Travel Package Purchase Prediction

By Jacob Siegel

Contents

- Introduction and Business Context
- Data Overview and Pre-Processing
- EDA & Multi Variable Analysis
- Ensemble Technique Models
- Conclusion and Recommendations

Business Problem Overview and Solution Approach

- **Context:** "Visit With Us" wants to expand its customer base and save costs with targeted marketing to customers more likely to purchase a travel package.
- **Goal:** Understand if a customer are more likely to purchase a travel package based on the data set provided.
- **Implications:** Use the insights from the analysis to create a competitive advantage and market to customers that are more likely to buy a travel package.

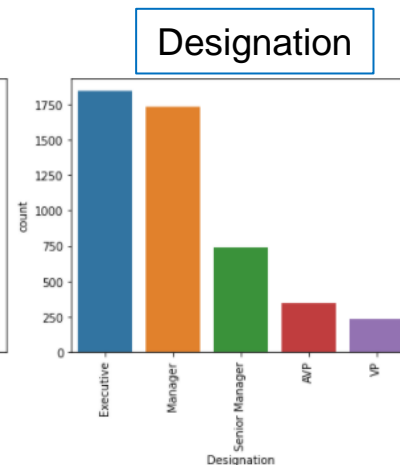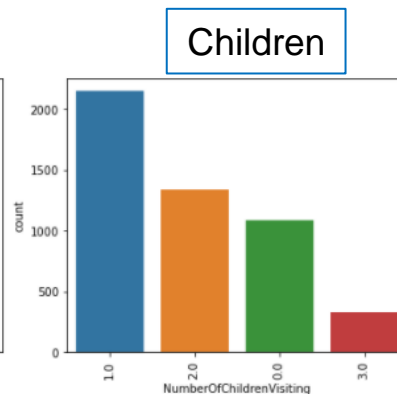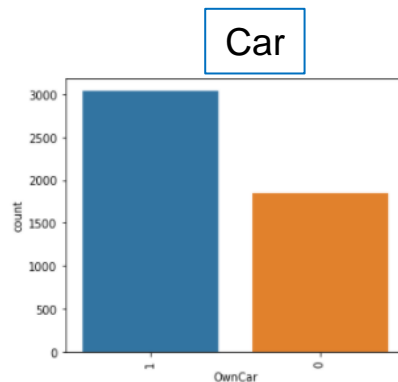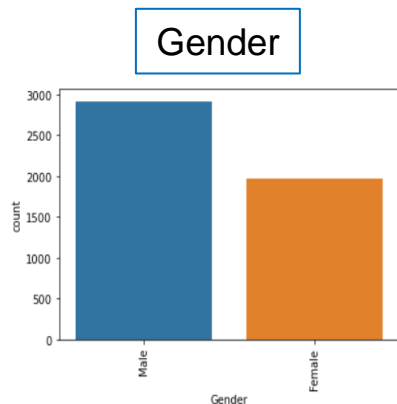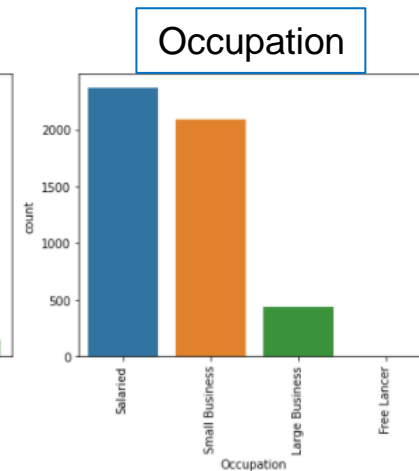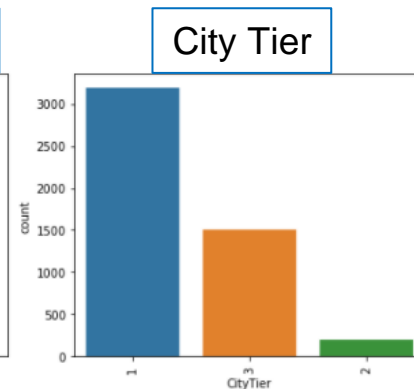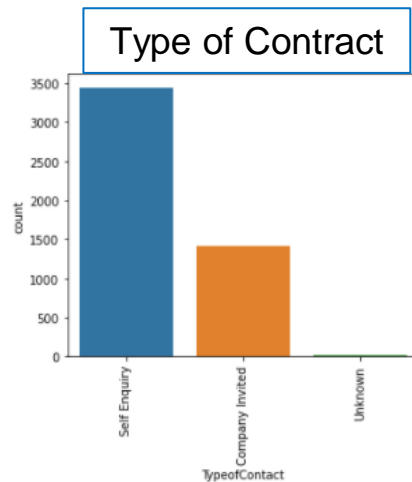# Data Overview & Pre-Processing

<u>Travel Package Purchase:</u>
- 920 people purchased the travel package
- 3968 did not purchase the package

- The raw data set contains 4888 rows and 20 columns.
- 15% of the data contains rows with missing values, see table. Two approaches were taken to deal with the missing values:
  1. Fill in missing values with averages.
  2. Remove rows with missing values.
- Gender category had a typo where several rows had 'Fe Male' instead of Female, this was corrected.
- The type of contact category had several missing values that were filled in with "Unknown."

Columns with Missing Data

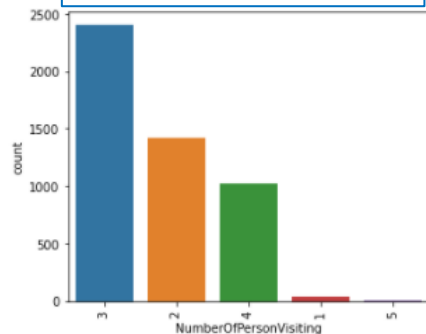|	Number of missing values
CustomerID	0
ProdTaken	0
Age	226
TypeofContact	25
CityTier	0
DurationOfPitch	251
Occupation	0
Gender	0
NumberOfPersonVisiting	0
NumberOfFollowups	45
ProductPitched	0
PreferredPropertyStar	26
MaritalStatus	0
NumberOfTrips	140
Passport	0
PitchSatisfactionScore	0
OwnCar	0
NumberOfChildrenVisiting	66
Designation	0
MonthlyIncome	233

Data Overview: Categorical Variables

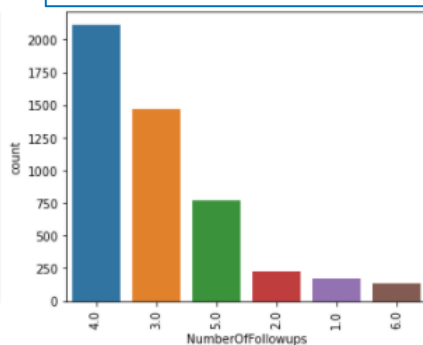- The majority of the data set contains categorical variables, the distributions are shown over the next two slides.

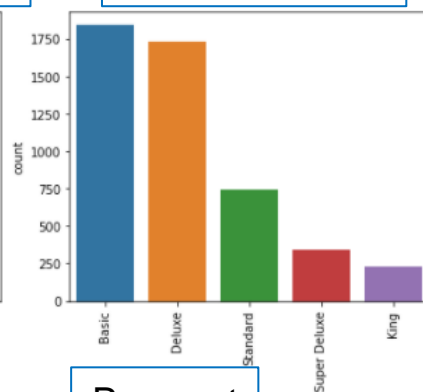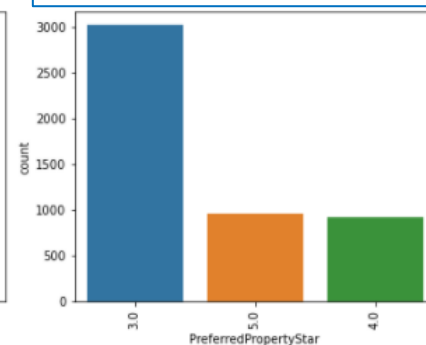# Data Overview: Categorical Variables (continued)
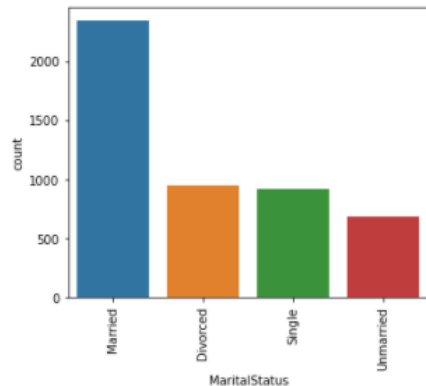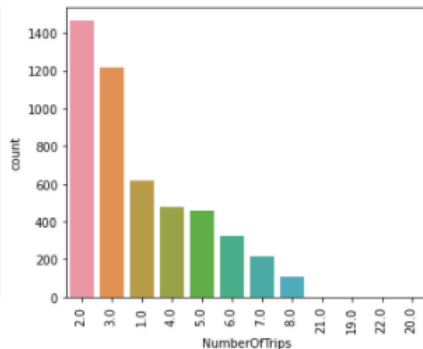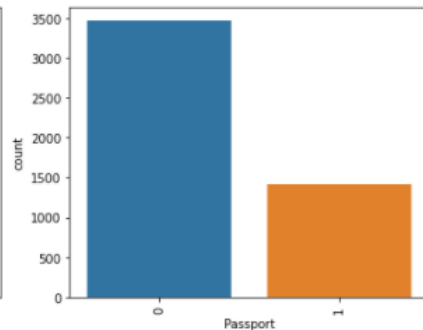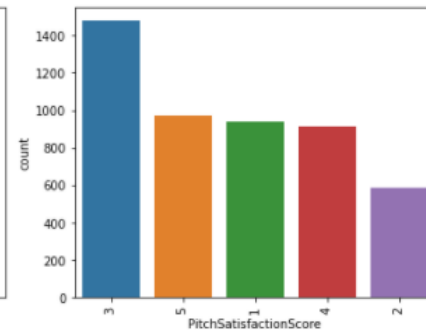
Data Overview: Continuous Variables & Pre-Processing

- The continuous data showed several outliers in 'DurationofPitch' and 'MonthlyIncome' these outliers were removed from the dataset.

Distribution plot after removing multiples

Age

Duration of Pitch

Monthly Income

outliers

EDA: Product Purchased Ratio in Categorical Variables

- Increasing the number of follow ups from the sales person increases the likelihood of purchasing a travel package.
- The basic package has the highest ratio of sales.
- Invited contracts have a higher success rate with purchased products.
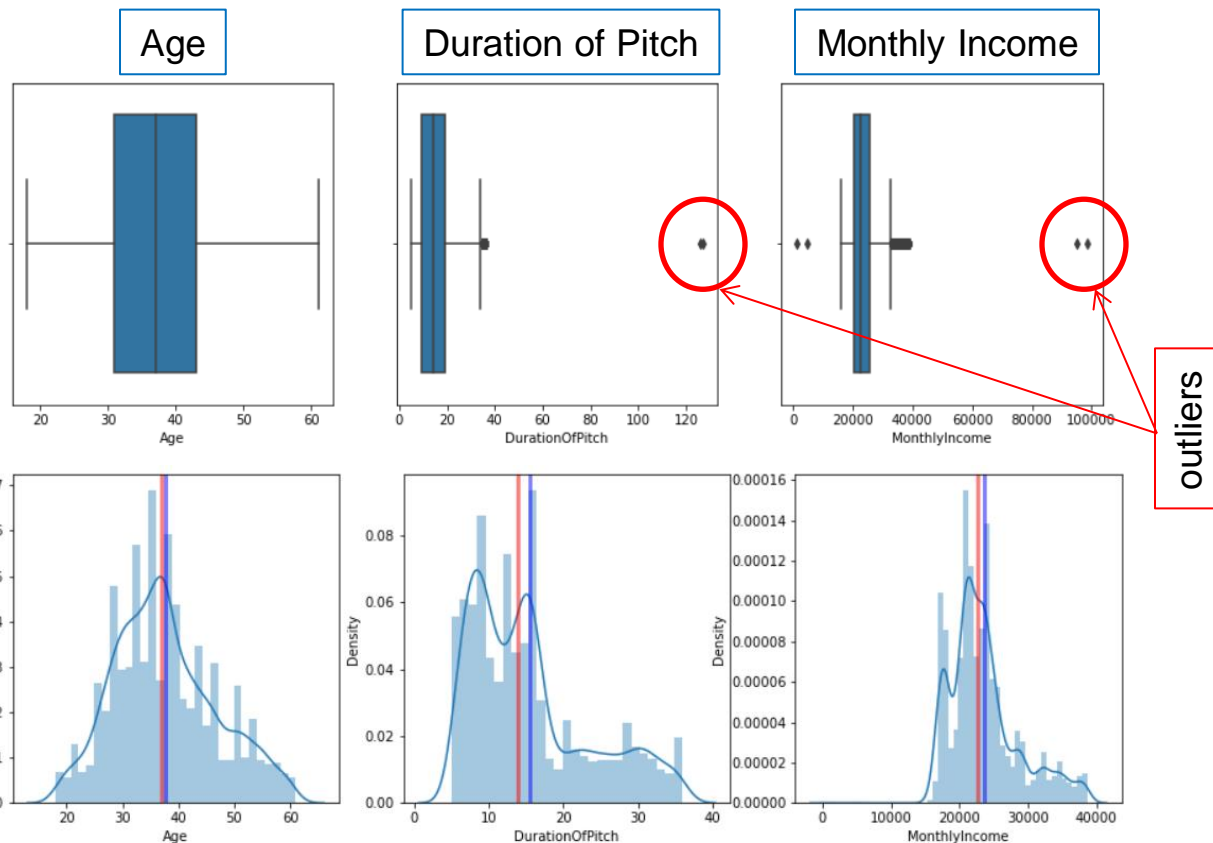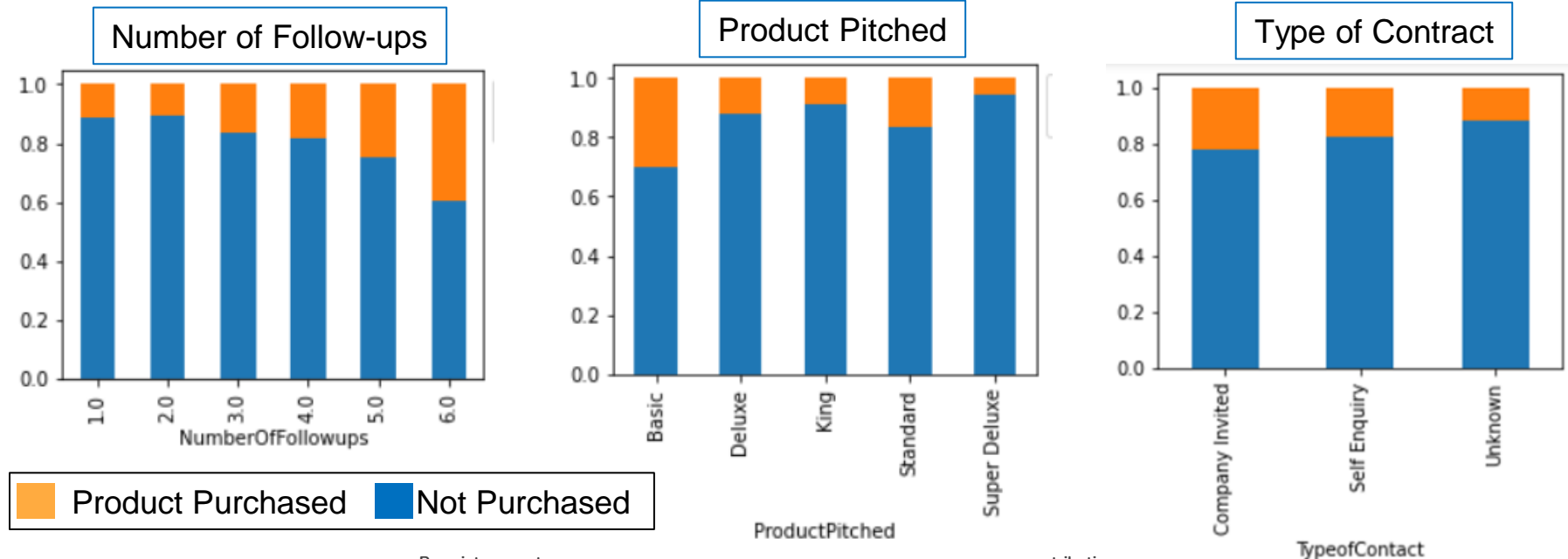

Number of Follow-ups

Product Pitched

Type of Contract

Product Purchased   Not Purchased

EDA: Product Purchased Ratio in Categorical Variables

- Customers that take more trips (19/20) per year always by the travel package, however, customers who travel more than 20 trips a year don't by the package at all.
- Group size between 2 – 4 purchases the most travel packages.
- Executives purchase the most amount of travel packages.
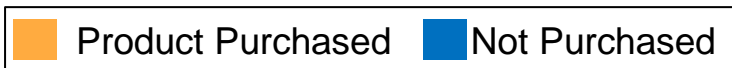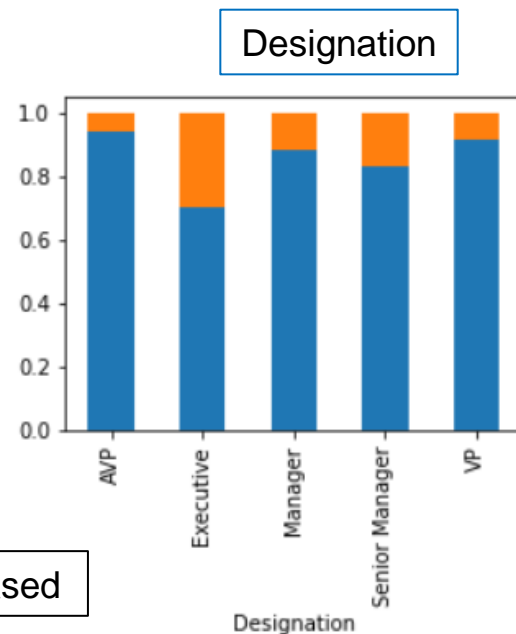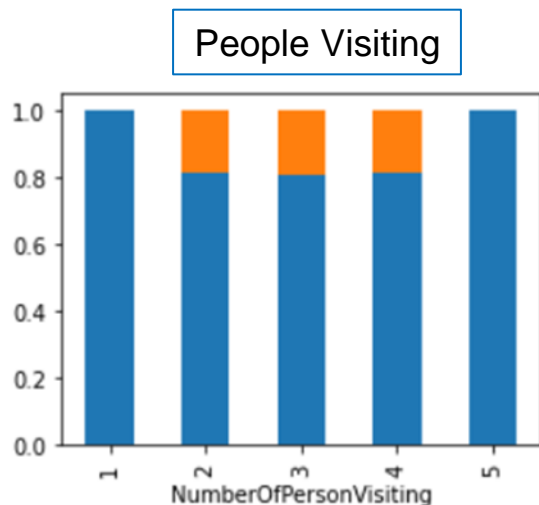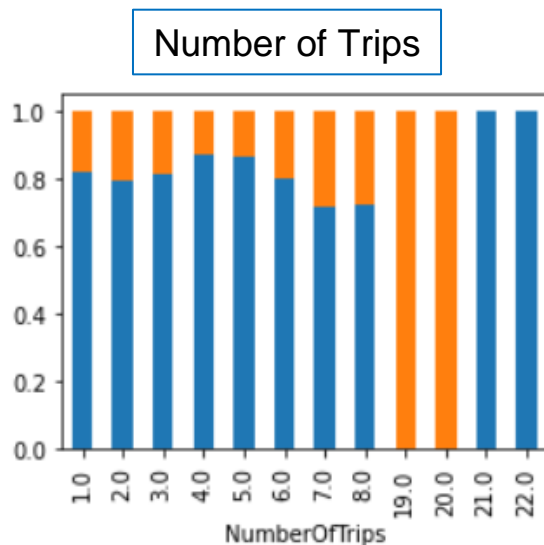
EDA: Product Purchased Ratio in Categorical Variables

- Free Lancer Occupation purchases the most travel packages, as do people with passports.
- Single people also purchase the most travel packages.
- Other categories did not have a significant difference and are not shown in this presentation.

Occupation

Passport

Marital Status

Product Purchased   Not Purchased
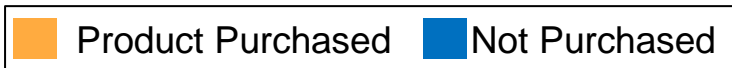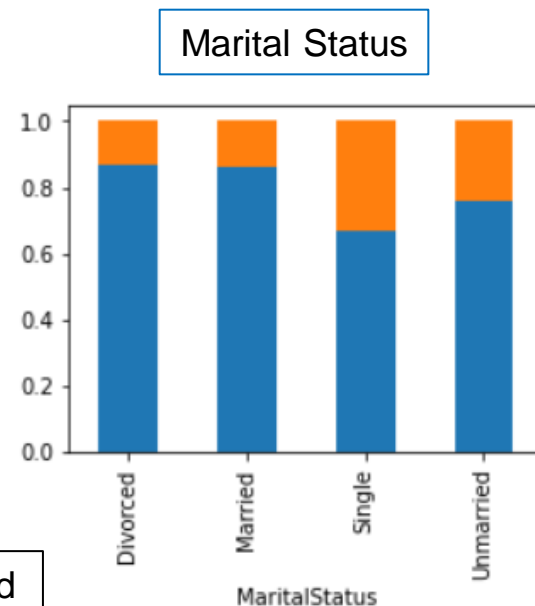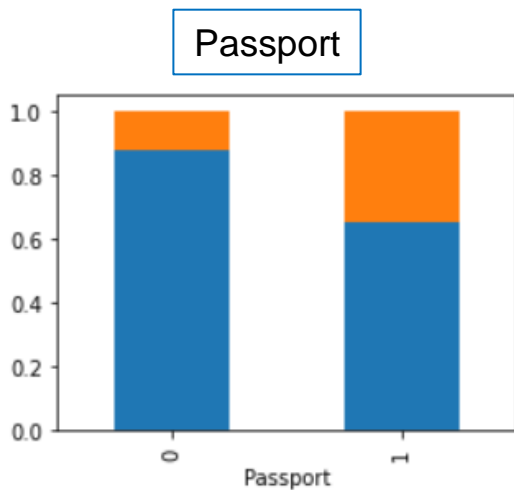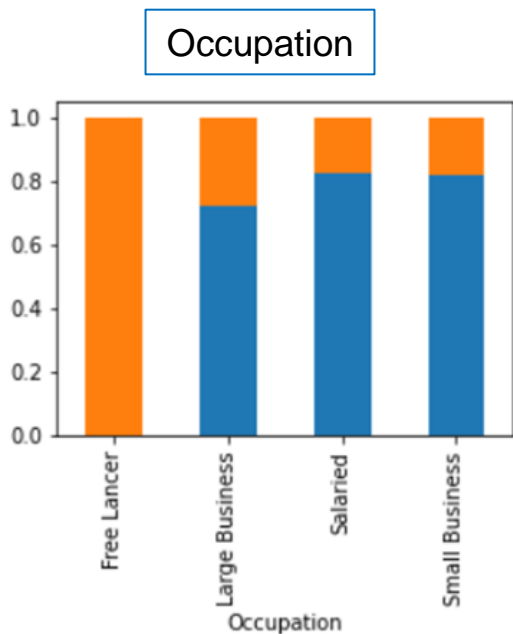
EDA: Income vs Age

- Monthly income increase with age. It doesn't have a clear distinction between people who purchased the product or not. However, surprisingly, people who did not purchase the travel package have a slightly higher income on average.
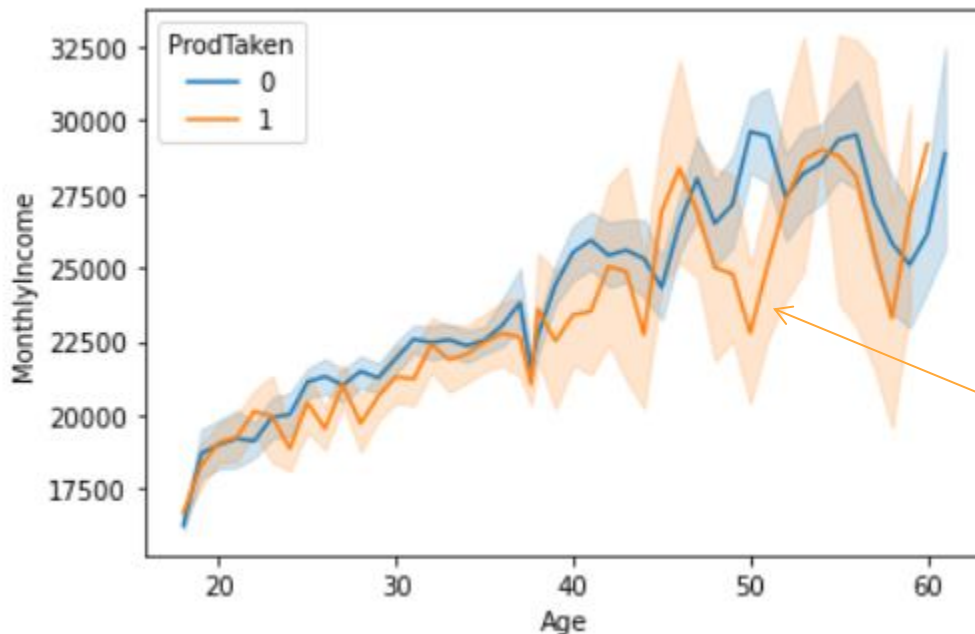

Slightly lower average income per age for people who purchased travel package
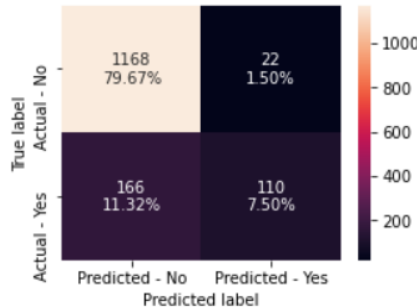
Model Results: Un-tuned Models

- 3 ensemble models with default hyper parameters were run.
- XGBoost model identified the most amount of potential purchasers.
- XGBoost V2 uses the data with all rows with missing values removed, this data will be used in the model tuning to follow.
- AdaBoost is the only model that identified Monthly Income as an important feature.
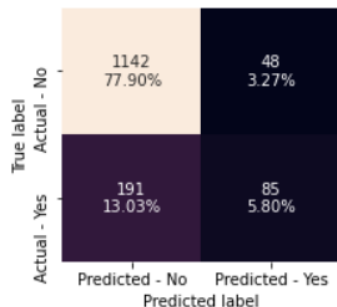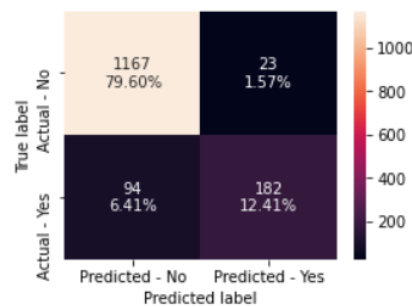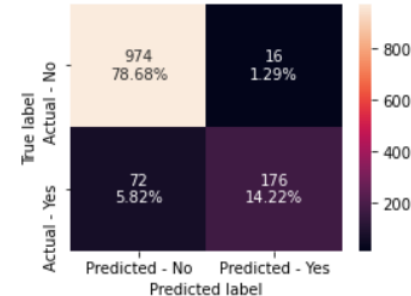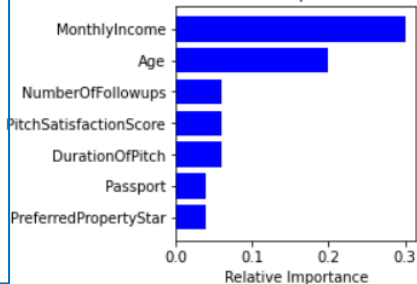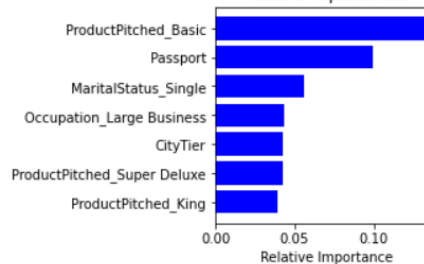
Model Results with Hyperparameter Tuning
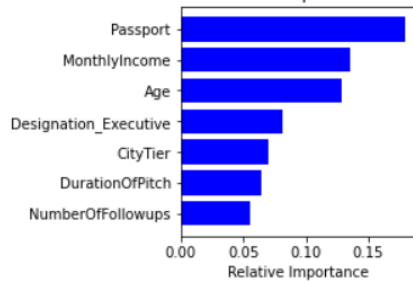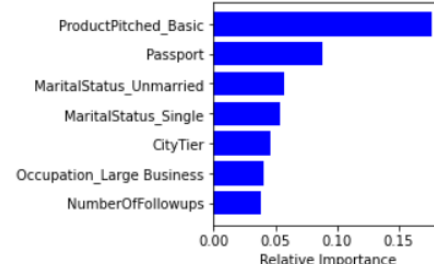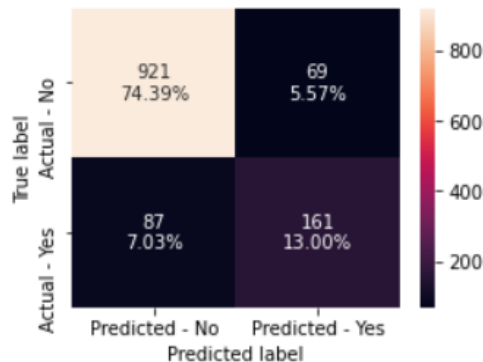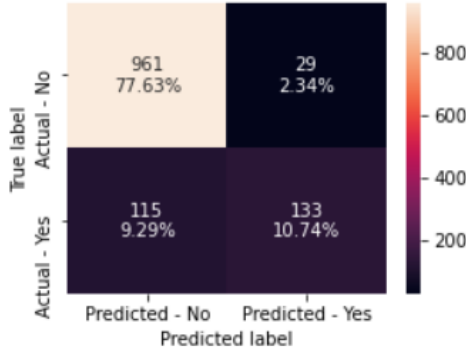
- XGBoost with hyperparameter tuning identified the most amount of people who purchased a travel package.
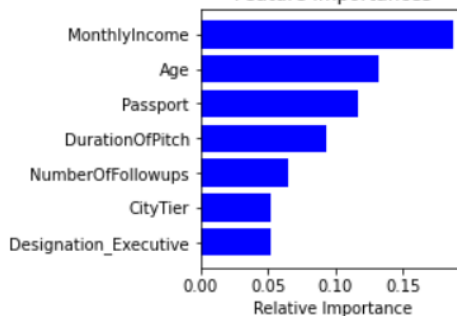
Model Review
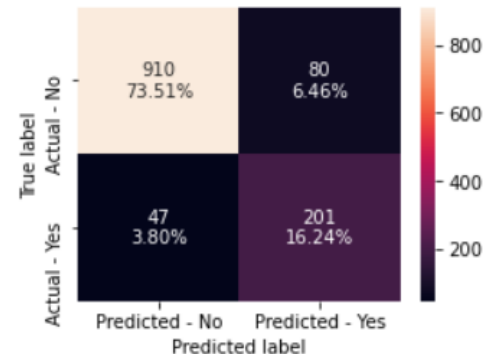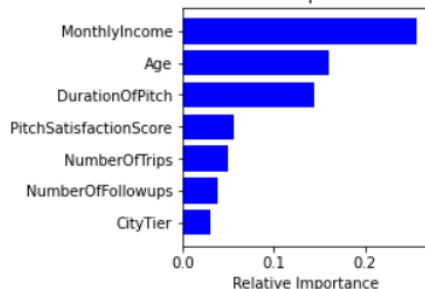
- The XGBoost model with default parameters, followed by the XGBoost hyperperparameter tuned model, performed the best in terms of test accuracy, recall, and precision.
- This model suggests that the "Basic" product and having a passport are the two more important features, followed by marital status single, and occupation large business, see table.

Most Important Features (from models)

- Product Pitched – Basic
- Having A Passport
- Marital Status – Single
- Occupation Large Business
- Designation – Executive
- Designation – Manager
- Product Pitched – super Deluxe
- Product Pitched – Deluxe
- City Tier

Model Results

	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
0	AdaBoost with default paramters	0.84	0.85	0.34	0.38	0.68	0.75
1	AdaBoost Tuned	0.99	0.87	0.96	0.65	0.99	0.70
2	Gradient Boosting with default parameters	0.88	0.89	0.46	0.48	0.86	0.90
3	Gradient Boosting with init=AdaBoost	0.90	0.87	0.52	0.43	0.90	0.84
4	Gradient Boosting Tuned	0.93	0.88	0.69	0.54	0.96	0.82
5	XGBoost with default parameters	1.00	0.93	1.00	0.71	1.00	0.92
6	XGBoost Tuned	0.95	0.90	0.98	0.81	0.81	0.72

Conclusion and Recommendations

Conclusions/Observations:

- To assist with the marketing research for "Visit With Us" an analysis was conducted with EDA and Ensemble Model techniques on a dataset provided by the company to understand the demographics of people more likely to purchase a travel package.
- **From the EDA**, the most important features are:
  - Having more follow ups from the sales person, Occupation Executive, Free Lancer, People with passports, and marital status single.
- **From the Ensemble Models**, the most important features are:
  - Having a Passport, Single Marital Status, and the basic travel package.

Recommendations to Grow Business:

- The ideal candidate for marketing is someone who: has a passport, is single, an executive and/or free lancer.
- The sales people should do several follow ups and suggest the basic travel package to increase the likelihood of someone purchasing.