

```
#Insurance factors identification
```

```
getwd()
```

```
setwd("~/Insurance factors identification")
```

```
library(readxl)
```

```
library(gdata)
```

```
Insurance <- read.csv  
( 'Insurance_factor_identification.csv' )
```

```
View(Insurance)
```

```
# To know each field of the data
```

```
summary(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.
```

```
lm1 <-
```

```
lm(Insurance$Payment~Insurance$Claims+Insurance$Insured)
```

```
lm1
```

```
summary(lm1)
```

```
#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.
```

```
cor(Insurance$Claims,Insurance$Payment)
```

```
cor(Insurance$Insured,Insurance$Payment)
```

```
# They also want to visualize the results for better understanding.
```

```
library(ggplot2)
```

```
plot(Insurance$Insured,Insurance$Payment)
```

```
plot(Insurance$Payment,Insurance$Insured)
```

```
# 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
```

```
lm2 <- lm(Insurance$Payment~., data = Insurance)
```

```
lm2
```

```
#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 gets increased.
```

```
new_branch <- apply(Insurance[,c(5,6,7)], 2,  
                    function(x)  
tapply(x,Insurance$Zone,mean))
```

```
new_branch
```

```
# find at what location, kilometer, and bonus level  
their  
# insured amount, claims, and payment gets increased.
```

```
high_claims <- apply(Insurance[,c(5,6,7)],2,  
function(x)tapply(x,Insurance$Kilometres,mean))
```

```
high_claims
```

```
max_pay <- apply(Insurance[,c(5,6,7)],2,  
function(x)tapply(x,Insurance$Bonus,mean))
```

```
max_pay
```

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

```
affect_claim <-  
lm(Claims~Kilometres+Zone+Bonus+Make+Insured,  
    data = Insurance)
```

```
summary(affect_claim)
```