MISM 6210 – Info Visuals and Dashboard for Business
Final Team Project (TP: 25%)
Spring 2024

Tourism Package Marketing Strategy Optimization Project Report

Research Problem

The research addresses the optimization of marketing strategies for wellness tourism packages. The primary question is how customer satisfaction metrics and engagement levels (such as the number of follow-ups) influence the likelihood of purchasing these packages. Understanding these factors is vital for enhancing marketing efficiency, improving customer retention, and boosting sales.

Introduction

The tourism industry faces significant challenges in converting inquiries into sales. Previous studies have shown that targeted marketing strategies, based on customer behaviour and satisfaction, can dramatically increase conversion rates. This project builds on existing research by applying advanced data analytics to uncover deeper insights into customer preferences and behaviour.

Key Columns

1. **CustomerID:** Unique identifier for each customer.
2. **ProdTaken:** Binary value indicating whether the product was purchased (1) or not (0).
3. **Age:** Age of the customer.
4. **TypeofContact:** How the contact was initiated (e.g., 'Self Enquiry', 'Company Invited').
5. **CityTier:** Represents the tier of the city from where the customer is (1, 2, 3).
6. **DurationOfPitch:** Duration in minutes of the pitch made to the customer.
7. **Occupation:** Customer's occupation (e.g., 'Salaried', 'Free Lancer').
8. **NumberOfPersonVisiting:** Number of people visiting.
9. **NumberOfFollowups:** Number of follow-ups made after the initial contact.
10. **ProductPitched:** Type of product pitched to the customer.
11. **PreferredPropertyStar:** Star rating of preferred properties.
12. **MaritalStatus:** Marital status of the customer.
13. **NumberOfTrips:** Number of trips the customer usually takes.
14. **Passport:** Whether the customer has a passport (1) or not (0).
15. **PitchSatisfactionScore:** Customer's satisfaction score from the pitch.
16. **OwnCar:** Whether the customer owns a car (1) or not (0).
17. **NumberOfChildrenVisiting:** Number of children visiting with the customer.
18. **Designation:** Customer's professional designation.
19. **MonthlyIncome:** Customer's monthly income.

Data and Methodology

Data Source and Nature: The data comes from an internal customer database, structured in a CSV format. It includes various attributes such as customer demographics, contact types, engagement metrics, and historical purchase information.

Data Preparation:

Handling Missing Values: The dataset was first cleansed of missing values, particularly in critical columns like Age, Monthly Income, and Duration of Pitch. Techniques such as imputation were used to fill in these gaps.

Data Splitting: The dataset was divided into training and testing subsets to validate the model's predictions effectively.

Feature Selection: Statistical tests like the Chi-squared test were employed to identify significant categorical predictors. Similarly, ANOVA was used for numerical features to determine their impact on the target variable.

Methodology:

Exploratory Data Analysis (EDA): Involved visualizing distributions and relationships in the data to understand the underlying patterns.

Predictive Modelling: Utilized the XGBoost algorithm to create a predictive model capable of assessing the likelihood of a customer purchasing a wellness package.

Threshold Adjustment: The decision threshold of the model was fine-tuned to optimize the balance between precision and recall, enhancing the model's practical applicability.

Scope

The scope includes analysing an internal customer database to identify characteristics that predict package purchase likelihood. This analysis will guide the development of targeted marketing strategies, potentially transforming the company's approach to customer engagement.

Visualizations

Refer Exhibits

Insights From the analysis

Analysis revealed a clear correlation between the number of follow-ups and customer satisfaction scores. This observation was quantitatively supported by the use of a box plot visualization that depicted pitch satisfaction scores across different follow-up frequencies. Statistical tests such as the ANOVA (Analysis of Variance) further substantiated this correlation, showing that increases in the number of follow-ups typically lead to higher satisfaction scores.

- **Statistical Evidence:** The mean pitch satisfaction score increased with the number of follow-ups. For instance, customers with more than five follow-ups showed a noticeable improvement in their satisfaction scores compared to those with fewer follow-ups.
- **Methodological Approach:** The relationship was investigated using descriptive statistics and visual tools like box plots to graphically represent data distributions, which highlighted the median, quartiles, and outliers in pitch satisfaction across different levels of customer engagement.

Model Predictions

The XGBoost model, a powerful machine learning algorithm known for its performance in classification tasks, was employed to identify the factors most predictive of a customer purchasing a wellness package. This model was

trained using features identified as significant through exploratory data analysis and statistical tests such as Chi-squared for categorical variables and ANOVA for continuous variables.

- **Key Predictors Identified:**
    - Age: Younger customers, particularly those aged 25-35, were more likely to purchase wellness packages.
    - Income: Customers with middle-range incomes (e.g., ₹15,000 - ₹25,000 per month) showed a higher propensity to purchase.
    - Passport Possession: Owning a passport was a strong predictor, suggesting that those who already travel internationally are good targets for such packages.
- **Model Validation:** The model's predictions were validated using a split of training and testing data, ensuring that the insights derived were robust and not overfitted to the training dataset. The predictive accuracy was measured using the AUROC (Area Under the Receiver Operating Characteristic) curve, where a higher score indicates better discriminative ability of the model.

Data-Driven Segmentation

The segmentation analysis was grounded on the model's ability to classify potential buyers effectively. Using clustering techniques and the model's predictive output, customers were segmented into groups based on their likelihood to purchase, informed by their demographic and behavioural characteristics.

- **Segment Characteristics:**
    - Young Professionals: Typically, in their mid-20s to mid-30s, often possessing a passport, and showing a clear preference for high-end travel experiences.
    - Middle-Income Earners: Demonstrated a specific interest in wellness packages, potentially due to a combination of disposable income and a lifestyle that values wellness.
- **Statistical Techniques Used**: Clustering algorithms alongside predictive modelling helped to differentiate these segments based on their scores and characteristics defined by the model.

Recommendations

1. Enhance Customer Engagement Strategies: Given the positive correlation between the number of follow-ups and customer satisfaction, it's advisable to:

- Increase Follow-Up Contacts: Develop a structured follow-up strategy where potential customers are contacted more frequently and systematically to increase engagement and conversion rates. Tailor the content of these interactions to reflect customers' previous interactions and feedback.
- Leverage Automation: Utilize automated marketing tools to schedule and send follow-up messages, reminders, or newsletters. Ensure that these tools integrate personalized touches based on customer data to maintain a high level of personal engagement.

2. Refine Targeting Based on Key Predictors: The XGBoost model identified age, income, and passport possession as significant predictors of purchase likelihood:

- Target Young Professionals: Create tailored marketing campaigns that appeal to young professionals who show a higher propensity to purchase wellness packages. These campaigns could highlight adventure and relaxation elements, aligning with their lifestyle and interests.
- Focus on Middle-Income Segments: Design affordable package options and payment plans for middle-income earners. Marketing messages for this segment should emphasize value, wellness benefits, and the transformative experience of the travel packages.

- Promote Pre-Travel Benefits: For customers identified as having a passport, emphasize the ease of booking and preparing for international travel. Offer special deals or incentives for early bookings to encourage commitment.

3. Utilize Data-Driven Personalization: Segmentation has shown distinct customer groups with specific preferences:

- Customized Packages: Develop and promote customized travel packages based on the identified preferences of different segments. For instance, offer luxury wellness retreats for those interested in high-end experiences and budget-friendly wellness getaways for cost-conscious consumers.
- Dynamic Content Marketing: Use data analytics to create dynamic content that adapts based on the user's past interactions and preferences. This could include personalized emails, targeted ads, and customized web experiences.

4. Implement Continuous Learning and Adaptation: The travel and tourism market is dynamic, and customer preferences can evolve:

- Regular Model Updates: Regularly update the predictive models with new customer data to ensure the insights remain relevant and accurate. This involves retraining the models semi-annually or annually to adapt to new trends and customer behaviours.
- Feedback Loops: Establish mechanisms to capture customer feedback post-trip to refine future offerings and personalize future communications. Use this feedback to continuously improve the customer experience and product offerings.

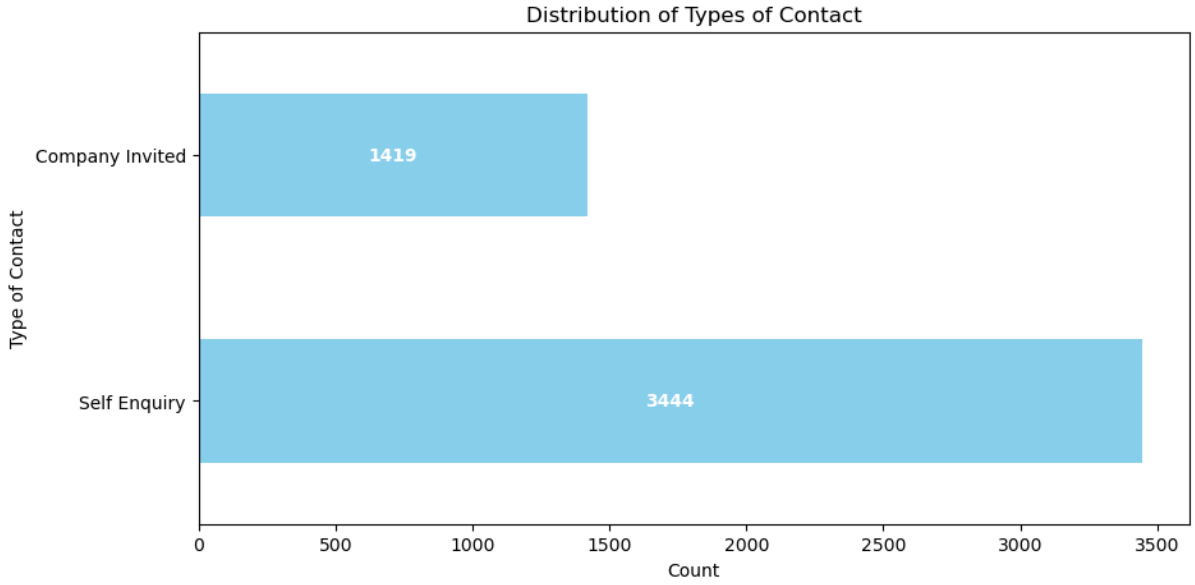
5. Invest in Advanced Analytics Capabilities

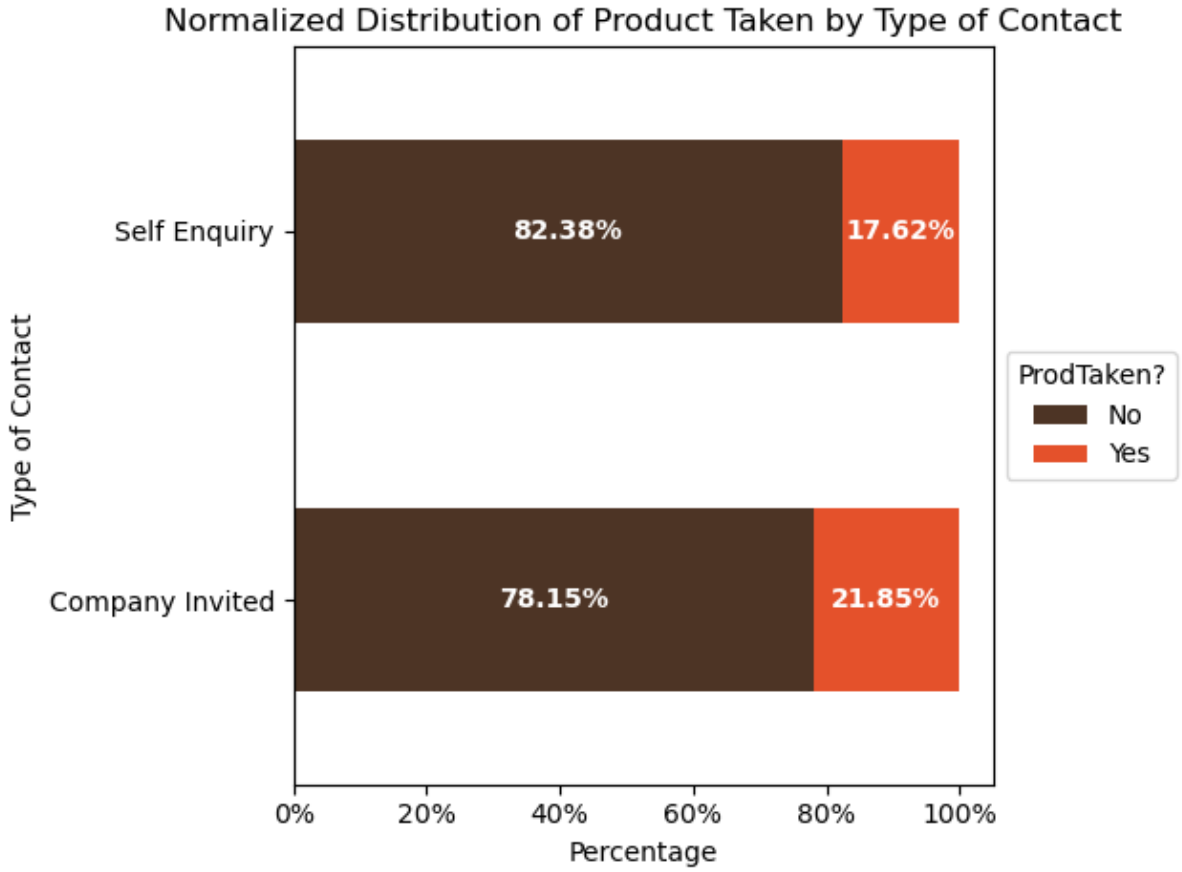
- Advanced Predictive Analytics: Invest in developing or enhancing predictive analytics capabilities to foresee future trends in customer preferences and market conditions.
- Real-time Data Processing: Implement systems that can process and analyse data in real-time to provide immediate insights that can be acted upon swiftly, enhancing responsiveness to market changes.

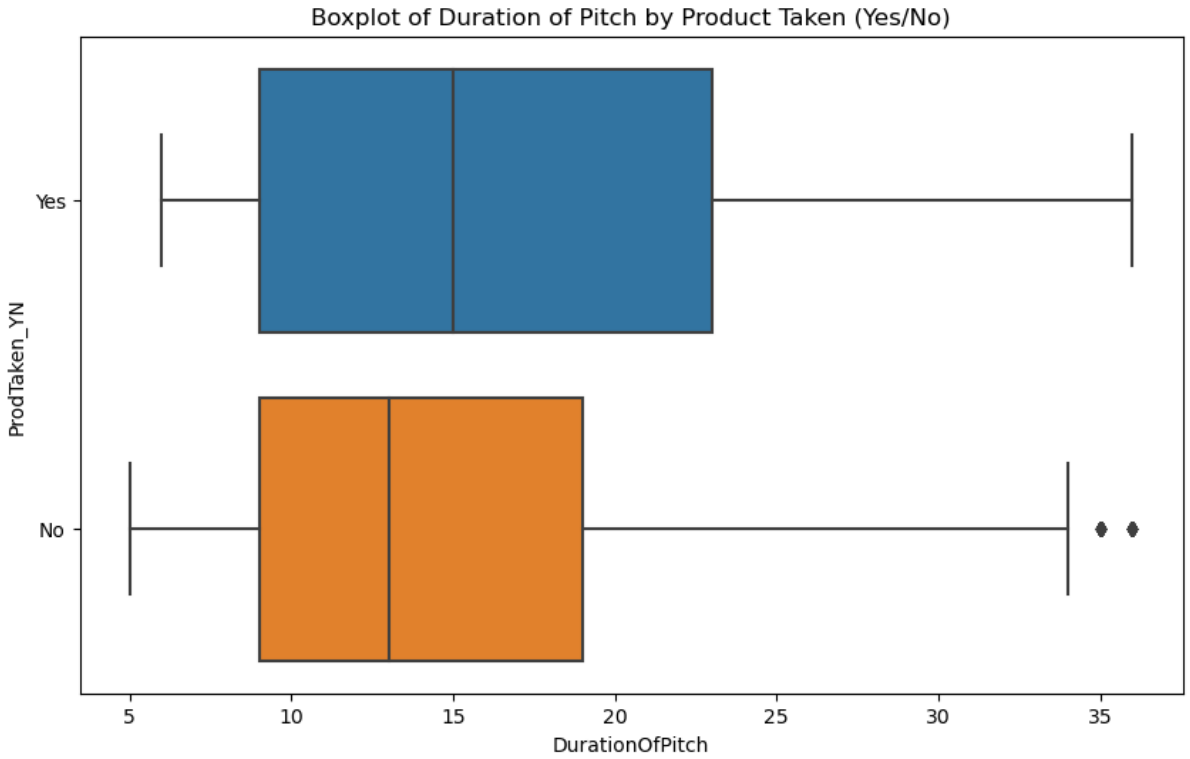
Conclusion

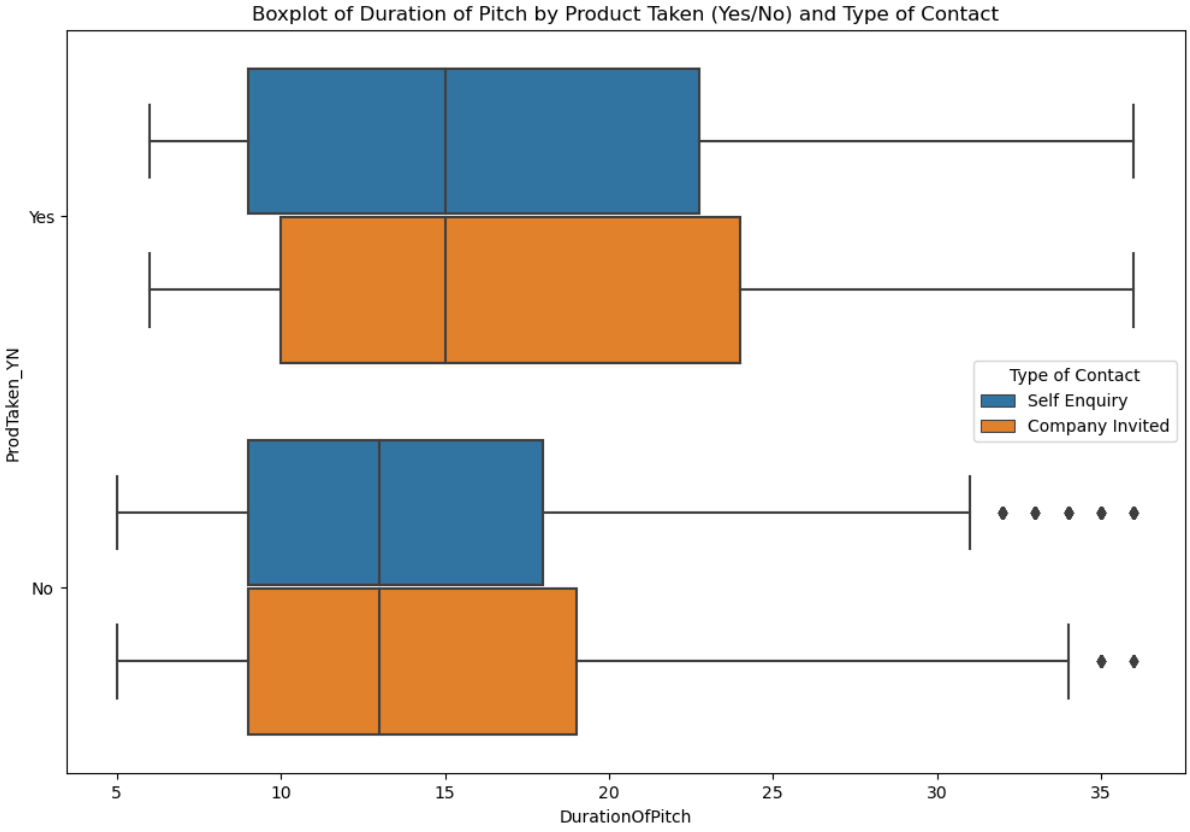
By following these recommendations, the company can not only enhance its marketing strategies but also significantly improve its overall effectiveness in attracting and retaining customers for its wellness tourism packages. This will ultimately lead to increased sales, customer satisfaction, and a stronger competitive position in the marketplace.

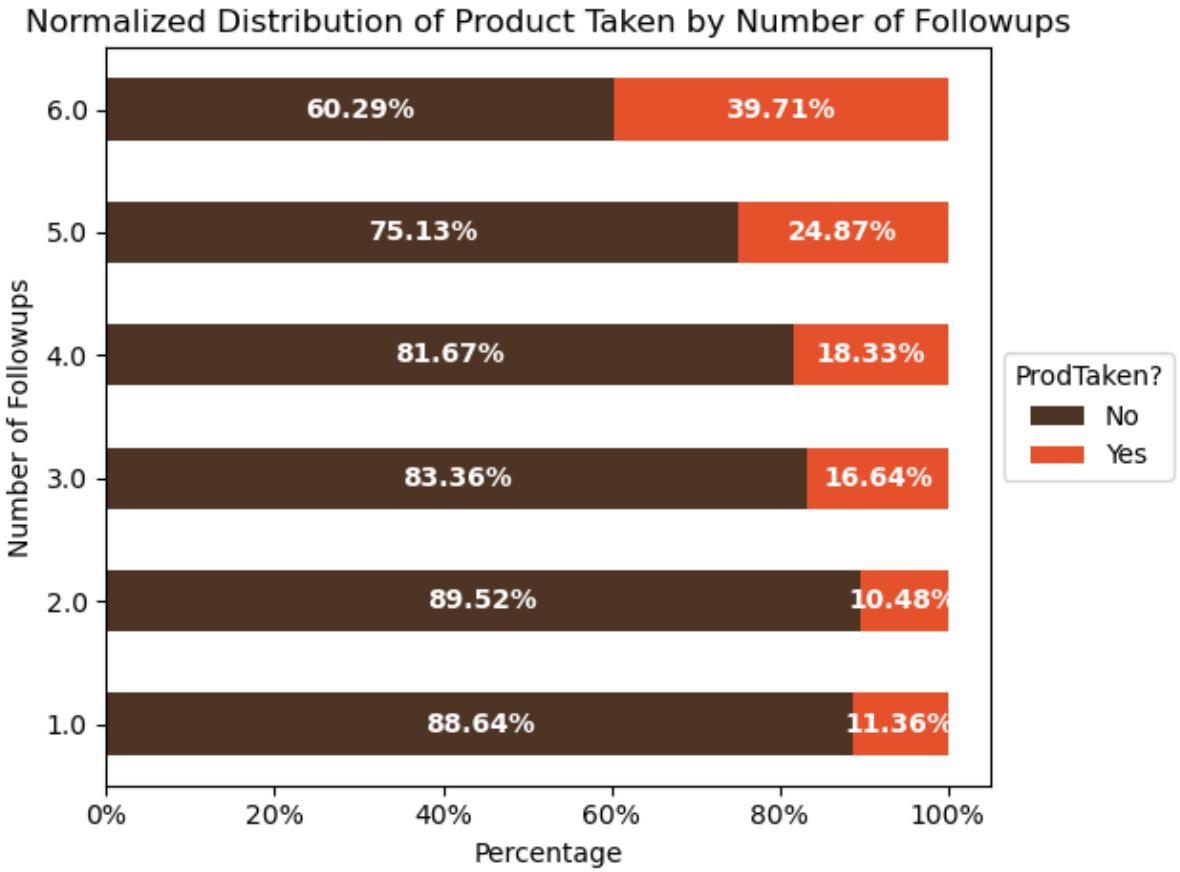
Exhibits

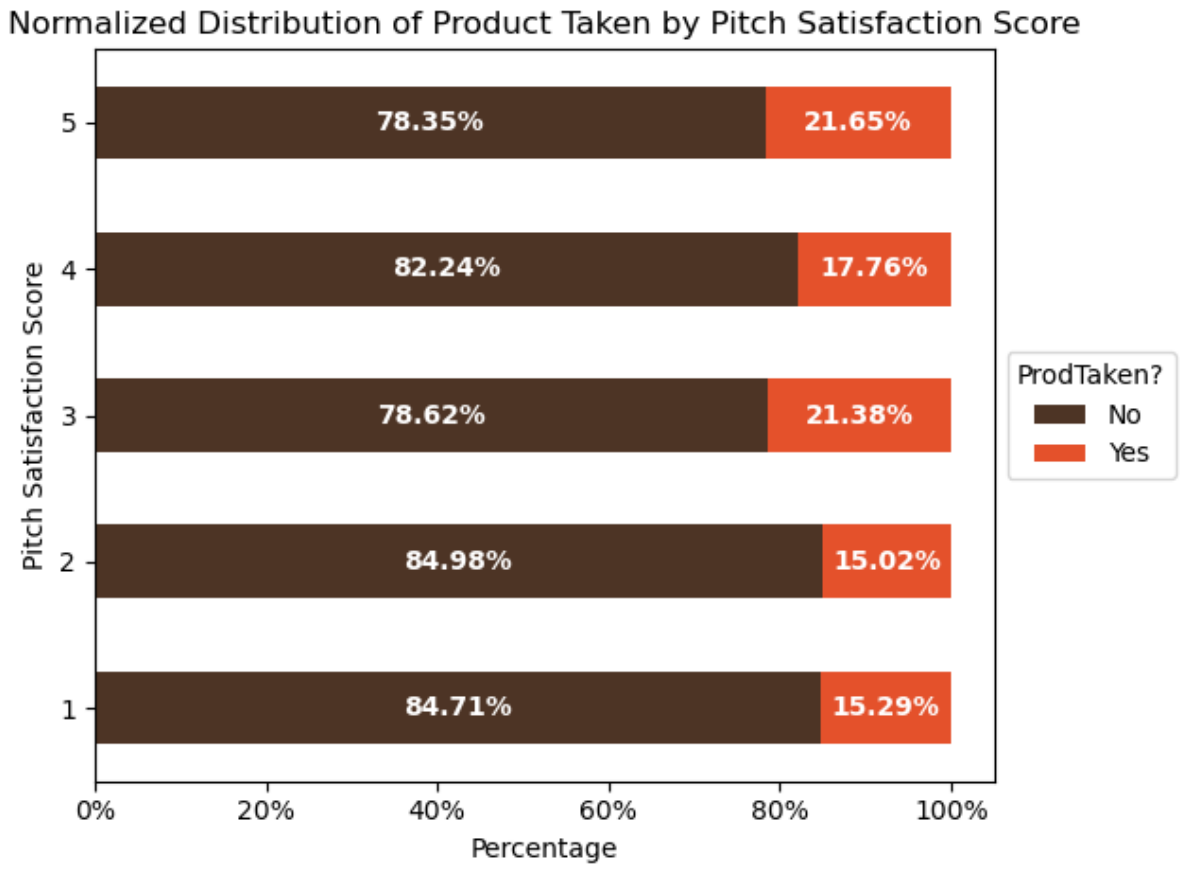

Distribution of Types of Contact
Company Invited
Self Enquiry
1419
3444
Type of Contact
0
500
1000
1500
2000
2500
3000
3500
Count

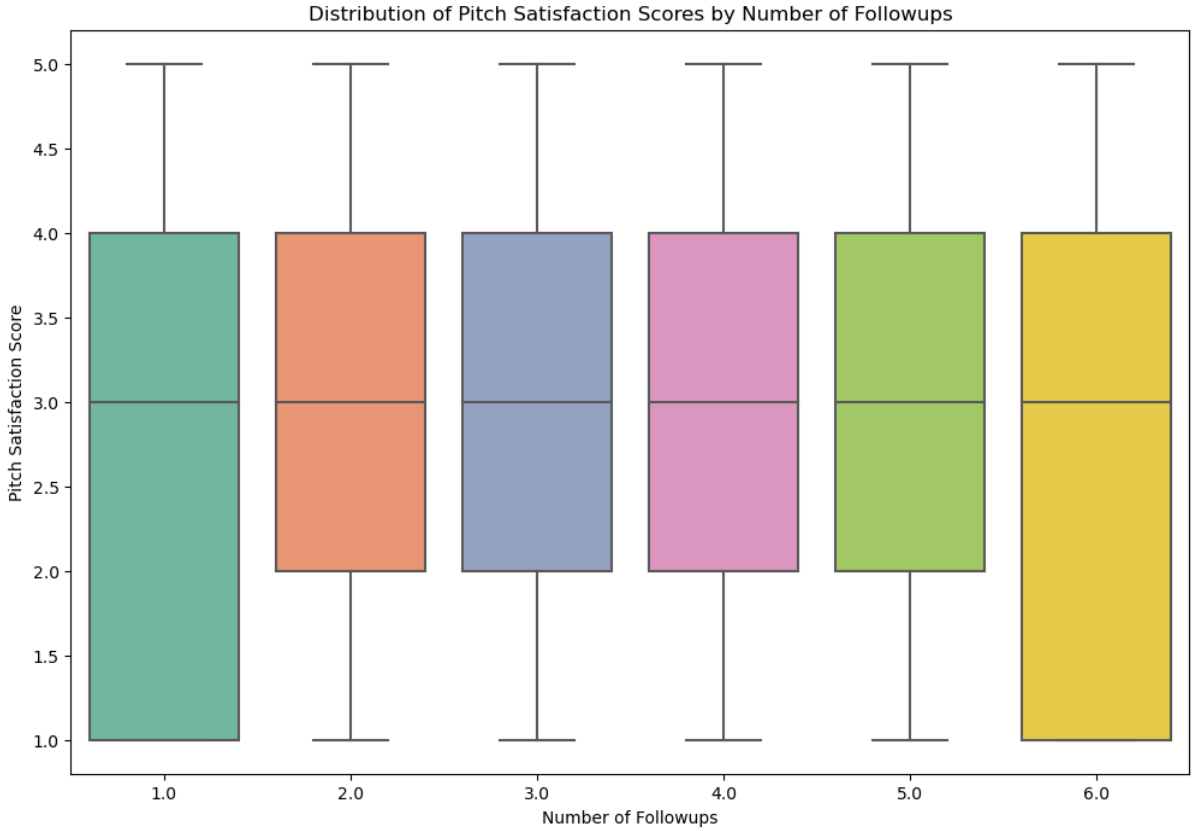

Normalized Distribution of Product Taken by Type of Contact
Self Enquiry
Company Invited
82.38%
17.62%
78.15%
21.85%
Type of Contact
ProdTaken?
No
Yes
0%
20%
40%
60%
80%
100%
Percentage

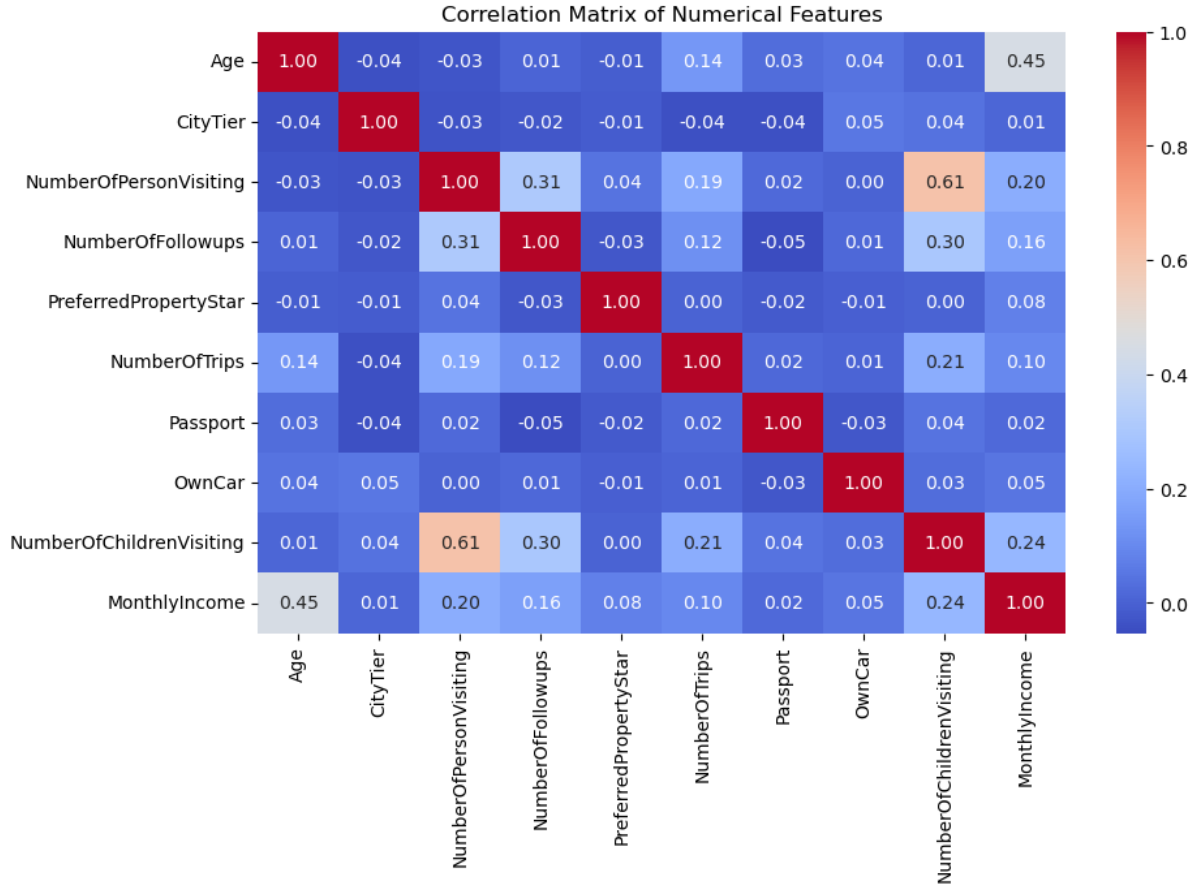

Boxplot of Duration of Pitch by Product Taken (Yes/No)
Yes
No
ProdTaken_YN
5
10
15
20
25
30
35
DurationOfPitch

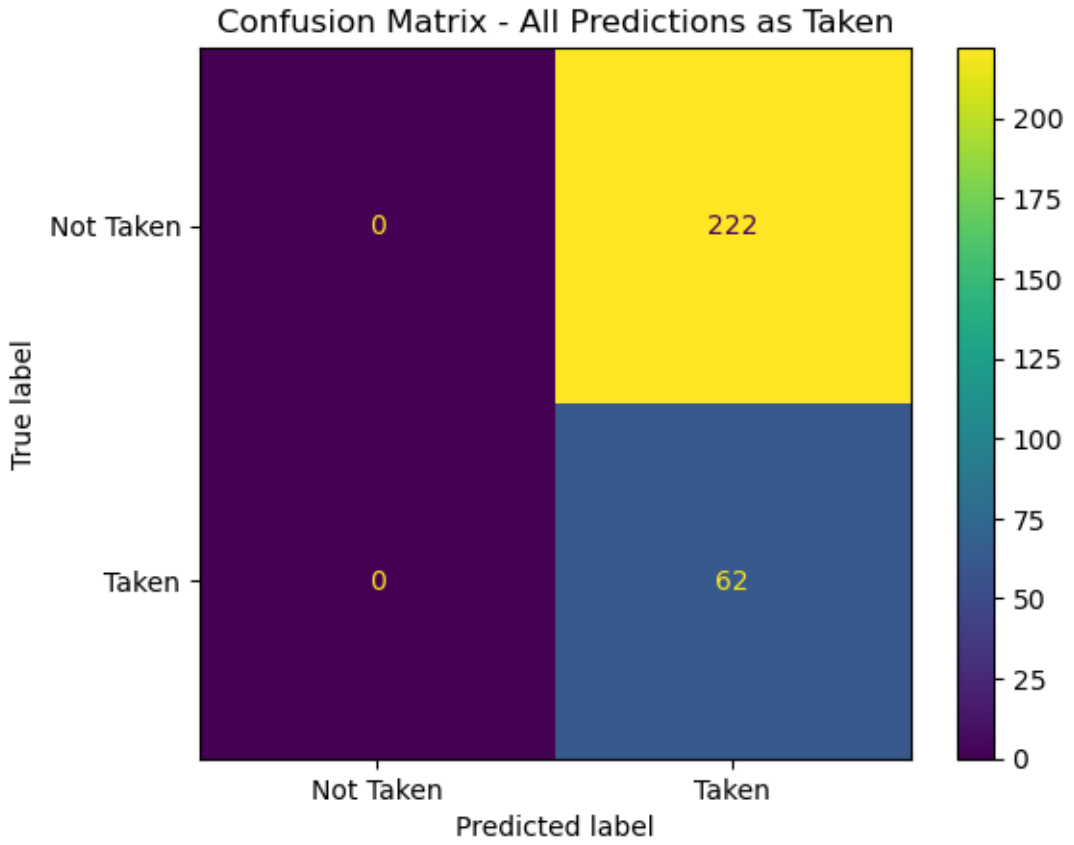

Boxplot of Duration of Pitch by Product Taken (Yes/No) and Type of Contact
Yes
No
ProdTaken_YN
Type of Contact
Self Enquiry
Company Invited
5
10
15
20
25
30
35
DurationOfPitch

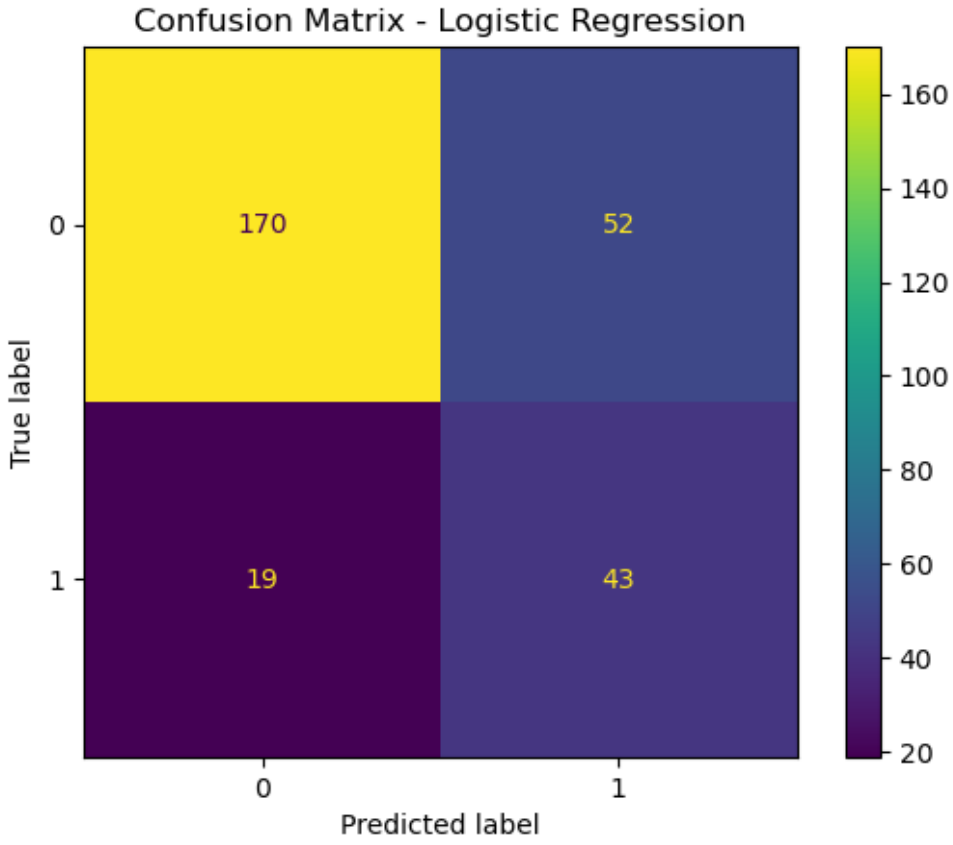

Normalized Distribution of Product Taken by Number of Followups
Number of Followups
6.0
5.0
4.0
3.0
2.0
1.0
60.29%
39.71%
75.13%
24.87%
81.67%
18.33%
83.36%
16.64%
89.52%
10.48%
88.64%
11.36%
0%
20%
40%
60%
80%
100%
Percentage
ProdTaken?
No
Yes

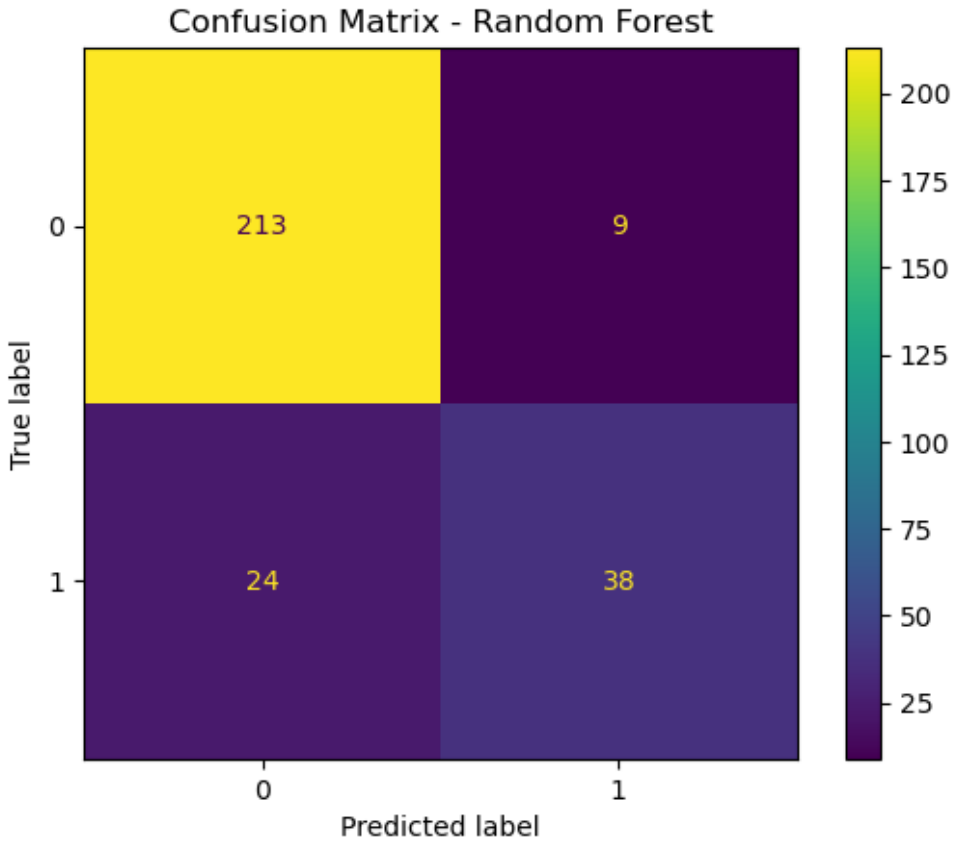

Normalized Distribution of Product Taken by Pitch Satisfaction Score
Pitch Satisfaction Score
5
4
3
2
1
78.35%
21.65%
82.24%
17.76%
78.62%
21.38%
84.98%
15.02%
84.71%
15.29%
0%
20%
40%
60%
80%
100%
Percentage
ProdTaken?
No
Yes

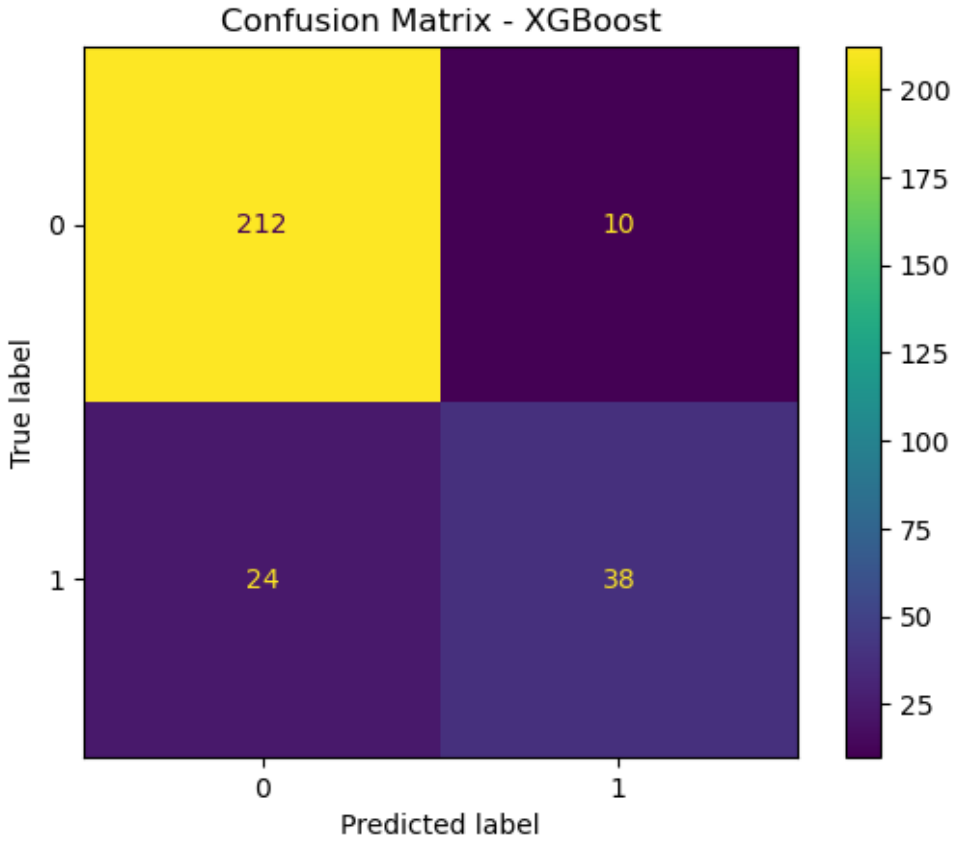

Distribution of Pitch Satisfaction Scores by Number of Followups
Pitch Satisfaction Score
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
1.0
2.0
3.0
4.0
5.0
6.0
Number of Followups

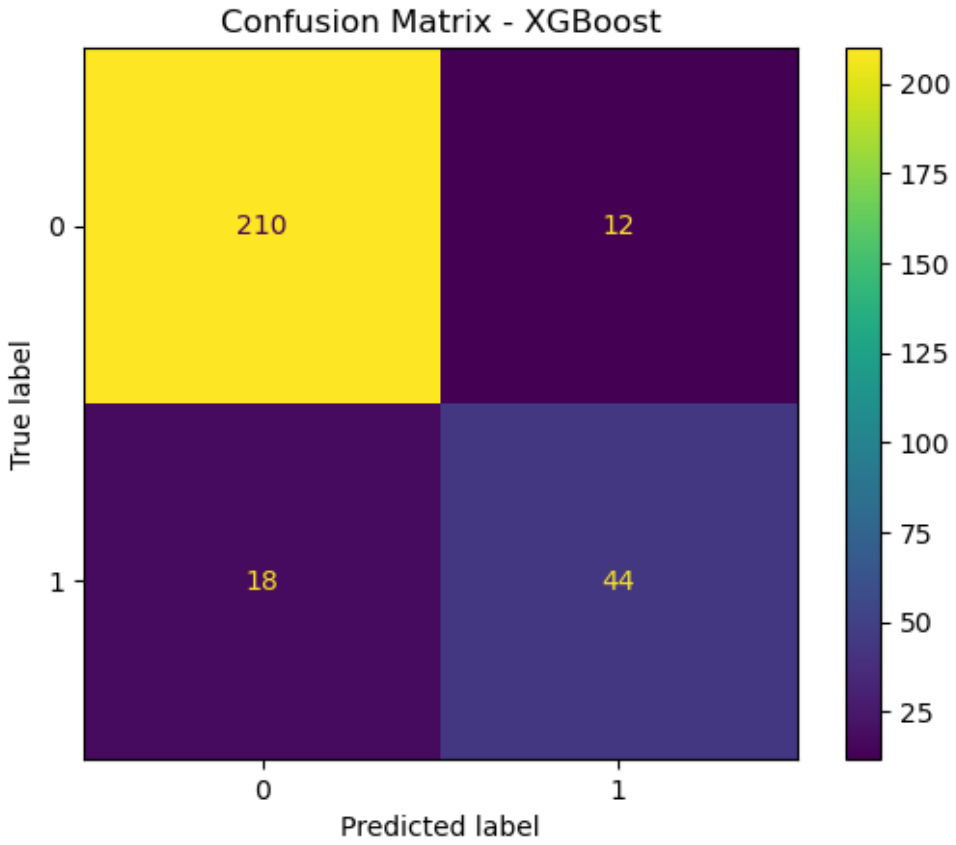

Correlation Matrix of Numerical Features
Age 1.00 -0.04 -0.03 0.01 -0.01 0.14 0.03 0.04 0.01 0.45
CityTier -0.04 1.00 -0.03 -0.02 -0.01 -0.04 -0.04 0.05 0.04 0.01
NumberOfPersonVisiting -0.03 -0.03 1.00 0.31 0.04 0.19 0.02 0.00 0.61 0.20
NumberOfFollowups 0.01 -0.02 0.31 1.00 -0.03 0.12 -0.05 0.01 0.30 0.16
PreferredPropertyStar -0.01 -0.01 0.04 -0.03 1.00 0.00 -0.02 -0.01 0.00 0.08
NumberOfTrips 0.14 -0.04 0.19 0.12 0.00 1.00 0.02 0.01 0.21 0.10
Passport 0.03 -0.04 0.02 -0.05 -0.02 0.02 1.00 -0.03 0.04 0.02
OwnCar 0.04 0.05 0.00 0.01 -0.01 0.01 -0.03 1.00 0.03 0.05
NumberOfChildrenVisiting 0.01 0.04 0.61 0.30 0.00 0.21 0.04 0.03 1.00 0.24
MonthlyIncome 0.45 0.01 0.20 0.16 0.08 0.10 0.02 0.05 0.24 1.00
Age
CityTier
NumberOfPersonVisiting
NumberOfFollowups
PreferredPropertyStar
NumberOfTrips
Passport
OwnCar
NumberOfChildrenVisiting
MonthlyIncome
1.0
0.8
0.6
0.4
0.2
0.0

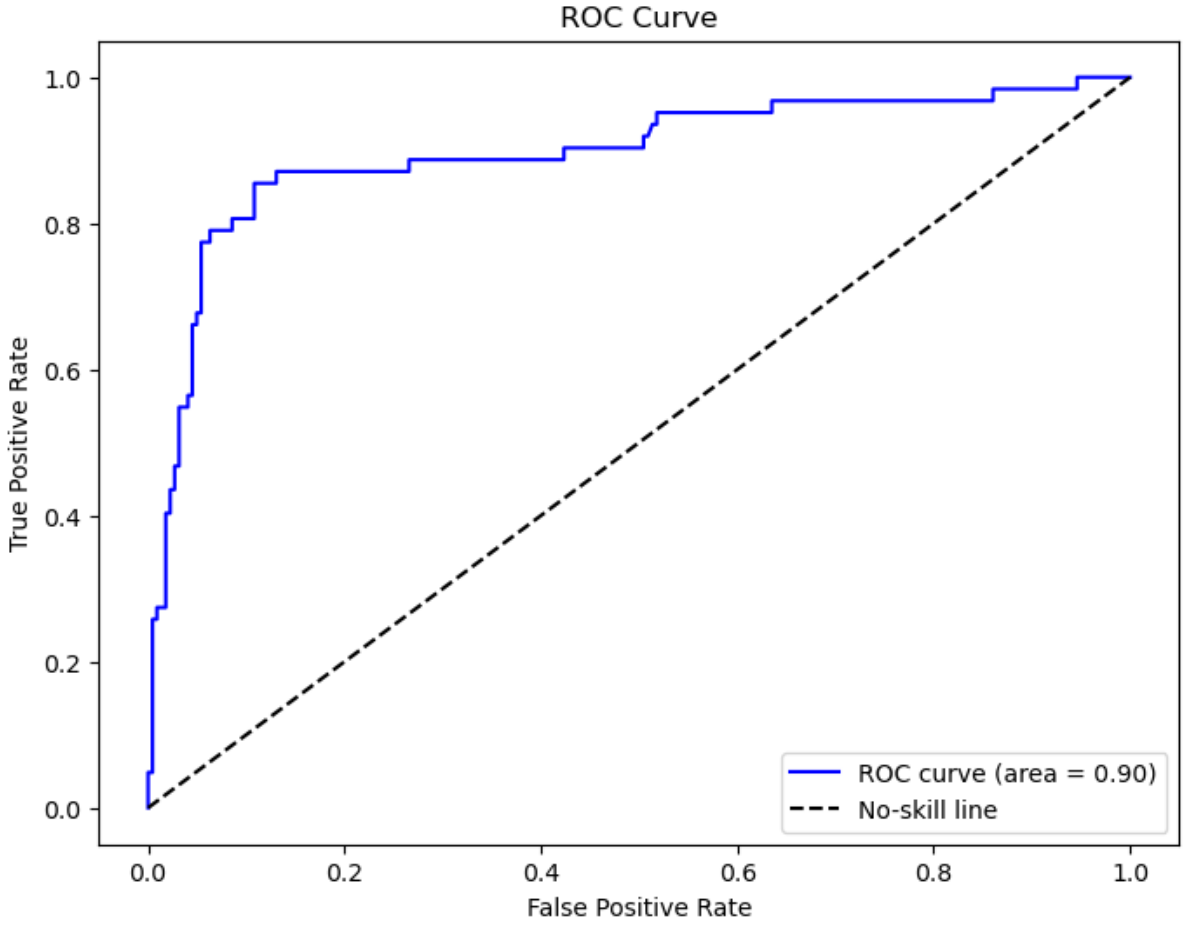

Confusion Matrix - All Predictions as Taken
Not Taken
Taken
True label
0
222
0
62
Not Taken
Taken
Predicted label
200
175
150
125
100
75
50
25
0

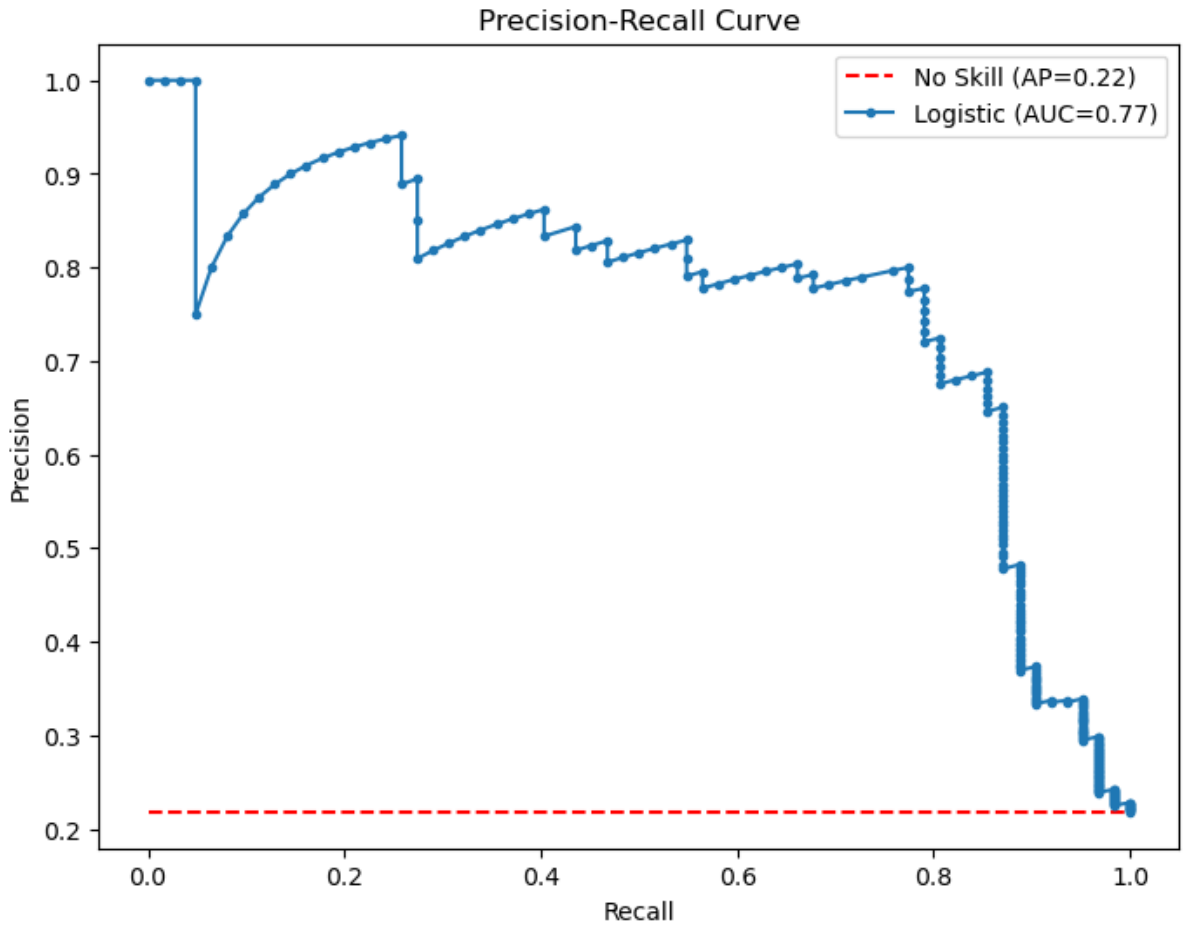

Confusion Matrix - Logistic Regression
0
1
True label
170
52
19
43
0
1
Predicted label
160
140
120
100
80
60
40
20

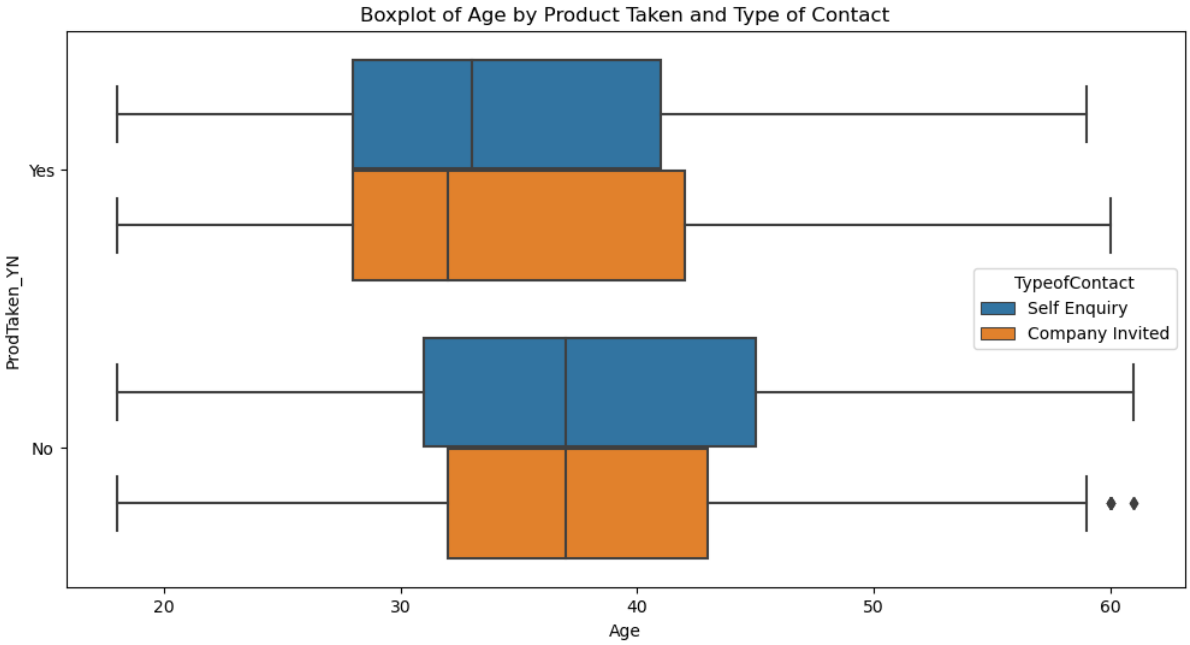

Confusion Matrix - Random Forest
213
9
24
38
0
1
True label
0
1
Predicted label
200
175
150
125
100
75
50
25

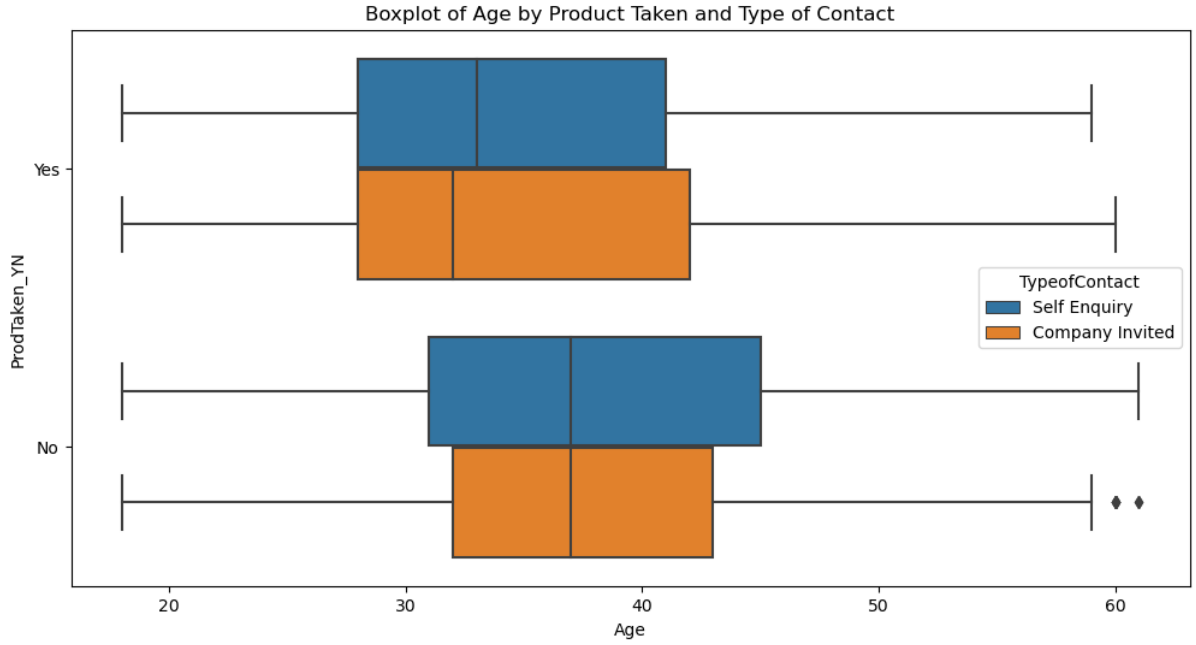

Confusion Matrix - XGBoost
212
10
24
38
0
1
True label
0
1
Predicted label
200
175
150
125
100
75
50
25

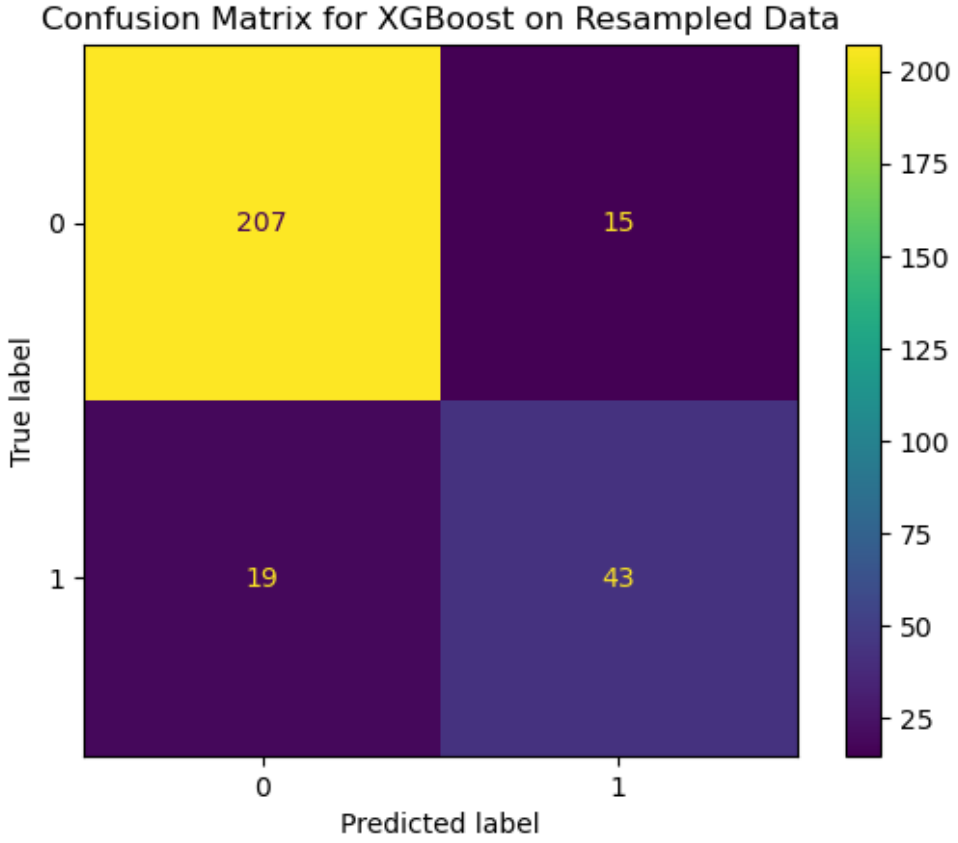

Confusion Matrix - XGBoost
210
12
18
44
0
1
True label
0
1
Predicted label
200
175
150
125
100
75
50
25

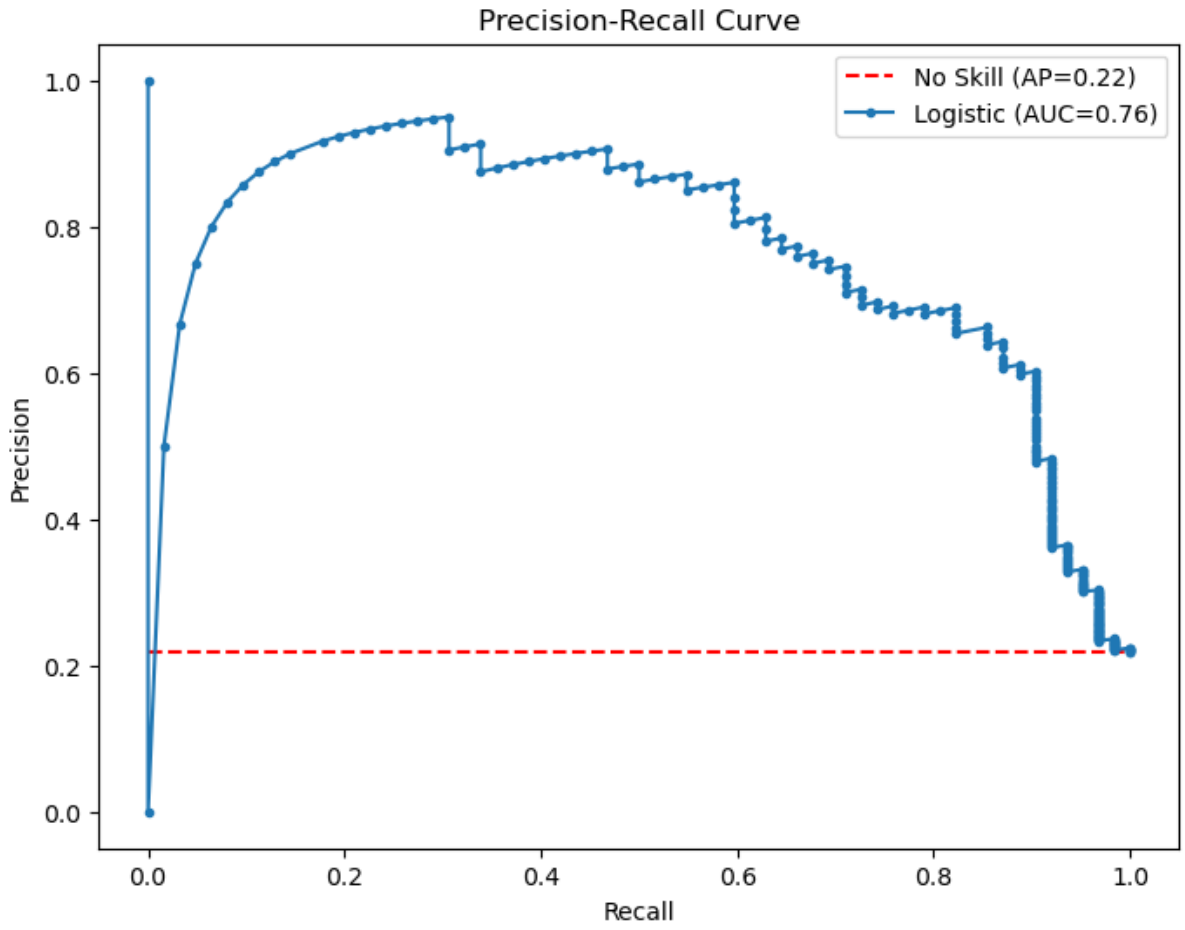

ROC Curve
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
ROC curve (area = 0.90)
No-skill line

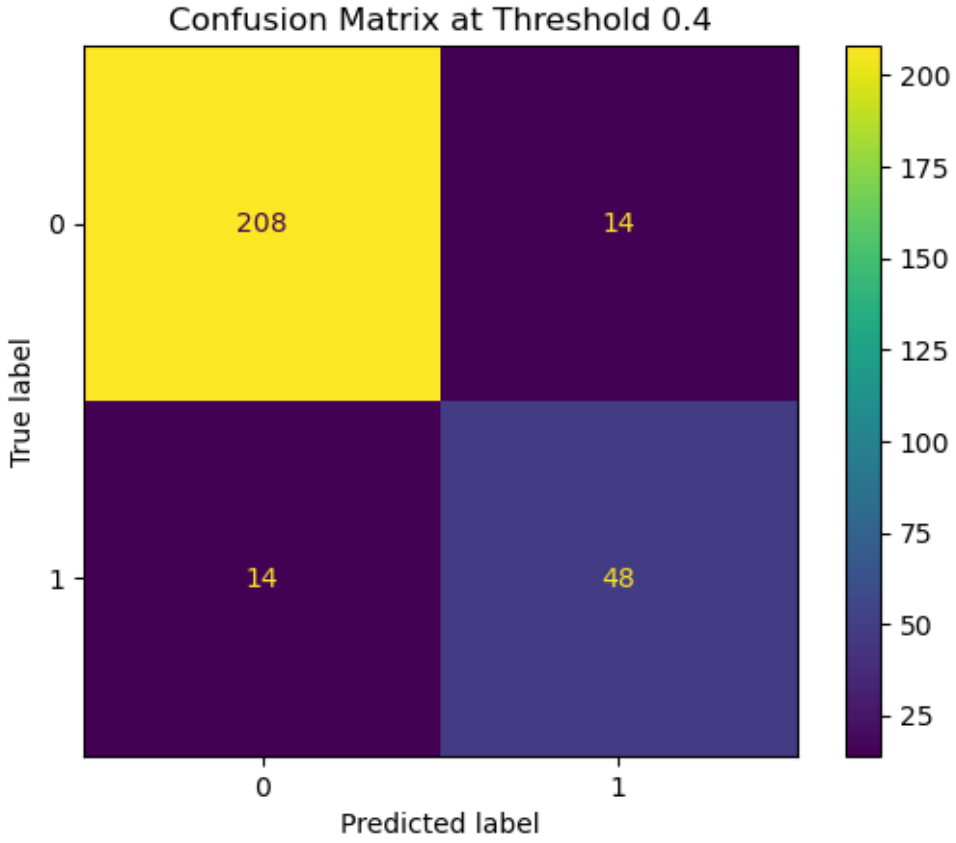

Precision-Recall Curve
No Skill (AP=0.22)
Logistic (AUC=0.77)
Precision
Recall
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.0
0.2
0.4
0.6
0.8
1.0

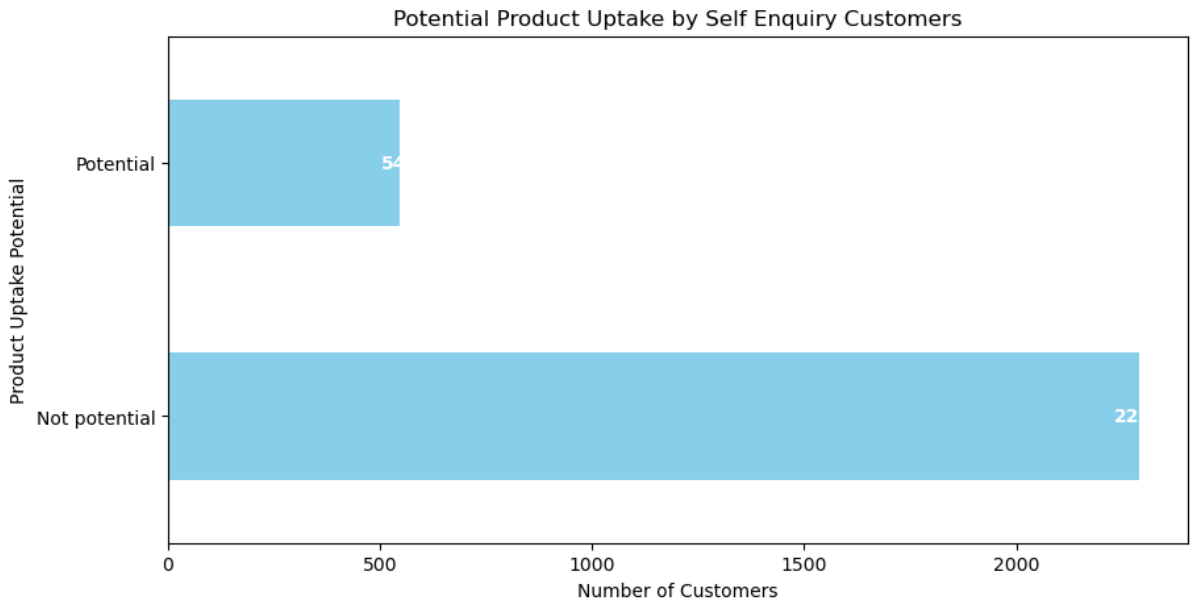

Boxplot of Age by Product Taken and Type of Contact
ProdTaken_YN
Yes
No
TypeofContact
Self Enquiry
Company Invited
20
30
40
50
60
Age


Boxplot of Age by Product Taken and Type of Contact
ProdTaken_YN
Yes
No
TypeofContact
Self Enquiry
Company Invited
20
30
40
50
60
Age


Confusion Matrix for XGBoost on Resampled Data
True label
0
1
207
15
19
43
0
1
Predicted label
200
175
150
125
100
75
50
25


Precision-Recall Curve
No Skill (AP=0.22)
Logistic (AUC=0.76)
Precision
Recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


Confusion Matrix at Threshold 0.4
208
14
14
48
True label
Predicted label
0
1
0
1
200
175
150
125
100
75
50
25


Potential Product Uptake by Self Enquiry Customers
Product Uptake Potential
Potential
Not potential
54
22
0
500
1000
1500
2000
Number of Customers