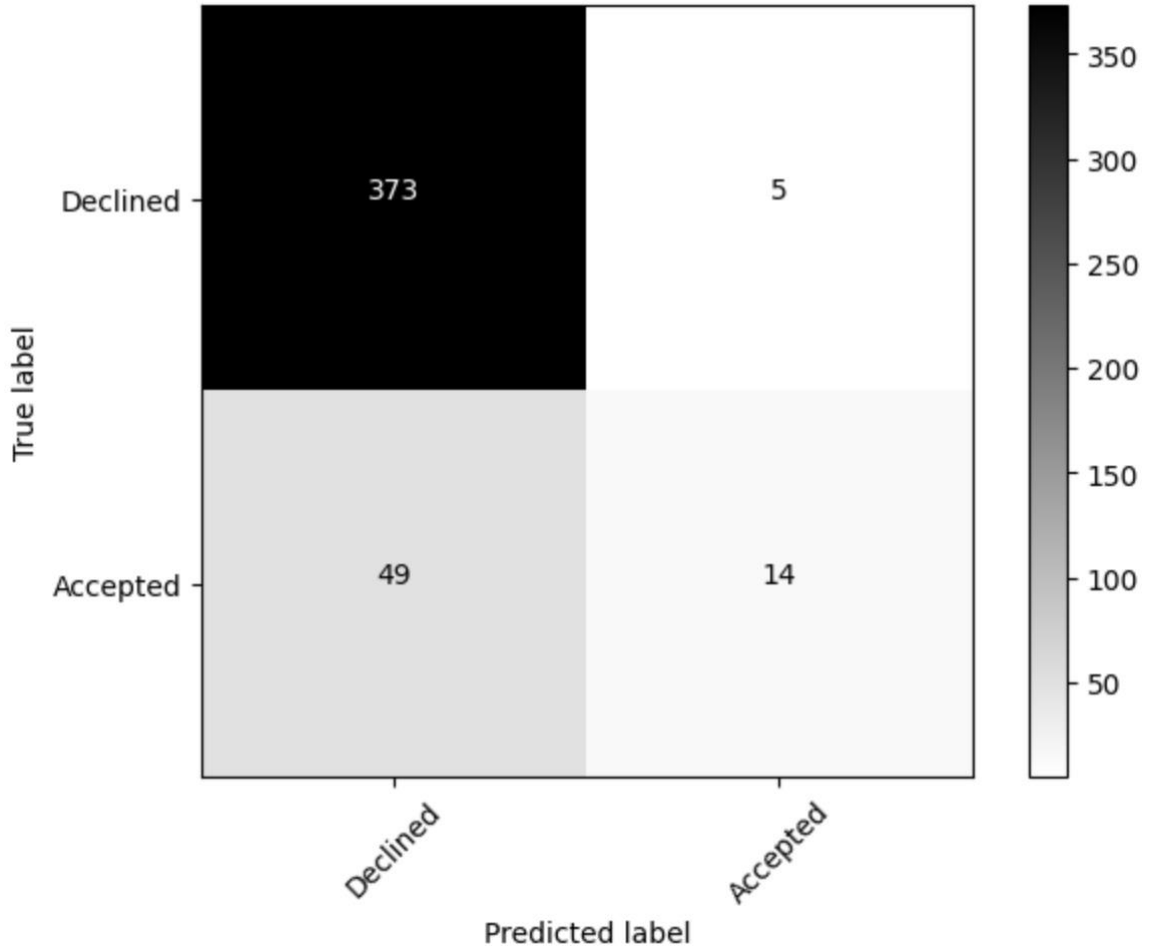

Accepted Target Offer
Count
0
20
40
60
80
100
Married
Together
Single
Divorced
Widow
Marital Status


Accepted Target Offer
Count
0
20
40
60
80
100
120
140
Basic
2n Cycle
Graduation
Master
PhD
Education

Model Building

- Define Variables
- Split Data
- Train Data
- Evaluate Model Metrics

```
Accuracy: 0.8775510204081632
Precision: 0.7368421052631579
Recall: 0.2222222222222222
Specificity: 0.9867724867724867
F1: 0.34146341463414637
AUC_ROC: 0.6044973544973545
```

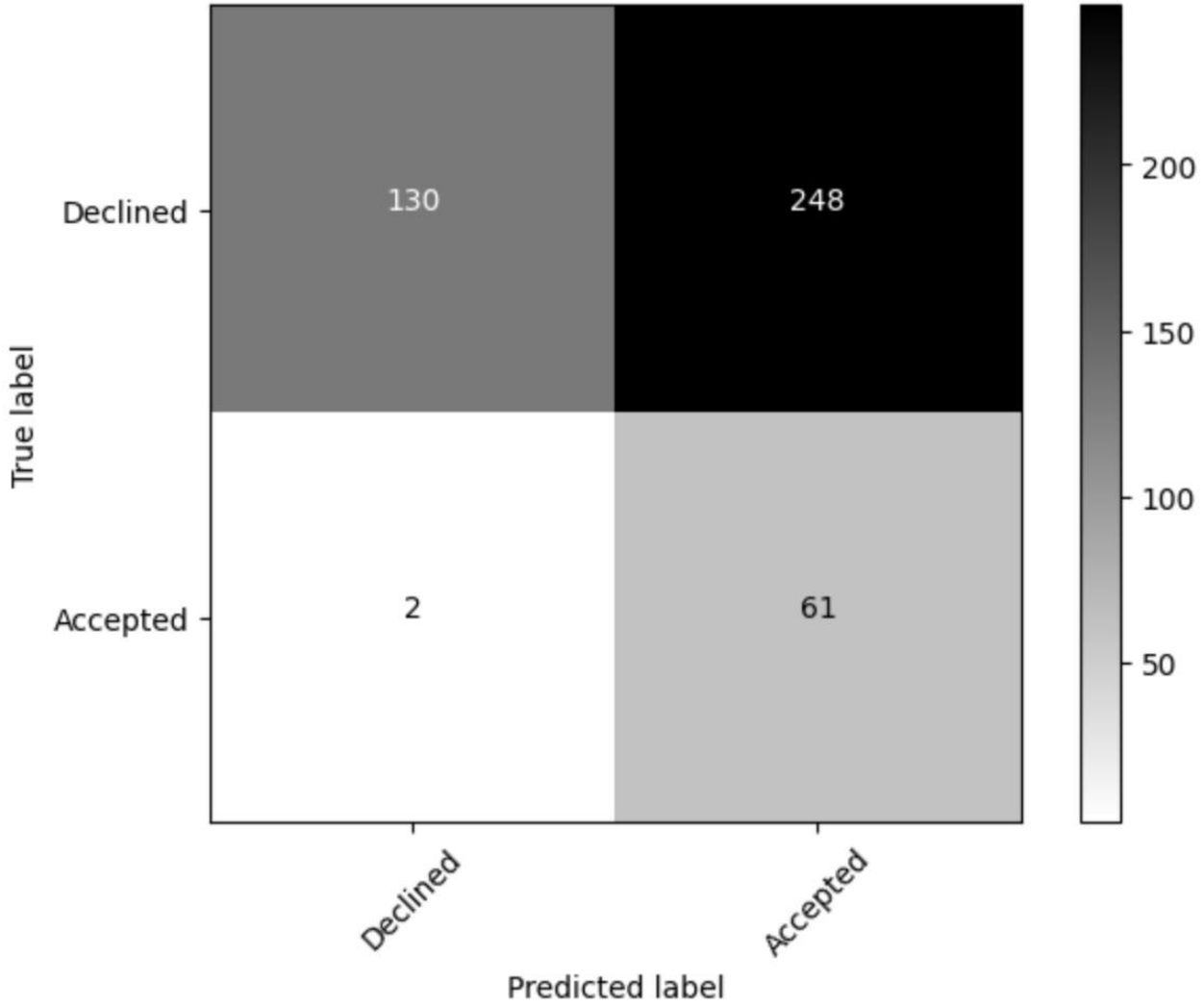


Recall

- We want Recall (True +/All +) to be as high as possible:
  - Don't want to miss out on potential customers
- SMOTE (Synthetic Minority Over-sampling Technique)
- Decision Threshold Adjustments

Recall

Accuracy: 0.4331
Precision: 0.1974
Recall: 0.9683
F1 score: 0.3280
Specificity: 0.3439
AUC-ROC: 0.6561



Recommendations

- Use the model with the highest recall
- Maintain nearly the same # of accepted offers
- Save 30% on marketing expenses