Hackathon Presentation

Will They Claim It

Nishant Pandey

Problem Statement

- Forecast whether the client would Claim the Travel Insurance. This would help in charging competitive premiums , thus helping the company to mitigate risks.

Why solve this problem?

- Business Impact
  - Improve Prediction -> Identify common features of Customers Claiming the insurance -> risk mitigation
  - Improve Prediction -> Identify sectors/ areas for targeted marketing
  - Improve Prediction -> Identify the probability of Claim being made -> Charge competitive premium and stay ahead of the competition
- Stakeholders
  - Chief Financial Officer
  - Chief Marketing Officer

Data

Dataset Information:

The dataset consists of data corresponding to 52310 customers with 11 features. There are 10 predictors and 1 target that describes whether the customer has claimed the insurance or not.

Below are some of the features and target variable:

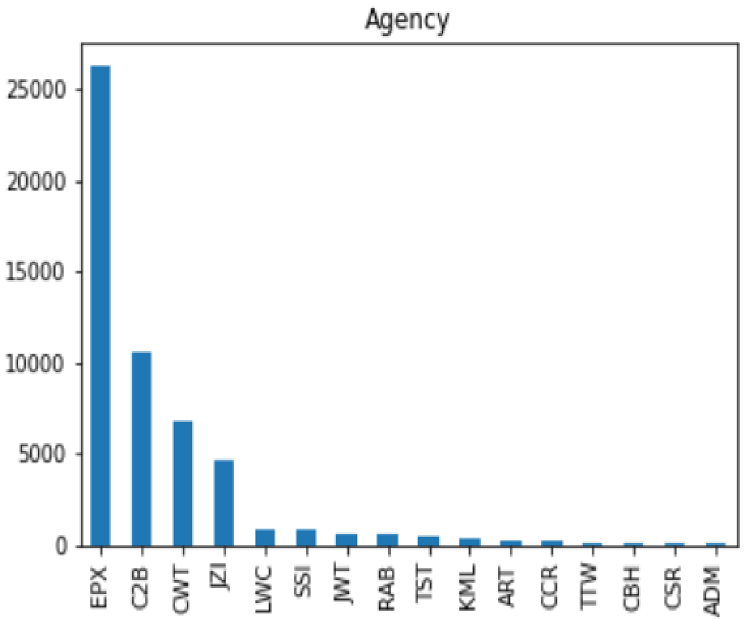
Feature	**Feature Type**	**Description**
ID	numeric	The identification record of every observation
Agency	categorical, nominal	Name of agency
Agency Type	categorical, nominal	Type of travel insurance agencies
Distribution Channel	categorical, nominal	Distribution channel of travel insurance agencies
Product Name	categorical, nominal	Name of the travel insurance products
Duration	numeric	Duration of travel
Destination	categorical, nominal	Destination of travel
Net Sales	numeric	Amount of sales of travel insurance policies
Commision (in value)	numeric	The commission received for travel insurance agency
Age	numeric	Age of insured

Feature	**Feature Type**	**Description**
Claim	Binary	Has the customer made the claim? (0,1)

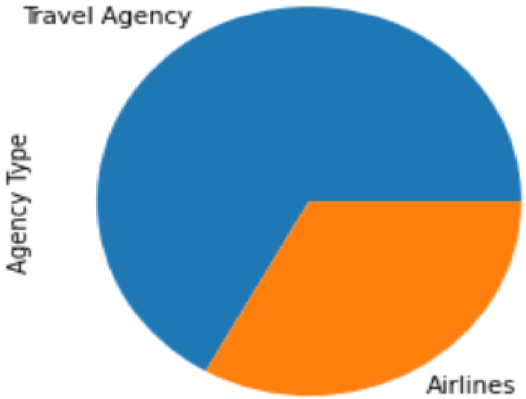
Evaluation Metric

- The evaluation metric for this project is precision score.
- The precision is intuitively the ability of the classifier not to label as positive a sample that is negative.
- False Positive - predicted Claimed the Insurance, but actually not Claimed.
- False Negative - predicted not Claimed the Insurance, but actually Claimed.
- For the use case, False Positive must be reduced. So Precision to be given more importance

EDA – Univariate Analysis

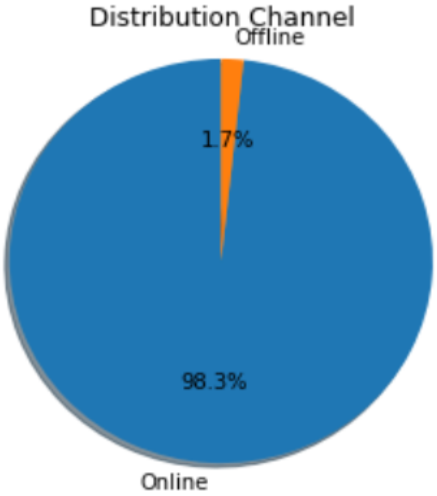


Agency : Majority of the insurance sold by a particular agency, roping them would help.

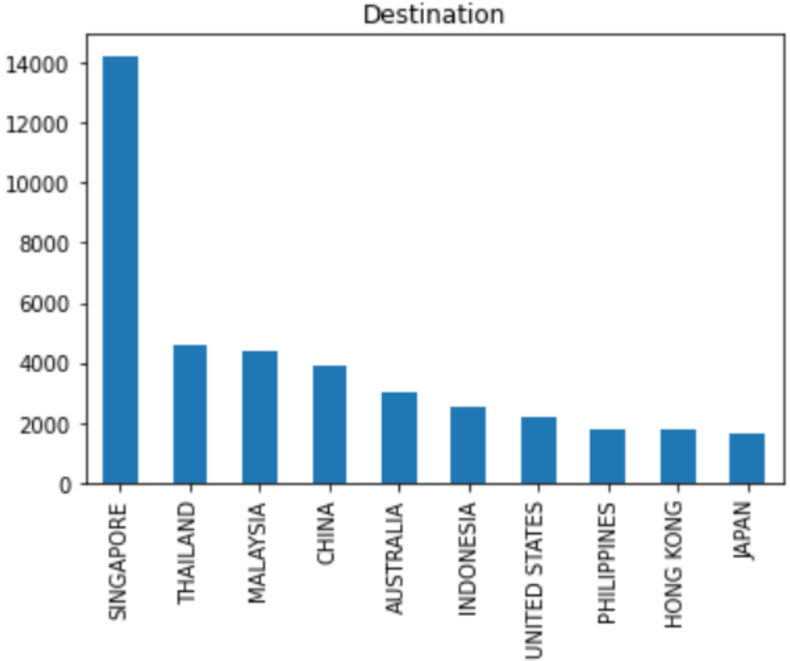


Agency Type : Surely Agencies are more aggressive in marketing than the Airlines

EDA – Univariate Analysis Cont.

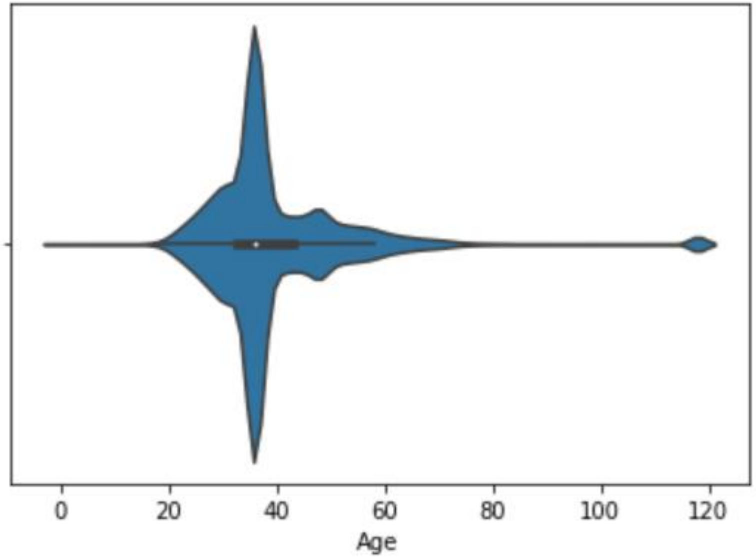


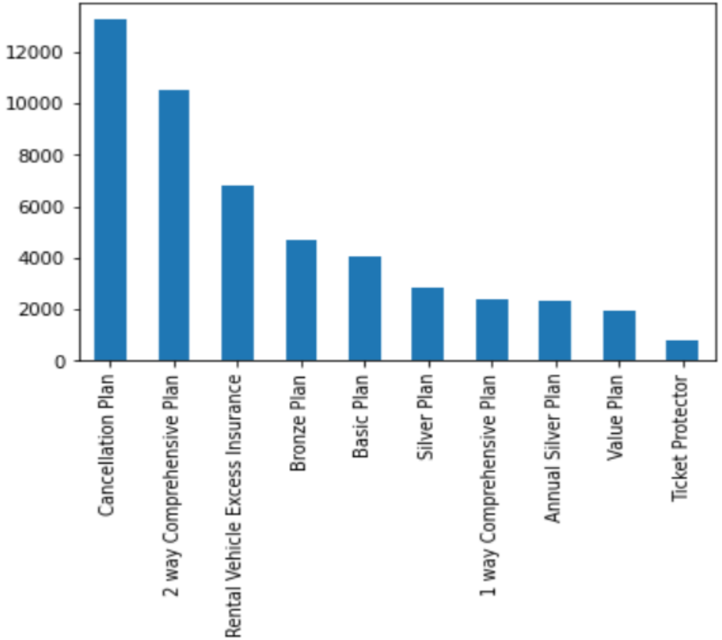
Distribution Channel : Increased internet penetration among the masses has decimated the offline mode. Better to go online.



Destination : Singapore has headquarters of large business and is also a preferred tourist destination.

EDA – Univariate Analysis

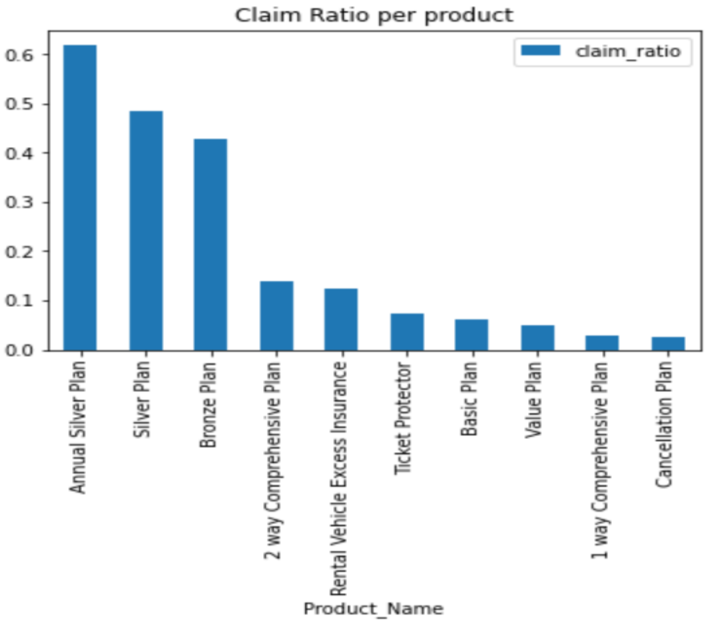


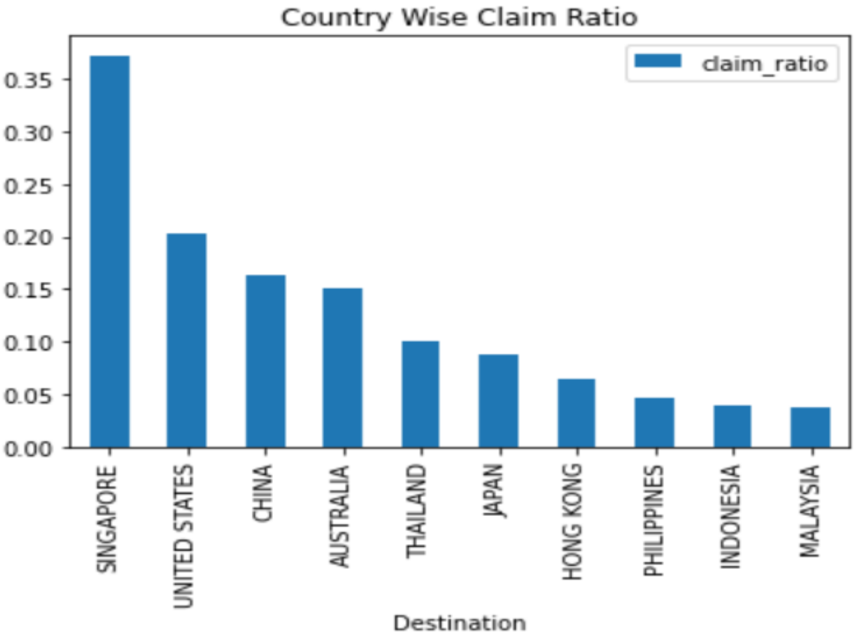


Age : Majority of travelers are in the age bracket 35 – 40, working population

Product Name : Lot of travelers purchase insurance for visa formality

EDA – Bivariate Analysis

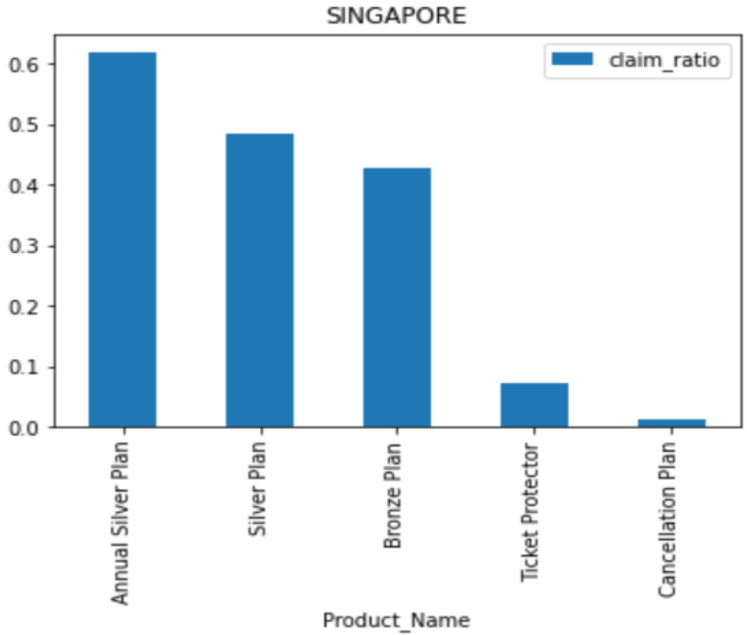




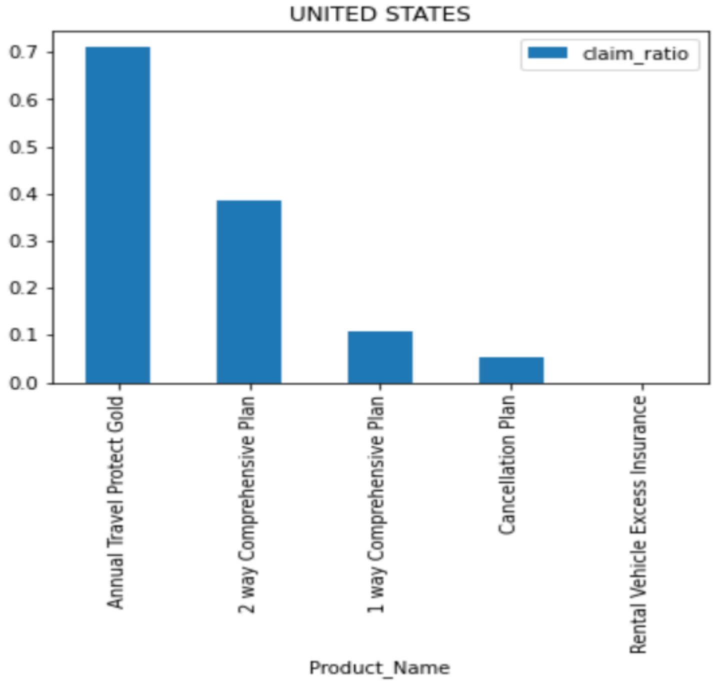
Lot of frequent traveler and purchase comprehensive package

Singapore seeing the max travelers also has highest claim ratio

EDA

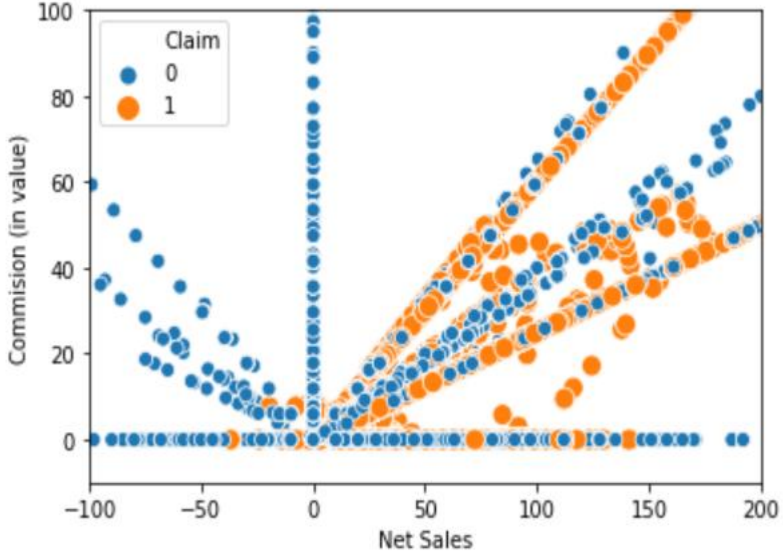


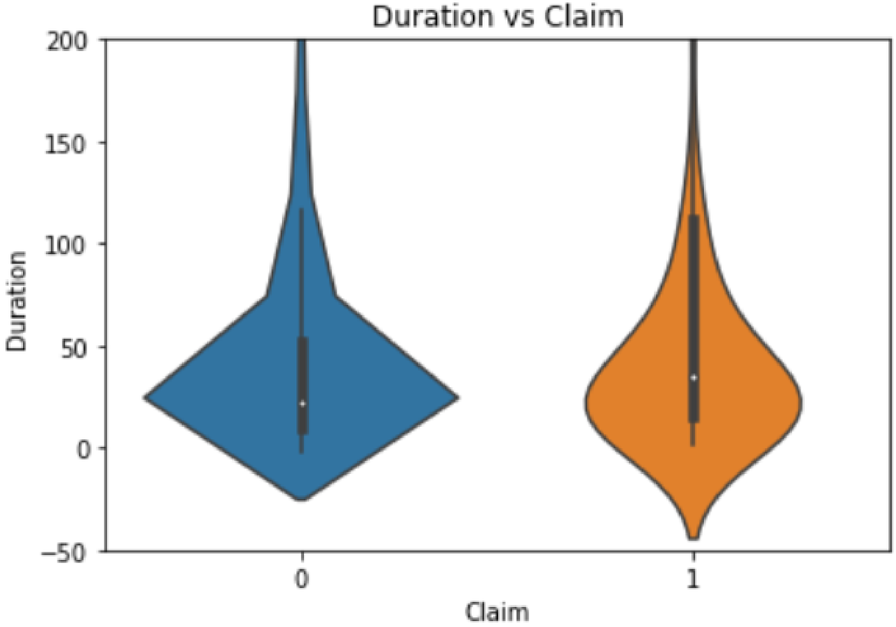
Travelers to Singapore are very Frequent and so Annual Silver Plan



Travelers to USA are less frequent and purchase comprehensive plans

EDA





Majority of Claim occurs for policies with Commission and Net Sales greater than zero

Max claim occurs for stay around 25 days , which is justified as the median length of travel duration is 24

Pipeline

Outlier Treatment:

Outliers in the continuous features(Age, Sales) were detected but no action was performed on them to maintain the variance in the data set.

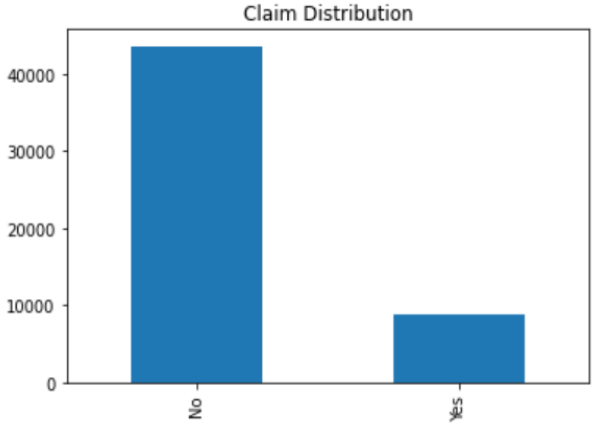
Missing Values:

There were no missing values in both the continuous and categorical features.

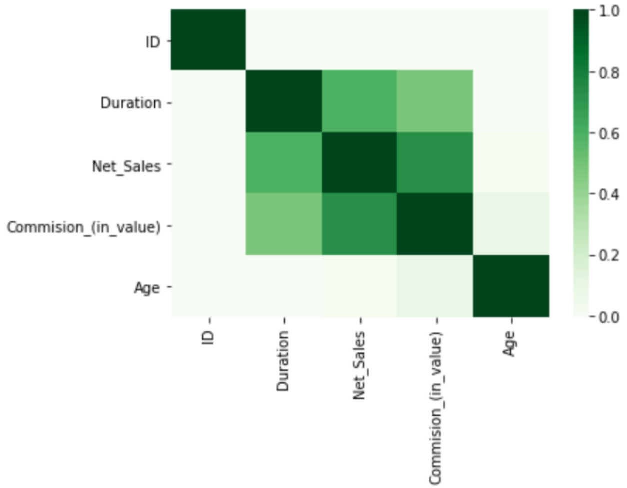
Pipeline

Class Imbalance

The distribution of target below shows a clear imbalance of the two classes.



Pipeline



Feature Selection

Following methods were used for Feature Selection:

- Correlation

- After estimating the Pearson Correlation coefficients between continuous features , following features was dropped
  - Commision(in value)

Models and Approaches

Following models were assessed without performing any hyperparameter tuning and without treatment of class imbalance of the target. The models were fitted to train dataset without performing feature selection. The models are:

- Logistic Regression
- Random Forest Classifier
- XGBoost Classifier
- Support Vector Machines
- KNN classifier
- Stacking Classifier

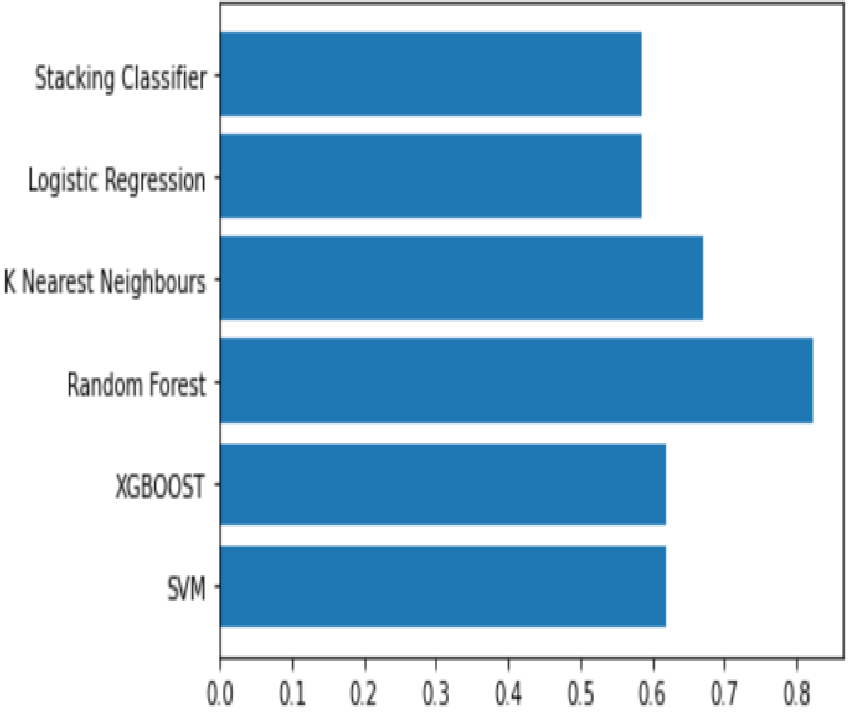
Out of the above models , Random Forest Classifier gave the best Precision Score of 82 %.

To further increase the score , hyperparameter tuning was performed using Grid Search .

Model and Approaches

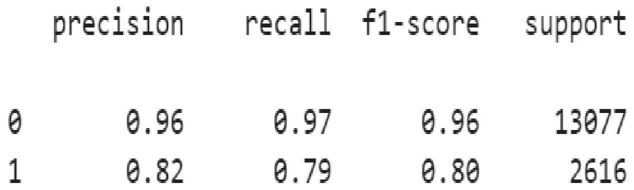
- The Vanilla model used yielded the following results:

Modelling Method	Precision Score
SVM	0.618
XGBoost	0.618
Random Forest	0.822
K Nearest Neighbors	0.671
Logistic Regression	0.585
Stacking Classifier	0.585



Model Tuning

Selected Random Forest Classifier on the basis of the precision score. Performed hyperparameter tuning using Grid Search on Random Forest Classifier, the following results were observed using all the features.

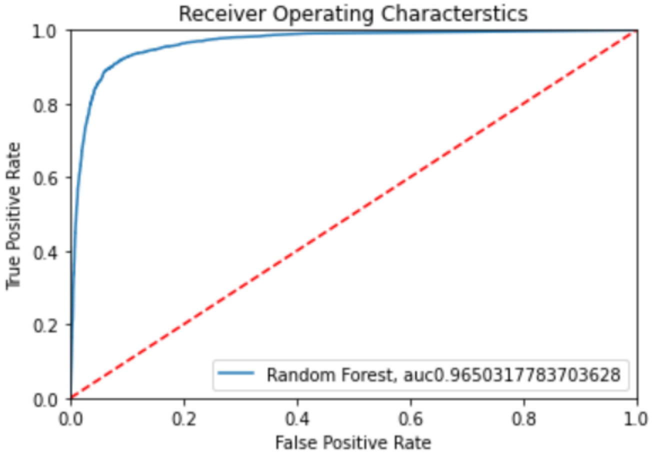


```
        precision    recall  f1-score   support

0          0.96      0.97      0.96     13077
1          0.82      0.79      0.80      2616
```

Evaluation & Results

Below is the AUC_ROC plots after the hyperparameter tuning



Final Results

- From the above observations and plotting it can be inferred that the best performing model was Random Forest Classifier giving an AUC_ROC score of 87.6%
- Confusion Matrix

	Predicted Negative	Predicted Positive
Actual Negative	12627	450
Actual Positive	555	2061

Insights & Decisions

Competitive pricing for certain sectors such as :

- Age : 35-40
- Destination : Singapore, Thailand, Malaysia
- Agency : EPX, C2B
- Plan : Cancellation Plan, 2 Way Comprehensive Plan

To mitigate risk , premium pricing should be done for sectors which have high claim ratio such as :

- Plan : Annual Silver Plan, Silver Plan , Bronze Plan
- Destination : Singapore, USA, China
- Destination wise Plan :
    - Singapore : Annual Silver Plan
    - USA : Annual Travel Protect Plan
    - China : 2 Way Comprehensive Plan

Next Steps

If time permitted, could have tried the following :

- Better feature engineering
- An ensemble of different models