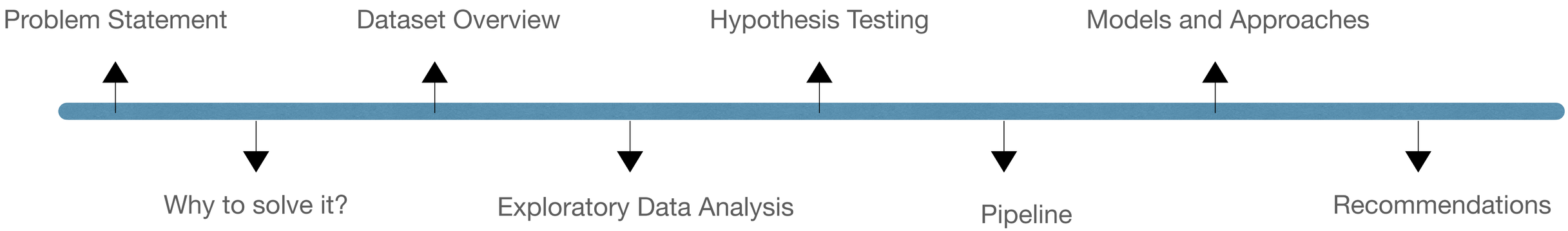
Will They Claim It?

Binary Pundits
(Monica Data, Prathmesh Mahadeshwar, Avani Dhatrak, Rounak Sharma

Overview


Problem Statement
Dataset Overview
Hypothesis Testing
Models and Approaches
Why to solve it?
Exploratory Data Analysis
Pipeline
Recommendations

Problem Statement

Business Overview

- SafeTravels is world's largest insurance companies specializing in travel insurance
- **Potential Customers** - Travelers wanting to insure themselves against travel related risks
- **Offerings** - 1-way travel insurance, 2-way insurance, insurance against cancellations, renting vehicles insurance, etc.

Business Problem

- 1000s of claims spread across different products are received weekly thereby increasing manual effort and time
- Automatically predicting the claims could lead less effort, time and operational costs
- Whenever a claim is registered and if accepted it is a cost to the company

Data Science Problem

- Building an ML model to predict whether the customer will apply for the claim or not

Why solve this problem?

Business Impact

- Improve Financial Implications
- Enhance Customer Experience
- Reduce Manual Effort

Key Stakeholders

- CFO, The company board
- Director Claims Division,
- Head of New Products
- COO
- We are going ahead with **CFO**

Data Science Metric

- Precision score

Key Assumptions

- Claimed/Claim or value 1 means Claim Registered
- Plan/Insurance/Product are same
- Sales and commission are considered to be in $

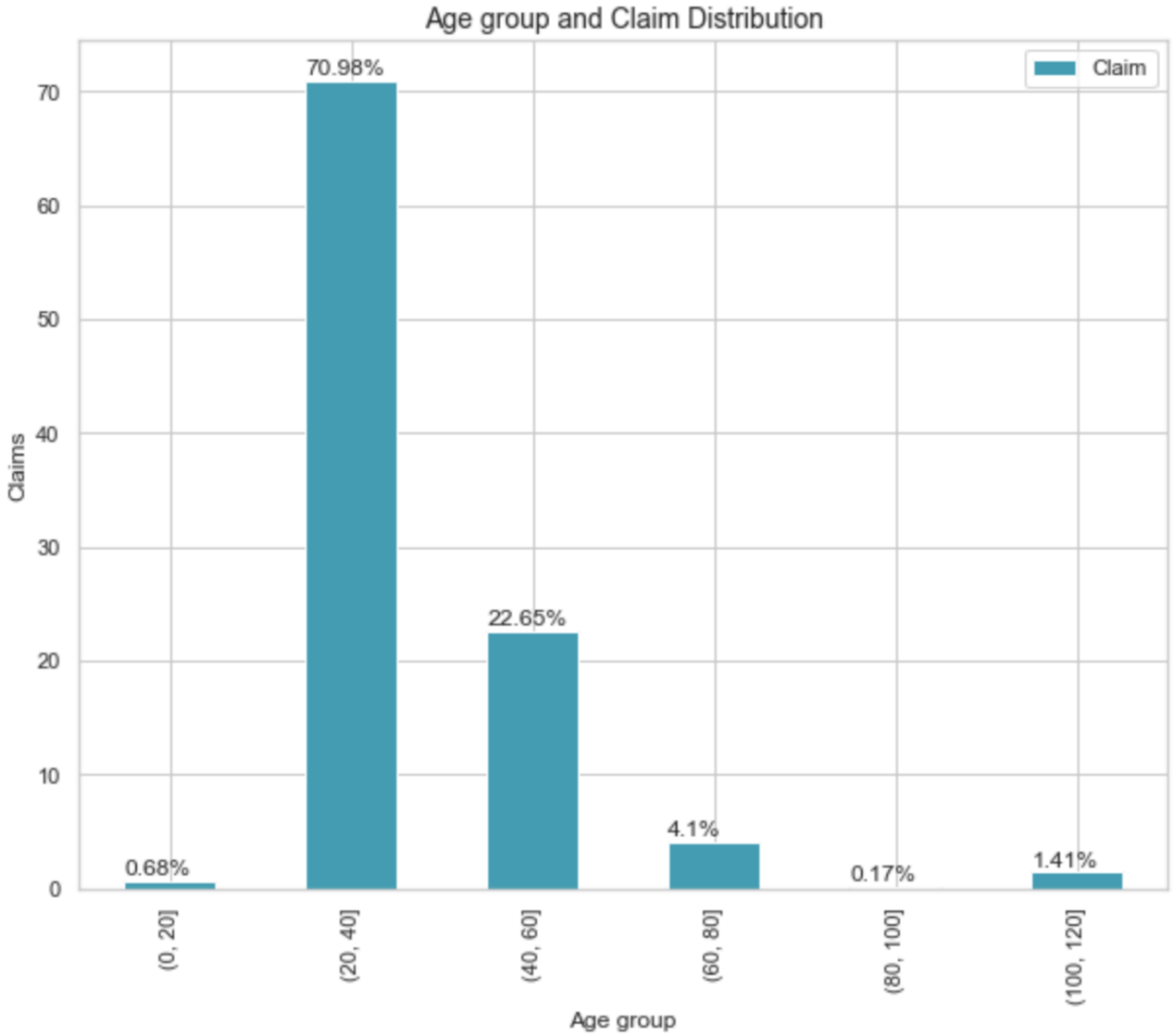
Data

Features	Description
ID	The identification record of every observation
Agency	Name of agency
Agency Type	Type of travel insurance agencies
Distribution Channel	Distribution channel of travel insurance agencies
Product Name	Name of the travel insurance products
Duration	Duration of travel
Destination	Destination of travel
Net Sales	Amount of sales of travel insurance policies
Commission (in value)	The commission received for travel insurance agency
Age	Age of insured
Claim	Claim Status

EDA :

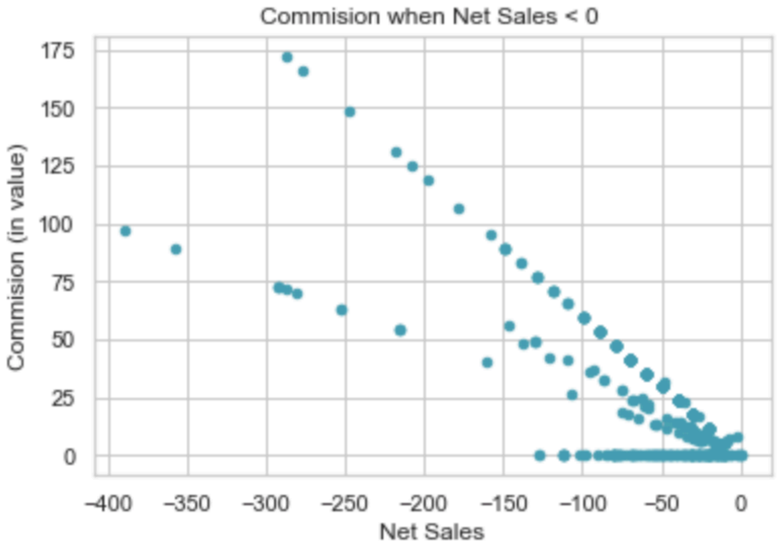
Customer Age Analysis

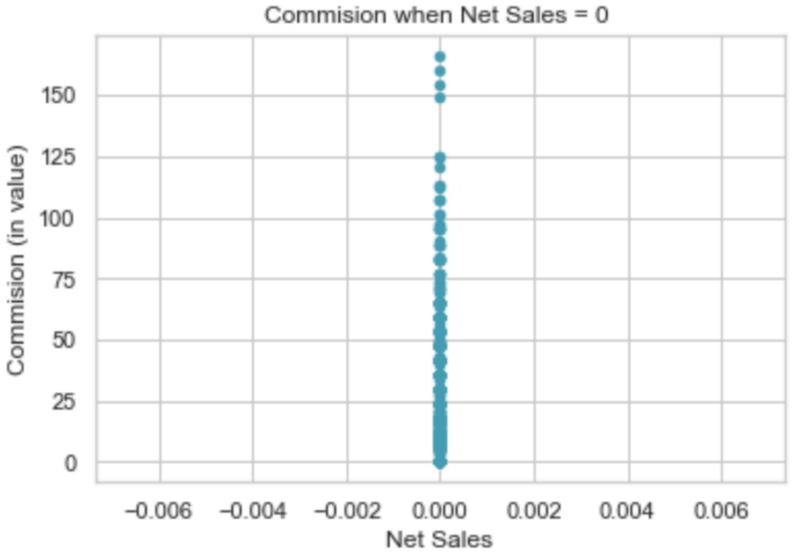
- 71% customers who apply for claims fall under the 20-40 age bracket
- Around approx 23% are into 40-60 age bracket

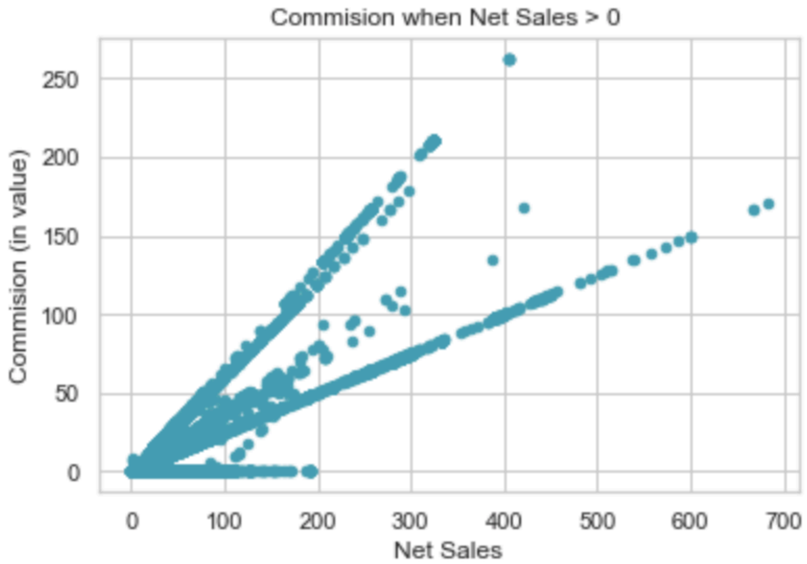


EDA:

Sales vs Commission Analysis



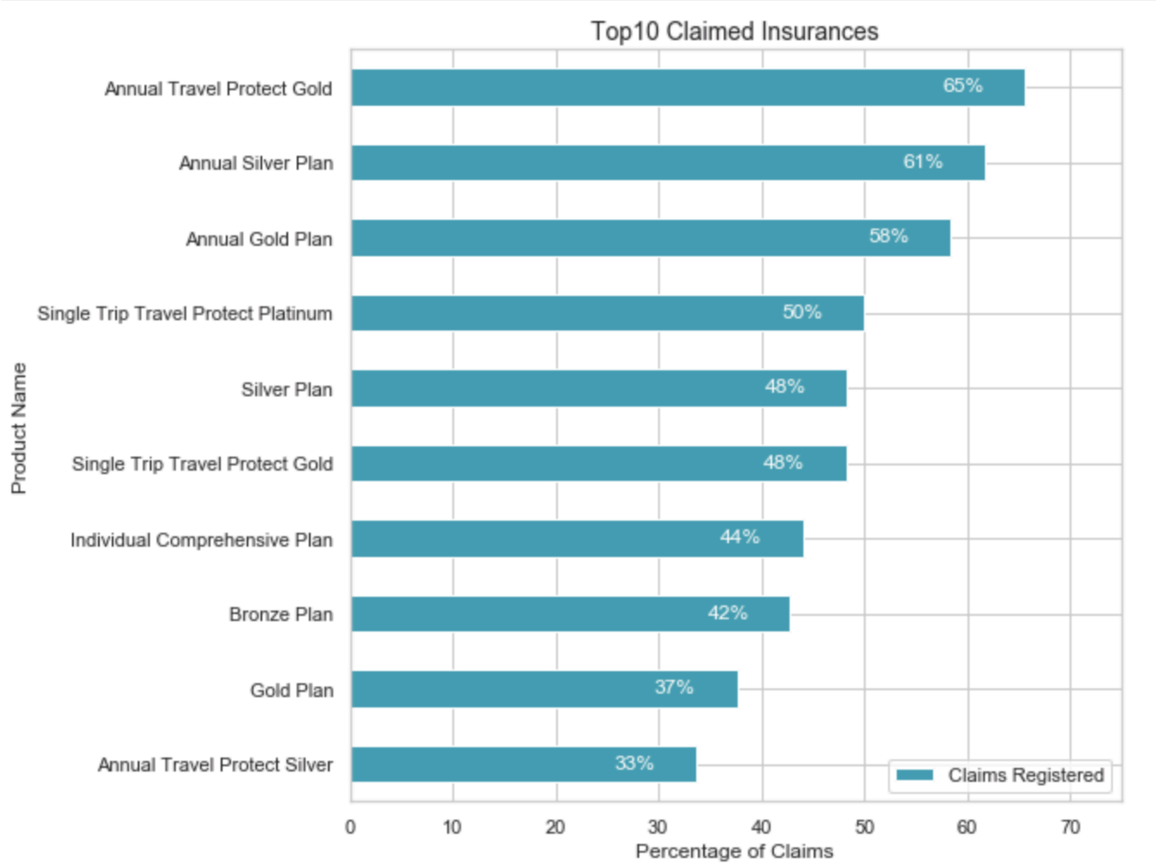


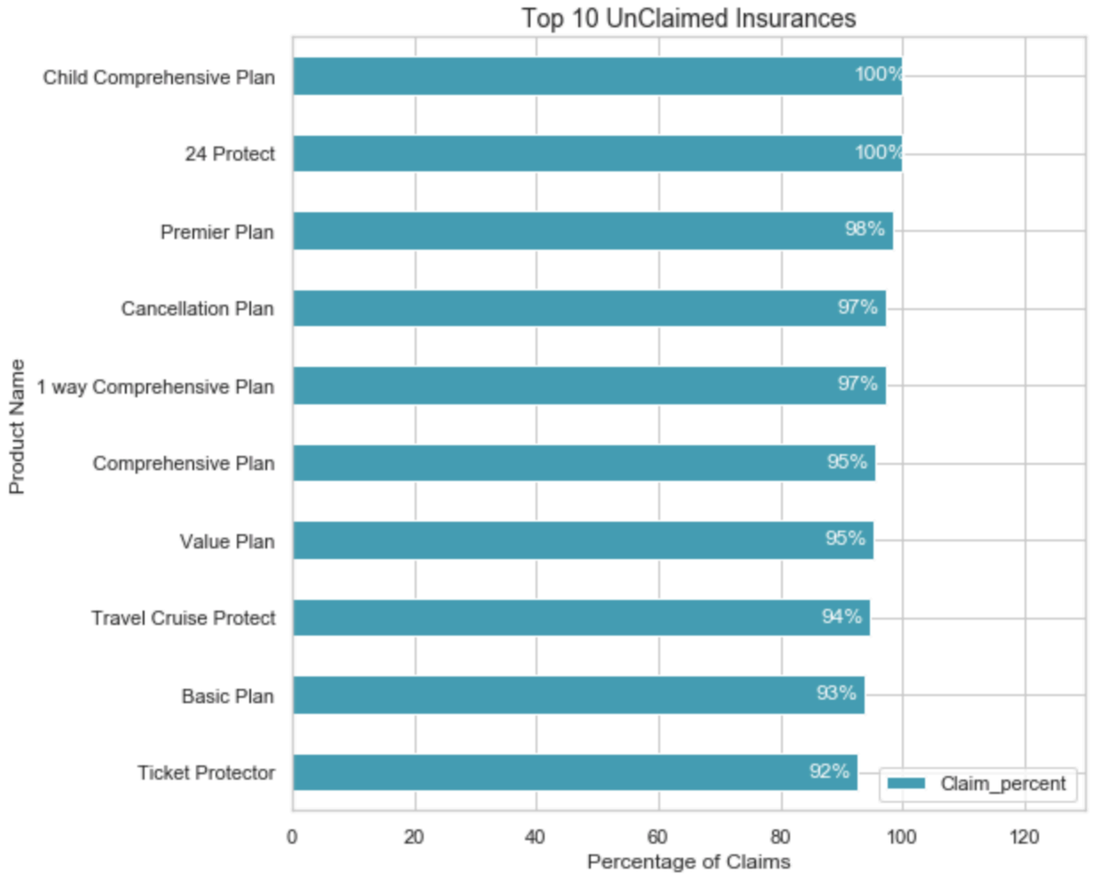


- Negative Sales from 0 to -400 have zero to positive commission with maximum commission going until 175 value.
- Zero Sales have maximum commission of 150+ value
- Positive Sales have maximum commission of 250+ value

EDA:

Portfolio Analysis

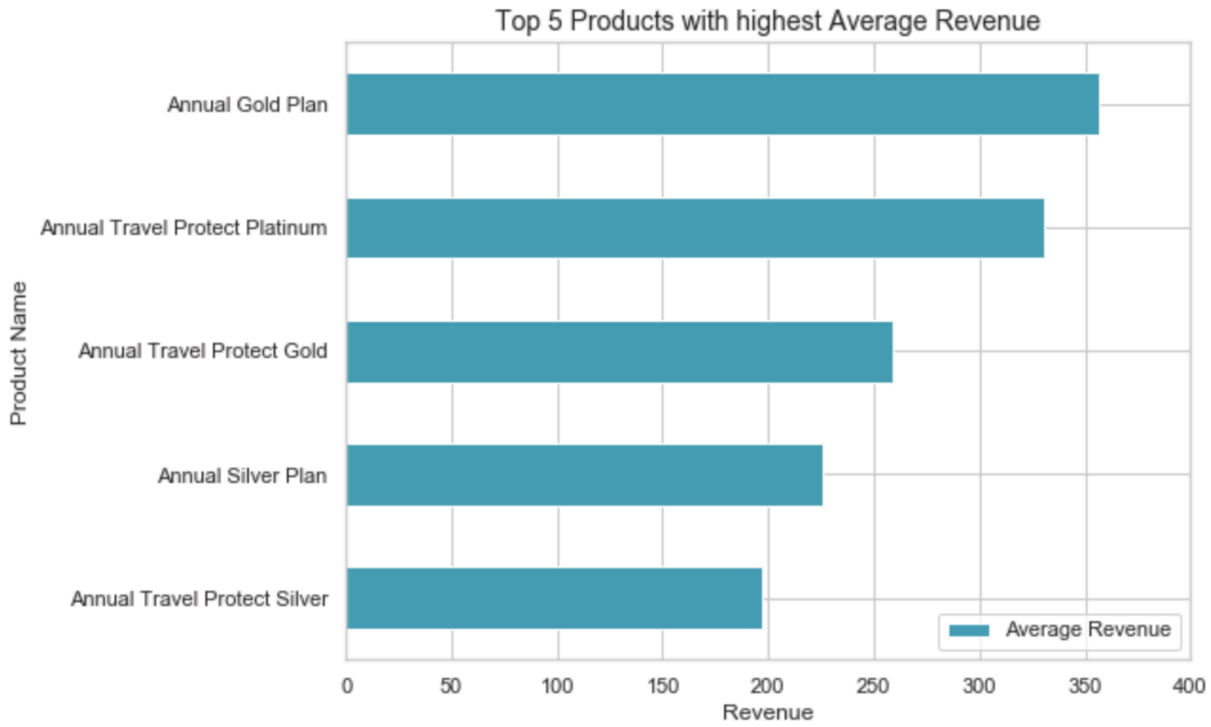


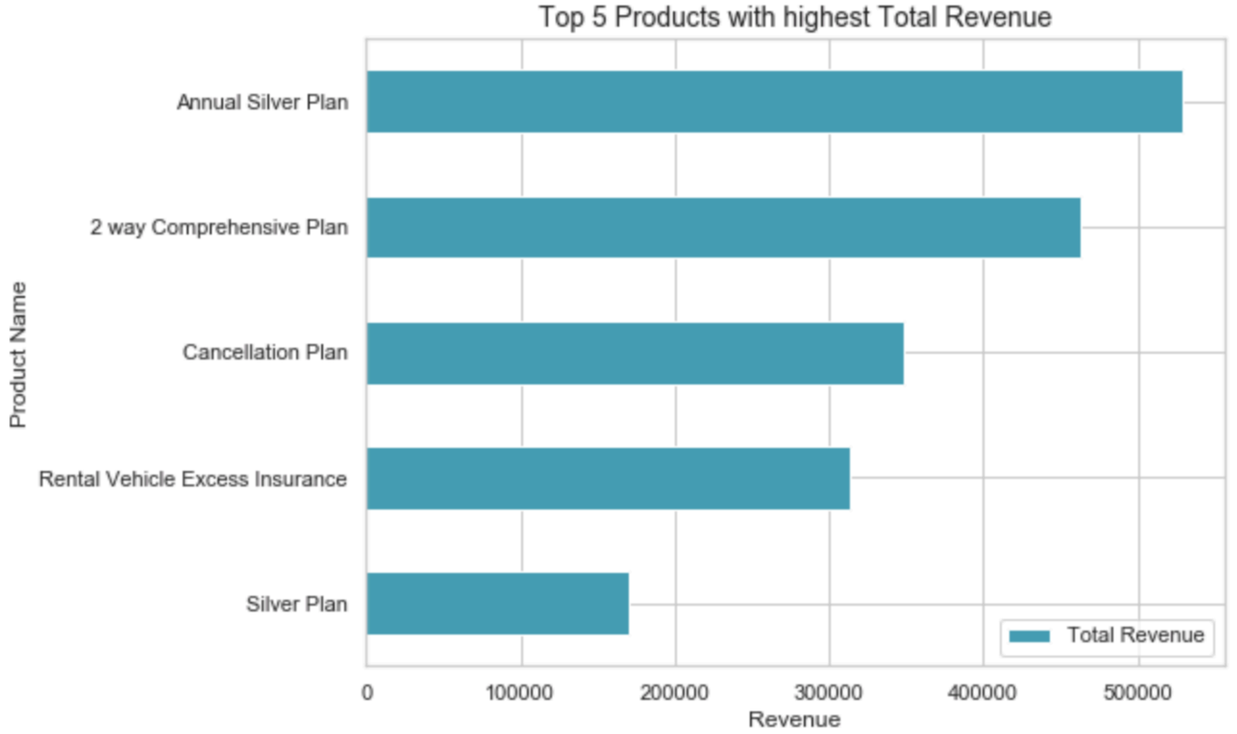


- Annual Gold, Silver and Travel protect Gold plan are claim rate greater than 50%
- The top 10 unclaimed insurances have an unclaimed rate > 90%

EDA:

Product- Sales Analysis

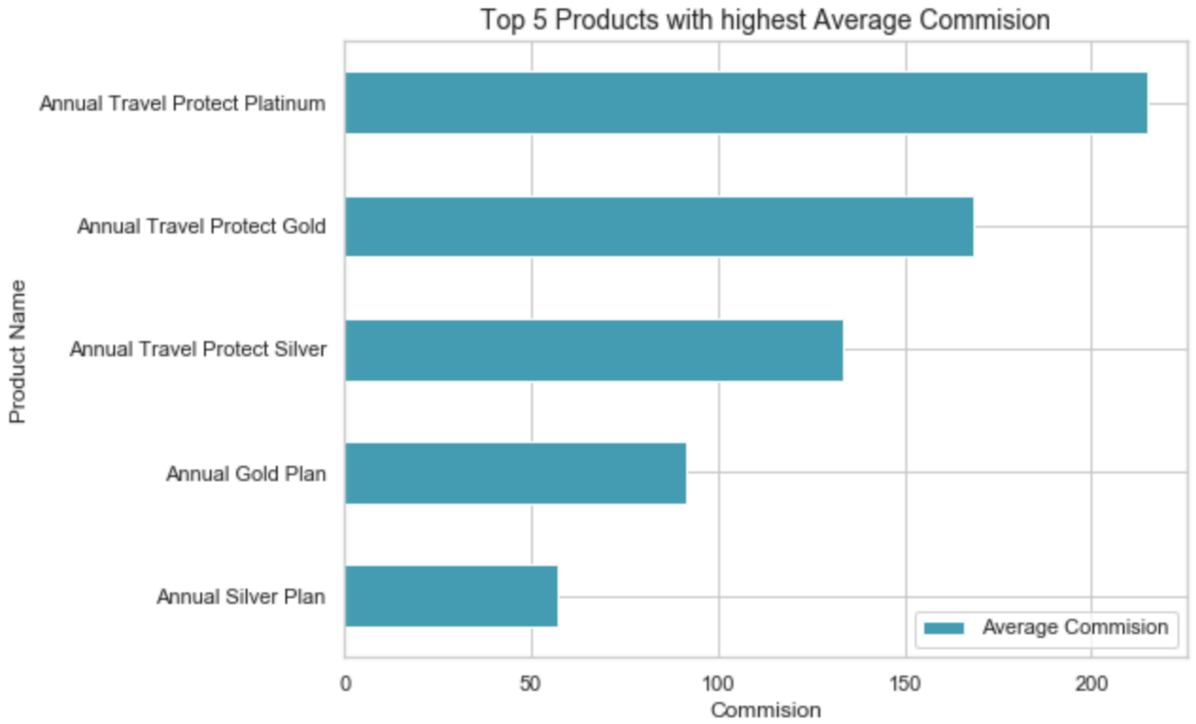


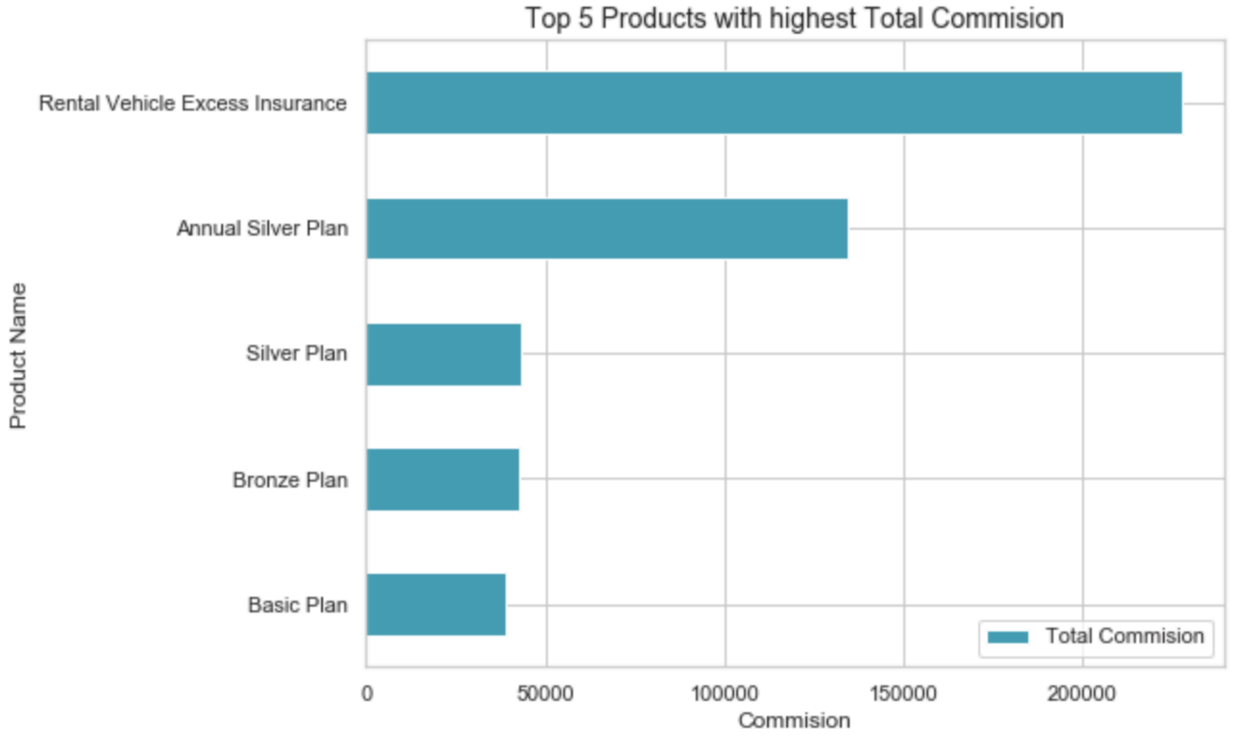


- Most claimed insurances have the maximum average sales such as Annual Gold and SIlver Plan, Travel Protect Gold
- Annual Silver Plan has overall highest Total Revenues

EDA:

Product- Commision Analysis





- Annual Plans have highest Average commission whereas Rental Vehicle Excess Insurance have maximum commission

EDA :

Product-Agency-Commision Analysis

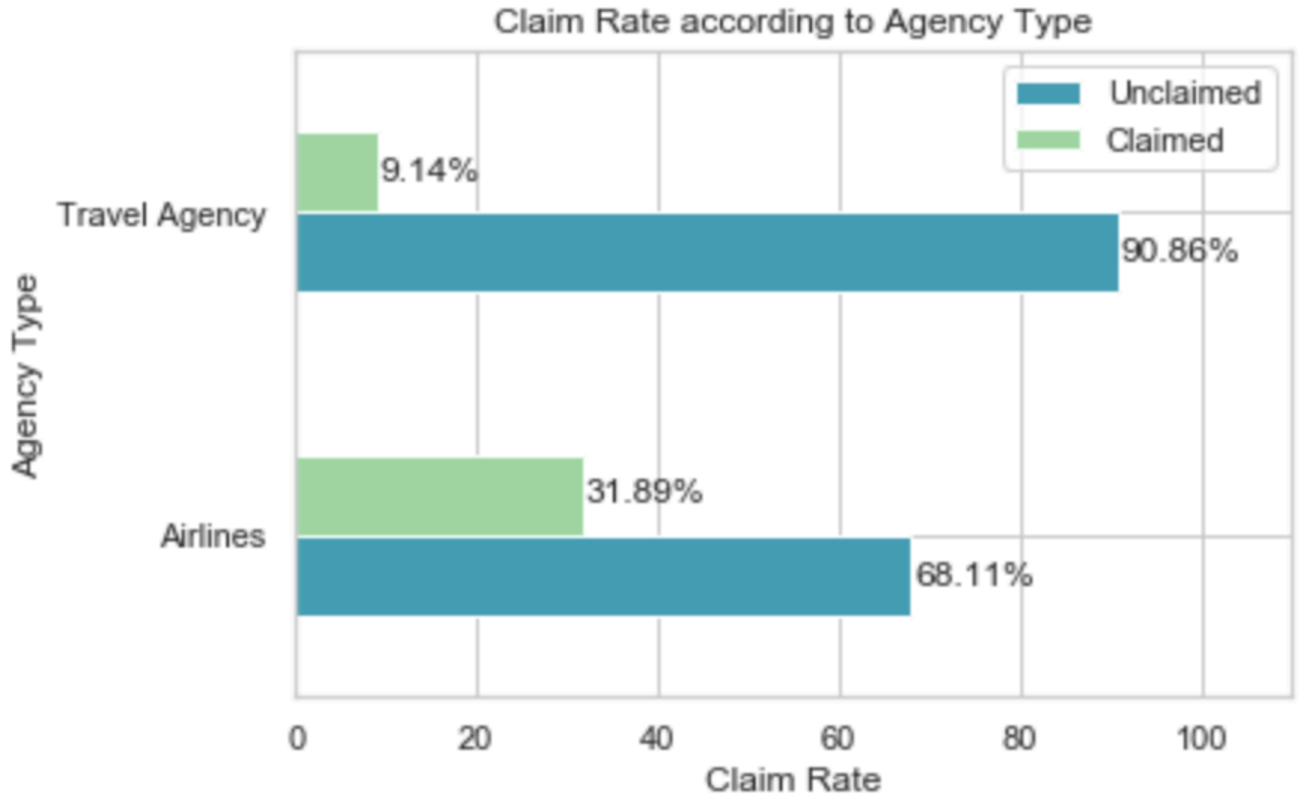
- Travel Agency offers more average commission compared to Airlines
- It can go higher upto 215 units

		commision_average
Product Name	**Agency Type**	
Annual Gold Plan	**Airlines**	91.491629
Annual Silver Plan	**Airlines**	57.291729
Annual Travel Protect Platinum	**Travel Agency**	215.064490
Rental Vehicle Excess Insurance	**Travel Agency**	33.443827

EDA :

Agency Type Analysis

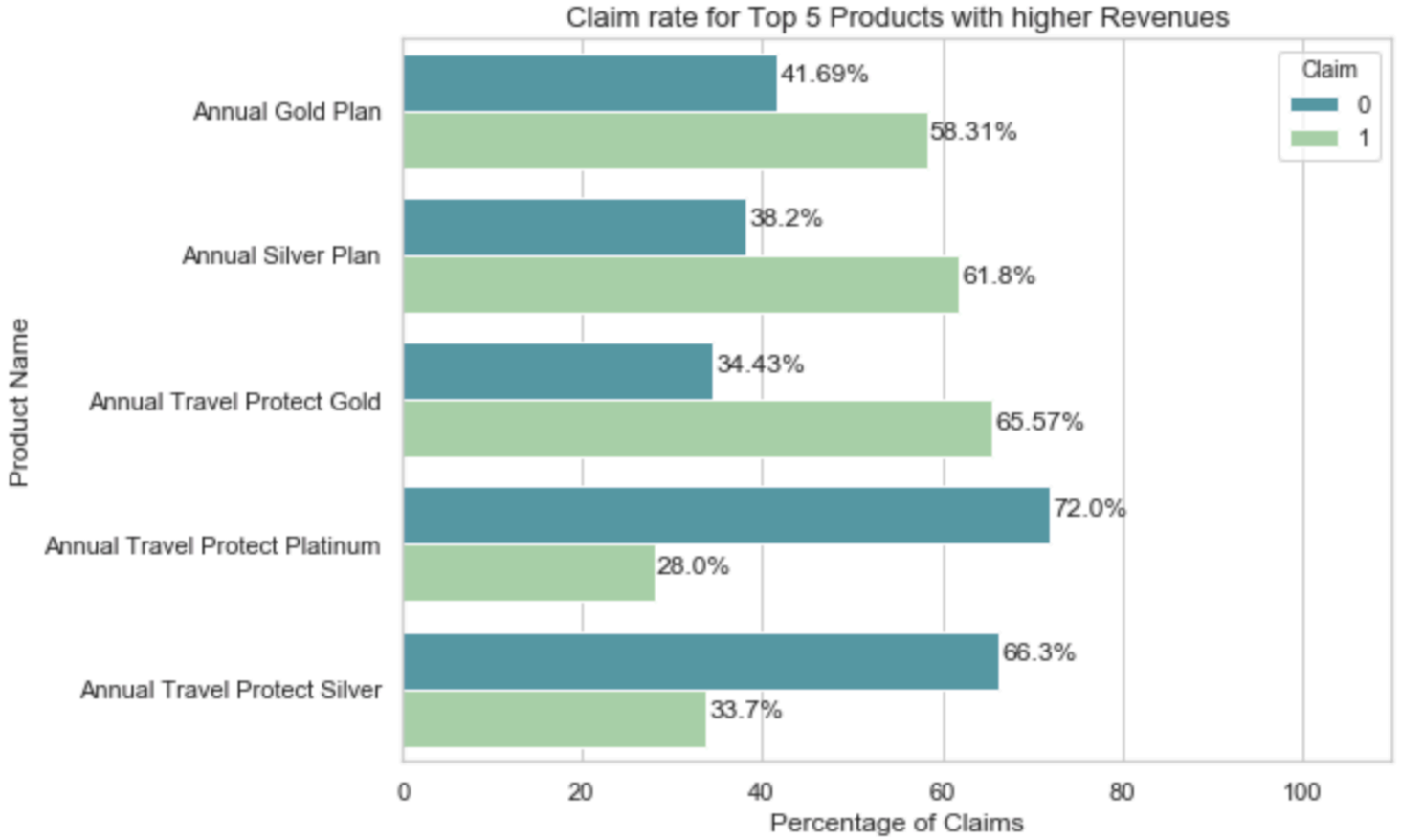
- Airlines have higher Claim ratio as compared to Travel Agency.
- Travel Agency have higher unclaimed insurances.



EDA :

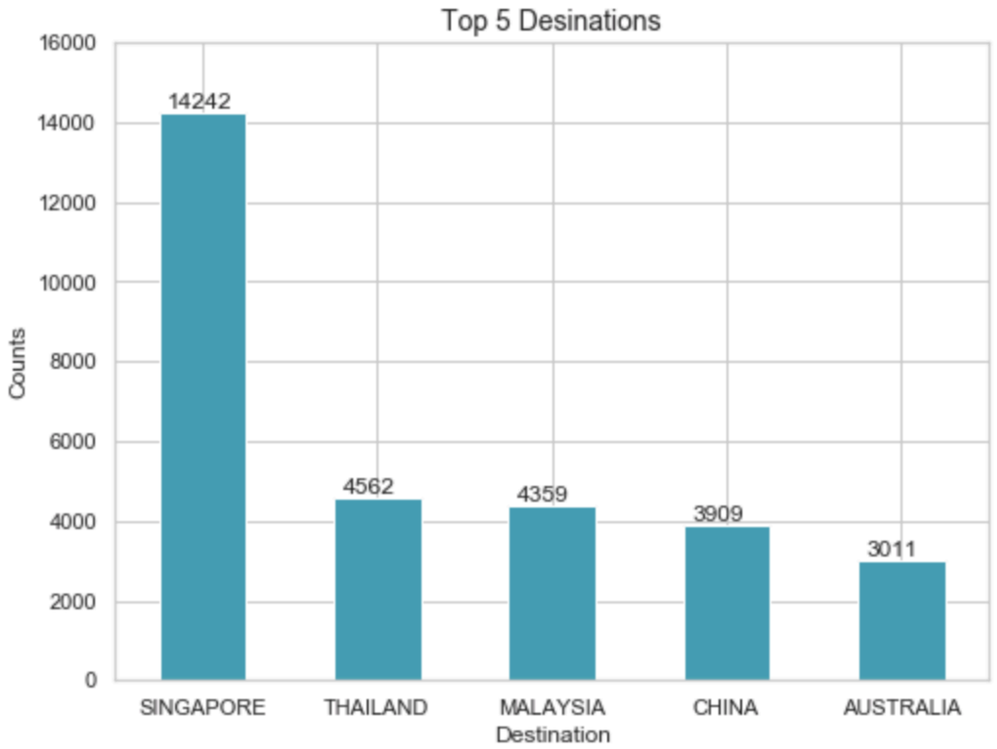
Product- Claim Rate Analysis

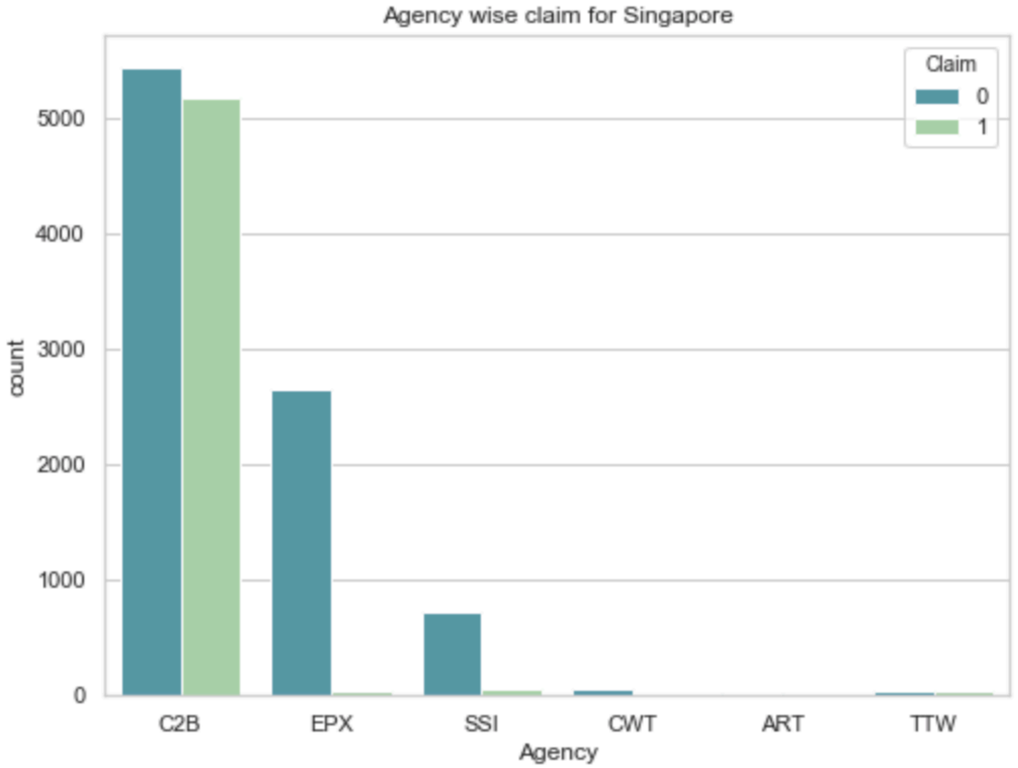
- Top 3 higher revenue generating plans have higher claim rate
- Last 2 products are not claimed minimum 65% of the times.



EDA :

Destination- Claim Rate Analysis



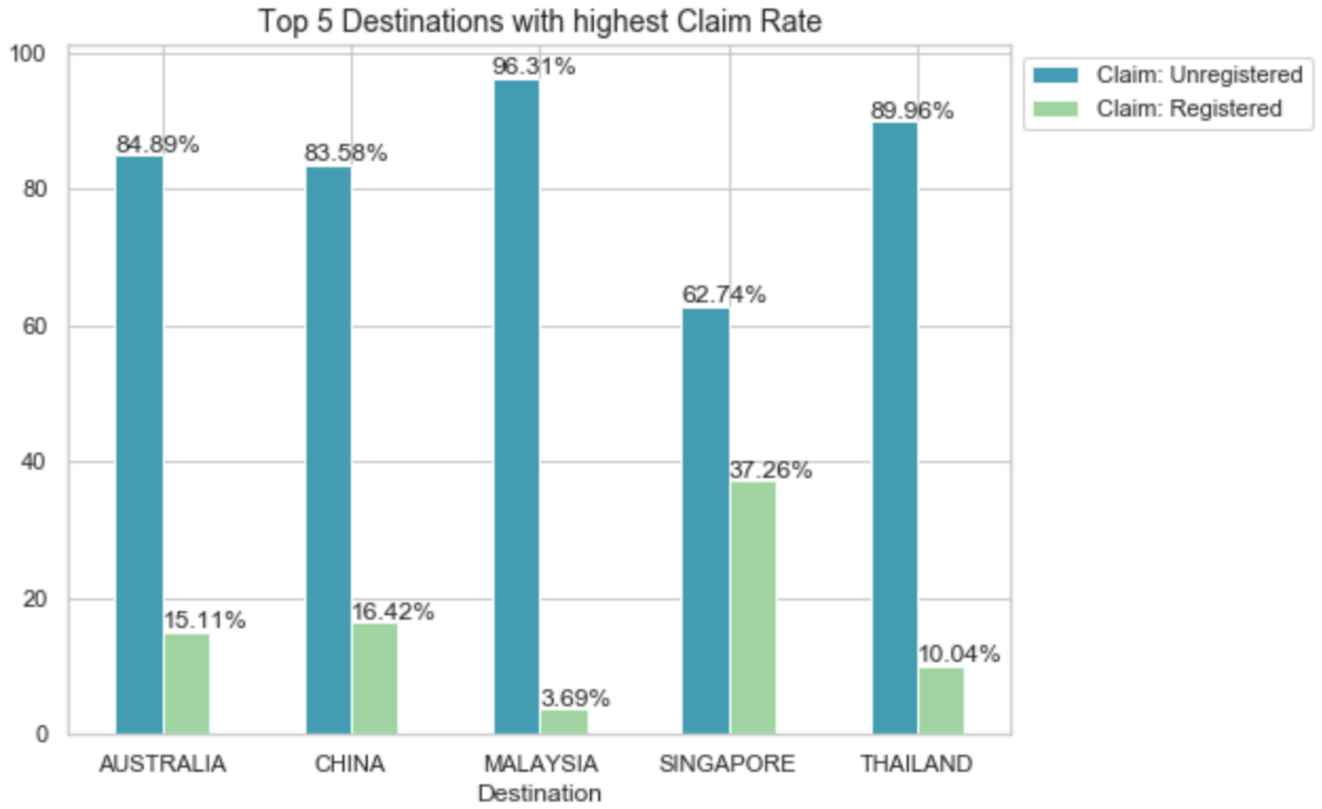


- Singapore is the most visited country and C2B has highest claim registered in Singapore

EDA :

Destination- Claim Rate Analysis

- Singapore have highest registered claims
- Malaysia and Thailand have least claimed insurances



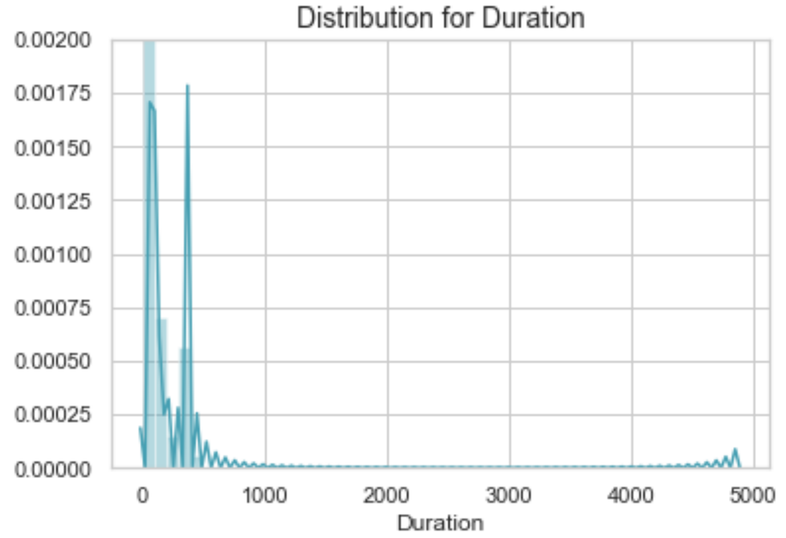
Hypothesis Testing :

Null Hypothesis — The unit of duration is in days.

Alternate Hypothesis — The unit for duration is in years

Hypothesis Testing — Distributing Dataset into 2 sets.

Min	-2
Max	4881
Mean	58.25
Median	24

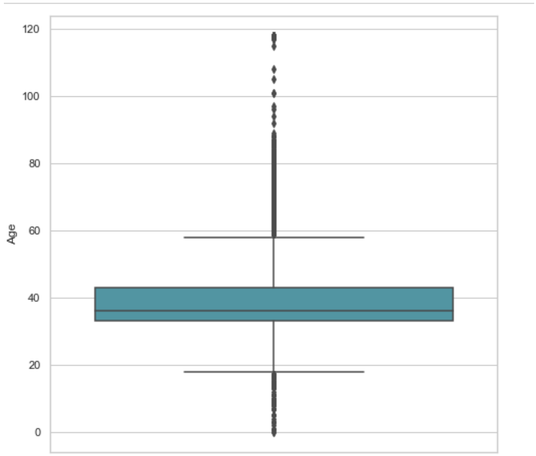


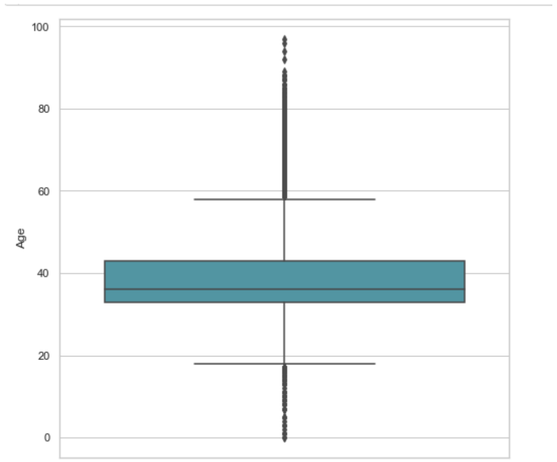
		Counts
Duration Year	**Product Name**	
1	**Annual Gold Plan**	176
	Annual Silver Plan	1483
	Annual Travel Protect Gold	139
	Annual Travel Protect Platinum	41
	Annual Travel Protect Silver	81
	Bronze Plan	3
	Child Comprehensive Plan	3
	Individual Comprehensive Plan	22
	Rental Vehicle Excess Insurance	1
	Spouse or Parents Comprehensive Plan	7
2	**Annual Silver Plan**	1
12	**Ticket Protector**	3
13	**Ticket Protector**	6

- Less than 4% customers had plans greater than 365 days
- Customers had annual plans, so this answers the assumption of unit of duration being in days
- Also the annual plans have duration of 364 days this proves that the duration is in unit of days.
- So we fail to reject our Null Hypothesis

		counts
Product Name	**Duration**	
Annual Gold Plan	**364**	41
Annual Silver Plan	**364**	361
Annual Travel Protect Gold	**364**	15
Annual Travel Protect Platinum	**364**	6
Annual Travel Protect Silver	**364**	4

Pipeline (Outlier Treatment)



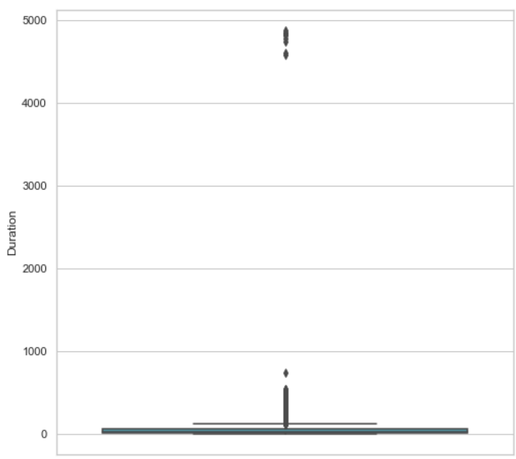


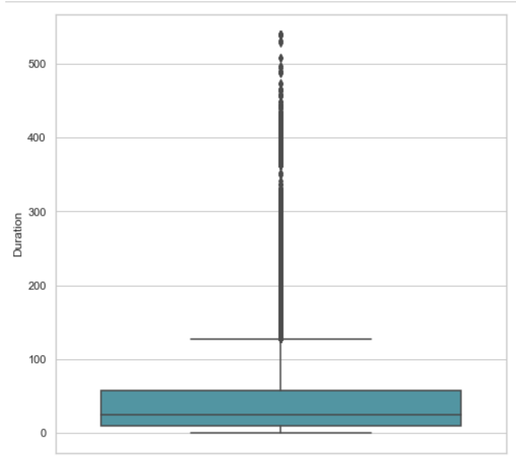
Age

Before Treatment

After Treatment

Before	After
737	0





Duration

Before Treatment

After Treatment

Before	After
16	0

Pipeline

Missing Value :

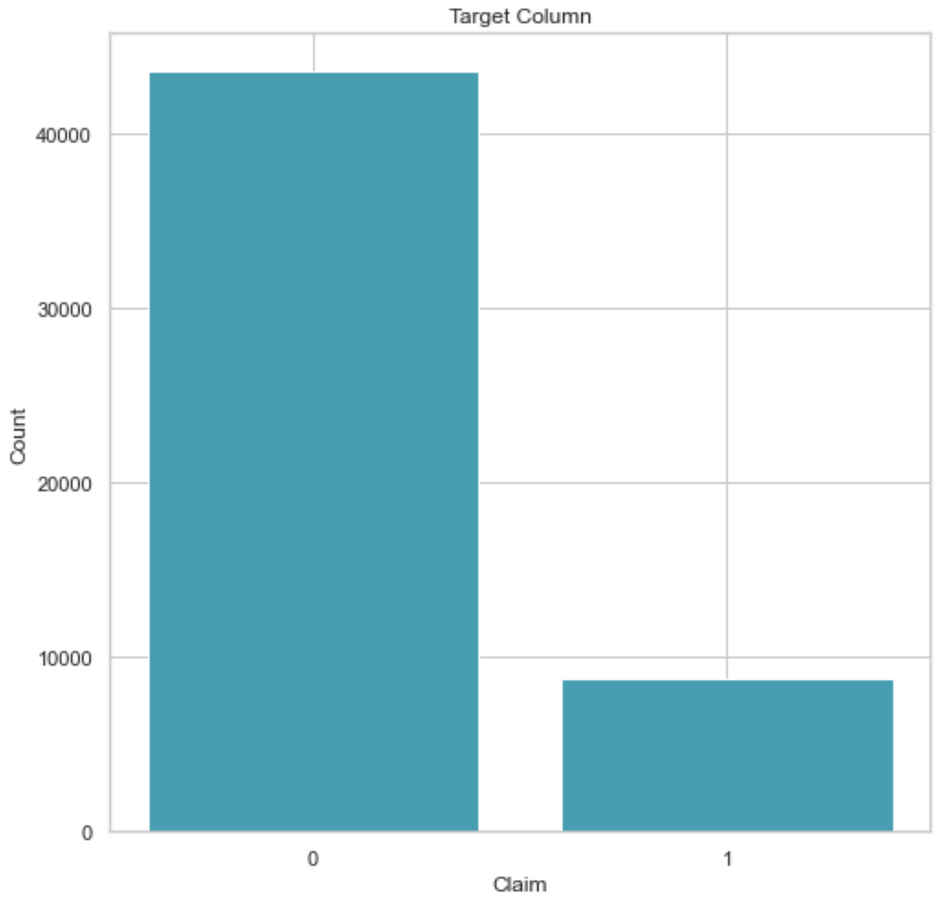
There were no missing values in the continuous features.

Null Value :

There were no null values in the given dataset.

Class Imbalance :

0	1
43590	8720



Feature Engineering

1. One Hot Encoding
2. Frequency Encoding

Approach 1

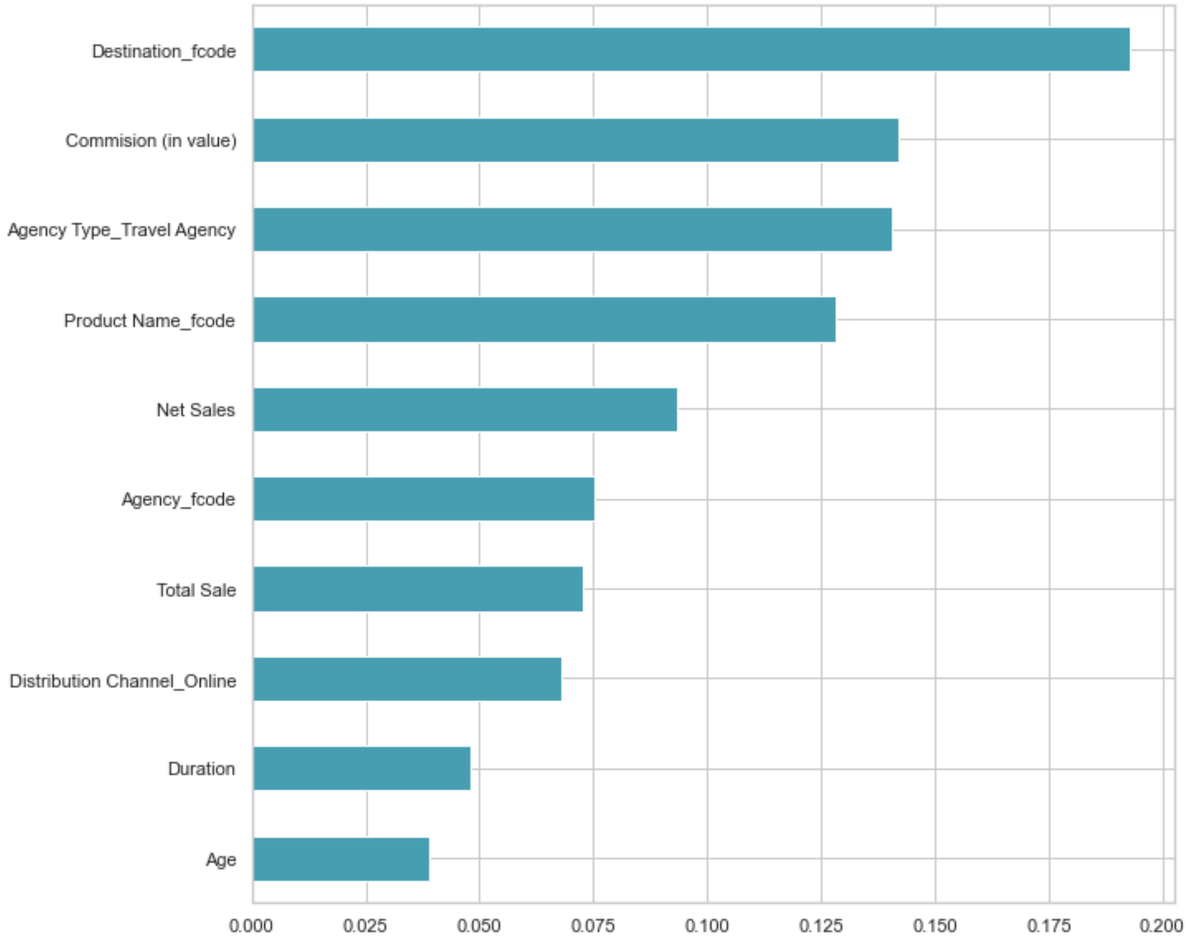
Column Name	Technique
Agency Type	One Hot Encoding
Distribution Channel	One Hot Encoding
Agency	Frequency
Product Name	Frequency
Destination	Frequency

Approach 2

Column Name	Technique
Agency Type	One Hot Encoding
Distribution Channel	One Hot Encoding
Agency	One Hot Encoding
Product Name	One Hot Encoding
Destination	One Hot Encoding

Feature Selection

1. Numerical Data (Co-relation)
2. Categorical Data (Chi Square)
3. RFE
4. Extra Trees Classifier for feature importance
5. XGBoost classifier with Grid search for feature importance



Model Selection

1. Logistic Regression
2. Decision Tree Classifier
3. Random Forest Classifier
4. XGBoost Classifier

Different Model Scores

Without Handling imbalance data and Hyper Parameter Tuning

Model	Precision	Recall
Logistic Regression	0.62	0.28
Decision Tree Classifier	0.82	0.76
Random Forest Classifier	0.72	0.69
XGBoost Classifier	0.8	0.74

Handling Imbalance Data and Hyperparameter Tuning

With Handling imbalance data and Hyper Parameter Tuning

Techniques to handle Imbalance data:

1. SMOTE
2. SMOTE Tomek

Hyper Parameter Techniques:

1. Grid SearchCV
2. Random Search CV

Different Model Scores

Model	Precision	Recall
Logistic Regression	0.78	0.78
Decision Tree Classifier	0.89	0.88
Random Forest Classifier	0.95	0.92
XGBoost Classifier	0.93	0.93

Final Confusion Matrix

Top 2 Algorithms :

1. Random Forest : 0.95.

2. XGBoost : 0.93

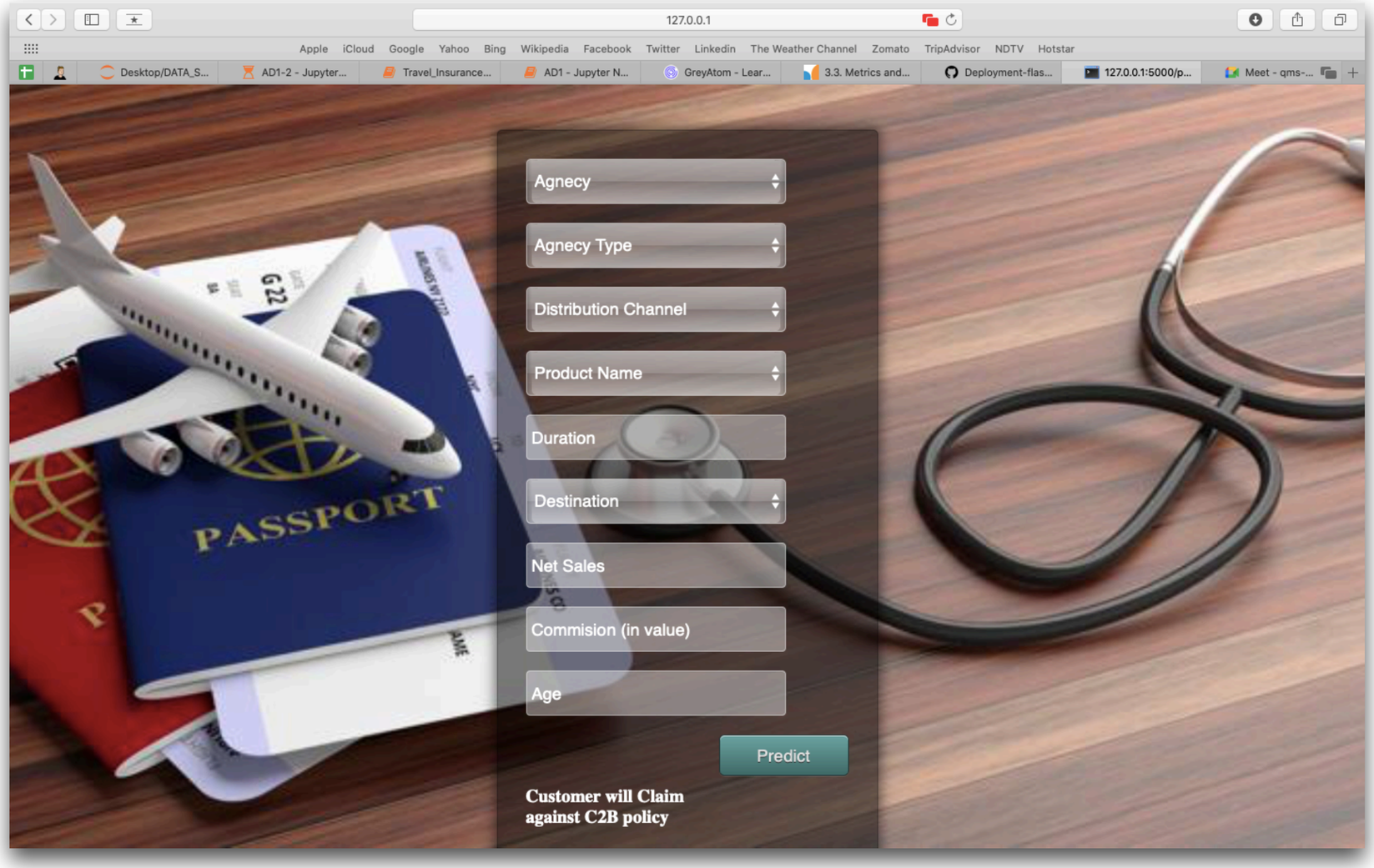
From the above observations it can be inferred that the best performing model was Random Forest giving the Precision score of **0.95**.

		Actual	
		Unclaimed	**Claim Registered**
Predicted	**Unclaimed**	12699	380
	Claim Registered	2057	557

Deployment

Steps:

1. Create Pickle (pkl) File.
2. Create web app in Flask
3. Integrate Pickle file in Flask App
4. Predict score with model



Recommendations

- More than 70% of the claims were done for customers in the age group of 20-40 years. Therefore, business should target Millenials and design products and offers targeting them.

- Approx 98% of the products were purchased online therefore we should increase our dominance in the digital space and try building new and growing existing presence on OTA platforms as well as our website.

- Top 3 highest claimed insurances are Annual plans. Premium of such plans can be increased.

- Airlines has 3 times more claim registered in comparison to travel. We can look forward to new / strengthen partnership opportunities(3rd party/OTA) in the airline space.

- Single usage products like Cancellation plans, 1 way Comprehensive plans are mostly unclaimed and in fact Cancellation plans overall has higher revenue generation. This product can have more targeted campaigns to increase business revenues.

Thank You For Your Patience!