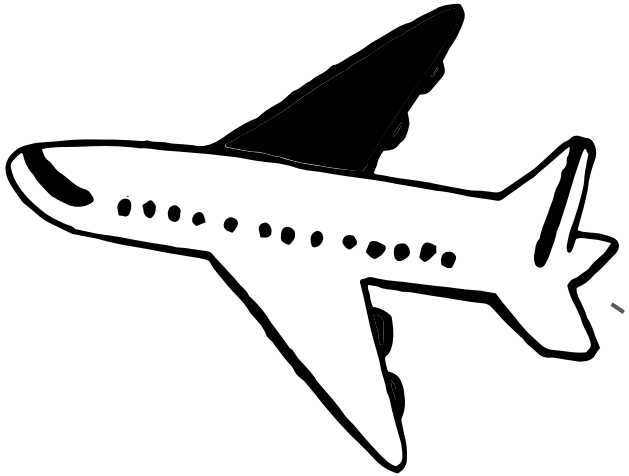
Travel Insurance

By Looping Group

PT Looping Tour & Travel

Just enjoy the trip.



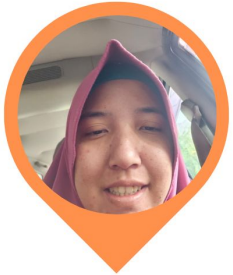
PT Looping Insurance is a Tour & Travel Company. We committed always provide protection on unexpected incidents which occur during travel (domestic and abroad).

Data Science Team

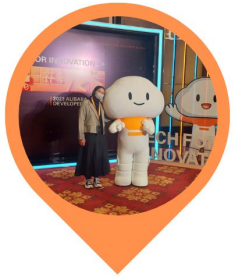
ANDI

ARYAN

IMAM



SEPTI



SITI

ROSIANA

VICKY

CURRENT ISSUE

PT Looping Tour & Travel want to increase buy rate of Travel Insurance and evaluate the performance of our product based on database history. The company need to know which customer which customer that likely to buy Travel Insurance.

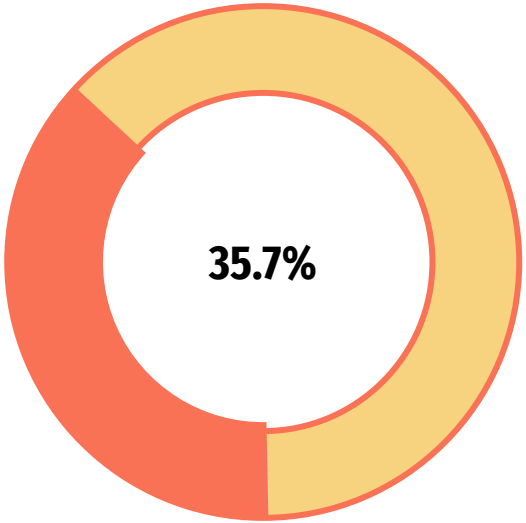
As a Data Scientist, we are responsible to build an intelligent model that can predict if the customer will be interested to buy the Travel Insurance.

TABLE OF CONTENTS

01

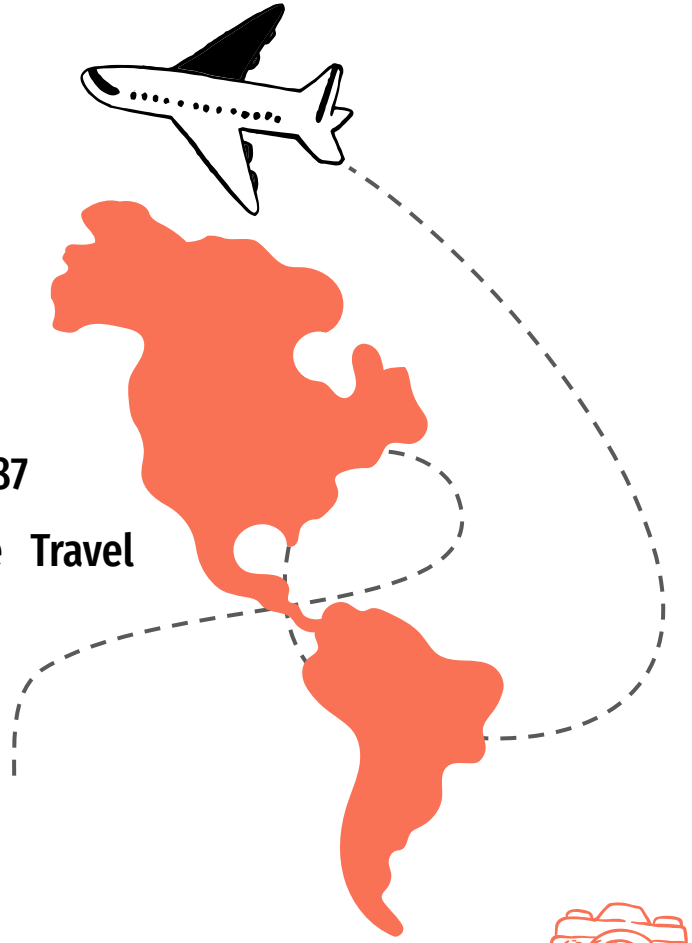
BUSINESS UNDER STANDING

What is the Problems?



ONLY

35.7% of the total 1987 customer that buy the Travel Insurance



710	1277
Buy Travel Insurance	Not Buy Travel Insurance

What is the Problems?

GOALS

Increasing buy rate at least 20%

OBJECTIVES

- Create a machine learning that can predict potential customer
- Determine customer segmentation based on the most influential features

BUSINESS METRICS

Buy rate

02

EXPLORATORY DATA ANALYSIS (EDA)

FEATURES

UNNAMED: 0

AGE

Age Of Customer

EMPLOYMENT TYPE

The Sector In Which Customer Is Employed

GraduateOrNot

Whether The Customer Is College Graduate Or Not

AnnualIncome

The Yearly Income Of The Customer

FamilyMembers

Number Of Members In Customer's Family

ChronicDisease

Whether The Customer Suffers From Any Major Disease Or Conditions

FrequentFlyer

Derived Data Based On Customer's History Of Booking Tickets

EverTravelled Abroad

Has The Customer Ever Travelled To A Foreign Country

TRAVELINSURANCE

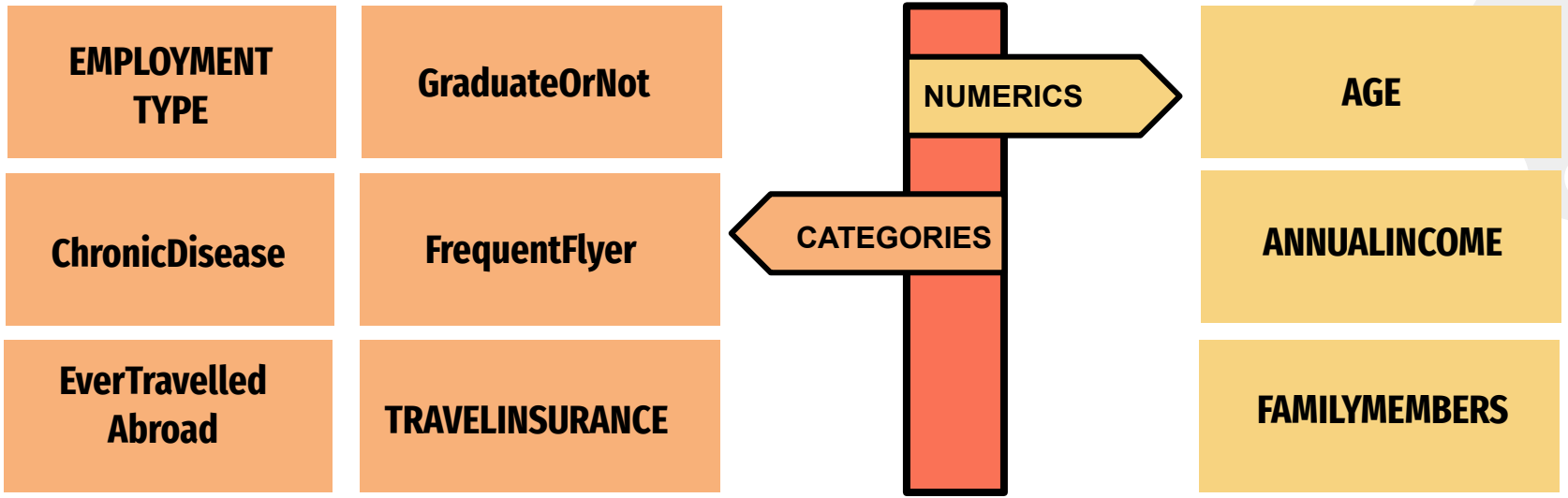
Did The Customer Buy Travel Insurance Package

Missing Data & Shape

- From data.info there is nothing missing data in each columns
- Shape data: 1987 rows & 10 columns (9 features and 1 target)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1987 entries, 0 to 1986
Data columns (total 10 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Unnamed: 0           1987 non-null   int64
 1   Age                  1987 non-null   int64
 2   Employment Type      1987 non-null   object
 3   GraduateOrNot        1987 non-null   object
 4   AnnualIncome         1987 non-null   int64
 5   FamilyMembers        1987 non-null   int64
 6   ChronicDiseases      1987 non-null   int64
 7   FrequentFlyer        1987 non-null   object
 8   EverTravelledAbroad  1987 non-null   object
 9   TravelInsurance      1987 non-null   int64
dtypes: int64(6), object(4)
memory usage: 155.4+ KB
```

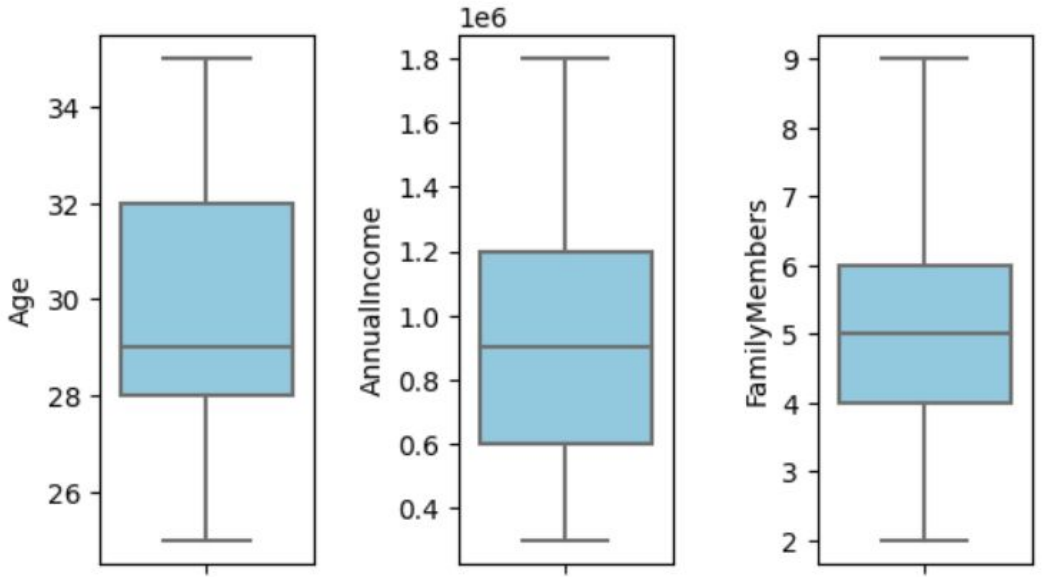
NUMERIC & CATEGORY FEATURES



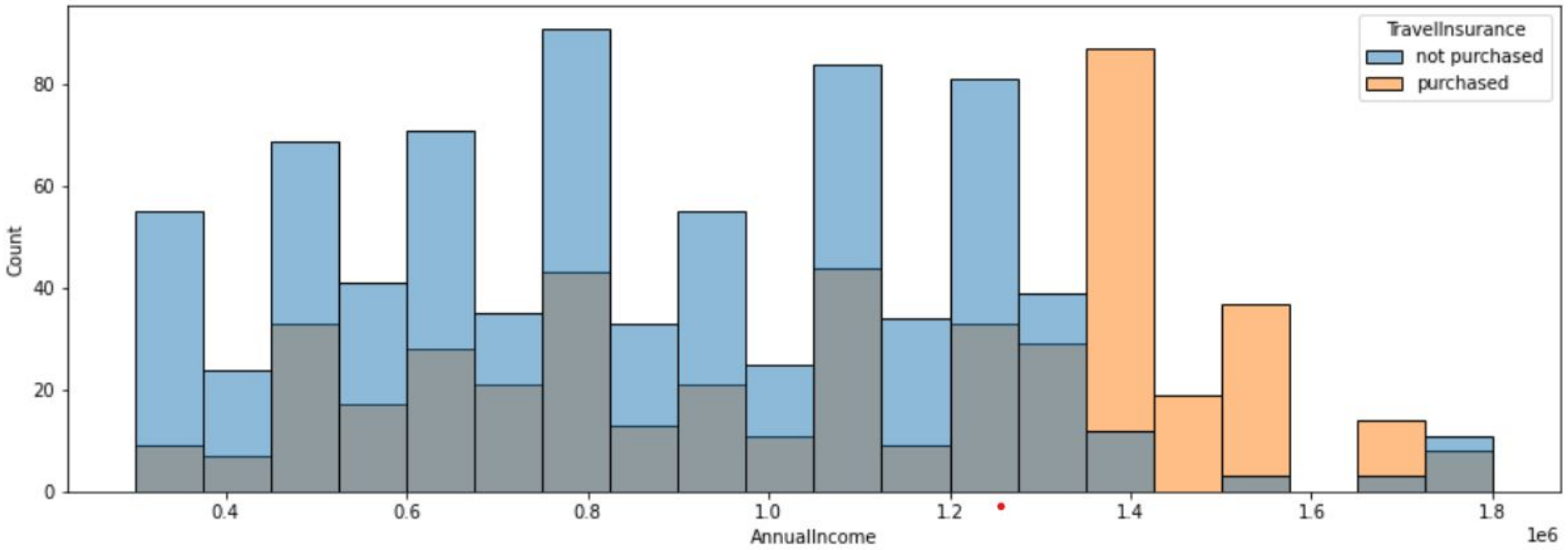
CHECK OUTLIER

OUTLIER

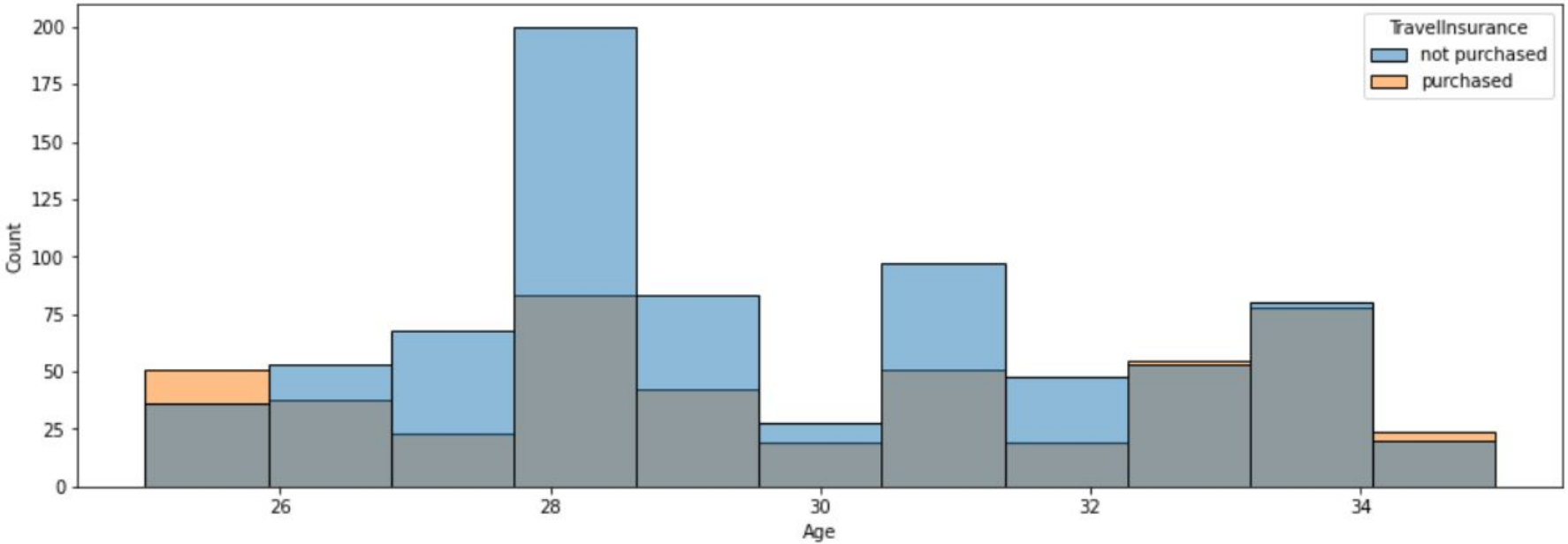
We can look there are nothing outlier from distribution numerical data (Age, Annualincome, Familymembers) with Boxplot



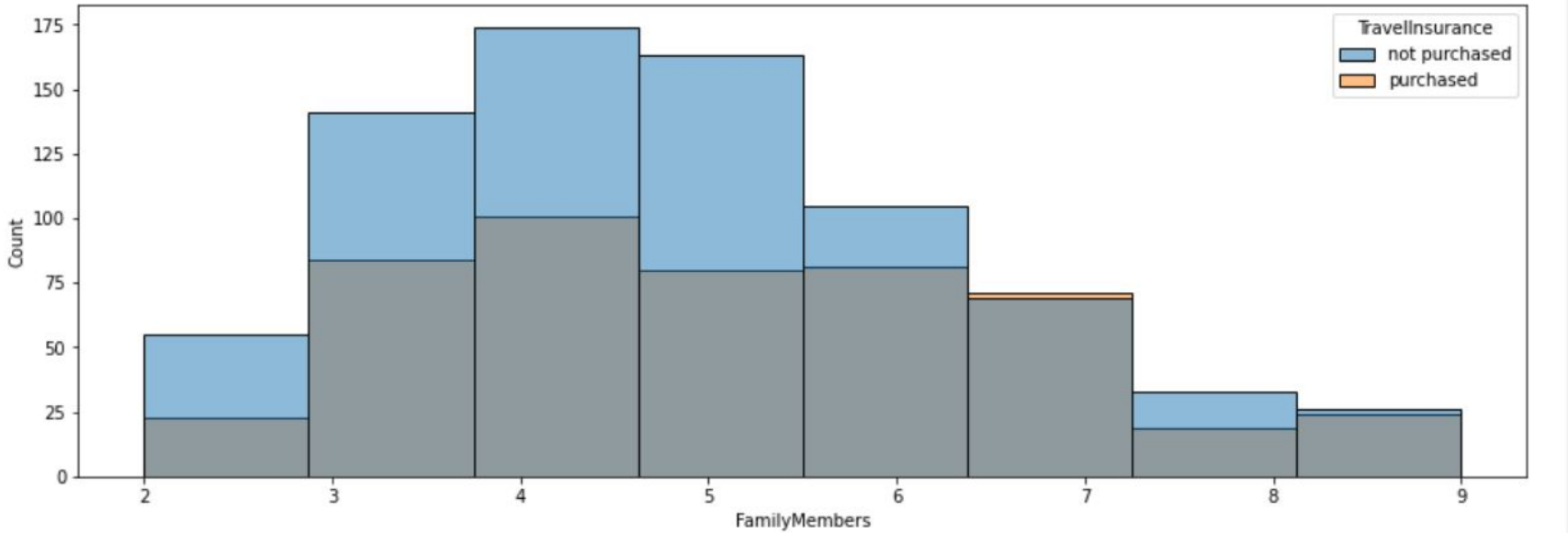
ANNUAL INCOME TO TRAVEL INSURANCE



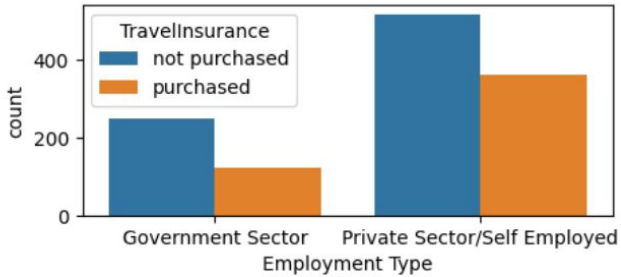
AGE TO TRAVEL INSURANCE

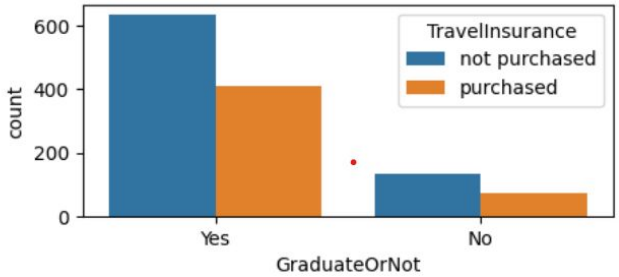


FAMILY MEMBERS TO TRAVEL INSURANCE

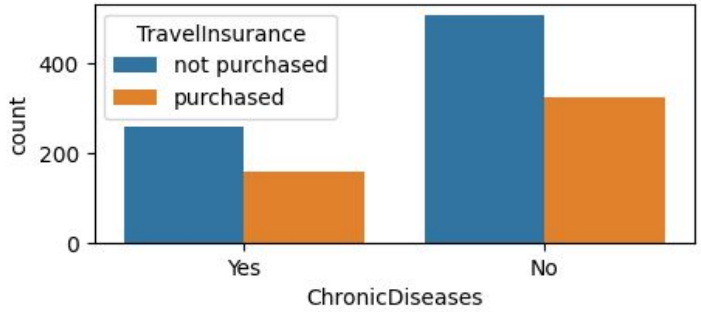


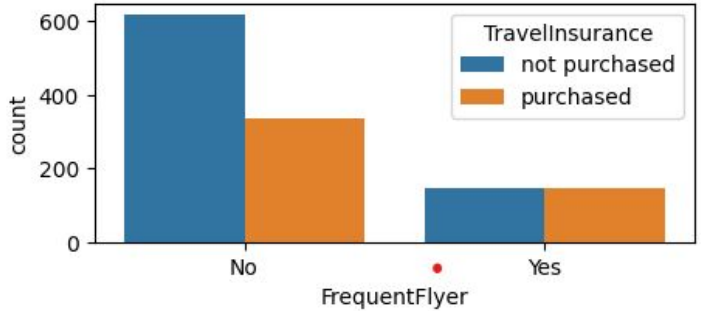
CATEGORY FEATURES DISTRIBUTION

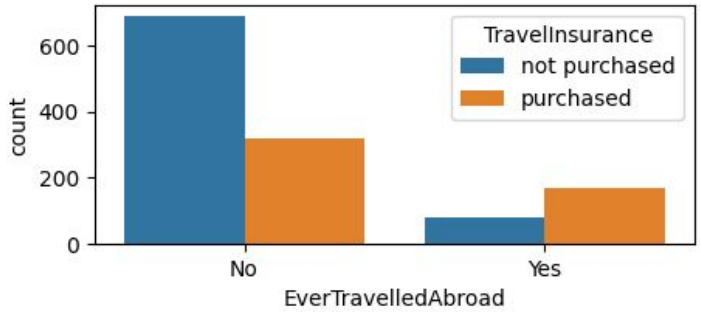




Based on the graph above, the type of work and experience of traveling abroad affects the level of travel insurance ownership





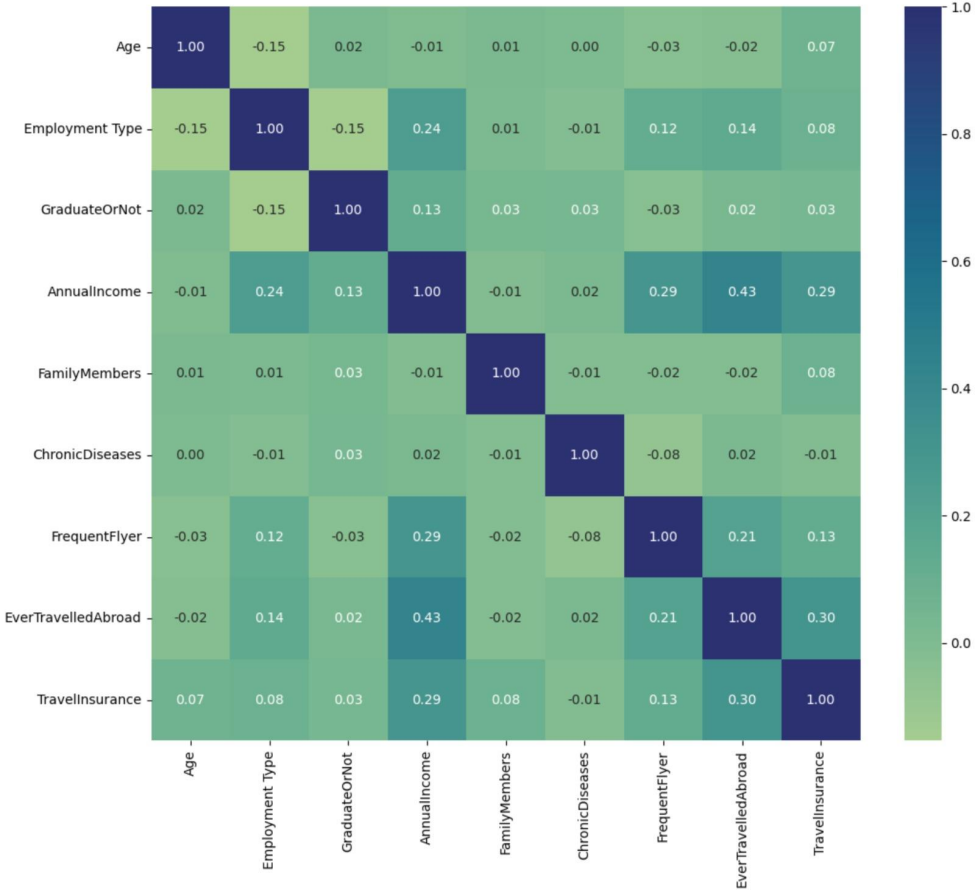


Insight from EDA

If seen from the distribution of numerical data, the majority of people who do not use travel insurance are people with an age range of 26-30 years, who have family members in the range of 3-5 people, and income below 1,500,000 per year (possibly in dollars).

Meanwhile, judging from the distribution of categorical data, the majority of people who do not use travel insurance are in the undergraduate category, work in the private sector, do not have chronic illnesses, do not travel frequently, and have never gone abroad.

CORRELATION FEATURES



➔ Variables that have a high correlation with the target variable "TravelInsurance": "EverTraveledAbroad", "FrequentFlyer" and "AnnualIncome'. Another variable that has a significant correlation with the target variable is "Employee type", "Age" and "FamilyMembers".

➔ Variables that have a low correlation or even no correlation with the target variable, such as "GraduateorNot" and "ChronicDiseases" can be considered less relevant. But it is necessary to do a more in-depth analysis first to ensure that these variables are indeed irrelevant in the modeling so that these two variables are still used.

03

DATA PRE-PROCESSING

Data Prepocessing

Handle Missing Value

There are no missing values

Handle Duplicated

There were 738 duplicate data after dropping the Unnamed feature. Therefore the duplicated data is deleted so that the remaining 1249 rows of data

Handle Outlier

There are no outliers

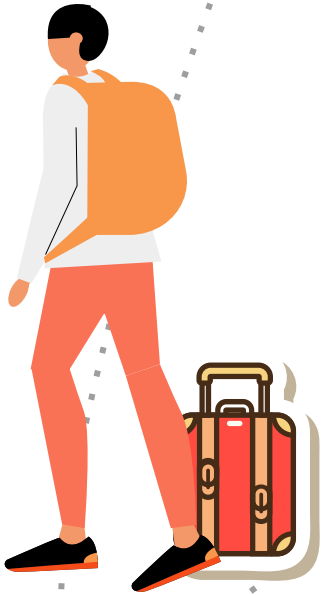
Data Prepocessing

Feature encoding

Label Encoder

Split Data

70 : 30
Train : Test



04

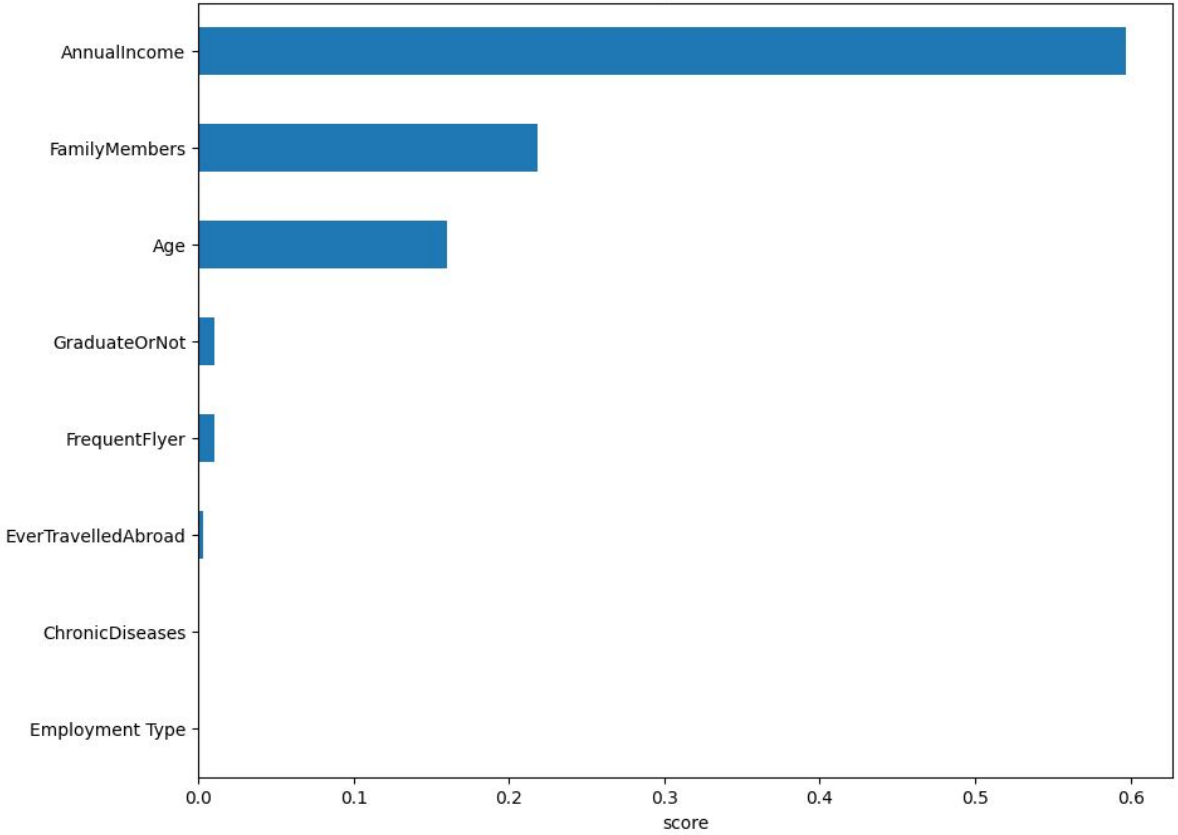
MODELING

Modeling Evaluation

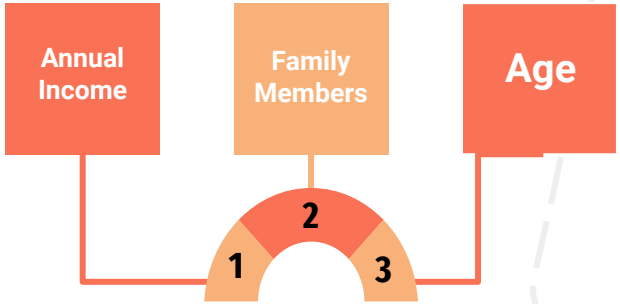
	Gradient Boosting	Ada Boost	Random Forest	Quadratic Discriminant Analysis
Accuracy	0.80	0.79	0.66	0.75
Precision	0.91	0.91	0.56	0.77
Recall	0.52	0.49	0.52	0.50
ROC AUC (Test)	0.73	0.73	0.68	0.74
ROC AUC (Train)	0.72	0.75	0.94	0.71

Based on testing result, model that has best performance is **Gradient Boosting**.

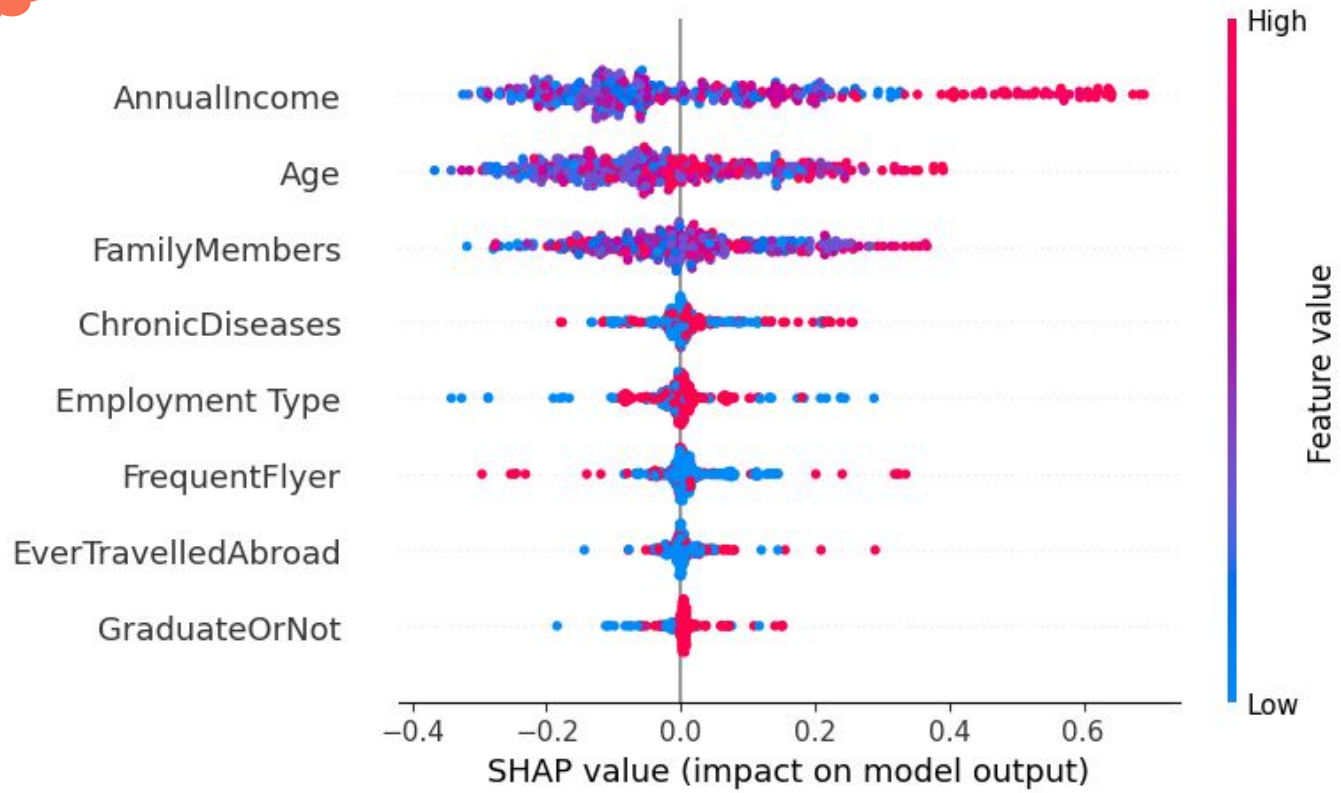
Feature Importance



The sequence of **features** that most influence the modeling results



SHAP Value


AnnualIncome
Age
FamilyMembers
ChronicDiseases
Employment Type
FrequentFlyer
EverTravelledAbroad
GraduateOrNot
−0.4
−0.2
0.0
0.2
0.4
0.6
SHAP value (impact on model output)
High
Feature value
Low

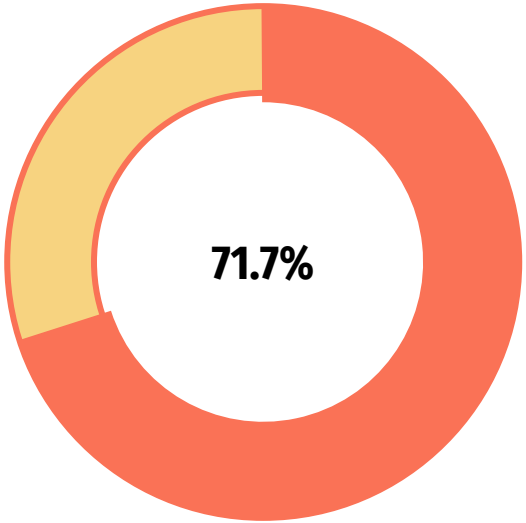
05

BUSINESS RECOMMENDATION

Business Recommendation

- Prioritizing customers based on high Annual Income
- Make special offers for customer who never buy Travel Insurance
- Collaborate with Airline companies to buy airplane tickets bundling with Travel Insurance
- Add new features to improve Machine Learning Modeling, such as:
  - Gender
  - Destinations (Foreign/Domestic)
  - Travel Duration
  - Travel Time (Holiday Season/Not)

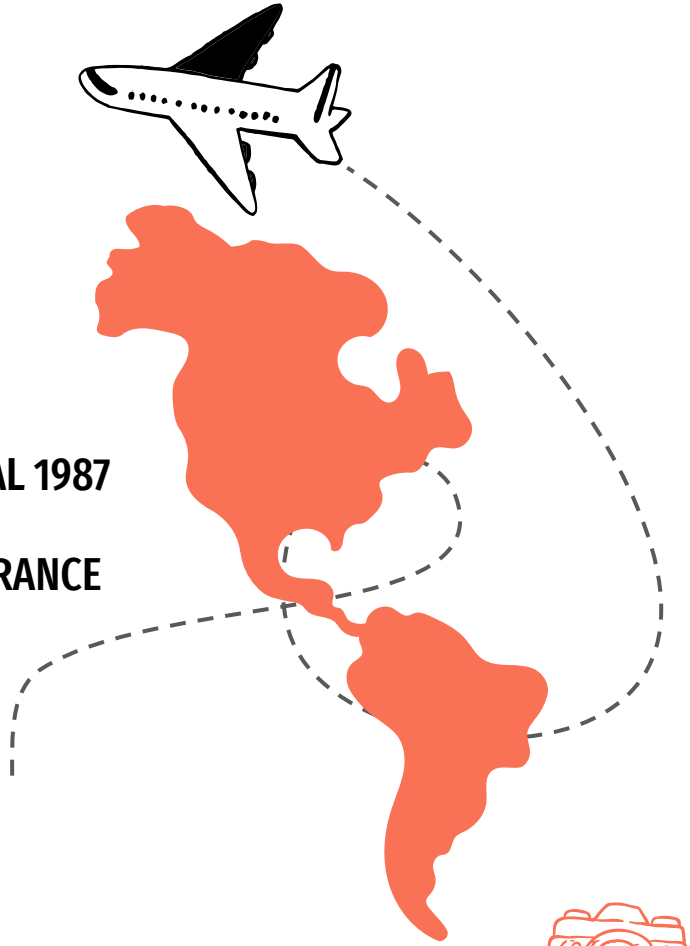
Prediction Result



INCREASE 36%

35.7% ⇨ 71.7% OF TOTAL 1987

CUSTOMERS BUY TRAVEL INSURANCE



1425	562
Buy Travel Insurance	Not Buy Travel Insurance


THANK YOU