Problem Statement

1. Predict whether to sanction the insurance or not?
2. Issuing of Credit Card/Loan ?
3. Identifying the fraud agency

Why solve this problem?

- Minimize financial loss of the company
- Improvement of company reputation in market
- Identifying the product & destination beneficial for the company while issuing the travel insurance
- Stakeholders - Risk Management, Underwriting and Product development teams.

Data

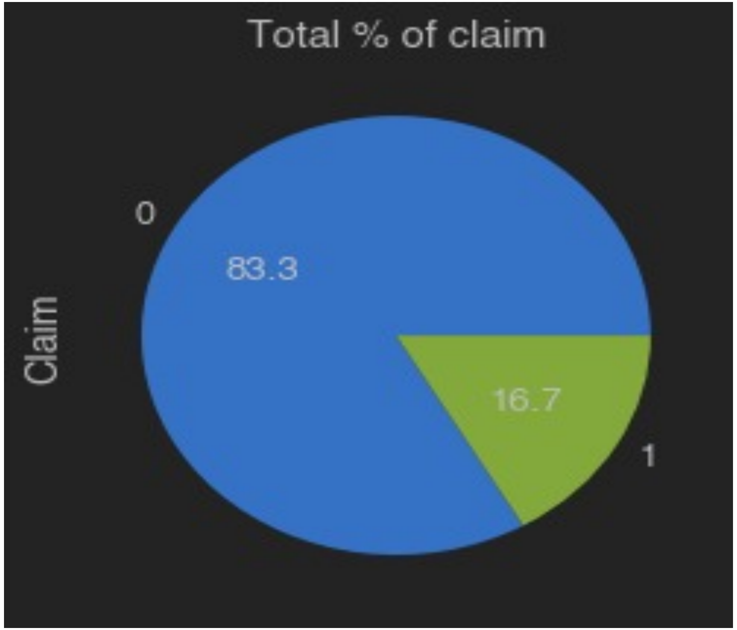
- **Dataset Info**: The data consists of roughly 52310 records and 11 features with 10 predictors and 1 target that describes whether the customer will claim or not.

- **data would like to have:**
  - Date of travel

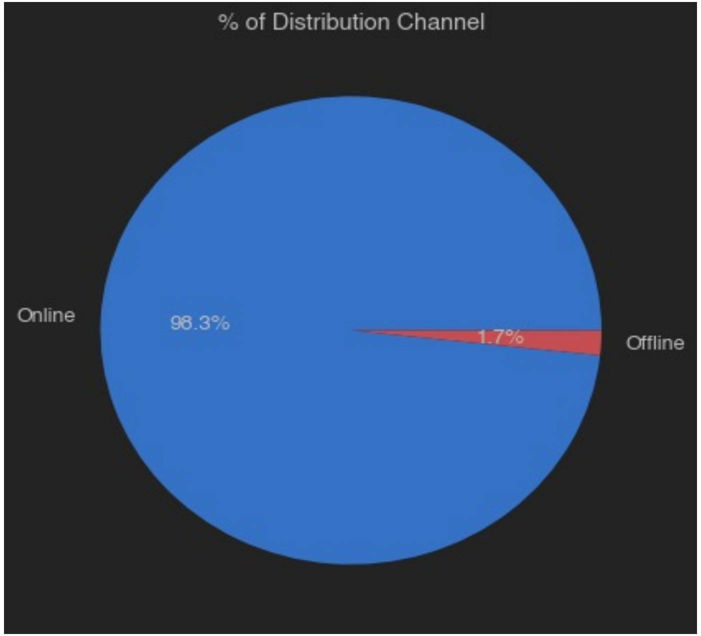
Feature	Feature Type	Description
ID	numeric	ID of customer
Agency	categorical	Name of agency
Agency Type	categorical	Type of travel insurance agencies
Distribution Channel	categorical	Distribution channel of travel insurance agencies
Product Name	categorical	Name of the travel insurance products
Duration	numeric	Duration of travel
Destination	categorical	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
Claim	Binary	Claim Status

Exploratory Data Analysis

Claim %

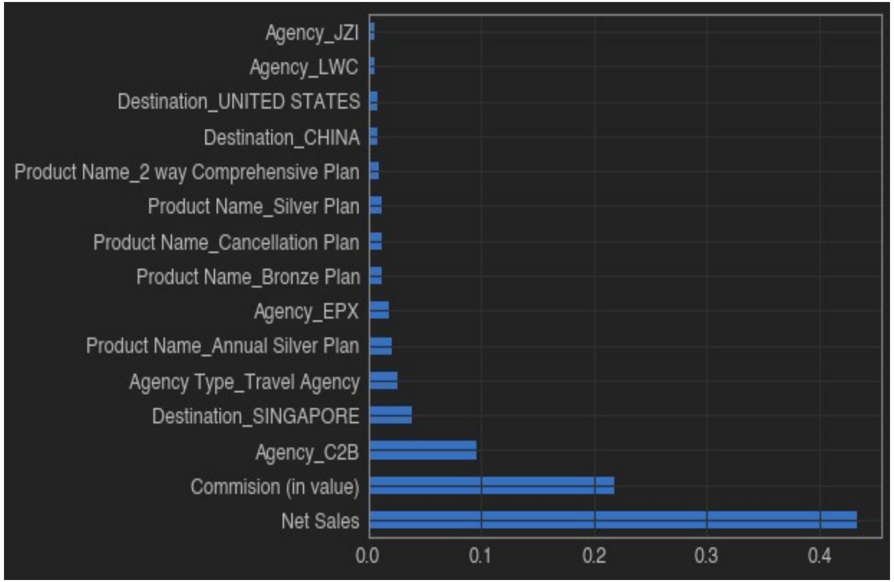


Distribution Channel %

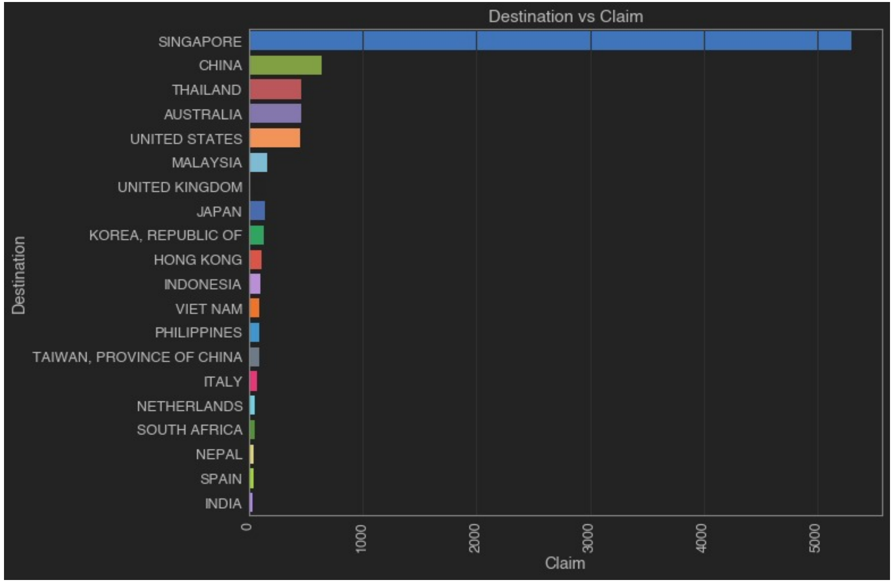


Feature Selection - Exploratory Data Analysis

Top 15 Best Features

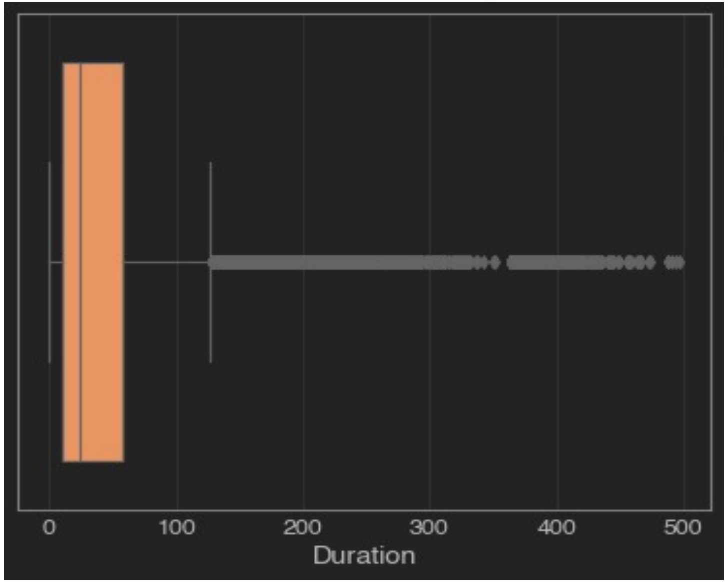


Claim For Top 15 Countries

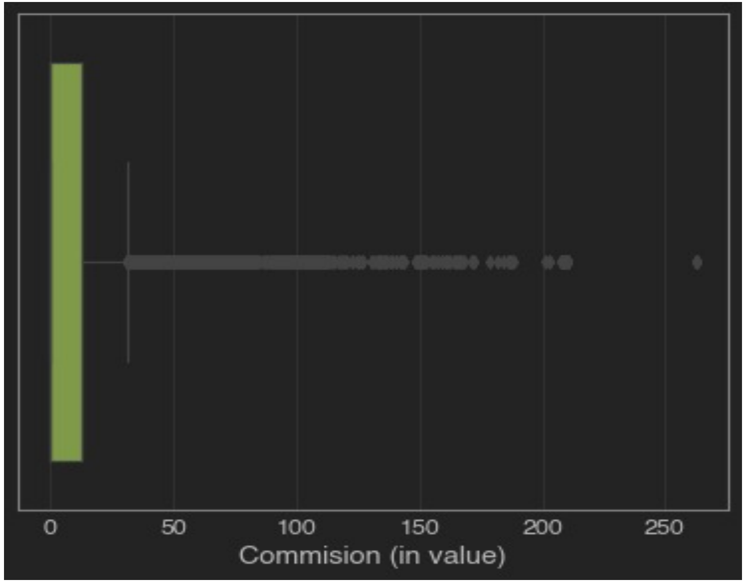


Exploratory Data Analysis

Outliers - Duration

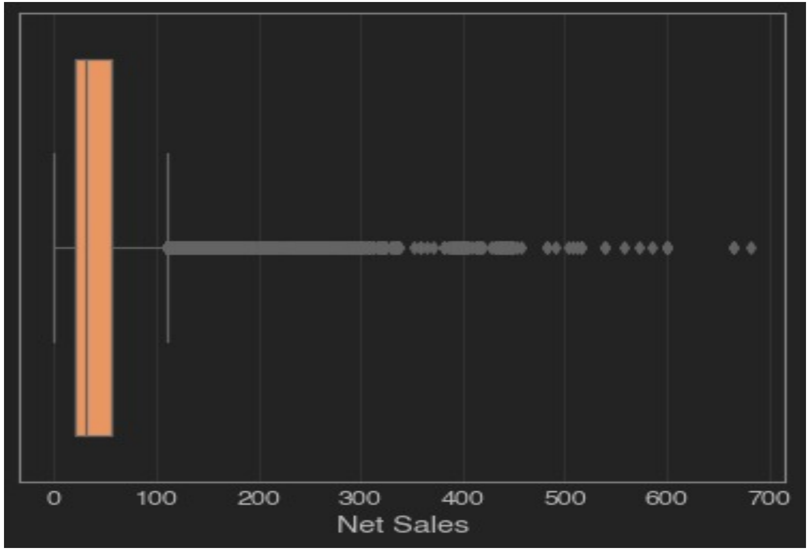


Outliers – Commission (in value)

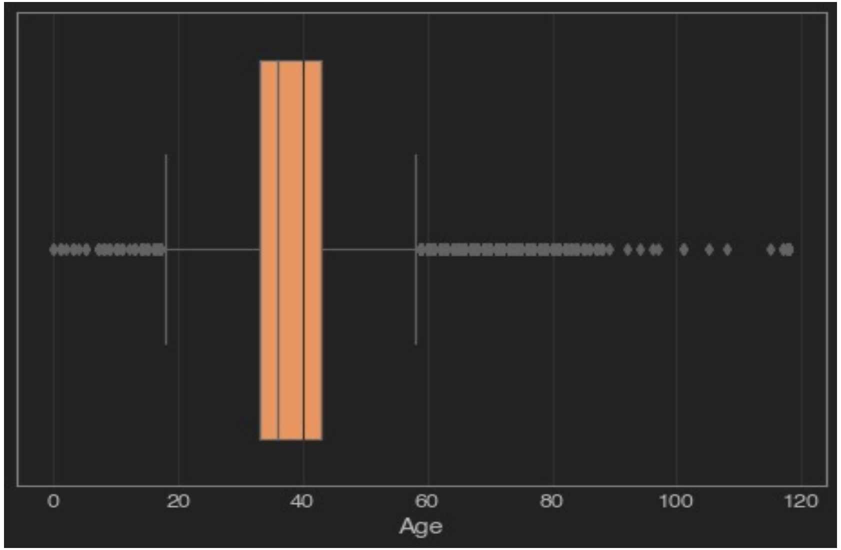


Exploratory Data Analysis

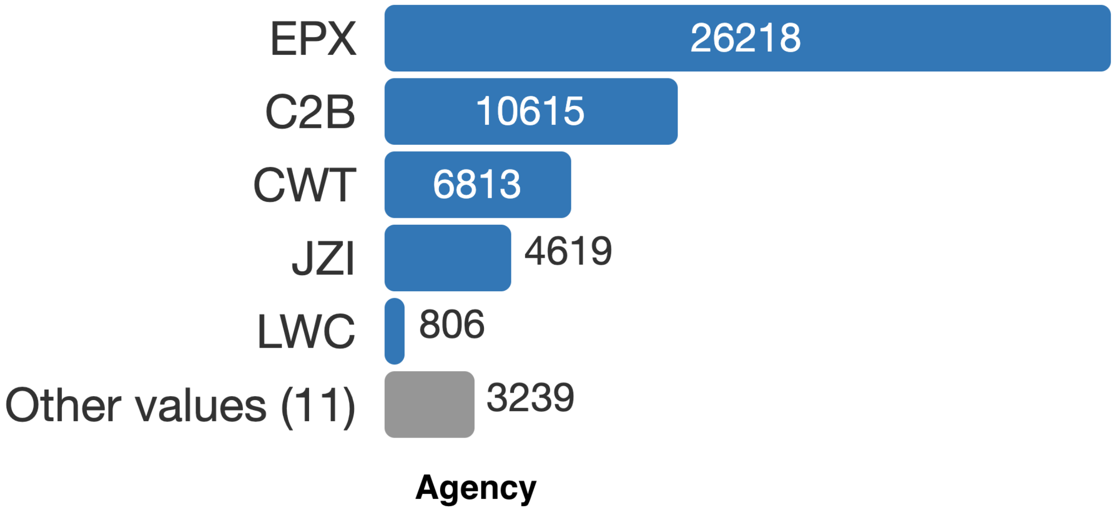
Outliers – Net Sales



Outliers - Age



EDA



EPX
26218
C2B
10615
CWT
6813
JZI
4619
LWC
806
Other values (11)
3239
Agency

Features Correlation Matrix

	Net Sales	Commision (in value)	Agency_C2B	Destination_SINGAPORE	Agency Type_Travel Agency	Product Name_Annual Silver Plan	Agency_EPX	Product Name_Bronze Plan	Product Name_Cancellation Plan	Product Name_Silver Plan	Product Name_2 way Comprehensive Plan	Destination_CHINA	Destination_UNITED STATES	Agency_LWC	Agency_JZI	Claim
Net Sales	1	0.8	0.4	0.3	-0.2	0.7	-0.3	-0.08	-0.2	0.04	-0.04	-0.06	0.1	0.2	-0.09	0.3
Commision (in value)	0.8	1	0.3	0.2	-0.2	0.4	-0.5	-0.04	-0.3	0.03	-0.3	-0.08	0.1	0.4	-0.005	0.3
Agency_C2B	0.4	0.3	1	0.8	-0.7	0.4	-0.5	0.6	-0.3	0.5	-0.3	-0.1	-0.1	-0.06	-0.2	0.4
Destination_SINGAPORE	0.3	0.2	0.8	1	-0.6	0.4	-0.4	0.5	-0.2	0.4	-0.3	-0.2	-0.1	-0.08	-0.2	0.3
Agency Type_Travel Agency	-0.2	-0.2	-0.7	-0.6	1	-0.3	0.7	-0.4	0.4	-0.3	0.4	0.06	0.1	0.09	-0.4	-0.3
Product Name_Annual Silver Plan	0.7	0.4	0.4	0.4	-0.3	1	-0.2	-0.07	-0.1	-0.05	-0.1	-0.06	-0.05	-0.03	-0.07	0.3
Agency_EPX	-0.3	-0.5	-0.5	-0.4	0.7	-0.2	1	-0.3	0.6	-0.2	0.5	0.1	0.08	-0.1	-0.3	-0.3
Product Name_Bronze Plan	-0.08	-0.04	0.6	0.5	-0.4	-0.07	-0.3	1	-0.2	-0.08	-0.2	-0.09	-0.07	-0.04	-0.1	0.2
Product Name_Cancellation Plan	-0.2	-0.3	-0.3	-0.2	0.4	-0.1	0.6	-0.2	1	-0.1	-0.3	0.03	0.04	-0.07	-0.2	-0.2
Product Name_Silver Plan	0.04	0.03	0.5	0.4	-0.3	-0.05	-0.2	-0.08	-0.1	1	-0.1	-0.07	-0.05	-0.03	-0.07	0.2
Product Name_2 way Comprehensive Plan	-0.04	-0.3	-0.3	-0.3	0.4	-0.1	0.5	-0.2	-0.3	-0.1	1	0.1	0.05	-0.06	-0.2	-0.04
Destination_CHINA	-0.06	-0.08	-0.1	-0.2	0.06	-0.06	0.1	-0.09	0.03	-0.07	0.1	1	-0.06	0.02	0.1	-0.002
Destination_UNITED STATES	0.1	0.1	-0.1	-0.1	0.1	-0.05	0.08	-0.07	0.04	-0.05	0.05	-0.06	1	0.2	-0.04	0.02
Agency_LWC	0.2	0.4	-0.06	-0.08	0.09	-0.03	-0.1	-0.04	-0.07	-0.03	-0.06	0.02	0.2	1	-0.04	0.09
Agency_JZI	-0.09	-0.005	-0.2	-0.2	-0.4	-0.07	-0.3	-0.1	-0.2	-0.07	-0.2	0.1	-0.04	-0.04	1	-0.09
Claim	0.3	0.3	0.4	0.3	-0.3	0.3	-0.3	0.2	-0.2	0.2	-0.04	-0.002	0.02	0.09	-0.09	1

1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00

Models and Approaches

Model Name	Precision
Random Forest Classifier	0.7752
Decision Tree Classifier	0.675
SGD Classifier	0.6534
Kneighbors Classifier	0.4729

Model Tuning

- **Model Parameters Used :**

  RandomForestClassifier(

  n_estimators = 60,

  max_depth = 40,

  class_weight = {0:30,1:1}

)

- **Feature selection/ Feature engineering :**

  We have used ExtraTreeClassifier for best feature selection

- **Precision score after tuning RFC: 0.9779**

Evaluation & Results

- **Evaluation matrix used :**

  Precision Score

- **Why precision score used ? :**

  Precision: It is implied as the measure of the correctly identified positive cases from all the predicted positive cases. Thus, it is useful when the costs of False Positives is high.

  This will be helpful for not to loose out genuine customers and minimize the loss

- **Validation result Vs Test result:**

  Precision Score for train is : 0.99979983

  Precision Score for test is :   0. 97791164

Final Results

From the above observations and plotting it can be inferred that the best performing model was Random Forest Classifier giving a precision score of 0.9779

Confusion Matrix :

	Predicted Positive	Predicted Negative
Actual Positive	449 (TP)	387 (FP)
Actual Negative	9 (FN)	4386 (TN)

Classification Report:

	precision	recall	f1-score	support
0	0.92	1.00	0.96	4395
1	0.98	0.54	0.70	836
accuracy			0.92	5231

Insights & Decisions

- Few countries like Singapore, China, Thailand are having very high claim ratios

- Age groups below 14 and above 90 are having less claim ratios

- Some agencies like C2B, LWC,TTW has more claims, around 43% . Such agencies should be verified for any fraud or not?

- Online sale is very high compare to offline hence should be focused on how to increase/ improve processes for online sale.

Next Steps

If time permitted, we could have tried the following :

- Better feature engineering
- An ensemble of different models
- Reduced False Positive(FP)

References

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html

https://www.dezyre.com/recipes/tune-hyper-parameters-using-grid-search-in-python

https://www.dezyre.com/recipes/select-model-using-grid-search-in-python