TOURISM PROJECT

2021

OBJECTIVES

- To predict which customer is more likely to purchase the newly introduced travel package.
- To establish which variables are most significant.
- To determine which segment of customers should be targeted more.

BUSINESS PROBLEM OVERVIEW

BACKGROUND

- Visit With Us ia a Tourism company whose Executive Management seeks to expand their customer base and equally establish a viable business model using existing data of customers and potential ones as well.
- As a senior data scientist at Visit with Us,, I am tasked with coming up with a predictive model using available data and through indepth analysis, profer insightful recommendations to the Policy Makers of the firm as well as the Marketing Team regarding potential targets for a new travel package introduction to the market

SOLUTION APPROACH (MACHINE LEARNING)

- Define the problem and perform an Exploratory Data Analysis
- Illustrate the insights based on EDA
- Data pre-processing
- Model building – Decision Tree, Random Forest and Bagging Classifier and then Adaboost, GCBoost and XGBoost and Stacking Classifier
- Establish Model Parameters
- Model performance evaluation and Improvement via Tuning of Models
- Return the Most Import Features/Variables influencing Personal Loan Purchase
- Actionable Insights & Recommendations

BUSINESS IMPLICATIONS

- Optimize processes fo the subscription for ease of convertibility of targeted prospects to subscribe to travel packages
- Maximize Revenue and effectively design new packages to endear to unharnessed potential subscribers through more robust campaigns
- Increased Customer patronage, retention, referral and traffic which ultimately impacts on Visit_with_Us bottomline
- Tracking and Trapping new prospects with Targeted Marketing and Media Campaigns
- Optimize administrative and overhead costs by applying the Predictive model smartly on perculiarity basis
- Optimize Sales and Profit Margin
- Effective Allocation and redistribution of resources especially for product design, pricing and Ad-campaigns
- Developing potential markets for more revenue

DATA MANIPULATION

- Identification of Missing Values
- Fixing and Dropping Columns: (ID and NumberofChildrenVisiting) were dropped ultimately prior to Modelling
- Fixing specific sub-variables , Umarried('MaritalStatus)
- Replacing and merging sub-features on investigating unique values in Gender i.e ('Female' and 'Fe male')
- Conversion of Data types
- Prior to modelling, NumberofChildrenVisiting was dropped as a result of correlation with NumberofPersonsVisiting for a cleaner prediction

DATA INFORMATION

Variable	Description
ID	Unique Customer ID
Age	Customer's age in completed years
ProdTaken	Whether the customer has purchased a package or not (0: No, 1:Yes)
TypeofContact	How customer was contacted (Company Invited or Self Inquiry)
CityTier	City tier depends on the development of a city, population, facilities, and living standards. The categories are ordered i.e. Tier 1 > Tier 2 > Tier 3
Occupation	Occupation of customer
Gender	Gender of customer
NumberOfPersonVisiting	Total number of persons planning to take the trip with the customer
PreferredPropertyStar	Preferred hotel property rating by customer
MaritalStatus	Marital status of customer
NumberOfTrips	Average number of trips in a year by customer
Passport	The customer has a passport or not (0: No, 1:Yes)
OwnCar	Whether the customers own a car or not (0: No, 1:Yes)
NumberOfChildrenVisiting	Total number of children with age less than 5 planning to take the trip with the customer
Designation	Designation of the customer in the current organization
MonthlyIncome	Gross monthly income of the customer
PitchSatisfsctionScore	Sales pitch satisfaction score
ProductPitched	Product pitched by the salesperson
NumberOfFollowups	Total number of follow-ups has been done by the salesperson after the sales pitch
DurationOfSpeech	Duration of the pitch by a salesperson to the

Observations	Variables
4888	20

Float64	Int64
7 Age DurationOfSpeech NumberOfFollowups NumberOfTrips PreferredPropertyStar NumberOfChildrenVisiting MonthlyIncome	8 CustomerID ProdTaken CityTier NumberOfPersonVisiting Passport PitchSatisfactionScore OwnCar New_Price

Object (6)

TypeofContact
Gender
MaritalStatus
Occupation
Product Pitched
Designation

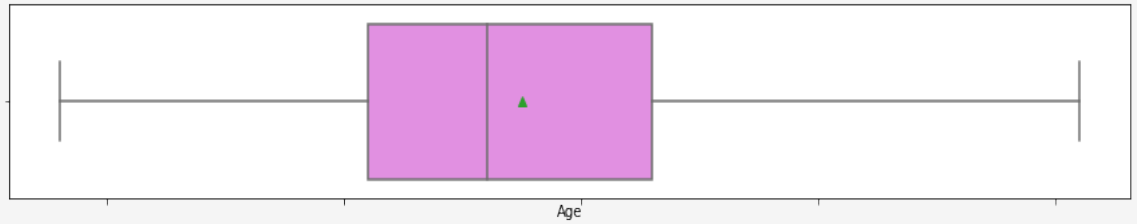
Missing Values in Data (Treated)

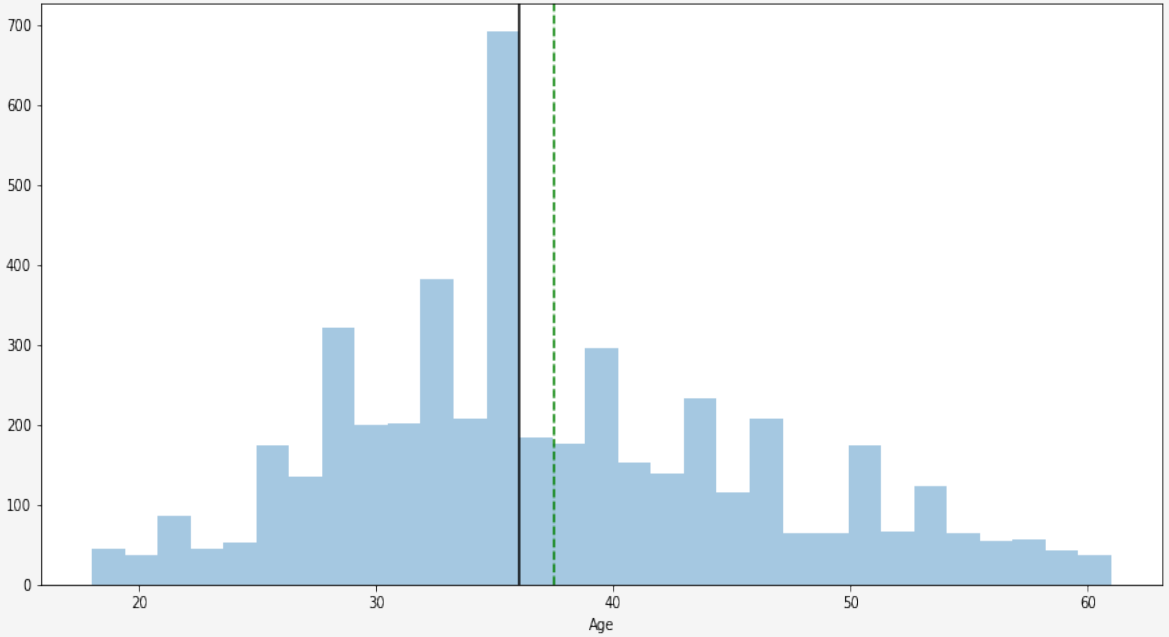
Arbitrary Values in Data
MaritalStatus ['Fe male']

Columns Dropped
ID and NumberofChildrenVisiting

EXPLORATORY DATA ANALYSIS

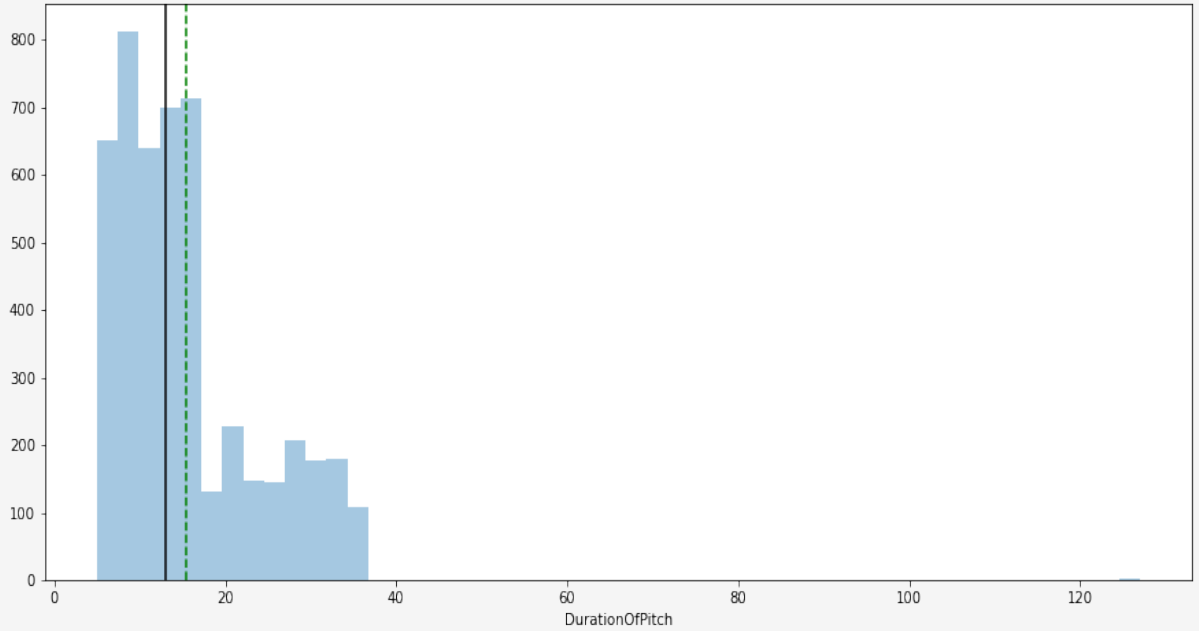
AGE





- Age is looking normally distributed, with a hint of right skew
- The customer distribution has a mean age of 37 yrs and a median of 36 yrs
- 75% of the customers are 43 yrs or less with the oldest at 61 years

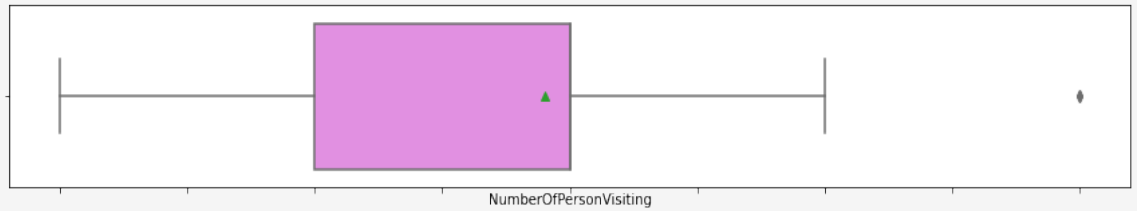
DURATION OF PITCH

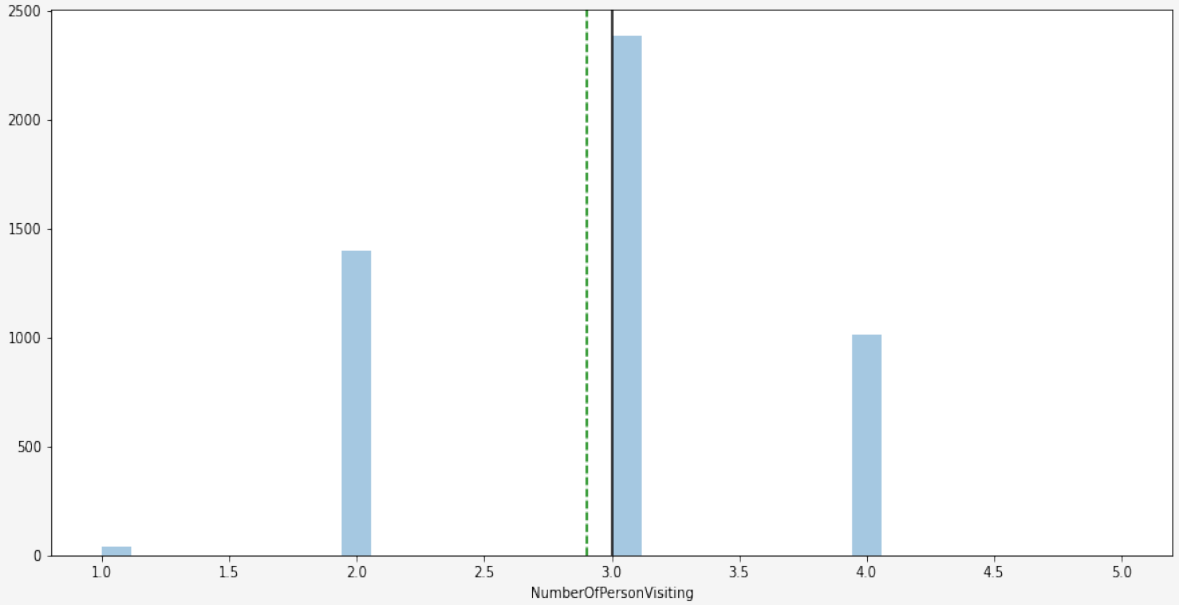


- The Duration of Pitch shows a presence of Outliers
- Some customers were pitched to for far longer periods than others
- The mean duration was 15 while 75% of the customers were accorded 19 or less with maximum at 127

EXPLORATORY DATA ANALYSIS

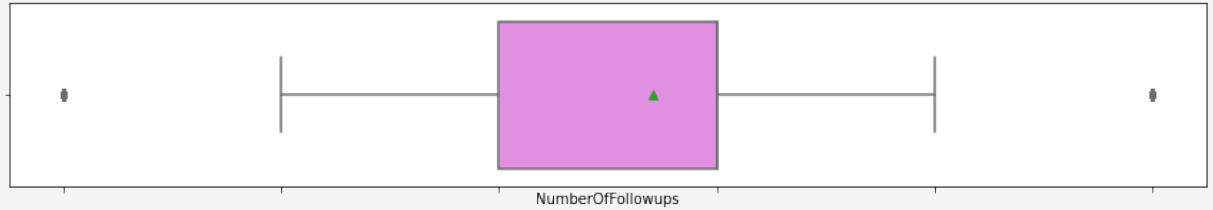
NUMBER OF PERSON VISITING

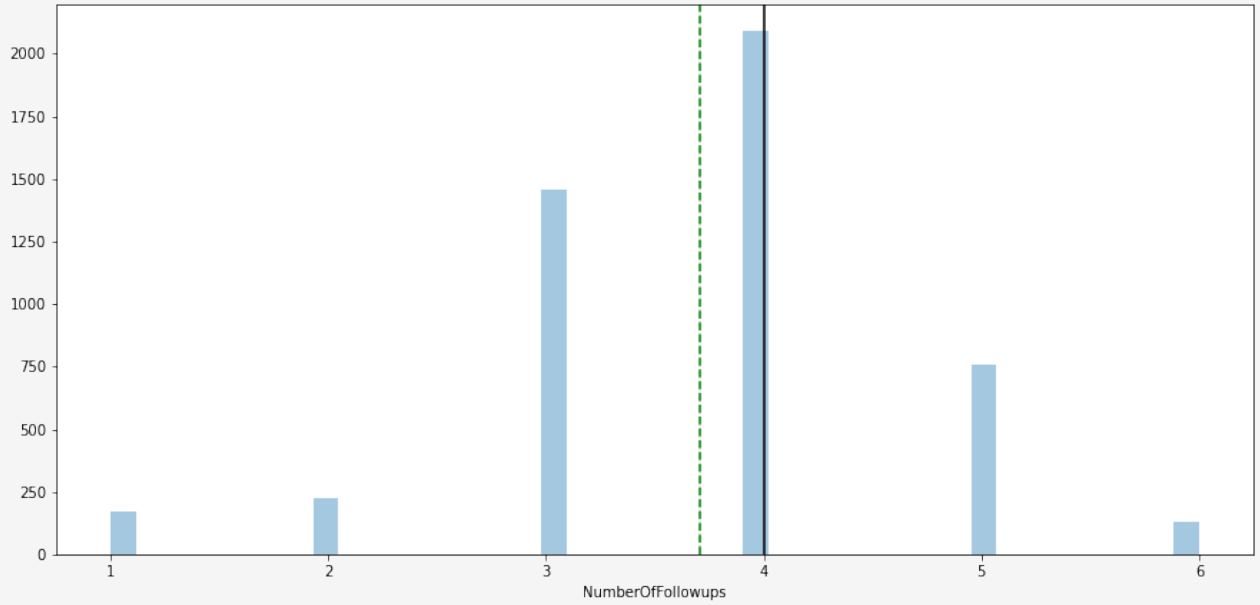




- This distribution has 3 peaks at 2,3 and 4.
- This clearly indicates that majority of the customers have between 2 to 4 people willing to go on a trip with them
- There is a presence of an outlier

NUMBER OF FOLLOWUPS

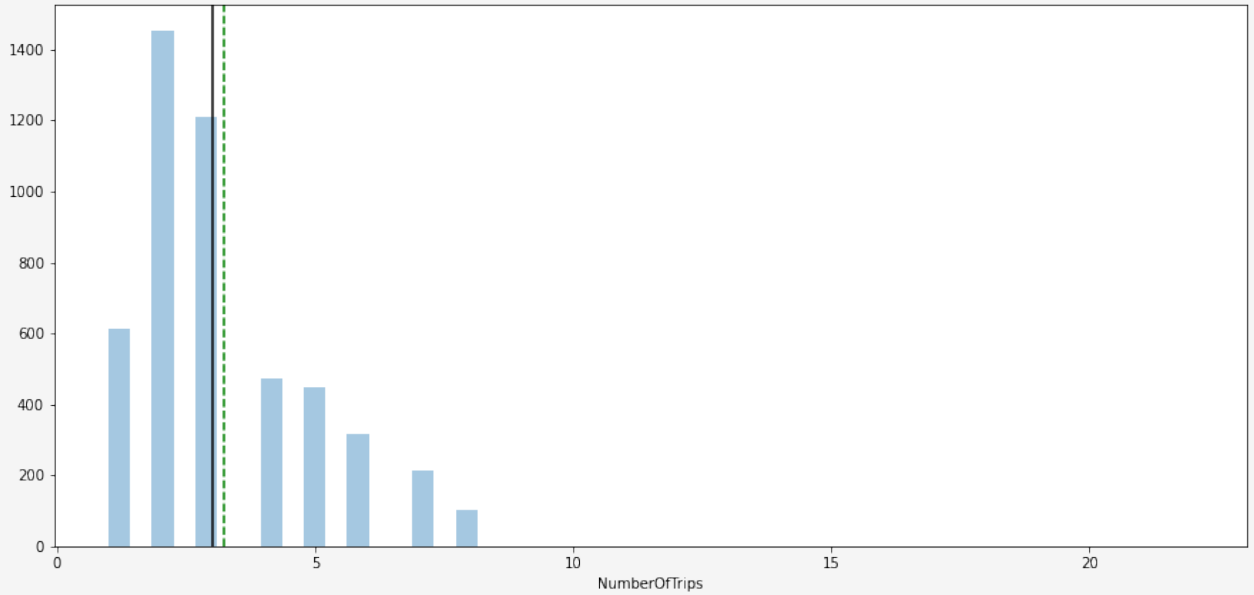




- The distribution for Number of followups has 3 peaks at 3, 4 and 5 with approximately 1400, 2000 and 750 customers respectively
- There is a presence of outliers

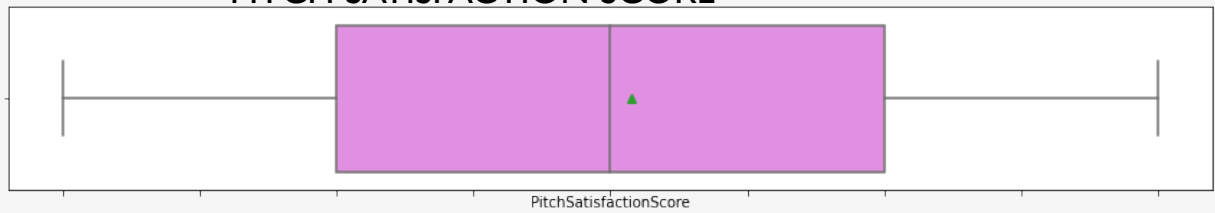
EXPLORATORY DATA ANALYSIS

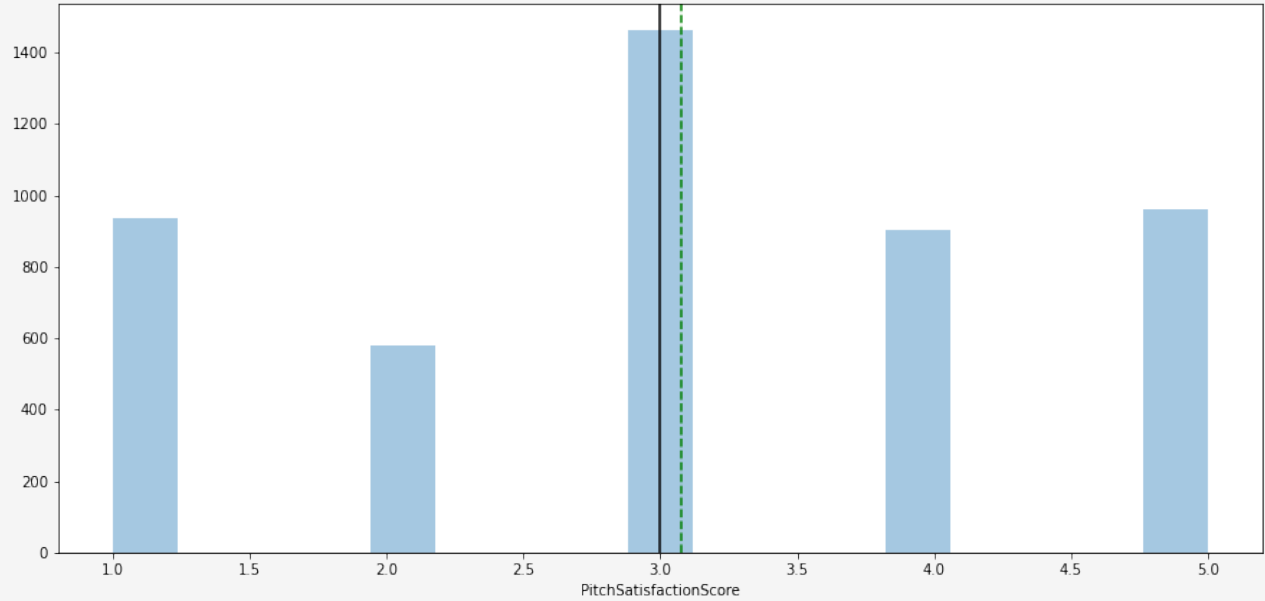
NUMBER OF TRIPS



- Number of Trips distribution is right-skewed with Mean > Median and 75% of the customers making a trip of 4 or less
- There are presence of outliers which irrefutably speaks to some customers taking far more trips than the lot

PITCH SATISFACTION SCORE

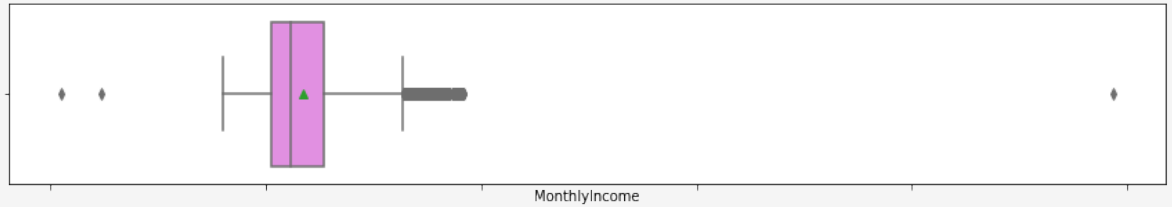


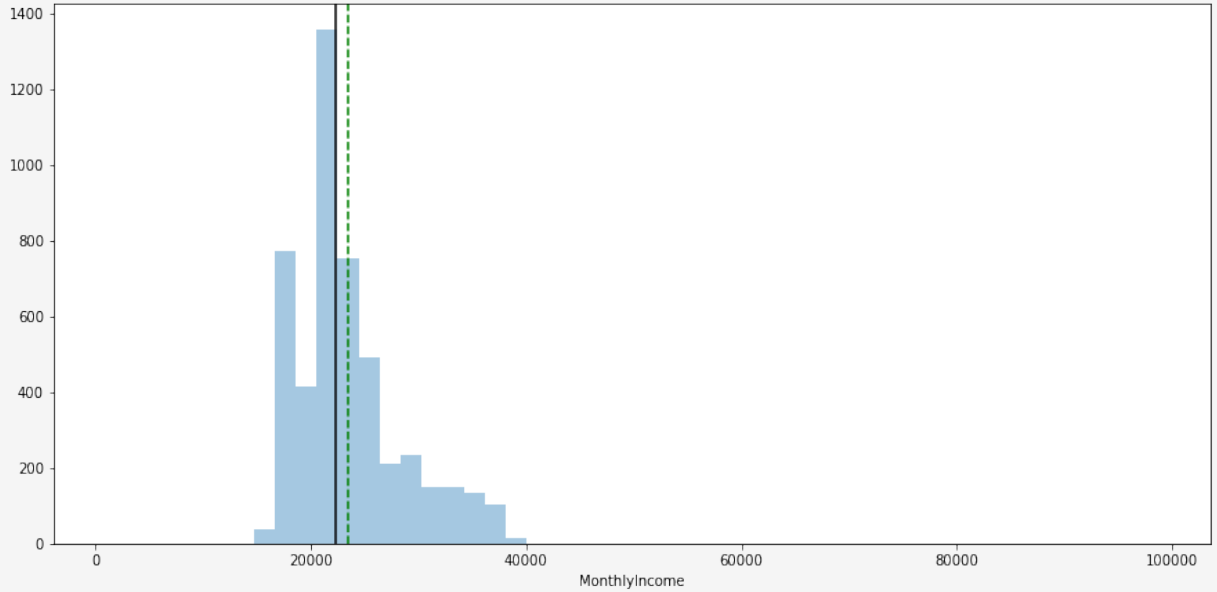


- We can infer that the average Pitch score accross the distribution is 3. The other two peaks are at 1 and 5
- This clearly portends a major influence on the target variable.
- It is also evident why a lot of resources have been expended on insignificant customer brackets hence spiking the cost of doing business

EXPLORATORY DATA ANALYSIS

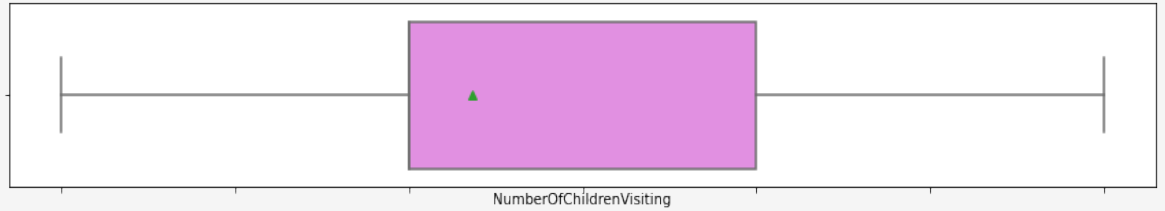
MONTHLY INCOME

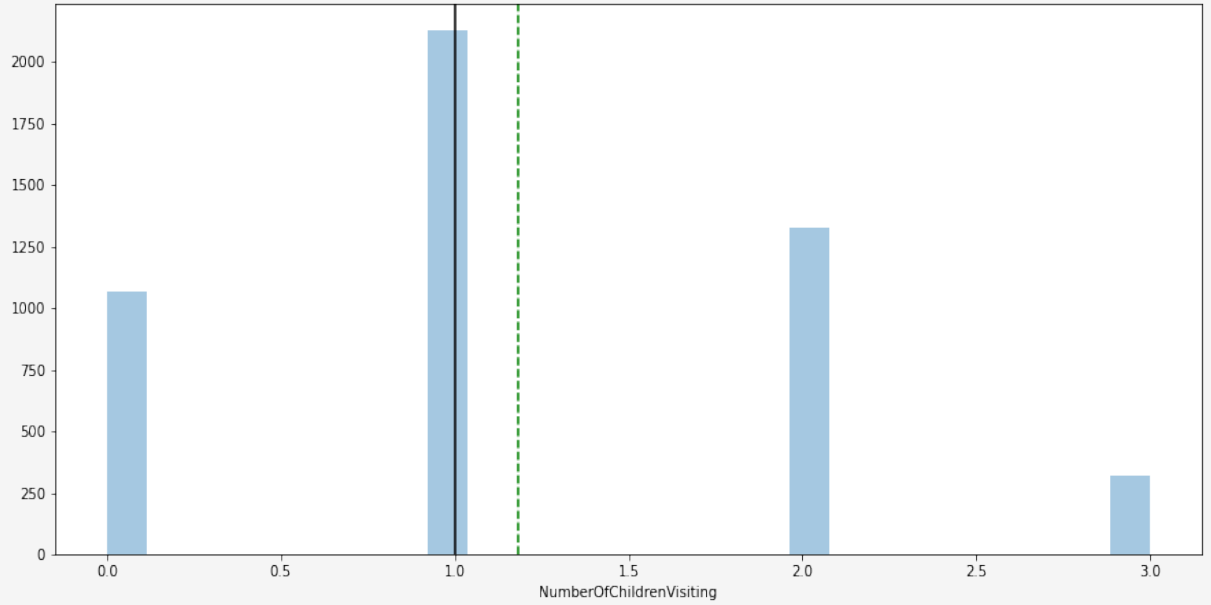




- Monthly Income is a normal ditribution but is Right skewed.
- Tremendous presence of outliers.This will be treated as we proceed.
- Mean Income is 23473 compared to median at 22347
- 75% of the customers earn a monthly income of 25374 or less

NUMBER OF CHILDREN VISITING

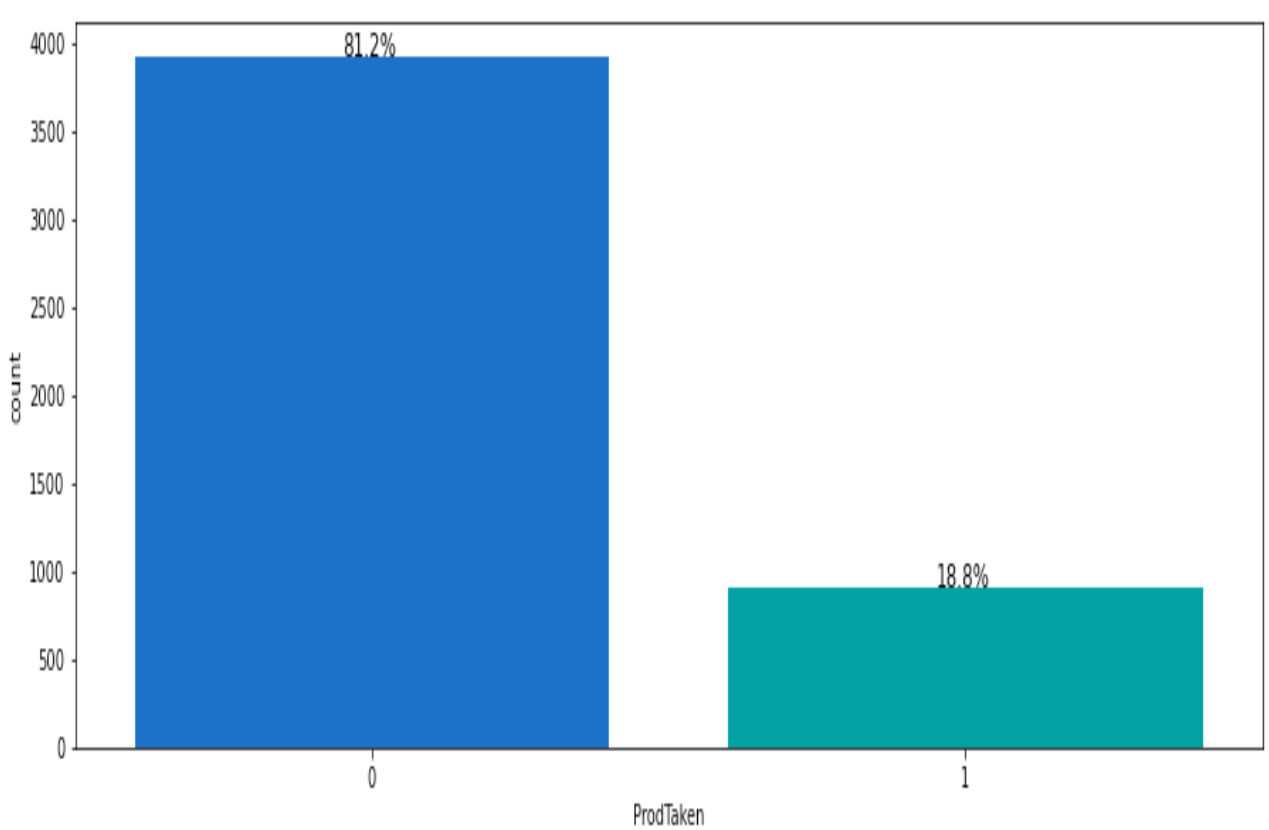




- The distribution peaks at 0, 1 and 2 respectively.
- On the average, 1 child is allowed to go on a visit with a customer.
- 75% of the distribution have 2 children or less who plan to go on a trip with customers

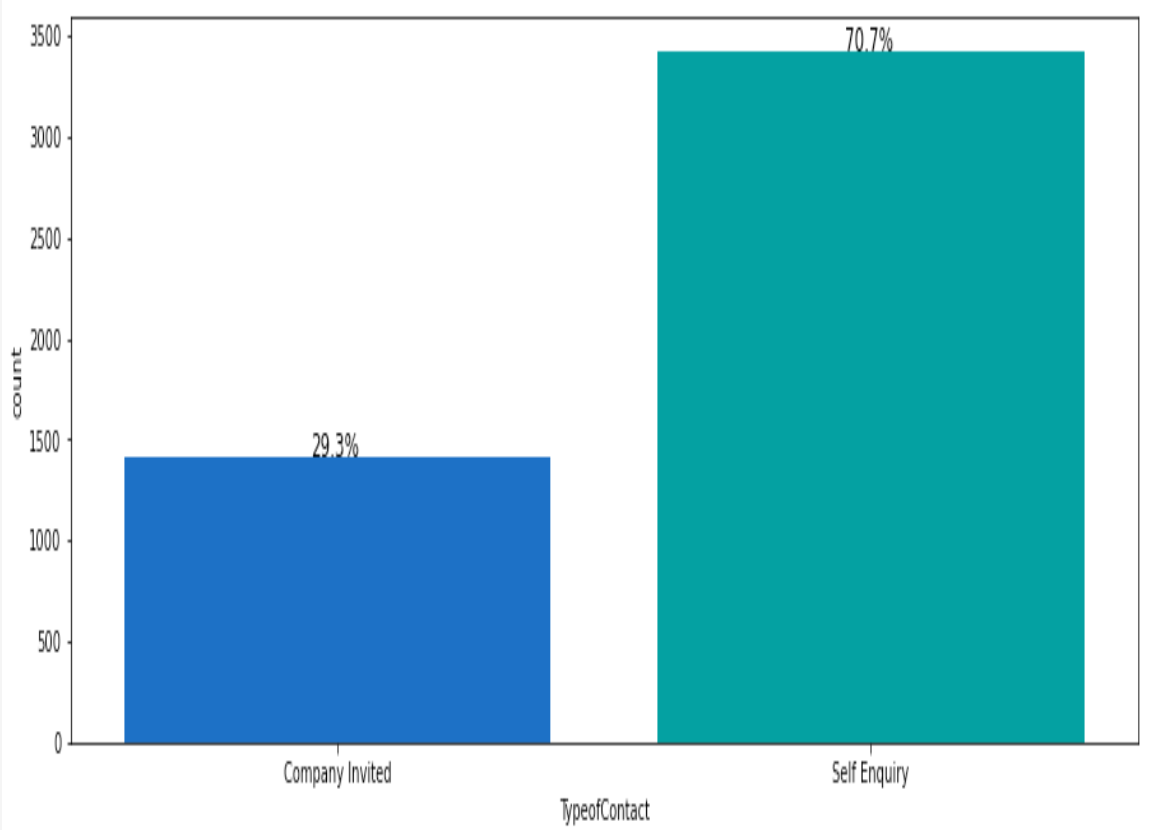
EXPLORATORY DATA ANALYSIS

PRODUCT TAKEN



- 81.2% of the customers did not subscribe to any packages while just 18.8% did as earlier captured in the problem statement

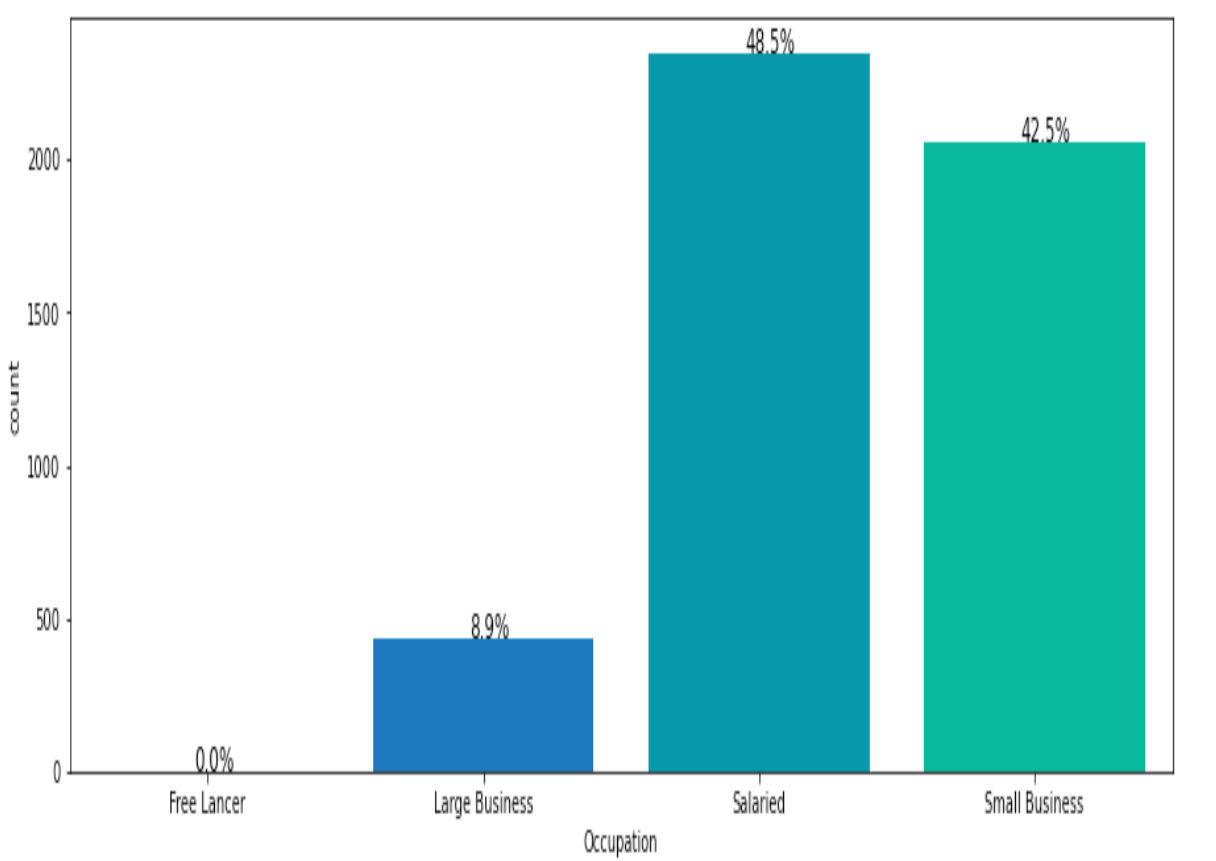
TYPE OF CONTACT



- 70.7% of customers in data made Self Enquiries about the Products on offer while 29.3% were invited by the Company

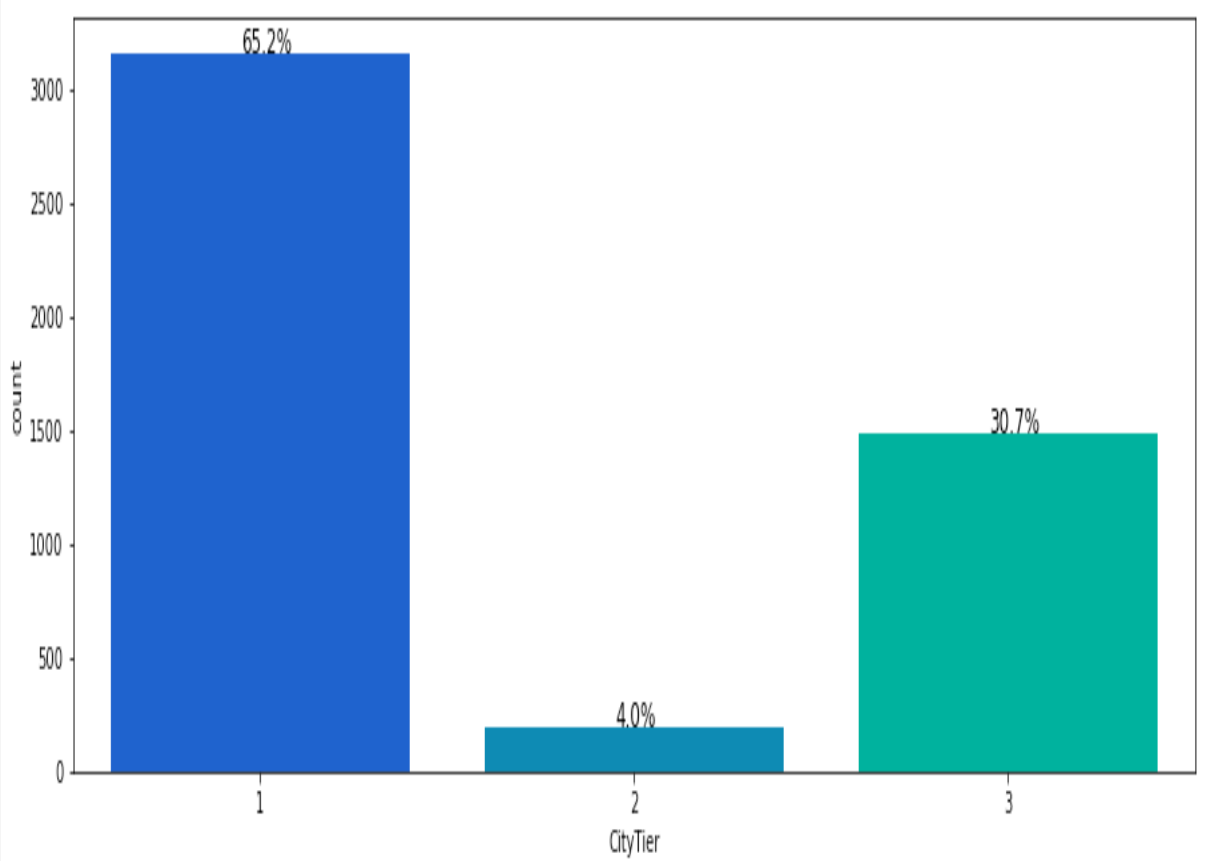
EXPLORATORY DATA ANALYSIS

OCCUPATION



- As can be inferred from the plot, 48.5% of the customers in the distribution are Salaried workers
- This is followed closely by Small Business owners with 42.5% and a distant 8.9% by Large Businesses.

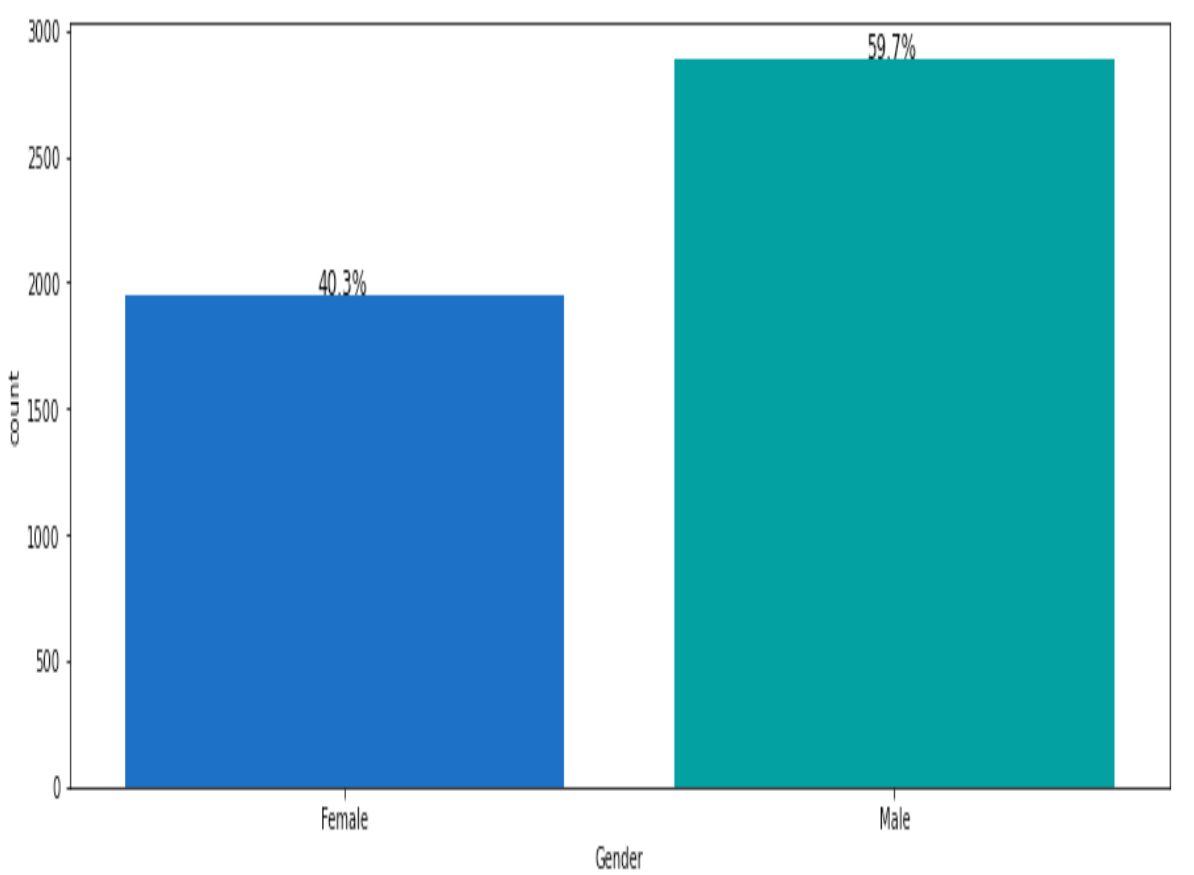
CITY TIER



- 65.2% of the customers are from Tier1 cities with a high standard of living and more population and commercial activities as against the Tier2 City with a very distant 4.0% of the customers and finally Tier3 with the lowest standards of living and dense population coming in at 30.7%.

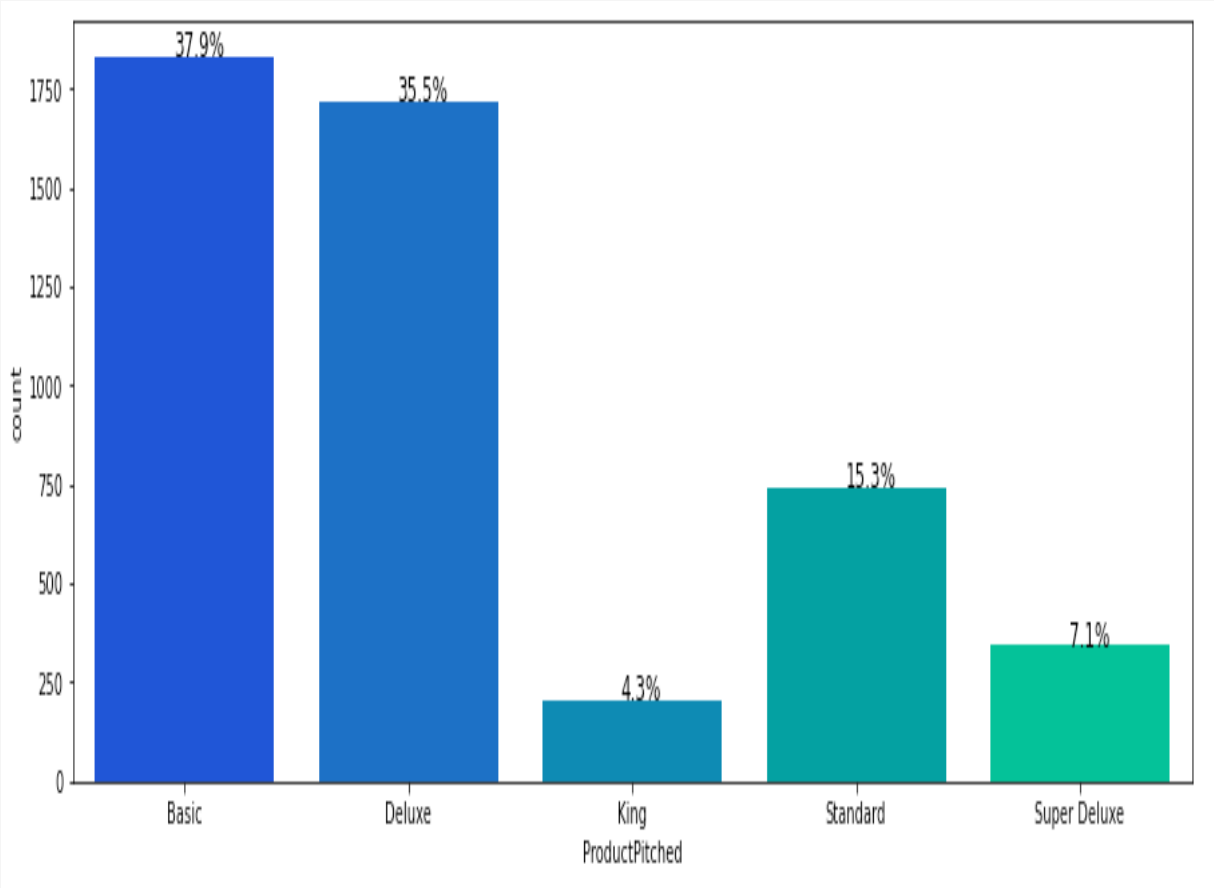
EXPLORATORY DATA ANALYSIS

GENDER



- 59.7% of the customers are Male while
- 40.3% are Females. This group equally holds potentials as well.

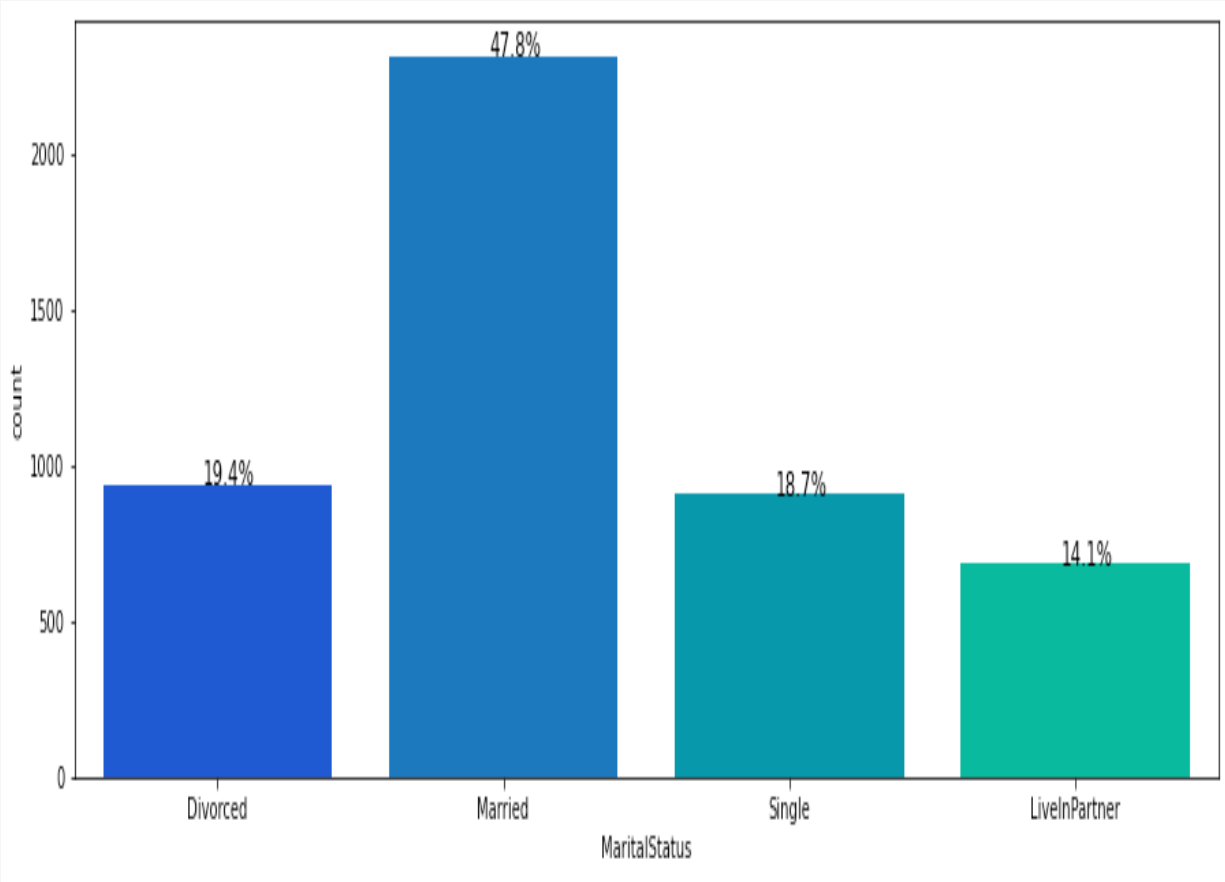
PRODUCT PITCHED



- The Basic package was the most subscribed at 37.9% (1831 customers) followed closely by Deluxe at 35.5%
- Standard, Super Deluxe and King respectively attracted low patronage at 15.3%, 7.1% and 4.3% accordingly
- This is a clear indication of a lapse in target and constructive marketing.

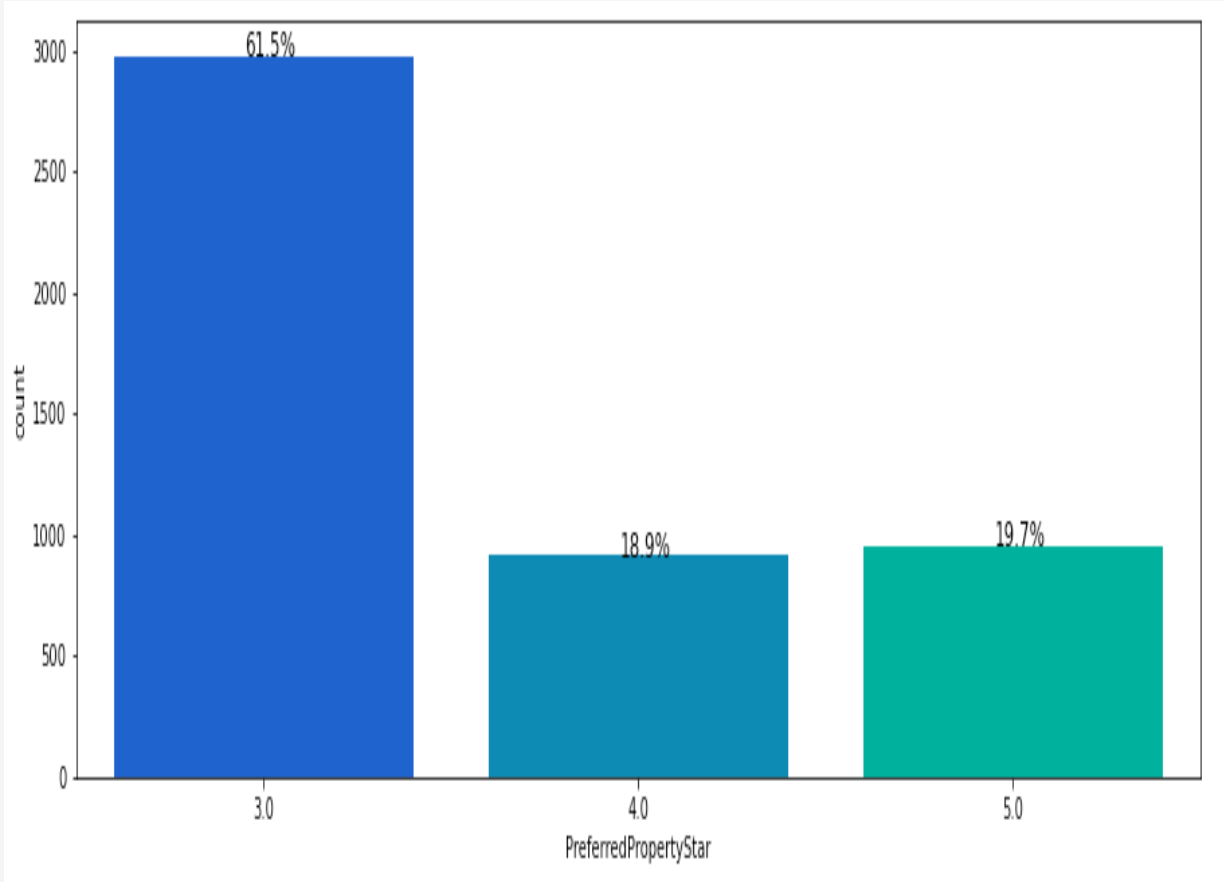
EXPLORATORY DATA ANALYSIS

MARITAL STATUS



Customers who are married top the chart of the Visit with Us database with 47.8% (2311)
Divorced customers come second with 19.4% and Singles following closely at 18.7%
LiveInPartners come last with 14.1% of the customers having this status

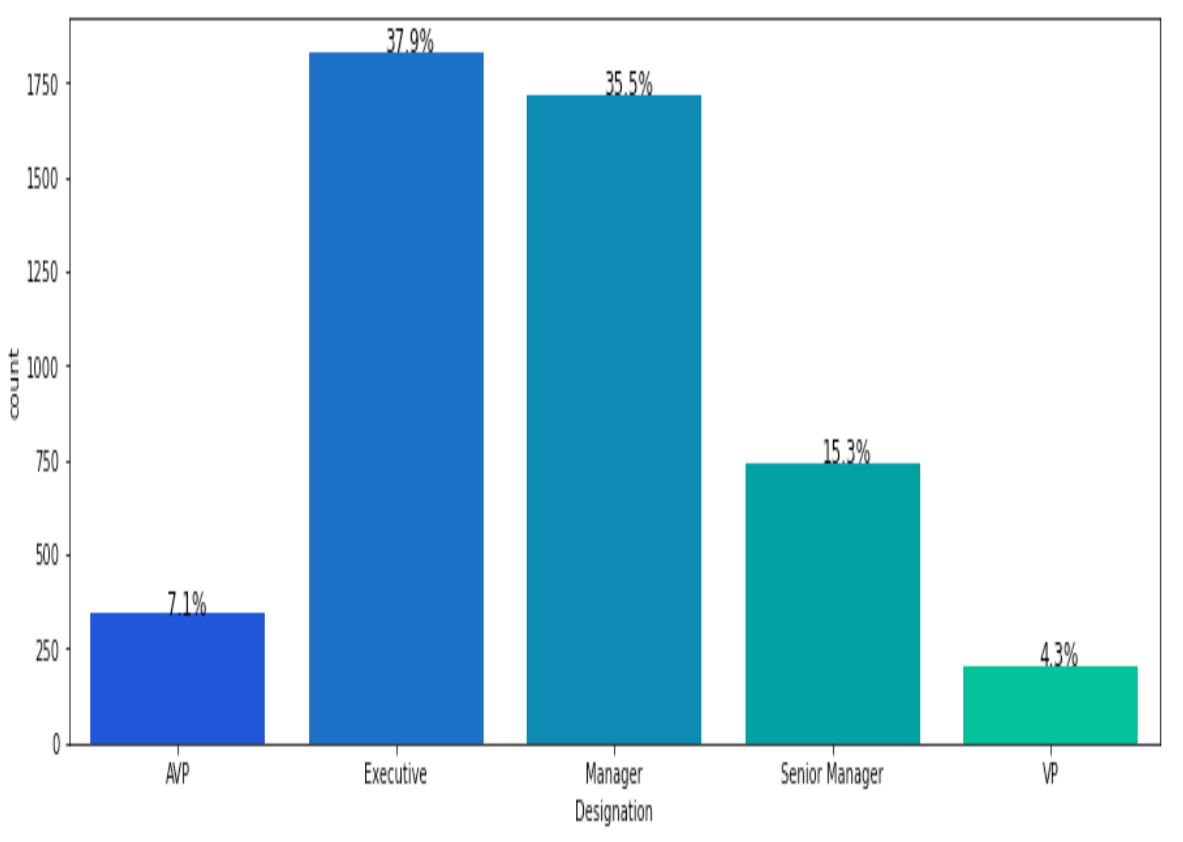
PREFERRED PROPERTY STAR



- 61.5% of the customers prefer 3 star property types when they make a trip
- 19.7% of the customers of Visit with Us are attracted most to 5 star properties
- 18.9% incline more to 4 star properties

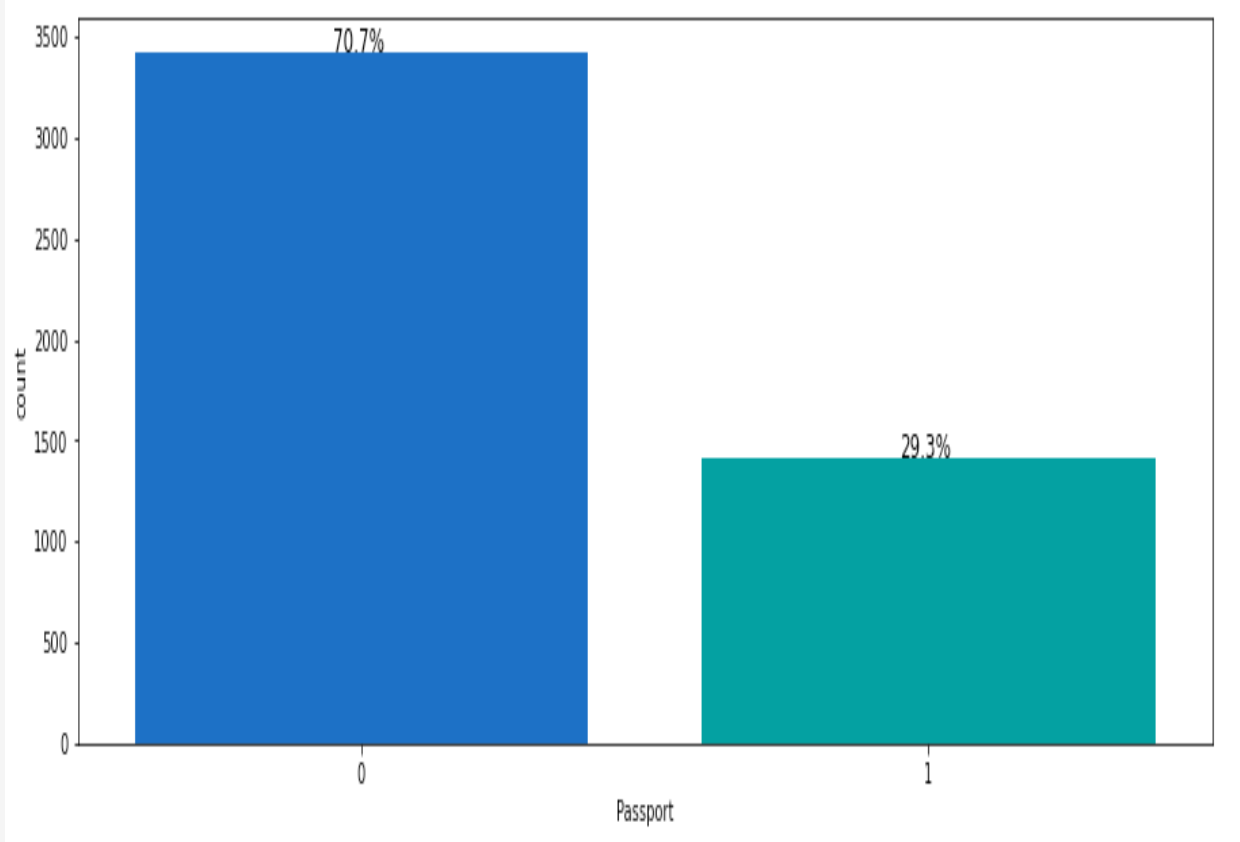
EXPLORATORY DATA ANALYSIS

DESIGNATION



- Executives lead the pack with 37.9% of the customer base followed closely by Managers at 35.5%
- Senior Managers come in at a distant 15.3%. Nearly half the size of the Managers
- Executive Management (AVP and VP) close in at 7.1% and 4.3% respectively

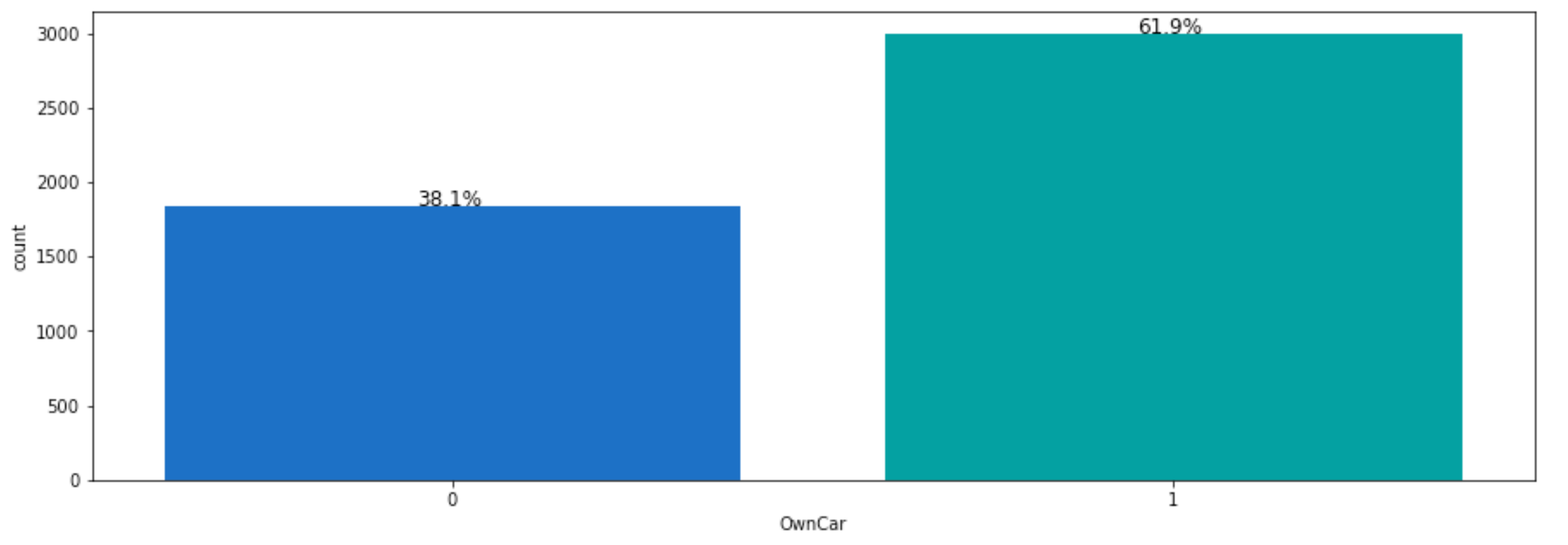
PASSPORT



- 70.7% of the customers own a Passport while 29.3% don't

EXPLORATORY DATA ANALYSIS

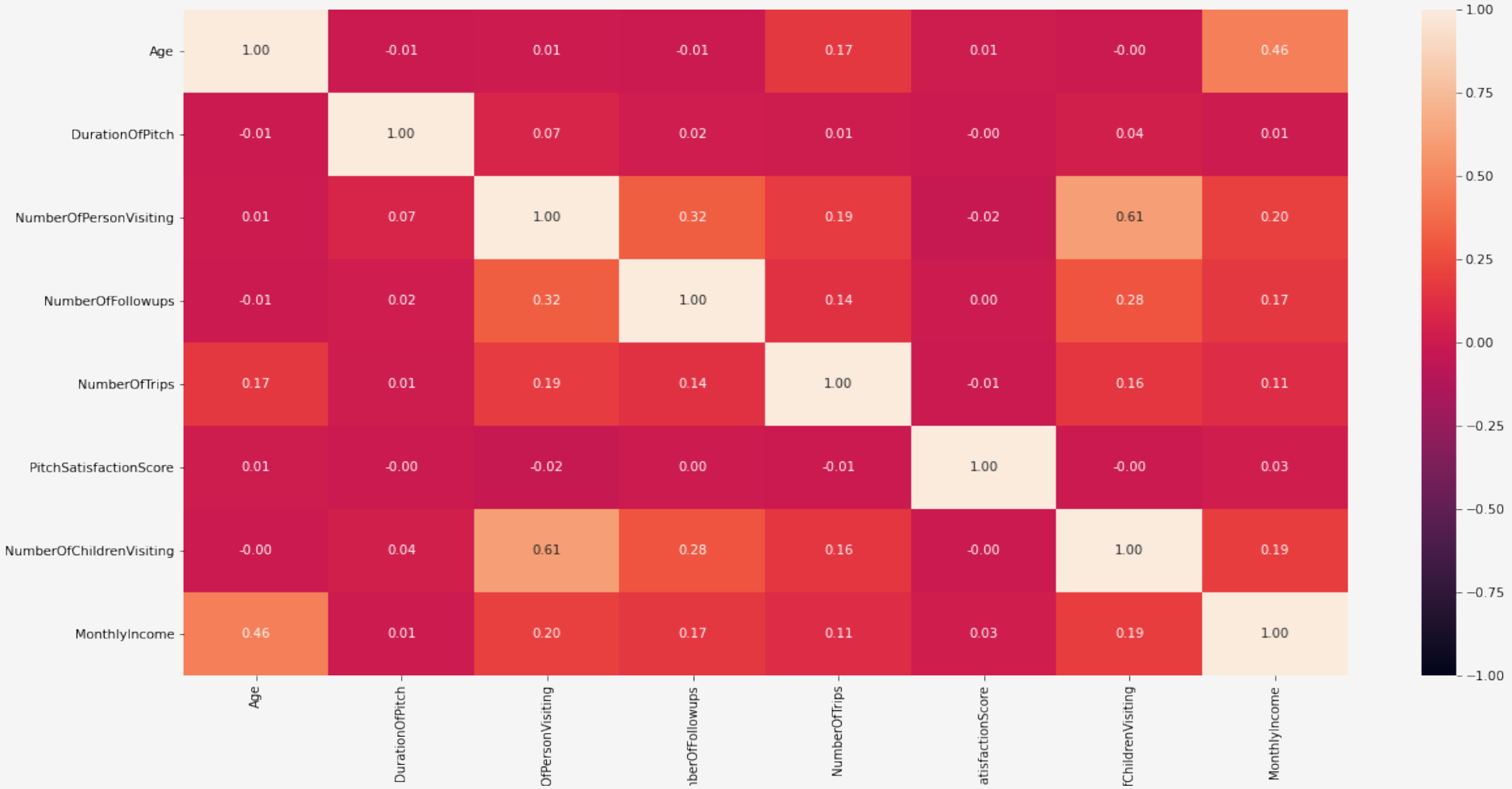
OWN CAR



- 61.9% of the customers at Visit with Us own a car
- 38.1% of the customers do not own one
- It is indeed quite clear that more resources should be allocated pitching to customers with cars
- Also, resources could be employed to monetize convenience for non-car owners.

EXPLORATORY DATA ANALYSIS

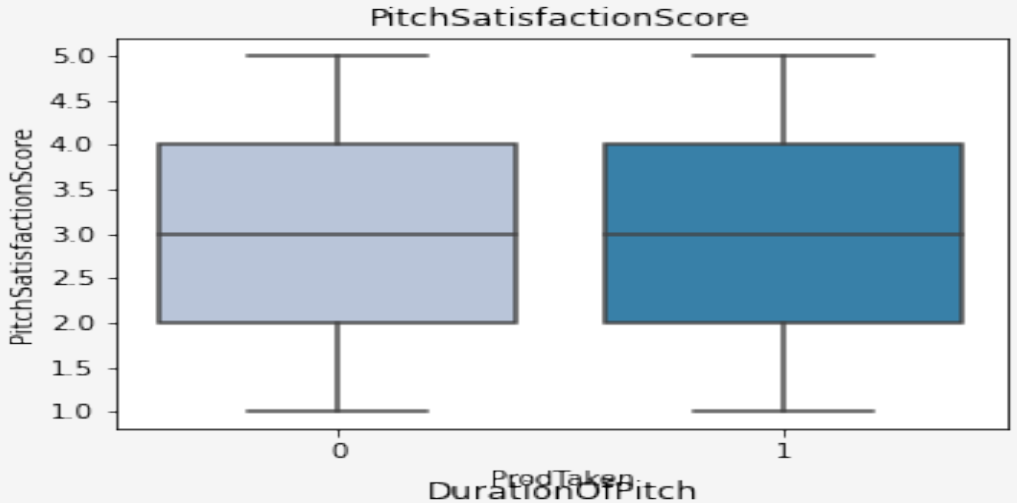
HEATMAP

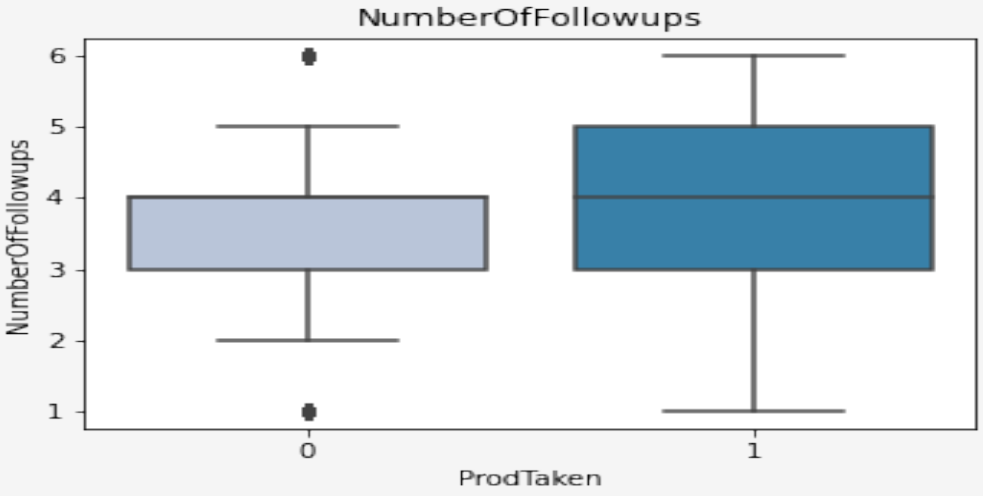


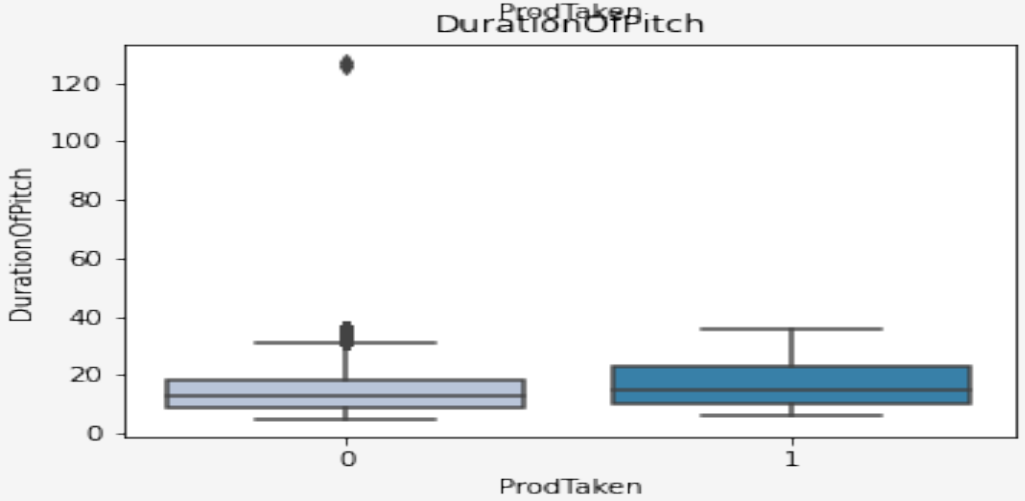
There is little or no correlation between most of the variables
We see a fairly reasonable correlation between Age and Income

EXPLORATORY DATA ANALYSIS

PRODUCT TAKEN VS CUSTOMER INTERACTION DATA)



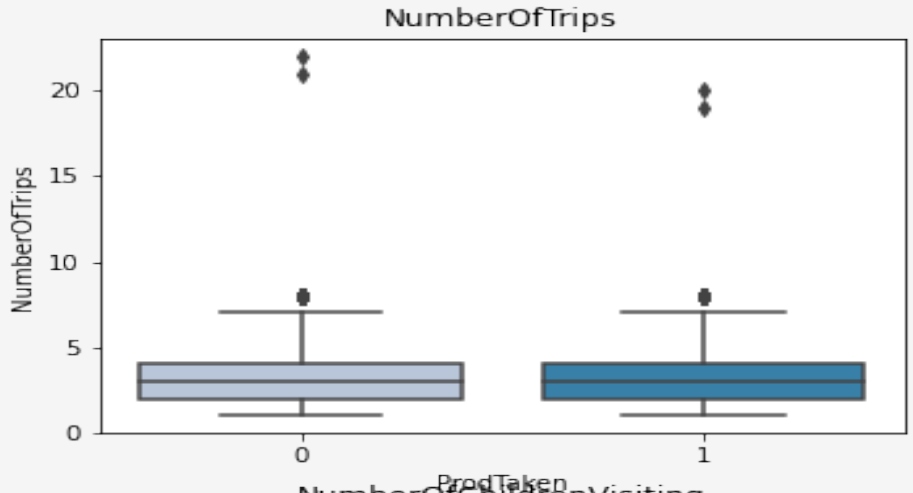


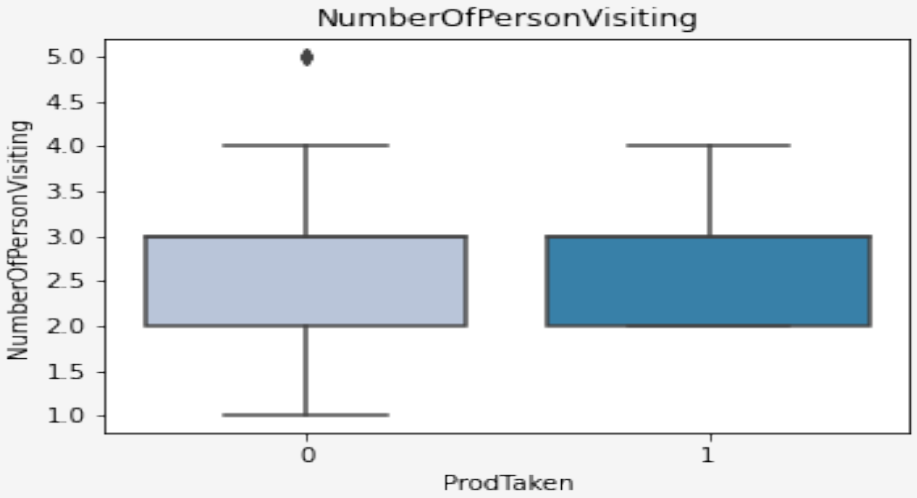


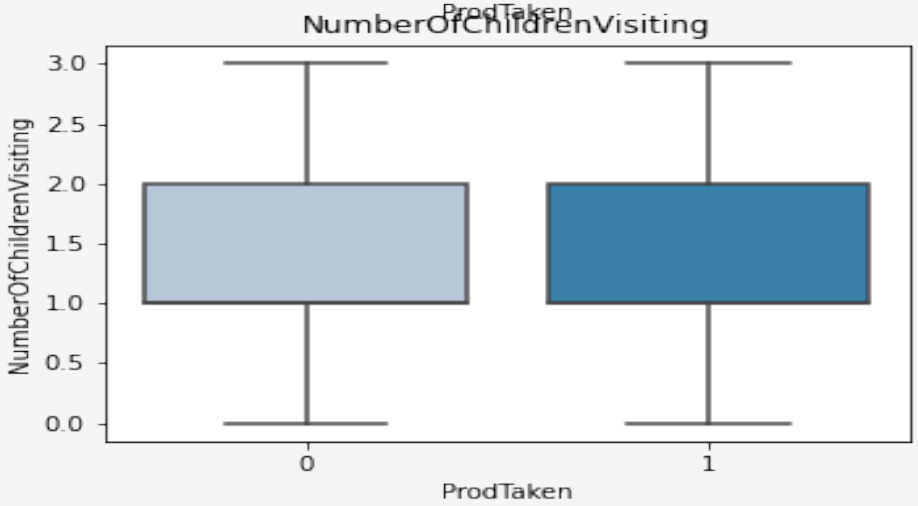
- The Pitch Satisfaction score clearly has no impact on the Subscription of Packages as both Customers who did or did not secure any package had relatively same scores
- Customers who subscribed to packages had a higher number of follow ups compared to others who didn't
- The Duration of Pitch clearly has an impact on subscription as customers who had higher durations subscribed

EXPLORATORY DATA ANALYSIS

PRODUCT TAKEN VS TRIP DATA



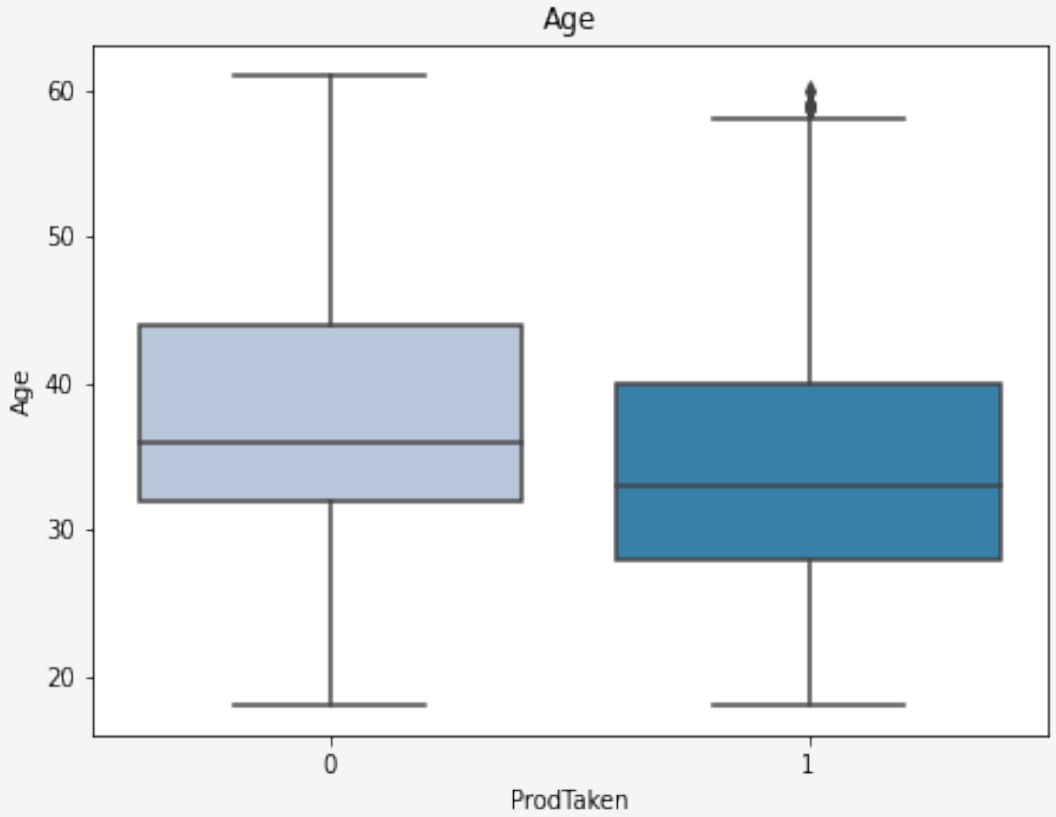


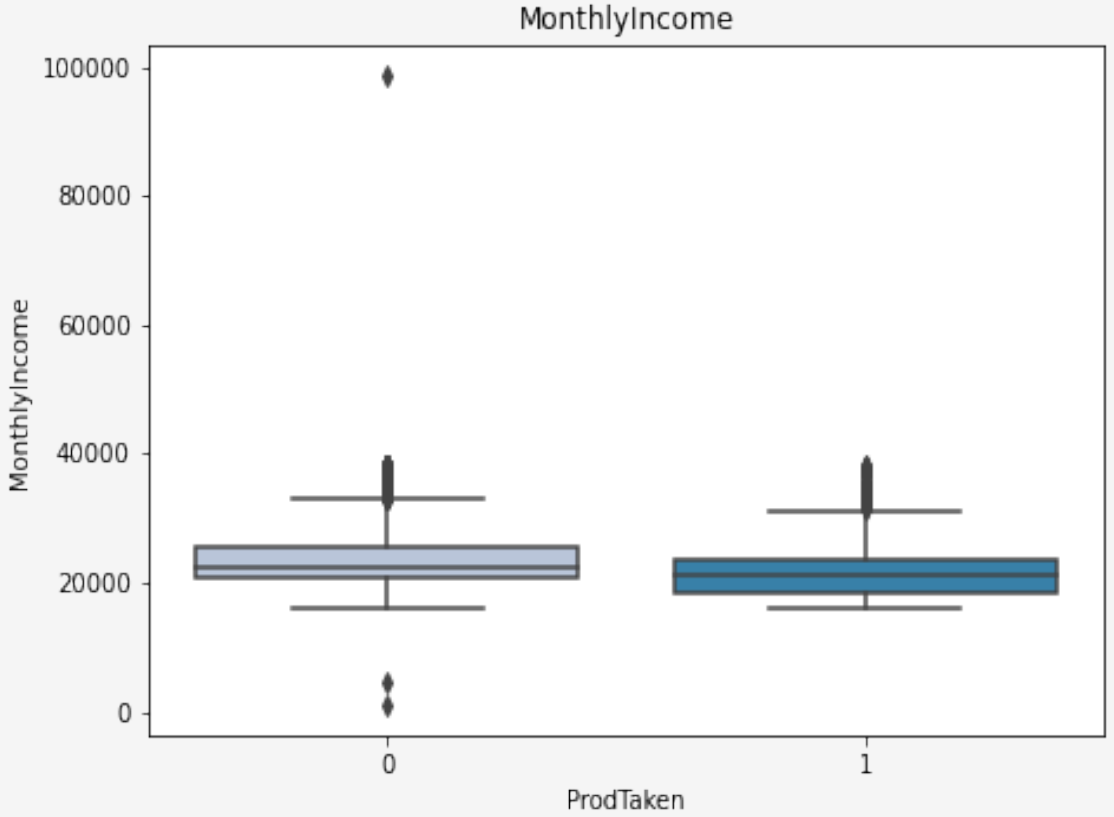


- The Number of Trips also appears to be an irrelevant feature as both customers who did or did not buy to package did not show any marked difference
- There is a clear correlation between Number of Children visiting and the Number of Persons Visiting .We would drop the former prior to modelling

EXPLORATORY DATA ANALYSIS

PRODUCT TAKEN VS AGE VS MONTHLY INCOME



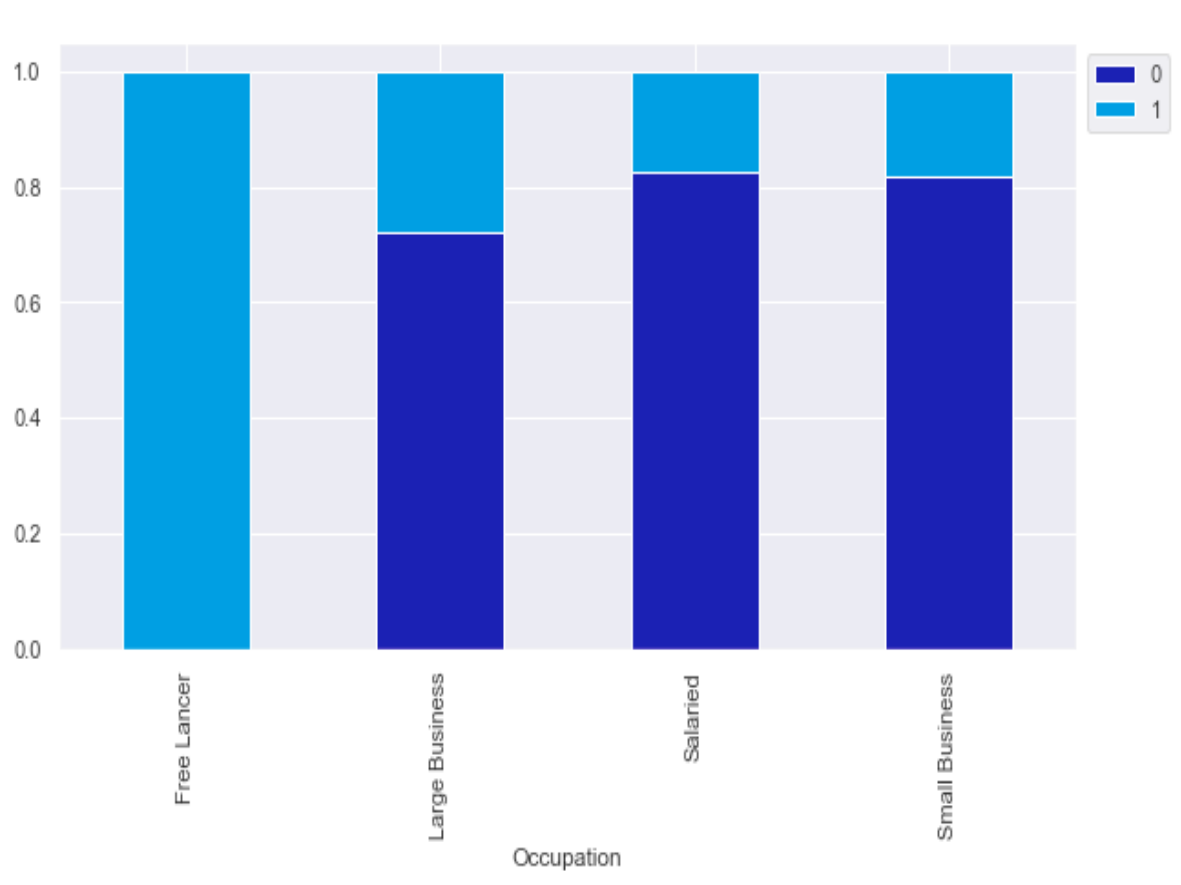


- 75% of those whose subscribed are below the age of 40 years as against 45 years for those who didn't
- Customers who did not take nor subscribe to any package are quite older than customers who did with the oldest at 60years

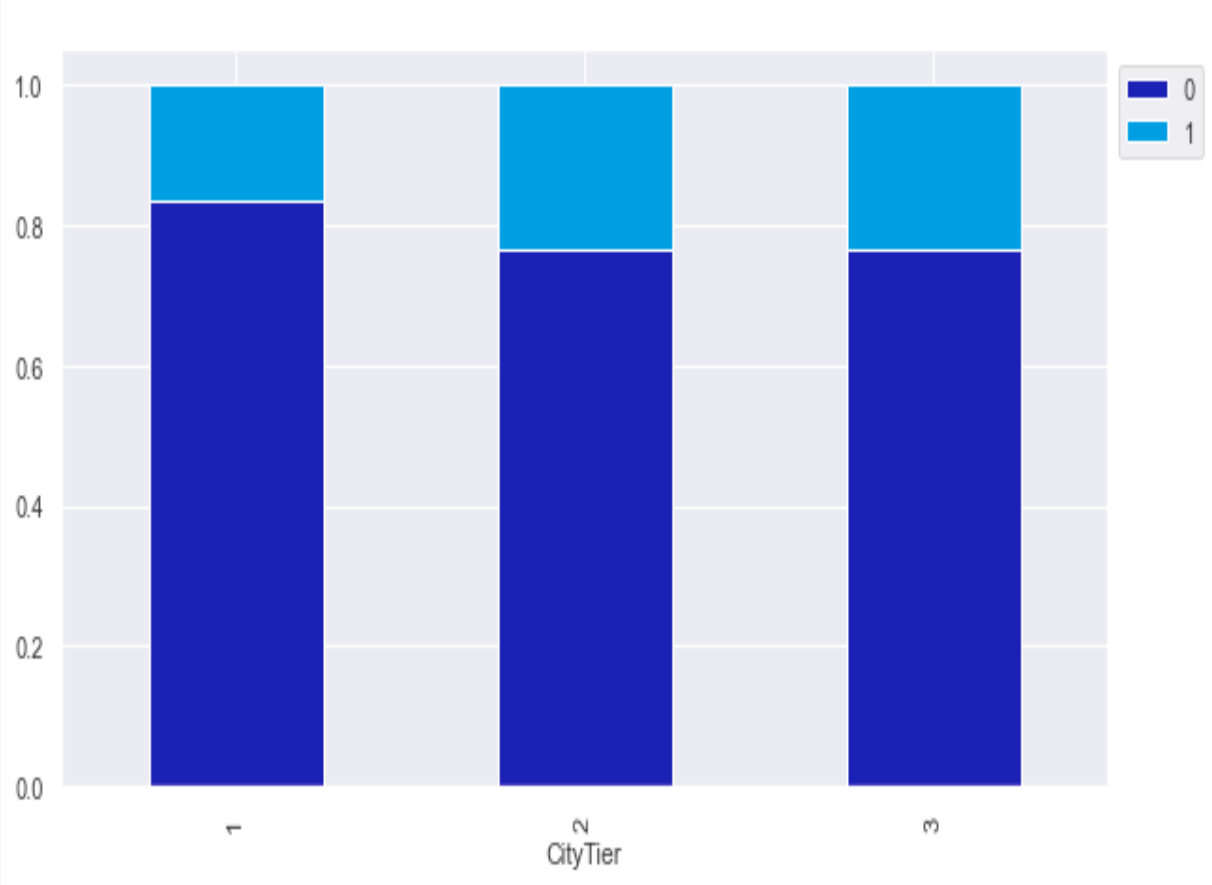
- Customers who did not subscribe to any package earn higher income than those who did.
- This relationship also speaks to an obvious impact on the Product taken (Packages)

EXPLORATORY DATA ANALYSIS

STACKED OCCUPATION



STACKED CITY TIER

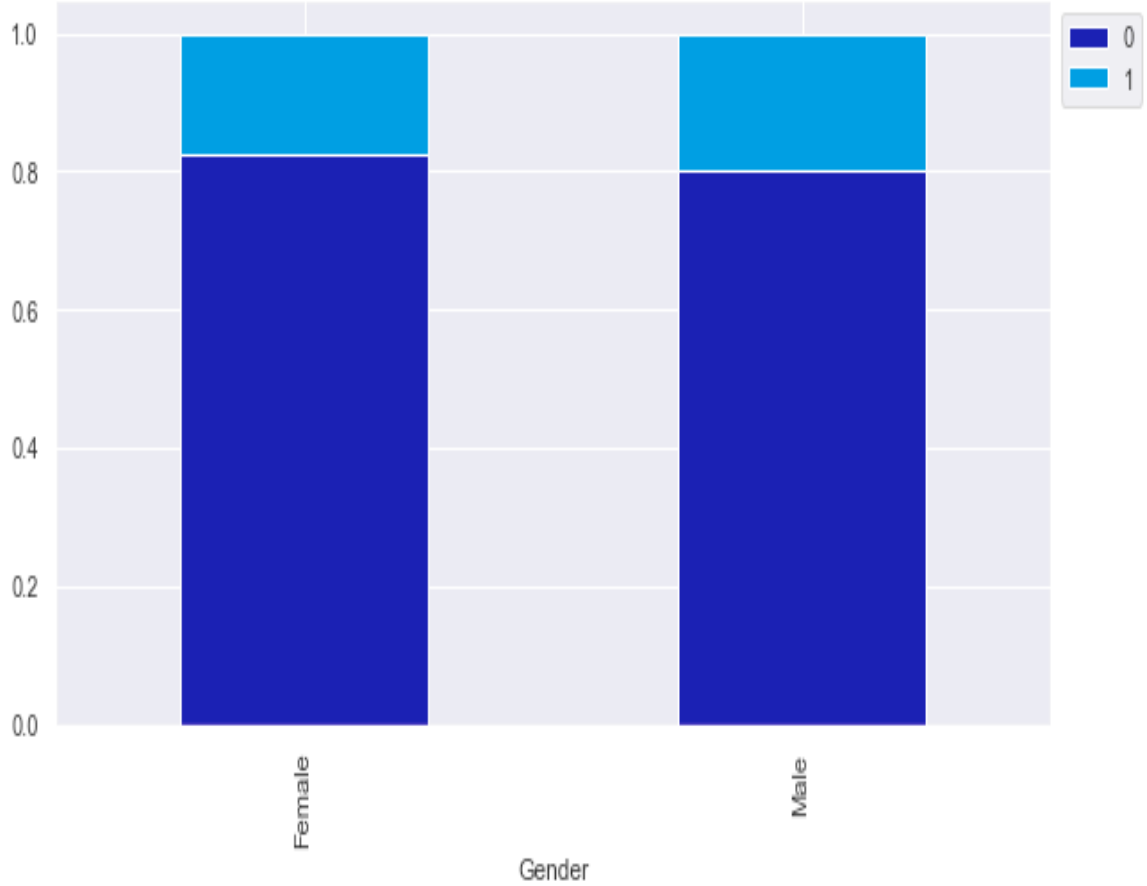


- Free Lancers have a 100% subscription. This is followed by Large Businesses with roughly 28% Small Businesses have the least subscriptions accross the distribution

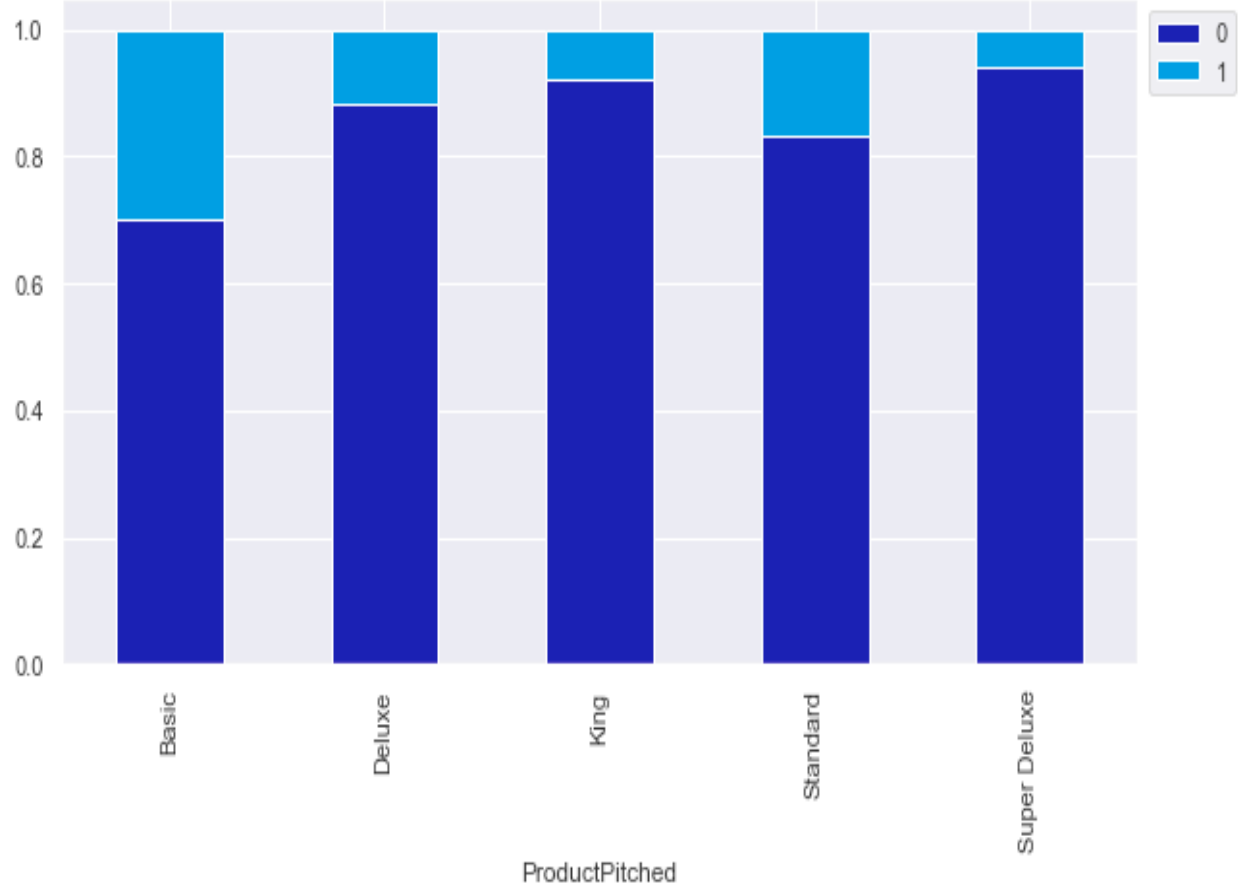
- Even though majority of the customers reside in Tier 1, customers from Tier 3 subscribed to more packages than their counterparts in other locations with about 22%

EXPLORATORY DATA ANALYSIS

STACKED GENDER



STACKED PRODUCT PITCHED

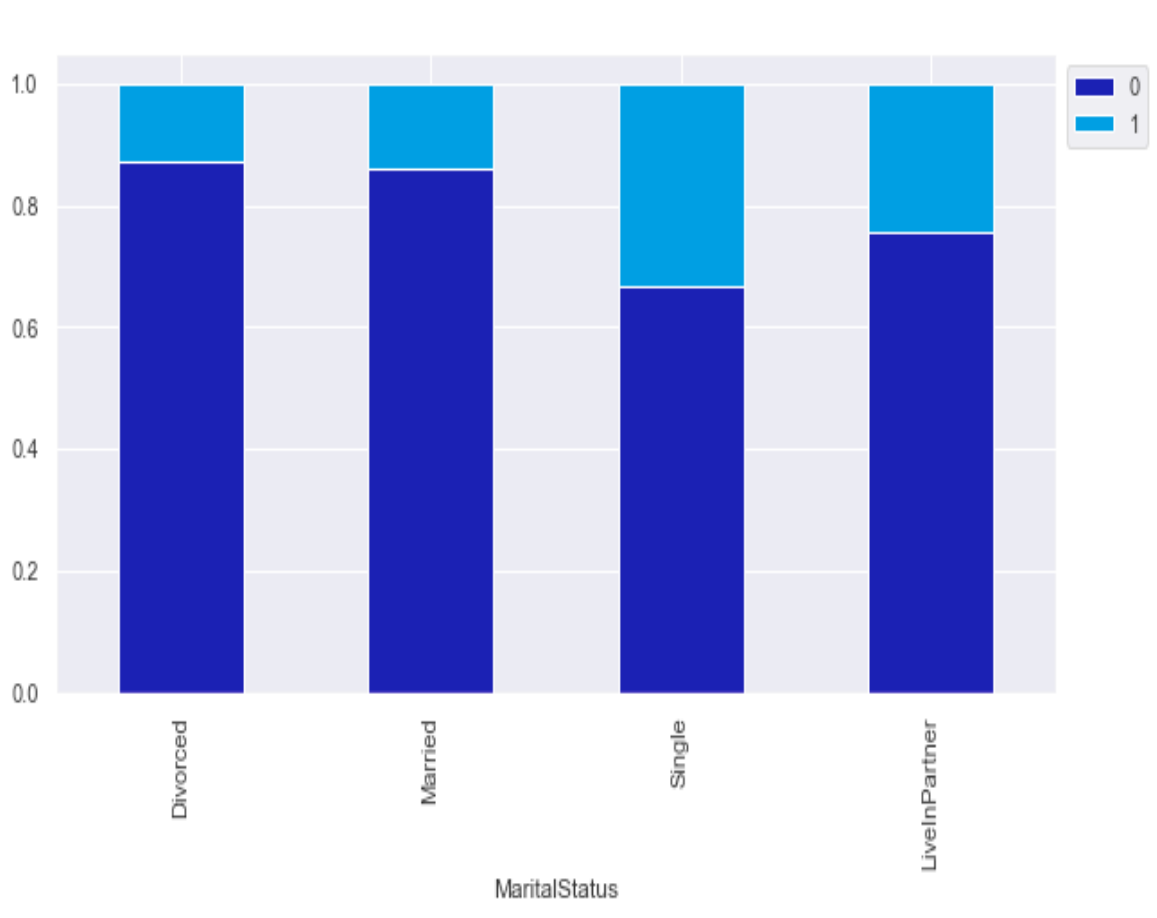


- Males have a higher subscription rate compared to females

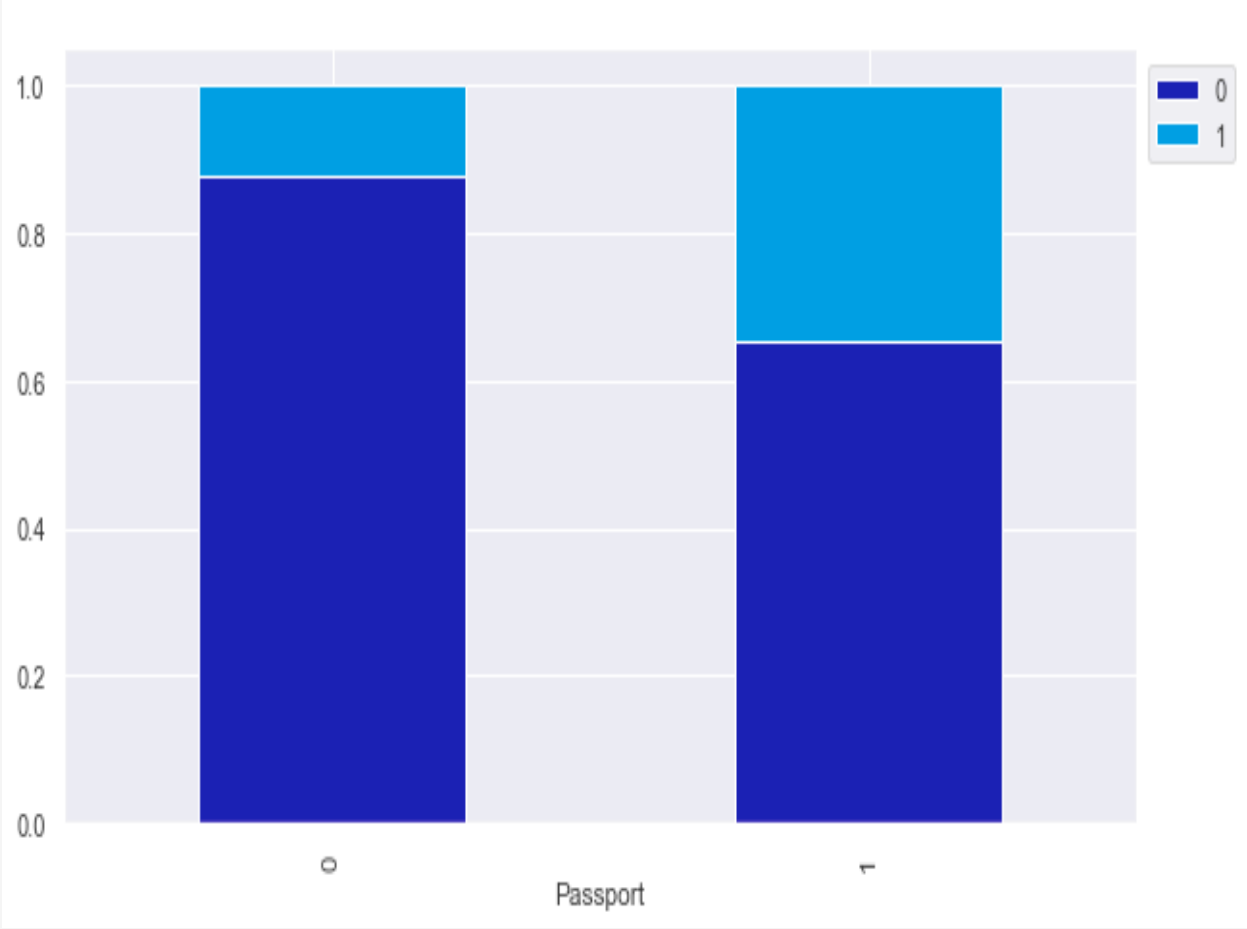
- Basic Package has the highest subscription rate at about 30% followed closely by Standard at roughly 16% Super Deluxe has the least subscription closely followed by King

EXPLORATORY DATA ANALYSIS

STACKED MARITAL STATUS



STACKED PASSPORT

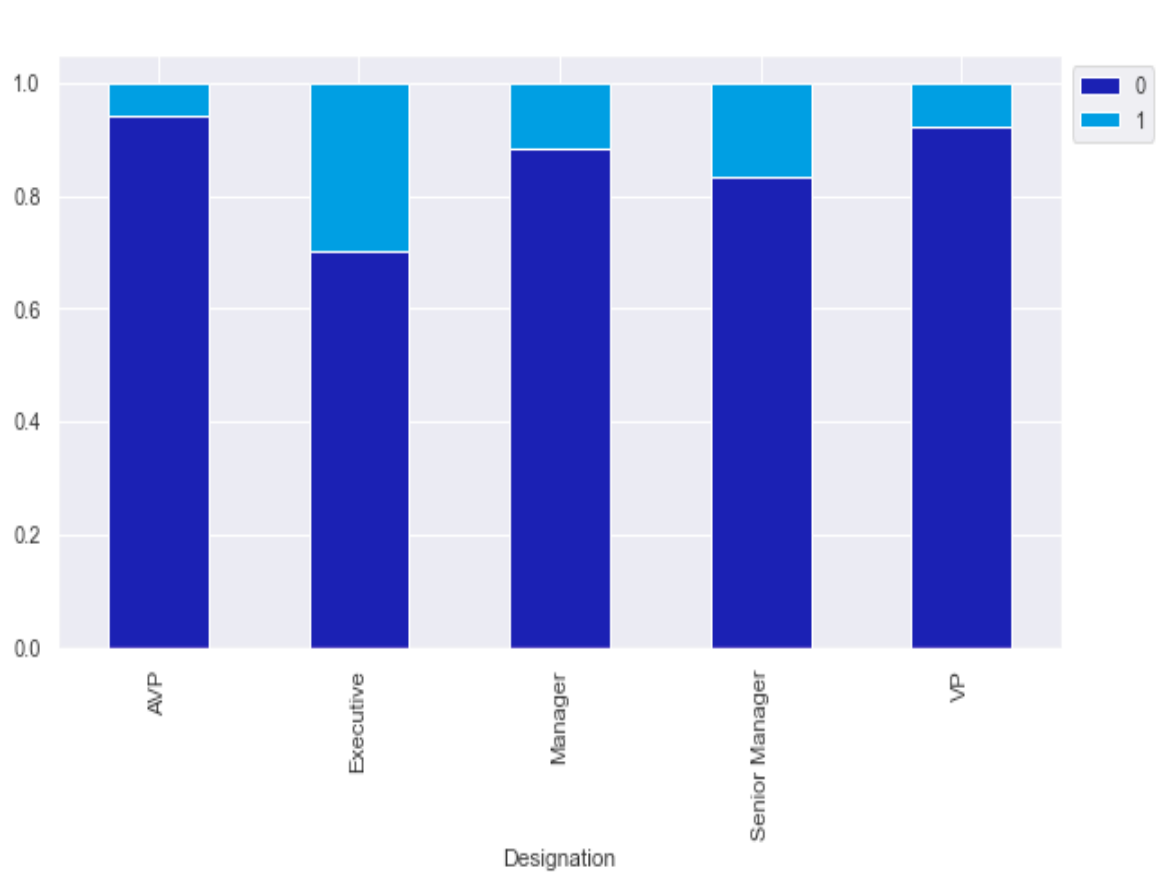


- As seen earlier, just as Visit with Us focuses more on the younger age bracket , it also tends to favor a lot more Singles with a 31% subscription. This is closely followed by LiveIn Partners at about 24%
- It is more than evident that Age and Marital status clearly impact on the potential to be offered a subscription

- Those with Passports have an estimated 35% subscription as against those without with approx. 12%
- This makes sense as major determinant for eligibilty to subscribe to a package

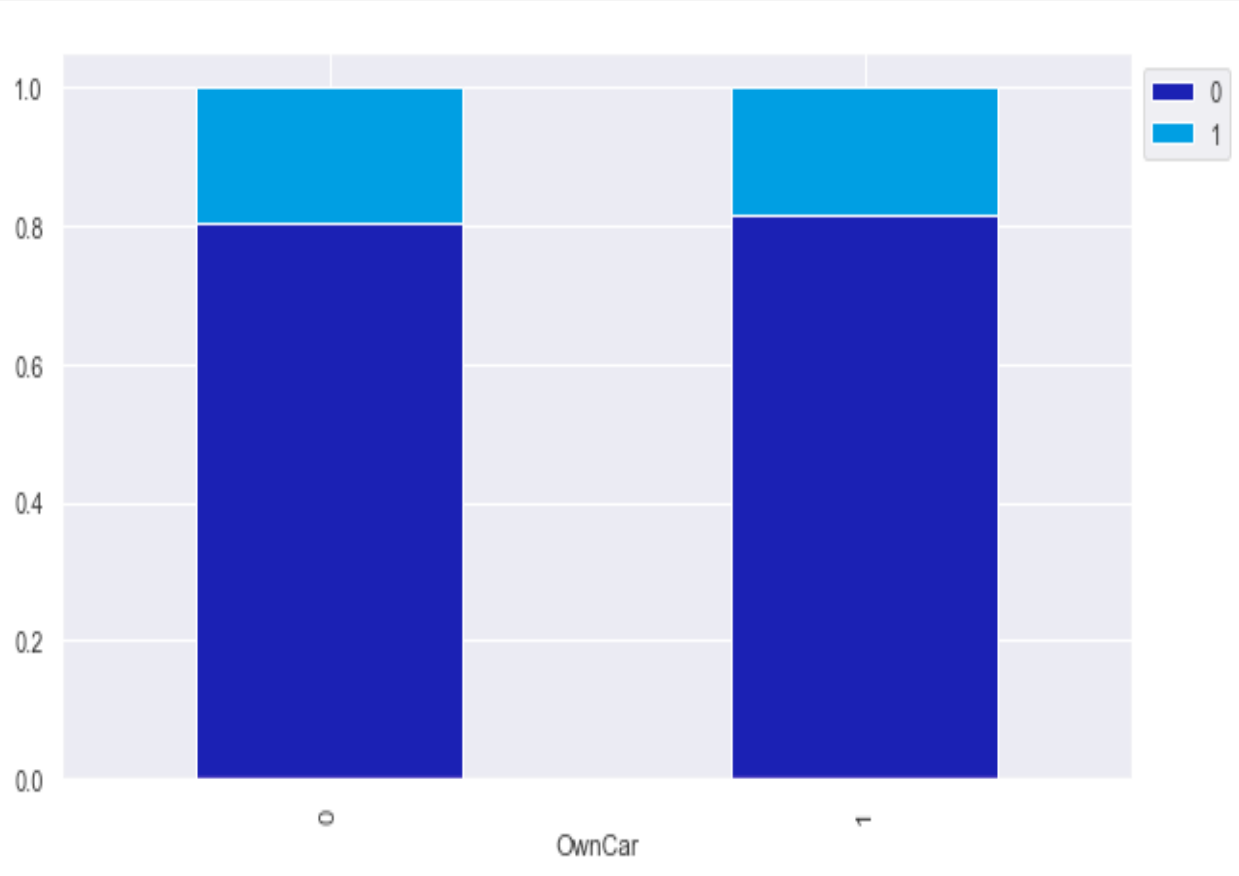
EXPLORATORY DATA ANALYSIS

STACKED DESIGNATION



- Executives subscribed more to the new travel package followed by Senior Managers at about 18%
- AVP tends to be the least. Its quite logical as his schedule is way tighter at the top than others. He would rather delegate

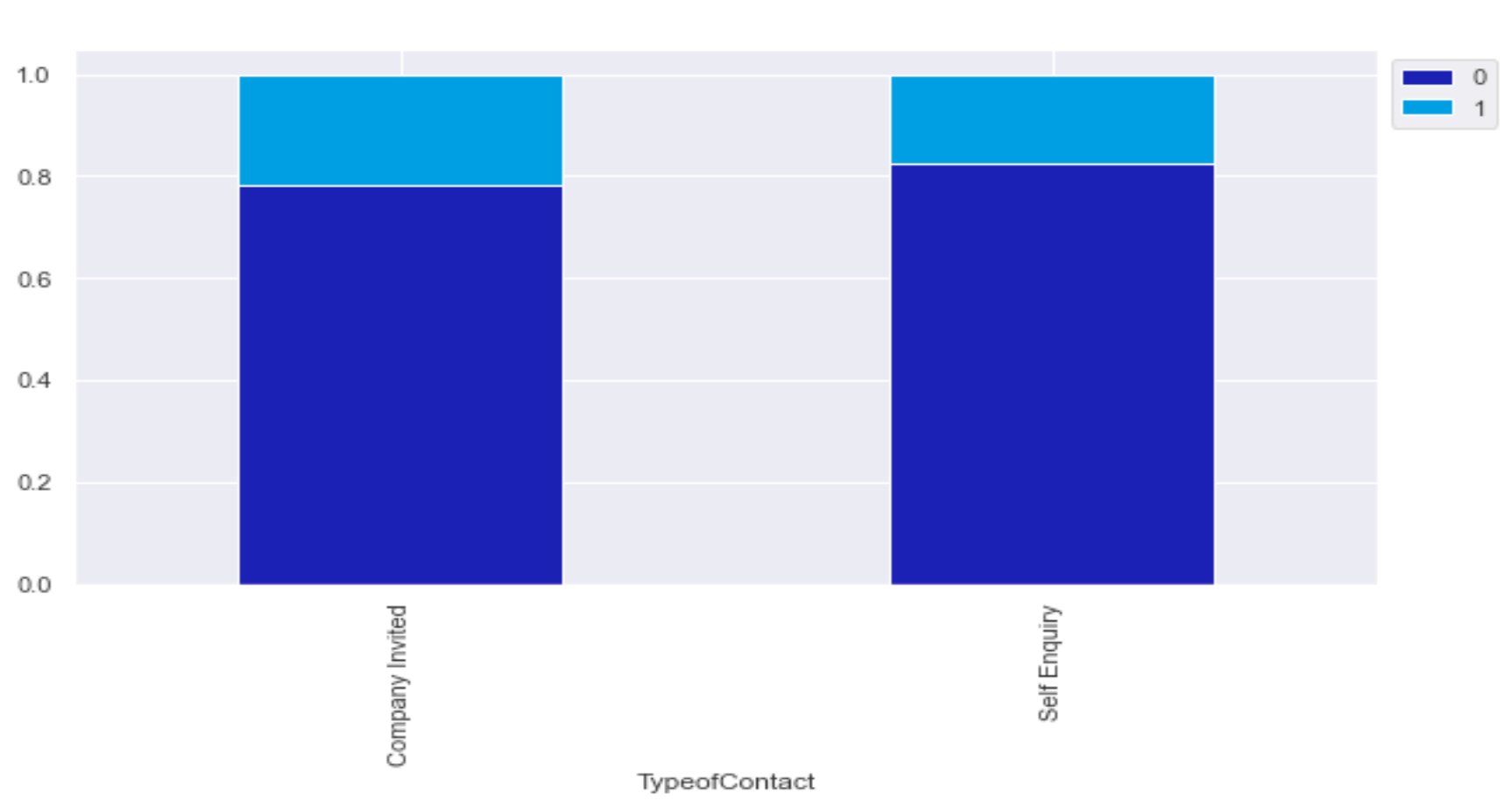
STACKED OWN CAR



- There is relatively a near insignificant impact on Products taken and Owncar as it concerns those who did or did not subscribe to any of the packages
- This is not a good predictor of who or who won't buy a package

EXPLORATORY DATA ANALYSIS

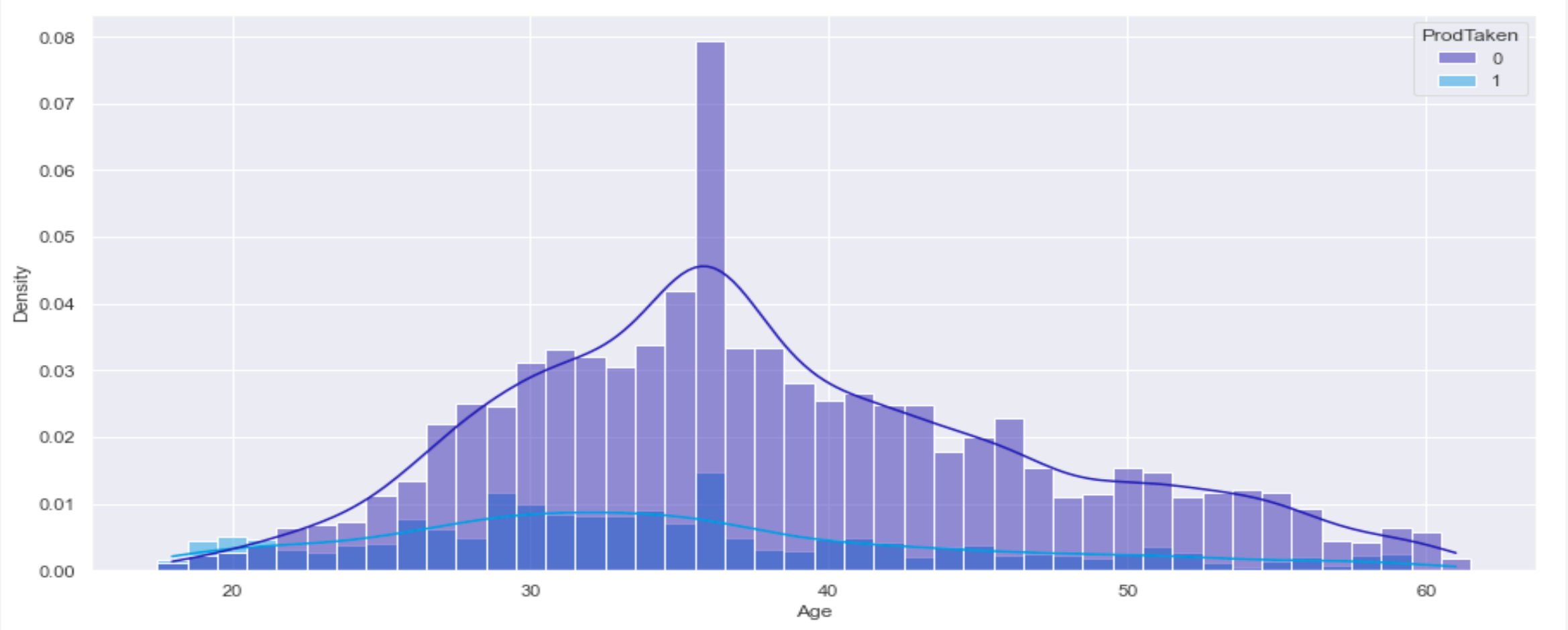
STACKED TYPE OF CONTACT



- Customers invited by Visit with Us had higher subscriptions, roughly 21% compared to their counterparts even though there are more customers who Self Enquired compared to company invitations
- Scarce reources can be minimised with taking advantage of Technology for faster engagements through Mails, Social Media and Instant messages

EXPLORATORY DATA ANALYSIS

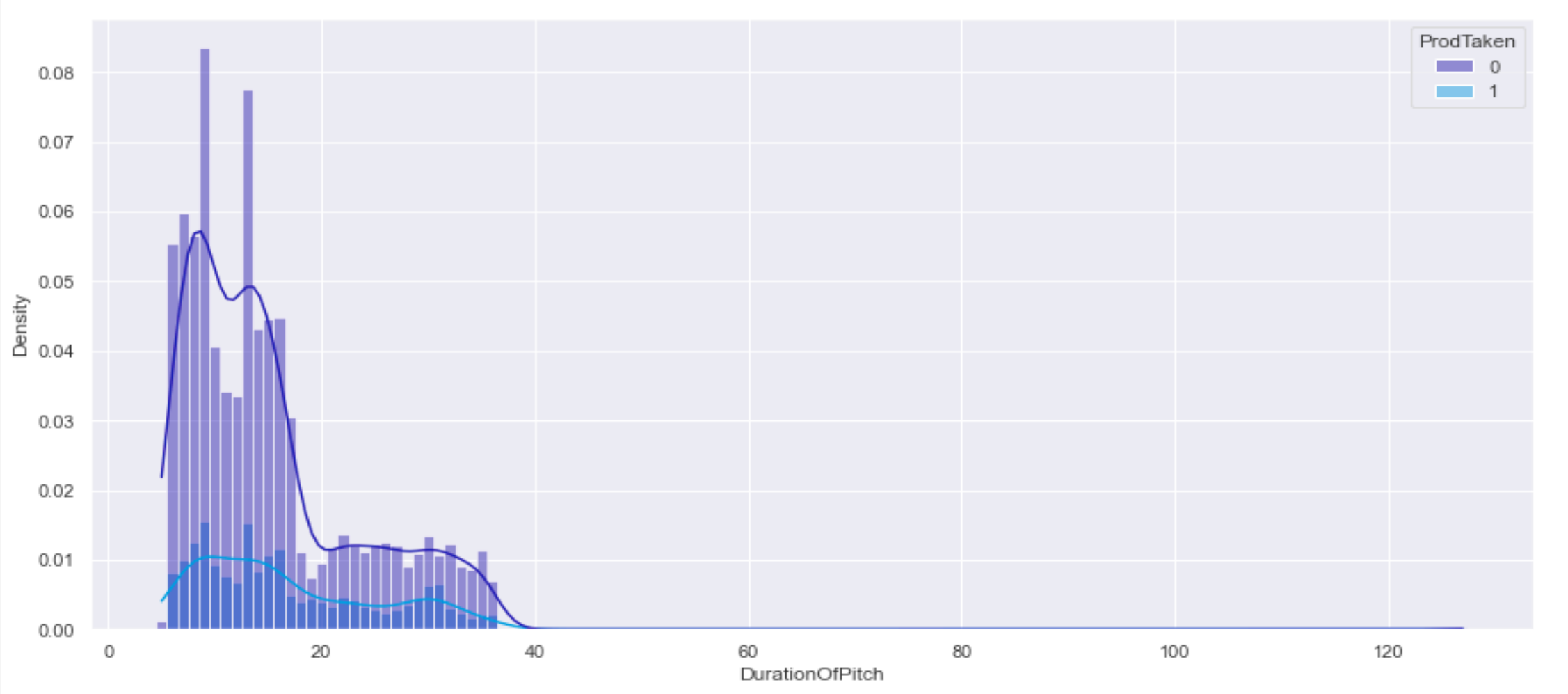
AGE DISTRIBUTION BY PRODUCT TAKEN



- This pattern here is quite revealing as we can see that much older customers make up a majority of non-subscribers to the travel packages
- It also indicates Visit with Us designs products which endears more to the younger generations

EXPLORATORY DATA ANALYSIS

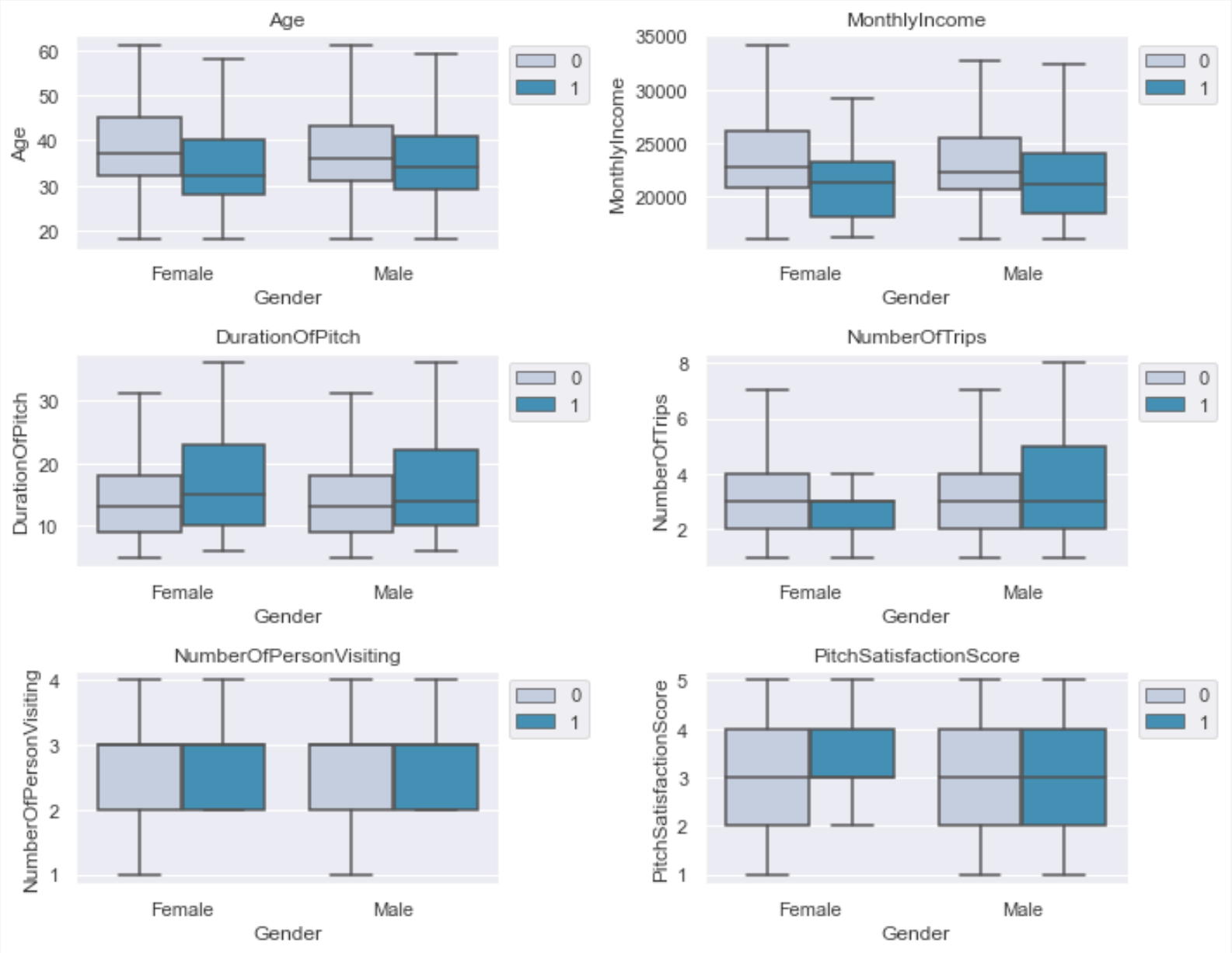
PRODUCT TAKEN BY DURATION OF PITCH



- Customers who subscribed to the new Travel package had shorter pitch durations than those who had much longer pitches. So it is indeed no guarantee that a customer will buy the package if he/she was pitched longer.

EXPLORATORY DATA ANALYSIS

CUSTOMER LEVEL ANALYSIS



- Female non-subscribers are older than their male counterparts. Together they trump the number of subscribers by population.
- The number of subscribers across both genders at almost at par in Age and younger as well.
- Male customers who subscribed dwarf their femal counterparts in monthly income earnings however both male and female non-subscribers earn far much higher income than the subscribers
- Male and Female customers who opted for the new travel package had much longer pitches compared to the non-subscribers
- Male subscribers have had twice more trips than their female counterparts, peaking at 8 as against 4
- Both genders who did not opt for the package have exactly same number of trips
- Both male and female customers accross divides clearly have same number of trips
- Males who did or did not buy the package had similar pitch scores which points to the fact that there are no guarantees even if you have a much higher pitch score
- It would seem that female subscribers had much higher Pitch scores.

MODEL PERFORMANCE SUMMARY

OVERVIEW OF ML MODELS AND PARAMETERS

- Prior to Modeling, Outliers were treated and further Data-Preprocessing was done to identify independent variables viable for the prediction process

- Next step was to create dummy variables using a function that automates On-Hot encoding for a more surgical approach. Education was hot encoded

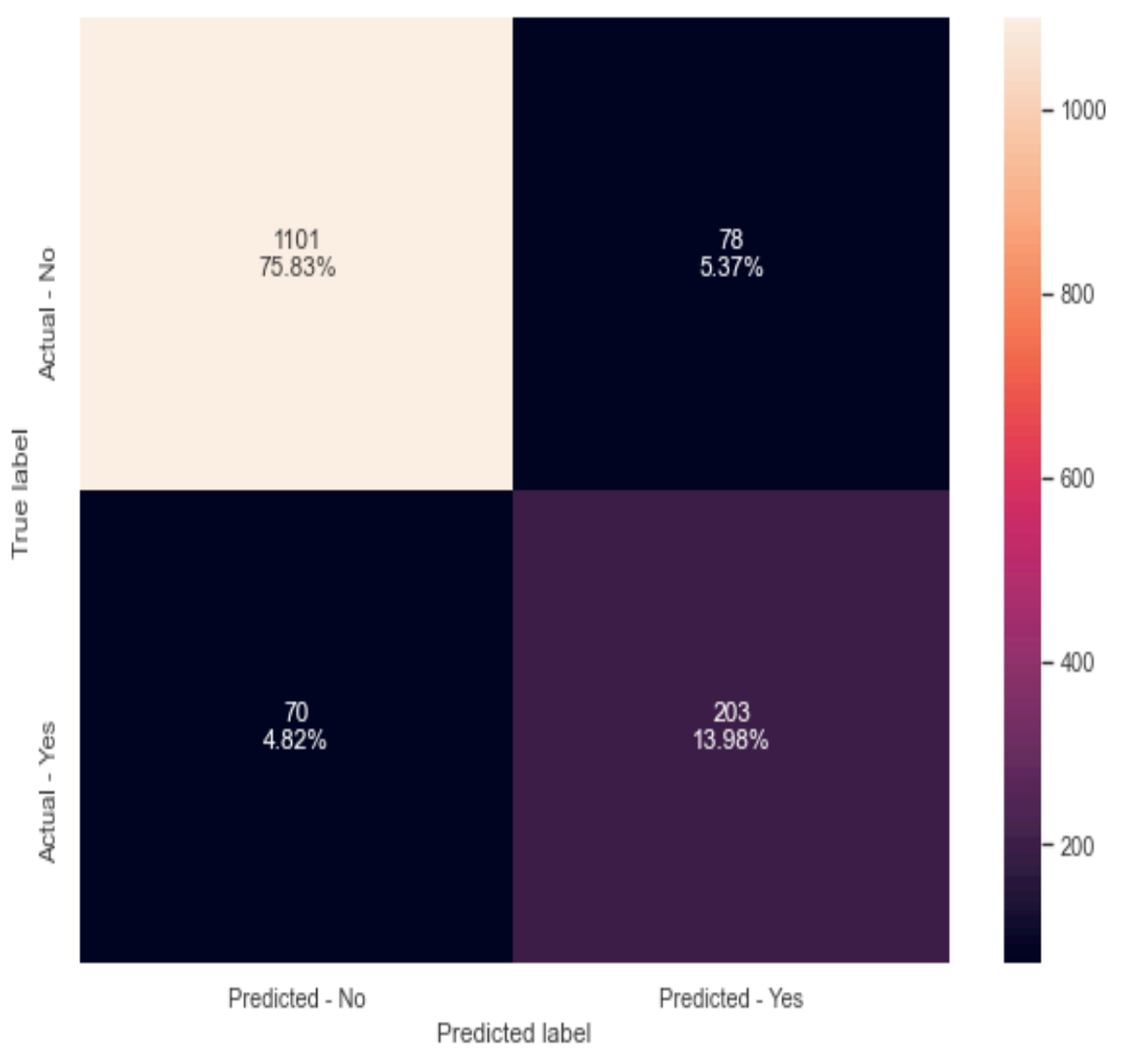
- Data was Preprocessed and Split into Train and Test sets at a 70:30 Ratio

- Modelling was executed using the Decision Tree, Random Forest and Bagging Classifier and then Adaboost, GCBoost and XGBoost and Stacking Classifier
- and results compared

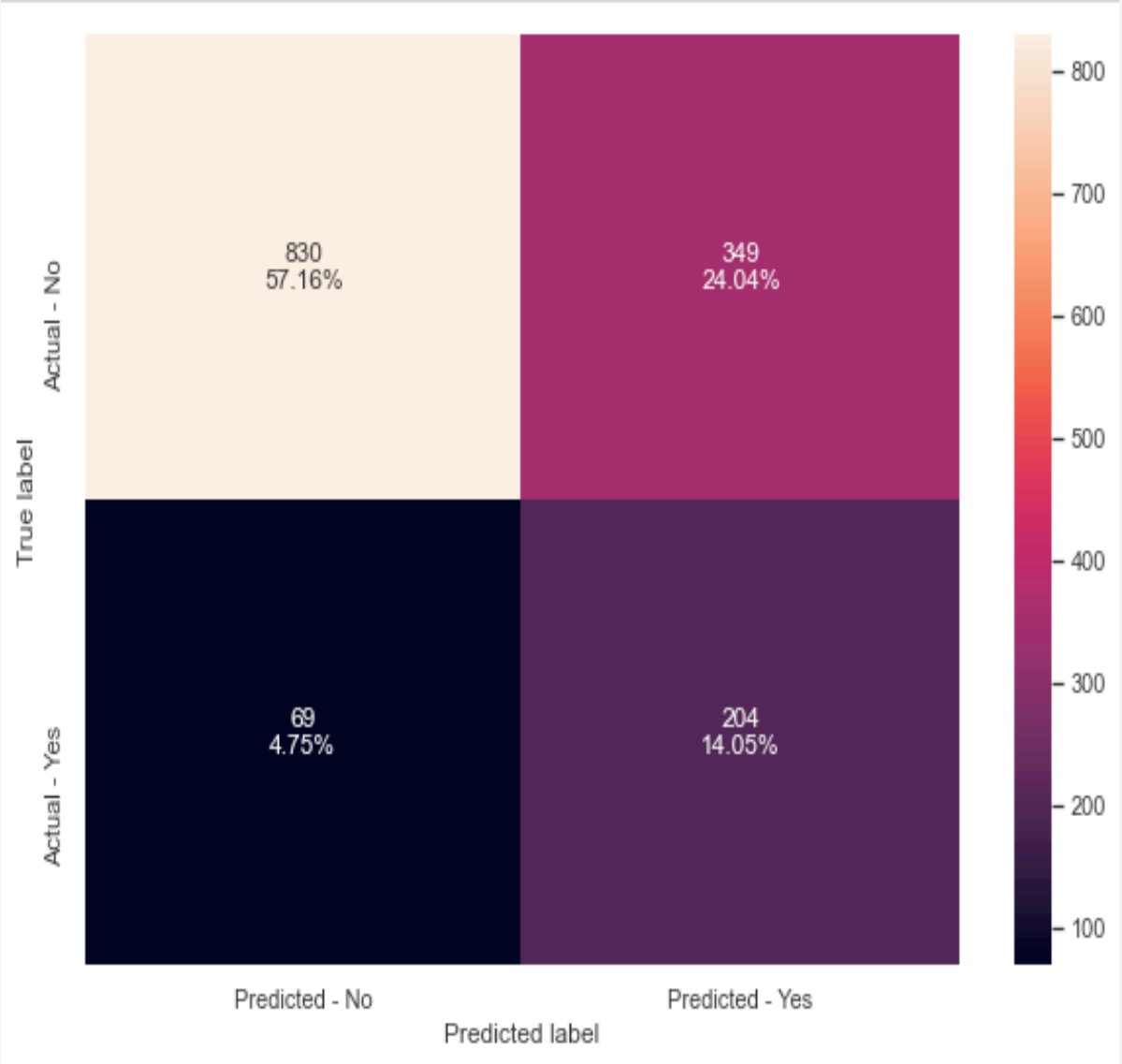
- Correlation between NumberofChildrenVisiting and NumberofPersonsVisiting was addressed and a model chosen with improved model performance parameters (Accuracy, Precision, Recall, RUC and f1scores)

- Independent variables (X) against a target or dependent variable(Y=ProdTaken) in this case.

CONFUSION MATRIX PRE & POST TUNING (DECISION TREE CLASSIFIER)

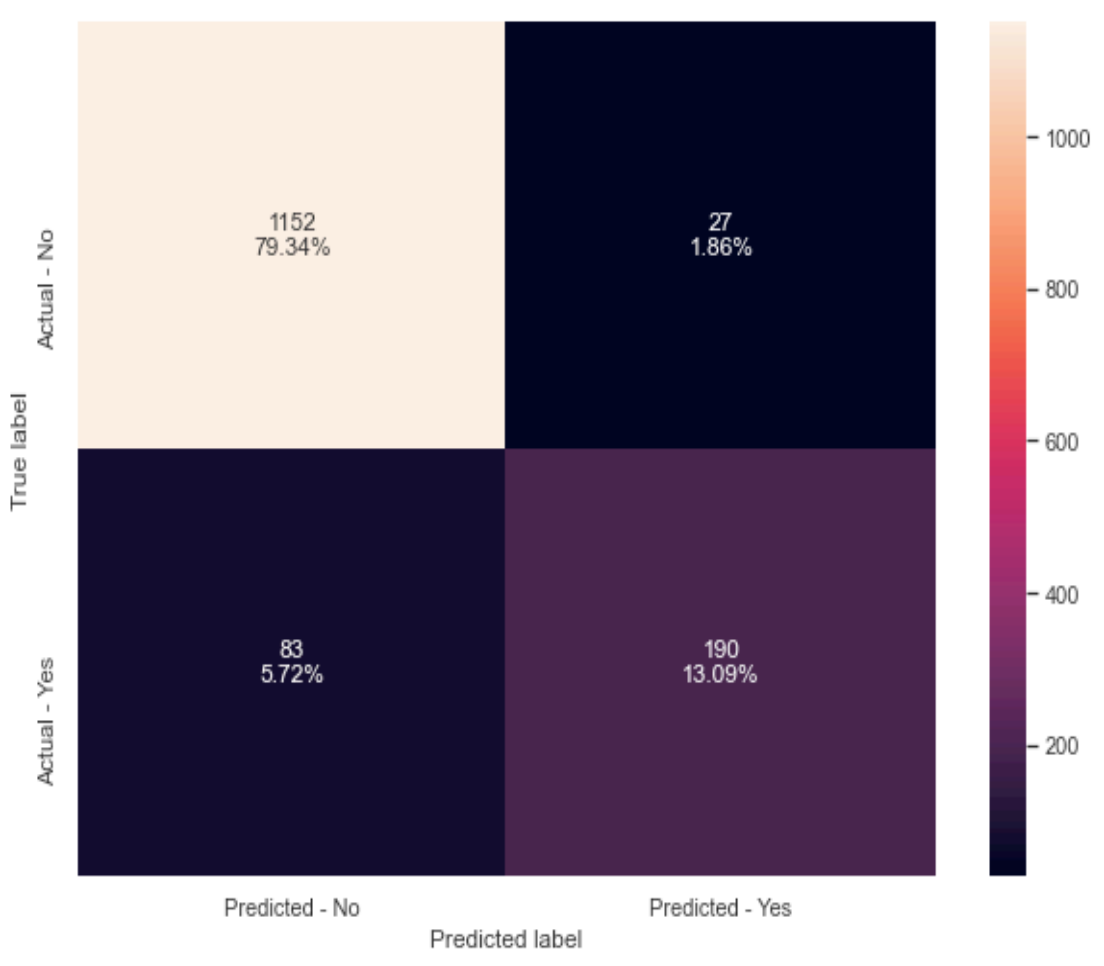


PRE-TUNE

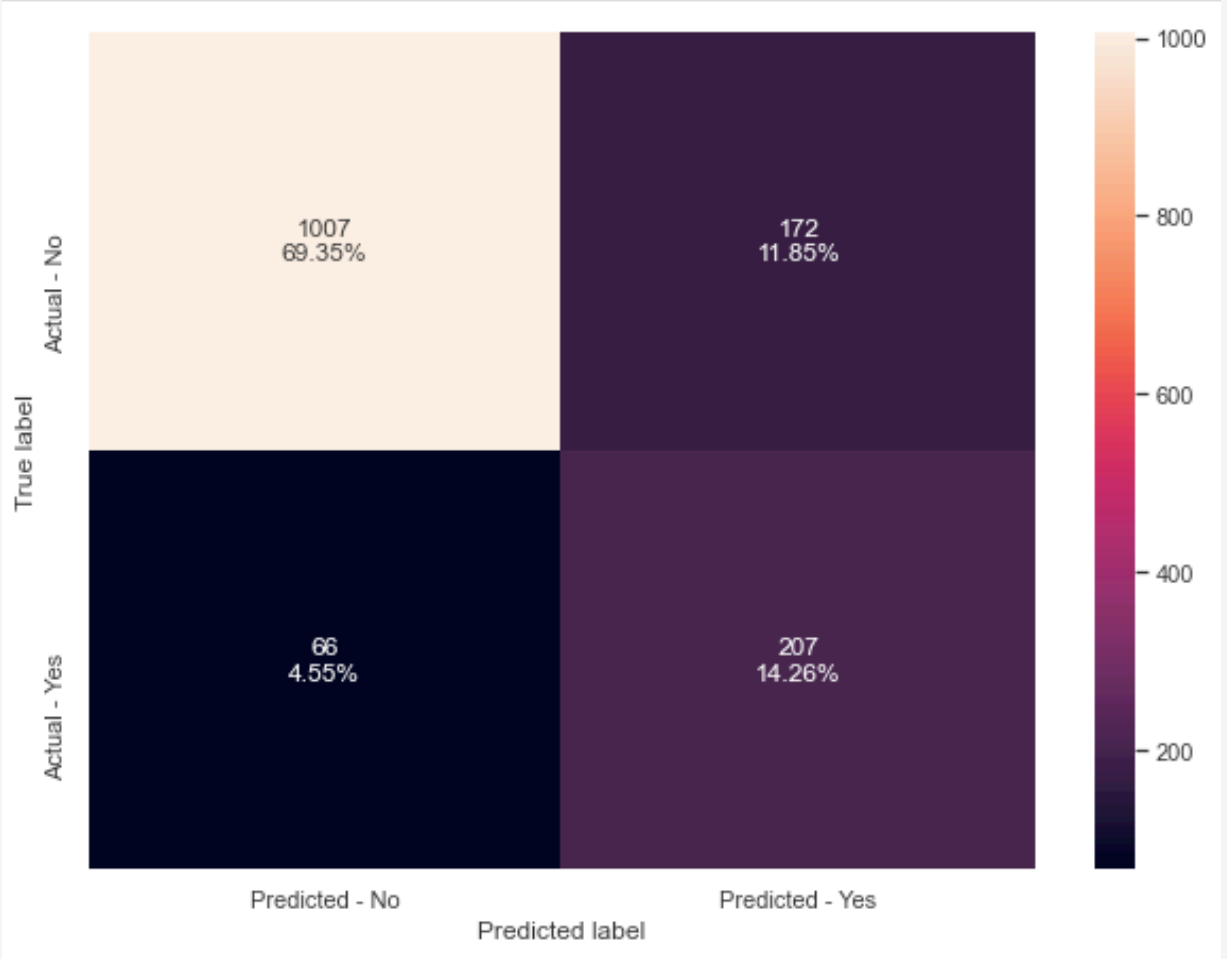


POST -TUNE

CONFUSION MATRIX PRE & POST TUNING (XGBOOST)



PRE-TUNE



POST TUNE

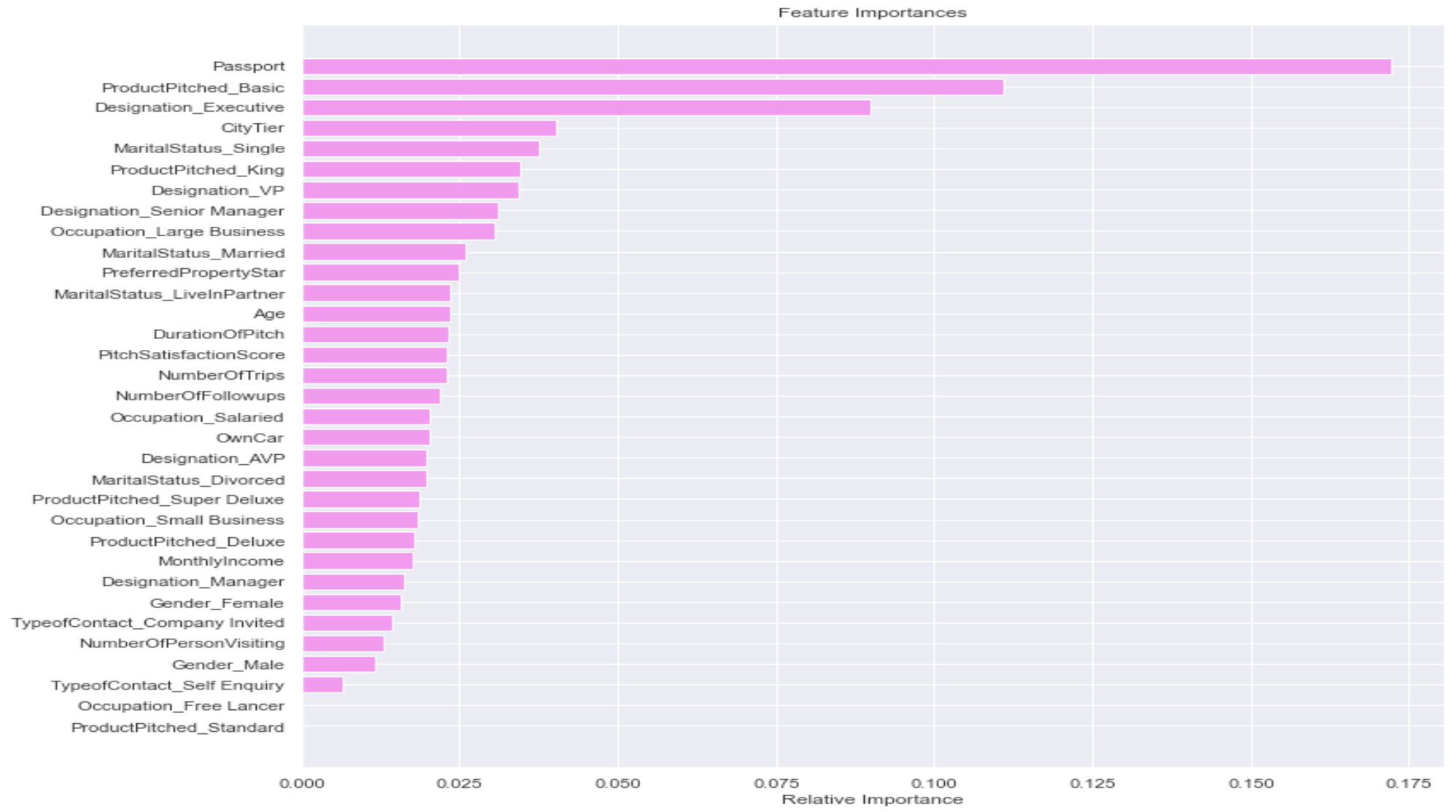
- As can be seen , the tuned XGBoost Classifier improved the Recall significantly though it could be far better if the dataset was exhaustive. Hence more effort should be geared toward gathering more data for a decent analysis

MODEL PERFORMANCE INSIGHTS- LOGISTICS REGRESSION

	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
13	Tuned XGBoost Classifier	0.909601	0.836088	0.921630	0.758242	0.696682	0.546174
1	Tuned Decision Tree	0.746529	0.712121	0.764890	0.747253	0.408027	0.368897
0	Decision Tree	1.000000	0.898072	1.000000	0.743590	1.000000	0.722420
4	Tuned Bagging Classifier	0.581684	0.557163	0.741379	0.739927	0.274362	0.260982
12	XGBoost Classifier	0.999409	0.924242	0.996865	0.695971	1.000000	0.875576
2	Bagging Classifier	0.993501	0.908402	0.965517	0.615385	1.000000	0.857143
5	Random Forest	1.000000	0.916667	1.000000	0.611722	1.000000	0.917582
9	Tuned Adaboost Classifier	0.990251	0.878788	0.963950	0.604396	0.984000	0.708155
6	Weighted Random Forest	1.000000	0.915289	1.000000	0.597070	1.000000	0.926136
3	Weighted Bagging Classifier	0.994387	0.907025	0.971787	0.593407	0.998390	0.870968
14	Stacking Estimator	0.955982	0.870523	0.846395	0.582418	0.913706	0.682403
11	Tuned Gradient Boost Classifier	0.927622	0.880165	0.666144	0.527473	0.929978	0.761905
7	Tuned Random Forest	0.922009	0.869835	0.623824	0.454212	0.943128	0.756098
10	Gradient Boost Classifier	0.892467	0.866391	0.501567	0.454212	0.874317	0.733728
8	AdaBoost Classifier	0.851699	0.831267	0.351097	0.340659	0.717949	0.588608

- A predictive model that can be used by the Visit with Us to target potential with a high Recall on the training set and formulate marketing strategies accordingly as well as allocate scarce resources toward maximising profit and turnover via fees and good customer traffic accordingly

IMPORTANT FEATURES AS RETURNED BY TUNED XGBOOST CLASSIFIER



- The most important features are Passport, Basic Travel Package, Executives by Designation, CityTier and Marital Status

CONCLUSIONS-KEY INSIGHTS FROM MODELING

- We built a predictive model that:

- a) Visit with Us can deploy to identify customers who are likely to subscribe to specific travel packages
- b) Visit with Us can use to find the catalysts of subscription to any Travel package
- c) As a result, the firm can take appropriate actions to build better marketing, advertisement and budgeting policies and recommendations

- Factors that influence Product Taken based on preferred model are - Passport,Product_Pitched_Basic,Designation Executive, CityTier, MaritalStatus_Single

- Passport: Customers with a Passport clearly have very high chances of buying a travel package. This is simply a given. Visit with Us should include a pointer in their subscription forms requesting availability of a passport and propose value added services to secure one in addition to the preferred travel package of sorts, This speaks to product design and customisation and would appeal mostly to the yet untapped customer brackets like singles and younger customers.

- Product_Pitched_Basic: Basic Travel package appeals a whole lot to potential subscribers as it is most likely a budget product and is quite affordable. Visit with Us should tweak their marketing policies to prioritize this package over others in order to gain an unprecedented spike in critical mass customer subscription

- Designation_Executive: This speaks to the top management of firms and organisations who maintain a relationship with Visit with Us. The company can take advantage of this to attract more customers in this bracket by incentivizing specific subscriptions to specific top management designations as they have a very busy schedule as such can only be captured with unique offerings

- CityTier: This is also an important factor for the retention and expansion of the customer base of Visit with Us. It is logical that the more developed a city is, the higher the standard of living and hence a state or country is. As such migration is inevitable within CityTiers and thus the need for the Policy team do do the needful.

- Even though it appears our data collection technique is quite rich in depth as it clearly reflects unique insightful patterns that greatly speaks to the propensity to delineate between both divides regarding Product Taken,reources can be allocated specifically to secure an exhaustive data that when analyzed would give a more generalized model performance.

- This will further boost our chances of building more robust predictive models

CONCLUSIONS-KEY INSIGHTS FROM EDA

- 75%% of the customers are 43 yrs or less with the oldest at 61 years
- 75% of the customers were accorded 19 or less with maximum at 127
- 2 to 4 people are willing to go on a trip with the customers
- The distribution for Number of follow-ups has 3 peaks at 3, 4 and 5 with
- approximately 400, 2000 and 750 customers respectively
- 75% of the customers earn a monthly income of 25374 or less
- 75% of the distribution have 2 children or less who plan to go on a trip with customers
- 81.2% of the customers did not subscribe to any packages while just 18.8% did as earlier captured in the problem statement
- 70.7% of customers in data made Self Enquiries about the Products on offer while 29.3% were invited by the Company
- Average Pitch score across the distribution is 3. The other two peaks are at 1 and 5
- 59.7% of the customers are Male while 40.3% are Females.
- The Basic package was the most subscribed at 37.9% (1831 customers) followed closely by Deluxe at 35.5%
- Customers who are married top the chart of the Visit with Us database with 47.8% (2311)
- Divorced customers come second with 19.4% and Singles following closely at 18.7%
- 61.5% of the customers prefer 3 star property types when they make a trip as against 19.7% for 5 stars
- 48.5% of the customers in the distribution are Salaried workers ,followed closely Small Business owners with 42.5% and a distant 8.9% by Large Businesses
- 65.2% of the customers are from Tier1 cities whileTier3 with the lowest standards of living and dense population coming in at 30.7%.
- Executives lead the pack with 37.9% of the customer base followed closely by Managers at 35.5%
- 70.7% of the customers own a Passport while 29.3% don't
- 61.9% of the customers at Visit with Us own a car
- 38.1% of the customers do not own a car
- Males have a higher subscription rate compared to females (Pls view Slide 27)
- Females spend longer Pitching times than Males
- Customers who did not subscribe to any package earn higher income than those who did.
- Males have had twice the Number of Trips by woman
- Those with Passports have an estimated 35% subscription as against those without with approx. 12%
- Customers who subscribed had shorter Pitches than those who didn’t
- Male customers who subscribed dwarf their femal counterparts in monthly income earnings however both male and female non-subscribers earn far much higher income than the subscribers

BUSINESS RECOMMENDATIONS

The following are actionable insight from our Analysis,

- These insights can be inferred to form the basis of a formidable marketing vis-à-vis advert campaigns to gain a competitive edge, more market share and ultimately increased revenue
- Resources should effectively be allocated for product designs or the new ones tweaked to be more robust with value added services like facilitating passport acquisition or renewal to endear prospects to Visit with Us
- The much needed critical mass needed to expand the customer base would come naturally as a result
- As evidenced in the analysis, an extensive and aggressing marketing initiative ought to be considered to tap into the significant chunk of customers with potentials for for subscription.
- Efforts should be galvanized to tap more into Executives and Senior Management customers and prospects as they present opportunities for referrals and huge corporate accounts.
- Exhaustive sacrifice must indeed be made to avail the management of Visit with Us Competition Data in order to enable a surgical counteractive initiative to design policies based on the predicted models to stay ahead of their peers.
- Effective and competitive Pricing of Travel packages with additional incentives will appeal a lot more to untapped brackets within the customer base and also enable customer retention and referral that will ultimately spike an attendant traffic in patronage and ultimately revenue
- It is quite apparent the training data is not exhaustive. More effort should be geared toward leveraging on a wider observation sample to draw more inclusive, extensive and impactful insights

THANK YOU