

SwedishMotorInsurance

Suprava Sahoo

26/06/2020

Analysis need to be done:-

#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.

#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 the number of claims and the number of insured policy years. They also want to visualize the results for better understanding.

#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 insurance company is planning to establish a new branch office, so they are interested to find at what location, kilometre, and bonus level their insured amount, claims, and payment gets increased. (Hint: Aggregate Dataset)

#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, kilometre, bonus, or make affects the claim rates and to what extent.

```
setwd ("C:/Users/suppy/Downloads")
data_insurance = read.csv("d:/dataset/SwedishMotorInsurance.csv")
View(data_insurance)
str(data_insurance) ##-- returns the data types and the total observations for the dataset
```

```
## 'data.frame': 2182 obs. of 7 variables:
## $ Kilometres: int 1 1 1 1 1 1 1 1 1 1 ...
## $ Zone : int 1 1 1 1 1 1 1 1 1 1 ...
## $ Bonus : int 1 1 1 1 1 1 1 1 1 2 ...
## $ Make : int 1 2 3 4 5 6 7 8 9 1 ...
## $ Insured : num 455.1 69.2 72.9 1292.4 191 ...
## $ Claims : int 108 19 13 124 40 57 23 14 1704 45 ...
## $ Payment : int 392491 46221 15694 422201 119373 170913 56940 77487 6805992 214011 ...
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```

require("psych")

## Loading required package: psych

##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##      %+%, alpha

describe(data_insurance)

##           vars      n      mean      sd    median trimmed      mad    min
## Kilometres    1 2182      2.99      1.41     3.00      2.98      1.48 1.00
## Zone          2 2182      3.97      1.99     4.00      3.96      2.97 1.00
## Bonus         3 2182      4.02      2.00     4.00      4.02      2.97 1.00
## Make          4 2182      4.99      2.59     5.00      4.99      2.97 1.00
## Insured       5 2182    1092.20    5661.16    81.53    226.88    110.62 0.01
## Claims        6 2182      51.87     201.71     5.00     12.92      7.41 0.00
## Payment       7 2182 257007.64 1017282.59 27403.50 63667.75 40628.43 0.00
##           max      range    skew kurtosis      se
## Kilometres    5.0        4.0  0.01    -1.29     0.03
## Zone          7.0        6.0  0.01    -1.24     0.04
## Bonus         7.0        6.0 -0.01    -1.25     0.04
## Make          9.0        8.0  0.00    -1.23     0.06
## Insured      127687.3 127687.3 13.94    249.69   121.19
## Claims       3338.0   3338.0  8.56     93.06     4.32
## Payment     18245026.0 18245026.0 9.10    108.56 21777.81

library(corrplot)

## Warning: package 'corrplot' was built under R version 4.0.2

## corrplot 0.84 loaded

require("corrplot")

cor(data_insurance[,c("Payment", "Claims", "Insured")])

##           Payment    Claims    Insured
## Payment 1.0000000 0.9954003 0.9332170
## Claims  0.9954003 1.0000000 0.9103478
## Insured 0.9332170 0.9103478 1.0000000

windows()

pairs.panels(data_insurance[,c("Payment", "Claims", "Insured")])

```



```
lineModel = lm(Payment ~ ., data = data_insurance)
```

```
summary(lineModel)
```

```
##
## Call:
## lm(formula = Payment ~ ., data = data_insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -806775  -16943   -6321   11528   847015
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.173e+04  6.338e+03  -3.429 0.000617 ***
## Kilometres   4.769e+03  1.086e+03   4.392 1.18e-05 ***
## Zone         2.323e+03  7.735e+02   3.003 0.002703 **
## Bonus        1.183e+03  7.737e+02   1.529 0.126462
## Make        -7.543e+02  6.107e+02  -1.235 0.216917
## Insured      2.788e+01  6.652e-01  41.913 < 2e-16 ***
## Claims       4.316e+03  1.895e+01 227.793 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 70830 on 2175 degrees of freedom
## Multiple R-squared:  0.9952, Adjusted R-squared:  0.9952
## F-statistic: 7.462e+04 on 6 and 2175 DF, p-value: < 2.2e-16
```

```

claimmod = lm(Claims ~., data = data_insurance)

summary(claimmod)

##
## Call:
## lm(formula = Claims ~ ., data = data_insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -181.330   -3.196    0.887    3.755   231.782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.327e+00  1.436e+00   4.405 1.11e-05 ***
## Kilometres  -1.220e+00  2.462e-01  -4.956 7.75e-07 ***
## Zone        -7.697e-01  1.752e-01  -4.394 1.17e-05 ***
## Bonus       -4.339e-01  1.755e-01  -2.473  0.01349 *
## Make         4.402e-01  1.383e-01   3.182  0.00148 **
## Insured      -4.918e-03  1.735e-04 -28.349 < 2e-16 ***
## Payment      2.224e-04  9.762e-07  227.793 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.08 on 2175 degrees of freedom
## Multiple R-squared:  0.9937, Adjusted R-squared:  0.9936
## F-statistic: 5.685e+04 on 6 and 2175 DF,  p-value: < 2.2e-16

aggre=aggregate(data_insurance[,c(5,6,7)], by = list(data_insurance$Zone),FUN
=mean, na.rm=TRUE)
aggre

##   Group.1   Insured   Claims   Payment
## 1      1 1036.17175  73.568254 338518.95
## 2      2 1231.48184  67.625397 319921.52
## 3      3 1362.95870  63.295238 307550.85
## 4      4 2689.38041 101.311111 537071.76
## 5      5  384.80188  19.047923  93001.84
## 6      6  802.68457  32.577778 175528.47
## 7      7   64.91071   2.108844   9948.19

aggre=aggregate(data_insurance[,c(5,6,7)], by = list(data_insurance$Kilometres),FUN
=mean, na.rm=TRUE)
aggre

##   Group.1   Insured   Claims   Payment
## 1      1 1837.8163  75.59453 361899.35
## 2      2 1824.0288  89.27664 442523.78
## 3      3 1081.9714  54.16100 272012.58
## 4      4  398.9632  20.79493 108213.41
## 5      5  284.9475  18.04215  93306.12

```

```
bonus_aggre=aggregate(data_insurance[,c(5,6,7)], by = list(data_insurance$Bonus),FUN =mean, na.rm=TRUE)
bonus_aggre
```

```
##   Group.1   Insured   Claims   Payment
## 1      1  525.5502  62.50489 282921.99
## 2      2  451.0754  34.23397 163316.62
## 3      3  397.4737  24.97419 122656.17
## 4      4  360.3867  20.35161  98498.12
## 5      5  437.3936  22.82109 108790.50
## 6      6  805.8167  39.94286 197723.82
## 7      7 4620.3728 157.22222 819322.48
```