Visit with us

The better way enjoy your world travel adventure

Contents

Business Problem Overview

- The Policy Maker of the company wants to enable and establish a viable business model to **expand the customer base**.
- The company in the last campaign contacted the customers at random without looking at the available information and we observed that **18%** of the customers purchased the packages.
- Currently, there are 5 types of packages the company is offering: Basic, Standard, Deluxe, Super Deluxe, King.
- One of the ways to expand the customer base is to **introduce a new offering** of packages. The company is now planning to launch a new product i.e. *Wellness Tourism Package.*

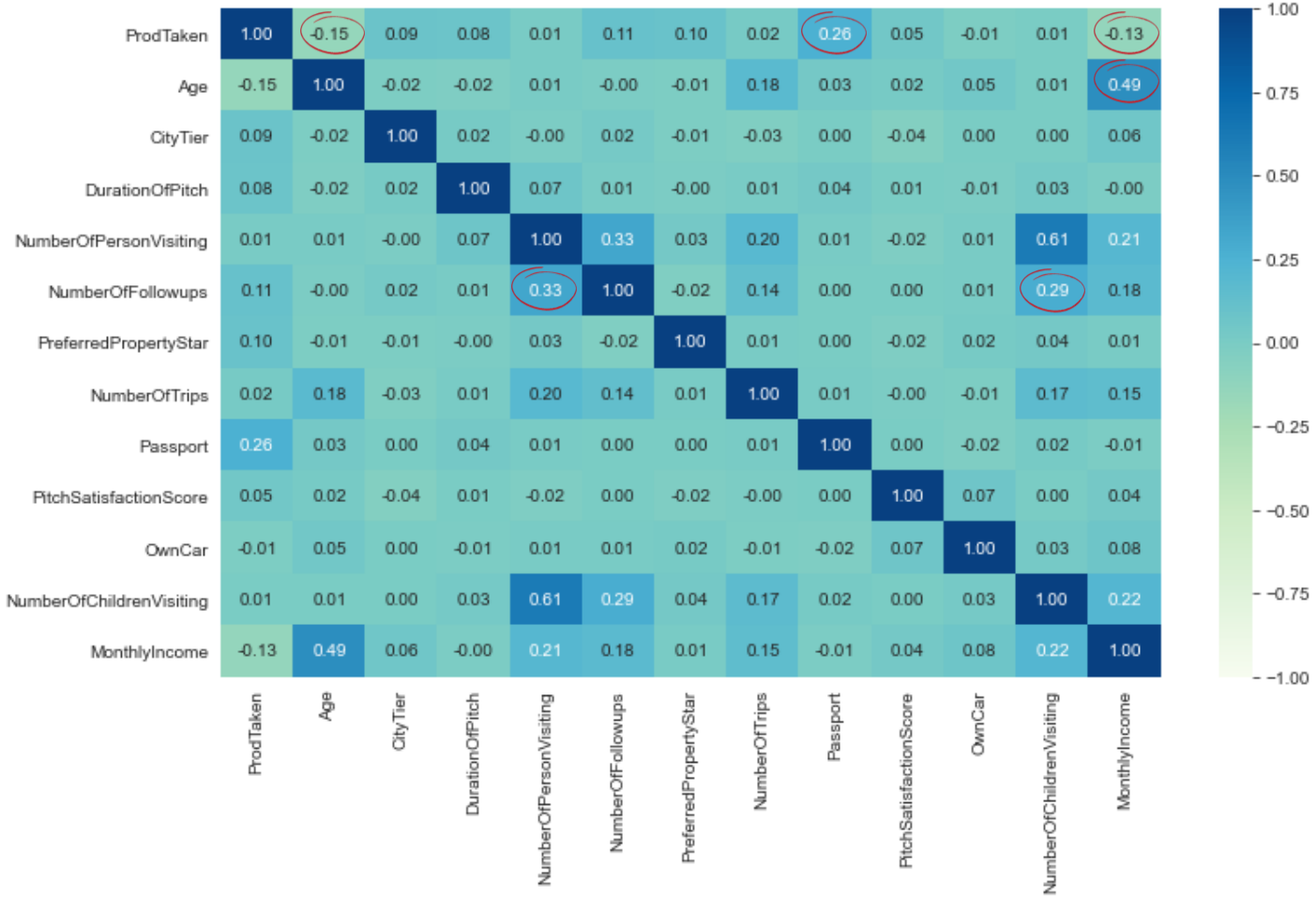
Data Overview

- The data contains information about 4888 customers and their 19 (rows) information as:
  - Designation, Type of Contact, Occupation, Gender, ProdPitched and Marital Status as categorical
  - City Tier (1, 2, 3) , Prefer Property Star (3, 4, 5) and Pitch Satisfaction Score (1, 2, 3, 4, 5) as Ordinal Categorical variables.
  - Binary Categorical variables such as Passport, Own Car and ProdTaken (Our Target Variable).
  - Monthly Income, Age, Number of person visiting, , Number of children visiting, Number of follow-ups and Number of trips as numerical variables

- Duration Of Pitch and Monthly Income had some extreme values that we analyzed and treated it. Number of trip have some high values but it seems real data, so we kept it.

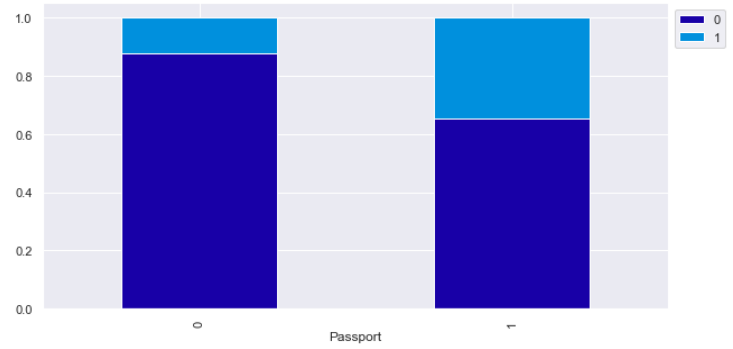
- All missing values we analyzed, trying to find pattern and group by with other information to replace it with the best information.
- We have an imbalanced data set, with 81% of customer as not a buyer and 19% that bought a travel package on the previews campaign.

Exploratory Data Analysis (EDA)

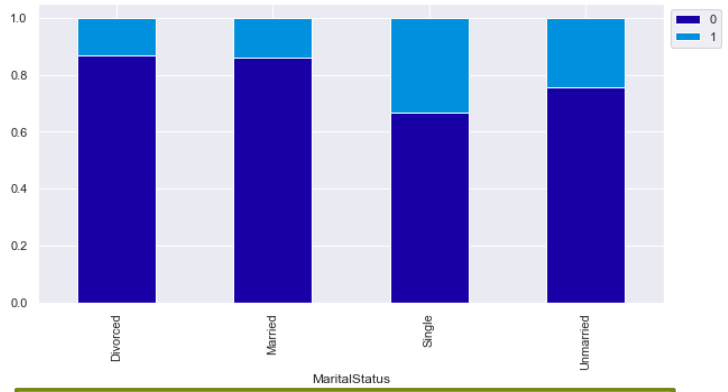


- ProdTaken has a **0.26** correlation with Passport, meaning that customer that have passport are more like hood to buy a travel package.
- ProdTaken has negative correlation with MonthlyIncome **-0. 13** and **-0.15** with Age, meaning that young customers, that usually has less income, are potential buyers.
- Age and MonthlyIncome has **0.49** correlation, what is king of expected.
- Number Of Follow-ups has correlation with Number of Person Visiting **0.33** and Number Of Children Visiting **0.29.**
- OwnCar, PitchSatisfactionScore, PreferredPropertyStar, DurationOfPitch,CityTier, does not show correlation with any of the other's variables.

Exploratory Data Analysis (EDA)

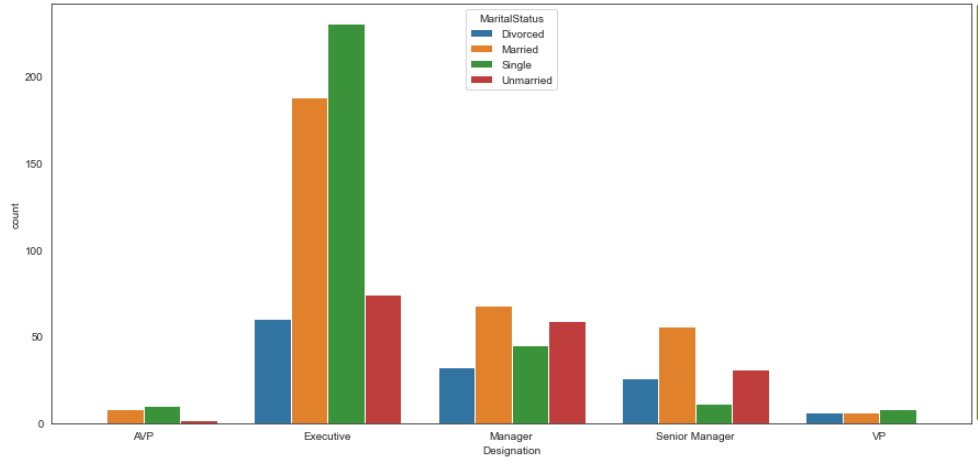


Customer with passport plays a big part in purchasing a travel package.



Single customer followed by Unmarried are our potential buyers.

Those customers who has a Designation as Executive are potential customers



Using the distribution between Marital Status VS Designation and knowing the packages they prefer; company can create packages specifics for the necessity of each profile of customer

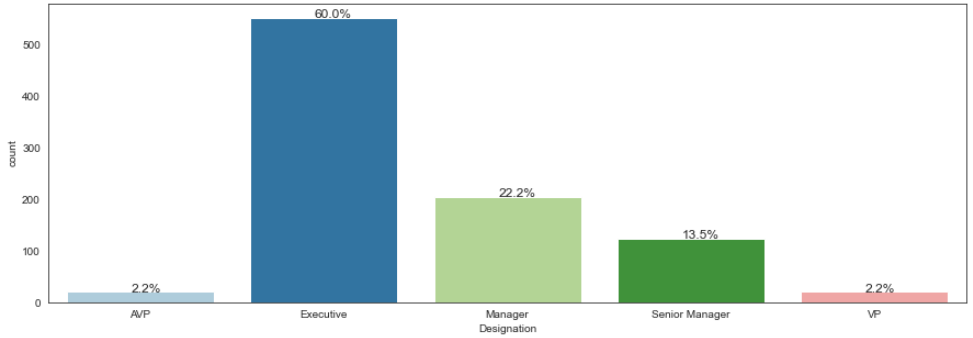
Customer Profile

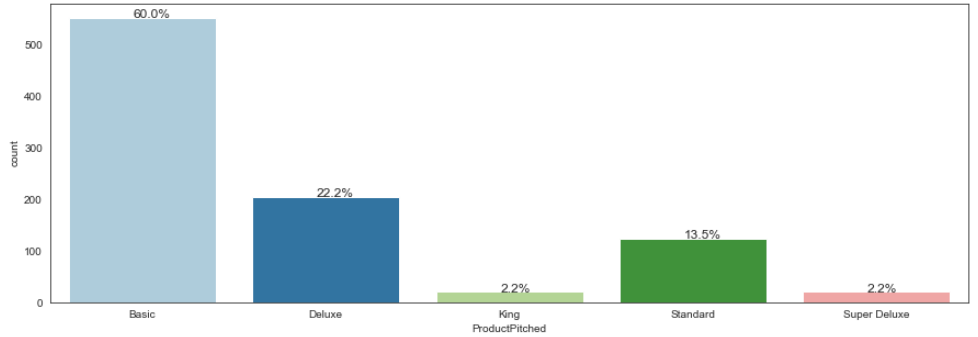
- Analyzing the behavior of previously buyer, we can come across this Profile:

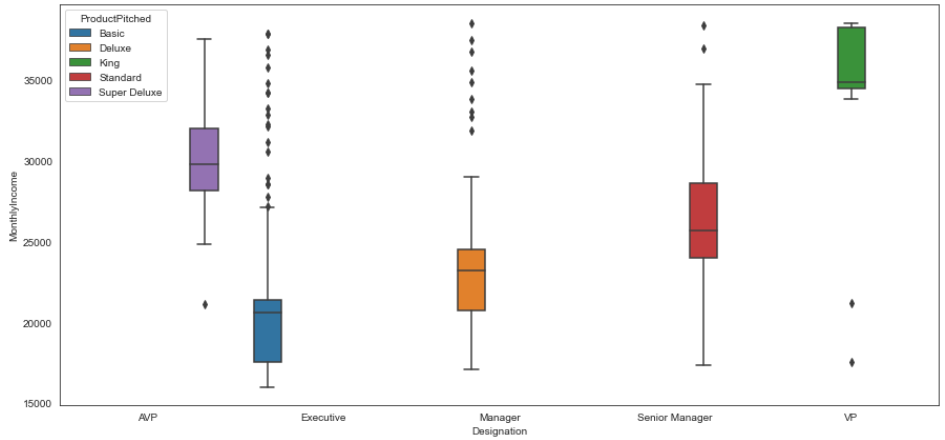
- ProdPitched is proportional a Customer Designation

- 60% of our buyers were **Executives**, buying Basic Travel Package.
  - They have mean Monthly Income around 20,162;
  - Mean of 31-year-old and most of them **single**, followed by married ;
  - 42% of them take 1 children, followed by 28% taking 2 children and 21% taking Zero children.

- 22.2% of customers were **Managers**, they prefer Deluxe Travel Package.
  - They have mean Monthly Income around 23,106;
  - Mean of 37 year old and most of them married, or Unmarried ;
  - 44% of them take 1 children, followed by 27.5% taking 2 children and 22% taking Zero children.

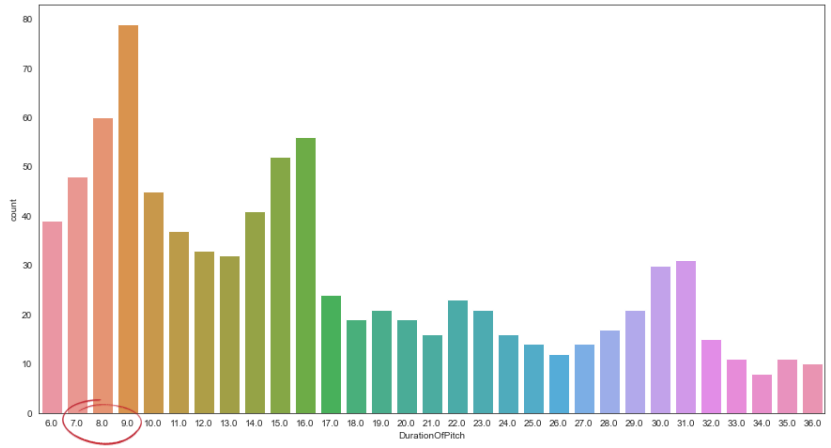




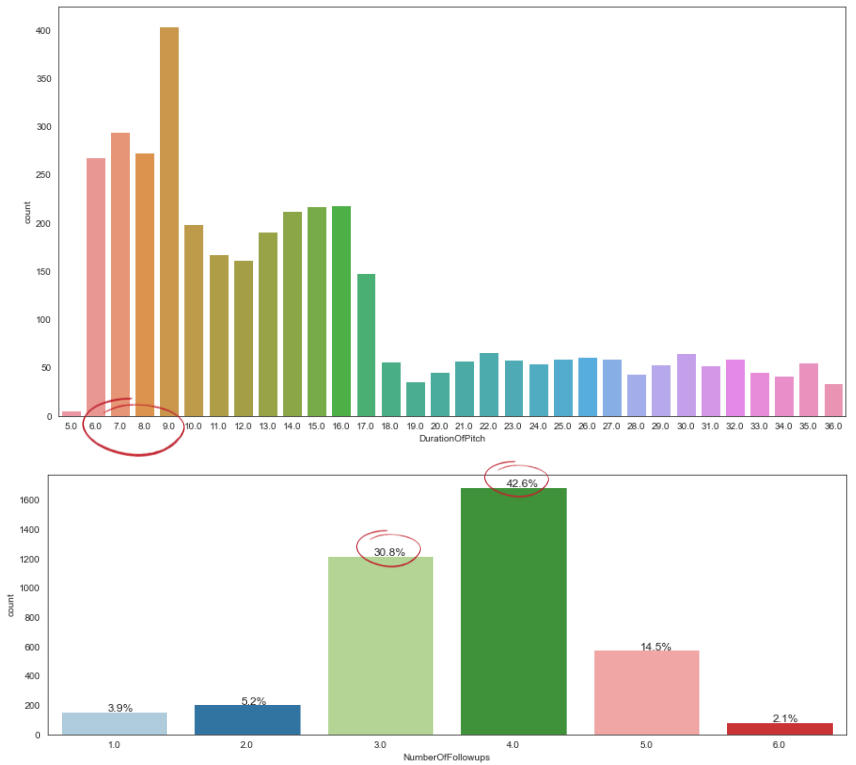


Customer Profile

Customer that bought the travel package:



Customer that DID NOT buy the package:



- Duration Of Pitch of customers who bought or not the travel package is frequently to between 6 to 9.
- This is the time our team have to introduce our product and get the customer to buy it.
- Looking to Number of follow-ups, we can see that 3 times, is where 30.8% of customers say NO, followed by 4 follow-ups where 42.6% say NO.
- That's the key for our team to convert this costumers as a buyers.
- Identify the customer profile and offering them the best package using approach of the time and number of follow-ups

Model Performance Summary

- We want to predict whether a customer will buy a Travel Package or not using their information initially provided to us.
- We will use Recall as the performance metric for our model.
- Predicting that customer will not buy but it would mean losing a potential buyer.

The built model can be used to predict if a customer is going to buy or not a Travel Package and to create a Customer Profile considering the significance of independent variables.

Model Performance Summary

	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision	Train_F1_Score	Test_F1_Score
0	Decision Tree - default parameters	1.000	0.885	1.000	0.714	1.000	0.689	1.000	0.701
1	Decision Tree - Tunned	0.796	0.803	0.648	0.663	0.469	0.483	0.544	0.559
2	Bagging classifier - default parameters	0.994	0.905	0.969	0.587	0.998	0.866	0.983	0.700
3	Bagging Classifier - weighted	0.994	0.898	0.967	0.547	0.998	0.858	0.983	0.668
4	Bagging Classifier - Tunned	0.716	0.714	0.666	0.703	0.362	0.365	0.469	0.480
5	Random Forest - deafult parameters	1.000	0.915	1.000	0.583	1.000	0.942	1.000	0.720
6	Random Forest - weighted	1.000	0.911	1.000	0.547	1.000	0.962	1.000	0.697
7	Random Forest - Tunned	0.903	0.869	0.848	0.674	0.699	0.646	0.766	0.660
8	AdaBoost - default parameters	0.849	0.847	0.331	0.322	0.712	0.701	0.452	0.442
9	AdaBoost - Tunned	0.990	0.864	0.964	0.576	0.981	0.660	0.973	0.615
10	Gradient Boosting - default parameters	0.888	0.868	0.449	0.395	0.909	0.807	0.601	0.530
11	Gradient Boosting - init	0.886	0.866	0.452	0.384	0.890	0.797	0.599	0.518
12	Gradient Boosting - Tunned	0.923	0.880	0.623	0.482	0.948	0.801	0.752	0.602
13	XGB - deafult parameters	1.000	0.926	0.998	0.692	1.000	0.888	0.999	0.778
14	XGB - Tunned	0.994	0.918	0.966	0.638	1.000	0.898	0.983	0.746
15	XGB2 - Tunned other parameters	0.906	0.863	0.935	0.815	0.684	0.600	0.790	0.691
16	XGB_2 - deafult (No treatment Missing values)	1.000	0.928	0.998	0.710	1.000	0.883	0.999	0.787
17	XGB_2 - Tunned (No treatment Missing values)	0.946	0.875	0.980	0.804	0.785	0.632	0.872	0.708
18	XGB2_2 - Tunned other parameters (No treatment...	0.909	0.862	0.929	0.826	0.694	0.595	0.794	0.692
19	stacking_classifier	0.987	0.921	0.950	0.699	0.982	0.854	0.966	0.769

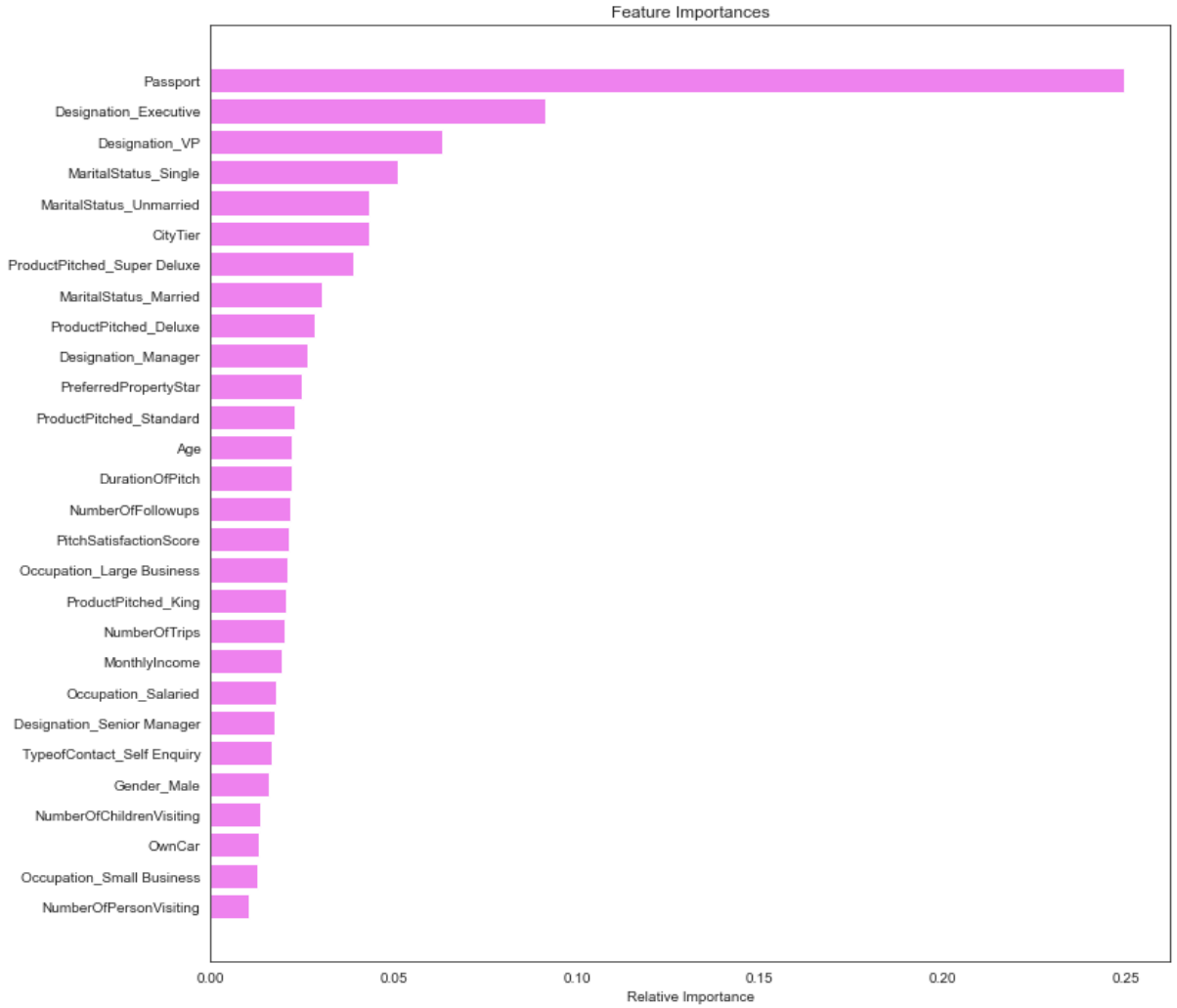
XGB2_2 - Tunned other parameters (No treatment missing values)

is our best model, it reduced the overfitting and is performing better compared to the other models.

Performance on Precision and F1_score is not performing that well. It still need improvement.

Insights we get from our Model

- We have been able to build a predictive model:
  - That the company can deploy to identify customers who will be interested in buy a travel package.
  - That the company can use to find the key factors that will have an impact on converting a customer as a buyer.
- Factors that have an impact: Passport, Designation_Executive, Designation_VP and Marital Status_Single.
- Customer with Passport is more like hood to buy it.
- If customer Designation is Executive higher are the chances for them to buy a travel package.
- Customer with Marital Status Single or Unmarried are also more like hood to buy the travel package.



Business Insights and Recommendations

Insights and Recommendations

Conclusion

- Overall, we can see that the XGBoost performs better on this dataset then other models.
- Considering that XGboost treats missing values, we can see that the model applied on data set without missing values treatment is the one performing better.
- We testes 2 different parameter on them and the second one come out performing a little bit better on Recall and reducing even more the overfitting.
- > XGB2_2 - Tunned other parameters (No treatment missing values) is our best model, it reduced the overfitting and is performing better compared to the other models. Performance on Precision and F1_score is not performing that well. It still need improvement considering that is performing 0.826 on test recall and 0.595 on Precision.

Recommendations

- We should understand the needs of each profile and provide the best travel package thinking in all points that will made the difference and Increase the customer satisfaction.
- Create strategies to best use our 5 to 10 time of pitch with our customer, considering this is our mean time spend with them before they say yes or no.
- Third and Fourth follow-ups is the KEY to convert this customer as a buyer, to do so we need to train our team to do the best presentation of the travel package and identify customer necessities.
- A new travel package BUSINESS should be considering, to target companies that have their employees traveling a lot.
- King and Super Delux need to be reviewed and improved to target more buyers.


Thank you!
Amanda Mendonca
PASSPORT
United States of America
Kodak
Greenland Sea
Barents Sea
Kara Sea
Laptev Sea
GREENLAND
(Denmark)
Labrador Sea
Davis Strait
SWEDEN
FINLAND
RUSSIA
EUROPE
KAZAKHSTAN
MONGOLIA
ASIA
CHINA
ATLANTIC OCEAN
Black Sea
Caspian Sea
UZBEKISTAN
TURKMENISTAN
AFGHANISTAN
IRAN
IRAQ
SYRIA
TURKEY
SAUDI ARABIA
EGYPT
LIBYA
ALGERIA
PAKISTAN
INDIA
MYANMAR
THAILAND
MALI
NIGER
CHAD
SUDAN
YEMEN
Arabian Sea
Bay of Bengal
Red Sea
Moscow
St. Petersburg
Kiev
Minsk
Warsaw
Budapest
Bucharest
Istanbul
Ankara
Athens
Tehran
Baghdad
Damascus
Cairo
Alexandria
Riyadh
Karachi
Delhi
Mumbai (Bombay)
Chennai (Madras)
Bangkok
Beijing
Shanghai
Tianjin
Seoul
Taipei
Hong Kong
Ho Chi Minh City
Manila