ENSEMBLE TECHNIQUES

TRAVEL PACKAGE PURCHASE PREDICTION

NAEEM SUFI

BACKGROUND AND CONTEXT

- Visit With Us is a travel and tourism company that is looking to expand into the wellness tourism market. Currently, there are 5 types of packages the company offers - Basic, Standard, Deluxe, Super Deluxe, and King. Data collected over the last year indicates that 18% of the customers purchased the packages. During that same period, marketing costs were quite high because customers were contacted at random without any data driven decision making.
- The company is now planning to launch the Wellness Tourism Package. Wellness Tourism is defined as travel that allows the traveler to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being. This time around, the company wants to harness the available data of existing and potential customers to make the expenditure more efficient.
- Here, we are going to analyze the company's data and information to provide recommendations for policy makers and the marketing team and also, build a model to predict potential customers who are more likely to purchase the wellness tourism package.

OVERVIEW OF DATASET PROBLEMS AND HOW THE ANALYSIS WILL BE APPROACHED TO OVERCOME

- The dataset that we have is relevant to the characteristics of customers as it relates to the current travel packages offered by the company, but we want to use this data to make predictions about a new travel package, wellness travel or tourism. Since the dataset that is going to be used in this analysis is not directly relevant to the predictions that are sought, the data will need to be weighted to conform it into a representative reflection of the population of interest. This will be done by making an "Educated Guess" about the demographics and characteristics of wellness tourism travelers, and then weights assigned employing a standardized process to give preference to those samples that most closely match our guesses. Utilizing the weighted dataset, predictions will be made against the data, and those predictions will be utilized to make inferences about which customers are ripe for marketing wellness tourism to and how best to pitch wellness tourism to customers based upon their demographics.

METRICS EMPLOYED AND MEASUREMENTS OF SUCCESS

- The ideal machine learning algorithm among DecisionTreeClassifier, BaggingClassifier, RandomForestClassifier, AdaBoost, GradientBoost, XGBoost, and Stacked models utilizing a sample weight array to conform the data so that it is applicable to wellness tourism will be selected based upon the following criteria:
- Consistency of its Accuracy score with a predicted accuracy score based upon greater than the mean sample weights.
- The highest available recall score to aim for the most correct predictions that lead to purchases of the wellness tourism package.
- Beyond predicting customers who are more likely to purchase the wellness travel package, the predictions made against the weighted dataset will be used to determine ideal parameters to guide customer interactions.

WHAT IS/ARE THE OBJECTIVES OF THE PROBLEM(S) THAT ARE ADDRESSED IN YOUR PROJECT (CLASSIFICATION/ REGRESSION/ CLUSTERING RULES/ TEXT ANALYTICS/ REINFORCEMENT LEARNING ETC...)?

- We choose 6 models for our classification problem because in our dataset target variable is in categorical format so, when class label is in categorical then this problem is related to classification. Models are given below that we have picked for our prediction.
- Decision Tree
- Bagging Classifier
- Random Forest Classifier
- Ada boost Classifier
- Gradient Boosting Classifier
- XGBoost Classifier

WE WILL MEASURE THE PERFORMANCE OF EACH MODELS ON DIFFERENT METRICS THAT ARE SHOWN BELOW:

- **Accuracy**

  Measure to evaluate how accurate model's performance is:

  $$\frac{TP + TN}{TP + FP + FN + FP}$$

- **Precision**

  Measure to evaluate how accurate model's performance is:

  $$\frac{TP}{TP + FP}$$

- **Recall**

  Measure to evaluate how accurate model's performance is:

  $$\frac{TP}{TP + FN}$$

- **F_1**

  Provides information of both sides TN and TP

$$2 * \frac{Precision * Recall}{Precision + Recall}$$

- $where\ TP = True\ Positive$
- $FP = False\ Positive$
- $TN = True\ Negative$
- $FN = False\ Negetive$
- **Confusion Matrix**

CHARACTERIZATION OF THE DATA SET: NUMBER OF ATTRIBUTES; HAS/ DOES NOT HAVE MISSING VALUES; NUMBER OF EXAMPLES ETC. CLEAN AND REMOVE THE MISSING VALUES FROM THE DATASET. PROVIDE A CLEAR STRATEGY?

- We have a data set of Tourism. This is not a huge dataset so, we will explore the dataset to know more about the data shape, descriptive analyses, Statistical analysis, Univariate ,bivariate analysis, correlation, missing values, dtypes, column names etc.
- Let's try to look the shape of data:

```
# lets try to check the shape of data
df.shape
```

```
(4888, 20)
```

- Lets check the column name and information about the data

```
['CustomerID', 'ProdTaken', 'Age', 'TypeofContact', 'CityTier', 'DurationOfPitch', 'Occupation', 'Gender', 'NumberOfPersonVisiting', 'NumberOfFollowups', 'ProductPitched', 'PreferredPropertyStar', 'MaritalStatus', 'NumberOfTrips', 'Passport', 'PitchSatisfactionScore', 'OwnCar', 'NumberOfChildrenVisiting', 'Designation', 'MonthlyIncome']
```

Now we Exploring the column names is an important aspect of EDA. • We can see that columns are not null. The data types of all columns are int, float and object data type. By closely observing the data and description given about each column attribute we can say that:

- Numeric data columns are **Age, MonthlyIncome and DurationOfPitch.**
- Ordinal Categorical columns are **TypeofContact , Occupation, Gender, ProductPitched, MaritalStatus**, and **Designation.**
- Nominal Categorical columns are **CustomerID , ProdTaken, CityTier, NumberOfPersonVisiting, NumberOfFollowups, PreferredPropertyStar, NumberOfTrips, Passport, PitchSatisfactionScore , NumberOfChildrenVisiting and Owncar**

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4888 entries, 0 to 4887
Data columns (total 20 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   CustomerID                4888 non-null   int64
 1   ProdTaken                 4888 non-null   int64
 2   Age                       4662 non-null   float64
 3   TypeofContact             4863 non-null   object
 4   CityTier                  4888 non-null   int64
 5   DurationOfPitch           4637 non-null   float64
 6   Occupation                4888 non-null   object
 7   Gender                    4888 non-null   object
 8   NumberOfPersonVisiting    4888 non-null   int64
 9   NumberOfFollowups         4843 non-null   float64
 10  ProductPitched            4888 non-null   object
 11  PreferredPropertyStar     4862 non-null   float64
 12  MaritalStatus             4888 non-null   object
 13  NumberOfTrips             4748 non-null   float64
 14  Passport                  4888 non-null   int64
 15  PitchSatisfactionScore    4888 non-null   int64
 16  OwnCar                    4888 non-null   int64
 17  NumberOfChildrenVisiting  4822 non-null   float64
 18  Designation               4888 non-null   object
 19  MonthlyIncome             4655 non-null   float64
dtypes: float64(7), int64(7), object(6)
memory usage: 763.9+ KB
```

From this we can see that too much missing values in every columns. So we will handle these missing values according our problem.

Now are moving toward the data cleaning part so that we can clean our data and move forward.

```
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
dtype: int64
```

DESCRIPTIVE ANALYSIS

- The Exploratory Data Analysis is more important method which shows basic statistical characteristics of each numerical feature (int64 and float64 types):
- Number of non-missing values, mean, standard deviation, range, median, 0.25 and 0.

	CustomerID	ProdTaken	Age	CityTier	DurationOfPitch	NumberOfPersonVisiting	NumberOfFollowups	PreferredPropertyStar	NumberOfTr
count	4888.000000	4888.000000	4662.000000	4888.000000	4637.000000	4888.000000	4843.000000	4862.000000	4748.000(
mean	202443.500000	0.188216	37.622265	1.654255	15.490835	2.905074	3.708445	3.581037	3.236!
std	1411.188388	0.390925	9.316387	0.916583	8.519643	0.724891	1.002509	0.798009	1.849(
min	200000.000000	0.000000	18.000000	1.000000	5.000000	1.000000	1.000000	3.000000	1.000(
25%	201221.750000	0.000000	31.000000	1.000000	9.000000	2.000000	3.000000	3.000000	2.000(
50%	202443.500000	0.000000	36.000000	1.000000	13.000000	3.000000	4.000000	3.000000	3.000(
75%	203665.250000	0.000000	44.000000	3.000000	20.000000	3.000000	4.000000	4.000000	4.000(
max	204887.000000	1.000000	61.000000	3.000000	127.000000	5.000000	6.000000	5.000000	22.000(

DESCRIPTIVE ANALYSIS ON NON-NUMERIC

- Previously we see that descriptive analysis on int or float type but not in numeric.
- In order to see statistics on non-numerical features, one must explicitly indicate data types of interest.

	TypeofContact	Occupation	Gender	ProductPitched	MaritalStatus	Designation
count	4863	4888	4888	4888	4888	4888
unique	2	4	3	5	4	5
top	Self Enquiry	Salaried	Male	Basic	Married	Executive
freq	3444	2368	2916	1842	2340	1842

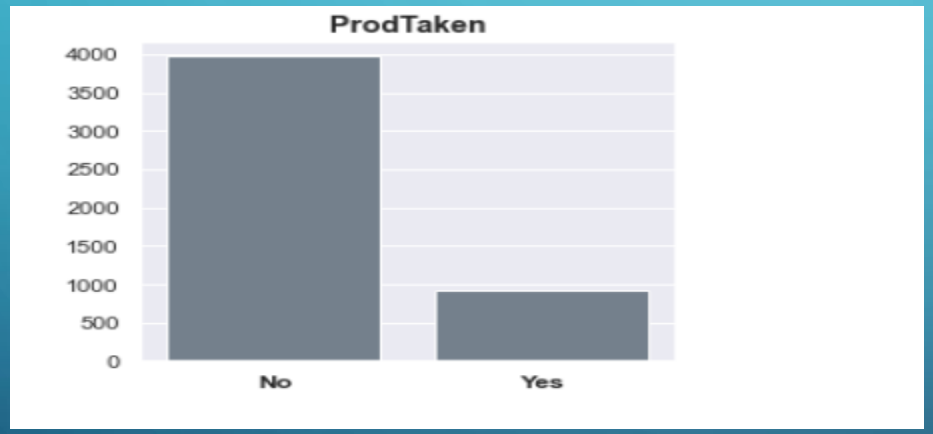
- This give us count of values, unique values, top and frequency of values

DATA VISUALIZATION(UNIVARIATE)

In this section, we will visualize all the columns and check the distribution and relationship with other columns and also, we will get more analysis from the graph.

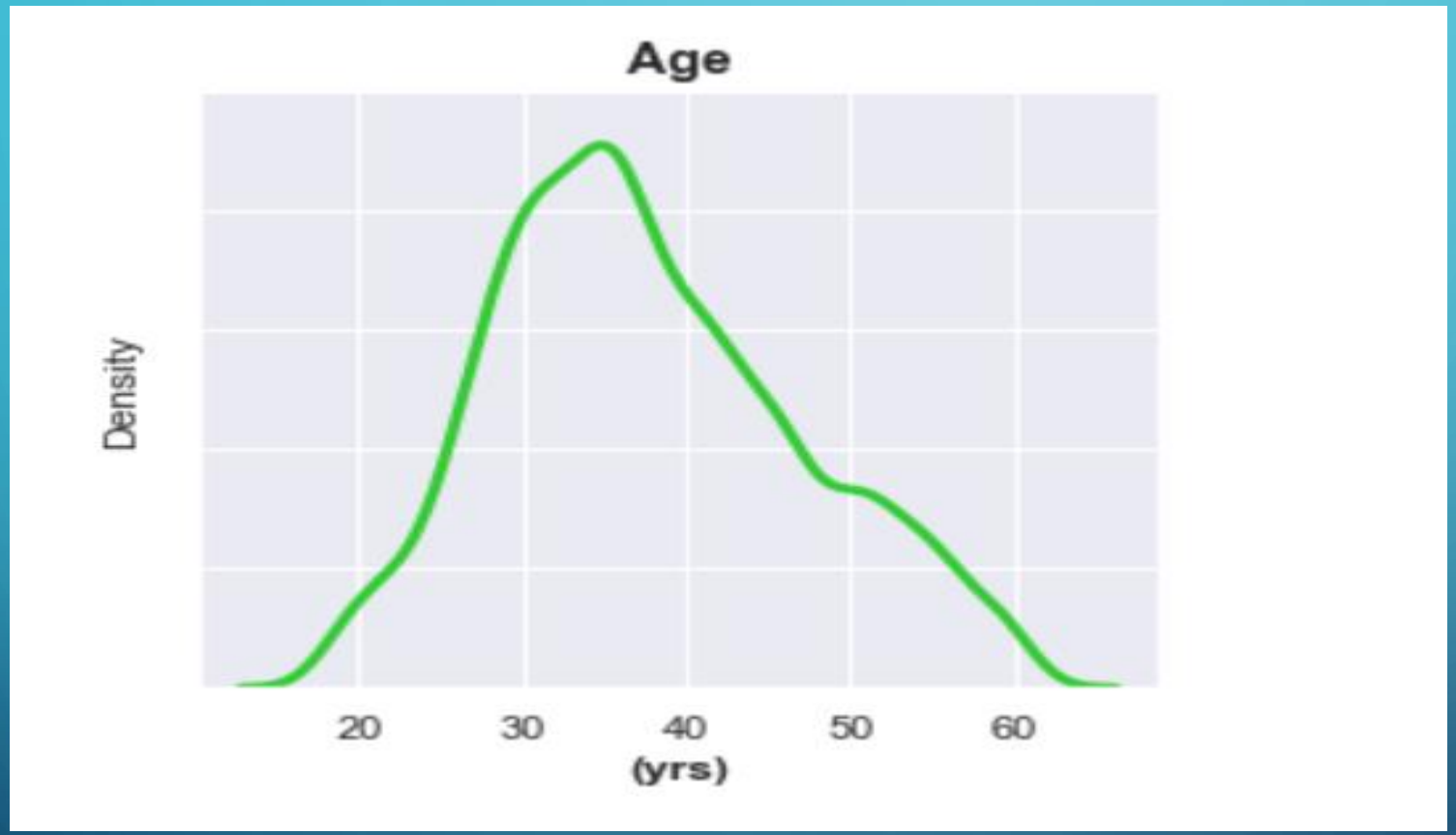
Uni stands for "one," meaning that, there is just one sort of variable in the data. Univariate analysis' main purpose is to characterize the data. The information will be collected, analyzed, and a pattern will be identified

- Now let us see that those customers which are visiting how many number of customer purchased the package.



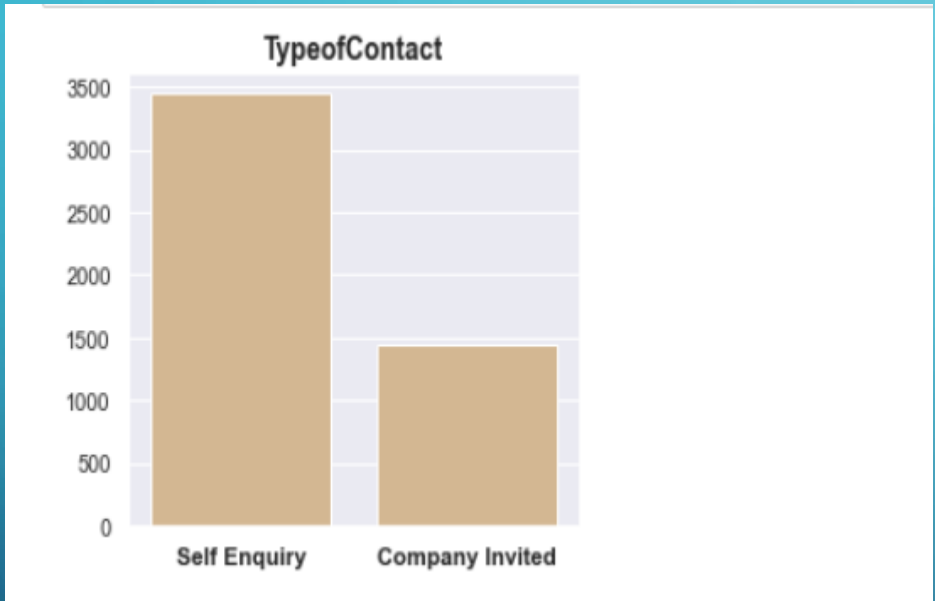
- Whether the customer has purchased a package or not (0: No, 1: Yes) Most people are not purchasing anything as you see.

- Now let us see through the curve that in which age people purchase most packages



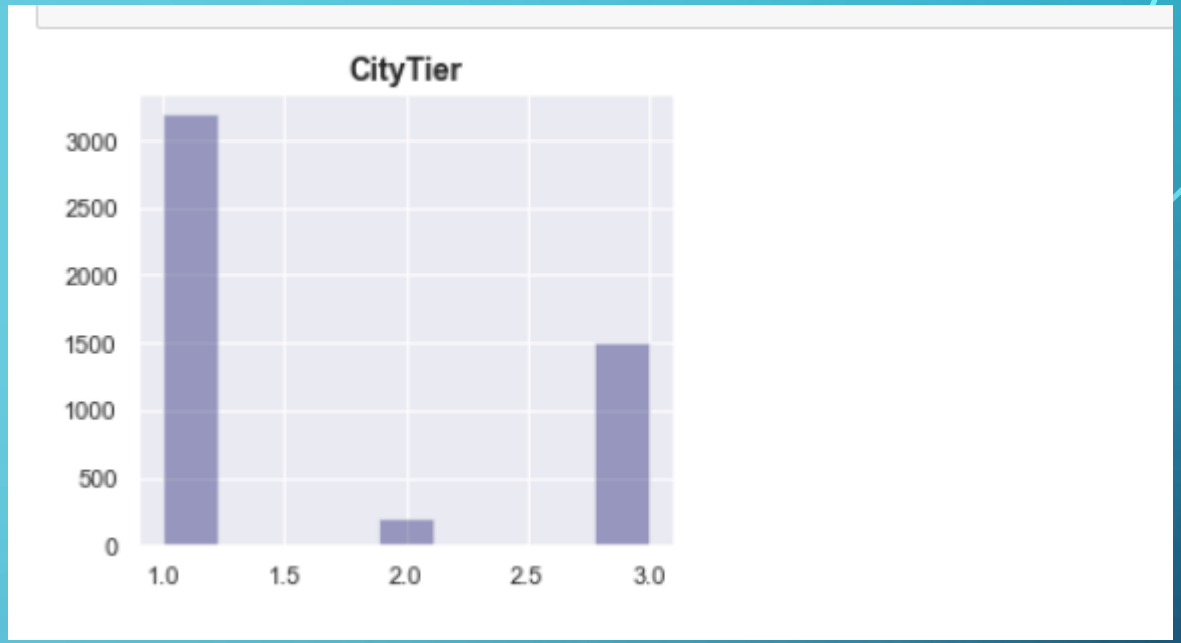
- From above figure Age of customer are mostly between 25 to 40 and a very small number people which is less then 20 years old and more than 60 years old.

Let’s Now try to look that how customer was contacted (Company Invited or Self Inquiry)

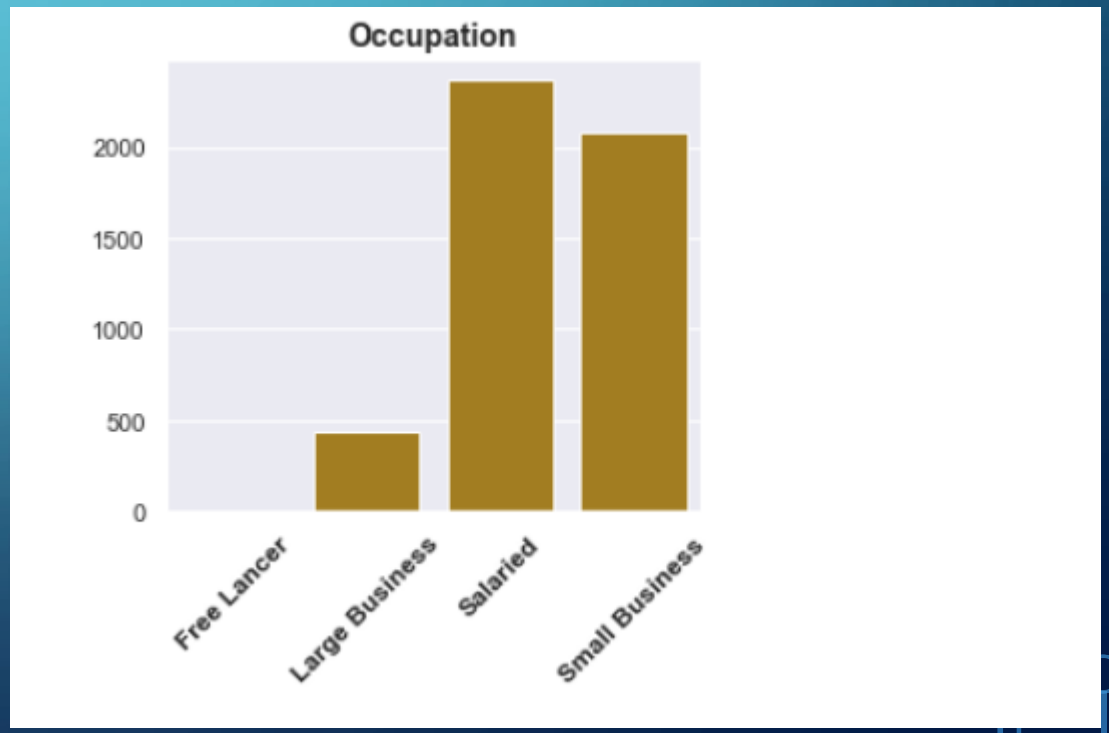


- Most of people are contracted as a self enquiry and average number of people are company invited.

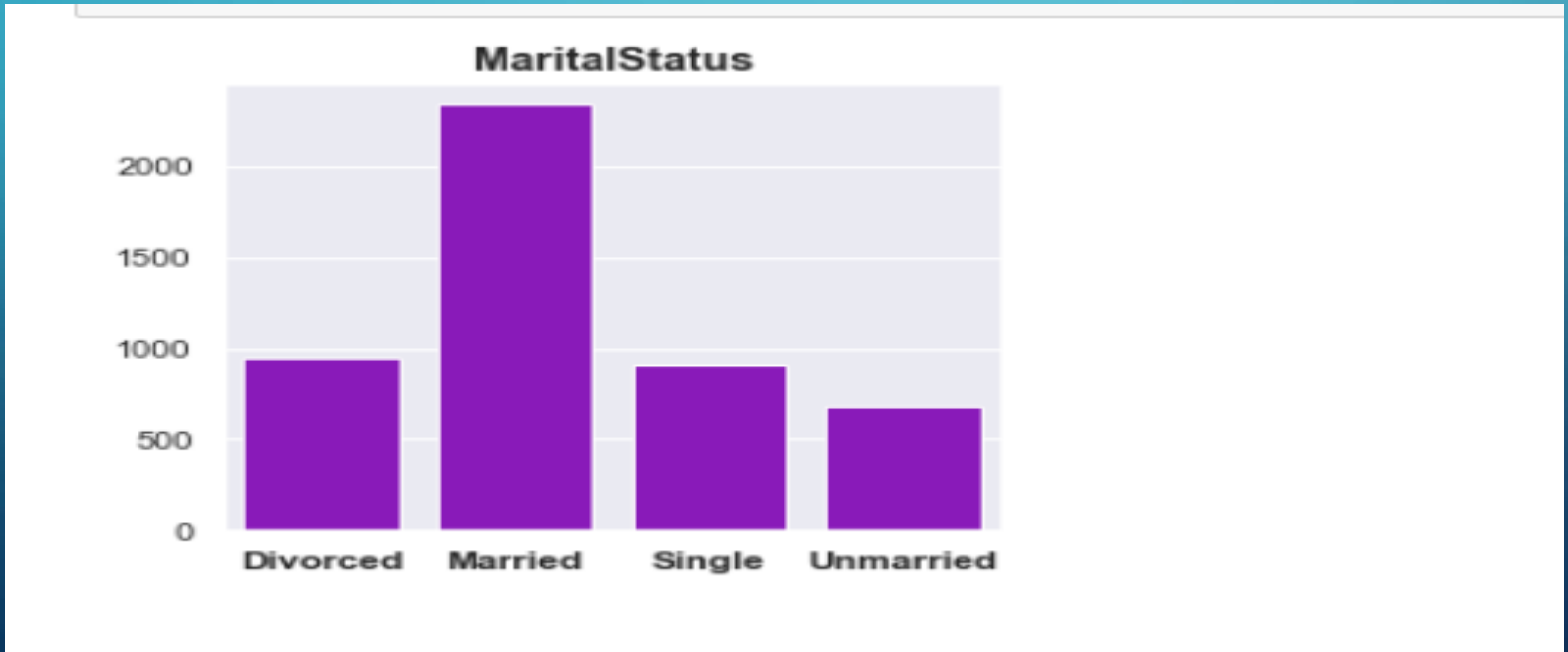
Now we can plot the bar graphs so that we can understand that city tier depends on the development of a city, population, facilities, and living and most of are come from the population side.



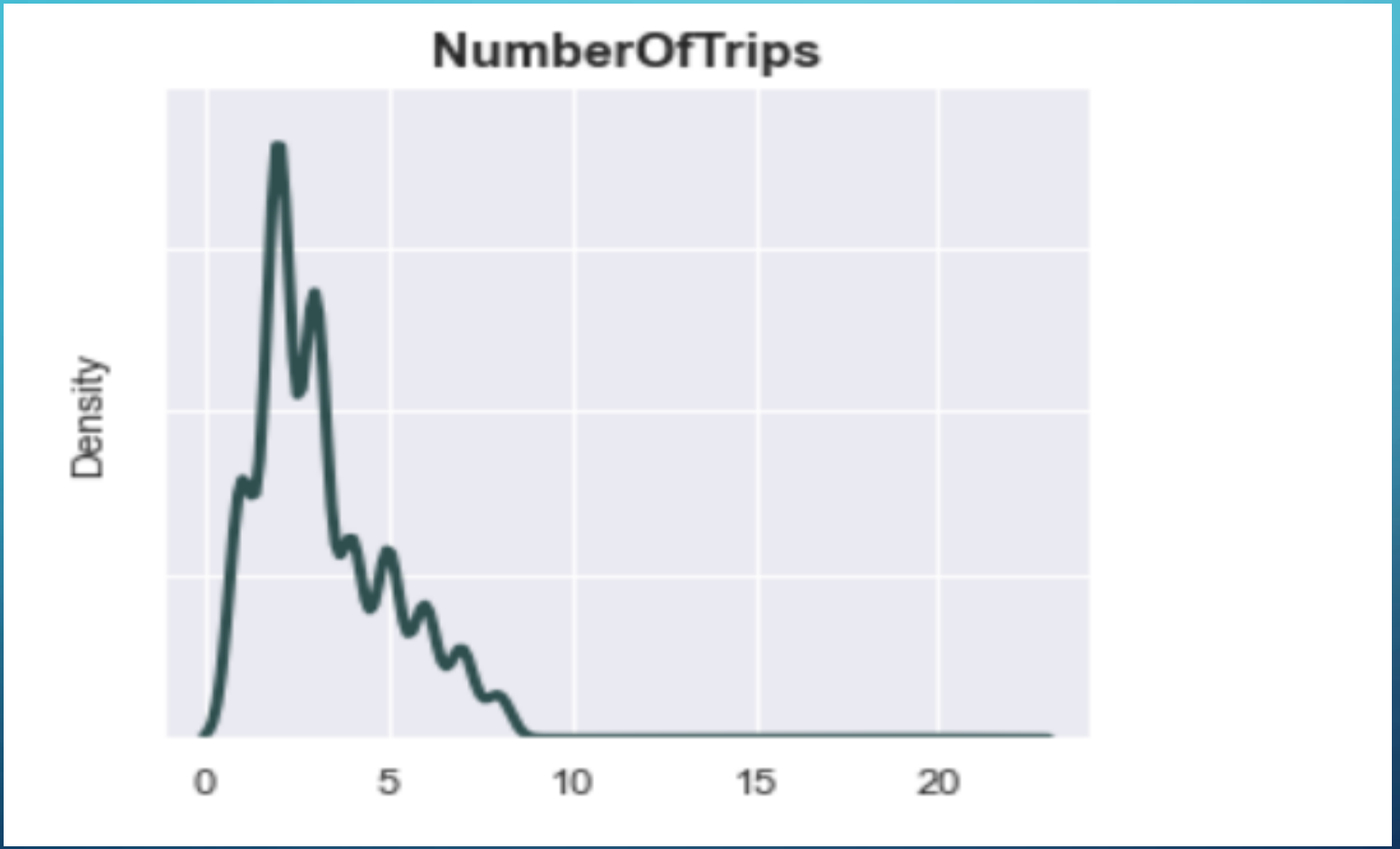
Again, through bar graph we are seeing that most of occupation of customers are Salaried based or have a small business. And a very low number of people have occupation of Large Business.



Marital status of customer which are going to trip is Married and others like Single, Divorced or Unmarried are approximately equally.
But Married people are most going to trip.
There are almost 70 percent of people are married.



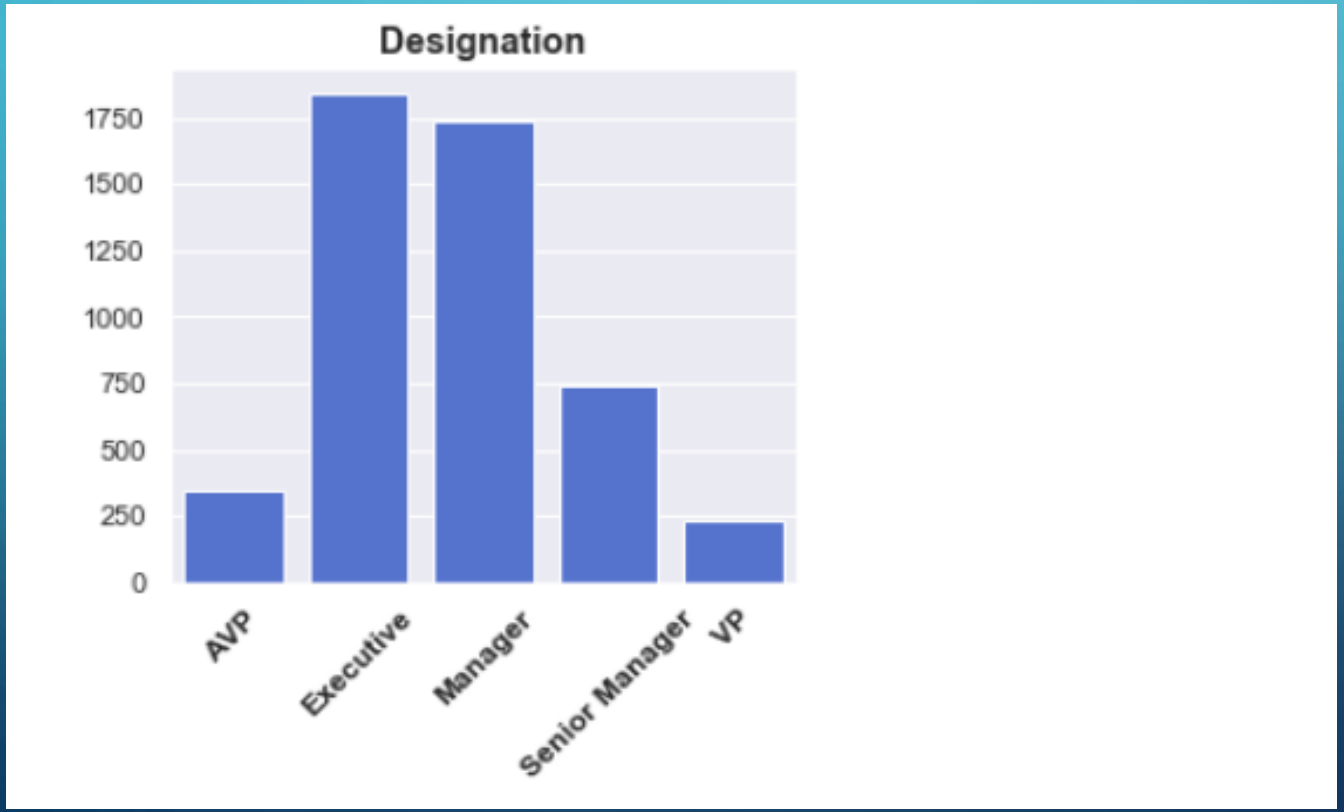
Average number of trips in a year by customer are between 2 to 5 and maximum number trips of people are around about 10.



- From above figure Total number of persons planning to take the trip with the customer are 2 to 4

- Total number of children with age less than 5 planning to take the trip with the customer.

- As you can see from the bar plot that Most designation of the customer in the current organization are Executive or as a Manager.
- And a very low at AVP or VP

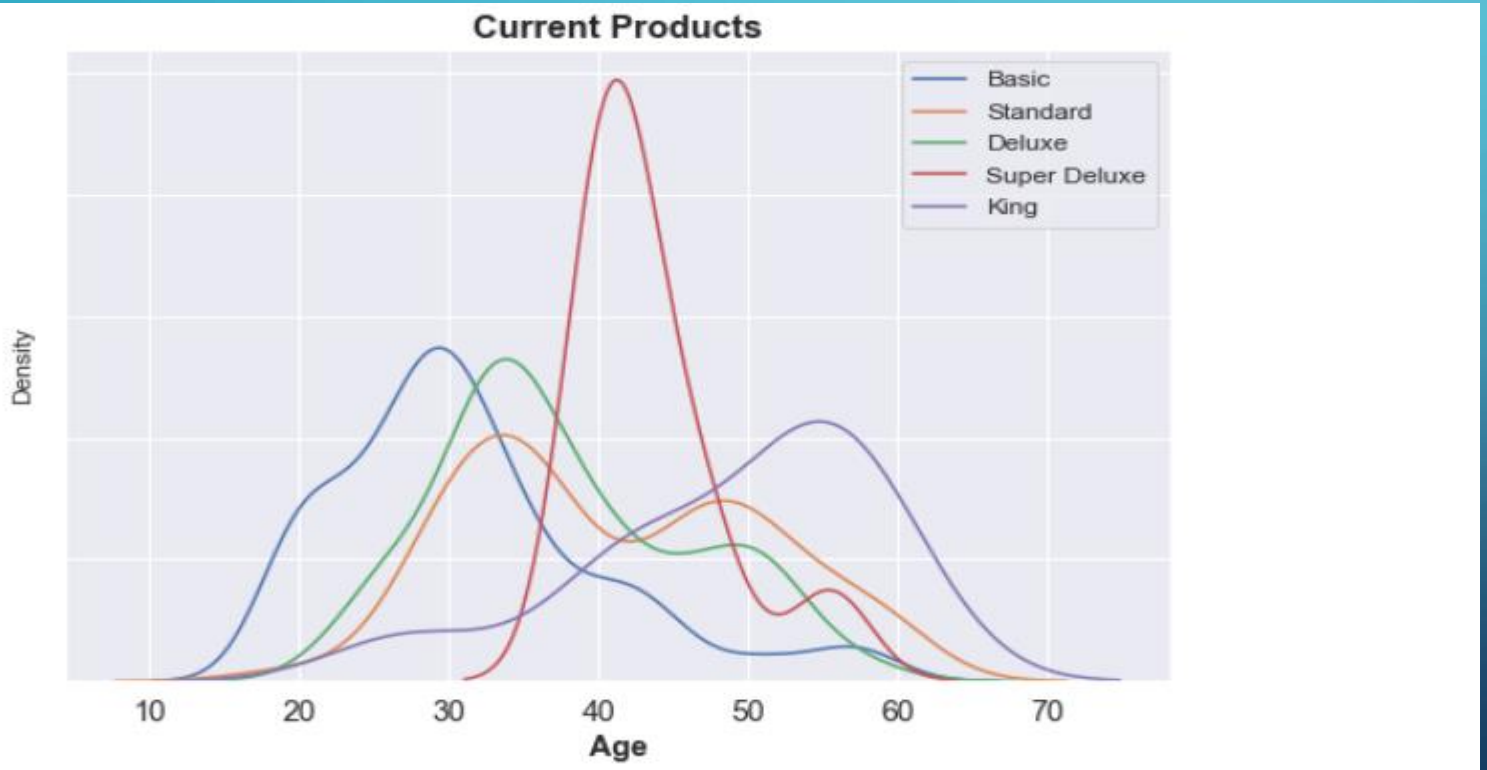


INFERENCES FROM UNIVARIATE ANALYSIS

- Most people are not purchasing anything.
- Average number of age of customer are mostly between 25 to 40 and a very small number people which is less then 20 years old and a smaller number of people more than 60 years old.
- Most people are contracted as a self Enquiry.
- Most number of people come from the population side.
- Occupation of customer are Salaried based or have a small business.
- Most Married people are going to trip.There are almost 70 percent of people who are married.
- Average number of trips in a year by customers are between 2 to 5.
- Total number of children with age less than 5 planning to take the trip with the customer.
- Most designation of the customer in the current organization are as an Executive or as a Manager.

EXPLORATORY DATA ANALYSIS OF CURRENT PRODUCT SALES AND CUSTOMER TRENDS

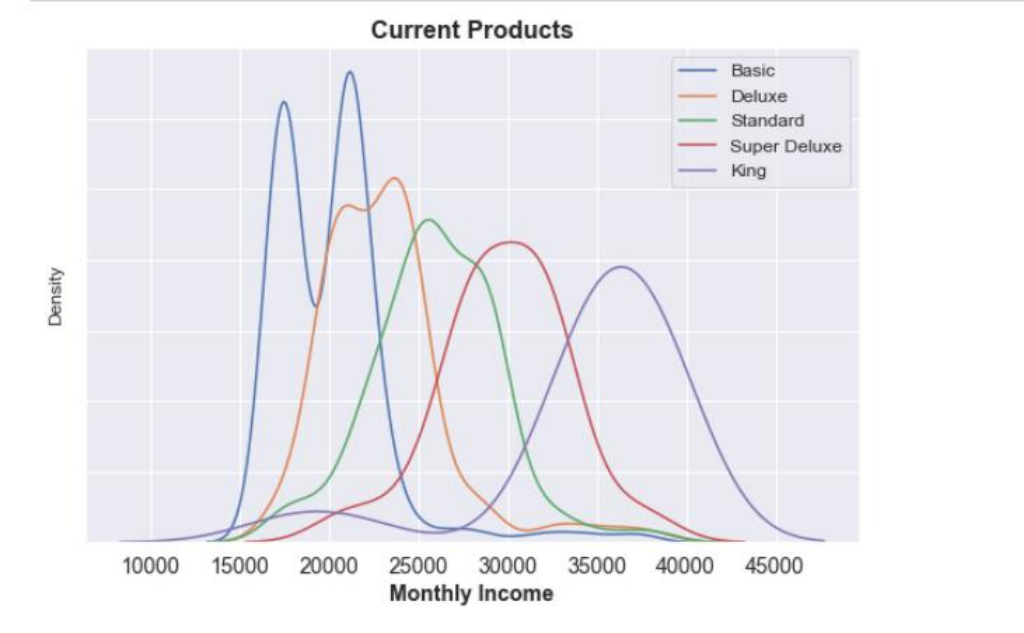
- Below, customer trends for current product sales are visualized and explored to gain insight into current sales operations. Later in this notebook, these same visualizations will be compared against predicted wellness tourism customers to attempt a differentiation.
- Below, customer trends for current product sales are visualized and explored to gain insight into current sales operations.



As we see from this graph that the customer has most purchased a package is Super Deluxe and age of those peoples who purchases most these packages are between 30 to 50.

OBSERVATIONS

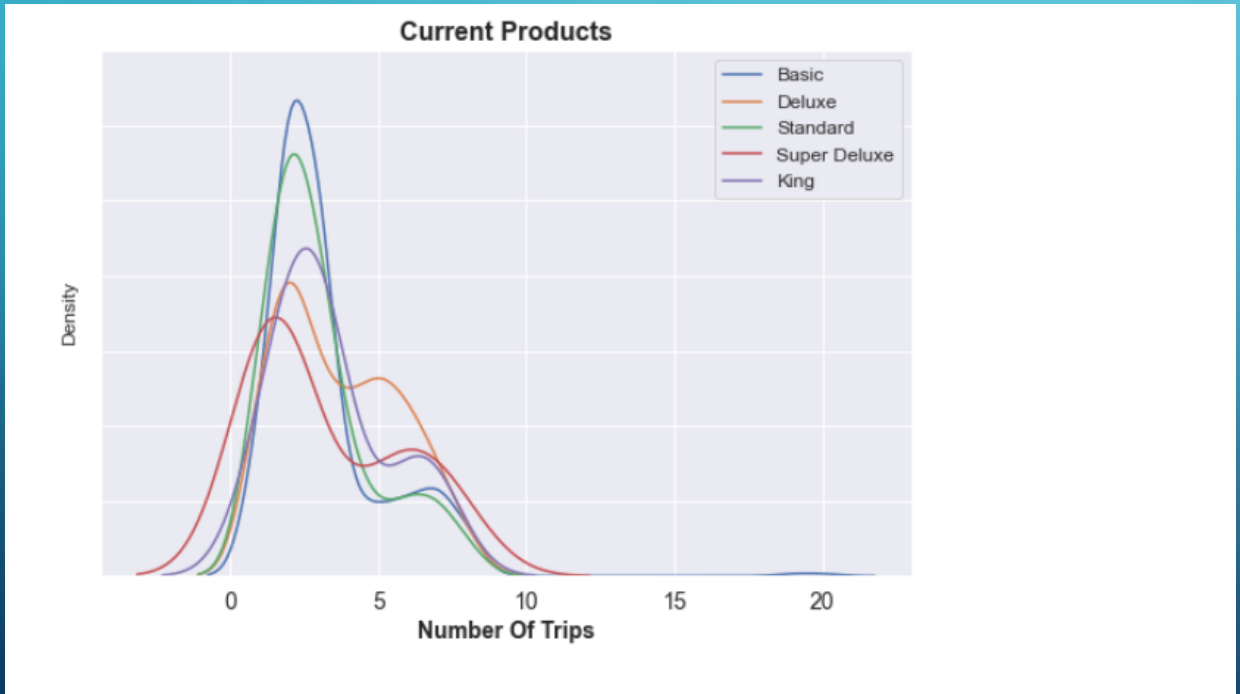
- Basic package: popular with customers in their 20s, 30s, and 40s Standard and Deluxe packages: skewed towards customers in their 30s and 40s Super Deluxe package: popular with customers in their 40s and 50s King package: popular with customers in their 40s and 50s



Those customers has most purchased a Basic package whose Monthly income is between 15000 to 25000. And those people purchase the king package whose monthly income between 35000 to 45000 which is quite high range.

OBSERVATIONS

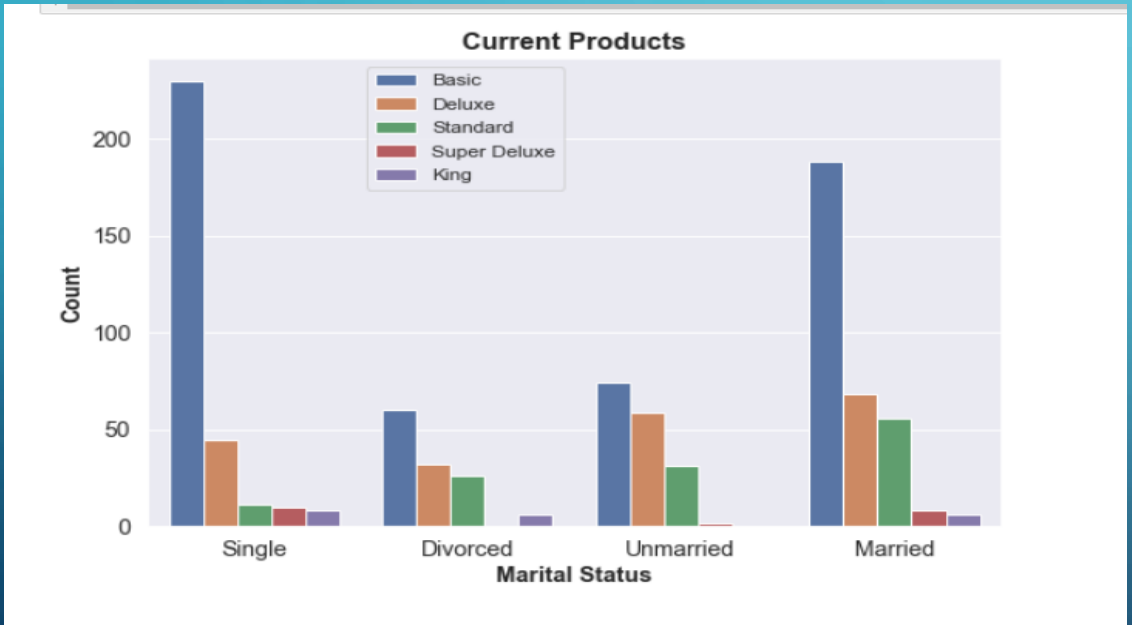
- Basic Package: Popular with customers with income range between 15k and 25k Deluxe Package: Popular with customers with income range between 20k and 30k Standard Package: Popular with customers with income range between 20k and 35k Super Deluxe Package: Popular with income range between 25k and 35k King: Popular with customers with income range between 30k and 45k.



Average number of trips are between 0 to 5 in a year and those trips are basic package trips.

OBSERVATIONS

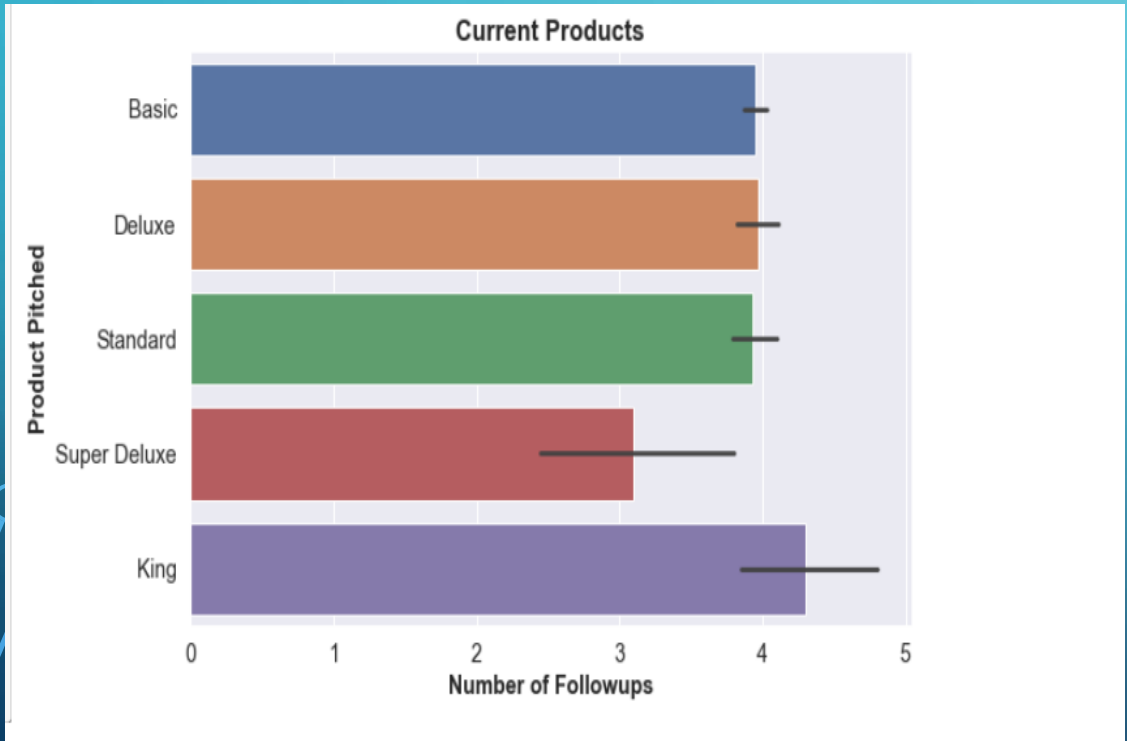
- This plot speaks for itself. Executives: Basic package, Manager: Deluxe Package, Senior Manager: Standard Package, AVP: Super Deluxe, VP: King. Real life isn't like this, but for the dataset provided, a customers Designation is all that would be needed to determine with level of package to pitch.



A majority of customers required 3-5 follow-ups before making a purchase

OBSERVATIONS

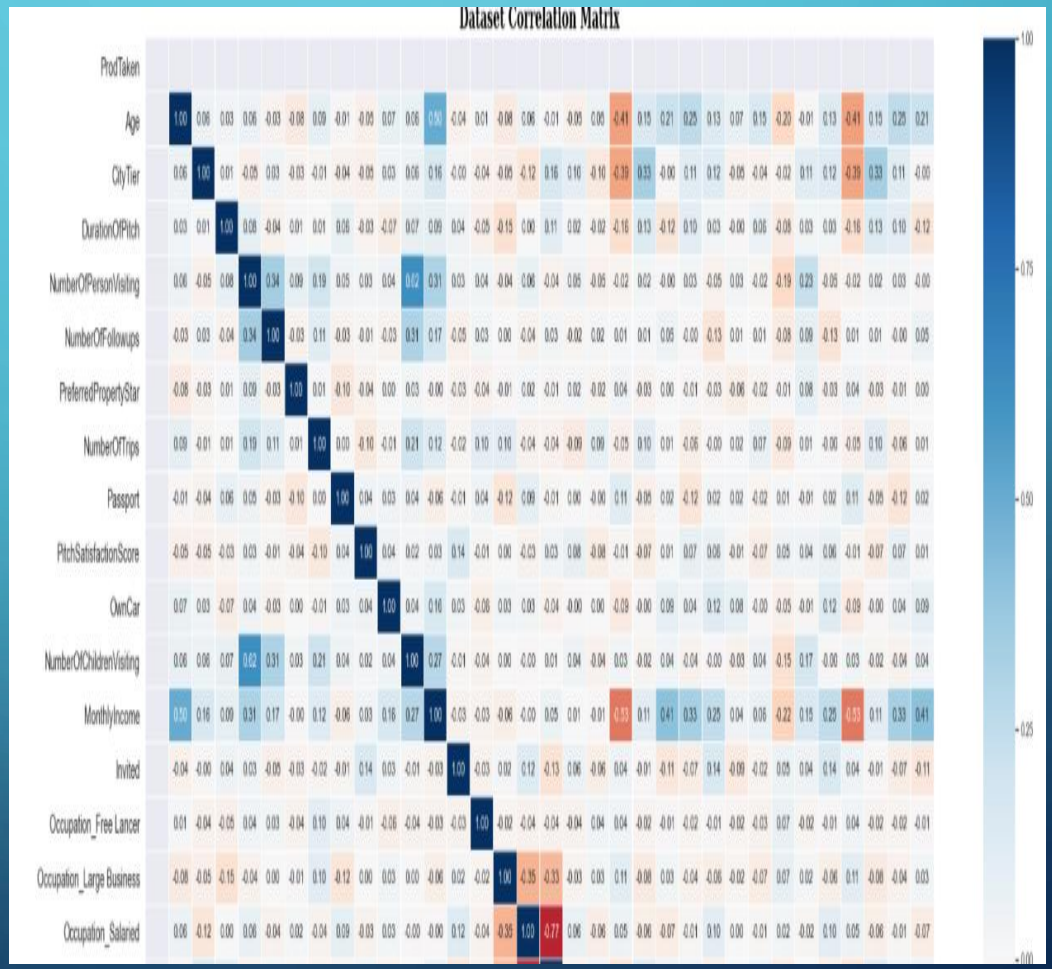
- Most customers required 3-5 follow-ups before making a purchase.



Marital status of customer who has purchased a basic package no Basic, Deluxe, and Standard packages on average required 4 follow-ups before purchase made. Super Deluxe on average required 3 follow ups before purchase made. And King on average required 4-5 follow-ups. ** there may be an indicator here on the Super Deluxe package performance - the lower-than-average follow-ups may indicate that low sales might be due to not following up enough on Single and Married.

EXPLORATORY DATA ANALYSIS (BIVARIATE)

- This graph is of correlation between variables in the form of heat map.
- There is a high correlation between Age ,Monthly income and Number of children visiting columns.



EXPLORATORY DATA ANALYSIS (BIVARIATE)

- Majority of customers own a car, but those that do not are likely to purchase lower end packages, especially the basic package.

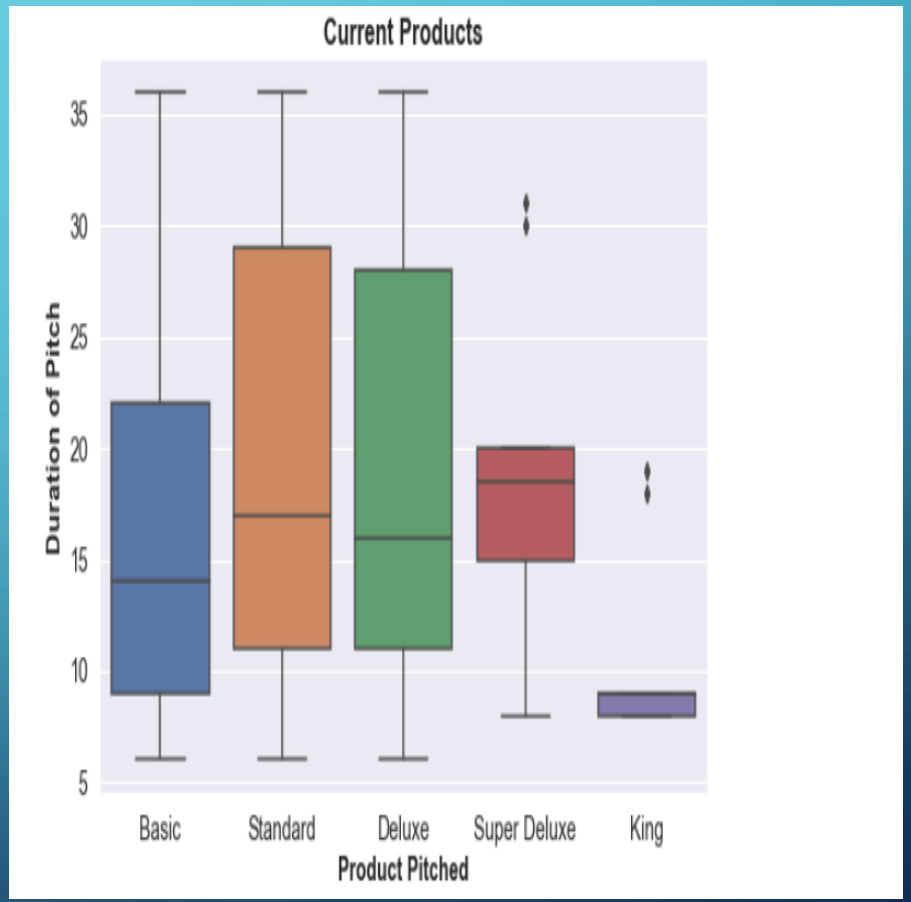
This plot demonstrates that there was not a set duration of pitch and that each pitch went from about 5 to 35 minutes. For super deluxe and king packages, it is likely that the durations would have followed suit had there been a greater sample of customers who purchased those packages.

EXPLORATORY DATA ANALYSIS (BIVARIATE)

The mean pitch durations for the standard travel package is statistically different than the mean duration for the Basic package but is not different than the Deluxe or Super Deluxe, even though those two packages have a statistically similar mean to the basic travel package.

This implies that there is not in-fact much difference between the means, but an overall difference in the IQR of the two distributions.

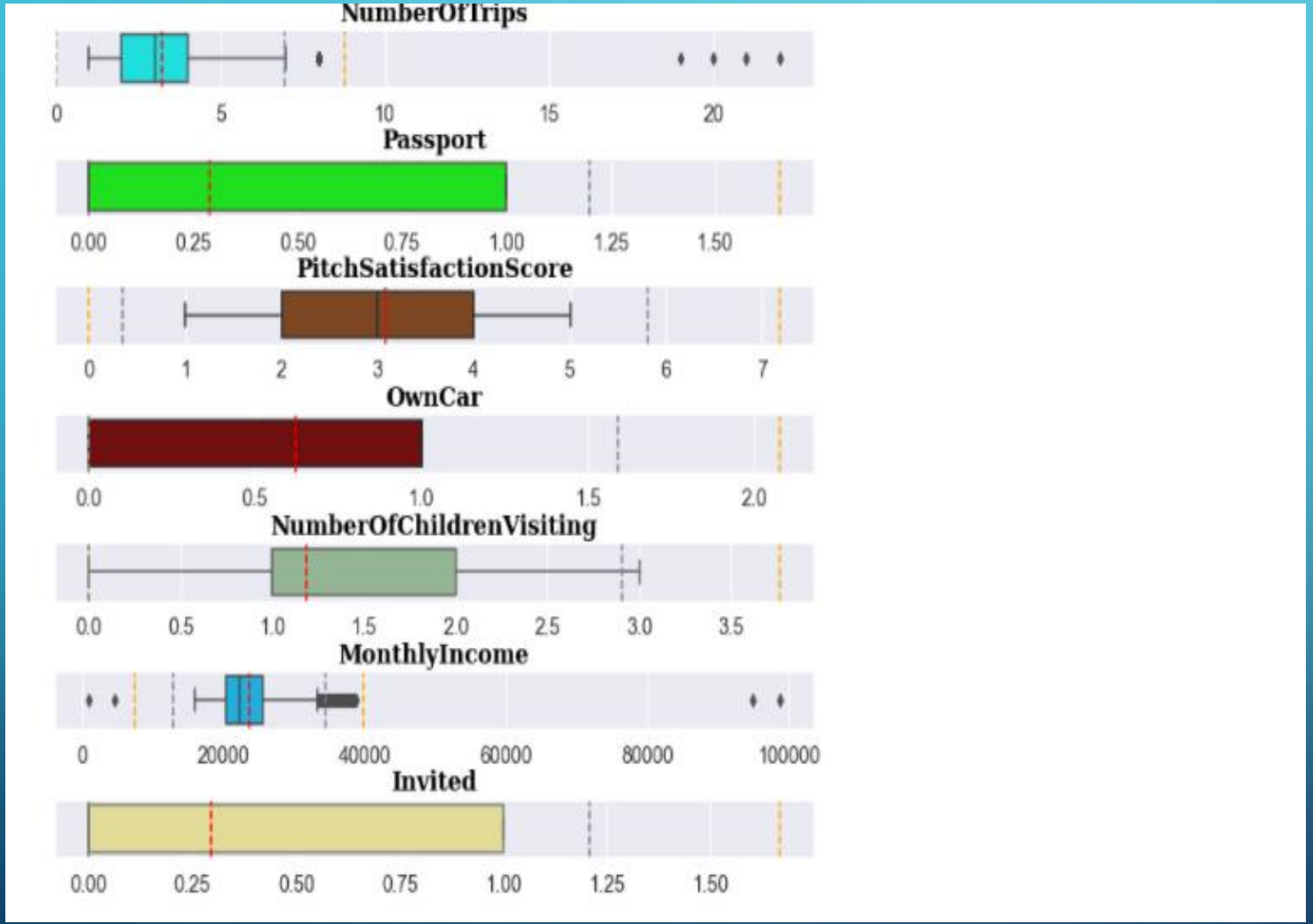
The mean pitch duration for the king travel package was statistically less than all other packages.



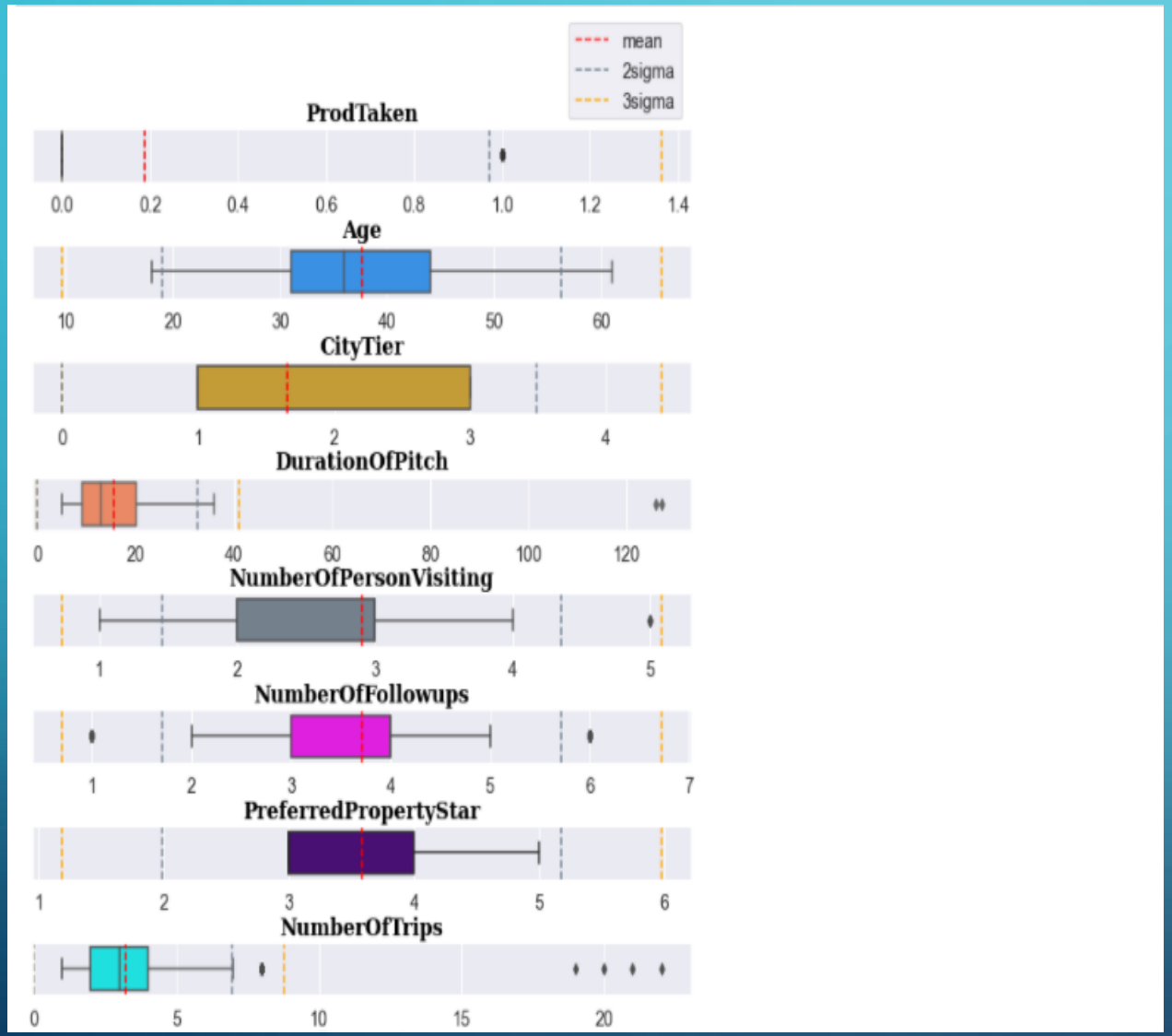
INFERENCE FROM BIVARIATE ANALYSIS

- There is a higher correlation in Product pitched basic, Product pitched Deluxe, Product pitched King, Product pitched Standard and Number of children visiting columns.it impact on performance as compared to others.
- Correlation coefficient of Age, city tier and product pitched basic variable is negative and close to zero so we can drop the variable. Correlation coefficient are close to zero so we can drop these variables as well.
- There was not a set duration of pitch and that each pitch went from about 5 to up to 35 minutes. For super deluxe and king packages, it is likely that the durations would have followed suit had there been a greater sample of customers who purchased those packages.
- The mean pitch durations for the standard travel package is statistically different than the mean duration for the Basic package but is not different than the Deluxe or Super Deluxe, even though those two packages have a statistically similar mean to the basic travel package.

Try to detect the outliers



There are some outliers here which don't need to be treated for decision tree, bagging, and random forest. Normally they should be treated when using Boosting models because those models are not robust to outliers due to the algorithm in which successive models place preference on the worst performing for successive iterations, which could lead to model overfitting.



For several reasons, the outliers are not going to be treated in this case:

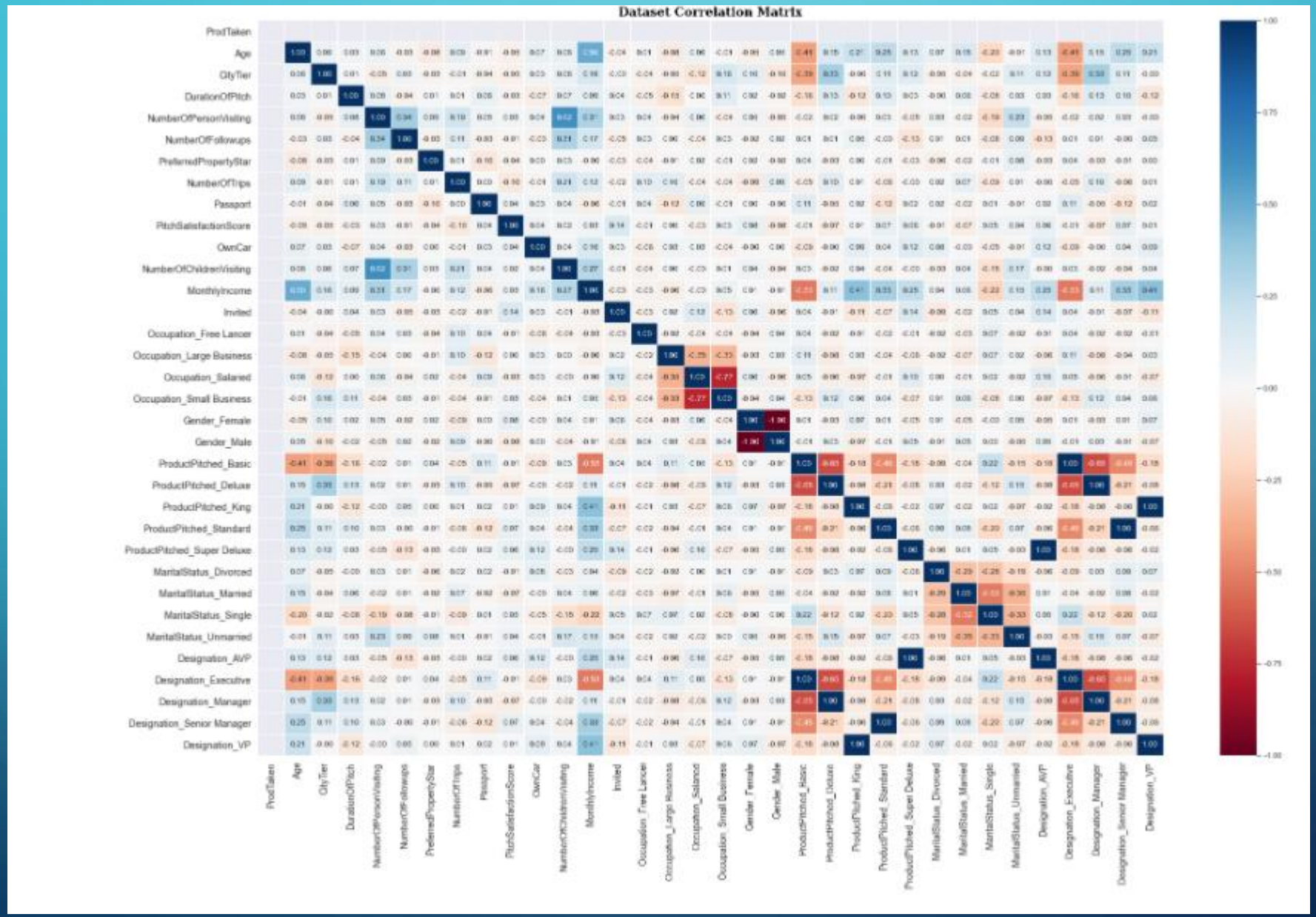
1) There's not many outliers, and importantly,

2) outlier treatment might affect the outcome of the weighting and analysis performed here.

Outliers will be left untouched though it will be presumed that Boosting may have more overfitting than they would if the outliers were treated.

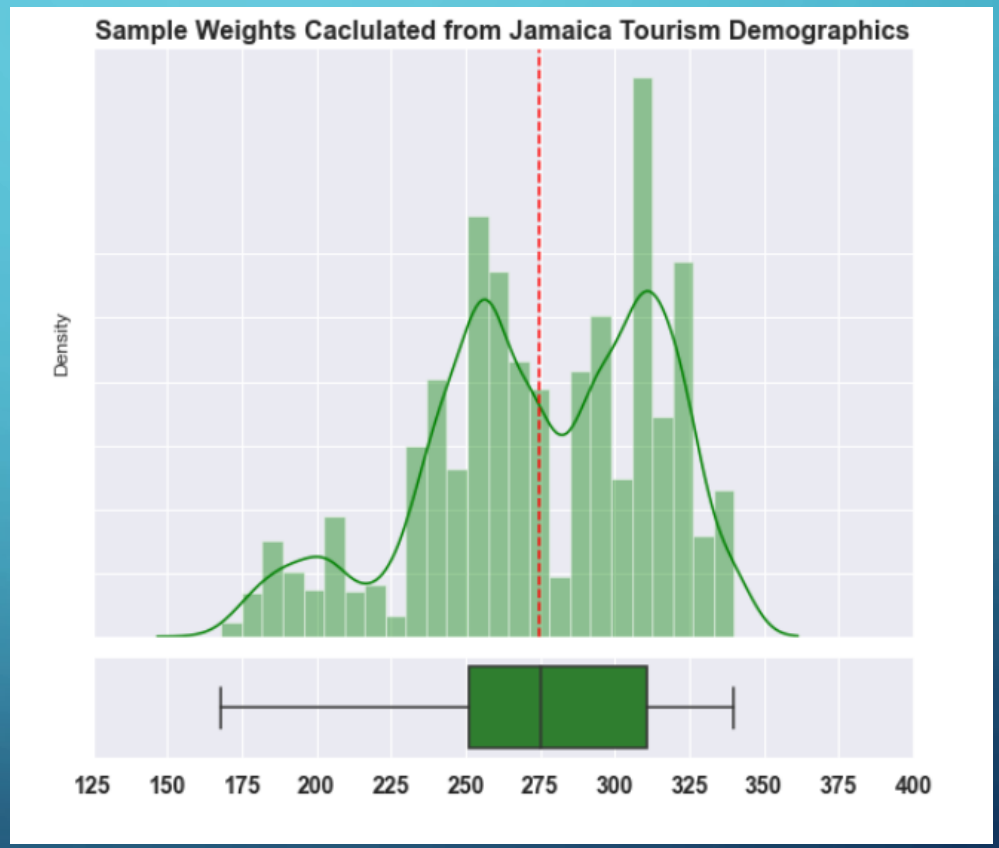
OBSERVATIONS

- The basic travel package has the highest conversion rate, at about 30% of total pitches, followed by the Standard at about 15%, then the Deluxe (10%) and King (8%). The super-deluxe package had the worst performance, though it is not clear why.
- It may be that the super deluxe package is not appealing to customers and should not be offered anymore, or the pitch needs to be improved upon and customers selected in a more methodical way. For all of the lower performing packages, it is recommended that their performance be further studied to find areas of improvement.


Dataset Correlation Matrix
ProdTaken
Age
CityTier
DurationOfPitch
NumberOfPersonVisiting
NumberOfFollowups
PreferredPropertyStar
NumberOfTrips
Passport
PitchSatisfactionScore
OwnCar
NumberOfChildrenVisiting
MonthlyIncome
Invited
Occupation_Free Lancer
Occupation_Large Business
Occupation_Salaried
Occupation_Small Business
Gender_Female
Gender_Male
ProductPitched_Basic
ProductPitched_Deluxe
ProductPitched_King
ProductPitched_Standard
ProductPitched_Super Deluxe
MaritalStatus_Divorced
MaritalStatus_Married
MaritalStatus_Single
MaritalStatus_Unmarried
Designation_AVP
Designation_Executive
Designation_Manager
Designation_Senior Manager
Designation_VP
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00

SAMPLE WEIGHTS CALCULATED FROM JAMAICA TOURISM DEMOGRAPHICS

- As demonstrated in the graph, the sample-weight distributions are spread out well, though they are left-skewed toward higher weights.



PREDICTION OF RELIABLE MODEL ACCURACY FROM SAMPLE WEIGHTS

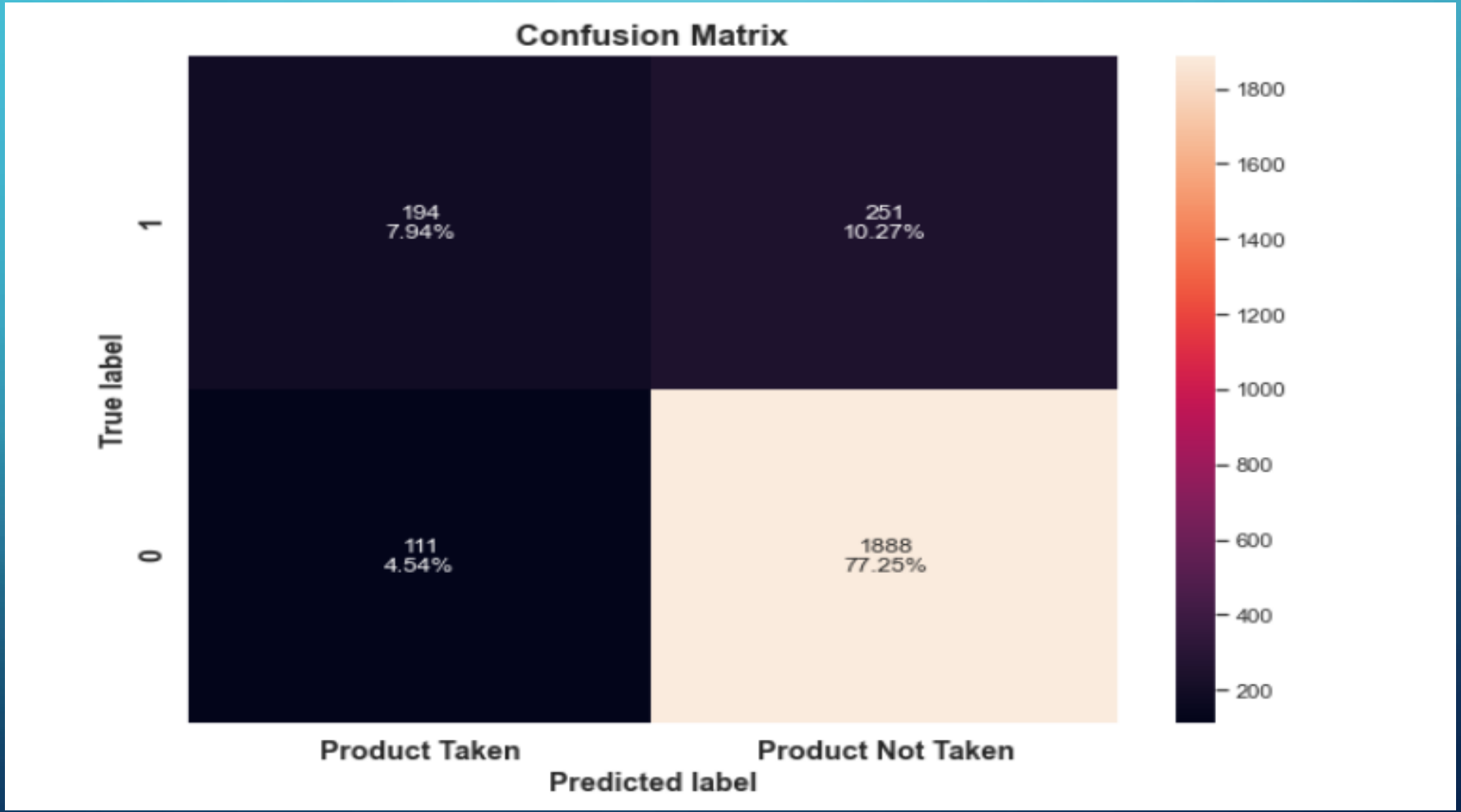
- Observation

It is presumed that a good model with good predictions for wellness tourism from the sample weights and current travel packages dataset should be about 90% accurate (meaning that about 10% of the predictions do not fit the educated guess of the wellness tourism demographics).

- Decision Tree Classification Model

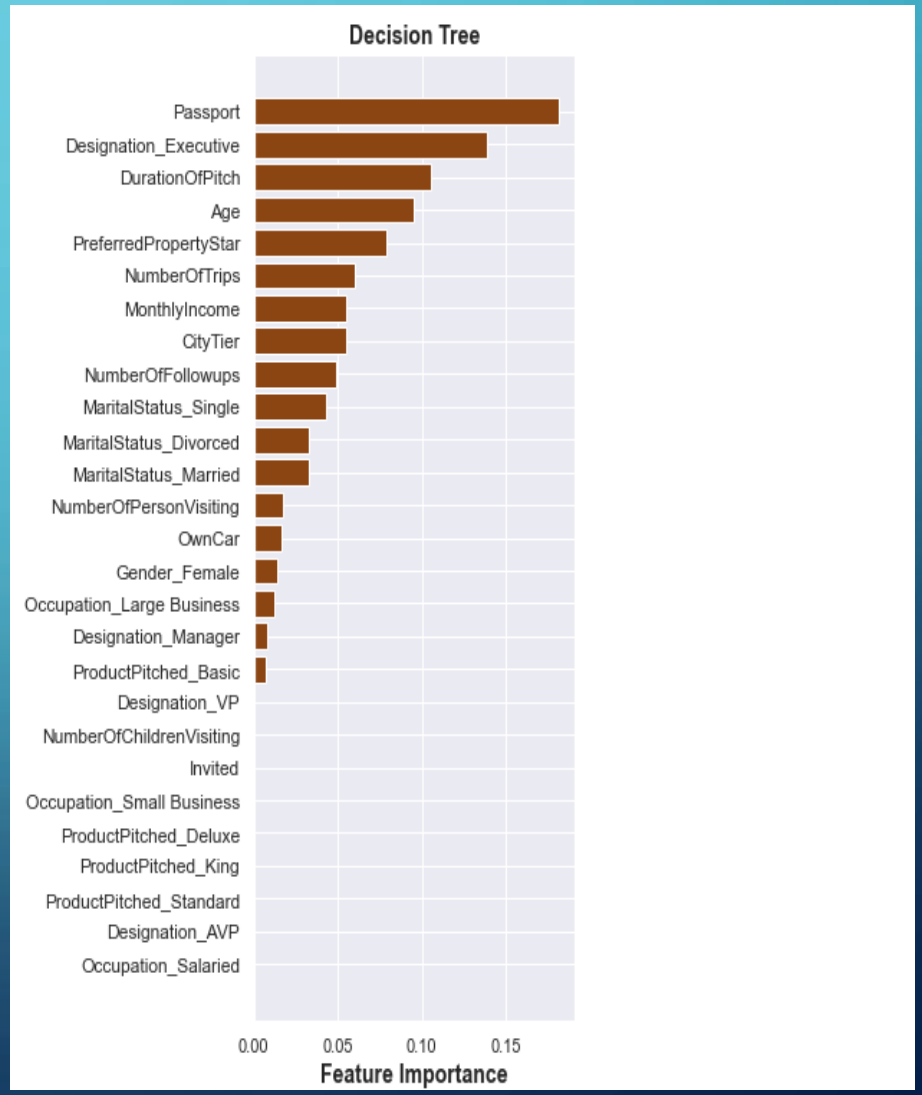
Decision Tree accuracy on the testing data was at about 85% with is slightly lower than the 90% expected rate. The recall is quite low, at around 40% (worse than what would be expected for a random / chance event). The Feature importance plot demonstrates that a little more than half of the features were used toward making predictions in this model.

```
Decision Tree -- Accuracy: 0.852 / Precision: 0.636 / Recall: 0.436 / Latency: 16.179ms
```



FEATURE IMPORTANCE

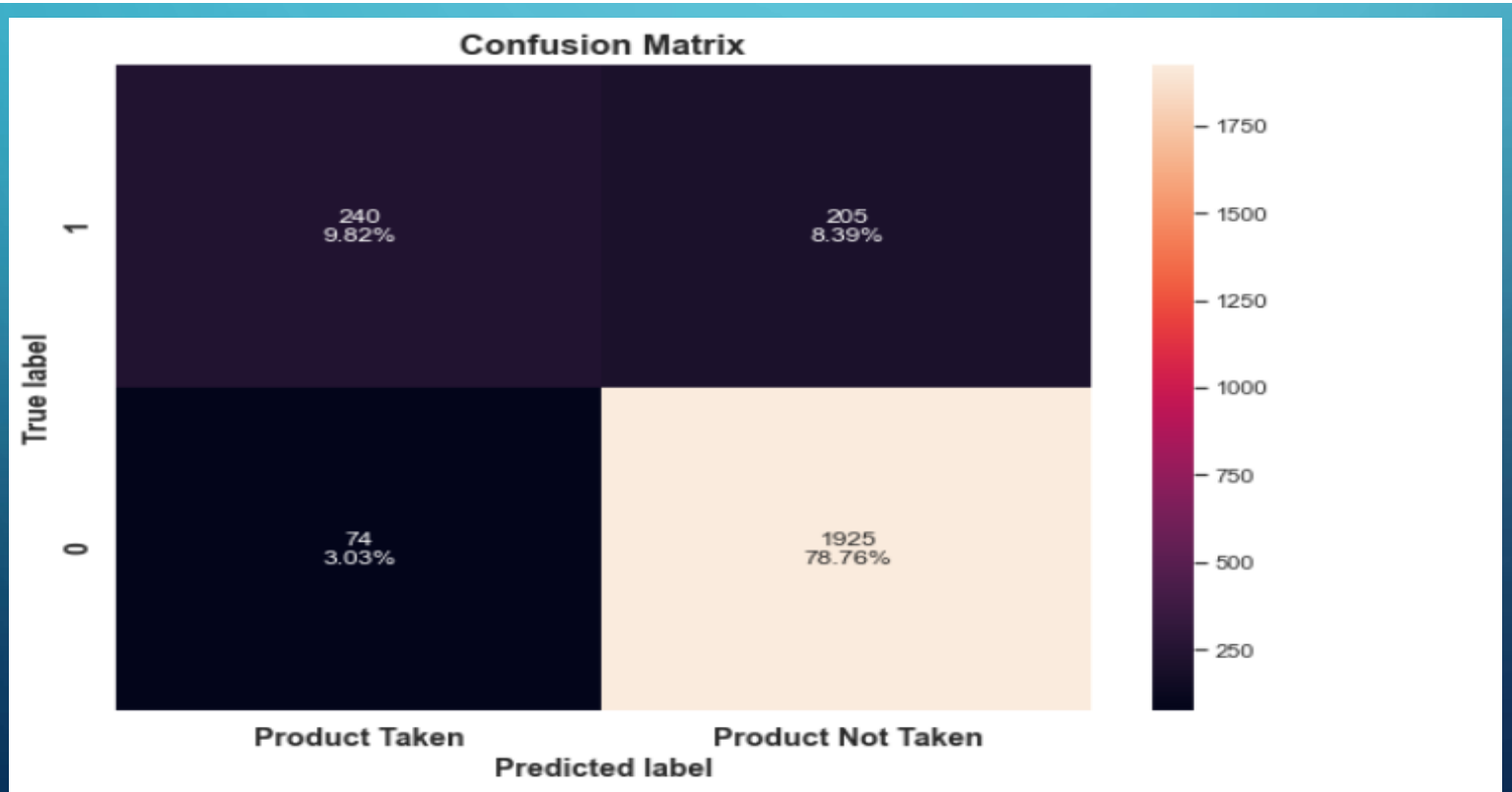
- Most important feature are Passport and Designation Executive and duration of pitch Age.
- And which are not important for Decision Tree are ProductPitched_Basic, Designation Manager and Gender Female



- Bagging Classifier

Accuracy of the Bagging Classifier model was at about 90% which was what was predicted as a reasonable accuracy for the weighted model. Recall saw an improvement over the Decision Tree model, but it is still quite low (50%) at about equal to a random event. Latency of the prediction is much greater than the decision tree, but it is not clear whether latency is important here.

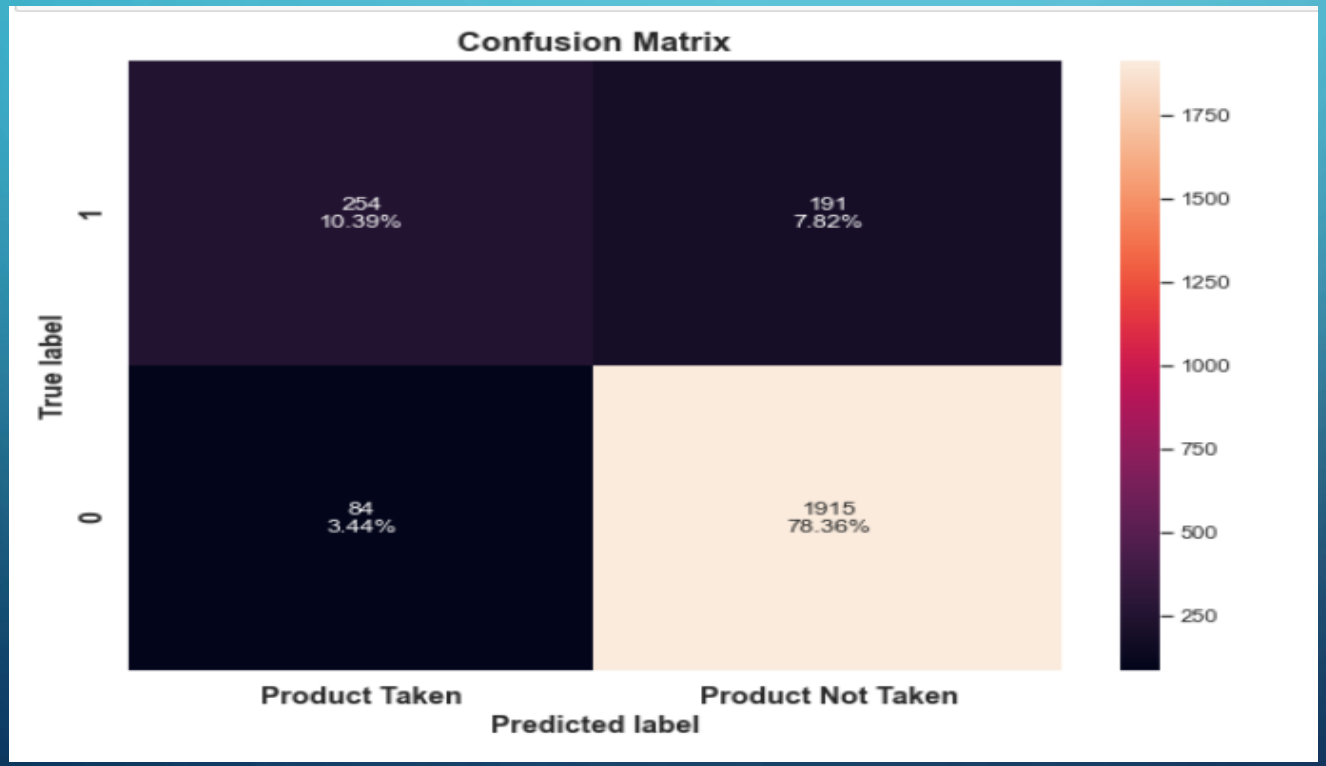
```
Bagging -- Accuracy: 0.886 / Precision: 0.764 / Recall: 0.539 / Latency: 166.581ms
```



Random Forest

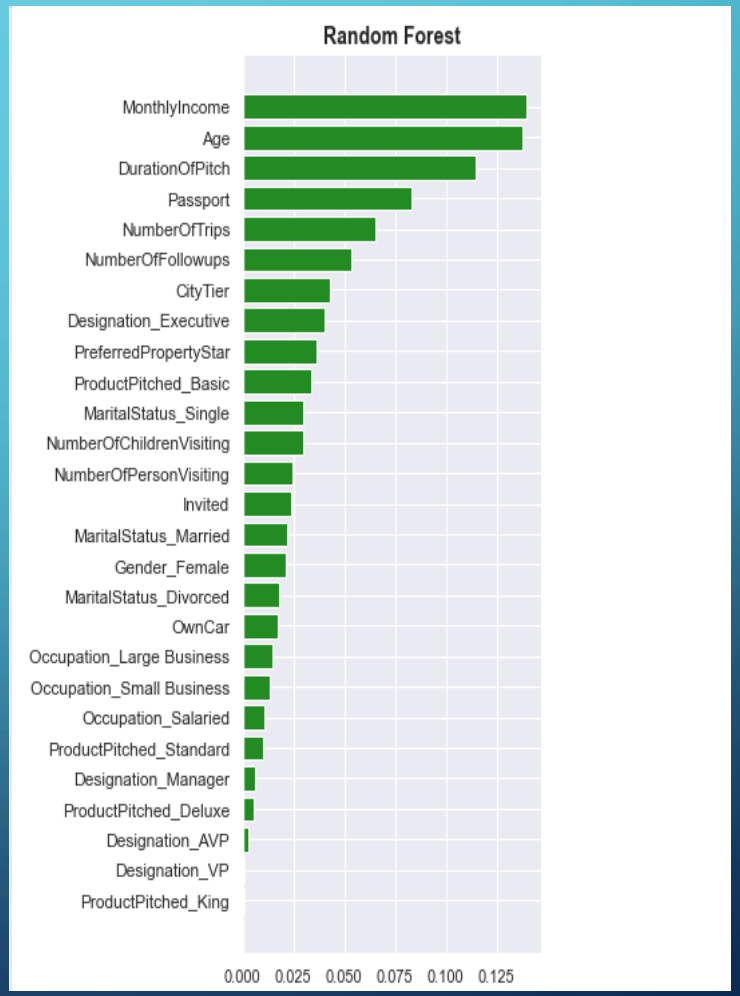
The Random Forest model has a near 90% accuracy rate consistent with the expected accuracy for the weighted model, like Bagging. Recall is better than random at 57%, though not by much. Analysis of feature importance indicates that all features contributed to the model, except Designation_VP (and most significantly contributed).

```
Random Forest -- Accuracy: 0.887 / Precision: 0.751 / Recall: 0.571 / Latency: 113.03ms
```



FEATURE IMPORTANCE

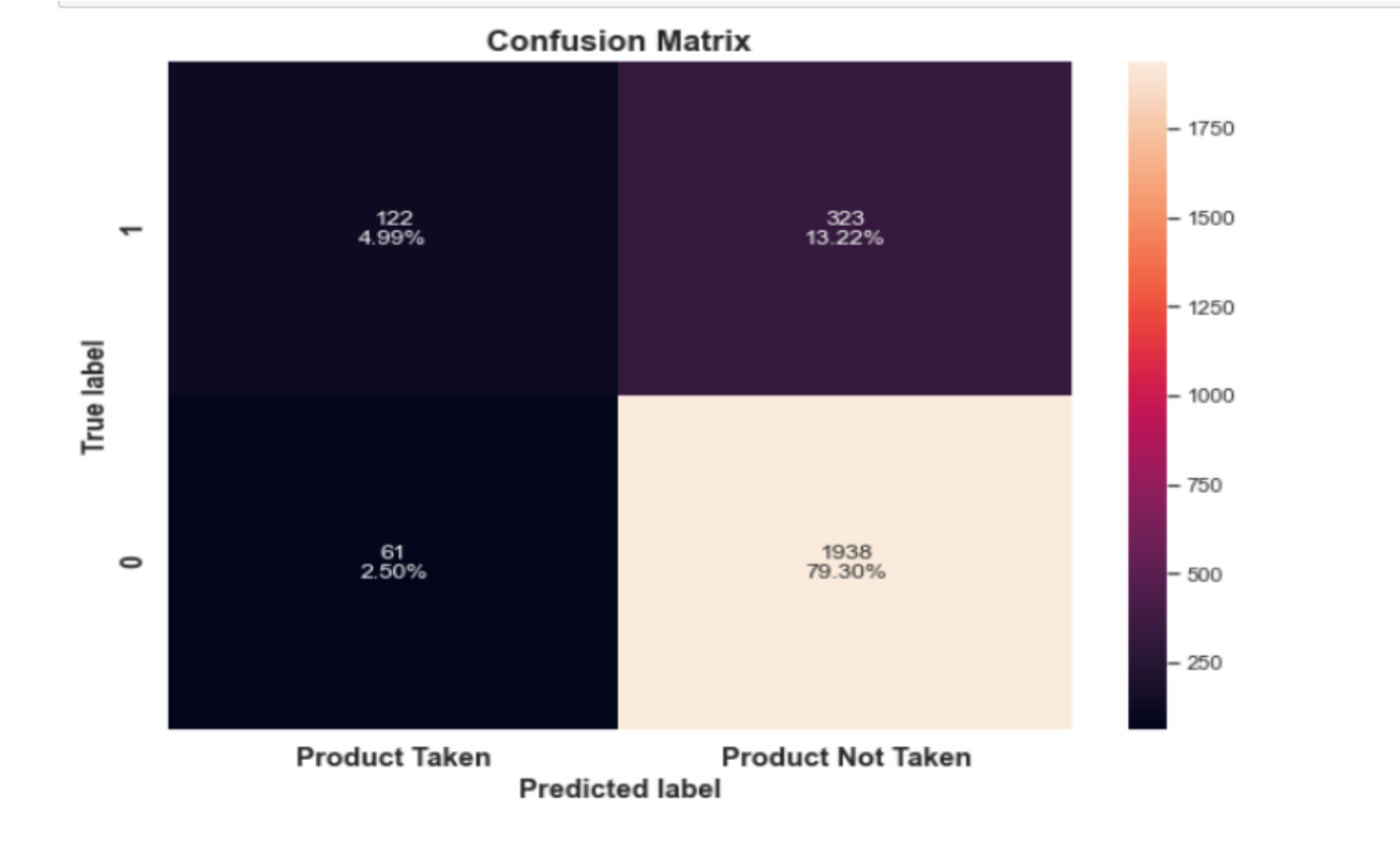
- Most important feature are Passport and Monthly Income and Age and duration of pitch.
- And which are not important for Random Forest are Own Car, Designation Manager and Gender Female.



- Ada boost

Model performance is terrible, with recall at 28%. This may be due to not having treated outliers, but more than likely, the algorithm performs in a way that just doesn't mesh well with the sample weights and distributions of the features of this dataset.

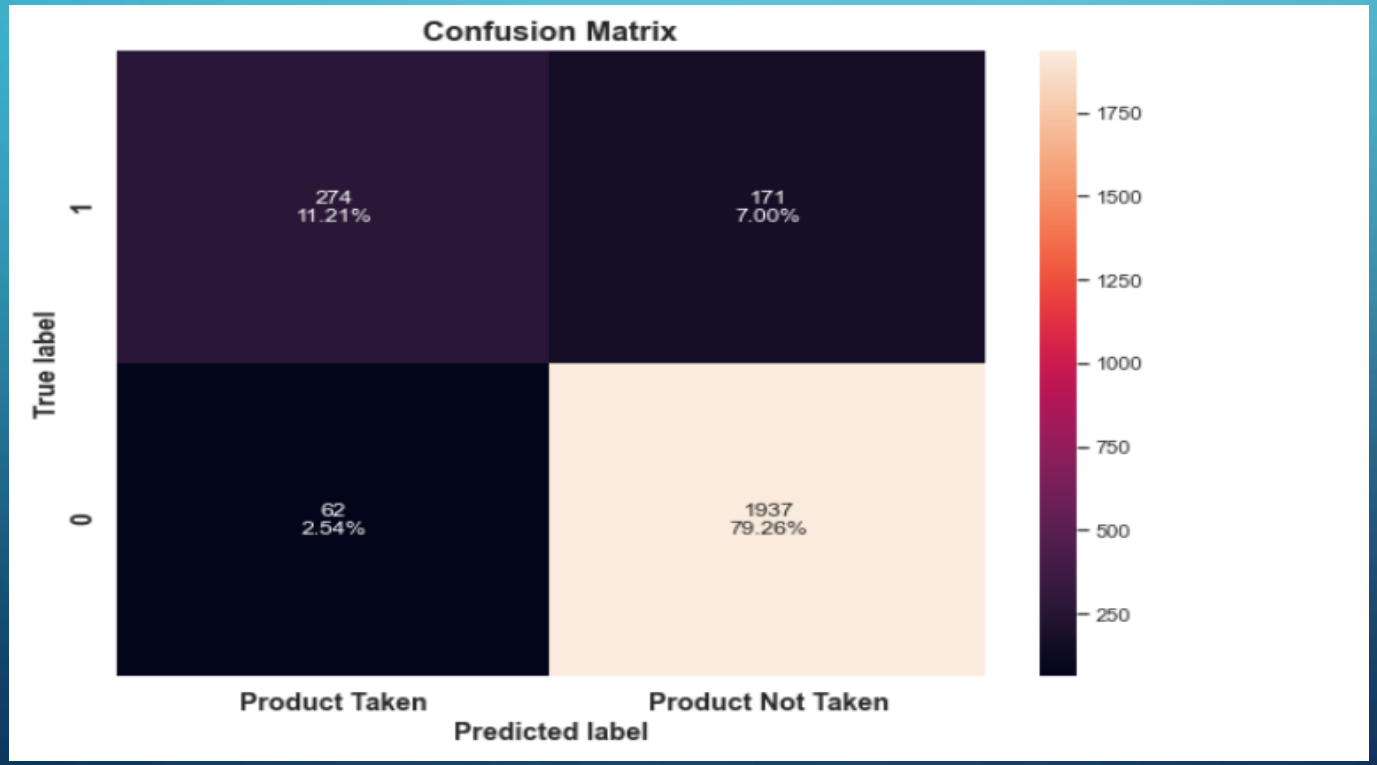
```
AdaBoost -- Accuracy: 0.843 / Precision: 0.667 / Recall: 0.274 / Latency: 132.205ms
```



- Gradient Boosting Classifier

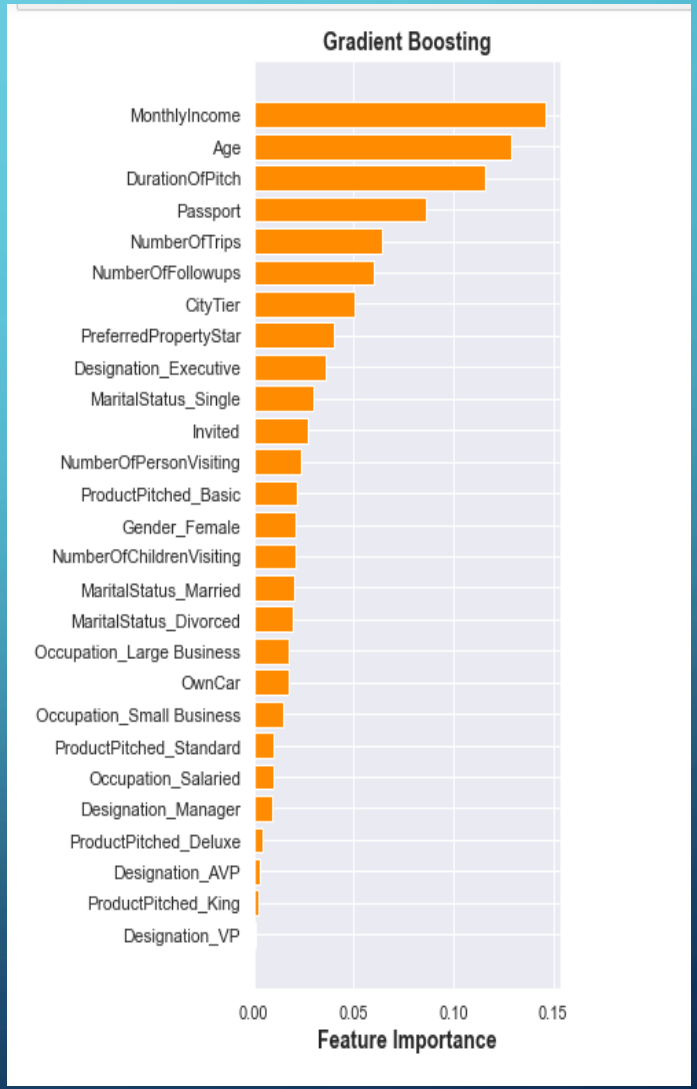
This model has slightly higher than predict accuracy but that is not necessary a bad thing. Recall at 63% is an improvement over other models (especially the 23% offered by AdaBoost). The model is using most features in its prediction.

```
Gradient Boosting -- Accuracy: 0.905 / Precision: 0.815 / Recall: 0.616 / Latency: 42.658ms
```



FEATURE IMPORTANCE

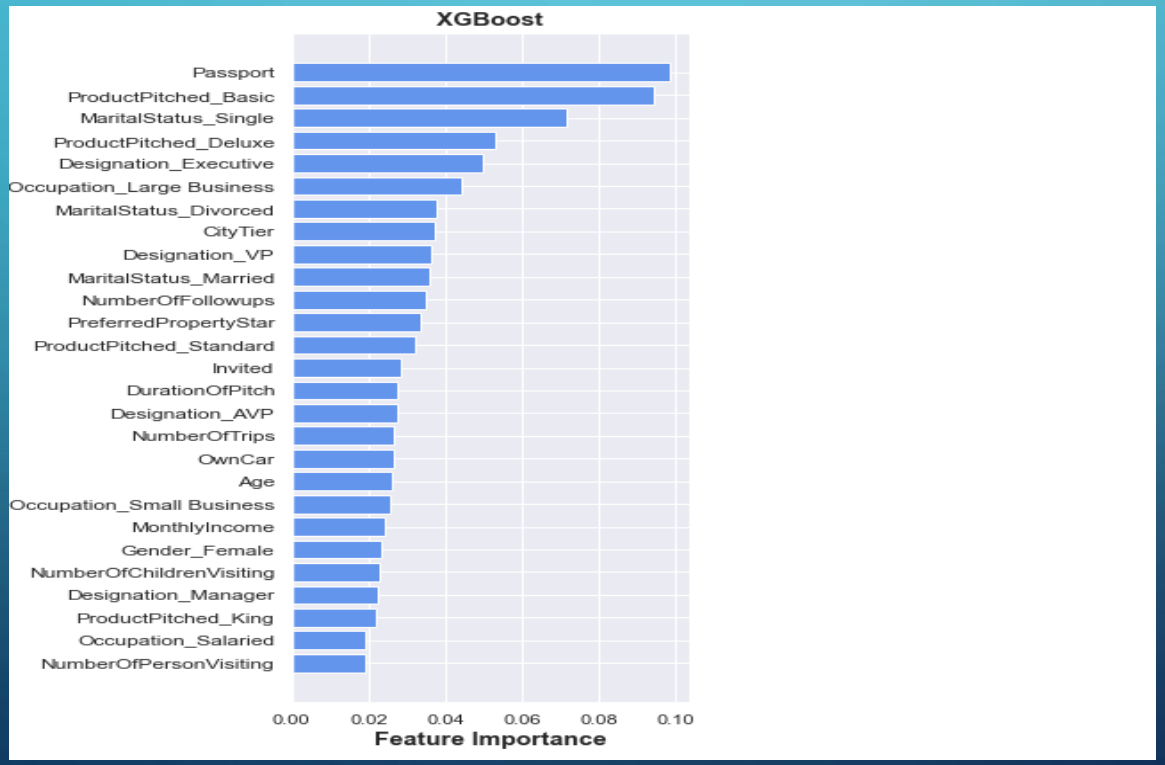
- Most important feature are Passport and Monthly Income and Age and duration of pitch.
- And which are not important for Gradient Boosting are Own Car, Designation Manager and Gender Female.



- XGBoost Classifier

Accuracy near prediction 90% but recall lower than the Gradient Boosting model. Feature importance indicates that the model is using all the features of the dataset in making its predictions.

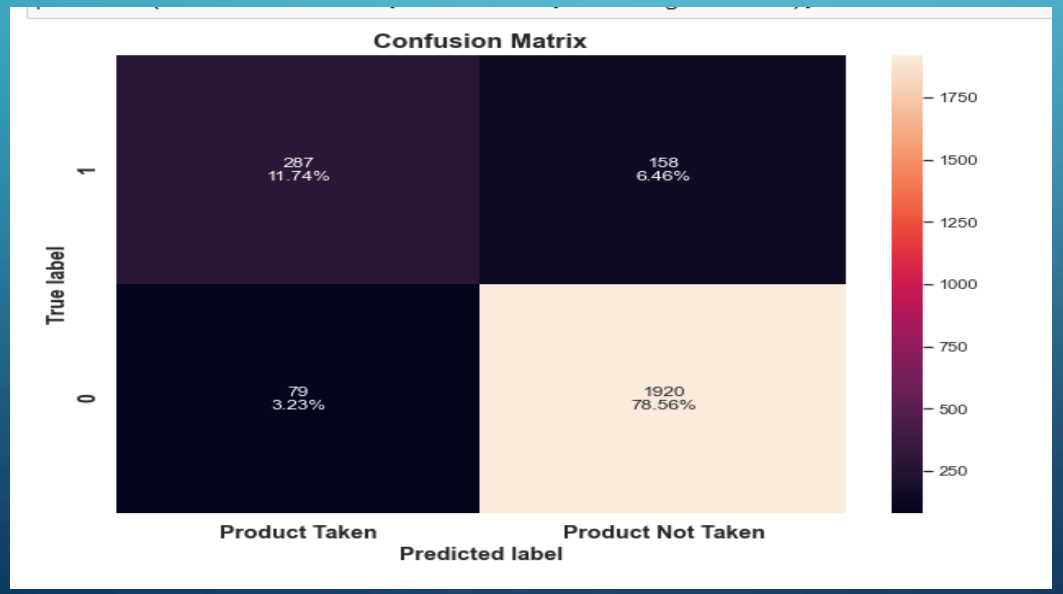
```
XGBoost -- Accuracy: 0.898 / Precision: 0.783 / Recall: 0.607 / Latency: 53.136ms
```



- Stacking Classifier

The stacking classifier is achieving a 90% accuracy rate, which is where it should be based upon predictions using the sample weights. Most import, the stacking classifier is achieving a near 72% Recall, which is the best model and therefore the model that will be used going forward. Latency was sacrificed in using the stacking classifier, which is expected considering that it is employing 4 models to make its predictions.

```
Stacking -- Accuracy: 0.903 / Precision: 0.784 / Recall: 0.645 / Latency: 216.693ms
```



SUMMARY OF MODEL PERFORMANCES

- All the models except the Decision Tree and Adaboost classifier appear to be slightly overfitting. Adaboost should not be seriously considered given its recall scoring.
- For the remainder of the models, the appearance of overfitting may indicate that the sample weights are being countered by the nature of the error correcting process of the algorithms.
- The Stacking model is the best performing model here and predictions from this model will be utilized going forward to Exploratory Analysis of the predicted subset.

SUMMARY AND RECOMMENDATIONS FOR THE FUTURE BUSINESS

- Age:
  - Basic - Predominately in 20s, also 30s and 40s
  - Standard & Deluxe - Predominately in 30s, also 40s
  - Super Deluxe - Predominately in 40s, also 50s
  - King - 30s, 40s, but predominately in 50s

- Income:
  - Basic - 15k - 25k
  - Deluxe - 20k - 30k
  - Standard - 20k - 35k
  - Super Deluxe - 25k - 35k
  - King - 30k - 45k

- Marital Status:
  - Basic - Single, Divorced, Unmarried, & Married
  - Deluxe - Single, Divorced, Unmarried, & Married
  - Standard - Divorced, Unmarried, & Married
  - Super Deluxe - Single & Married
  - King - Single, Divorced, & Married

- City Tier
  - Basic - 1, 2, & 3
  - Deluxe - 1 & 3
  - Standard - 1 & 3
  - Super Deluxe - 3
  - King - 1 & 3
  - Gender No Gender Specificity

- Person
  - Basic, Deluxe, Standard, & King - 2, 3, & 4
  - Super Deluxe - 2 & 3
- Occupation
  - Basic, Deluxe, Standard, & King - Salaried, Small Business, and Large Business
  - Super Deluxe - Salaried & Small Business
  - Freelancers very unlikely to purchase travel package
- Preferred Property Scores
  - Basic - 3, 4, & 5
  - Deluxe, Standard, and Super Deluxe - Predominately 3
  - King - 3 & 4 but predominately 4

- Passport
  - Basic more likely for International Travel
  - Designation
  - Basic - Executive
  - Deluxe - Manager
  - Standard - Senior Manager
  - Super Deluxe - AVP
  - King - VP

- Follow-ups
  - Expect 3-5 follow-ups before purchase
  - On Average:
  - Basic, Deluxe, Standard, & King - 4 follow-ups
  - Super Deluxe - 3 ** There is some indication that by increasing follow-ups to at least 4, this may result in an uptick in sales.

- Pitch Satisfaction Score

  There appears to be a problem with the survey for pitch satisfaction which is indicating satisfaction scores that are inconsistent with purchasing behavior. This may be due to a misunderstanding in which rating is good or bad. Recommend switch to Excellent, Good, Average, Poor scoring to get better feedback.

- # of children visiting
  - More sales when <= 2, but overall indeterminate.
  - Own Car
  - Super Deluxe & King - Yes
  - All others, indeterminate

- Pitch Time Recommendations: Basic, Standard, Deluxe, and Super Deluxe - ~20 minutes
- King - ~10 minutes

- Conversion Rates
- Basic - 30%
  - Standard - 15%
  - Deluxe - 10%
  - Super Deluxe - 5% -->Uptick may be obtained by increasing # of follow-ups to at least 4
  - King - 8%

CONCLUSION

- As far as the model is concerned, 90% accuracy, 72% Recall, and 74% Precision was obtained, which was the best scores out of all the models that were evaluated.
- While this model may be sufficient to obtain better than random results for an initial marketing program, it is recommended that Visit With Us obtain data on Wellness Tourism package sales and customers as the product rolls out so that the model may be updated with actual data to obtain better predictions as soon as practicable.