Travel Insurance

Will they claim it ?

Team : *Outstanding Outliers*

Meet the Team

Sarika Nangare

Nikhil Dalvi

Chandana Nighut

Aniket Naik

What is Travel Insurance? And Why do you Need it?

• Travel insurance provides you with protection against any financial losses due to sudden and unfavourable events related to your journey.

•We don’t want to think the worst, but it’s better to be prepared than sorry!

• Without travel insurance, all your plans could go up in smoke and cost you more than what you’ve already invested into the trip.

Insurance Benefits

Here are some more benefits of having travel insurance:

- Trip Cancellation Protection
- Emergency Medical Coverage
- Trip Delay Protection

Problem Statement

To predict whether the customers will report the claim on travel insurance.

Business Problem

In Travel Insurance industry companies take risks over customers.

Automatically predicting the claims from the various travel insurance-related attributes, the model will predict the future claims and can help pricing and risk management team to sanction the claims.

Stakeholders :

- CEO,Senior Management
- CFO,Head of Finance
- Risk Manager
- Director of Insurance

Pain Point

Sometimes **False claims** are getting reported and hence company is experiencing revenue loss however at some time genuine claims are getting reported also leading to lawsuit against the company.

False Positives is harmful for the business and we should look out for it to minimize financial loss.

Business and Data Science Metric

Business Metric: Rate of claims sanctioned not more than 10%

Data Science Metric: Precision_score

$$\text{Precision} = \frac{TP}{TP + FP}$$

Data

Dataset Information : The training dataset consists of data corresponding to **52310** customers and the test dataset consists of **22421** customers. Following are the features of the dataset.

Feature	Datatype	Description
ID	int64	The identification record of every observation.
Agency	object	Name of agency .
Agency Type	object	Type of travel insurance agencies .
Distribution Channel	object	Distribution channel of travel insurance agencies .
Product Name	object	Name of the travel insurance products .
Duration	int64	Duration of travel
Destination	object	Destination of travel .
Net Sales	float64	Amount of sales of travel insurance policies
Commision (in value)	float64	The commission received for travel insurance agency.
Age	int64	Age of insured (Age)

Target Column y	Datatype	Description
Claim	int64	Claim Status (Claim) 1(Yes) / 0(No)

Evaluation Metric

The evaluation metric for this project is **Precision_score.**

False Positive - Predicted as claimed but actually it is not claimed

False Negative - Predicted as not claimed but actually it is claimed

		Actual	
		Claimed	**Not Claimed**
Predicted	**Claimed**	True Positive	**False Positive**
	Not Claimed	False Negative	True Negative

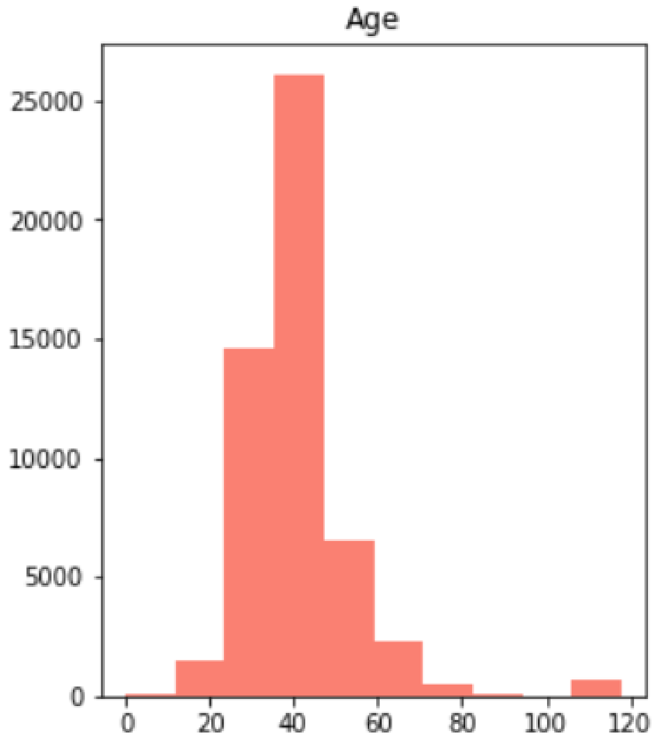
Exploratory Data Analysis

Class Imbalance : The distribution of the target variable shows a clear imbalance in the two classes.

SMOTE and class_weights is used is used to balance the distribution of target variable.

Exploratory Data Analysis

Column : Age



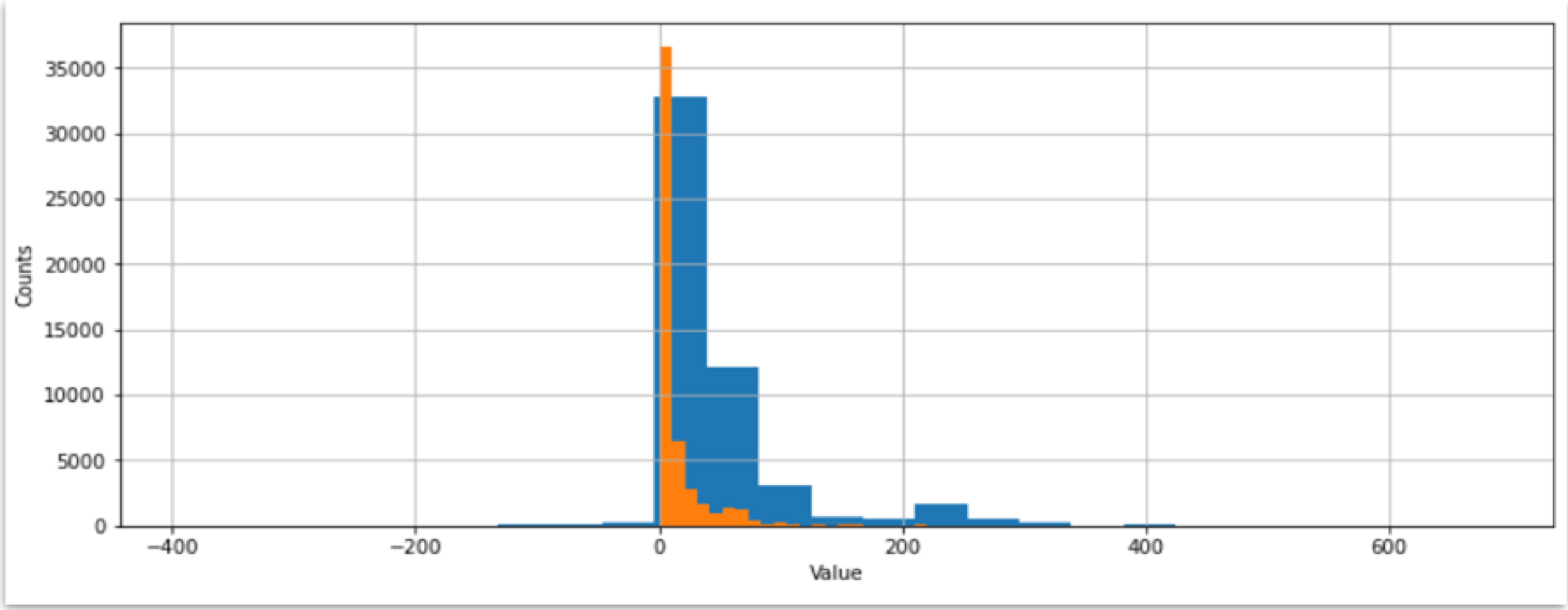
Minimum value = 0
Maximum value = 118
We have assumed that any individual upto age 100 is valid anything above is replaced with median.

Values in Age Column are Categorized as :

Age	Category
0 - 17	Child
18 - 24	Youth
25 - 34	Professionals
35 - 44	Adult(35 - 44)
45 - 54	Adult(45 - 54)
>54	Senior

Exploratory Data Analysis

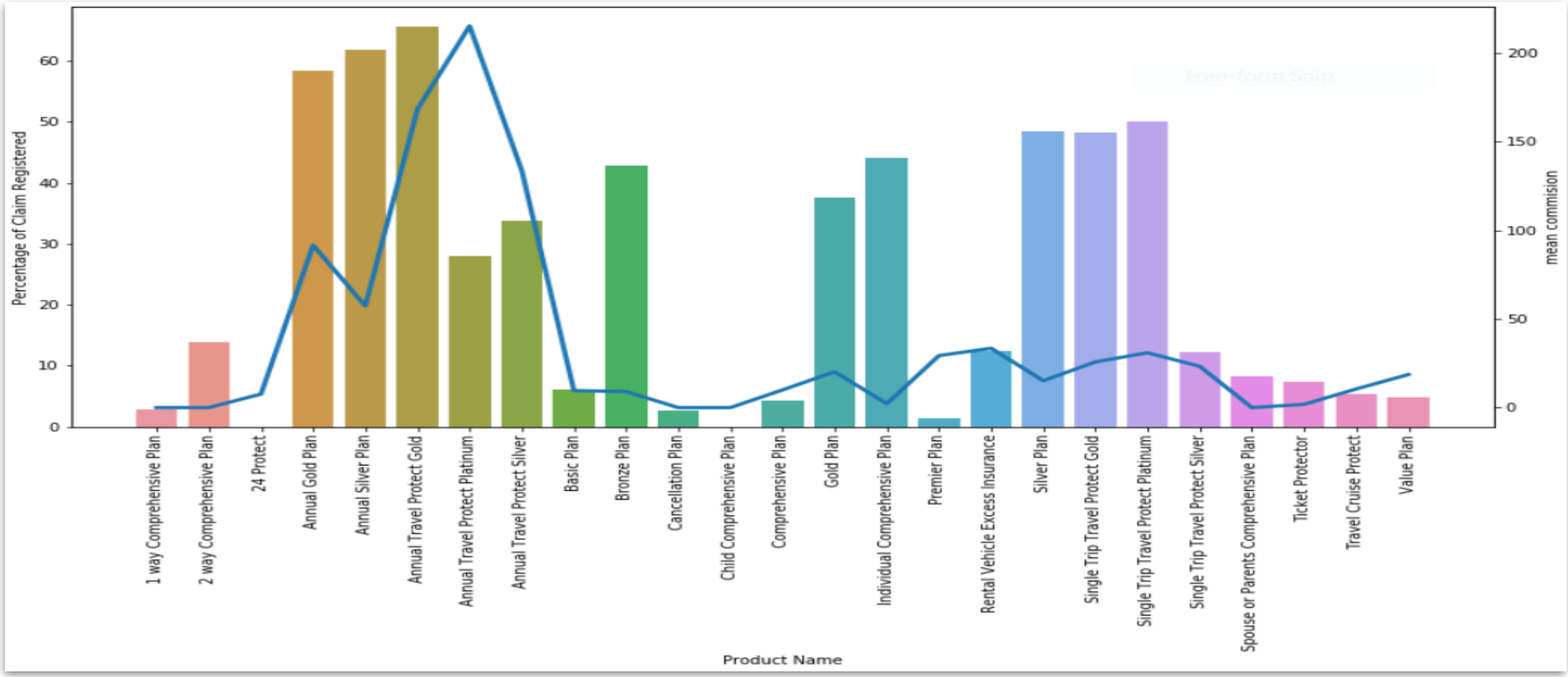
Net Sales vs Commission reported



These both column seems to related but the graph plot shows disparency as low net sales shows high commission.

Exploratory Data Analysis

Distribution of Claim and Commission as per Product Name

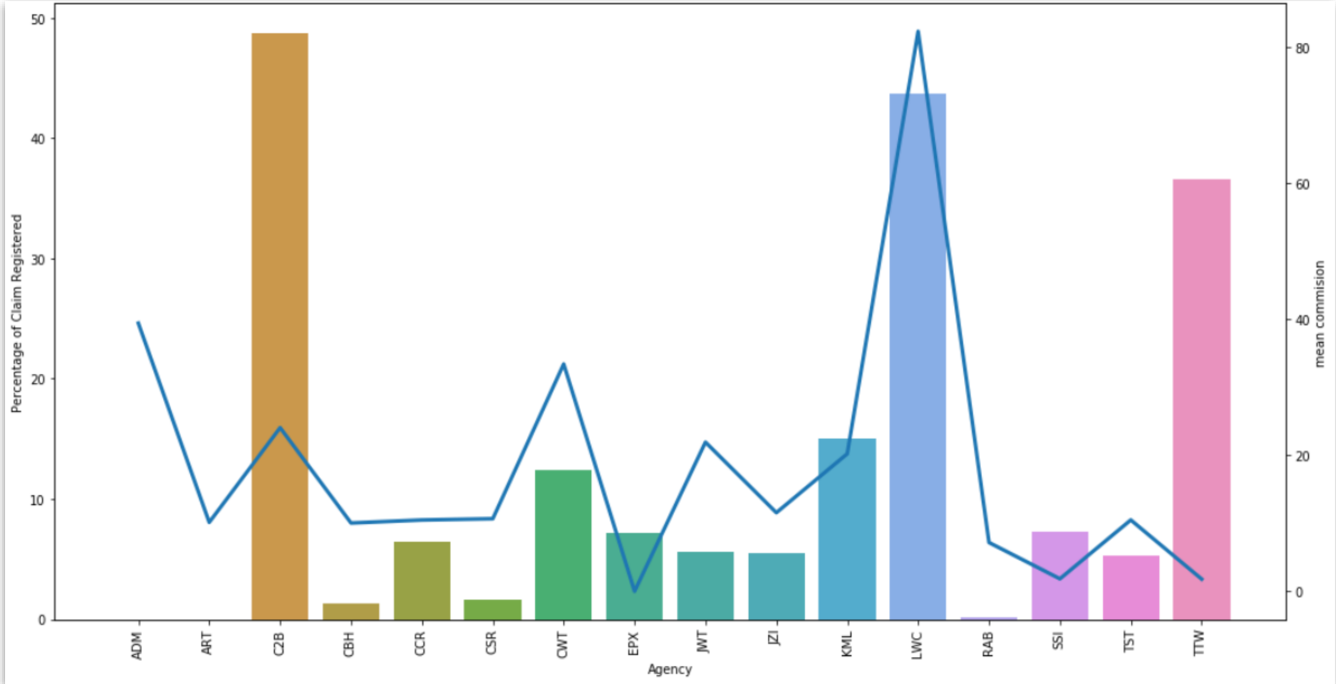


Annual Gold and Annual Silver Plans have higher no. of Claims registered whereas commission drawn is very low.

Annual Travel Protect Platinum Plan has low no. of Claims registered whereas commission drawn is very high

Exploratory Data Analysis

Distribution of Claim and Commission as per Agency



- Certain agencies have less registered claims but they draw commission on the Insurance Policies.(CBH, CSR, CWT, JWT)

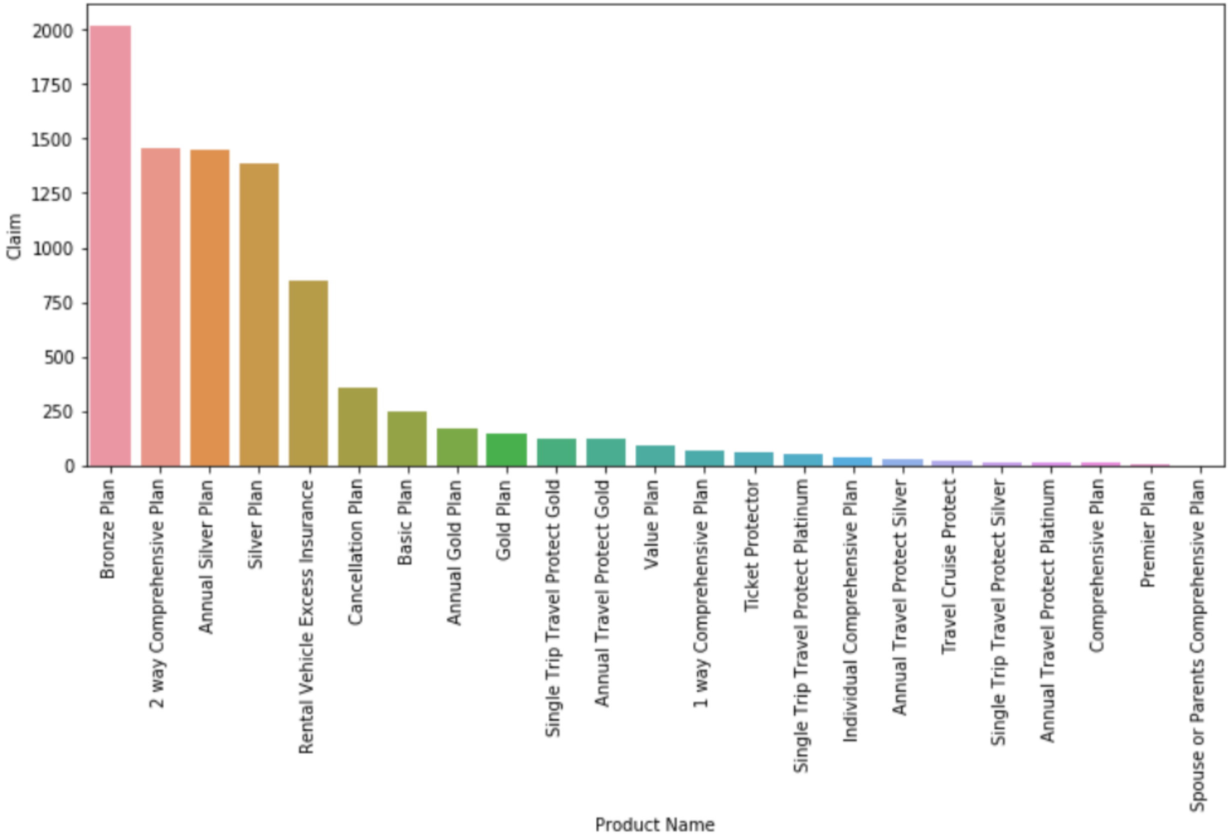
- ADM, ART has no registered claim but commission is high

- Certain agencies have higher percent of registered claims but the mean commission is very low(C2B, TTW)

- Claim and Mean Commission is high for LWC.

Exploratory Data Analysis

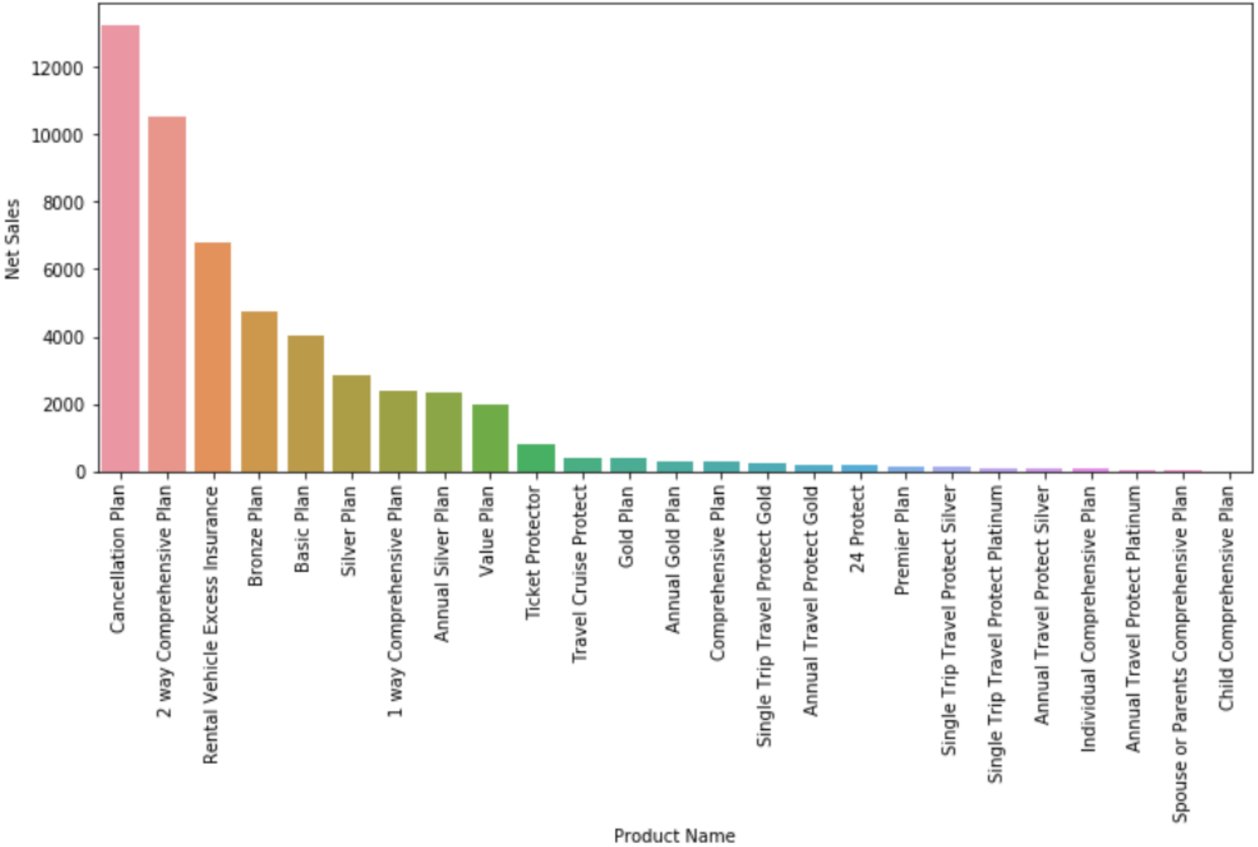
Relation between registered Claims and Product Name



Bronze Plan, 2 way Comprehensive Plan, Annual Silver Plan has highest claims reported.

Exploratory Data Analysis

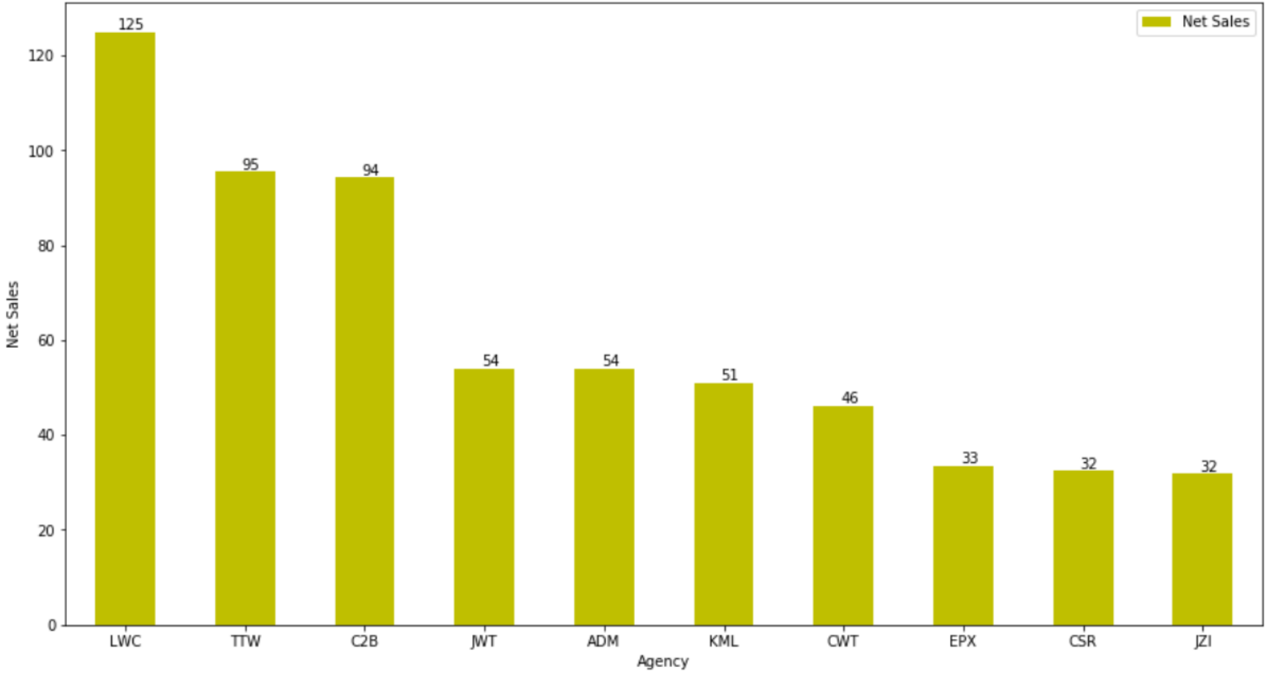
Products vs Net Sales.



- As Cancellation Plan and 2 way comprehensive plans are contributing more in Sales.
- Even Cancellation plan has less claims so Agencies should focus more on this plan.

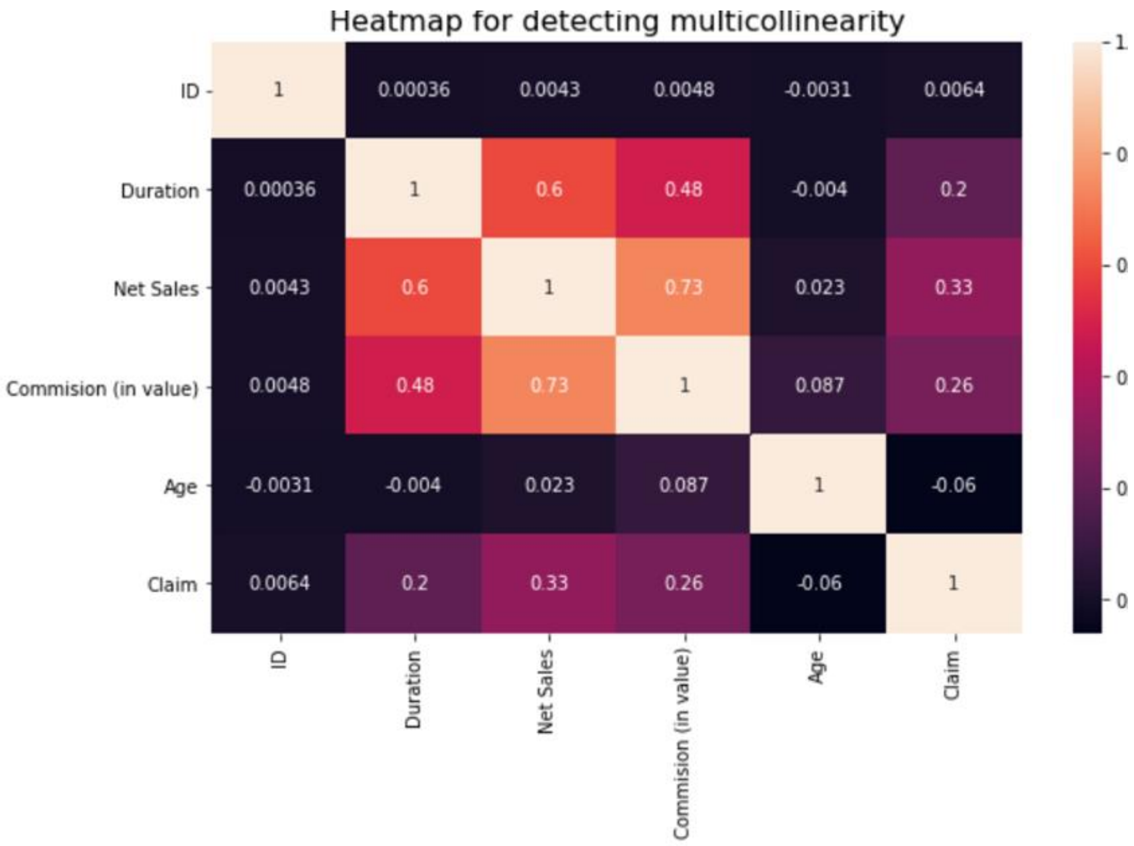
Exploratory Data Analysis

Agencies with the maximum number of Net Sales.



The Agencies with Top 5 Net Sales were LWC,TTW and C2B.

Feature Selection



- There is no high correlation between any features and hence we can continue with all of the above features to train our model.

- ID column has very low correlation with other features, so it has been dropped.

Models and Approaches

Four vanilla models were assessed without performing any hyperparameter tuning and without treatment of class imbalance of the target. The models were

- Logistic Regression
- Random Forest Classifier
- XGBoost Classifier
- AdaBoost

None of the four vanilla models were able to give an Precision score above 75% on actual test data.

This called for performing hyperparameter tuning using Grid Search and also treatment of class imbalance using SMOTE and Class weights for further improvement of the Precision score and ROC_AUC score.

Model Score

Models Assessed : The vanilla models used yielded the following results below.

Modelling Method	Precision	Recall	AUC_ROC
Logistic Regression	• **0** - 0.92 • **1** - 0.68	• **0** - 0.94 • **1** - 0.61	93.65%
Random Forest Classifier	• **0** - 0.96 • **1** - 0.74	• **0** - 0.94 • **1** - 0.81	95.82 %
XGBClassifier	• **0** - 0.95 • **1** - 0.80	• **0** - 0.96 • **1** - 0.72	96.05 %
AdaBoost Classifier	• **0** - 0.60 • **1** - 0.88	• **0** - 0.95 • **1** - 0.37	86.53 %

Model Tuning

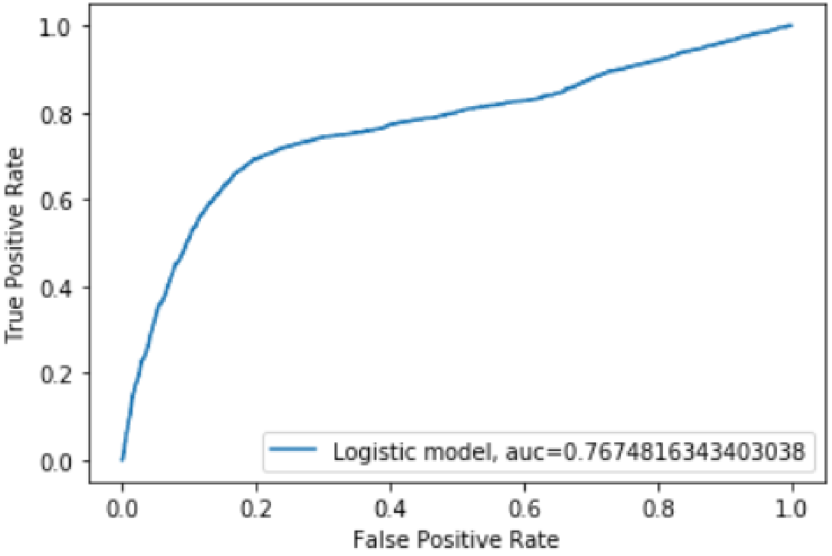
After performing hyperparameter tuning using Grid Search and treating imbalanced classes also RF with Class Weights and **SMOTE** along with Ensemble model of **GradientBoostingClassifier** and Adaboost yielded the following results:

Modelling Method	Precision	Recall	ROC_AUC
Logistic Regression with GridsearchCV	• **0** - 0.86 • **1** - 0.61	• **0** - 0.97 • **1** - 0.22	76.7%
Random Forest Classifier with SMOTE	• **0** - 0.96 • **1** - 0.75	• **0** - 0.95 • **1** - 0.80	95.98 %
Random Forest Classifier with Class weights	• **0** - 0.94 • **1** - 0.84	• **0** - 0.97 • **1** - 0.70	83.51 %
Ensembling(GBC on Adaboost Classifier)	• **0** - 0.95 • **1** - 0.87	• **0** - 0.98 • **1** - 0.75	96.69 %

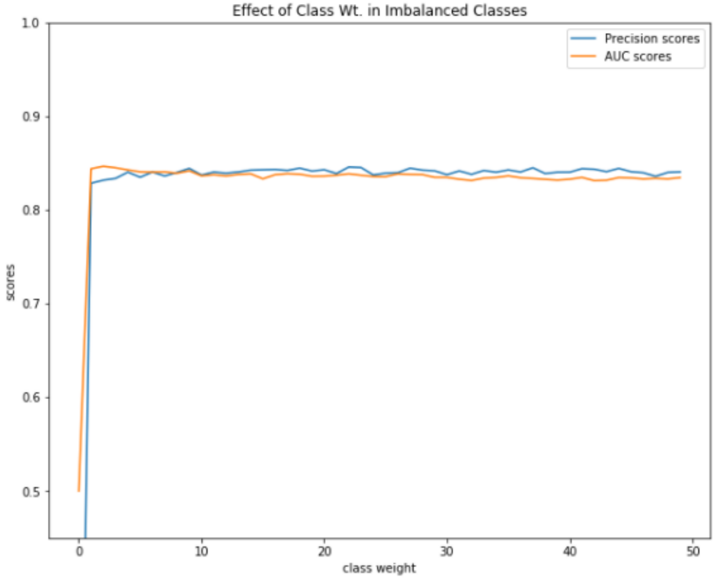
Evaluation & Results

Below are the precision score plots for the after hyperparameter tuning.

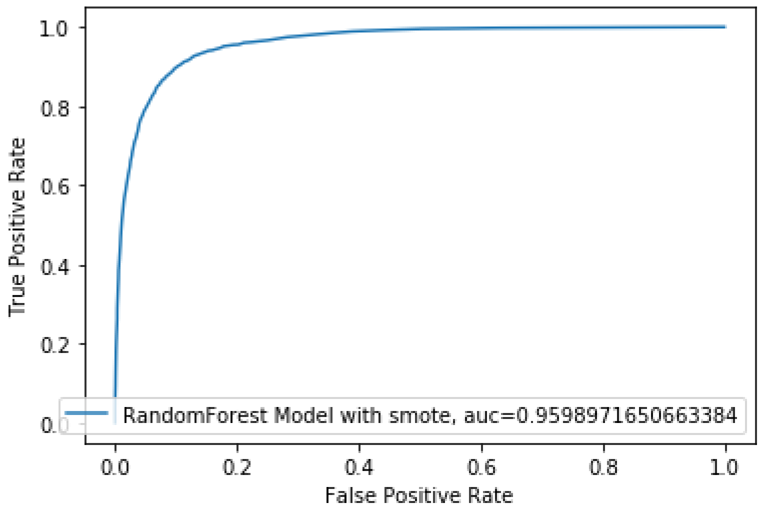
Logistic Regression with GRidSearchCV



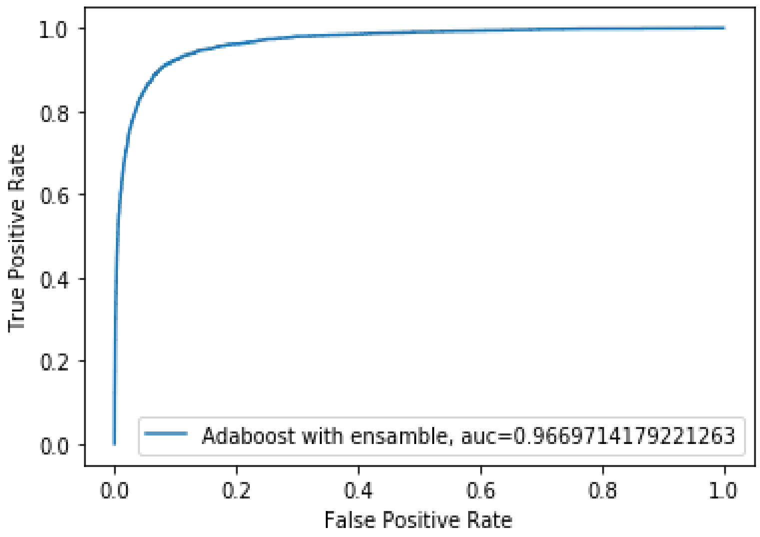
Random Forest with Class weight



Random Forest with SMOTE



Ensemble model with GBM and Adaboost



Final Results

From the above observations and plottings it can be inferred that the best performing model was **AdaBoost with GBC (Ensemble)**giving an **precision score** of **88 %**. While **AdaBoost** is used , it is always prudent to start from simpler algorithms and then go to complex ones.

Confusion Matrix :

	Predicted Positive	Predicted Negative
Actual Positive	2175	707
Actual Negative	357	14024

Insights and Recommendations

- **Insight:** From the Age Claim visualizations it is evident that 30-45 age group travel most and have maximum claim registered.

- **Recommendation**:Insight So for this age group we can increase the premium and commission charges on the claimed insurance.
  Offer more lucrative discounts and schemes to other age groups for boosting the Net Sales.

- **Insight :**Cancellation Plan is the most selling one and the claims registered are less.

- **Recommendation:**Agencies with low Net Sales should sell the products which are highly sold and has low claim rates.(Cancellation Plan).This will maximize profit.

Thank you .

Outstanding Outliers:

Sarika Nangare
Chandana Nighut
Nikhil Dalvi
Aniket Naik