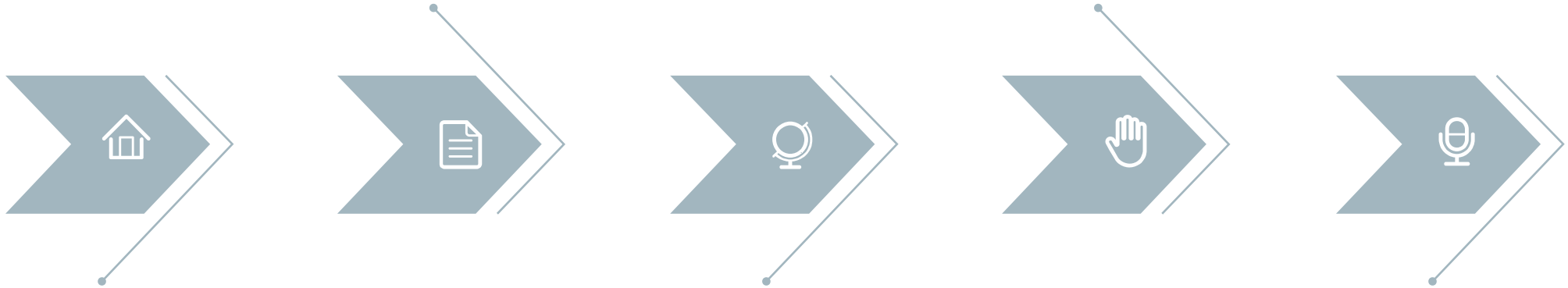
Travel Insurance Prediction Data

Predict Whether A Customer Will Be Interested In Buying Travel Insurance

Tiara Sekar Monica

CONTENTS



Business Background & Objective

Exploratory Data Analysis

Feature engineering

Modelling

Result Prediction

Business Background
&
Objective

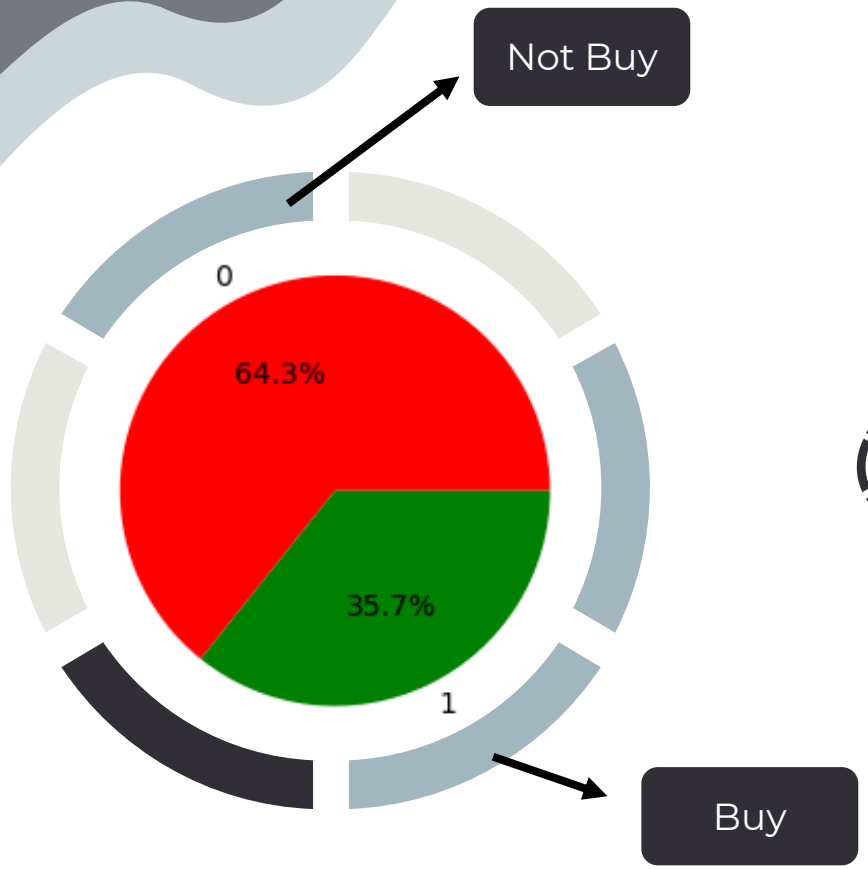
Business Background

A Tour & Travels Company is offering travel insurance package to their customers. The new insurance package also includes covid cover. The company requires to know the which customers will be interested to buy it based on its database history. The insurance was offered to some of the customers in 2019 and the given data has been extracted from the performance/sales of the package during that period.

The data is provided for almost 2000 of its previous customers. A Tour & Travel company requests to create a machine learning model that can predict if customers will be interested in buying a travel insurance package based on certain parameters.

Objectives

1. Does age affect customer interest in buying travel insurance?
2. How much income do customers who buy insurance have?
3. Does the number of family members affect the purchase of insurance?
4. Do customers with chronic diseases buy insurance?
5. What type of Employment buys insurance?
6. Does graduation affect the purchase of insurance?
7. Are customers who travel frequently are more likely to buy insurance?
8. Do people who frequently travel abroad buy more insurance?
9. How much annual income is obtained from each type of employment?



Imbalanced data in the target variable column

It can tell that majority of customers are not interested (class 0) in the service.

The data contains 1987 rows, 10 columns and 1 `TravelInsurance` binary column as target. Features contains 5 numerical features and 4 categorical features (Column `Unnamed: 0` has been dropped).

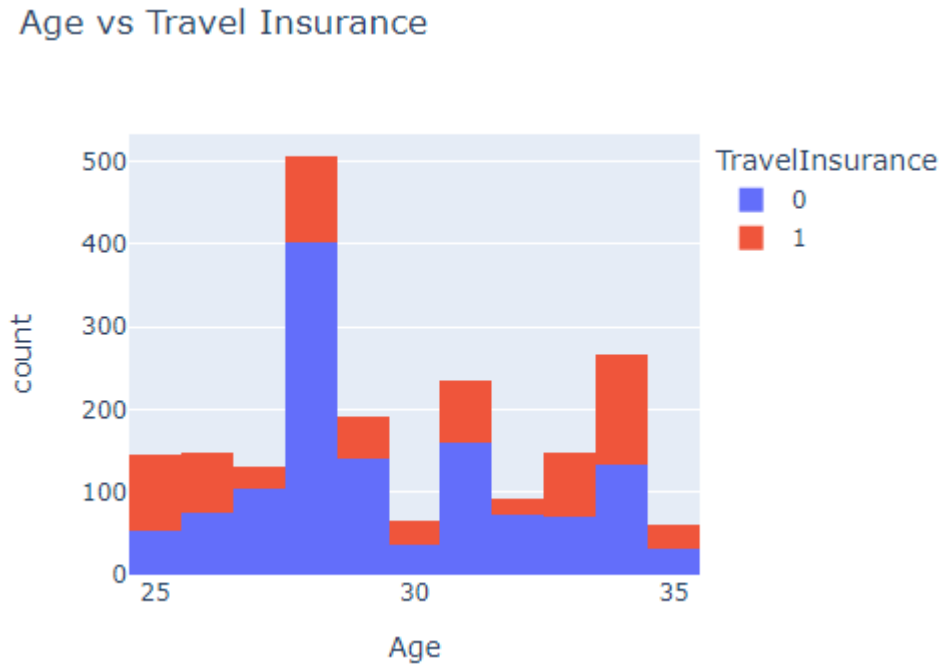
Numerical Features

- Age - Age Of The Customer
- AnnualIncome - The Yearly Income Of The Customer In Indian Rupees [Rounded To Nearest 50 Thousand Rupees]
- FamilyMembers - Number Of Members In Customer's Family
- ChronicDiseases - Whether The Customer Suffers From Any Major Disease Or Conditions Like Diabetes/High BP or Asthama,etc.
- TravelInsurance - Did The Customer Buy Travel Insurance Package During Introductory Offering Held In The Year 2019.

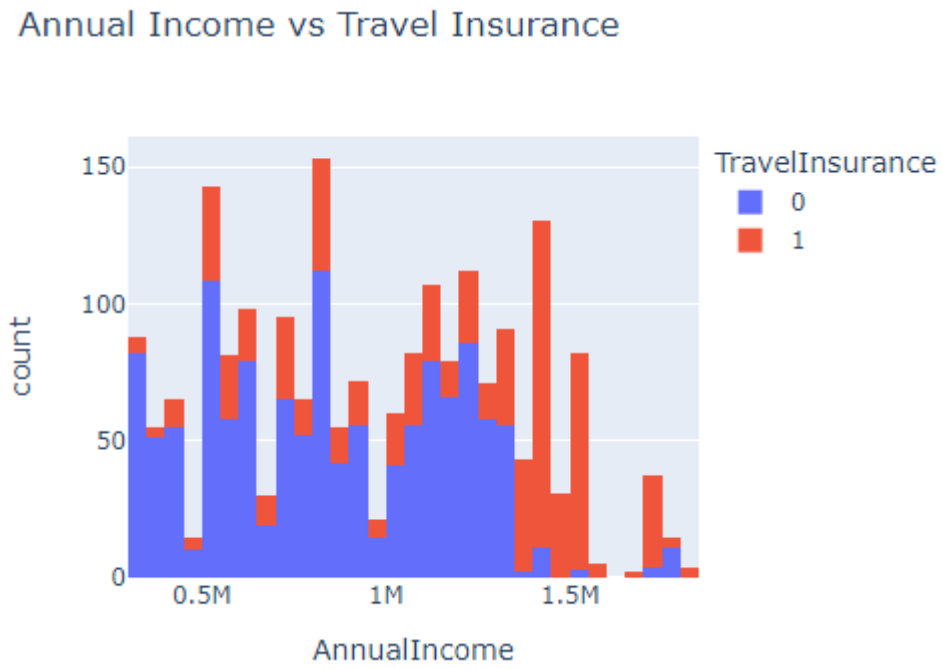
Categorical Features

- Employment Type - The Sector In Which Customer Is Employed
- FrequentFlyer - Derived Data Based On Customer's History Of Booking Air Tickets On Atleast 4 Different Instances In The Last 2 Years[2017-2019]
- EverTravelledAbroad - Has The Customer Ever Travelled To A Foreign Country[Not Necessarily Using The Company's Services]
- GraduateOrNot- Whether The Customer Is College Graduate Or Not

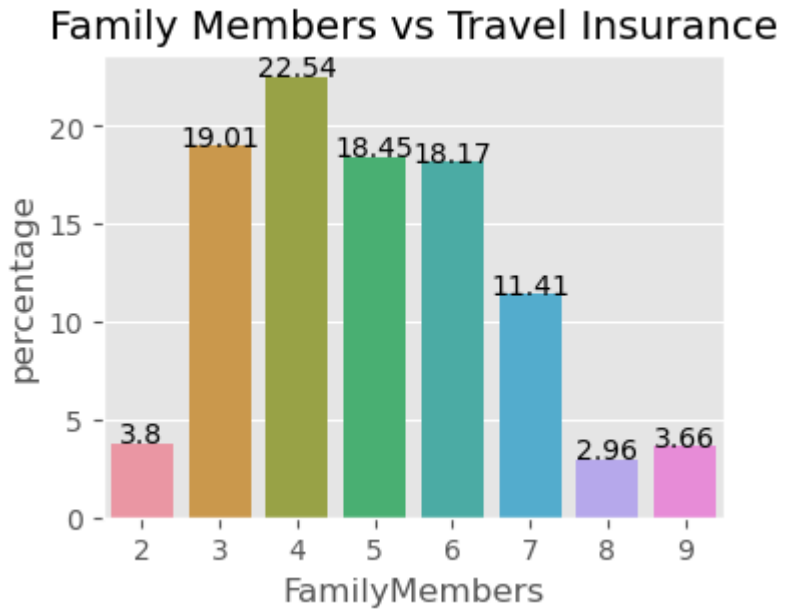
Exploratory Data Analysis



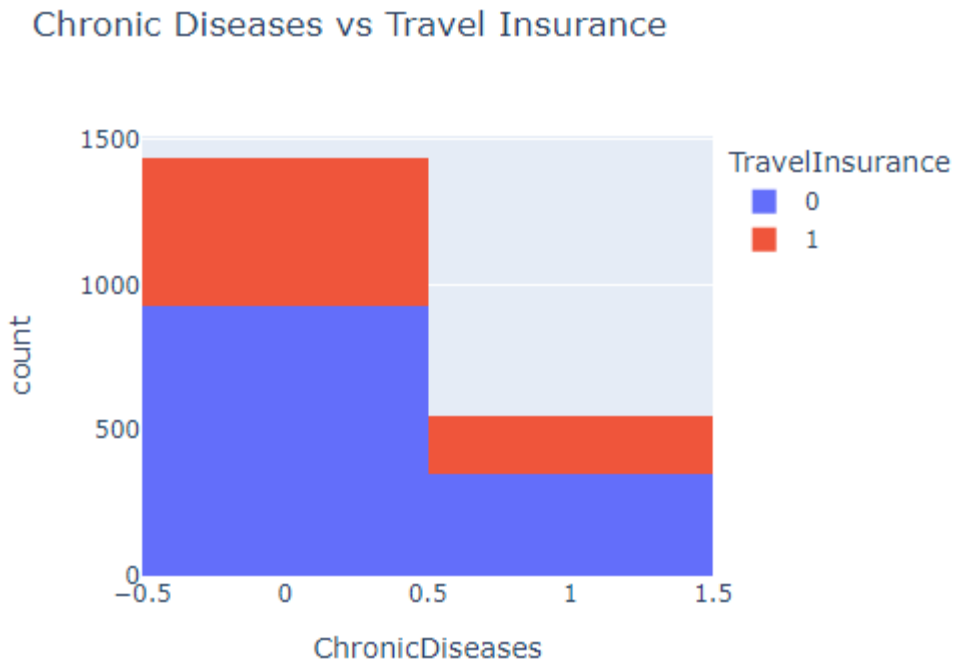
In the graph above, the 27, 29, 30, 32, and 35 age groups are not very interested in buying travel insurance. However, people other than that age are still interested in buying this insurance. It can be seen that age doesn't significantly affect the purchase of travel insurance.



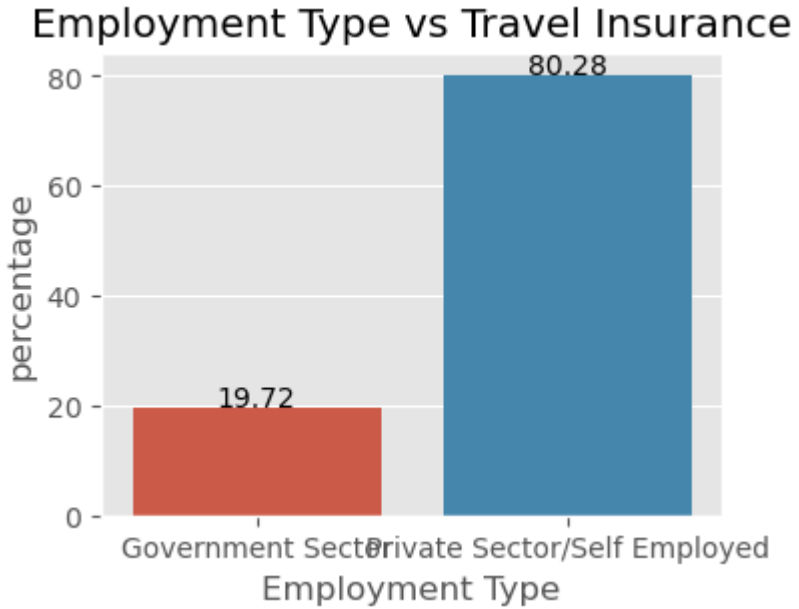
From the graph above it is clear that people whose annual income is greater than 1.3 million have purchased more travel insurance than people whose annual income is less than 1.3 million.



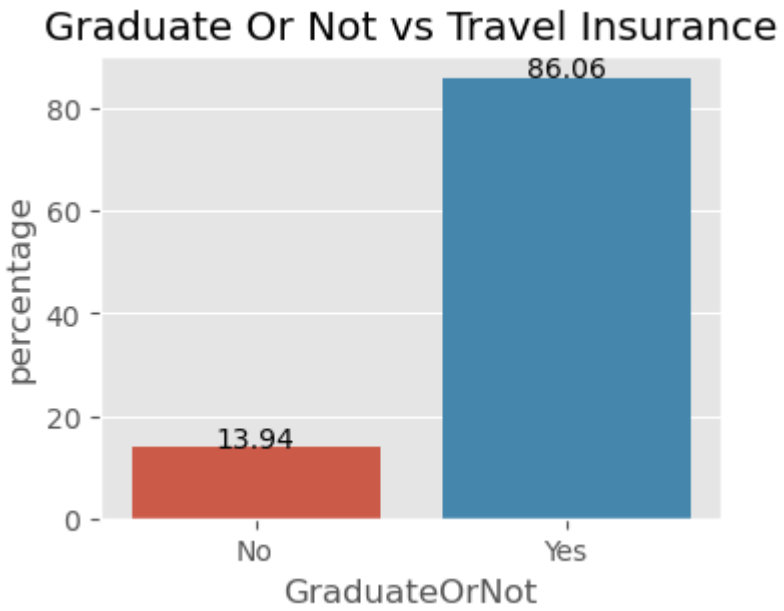
From the graph above it can be concluded that the number of family members doesn't significantly affect the purchase of travel insurance. In fact, families of 8 to 9 people don't buy travel insurance much more than other families.



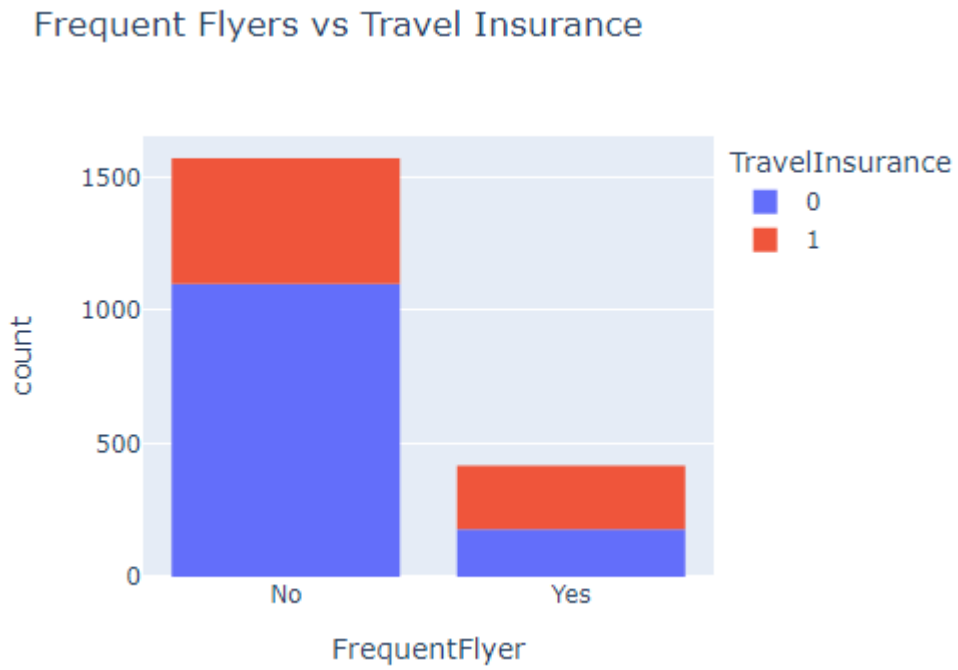
From the graph above it can be assumed that chronic disease has no impact on buying travel insurance, because based on the graph above it shows that many customers who have chronic disease don't buy insurance.



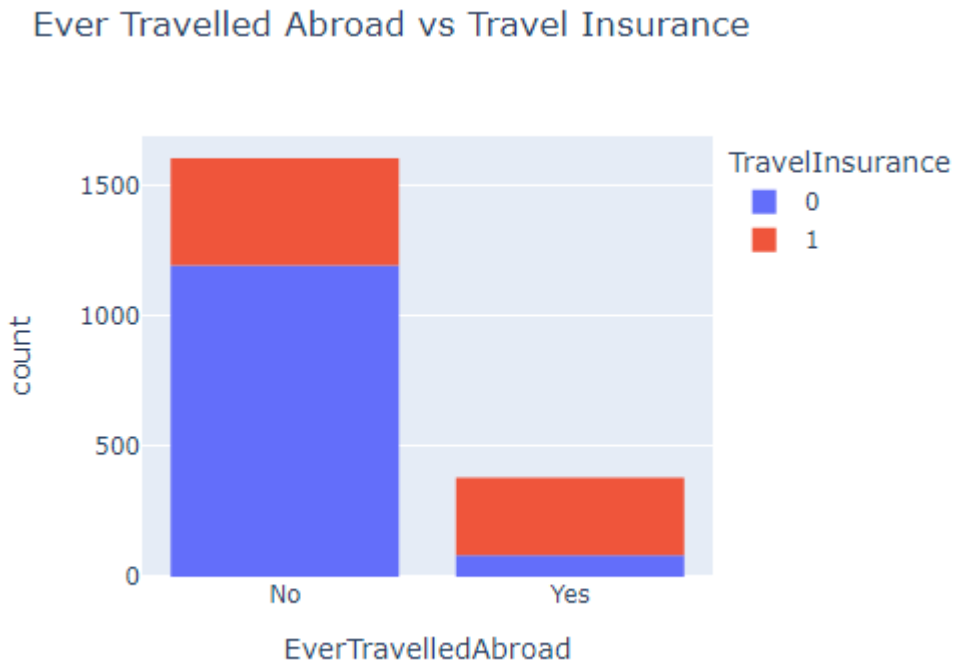
People who work in the government sector, the number of people who buy travel insurance is smaller with a percentage of 19.72% than in the private sector/self employed with a percentage of 80.28%.



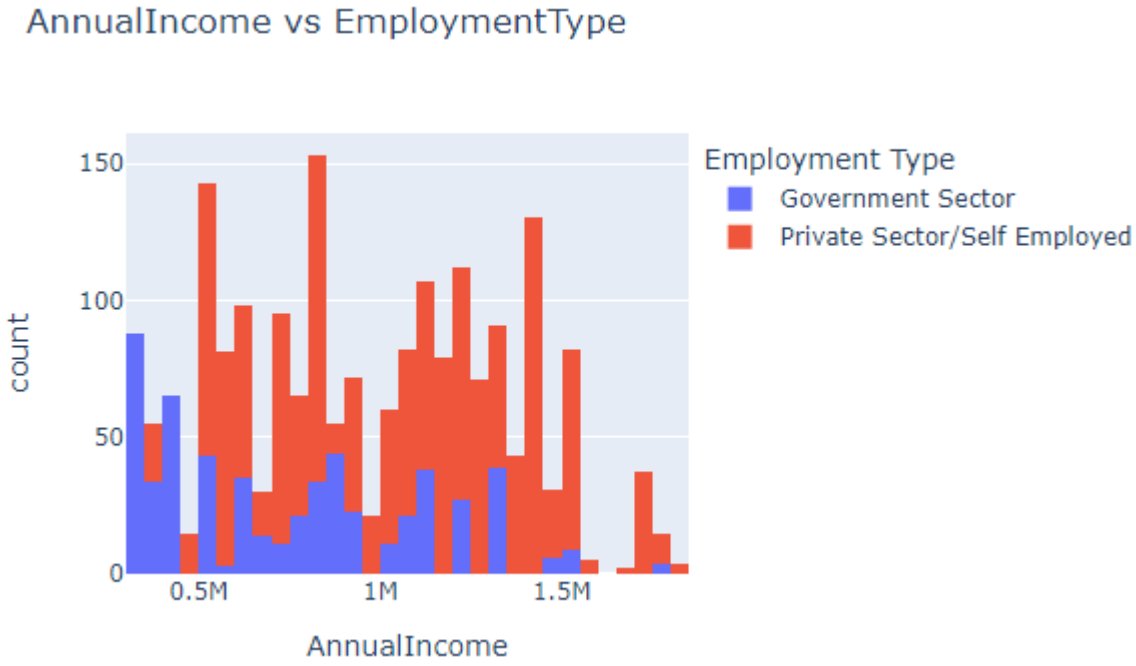
It can be seen that there are more graduates who buy insurance with a percentage of 86.06%.



From the results above it can be said that people who don't travel frequently are more likely to buy travel insurance while there are a small number of people who travel frequently but buy insurance.



From the graph above, it can be concluded that there are still a large proportion of people who travel abroad and buy insurance. In addition, the majority of people have not traveled abroad and have not purchased insurance. So, one thing can be assumed that people who have not traveled abroad don't need insurance when traveling within the country because they may find it less useful.



This graph shows that the government sector has a lower annual income, so one might expect that the government sector travels less or at least not far enough to consider insurance.

Feature Engineering

Check Missing Values

```
ID                   0
Age                  0
Employment Type      0
GraduateOrNot        0
AnnualIncome         0
FamilyMembers        0
ChronicDiseases      0
FrequentFlyer        0
EverTravelledAbroad  0
TravelInsurance      0
dtype: int64
```

No missing values

Check Duplicated Data

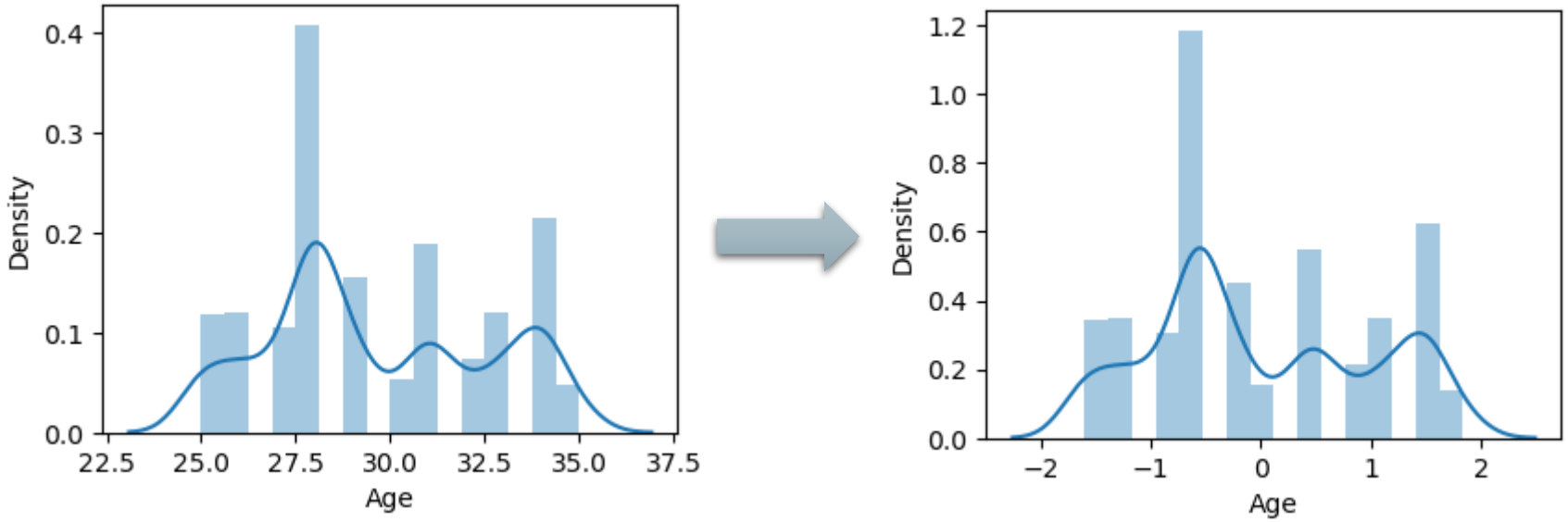
```
df_drop.duplicated().sum()
```

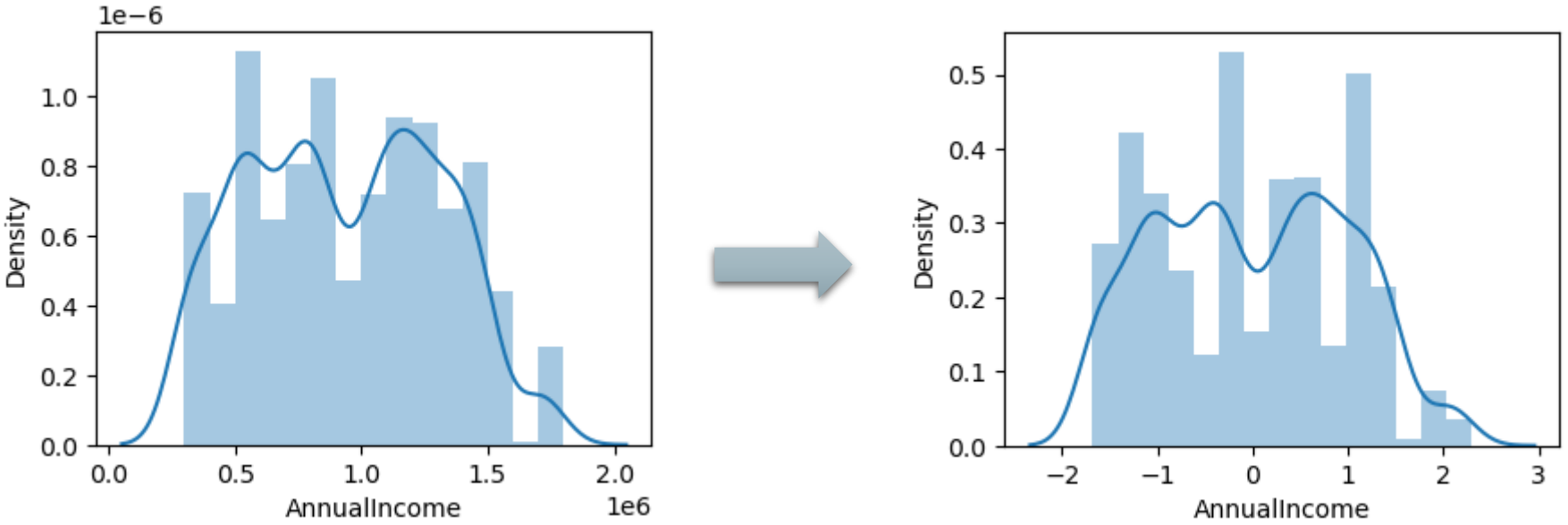
```
738
```

	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravelInsurance
40	25	Private Sector/Self Employed	Yes	1400000	3	1	Yes	Yes	1
44	25	Private Sector/Self Employed	Yes	1400000	7	0	No	Yes	1
46	25	Private Sector/Self Employed	Yes	1400000	4	0	No	Yes	1

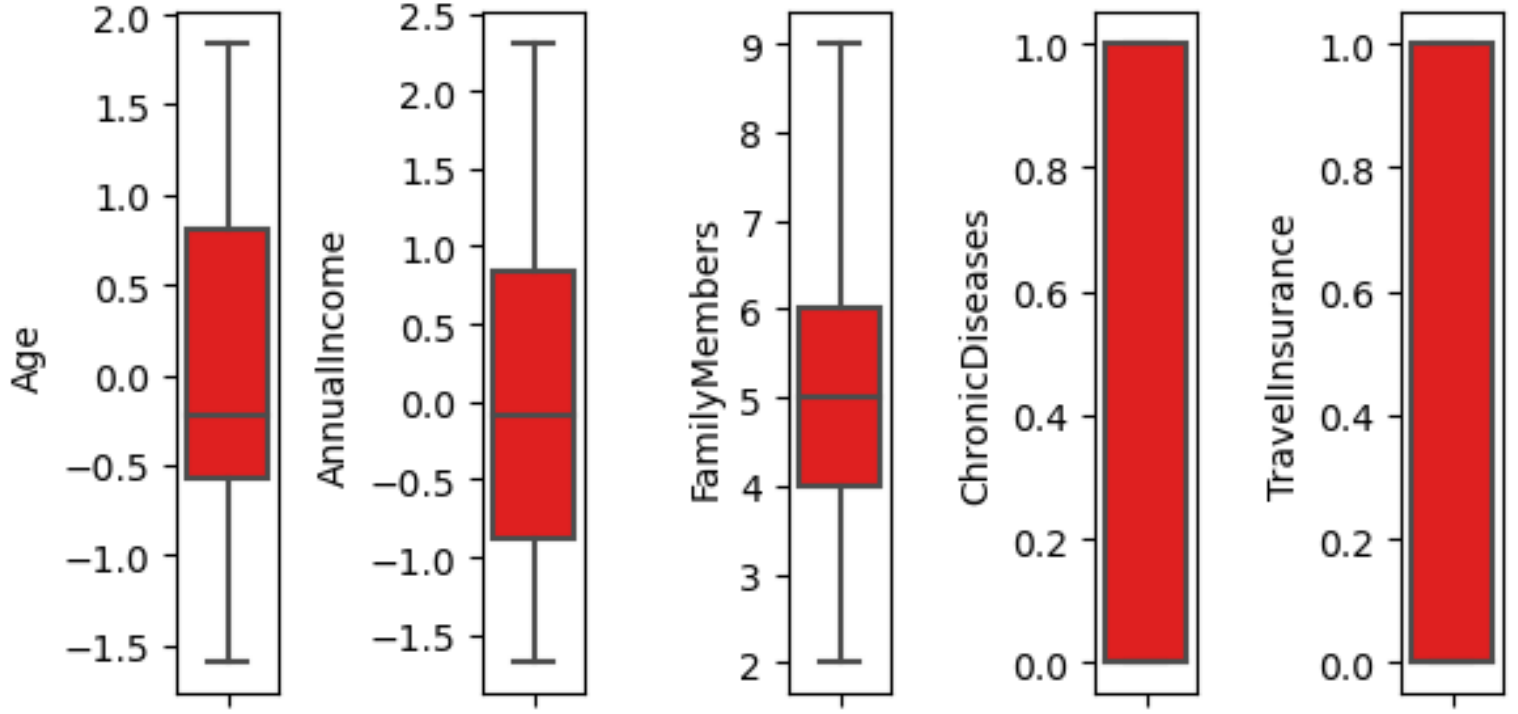
The data above shows that what is considered duplicate data is only from several columns with the same value, not all columns contained in the dataframe. Then the data doesn't need to be dropped.

Standardized





Boxplot to Detect Outliers



No outliers

Categorical Features

1. One Hot Encoding (OHE) :
   Employment Type
2. Dictionary Encoding :
   GraduateOrNot, FrequentFlyer,
   EverTravelledAbroad

Imbalanced Data

Handling Imbalanced Data with
Undersampling and Oversampling

Modelling

Baseline Modelling

The dataset becomes Train and Test data with a proportion of 75:25

Baseline Modelling				
Model Name	**Accuracy**	**Precision**	**Recall**	**F1-Score**
KNN	83%	92%	57%	71%
Logistic Regression	78%	81%	53%	64%
Decision Tree	77%	72%	61%	66%
Random Forest	80%	77%	66%	71%

Handling Imbalance Model and Tunning Results

Handling Imbalance Model and Tunning Results				
Model Name	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Under KNN	79%	74%	64%	69%
Over KNN	77%	69%	66%	67%
Smoteenn KNN	81%	81%	61%	69%
SMOTETOmek KNN	77%	70%	65%	67%
Under Logistic Regression	75%	66%	64%	65%
Over Logistic Regression	75%	67%	63%	65%
Smoteenn Logistic Regression	77%	71%	62%	66%
SMOTETOmek Logistic Regression	74%	64%	65%	64%
Under Decision Tree	73%	63%	65%	64%
Over Decision Tree	78%	71%	66%	69%
Smoteenn Decision Tree	80%	76%	67%	71%
SMOTETOmek Decision Tree	78%	74%	63%	68%
Under Random Forest	75%	64%	69%	66%
Over Random Forest	80%	75%	68%	72%
Smoteen Random Forest	83%	83%	66%	73%
SMOTETOmek Random Forest	78%	71%	67%	69%
Optuna Smoteen Random Forest	84%	91%	61%	73%

The dataset becomes Train and Test data with a proportion of 75:25

Hyperparameter Tuning using Optuna Random Forest Smoteen is best model

Confusion Matrix from Tuning Optuna Random Forest Smoteenn

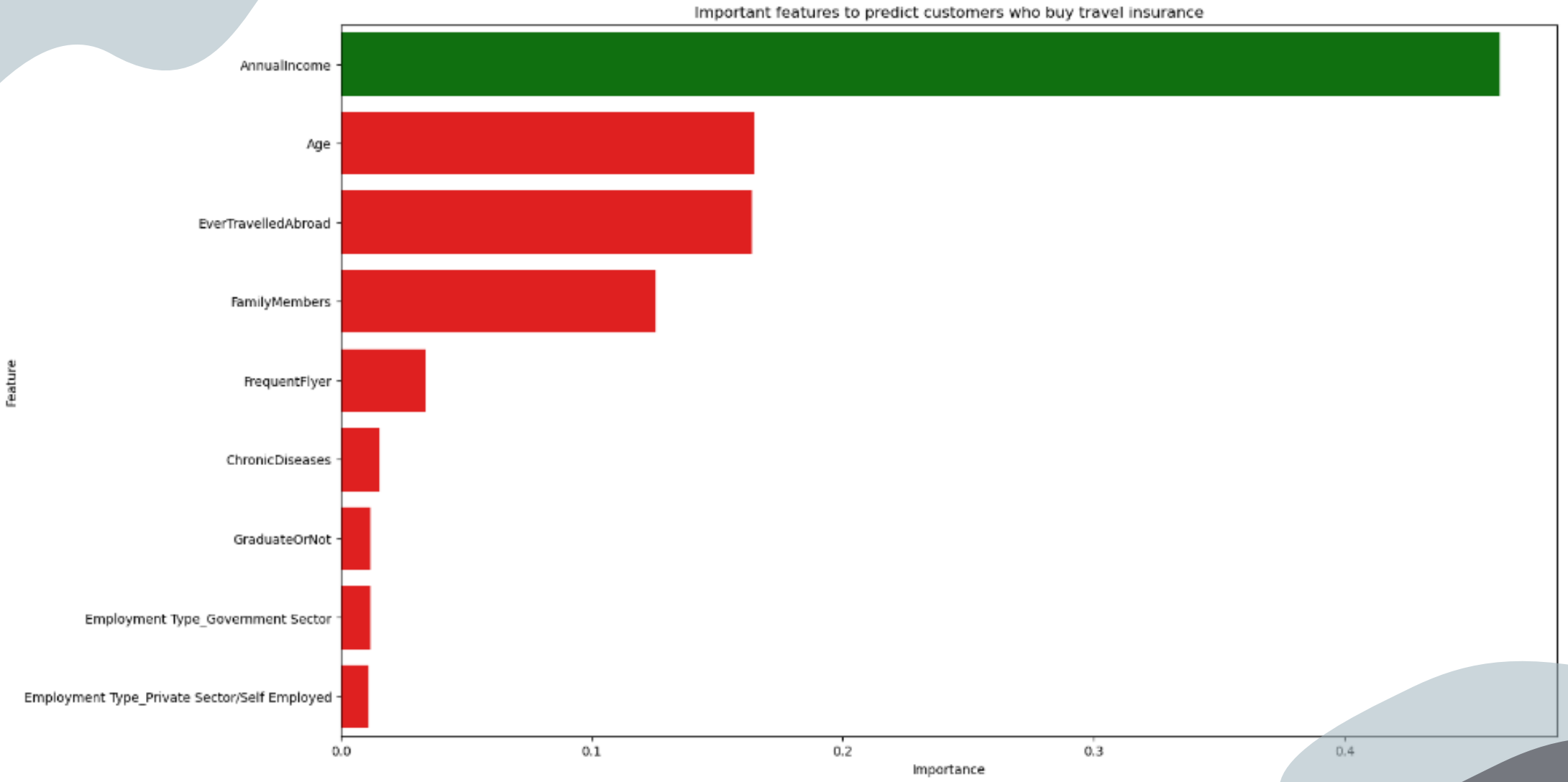
1. True Postive : 110
2. True Negative : 306
3. False Positive : 11
4. False Negative : 70

Result Prediction

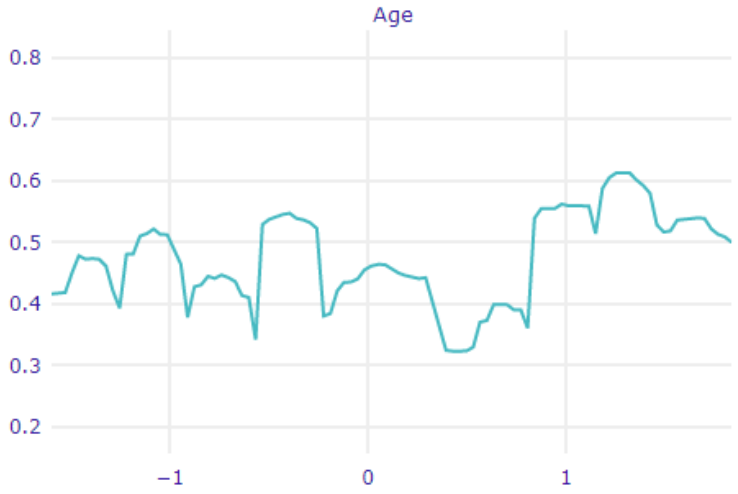
Sample Prediction
Results Best Model Hyperparameter
Tuning using Optuna Random Forest Smoteen

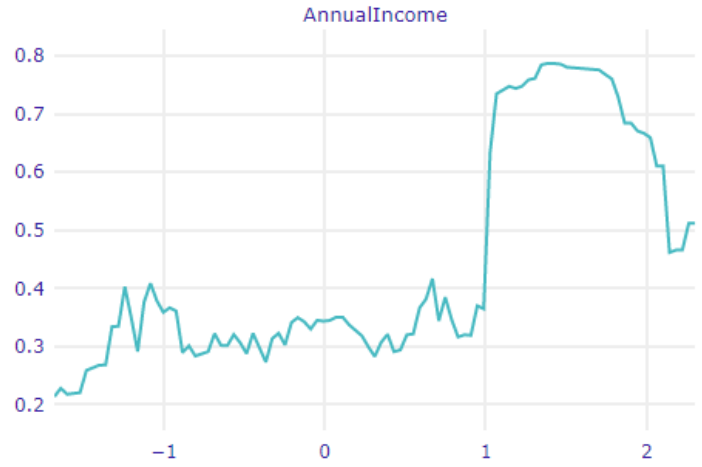
	ID	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	Prediction results for purchasing travel insurance
114	1315	25	Private Sector/Self Employed	No	1150000	3	0	No	No	0
381	877	26	Private Sector/Self Employed	Yes	1400000	3	0	No	Yes	1
460	546	28	Private Sector/Self Employed	Yes	550000	3	0	No	No	0

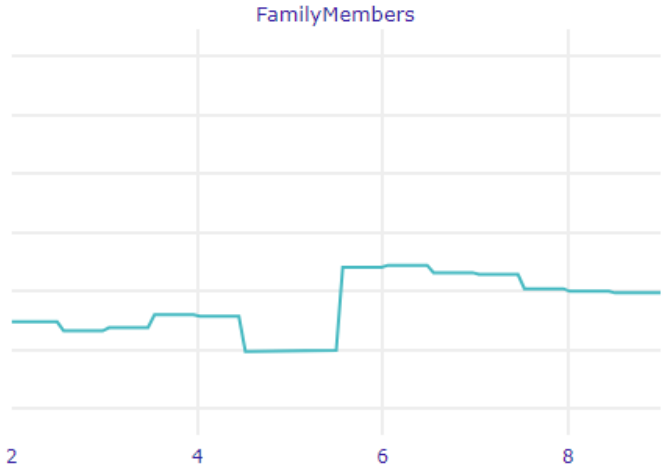
Feature Importances

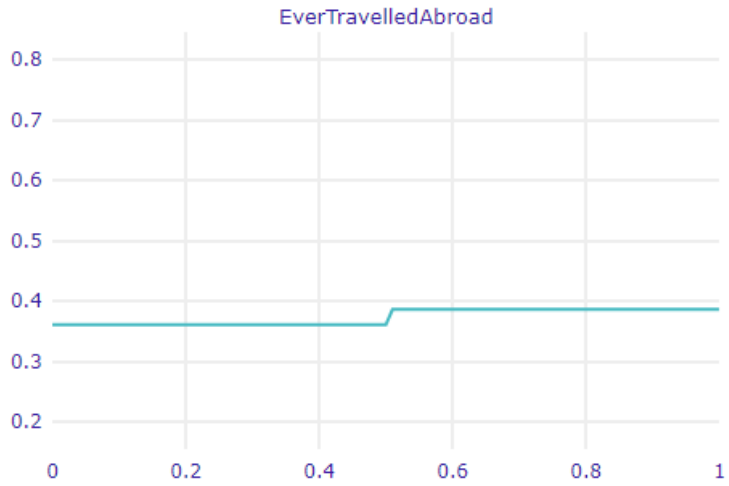


Explainability Model









Conclusion

From the observations made, it can be concluded that the best model for predicting whether a customer buys a travel insurance package or not is Hyperparameter Tuning using Optuna Random Forest Smoothen, because it has the best precision value of 91%.

Recommendation

1. The Company can make new offers, travel insurance packages at more suitable prices if needed, to people with incomes below 1.3 million such as people who work in the government sector
2. The Company need to do marketing and provide information about the packages provided
3. The company can get a smaller insurance package for ordinary travel or a coupon for users so that those who don't buy a travel insurance package can be enticed to buy it

Thank you

Gmail tsmon44@gmail.com

GitHub https://github.com/tsmon44

LinkedIn https://www.linkedin.com/in/tiara-sekar-monica/