Project Report

By: Priyanka Joshi

Analyze the report of Swedish Motor Insurance

Business Scenario

The data gives the details of third party motor insurance claims in Sweden for the year 1977. In Sweden, all motor insurance companies apply identical risk arguments to classify customers, and thus their portfolios and their claims statistics can be combined. The data were compiled by a Swedish Committee on the Analysis of Risk Premium in Motor Insurance. The Committee was asked to look into the problem of analyzing the real influence on the claims of the risk arguments and to compare this structure with the actual tariff.

Expectations or goals:

After understanding the data, you need to help the committee with the following by the use of the R tool:

1. The committee is interested to know each field of the data collected through descriptive analysis to gain basic insights into the data set and to prepare for further analysis.
2. The total value of payment by an insurance company is an important factor to be monitored. So the committee has decided to find whether this payment is related to number of claims and the number of insured policy years. They also want to visualize the results for better understanding.
3. The committee wants to figure out the reasons for insurance payment increase and decrease. So they have decided to find whether distance, location, bonus, make, and insured amount or claims are affecting the payment or all or some of these are affecting it.
4. The insurance company is planning to establish a new branch office, so they are interested to find at what location, kilometer, and bonus level their insured amount, claims, and payment get increased. (Hint: Aggregate Dataset)
5. The committee wants to understand what affects their claim rates so as to decide the right premiums for a certain set of situations. Hence, they need to find whether the insured amount, zone, kilometer, bonus, or make affects the claim rates and to what extent.

Analysis:

1. we can see that the minimum claims and payments are 0 so there are some datapoints with zero claims
2. From the results we can infer
   -99% correlation means claims are highly correlated with payment
   -93% correlation indicated that insured policy years is highly correlated to payment

   -From the plots we can see that with increase in claims and insured years payment also increases.
3. From the output we can infer that except bonus and make all other variables are significant variables and has impact on the variable payment as they have less p-value.
4. From the results we can infer that

   -Zone 4 has the highest claims and payments and zone 7 has the lowest claims and payments also

   -Most of the payments and claims come from zone 1 to 4

   -Kilometer group 5 has the lowest payment and group 2 has maximum payment

   -bonus group 7 has drastically high insured, claims and payment compared to other groups.
5. From the results and looking at the p-values we can infer that all the independent variables have significant impact on the claims