Travel Package Purchase Prediction Project 5

Objectives

- To predict which customer is more likely to purchase the newly introduced travel package.
- Identify which variables are most significant.
- Analyze the customers' data and information to provide recommendations to the Policy Maker and Marketing Team and build a model to predict the potential customer who is going to purchase the newly introduced travel package.

Data Information

Customer details:

Variable	Description
CustomerID	Unique customer ID
ProdTaken	Whether the customer has purchased a package or not (0: No, 1: Yes)
Age	Age of customer
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

Customer interaction data:

Variable	Description
PitchSatisfactionScore	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
DurationOfPitch	Duration of the pitch by a salesperson to the customer

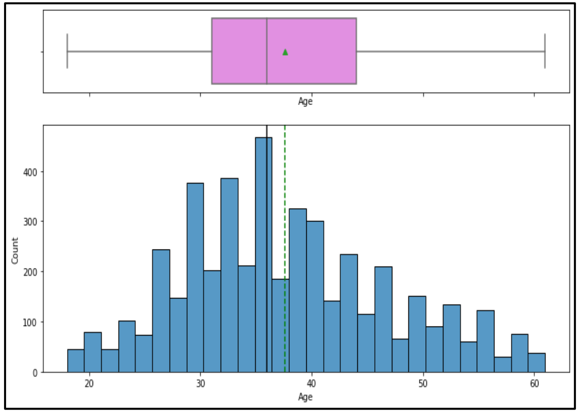
Observations	Variables
4888	20

Note:

- There are 8 variables with missing values:
- [Age, TypeofContact, DurationOfPitch, NumberOfFollowups, PreferredPropertyStar, NumberOfTrips, NumberOfChildrenVisiting, MonthlyIncome]
- We noticed two spellings for 'Female'
- The CustomerID attribute does not add any information to our analysis as all the values are unique. We can neglect this information for our model prediction.
- We will be changing the TypeofContact column to categorical to make our analysis easier

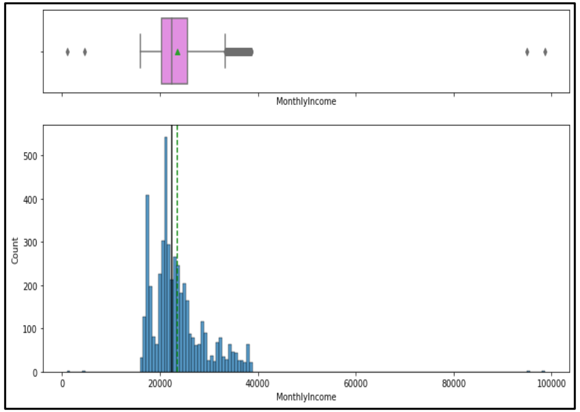
Exploratory Data Analysis – Age, Monthly Income, Duration of Pitch

Age



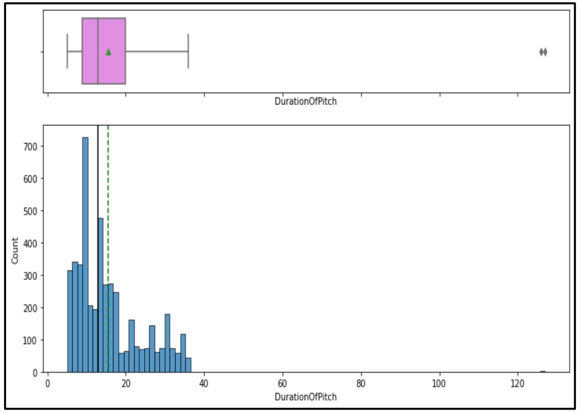
- The distribution of Age is mostly normally distributed.
- The mean and median are fairly close.
- There is a slight tail on the right.

Monthly Income



- The distribution of monthly income appears to be mostly normally distributed.
- The distribution has outliers on both ends.
- The mean and the median are very close together.

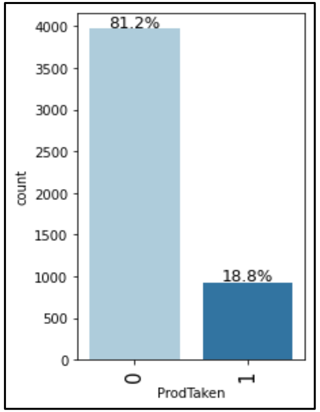
Duration of Pitch



- The distribution of Duration of Pitch is skewed to left.
- There are outliers to the left of the distribution.
- The mean and the median are also close together.

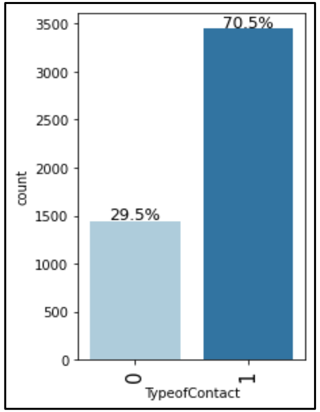
Exploratory Data Analysis – ProductTaken, TypeofContact, CityTier, Occupation, Gender

Product Taken



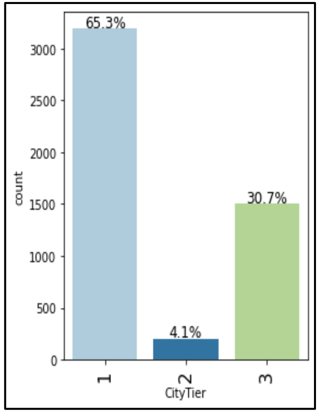
- 18.8% of customers did buy the vacation package.

Type of Contact



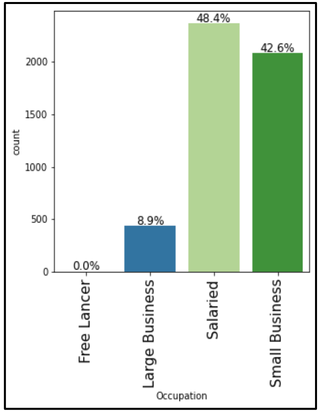
- 70.5% of customers reached out themselves about information on travel packages.

City Tier



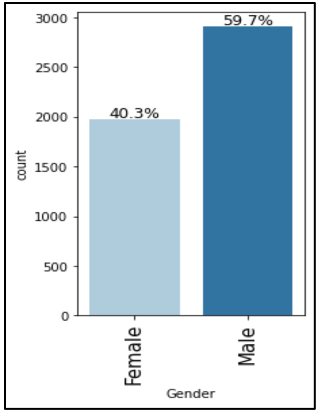
- 65.3% of customers are from tier 1 cities.
- 30.7% of customers are from tier 3 cities.

Occupation



- 48.4% of customers are salaried.
- 42.6% of customers own small businesses
- Free Lancer had 0.0% of customers (2 customers)

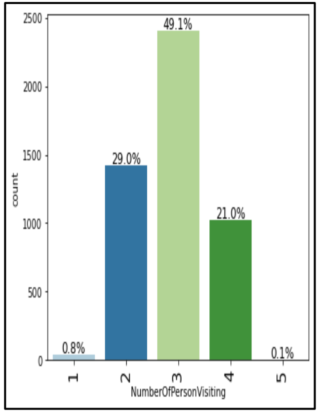
Gender



- 59.7% of customers are male.

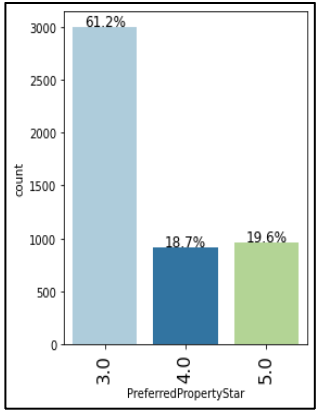
Exploratory Data Analysis – NumberOfPeopleVisiting, PreferredPropertyStar, MaritalStatus, Occupation, Passport

Number of Person Visiting



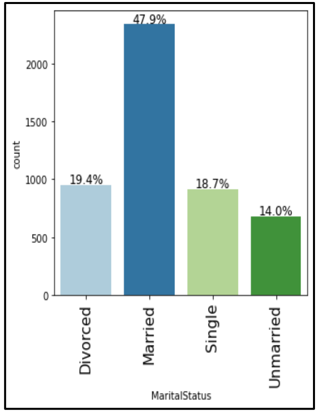
- Most customers have 3 people visiting at 49.1%.
- The next highest number of people visiting are 2 people at 29.0%.

Preferred Property Star



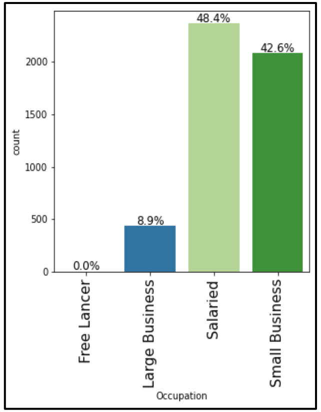
- Most customers prefer 3 star properties at 61.8%.
- This suggests most customers might prefer a more budget friendly property.

Marital Status



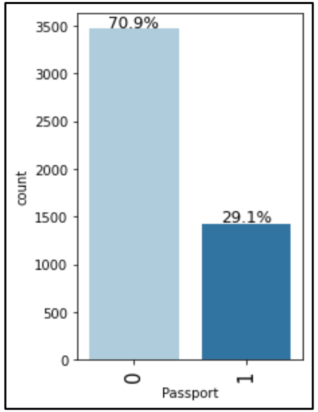
- Almost half of the customers are married at 47.9%.
- Other customer marital status are almost evenly distributed.

Occupation



- 48.4% of customers are salaried.
- 42.6% of customers own small businesses
- Free Lancer had 0.0% of customers (2 customers)

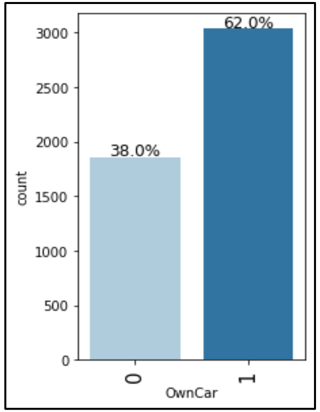
Passport



- A majority of customers do not have a passport at 70.9%.

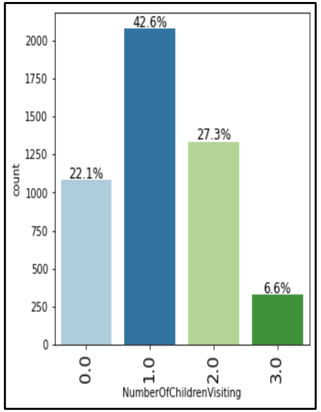
Exploratory Data Analysis – OwnCar, NumberofChildrenVisiting, Designation, PitchSatisfactionScore

Own Car



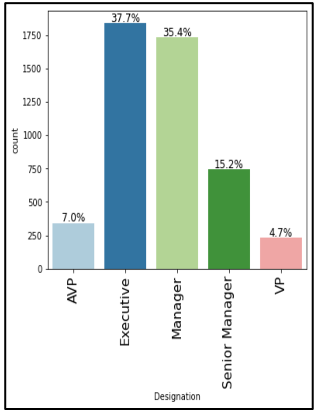
- Most customers own a car at 62.0%.

Number of Children Visiting



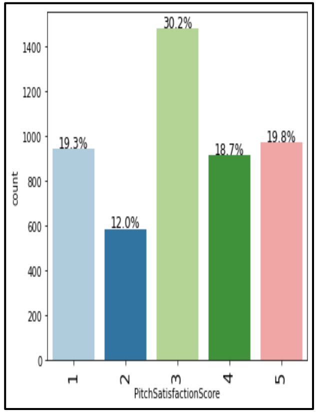
- Most customers have 1 child visiting at 43.9%.
- 27.3% of customers have 2 children visiting.
- 22.1% of customers do not have any children visiting.

Designation



- Most customers are either Executive or Manager at 37.7% and 35.4% respectively.

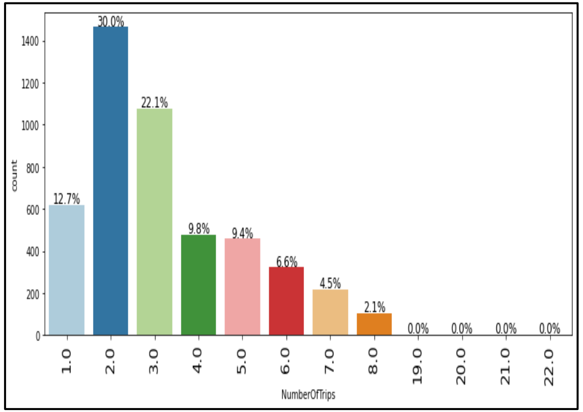
Pitch Satisfaction Score



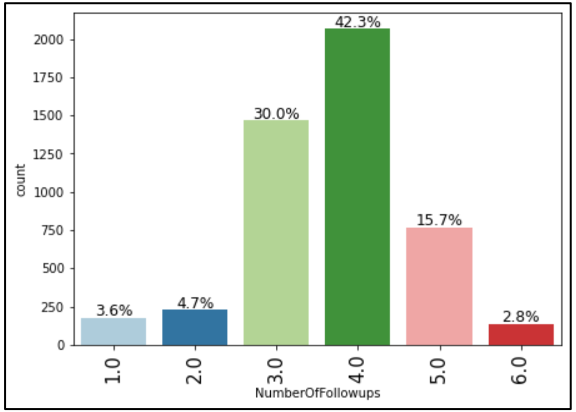
- Most customers gave a satisfaction score of 3 at 30.2%
- 19.3% of customers gave a rating of 1.
- The amount of low satisfaction is concerning; we will note factors in our analysis.

Exploratory Data Analysis – NumberOfTrips, NumberOfFollowups, ProductPitched

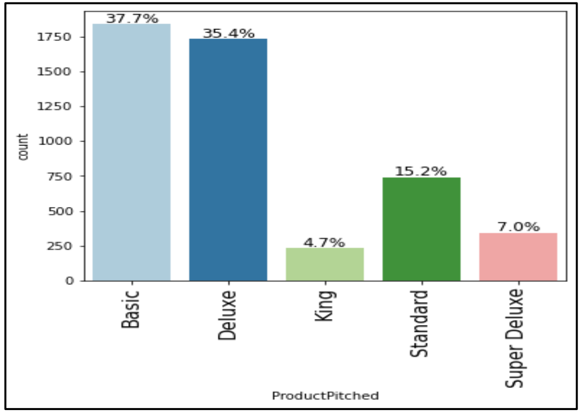
Number of Trips



Number of Follow-ups



Product Pitched



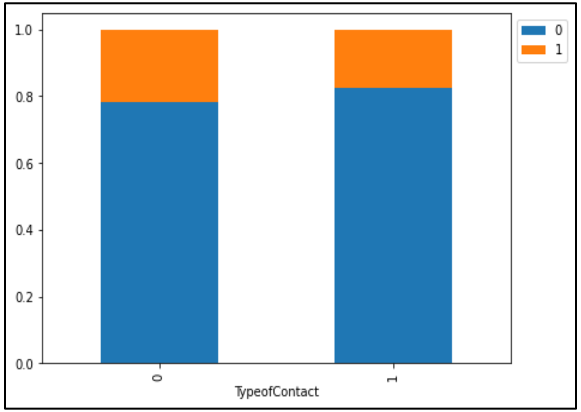
- The distribution of personal loans by the number of family members is almost evenly distributed

- Most customers receive 3 or 4 follow ups at 30.0% and 43.2% respectively.

- Most customers were pitched the basic and deluxe package at 37.7% and 35.4%.
- Most salespeople pitch the more budget friendly options.

Exploratory Data Analysis – TypeOfContact, CityTier, Occupation vs ProductTaken

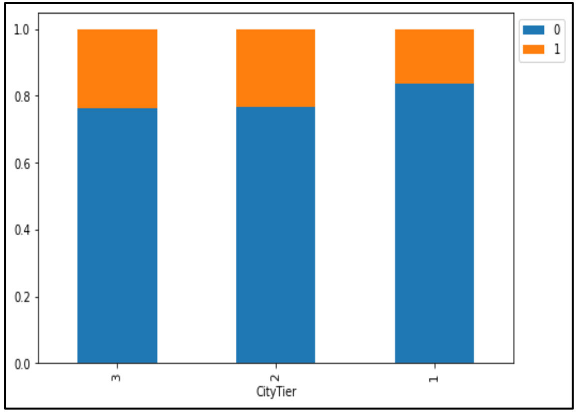
Type of Contact vs Product Taken



ProdTaken	0	1	All
TypeofContact			
All	3968	920	4888
1	2837	607	3444
0	1131	313	1444

- Almost 2/3rd of customers who self inquired on information also took the travel package

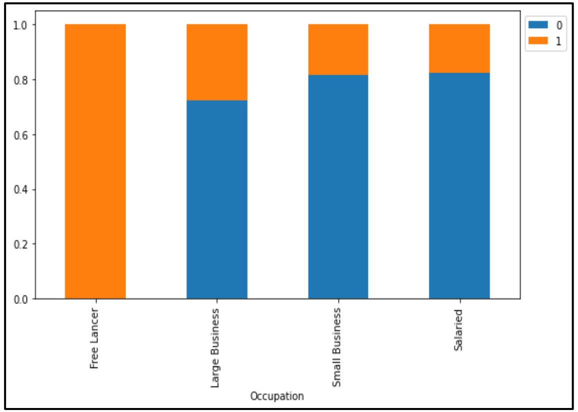
City Tier vs Product Taken



ProdTaken	0	1	All
CityTier			
All	3968	920	4888
1	2670	520	3190
3	1146	354	1500
2	152	46	198

- There were more customers in city tier 1 that took the travel package
- This could be interesting to investigate

Occupation vs Product Taken

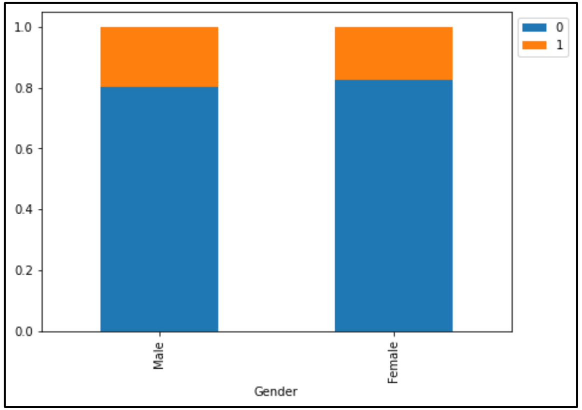


ProdTaken	0	1	All
Occupation			
All	3968	920	4888
Salaried	1954	414	2368
Small Business	1700	384	2084
Large Business	314	120	434
Free Lancer	0	2	2

- Most of the customers who took the travel package were either salaried or small business owners.
- All free lancer customers bought the travel package but there were only two free lancers in the data set.

Exploratory Data Analysis – Gender, NumberOfPersonVisiting, PreferredPropertyStar vs ProductTaken

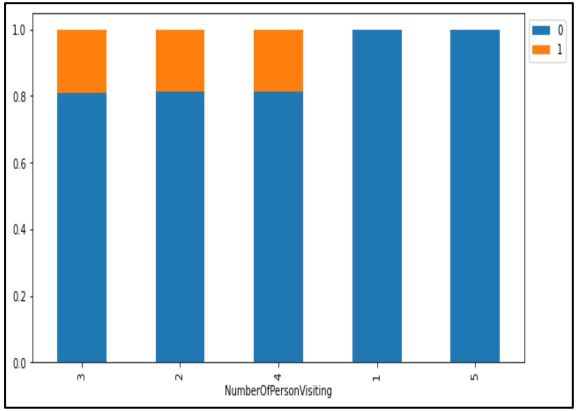
Gender vs Product Taken



ProdTaken	0	1	All
Gender			
All	3968	920	4888
Male	2338	578	2916
Female	1630	342	1972

- The data suggests that more male customers take the travel package

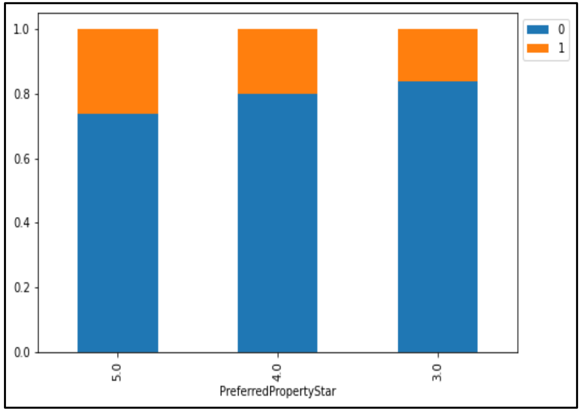
Number of Person Visiting vs Product Taken



ProdTaken	0	1	All
NumberOfPersonVisiting			
All	3968	920	4888
3	1942	460	2402
2	1151	267	1418
4	833	193	1026
1	39	0	39
5	3	0	3

- Customers who have one or five visitors do not take the travel package

Preferred Property Star vs Product Taken

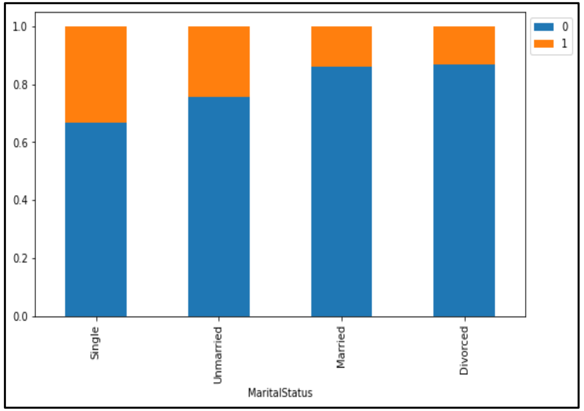


ProdTaken	0	1	All
PreferredPropertyStar			
All	3948	914	4862
3.0	2511	482	2993
5.0	706	250	956
4.0	731	182	913

- Most customers prefer 3 star properties compared to higher star properties
- This could suggest that more budget friendly properties are more appealing

Exploratory Data Analysis – MaritalStatus, NumberOfTrips, Passport vs ProductTaken

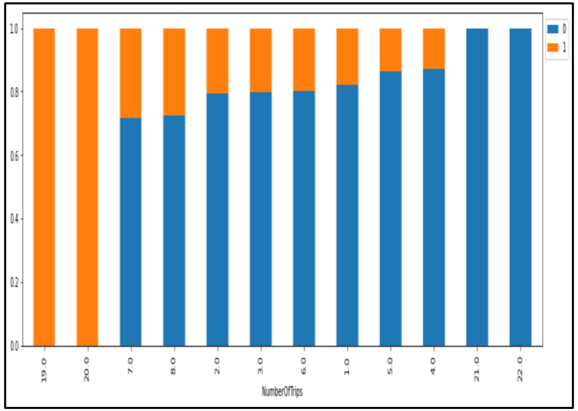
Marital Status vs Product Taken



ProdTaken	0	1	All
MaritalStatus			
All	3968	920	4888
Married	2014	326	2340
Single	612	304	916
Unmarried	516	166	682
Divorced	826	124	950

- Most customers who take the travel package are either Single or Married

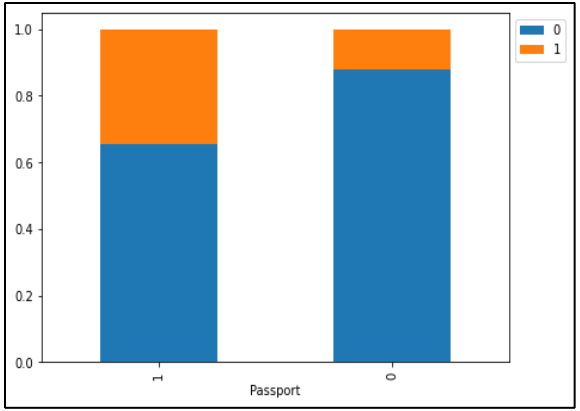
Number of Trips vs Product Taken



ProdTaken	0	1	All
NumberOfTrips			
All	3840	908	4748
2.0	1165	299	1464
3.0	862	217	1079
1.0	508	112	620
6.0	258	64	322
5.0	396	62	458
7.0	156	62	218
4.0	417	61	478
8.0	76	29	105
19.0	0	1	1
20.0	0	1	1
21.0	1	0	1
22.0	1	0	1

- Customers who take two or three trips are more likely to take the travel package according to this sample.

Passport vs Product Taken

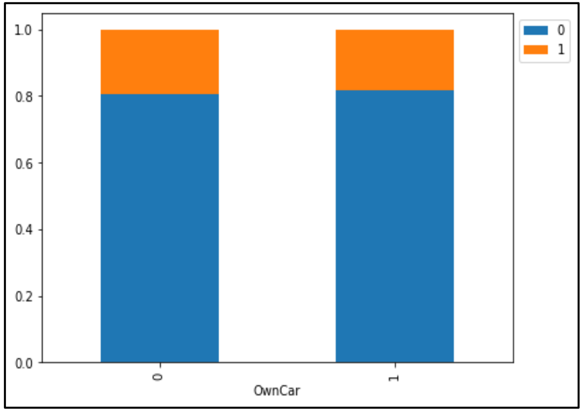


ProdTaken	0	1	All
Passport			
All	3968	920	4888
1	928	494	1422
0	3040	426	3466

- The number of customers who took the travel package are almost evenly split between customers who have a passport and customers who do not.

Exploratory Data Analysis – OwnCar, NumberOfChildrenVisiting, Designation vs ProductTaken

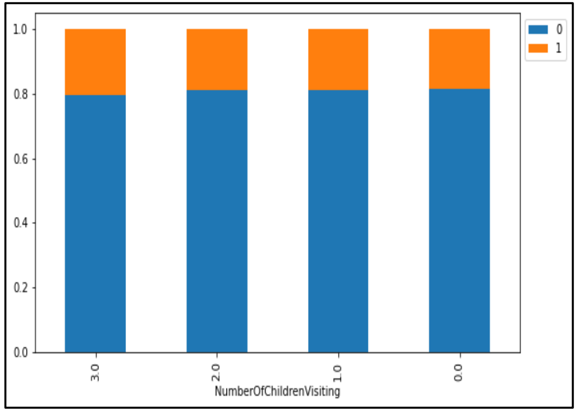
Own Car vs Product Taken



ProdTaken	0	1	All
OwnCar			
All	3968	920	4888
1	2472	560	3032
0	1496	360	1856

- More customers who own a car also took the travel package.

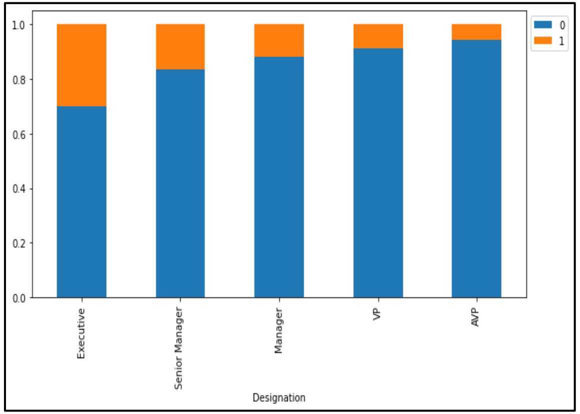
Number of Children Visiting vs Product Taken



ProdTaken	0	1	All
NumberOfChildrenVisiting			
All	3909	913	4822
1.0	1688	392	2080
2.0	1082	253	1335
0.0	880	202	1082
3.0	259	66	325

- More customers with one child visiting also bought a travel package
- Less customers with three children bought a travel package

Designation vs Product Taken

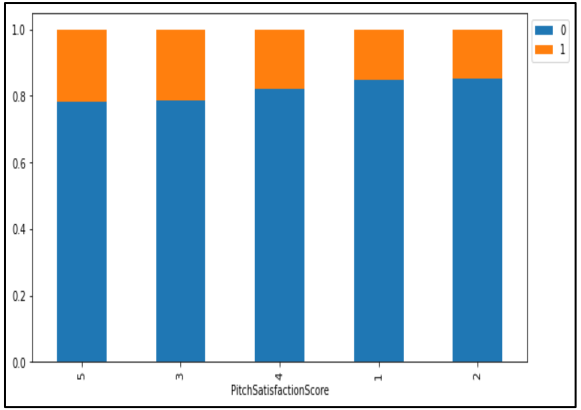


ProdTaken	0	1	All
Designation			
All	3968	920	4888
Executive	1290	552	1842
Manager	1528	204	1732
Senior Manager	618	124	742
AVP	322	20	342
VP	210	20	230

- There were more customers who were executives that bought a travel package
- AVP and VP had the least number of customers who bought packages, but they also had significantly smaller representation.

Exploratory Data Analysis – PitchSatisfactionScore, ProductPitched, NumberOfFollowups vs ProductTaken

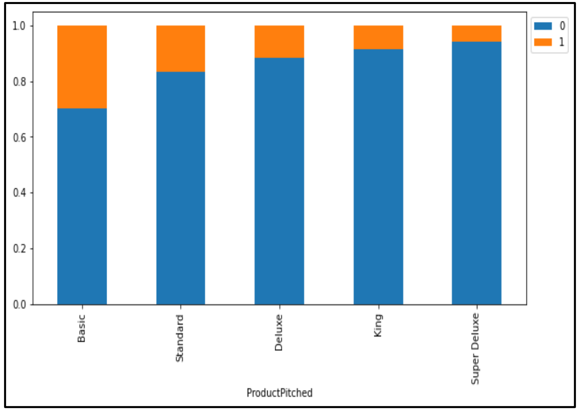
Pitch Satisfaction vs Product Taken



ProdTaken	0	1	All
PitchSatisfactionScore			
All	3968	920	4888
3	1162	316	1478
5	760	210	970
4	750	162	912
1	798	144	942
2	498	88	586

- On average most pitch scores were 3 for satisfaction

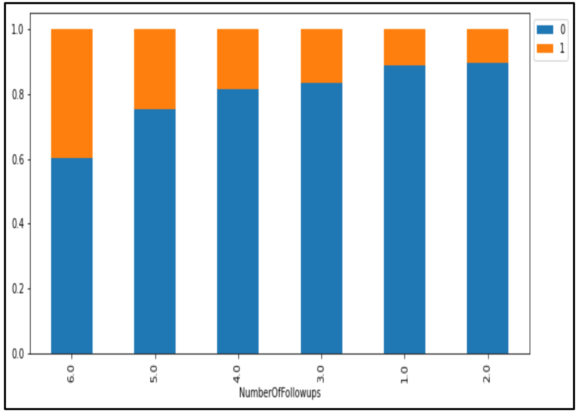
Product Pitched vs Product Taken



ProdTaken	0	1	All
ProductPitched			
All	3968	920	4888
Basic	1290	552	1842
Deluxe	1528	204	1732
Standard	618	124	742
King	210	20	230
Super Deluxe	322	20	342

- Customers who were pitched the Basic or Deluxe package were more likely to buy the package
- Less customers who were pitched the higher end packages were less likely to buy the travel package

Number of Follow-ups vs Product Taken

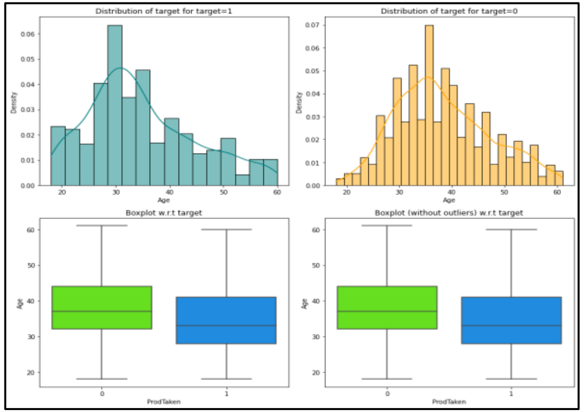


ProdTaken	0	1	All
NumberOfFollowups			
All	3931	912	4843
4.0	1689	379	2068
3.0	1222	244	1466
5.0	577	191	768
6.0	82	54	136
2.0	205	24	229
1.0	156	20	176

- Customers who had three to four follow ups bought travel package.
- The data would suggest that more follow ups are required to increase the likelihood of a package being purchased.

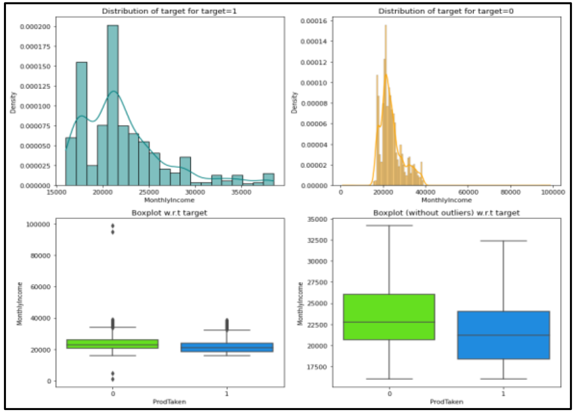
Exploratory Data Analysis

Age vs Product Taken



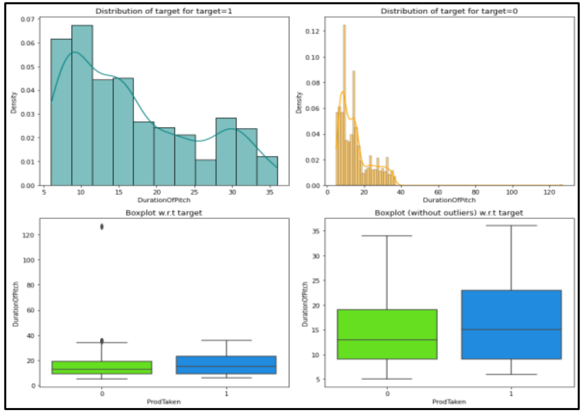
- The distribution for both outcomes are similarly distributed
- There are more customers with median age who did not take the product

Personal Monthly Income vs Product Taken



- There is a wider spread of the data for customers who took the product.
- Customers who did not take the product are mostly centered around a specific range of income.

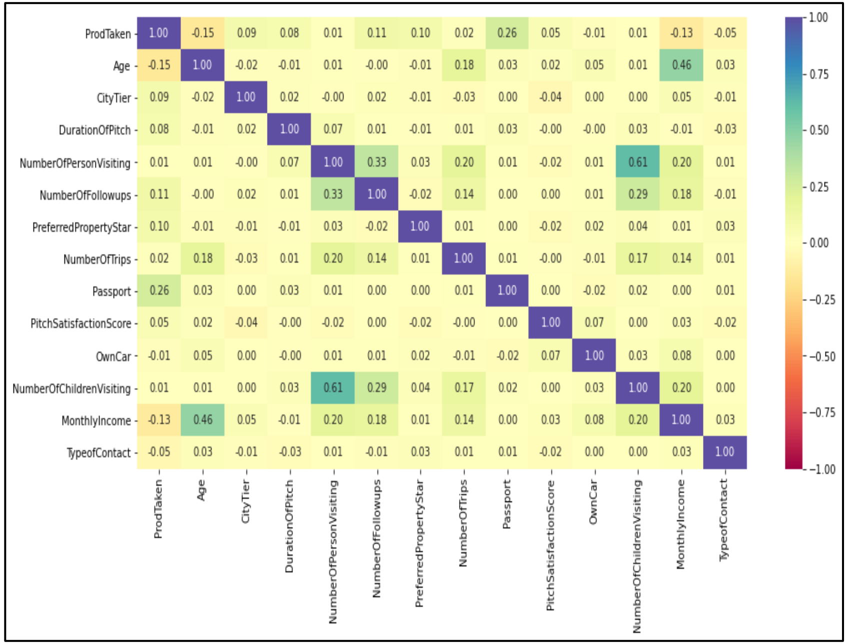
Personal Duration of Pitch vs Product Taken



- The duration of the sales pitch for customers who took the product was within a narrow range.
- There was a much wider range for customers who did not take the product with a large concentration that overlapped the duration for customers who did take the product.

Exploratory Data Analysis – Correlation

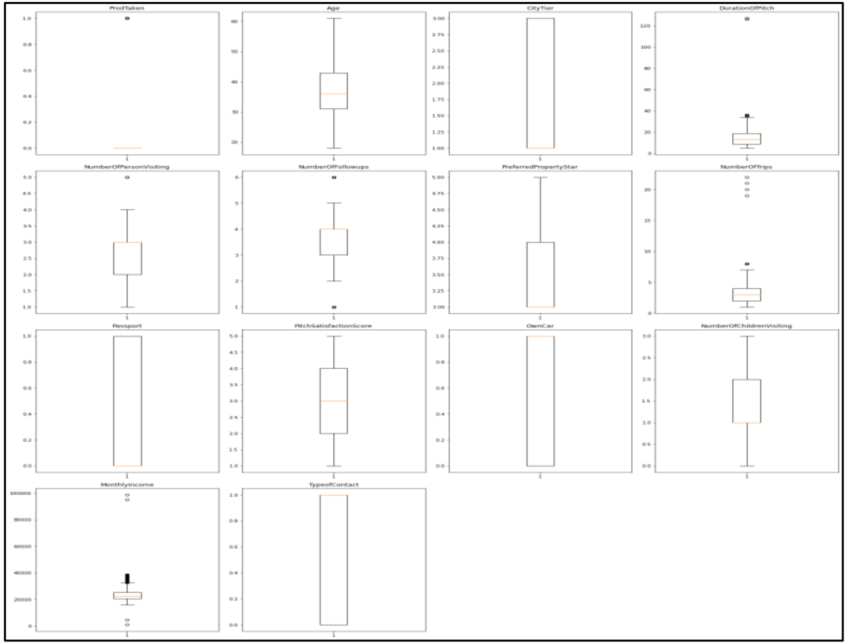
Heat Map Correlation



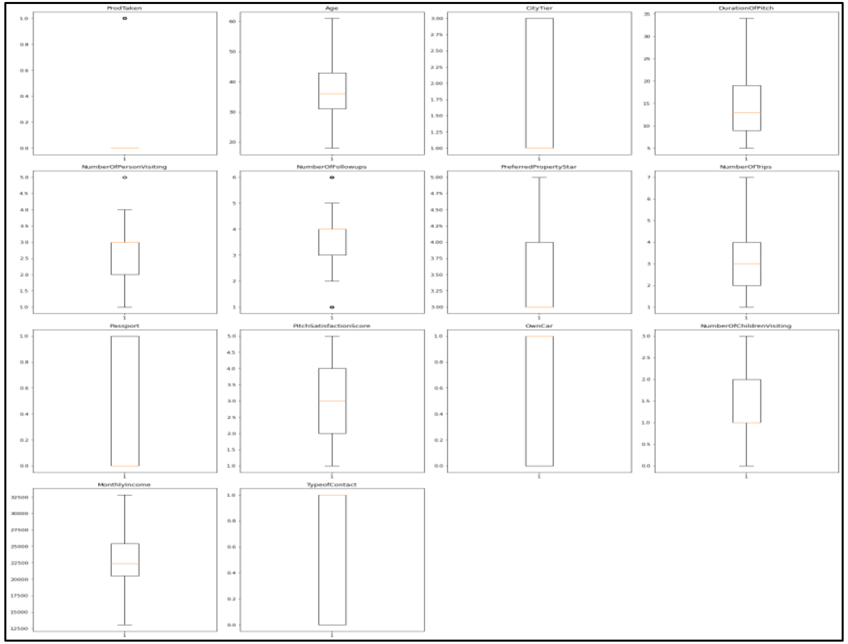
- Number of people visiting and the number of children visiting are highly correlated. There could be overlap in these two variables.
- We do also notice some correlation between monthly income and age.
- Monthly income and product taken have a negative correlation of -0.13.

Outlier Treatment

Pre Treatment



Post Treatment



- Three of our variables appear to have outliers: Duration of Pitch, Number of Trips, and Monthly Income
- We will treat only these three variable for outliers.
- We will exclude the other variables as they either do not have outliers or they have values of 1 or 0.

- All numerical variables that had outliers have been treated.
- Only the target variables have been treated for outliers.

Model Evaluation Criteron

- **The model can make wrong predictions as:**

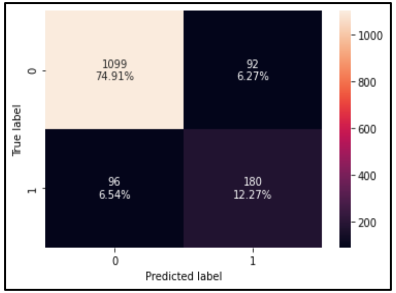
1. Predicting a customer will buy a travel package when they did not buy a travel package.
2. Predicting a customer will not buy a travel package when they did buy a travel package.

- **Which case is more important?**

1. If the model predicts a customer won't buy a travel package but it the customer indeed would then the company would incur the loss of a customer.
2. If the model predicts a customer will buy a travel package but the customer does not the company would lose time and resources that could be focused on potential customers.

- **Which metric to optimize?**
- We would want F1-Score to be maximized, the greater the F1-Score higher the chances of predicting both the classes correctly.

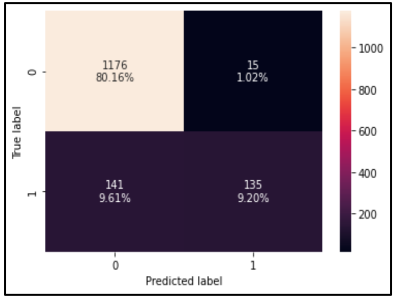
Decision Tree




```
Training performance:
    Accuracy  Recall  Precision   F1
0       1.0     1.0        1.0  1.0
Testing performance:
    Accuracy    Recall  Precision        F1
0  0.871847  0.652174   0.661765  0.656934
```

- The decision tree is overfitting the training data.
- We will try hyperparameter tuning and see if the model performance improves.

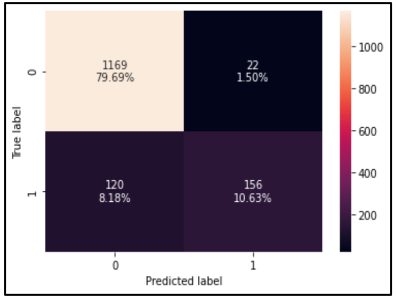
Random Forest




```
Training performance:
    Accuracy  Recall  Precision   F1
0       1.0     1.0        1.0  1.0
Testing performance:
    Accuracy   Recall  Precision        F1
0  0.893661  0.48913        0.9  0.633803
```

- Random forest is overfitting the training data as there is a huge difference between training and test scores for most of the metrics.
- The test recall is even lower than the decision tree but has a higher test precision.

Bagging Classifier

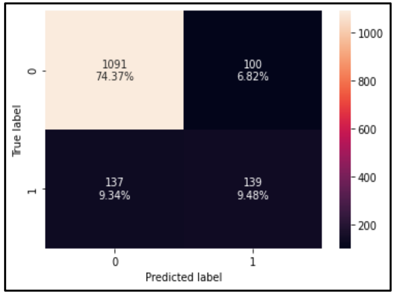



```
Training performance:
    Accuracy    Recall  Precision        F1
0  0.994154  0.970497   0.998403  0.984252
Testing performance:
    Accuracy    Recall  Precision        F1
0  0.903204  0.565217   0.876404  0.687225
```

- Bagging classifier giving a similar performance as random forest.
- The Bagging classifier is not overfitting the training as much as the random forest or the decision tree.
- It is also overfitting the training data but higher test recall than decision trees.

Hypertuning the models

Decision Tree - Tuned



Training performance:

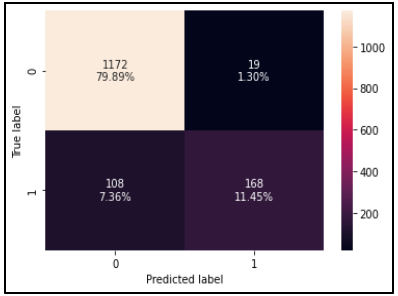
	Accuracy	Recall	Precision	F1
0	0.82432	0.487578	0.536752	0.510985

Testing performance:

	Accuracy	Recall	Precision	F1
0	0.838446	0.503623	0.58159	0.539806

- The test recall has decreased significantly after hyperparameter tuning and the decision tree is giving a generalized performance.

Random Forest - Tuned



Training performance:

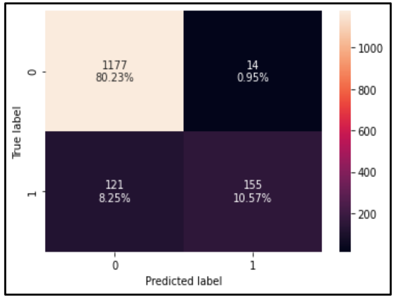
	Accuracy	Recall	Precision	F1
0	1.0	1.0	1.0	1.0

Testing performance:

	Accuracy	Recall	Precision	F1
0	0.913429	0.608696	0.898396	0.725702

- We can see that random forest's performance has increased as compared to the random forest model with default parameters.
- Model is still overfitting the data but not as much as the tuned bagging classifier.
- The test recall is still very low but improved compared to the random forest with the default parameters.

Bagging Classifier - Tuned



Training performance:

	Accuracy	Recall	Precision	F1
0	0.999708	0.998447	1.0	0.999223

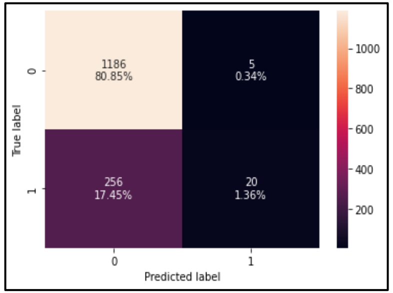
Testing performance:

	Accuracy	Recall	Precision	F1
0	0.907975	0.561594	0.91716	0.696629

•We can see that train accuracy and recall for the bagging classifier have mostly remained the same.
•The model is overfitting the data, as train recall are much higher than the test recall.

Logistic regression for Bagging Classifier and weighted class for Random Forest

Bagging Classifier – logistic regression



Training performance:

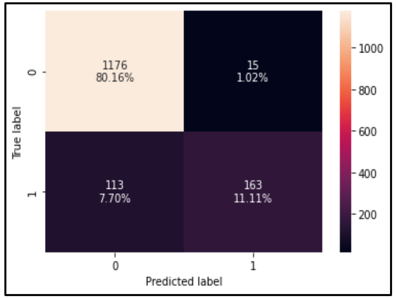
	Accuracy	Recall	Precision	F1
0	0.82549	0.082298	0.898305	0.150782

Testing performance:

	Accuracy	Recall	Precision	F1
0	0.822086	0.072464	0.8	0.13289

- Bagging classifier with logistic regression as base_estimator is not overfitting the data but the test recall is low.
- Ensemble models are less interpretable than decision tree but bagging classifier is even less interpretable than random forest. It does not even have a feature importance attribute.

Random Forest – class_weights



Training performance:

	Accuracy	Recall	Precision	F1
0	1.0	1.0	1.0	1.0

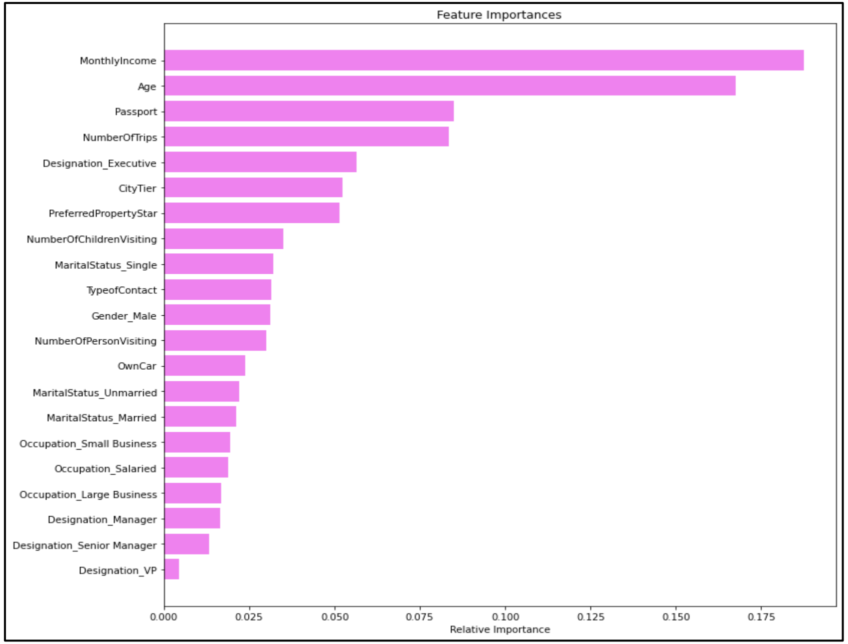
Testing performance:

	Accuracy	Recall	Precision	F1
0	0.912747	0.59058	0.91573	0.718062

- The model accuracy has mostly remained the same and is overall still overfitting.
- We see improved test precision and a drop in recall.

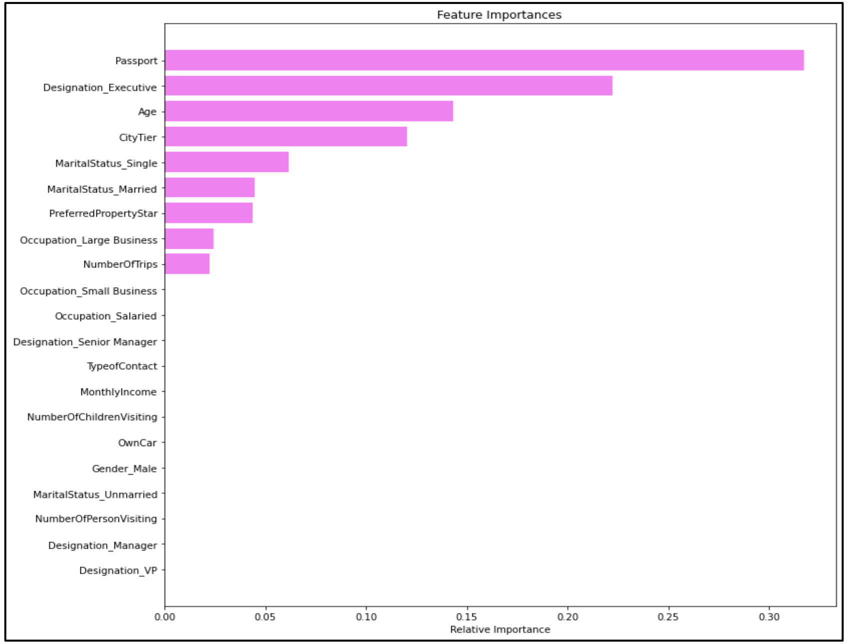
Feature Importance

Feature Importance



- Monthly income, age, and passport are the top 3 features.

Decision Tree
Feature Importance



- Passport, designation - executive, and age are the top 3 features.

Comparing all the models

Train Comparison Table

	Bagging classifier with default parameters	Tuned Bagging Classifier	Bagging classifier with base_estimator=LR	Decision Tree with default parameters	Tuned Decision Tree Classifier	Random Forest with deafult parameters	Tuned Random Forest Classifier	Random Forest with class_weights
Accuracy	0.994154	0.825490	0.825490	1.0	0.824320	1.0	1.0	1.0
Recall	0.970497	0.082298	0.082298	1.0	0.487578	1.0	1.0	1.0
Precision	0.998403	0.898305	0.898305	1.0	0.536752	1.0	1.0	1.0
F1	0.984252	0.150782	0.150782	1.0	0.510985	1.0	1.0	1.0

Test Comparison Table

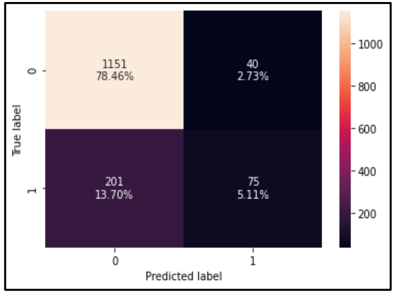
	Bagging classifier with default parameters	Tuned Bagging Classifier	Bagging classifier with base_estimator=LR	Decision Tree with default parameters	Tuned Decision Tree Classifier	Random Forest with deafult parameters	Tuned Random Forest Classifier	Random Forest with class_weights
Accuracy	0.903204	0.822086	0.822086	0.871847	0.838446	0.893661	0.913429	0.912747
Recall	0.565217	0.072464	0.072464	0.652174	0.503623	0.489130	0.608696	0.590580
Precision	0.876404	0.800000	0.800000	0.661765	0.581590	0.900000	0.898396	0.915730
F1	0.687225	0.132890	0.132890	0.656934	0.539806	0.633803	0.725702	0.718062

Conclusion:

- We can see that three variables - Passport, Age, and Monthly Income are the most important factors in identifying persons who are likely to buy a travel package. Other variables' importance is not as significant.
- Identifying what features influence customers to purchase travel packages will help identify and focus sales to certain demographics of customers.
- This will help sales reduce wasted time pitching to customers who are less likely to purchase a travel package.
- As per the decision tree business rules:
- Customers who do not have a passport and who are not single and who live in a city tier >1 and who prefer property star >4 and are age > 43 are more likely to buy a travel package.
- Customers who do not have a passport who are single and are age <32 who own large business are more likely to buy a travel package.
- Customers with a passport who are executives are likely to buy a travel package regardless of their age
- Based on the above analysis, we can say that:
- Middle-aged to older customers who are looking to stay at a higher rated property are more likely to purchase a travel package.
- Younger single customers without a passport find travel packages more appealing and are more likely to purchase one if they own a large business.
- Customers who are executives and own a passport are highly likely to buy a travel package.

Boosting Classifier

AdaBoost Classifier

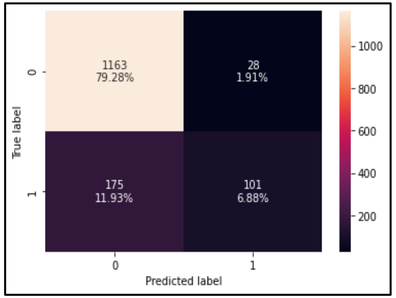


	Accuracy	Recall	Precision	F1
0	0.844782	0.27795	0.730612	0.4027

	Accuracy	Recall	Precision	F1
0	0.835719	0.271739	0.652174	0.383632

- Adaboost is giving more generalized performance than previous models but the test f1-score is too low.

Gradient Boosting Classifier



Training performance:

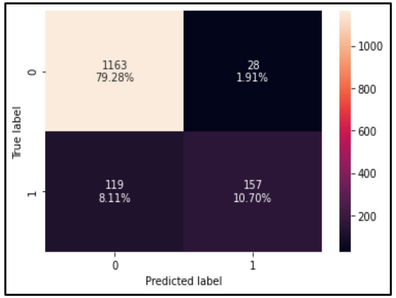
	Accuracy	Recall	Precision	F1
0	0.879567	0.43323	0.855828	0.575258

Testing performance:

	Accuracy	Recall	Precision	F1
0	0.861622	0.365942	0.782946	0.498765

- The gradient boosting classifier is overfitting the training data slightly.
- F1 scores are still low.

XGBoost Classifier



Training performance:

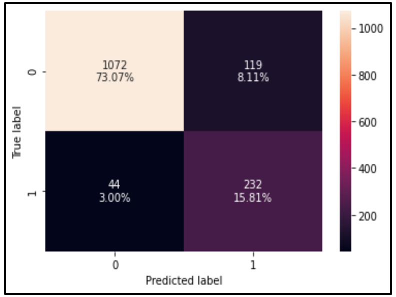
	Accuracy	Recall	Precision	F1
0	0.997077	0.984472	1.0	0.992175

Testing performance:

	Accuracy	Recall	Precision	F1
0	0.899796	0.568841	0.848649	0.681128

•Xgboost classifier is slightly overfitting the training data.
•We will try hyperparameter tuning and see if the model performance improves.

Stacking Classifier



Training performance:

	Accuracy	Recall	Precision	F1
0	0.997077	1.0	0.984709	0.992296

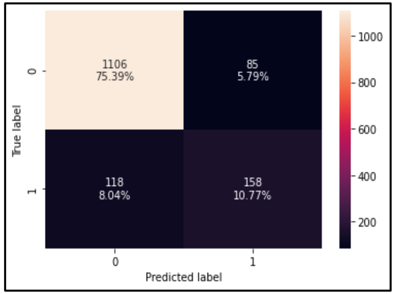
Testing performance:

	Accuracy	Recall	Precision	F1
0	0.888889	0.84058	0.660969	0.740032

- The stacking classifier is giving a similar performance as compared to XGBoost with slightly less overfitting.

Hypertuning the models

AdaBoost - Tuned

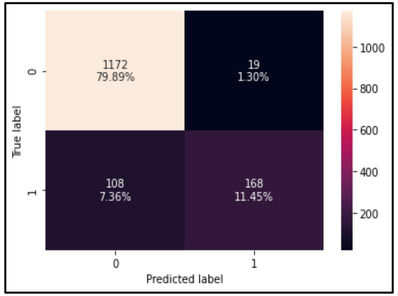


	Accuracy	Recall	Precision	F1
0	0.975738	0.903727	0.965174	0.93344

	Accuracy	Recall	Precision	F1
0	0.861622	0.572464	0.650206	0.608863

- The model performance has increased slightly but the model has started to overfit the training data.

Gradient Boosting - Tuned



Training performance:

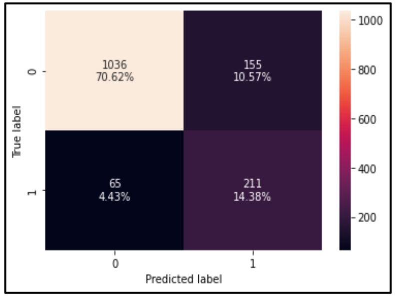
	Accuracy	Recall	Precision	F1
0	0.879567	0.43323	0.855828	0.575258

Testing performance:

	Accuracy	Recall	Precision	F1
0	0.861622	0.365942	0.782946	0.498765

- The gradient boosting classifier is overfitting the training data slightly.
- F1 scores are still low.

XGBoost - Tuned



Training performance:

	Accuracy	Recall	Precision	F1
0	0.940953	0.968944	0.774194	0.86069

Testing performance:

	Accuracy	Recall	Precision	F1
0	0.850034	0.764493	0.576503	0.657321

•The overfitting has reduced slightly.
•Testing recall has increased while testing precision has decreased.

Comparing all the models

Train Comparison Table

	Decision Tree	Decision Tree Estimator	Random Forest Estimator	Random Forest Tuned	Bagging Classifier	Bagging Estimator Tuned	Adaboost Classifier	Adabosst Classifier Tuned	Gradient Boost Classifier	Gradient Boost Classifier Tuned	XGBoost Classifier	XGBoost Classifier Tuned	Stacking Classifier
Accuracy	1.0	0.824320	1.0	1.0	0.994154	0.825490	0.844782	0.975738	0.879567	0.902952	0.997077	0.940953	0.997077
Recall	1.0	0.487578	1.0	1.0	0.970497	0.082298	0.277950	0.903727	0.433230	0.538820	0.984472	0.968944	1.000000
Precision	1.0	0.536752	1.0	1.0	0.998403	0.898305	0.730612	0.965174	0.855828	0.908377	1.000000	0.774194	0.984709
F1	1.0	0.510985	1.0	1.0	0.984252	0.150782	0.402700	0.933440	0.575258	0.676413	0.992175	0.860690	0.992296

Test Comparison Table

	Decision Tree	Decision Tree Estimator	Random Forest Estimator	Random Forest Tuned	Bagging Classifier	Bagging Estimator Tuned	Adaboost Classifier	Adabosst Classifier Tuned	Gradient Boost Classifier	Gradient Boost Classifier Tuned	XGBoost Classifier	XGBoost Classifier Tuned	Stacking Classifier
Accuracy	0.871847	0.838446	0.893661	0.913429	0.903204	0.822086	0.835719	0.861622	0.861622	0.875937	0.899796	0.850034	0.888889
Recall	0.652174	0.503623	0.489130	0.608696	0.565217	0.072464	0.271739	0.572464	0.365942	0.438406	0.568841	0.764493	0.840580
Precision	0.661765	0.581590	0.900000	0.898396	0.876404	0.800000	0.652174	0.650206	0.782946	0.817568	0.848649	0.576503	0.660969
F1	0.656934	0.539806	0.633803	0.725702	0.687225	0.132890	0.383632	0.608863	0.498765	0.570755	0.681128	0.657321	0.740032

- The majority of the models are overfitting the training data.
- The staking classifier is giving the highest f1-score on the test data and is giving more generalized performance.
- The tuned random forest has given the second-highest test f1-score but is overfitting the training data..

Conclusion

After all the analysis, we have been able to conclude:

- The CustomerID attribute does not add any information to our analysis as all the values are unique. There is no association between a person's customer ID and ProdTaken, also it does not provide any general conclusion for future potential travel package customers. We can neglect this information for our model prediction.
- We notice several missing values that need to be treated in 8 different columns. Income, CC Average, and Mortgage had outliers that were treated.
- We can see that three variables - Passport, Age, and Monthly Income are the most important factors in identifying persons who are likely to buy a travel package. Other variables' importance is not as significant.
- Identifying what features influence customers to purchase travel packages will help identify and focus sales to certain demographics of customers.
- This will help sales reduce wasted time pitching to customers who are less likely to purchase a travel package.
- Customers who do not have a passport and who are not single and who live in a city tier >1 and who prefer property star >4 and are age > 43 are more likely to buy a travel package.
- Customers who do not have a passport who are single and are age <32 who own large business are more likely to buy a travel package.
- Customers with a passport who are executives are likely to buy a travel package regardless of their age
- Middle-aged to older customers who are looking to stay at a higher rated property are more likely to purchase a travel package.
- Younger single customers without a passport find travel packages more appealing and are more likely to purchase one if they own a large business.
- Customers who are executives and own a passport are highly likely to buy a travel package.
- The majority of the models are overfitting the training data.
- The staking classifier is giving the highest f1-score on the test data and is giving more generalized performance.
- The tuned random forest has given the second-highest test f1-score but is overfitting the training data.

Recommendations

After all the analysis, we suggest the following recommendations:

- Based on the different models, we would recommend the stacking classifier model as well as the decision tree model for making business decisions.
- We saw our analysis that customers who do not have a passport, not single, in a city tier of >1, prefer a property star of >4 and are Age >43 but <46 are more likely to purchase a travel package.
- Also, customers who do not have a passport who are single, age <=31, and own a large business are more likely to buy a travel package.
- Based on the above two points, if the customer does not have a passport, they should take the age factor and if they are single into consideration.
- If a customer has a passport, the sales team should focus on executive of any age. Based on the decision tree feature importance, they are more likely to purchase a travel package.
- Also unmarried customers who have a passport and are in a city tier >1 are more likely to purchase a travel package.
- It would appear customers with passports are more likely to purchase a travel package.
- As a result, the company should target customers with passports.
- Sales team should also avoid customers who are in city tier 1.

greatlearning

Power Ahead

Happy Learning !