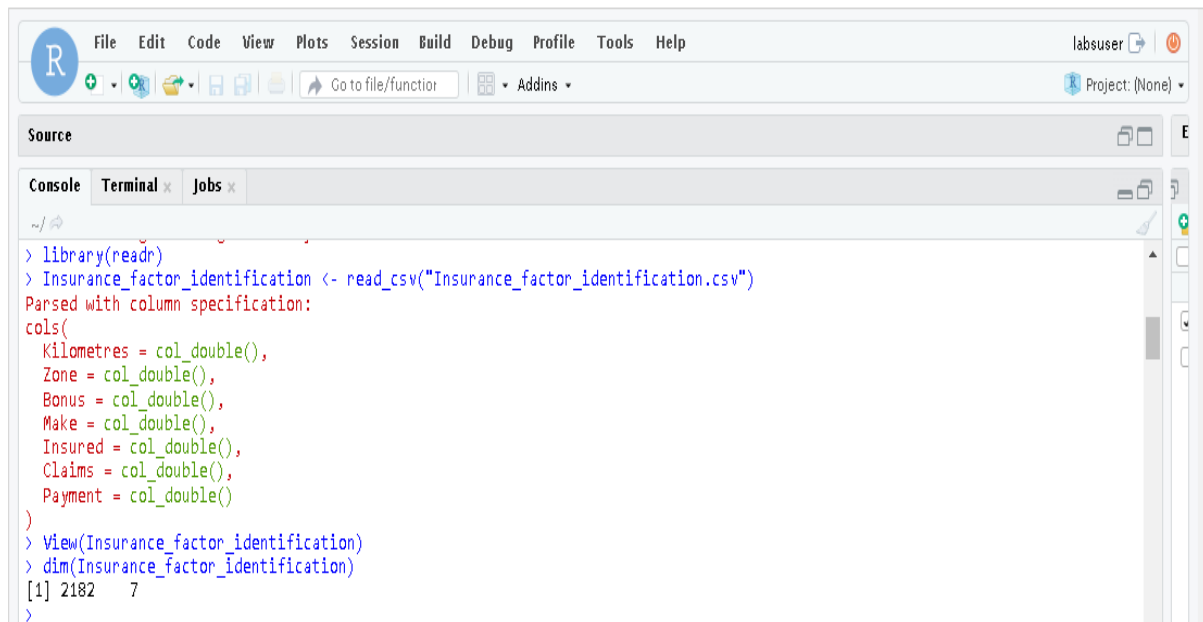


Data Science with R Insurance Project

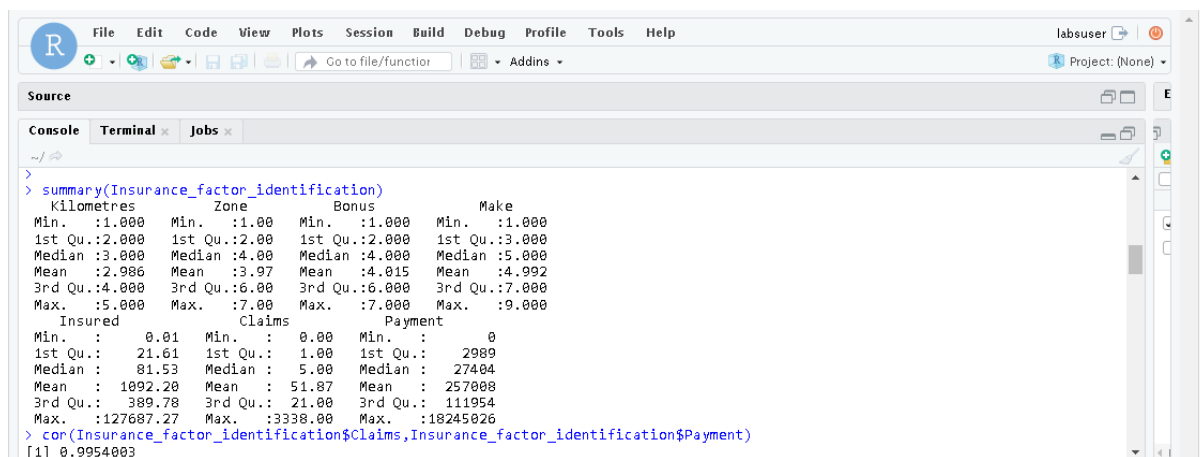
By Mpho Tsotetsi

Loading the data



The screenshot shows the RStudio interface with the following code in the console:

```
> library(readr)
> Insurance_factor_identification <- read_csv("Insurance_factor_identification.csv")
Parsed with column specification:
cols(
  Kilometres = col_double(),
  Zone = col_double(),
  Bonus = col_double(),
  Make = col_double(),
  Insured = col_double(),
  Claims = col_double(),
  Payment = col_double()
)
> View(Insurance_factor_identification)
> dim(Insurance_factor_identification)
[1] 2182 7
>
```



The screenshot shows the RStudio interface with the following code in the console:

```
> summary(Insurance_factor_identification)
  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
> cor(Insurance_factor_identification$Claims,Insurance_factor_identification$Payment)
[1] 0.9954003
```

1.Interpretation of the results:

The result gives 5 point summary of each column of the data. This gives the spread of data. There are some 0 values in Claims and Payment though the car has been Insured. This means no Claim or Payment has been made for that datapoint(s)

```
File Edit Code View Plots Session Build Debug Profile Tools Help labsuser
Go to file/function Addins
Project: (None)

Source

Console Terminal Jobs

~/
> cor(Insurance_factor_identification$Insured,Insurance_factor_identification$Payment)
[1] 0.933217
> plot(Insurance_factor_identification$Insured,Insurance_factor_identification$Payment)
> plot(Insurance_factor_identification$Claims,Insurance_factor_identification$Payment)
>
> lineModel <- lm(Payment ~ ., data = Insurance_factor_identification)
> summary(lineModel)

Call:
lm(formula = Payment ~ ., data = Insurance_factor_identification)

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 ***

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
File Edit Code View Plots Session Build Debug Profile Tools Help labsuser
Go to file/function Addins
Project: (None)

Source

Console Terminal Jobs

~/
            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

> ?apply
> ZoneResult <- apply(Insurance_factor_identification[,c(5,6,7)],2, function(x) tapply(x, Insurance_factor_identification$Zone, m
ean))
```

```
File Edit Code View Plots Session Build Debug Profile Tools Help labsuser
Go to file/function Addins
Project: (None)

Source

Console Terminal Jobs

~/
> ZoneResult
      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
> KmResult <- apply(Insurance_factor_identification[,c(5,6,7)],2, function(x) tapply(x, Insurance_factor_identification$Kilometre
s, mean))
> KmResult
      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
```

```
File Edit Code View Plots Session Build Debug Profile Tools Help labsuser
Source
Console Terminal Jobs
~/
> BonusResult <- apply(Insurance_factor_identification[,c(5,6,7)],2, function(x)tapply(x, Insurance_factor_identification$Bonus,
mean))
> BonusResult
      Insured   Claims  Payment
1  525.5502  62.50489 282921.99
2  451.0754  34.23397 163316.62
3  397.4737  24.97419 122656.17
4  360.3867  20.35161  98498.12
5  437.3936  22.82109 108790.50
6  805.8167  39.94286 197723.82
7 4620.3728 157.22222 819322.48
> md <- lm(Insurance_factor_identification$Claims ~ Insurance_factor_identification$Kilometres + Insurance_factor_identification
$Zone + Insurance_factor_identification$Bonus + Insurance_factor_identification$Make + Insurance_factor_identification$Insured)
> summary(md)
```

```
File Edit Code View Plots Session Build Debug Profile Tools Help labsuser
Source
Console Terminal Jobs
~/
Call:
lm(formula = Insurance_factor_identification$Claims ~ Insurance_factor_identification$Kilometres +
  Insurance_factor_identification$Zone + Insurance_factor_identification$Bonus +
  Insurance_factor_identification$Make + Insurance_factor_identification$Insured)

Residuals:
    Min       1Q   Median       3Q      Max
-1214.57  -25.18   -9.41   10.04  1301.78

Coefficients:
              Estimate Std. Error
(Intercept)  37.1230027   7.1270679
Insurance_factor_identification$Kilometres -3.9648601   1.2255209
Insurance_factor_identification$Zone -6.2924300   0.8647405
Insurance_factor_identification$Bonus -4.2468101   0.8707236
Insurance_factor_identification$Make  6.7725342   0.6755390
Insurance_factor_identification$Insured  0.0318697   0.0003158
```

```
File Edit Code View Plots Session Build Debug Profile Tools Help labsuser
Source
Console Terminal Jobs
~/
Insurance_factor_identification$Bonus  4.2468101  0.8707236
Insurance_factor_identification$Make  6.7725342  0.6755390
Insurance_factor_identification$Insured 0.0318697  0.0003158
              t value Pr(>|t|)
(Intercept)  5.209 2.08e-07 ***
Insurance_factor_identification$Kilometres -3.235 0.00123 ***
Insurance_factor_identification$Zone -7.277 4.75e-13 ***
Insurance_factor_identification$Bonus -4.877 1.15e-06 ***
Insurance_factor_identification$Make 10.025 < 2e-16 ***
Insurance_factor_identification$Insured 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
```