Problem Statement

A key challenge for the insurance industry is to charge each customer an appropriate price for the risk they represent as we know that risk varies widely from customer to customer. A deep understanding of different risk factors helps predict the likelihood and cost of insurance claims. National Insurance wants to take your help in identifying whether a customer claims for the medical insurance or not and if he/she claims for insurance, predict the claim amount.

In this context, you'll work with the data provided by the company to solve the problem. As the information is financial related, to maintain the confidentiality, the feature names and the values provided are masked. It is up to you how you would want to treat these features.

Data Set

You are provided with two csv files- “train.csv” and “test.csv”. The “train.csv” has the target variable (whether the customer is satisfied or not) and the test.csv has the target for you to test.

Evaluations

As specified in the problem statement, this would be a two-stage problem. **You would be classifying whether a customer claims for insurance or not.** For

We aim at better F1 statistic

Step1: **Visualizations**
Since this forms an important aspect in data science problems, we would want you to use visualizations

to obtain any insights from the data that could be a value add to the company Step2: **Benchmark F1 statistic**

You are required to submit your predictions on data test through the shiny app provided and check the F1 statistic obtained. The benchmark for F1 is 42%