My Code:{Solution}

#load essential libraries

> library(readxl)

> library(plyr)

> #setworking dir

> setwd("/Users/rushikeshkhankar/Desktop/R")

> getwd()

[1] "/Users/rushikeshkhankar/Desktop/R"

> #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.

> SwedishMotorInsurance <- read.csv("~/Desktop/R/Project/Projects for Submission/Insurance/Insurance/SwedishMotorInsurance.csv")

> View(SwedishMotorInsurance)

> summary(SwedishMotorInsurance)

Kilometres Zone Bonus Make

Min. :1.000 Min. :1.00 Min. :1.000 Min. :1.000

1st Qu.:2.000 1st Qu.:2.00 1st Qu.:2.000 1st Qu.:3.000

Median :3.000 Median :4.00 Median :4.000 Median :5.000

Mean :2.986 Mean :3.97 Mean :4.015 Mean :4.992

3rd Qu.:4.000 3rd Qu.:6.00 3rd Qu.:6.000 3rd Qu.:7.000

Max. :5.000 Max. :7.00 Max. :7.000 Max. :9.000

Insured Claims Payment

Min. : 0.01 Min. : 0.00 Min. : 0

1st Qu.: 21.61 1st Qu.: 1.00 1st Qu.: 2989

Median : 81.53 Median : 5.00 Median : 27404

Mean : 1092.20 Mean : 51.87 Mean : 257008

3rd Qu.: 389.78 3rd Qu.: 21.00 3rd Qu.: 111954

Max. :127687.27 Max. :3338.00 Max. :18245026

> #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.

> lm1 <- lm(SwedishMotorInsurance$Payment~SwedishMotorInsurance$Insured)

> summary(lm1)

Call:

lm(formula = SwedishMotorInsurance$Payment ~ SwedishMotorInsurance$Insured)

Residuals:

Min 1Q Median 3Q Max

-5946157 -75828 -70260 -30246 5343552

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 73852.388 7971.250 9.265 <2e-16 \*\*\*

SwedishMotorInsurance$Insured 167.695 1.383 121.266 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 365600 on 2180 degrees of freedom

Multiple R-squared: 0.8709, Adjusted R-squared: 0.8708

F-statistic: 1.471e+04 on 1 and 2180 DF, p-value: < 2.2e-16

> lm2<- lm(SwedishMotorInsurance$Payment ~ .,data = SwedishMotorInsurance)

> summary(lm2)

Call:

lm(formula = SwedishMotorInsurance$Payment ~ ., data = SwedishMotorInsurance)

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

> group\_zone <- apply(SwedishMotorInsurance[,c(5,6,7)], 2, function(x)tapply(x, SwedishMotorInsurance$Zone, mean))

> group\_zone

Insured Claims Payment

1 1036.17175 73.568254 338518.95

2 1231.48184 67.625397 319921.52

3 1362.95870 63.295238 307550.85

4 2689.38041 101.311111 537071.76

5 384.80188 19.047923 93001.84

6 802.68457 32.577778 175528.47

7 64.91071 2.108844 9948.19

> group\_kilm <- apply(SwedishMotorInsurance[,c(5,6,7)],2,function(x)tapply(x, SwedishMotorInsurance$Kilometres, mean))

> group\_kilm

Insured Claims Payment

1 1837.8163 75.59453 361899.35

2 1824.0288 89.27664 442523.78

3 1081.9714 54.16100 272012.58

4 398.9632 20.79493 108213.41

5 284.9475 18.04215 93306.12

> group\_bonus <- apply(SwedishMotorInsurance[,c(5,6,7)],2,function(x)tapply(x, SwedishMotorInsurance$Bonus, mean))

> group\_bonus

Insured Claims Payment

1 525.5502 62.50489 282921.99

2 451.0754 34.23397 163316.62

3 397.4737 24.97419 122656.17

4 360.3867 20.35161 98498.12

5 437.3936 22.82109 108790.50

6 805.8167 39.94286 197723.82

7 4620.3728 157.22222 819322.48

> #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 .

> reg <- lm(Claims~Kilometres+Zone+Bonus+Make+Insured, data = SwedishMotorInsurance)

> summary(reg)

Call:

lm(formula = Claims ~ Kilometres + Zone + Bonus + Make + Insured,

data = SwedishMotorInsurance)

Residuals:

Min 1Q Median 3Q Max

-1214.57 -25.18 -9.41 10.04 1301.78

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 37.1230027 7.1270679 5.209 2.08e-07 \*\*\*

Kilometres -3.9648601 1.2255209 -3.235 0.00123 \*\*

Zone -6.2924300 0.8647405 -7.277 4.75e-13 \*\*\*

Bonus -4.2468101 0.8707236 -4.877 1.15e-06 \*\*\*

Make 6.7725342 0.6755390 10.025 < 2e-16 \*\*\*

Insured 0.0318697 0.0003158 100.933 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 80.14 on 2176 degrees of freedom

Multiple R-squared: 0.8425, Adjusted R-squared: 0.8421

F-statistic: 2328 on 5 and 2176 DF, p-value: < 2.2e-16