



Third party Auto Insurance Claims Analysis

Data Science with R Programming - Course end project1

Student Name – Manohari Wijesooriya

11/10/2023



Contents

1. Problem Statement	3
2. Data	3
3. Descriptive Analysis	4
4. Total Payments	5
5. Variables impacting insurance payment.....	7
6. Deciding location for new Branch.....	8
7. Insurance factors Identification	10

1. Problem Statement

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.

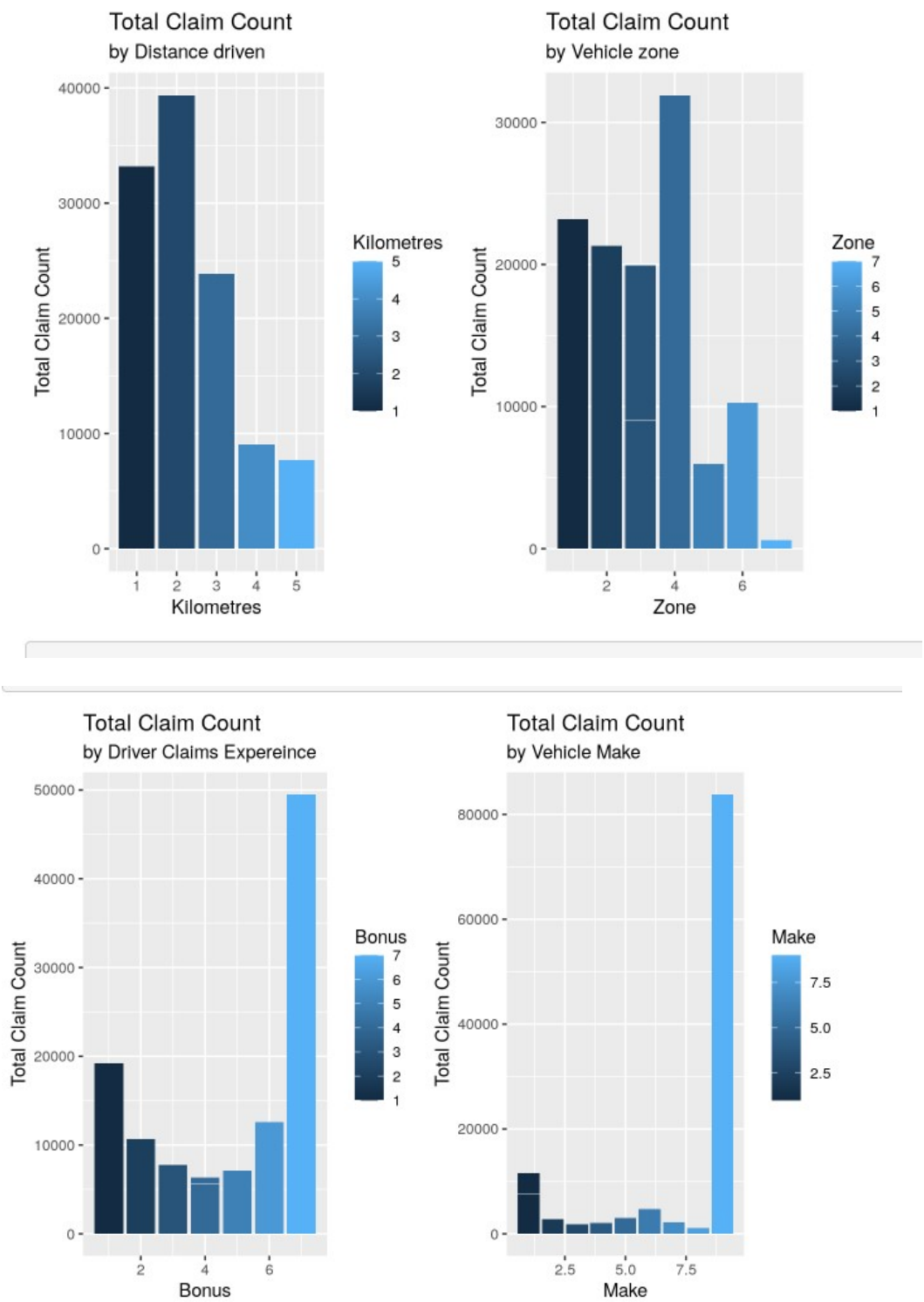
2. Data

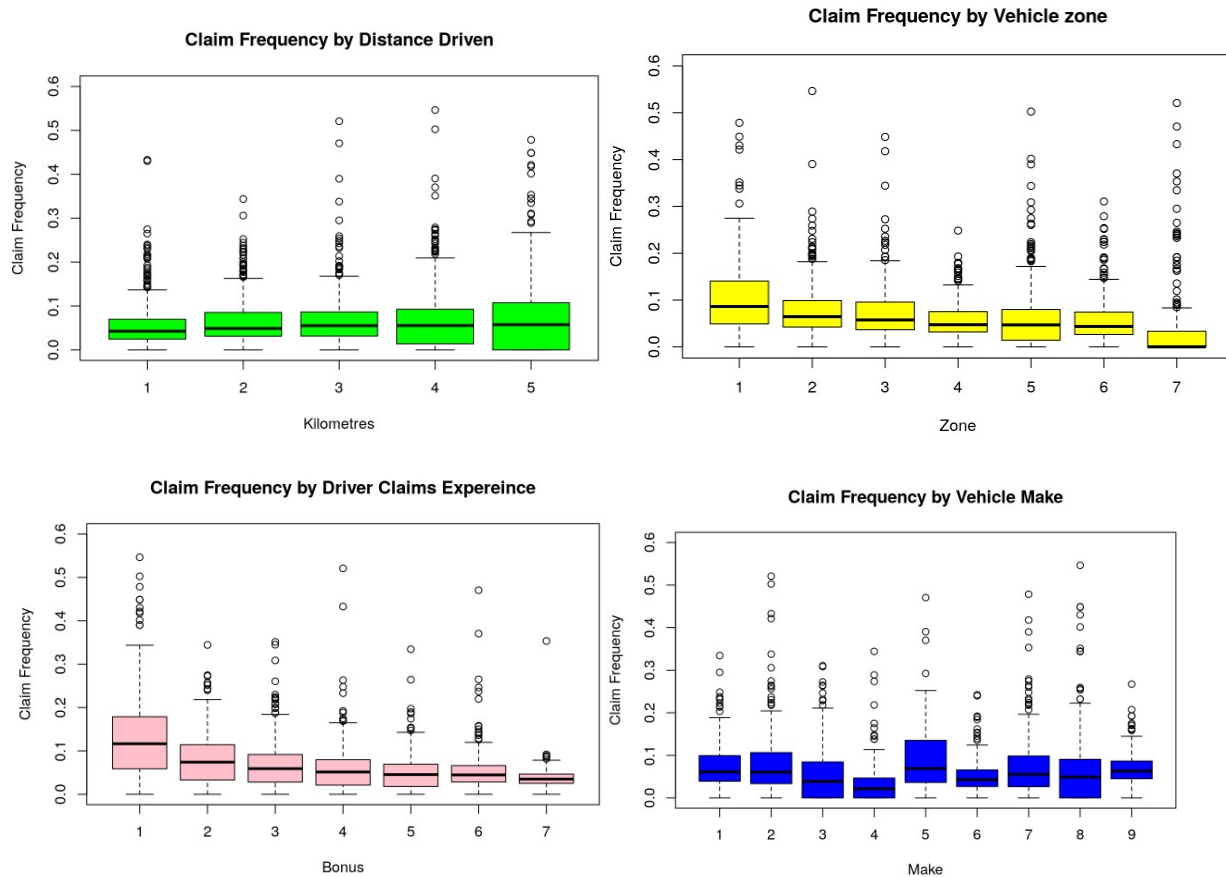
The insurance dataset holds 7 variables and the description of these variables are given below:

Variable	Description
Kilometres	Distance driven by a vehicle, grouped into five categories. Kilometers travelled per year 1: < 1000 2: 1000-15000 3: 15000-20000 4: 20000-25000 5: > 25000
Zone	Geographical zone of a vehicle, grouped into 7 categories. 1: Stockholm, Göteborg, and Malmö with surroundings 2: Other large cities with surroundings 3: Smaller cities with surroundings in southern Sweden 4: Rural areas in southern Sweden 5: Smaller cities with surroundings in northern Sweden 6: Rural areas in northern Sweden 7: Gotland
Bonus	Driver claim experience, grouped into 7 categories. No claims bonus; equal to the number of years, plus one, since the last claim
Make	1-8 represents eight different common car models. All other models are combined in class 9.
Insured	Number of insured in policy-years
Claims	Number of claims
Payment	Total value of payments in Skr (Swedish Krona)

3. Descriptive Analysis

This report is to gain basic insights into the data set and to prepare for the further analysis.





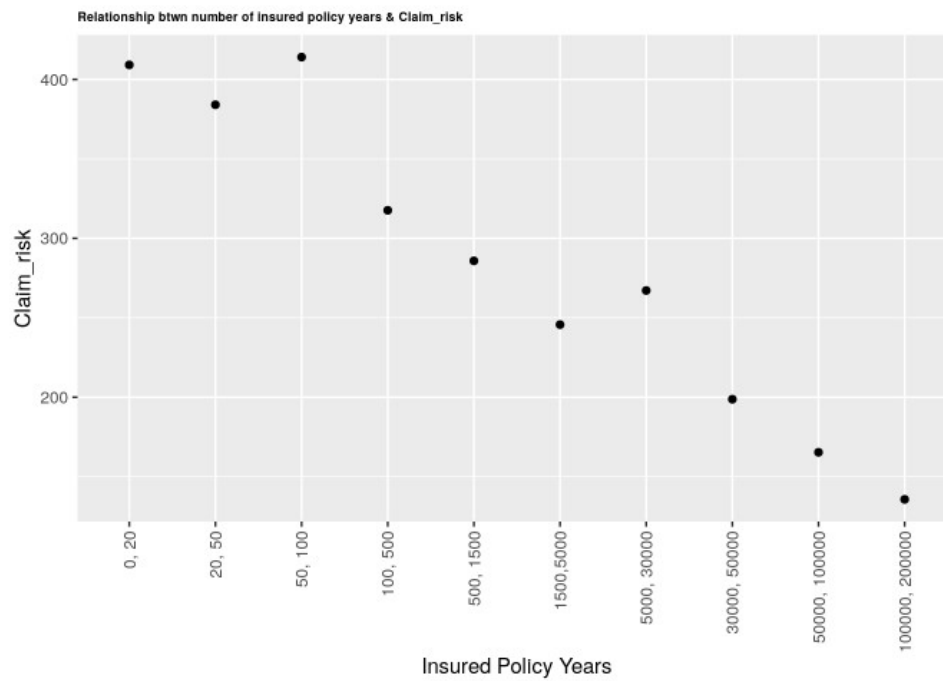
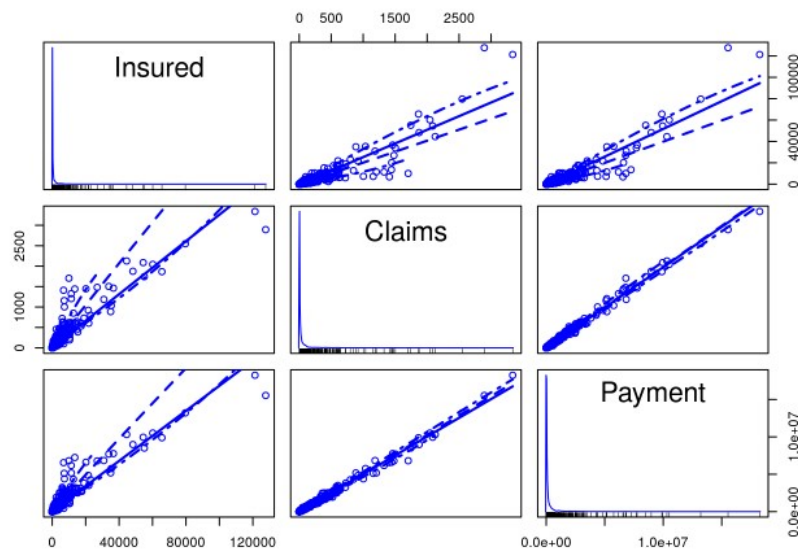
4. Total Payments

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.

Aggregating data by number of insured policy years :

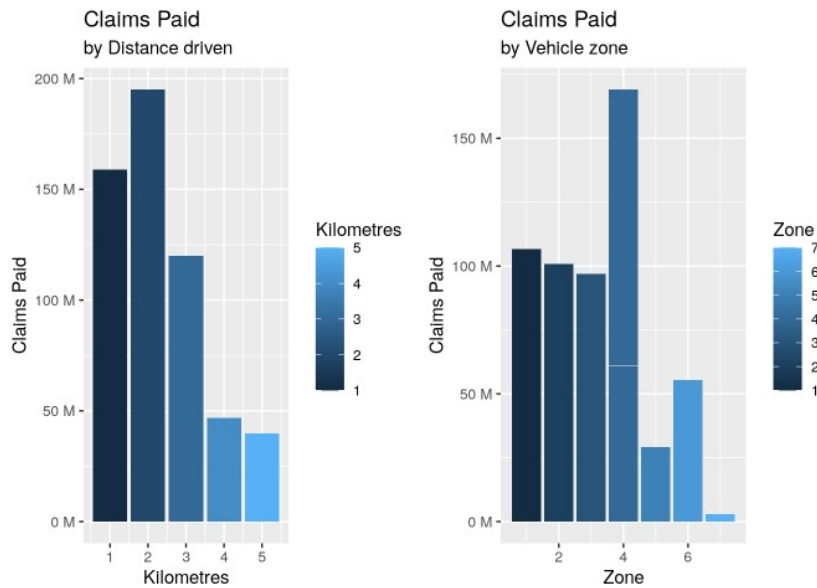
```
##   Insured_factor Insured Claims   Payment Severity Frequency Claim_risk
##   <fct>         <dbl> <int>    <int>    <dbl>    <dbl>    <dbl>
## 1 0, 20         4256.   326    1741500  5342.    0.0766   409.
## 2 20, 50        11792.  937    4528985  4833.    0.0795   384.
## 3 50, 100       20013. 1570    8286232  5278.    0.0784   414.
## 4 100, 500     126385. 8067   40142284 4976.    0.0638   318.
## 5 500, 1500    197563. 11551  56465555 4888.    0.0585   286.
## 6 1500,5000    371759. 18742  91331394 4873.    0.0504   246.
## 7 5000, 30000  823166. 45381 219915266 4846.    0.0551   267.
## 8 30000, 50000 264378. 10077  52537933 5214.    0.0381   199.
## 9 50000, 100000 314879. 10288  52056344 5060.    0.0327   165.
## 10 100000, 200000 248980.  6232  33785188 5421.    0.0250   136.
```

Swedish Motor Insurance



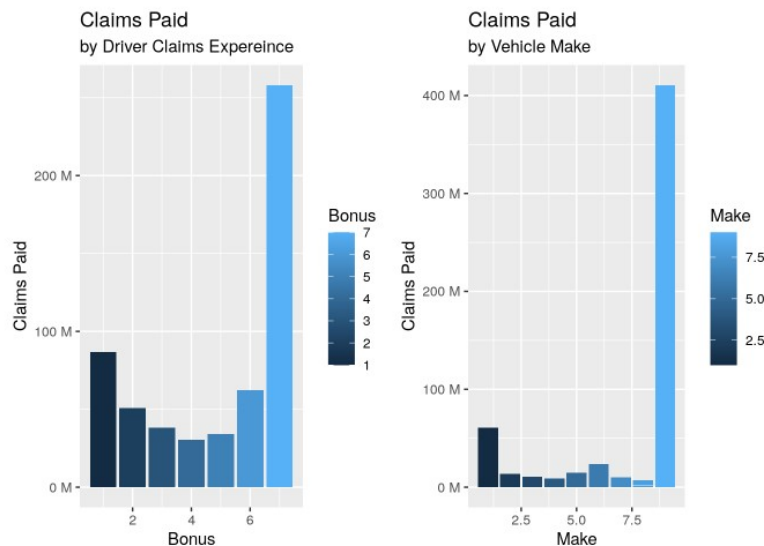
5. Variables impacting insurance payment

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.

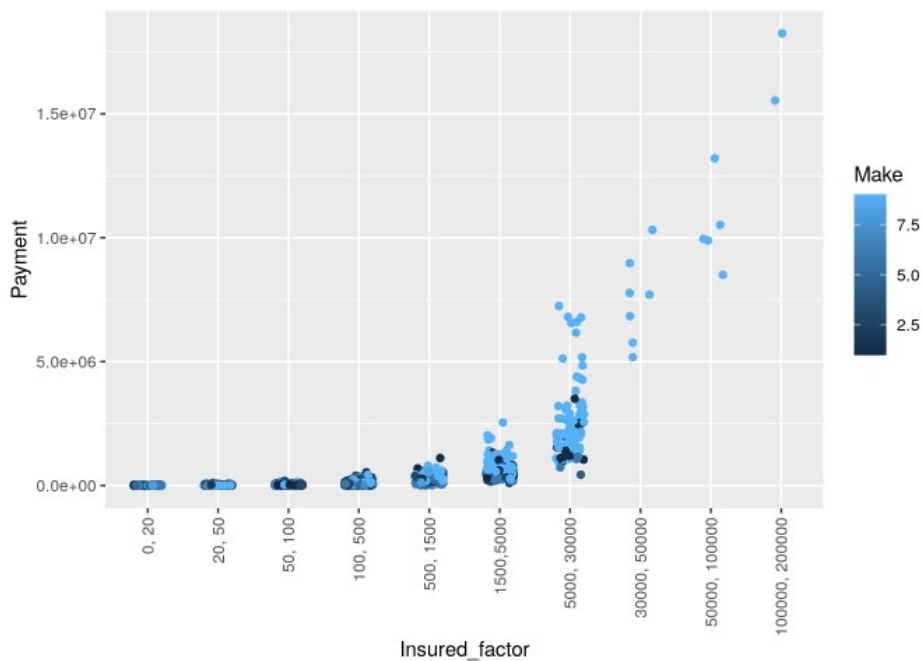


The most claims are from 1000-15000 km driven vehicles

The least amount of claims are paid in zone 7



Most claims are paid for policies with drivers with bonus = 7; ie 7 years from last claim



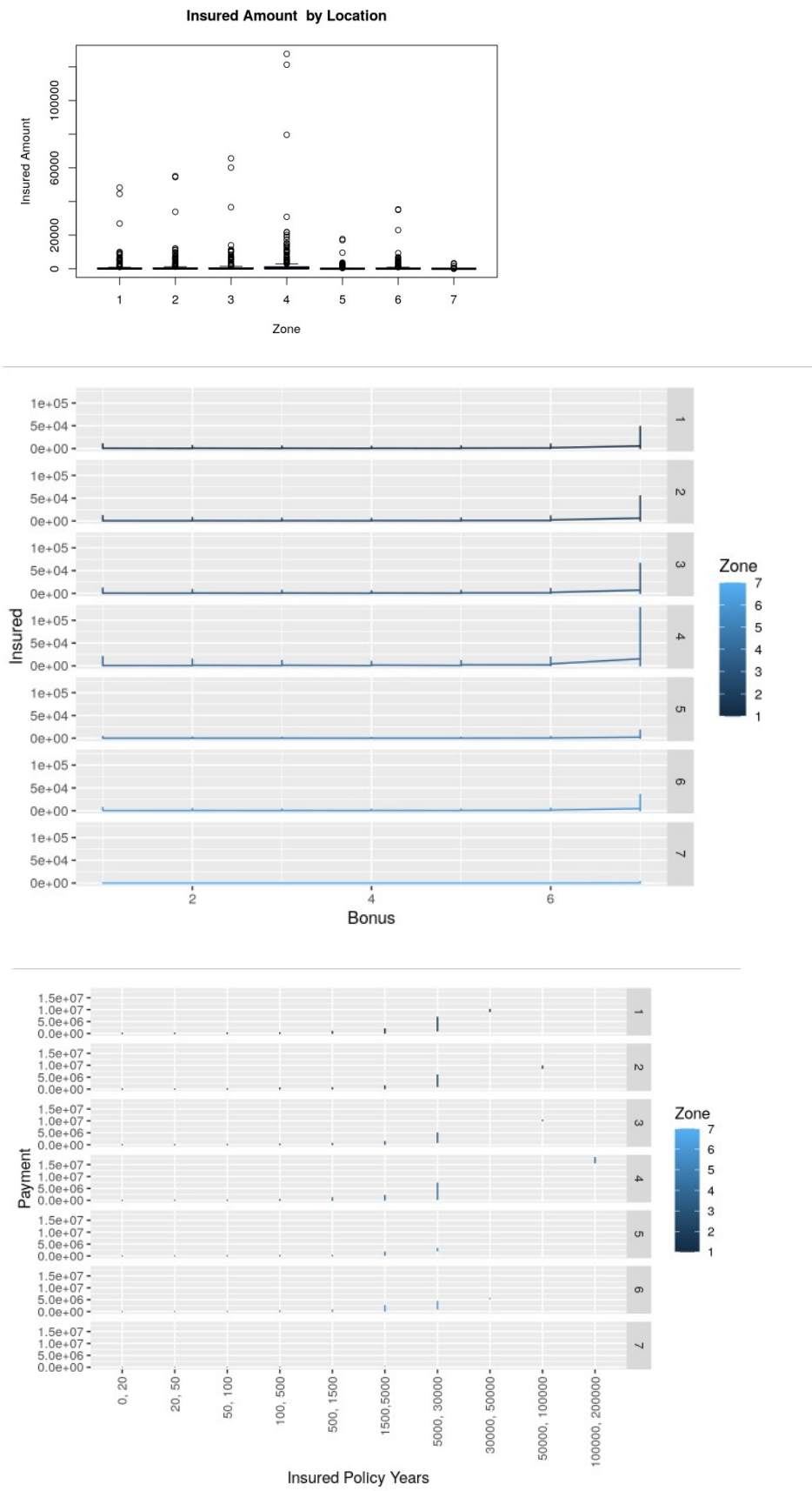
6. Deciding location for new Branch

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.

Summarising values by zone

##	Zone_f	Insured	Claims	Payment	Severity	Frequency	Claim_risk
## 1	1	326,394.10	23,174	106,633,468	1,463,024.3	32.61842	161,310.04
## 2	2	387,916.78	21,302	100,775,278	1,347,868.4	25.03624	111,987.36
## 3	3	429,331.99	19,938	96,878,519	1,514,008.4	22.74706	112,860.09
## 4	4	847,154.83	31,913	169,177,603	1,655,847.9	18.11293	95,291.90
## 5	5	120,442.99	5,962	29,109,577	1,327,481.7	19.59692	117,149.77
## 6	6	252,845.64	10,262	55,291,468	1,505,368.0	17.92239	107,957.10
## 7	7	19,083.75	620	2,924,768	541,679.8	14.83013	58,724.44

Swedish Motor Insurance



- High claim payments and high insured amounts in zones 4, 1 and 2.
- Good driving record in zones 4, 3, 2 and 1.
- High claim frequency in zone 7, 1 and 2.

7. Insurance factors Identification

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.

Correlation among variables

##	Kilometres	Zone	Bonus	Make	Insured	Claims	Payment
## Kilometres	1.00	-0.01	0.01	0.00	-0.33	-0.26	-0.24
## Zone	-0.01	1.00	0.01	-0.01	-0.32	-0.39	-0.36
## Bonus	0.01	0.01	1.00	0.00	0.35	0.20	0.20
## Make	0.00	-0.01	0.00	1.00	0.11	0.11	0.12
## Insured	-0.33	-0.32	0.35	0.11	1.00	0.93	0.90
## Claims	-0.26	-0.39	0.20	0.11	0.93	1.00	0.96
## Payment	-0.24	-0.36	0.20	0.12	0.90	0.96	1.00

A linear regression model is fitted to predict Payment.

The regression equation can be written as

$$\text{Payment} = -14704.39 + 16839.78 \cdot \text{Kilometres}_2 + 22285.52 \cdot \text{Kilometres}_3 + 15017.35 \cdot \text{Kilometres}_4 + 22107.93 \cdot \text{Kilometres}_5 + -12761.1 \cdot \text{Zone}_2 + -2312.81 \cdot \text{Zone}_3 + 39940.93 \cdot \text{Zone}_4 + 8851.57 \cdot \text{Zone}_5 + 14551.98 \cdot \text{Zone}_6 + 8405.92 \cdot \text{Zone}_7 + 3778.04 \cdot \text{Bonus} + -16167.04 \cdot \text{Make}_2 + -19636.57 \cdot \text{Make}_3 + -27087.48 \cdot \text{Make}_4 + -30199.86 \cdot \text{Make}_5 + -19971.42 \cdot \text{Make}_6 + -29646.97 \cdot \text{Make}_7 + -17026.41 \cdot \text{Make}_8 + -54497.24 \cdot \text{Make}_9 + 4939.75 \cdot \text{Claims}$$