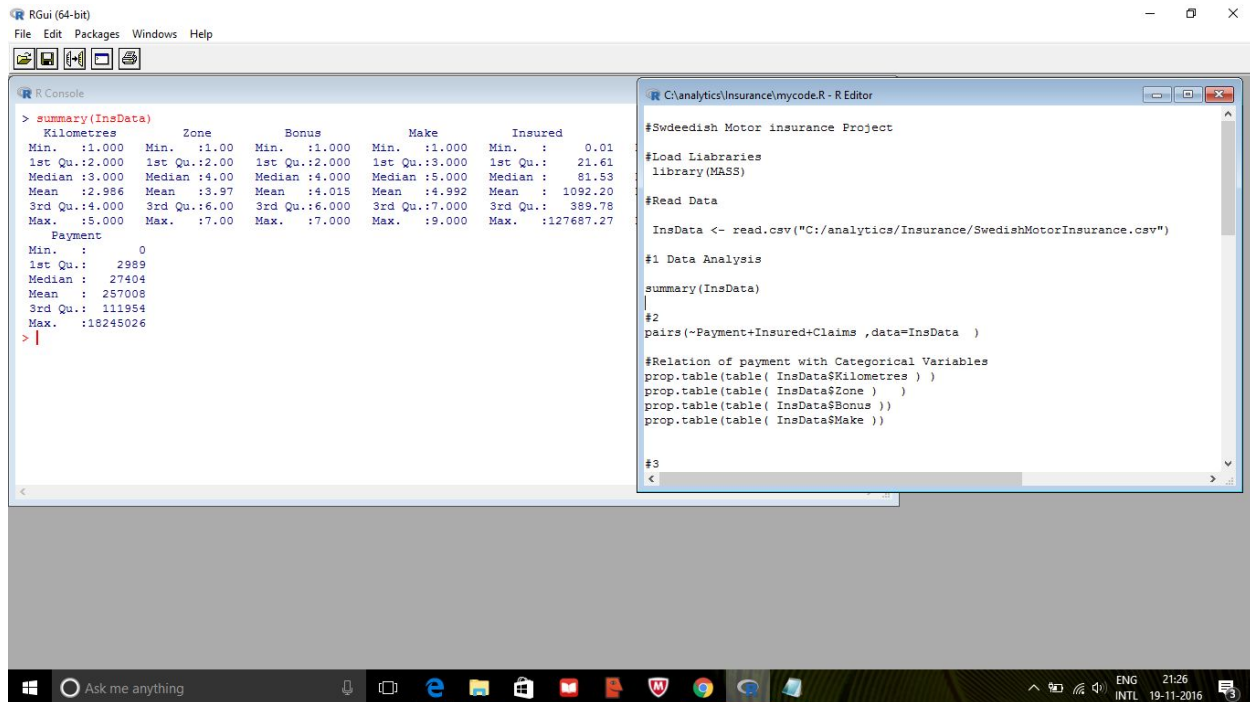


## Analysis report of Swedish Motor Insurance

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.

Code :

```
summary(InsData)
```



The screenshot shows the RGui (64-bit) interface. The R Console on the left displays the output of the `summary(InsData)` command, showing summary statistics for five variables: Kilometres, Zone, Bonus, Make, and Insured. The R Editor on the right shows the R script used to load the data and perform the analysis.

```
> summary(InsData)
      Kilometres      Zone      Bonus      Make      Insured
Min.   :1.000   Min.   :1.00   Min.   :1.000   Min.   :1.000   Min.   : 0.01
1st Qu.:2.000   1st Qu.:2.00   1st Qu.:2.000   1st Qu.:3.000   1st Qu.: 21.61
Median :3.000   Median :4.00   Median :4.000   Median :5.000   Median : 81.53
Mean   :2.986   Mean   :3.97   Mean   :4.015   Mean   :4.992   Mean   :1092.20
3rd Qu.:4.000   3rd Qu.:6.00   3rd Qu.:6.000   3rd Qu.:7.000   3rd Qu.:389.78
Max.   :15.000   Max.   :7.00   Max.   :7.000   Max.   :9.000   Max.  :127687.27

      Payment
Min.   : 0
1st Qu.: 2989
Median : 27404
Mean   : 257008
3rd Qu.: 111954
Max.   :18245026
> |
```

```
#Swedish Motor insurance Project
#Load Libraries
library(MASS)

#Read Data

InsData <- read.csv("C:/analytics/Insurance/SwedishMotorInsurance.csv")

#1 Data Analysis

summary(InsData)

#2
pairs(~Payment+Insured+Claims ,data=InsData )

#Relation of payment with Categorical Variables
prop.table(table( InsData$Kilometres ) )
prop.table(table( InsData$Zone ) )
prop.table(table( InsData$Bonus ))
prop.table(table( InsData$Make ))

#3
```

Analysis:

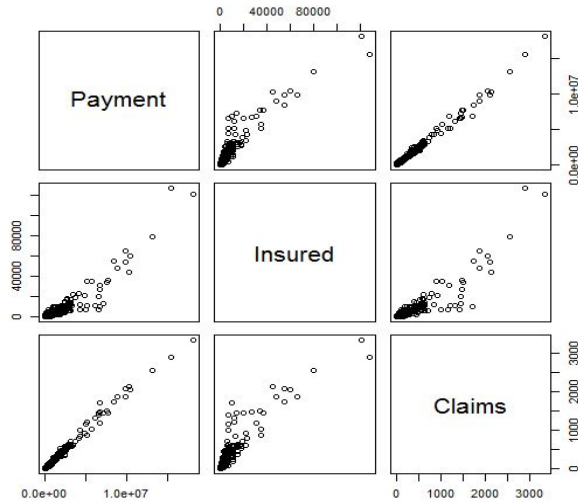
Avg Insure : 1092

Avg Payment : 257008 SKR

Avg Claims : 51

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

```
pairs(~Payment+Insured+Claims ,data=InsData )
```



#### Interpretation :

Payment and Insured is a linear relation , Most of the insured are less than 40000

Payment and claims are in Linear relation , Most of the claims are less than or equal to 2000

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 them are affecting it.

#### Code:

```
Rel <- lm(InsData$Payment~InsData$Kilometres+InsData$Zone+InsData$Bonus+InsData$Make+InsData$Insured )
summary(Rel)
```

```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console

---
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

> Rel <- lm(InsData$Payment~InsData$Kilometres+InsData$Zone+InsData$Bonus+InsData$Make+InsData$Insured+InsData$Claims )
> summary(Rel)

Call:
lm(formula = InsData$Payment ~ InsData$Kilometres + InsData$Zone +
    InsData$Bonus + InsData$Make + InsData$Insured + InsData$Claims)

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 ***
InsData$Kilometres  4.769e+03  1.086e+03  4.392 1.18e-05 ***
InsData$Zone       2.323e+03  7.735e+02  3.003 0.002703 **
InsData$Bonus      1.183e+03  7.737e+02  1.529 0.126462
InsData$Make      -7.543e+02  6.107e+02  -1.235 0.216917
InsData$Insured    2.788e+01  6.652e-01  41.913 < 2e-16 ***
InsData$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

> |
```

#### Analysis:

- A. P value for Bonus and make are significantly high. They will not affect payment
- B. Insured ,claims , distance has Significant impact on Payment
- C. Zone has moderate impact

**4. 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 get increased. ( Hint: Aggregate Dataset)**

I have created 2 new features “Claimratio” and “cost per insured”. It is proven assumption that

As claim ratio decreases profitability increases,

As Cost per insured decreases Risk decreases

So our objective is to find category with lower risk and higher profitability

#### 4.1 Analysis for Distance

##### Code:

```
InsSub <- subset(InsData , Select=C(
  FDist<- factor(InsData$Kilometres)
  InsDist <- subset(InsData, select= c (Insured,Claims,Payment,Kilometres) )
  AggrbyKM <- aggregate(.~FDist , InsDist ,sum)
  AggrbyKM$Claimratio <- (AggrbyKM$Claims/AggrbyKM$Insured)*100
  AggrbyKM$CostPerInsured <-AggrbyKM$Payment /AggrbyKM$Insured
  head(AggrbyKM)
  summary(AggrbyKM)
```

```

RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console
> head(AggrbyID)
  FDist Insured Claims Payment Kilometres Claimratio CostPerInsured
1 1 8048001.3 33186 158873815 439 4.113200 196.9183
2 2 804396.7 39371 195152987 882 4.894475 242.6079
3 3 477149.4 23885 119957549 1323 5.005770 251.4046
4 4 173150.0 8023 46964618 1736 5.212243 271.2365
5 5 121672.6 7704 39841712 2135 6.331747 327.4502

> summary(AggrbyID)
      FDist      Insured      Claims      Payment      Kilometres      Claimratio      CostPerInsured
1:1 Min. :121673 Min. : 7704 Min. : 39841712 Min. : 439 Min. : 4.113 Min. :196.9
2:1 1st Qu.:173150 1st Qu.: 8023 1st Qu.: 46964618 1st Qu.: 882 1st Qu.:4.894 1st Qu.:242.6
3:1 Median :477149 Median :23885 Median :119957549 Median :1323 Median :5.006 Median :251.4
4:1 Mean :476634 Mean :122634 Mean :112158136 Mean :1303 Mean :5.112 Mean :257.9
5:1 3rd Qu.:1804997 3rd Qu.:133186 3rd Qu.:158873815 3rd Qu.:1736 3rd Qu.:5.212 3rd Qu.:271.2
      Max. :1804901 Max. :139371 Max. :1295152987 Max. :2135 Max. :6.332 Max. :327.5

```

**Analysis : Observation indicates that**

**Distance 1 (<1000) km has lowest claim ratio and lowest cost Per insured**

**Claim ratio :4.11**

**Cost :196.91 SKR/Insured**

**And 2nd best is Distance 2 (1000-15000)**

**Claim ratio :4.89**

**Cost :242.91 SKR/Insured**

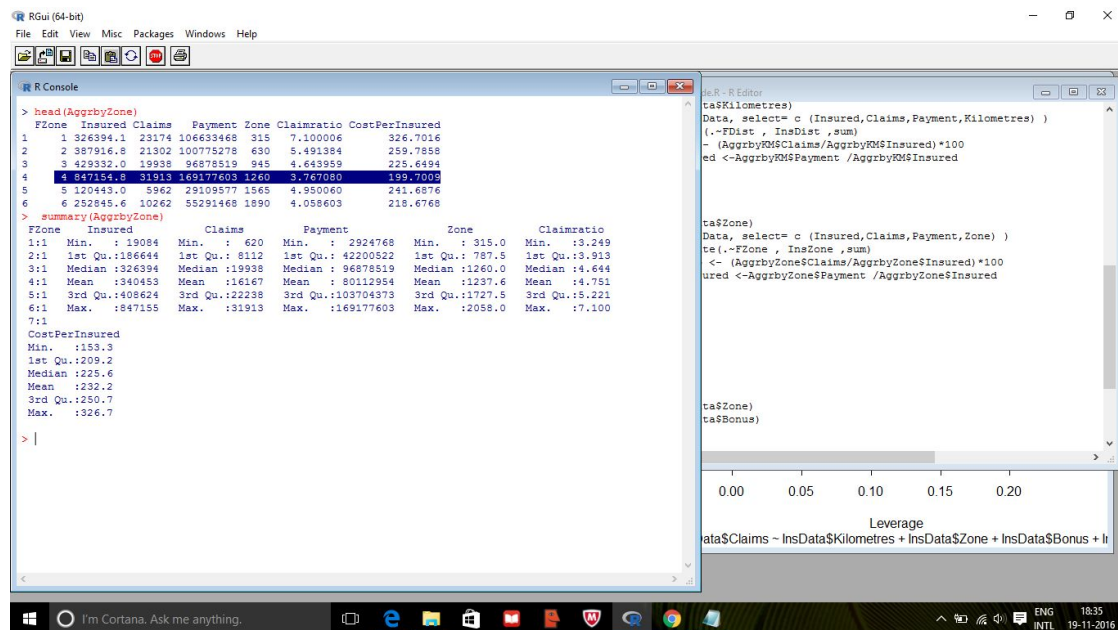
## 4.2 By Zone

**Code:**

```

FZone<- factor(InsData$Zone)
InsZone <- subset(InsData, select= c (Insured,Claims,Payment,Zone) )
AggrbyZone <- aggregate(~FZone , InsZone ,sum)
AggrbyZone$Claimratio <- (AggrbyZone$Claims/AggrbyZone$Insured)*100
AggrbyZone$CostPerInsured <-AggrbyZone$Payment /AggrbyZone$Insured
head(AggrbyZone)
summary(AggrbyZone)

```



**Analysis : Zone 4 has lowest claim ratio and lowest cost per insured.**

**Claim ratio :3.767**

**Cost :199.7 SKR/Insured**

### 4.3 By Bonus

**Code:**

```
FBonus<- factor(InsData$Bonus)
```

```
InsBonus <- subset(InsData, select= c (Insured,Claims,Payment,Bonus) )
```

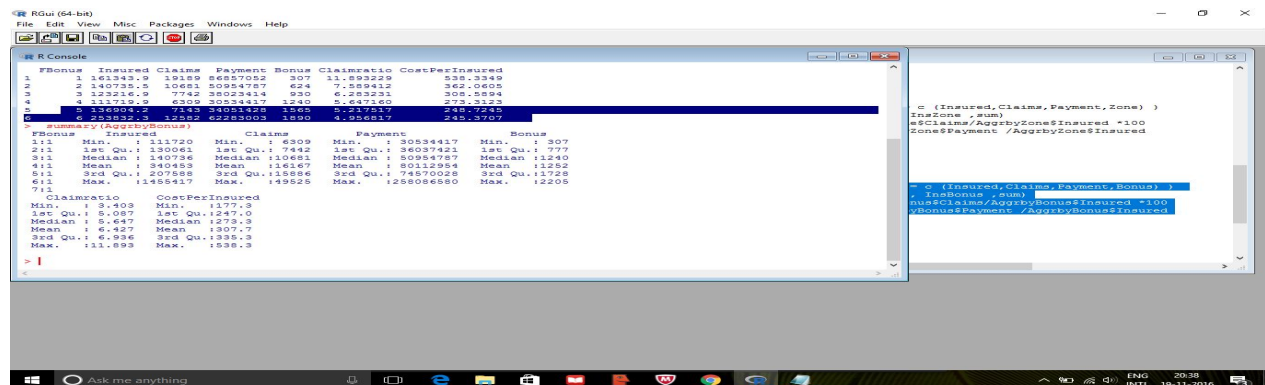
```
AggrbyBonus <- aggregate(.~FBonus , InsBonus ,sum)
```

```
AggrbyBonus$Claimratio <- AggrbyBonus$Claims/AggrbyBonus$Insured *100
```

```
AggrbyBonus$CostPerInsured <-AggrbyBonus$Payment /AggrbyBonus$Insured
```

```
head(AggrbyBonus)
```

```
summary(AggrbyBonus)
```



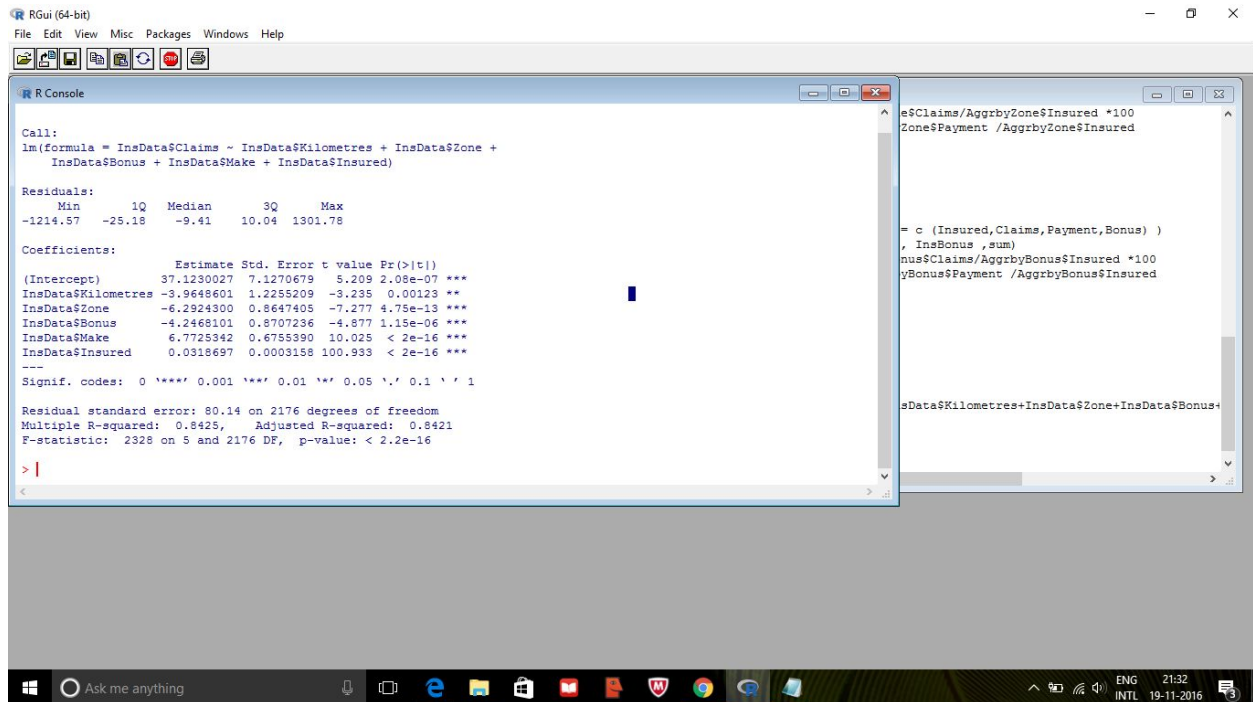
**Analysis : Bonus 6 has lowest claim ratio and lowest cost per insured.**

**Claim ratio :5.21 , Cost :245.37 SKR/Insured**

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.

Code:

```
ModelClaims <- lm(InsData$Claims~InsData$Kilometres+InsData$Zone+InsData$Bonus+InsData$Make+InsData$Insured)
summary(ModelClaims)
```



```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console
Call:
lm(formula = InsData$Claims ~ InsData$Kilometres + InsData$Zone +
    InsData$Bonus + InsData$Make + InsData$Insured)

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 ***
InsData$Kilometres -3.9648601   1.2255209  -3.235  0.00123 **
InsData$Zone      -6.2924300   0.8647405  -7.277 4.75e-13 ***
InsData$Bonus     -4.2468101   0.8707236  -4.877 1.15e-06 ***
InsData$Make       6.7725342   0.6755990  10.025 < 2e-16 ***
InsData$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

> |

Environment
e$Claims/AggrbyZone$Insured *100
Zone$Payment /AggrbyZone$Insured

= c (Insured,Claims,Payment,Bonus) )
, InsBonus ,sum)
nus$Claims/AggrbyBonus$Insured *100
yBonus$Payment /AggrbyBonus$Insured

sData$Kilometres+InsData$Zone+InsData$Bonus+
```

Analysis: As P value is Far less than 0.05 for all variables we are able to conclude that Claim rates are significantly dependent on insured amount, zone, kilometer, bonus, and make