## Data Collection

CSV File

Kaggle Competitions



Business Understanding



What type of customers buy travel insurance?

### Problem Identification



How to utilize Data Science Techniques to Predict potential customers that will buy travel insurance? Perform Exploratory Analysis to get Insight about the Data and find underlying patterns:

- > Exploring frequency of purchases
- > Look at **Age** distribution and percentage of purchases between Age Groups



- > Find Average Household **Income** of potential buyers
- > Effect of **customer's graduation** status on purchasing travel insurance
- > Analyze whether being **Frequent Flyer** or having **Chronic Diseases** effects probability to purchasing travel insurance.

## Feature Engineering and Pre-Processing

> Create a feature called wealth\_groups derived from household income



Splitting the Dataset



- > Class imbalance was not observed
- Dataset was splitted into75% Training and 25% Validation.

# Model and Evaluate Performace

Algorithms picked for Model Prediction:

- 1) Decision tree
- 2) AdaBoost Decision Tree
- 3) Random Forest
- 4) Logistic Regression
- 5) Neural Network
- 6) K-Nearest Neighbors
- 7) Multinomial Naive Bayes
- 8) Linear Discriminant Analysis
- 9) Ensemble Voting Classifier

## Hyper-Parameter Tuning

Below are some of the Hyper-Parameter Tuning Performed for each model to maximize F1-Score.

#### Decision Tree:

min\_impurity\_decrease, min\_samples\_split

ADA Boost:

Apply **Adaptive Boosting** with the decision tree as base estimator



Random Forest Model:

max\_depth, min\_samples\_leaf, n\_estimators
Perform Feature Importance technique to decide which columns to include.

Logistic Regression:

stepwise\_selection by specifying different Penalty Types, Constrain, and Solvers

Neural Network:

hidden\_layer\_sizes

K-Nearest Neighbors:

Optimizing K-value by looking at its effect on F1-Score

#### Evaluate Model Performance

Compare Performance of models using following metrics:

- > Accuracy
- > Precision
- > Recall
- > F1-Measure

Gather Findings and Give Recommendations

Ensemble Learning

Summarize insights gathered from the data.

Construct an Ensemble

**Voting Classifier** 

using the top 2

models that had

the highest F1-Score

(KNN and Random Forest).

Give recommendations of how the business can benefit from the findings.



Gains Chart was graphed to aid with finding the optimum model that will make most out of the advertising campaign fundings.

Measuring Benefit

of All Models