`

Date: 28/Mar/2024 EXPERIMENT – 09	LOGISTIC REGRESSION

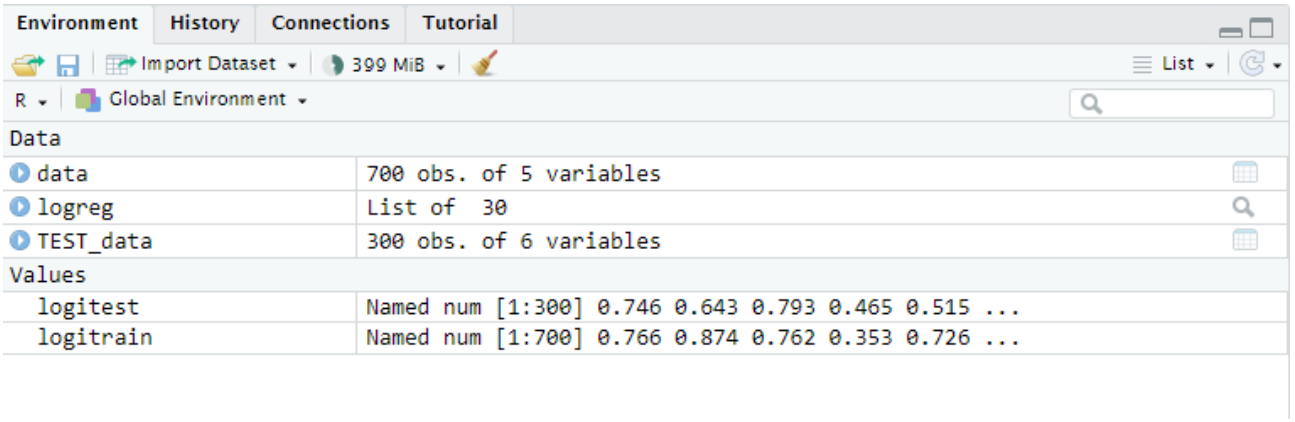
AIM: To perform Logistic Regression

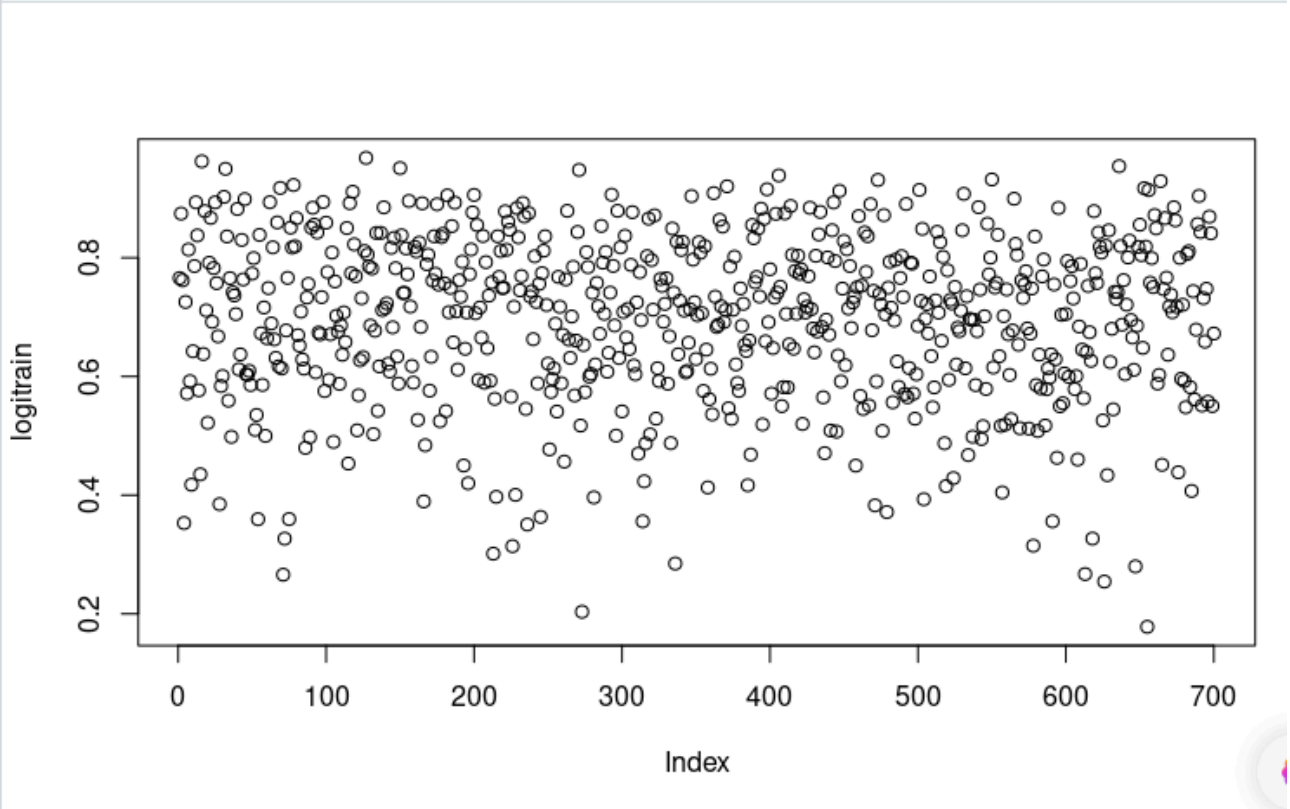
SOFTWARE REQUIRED: RStudio

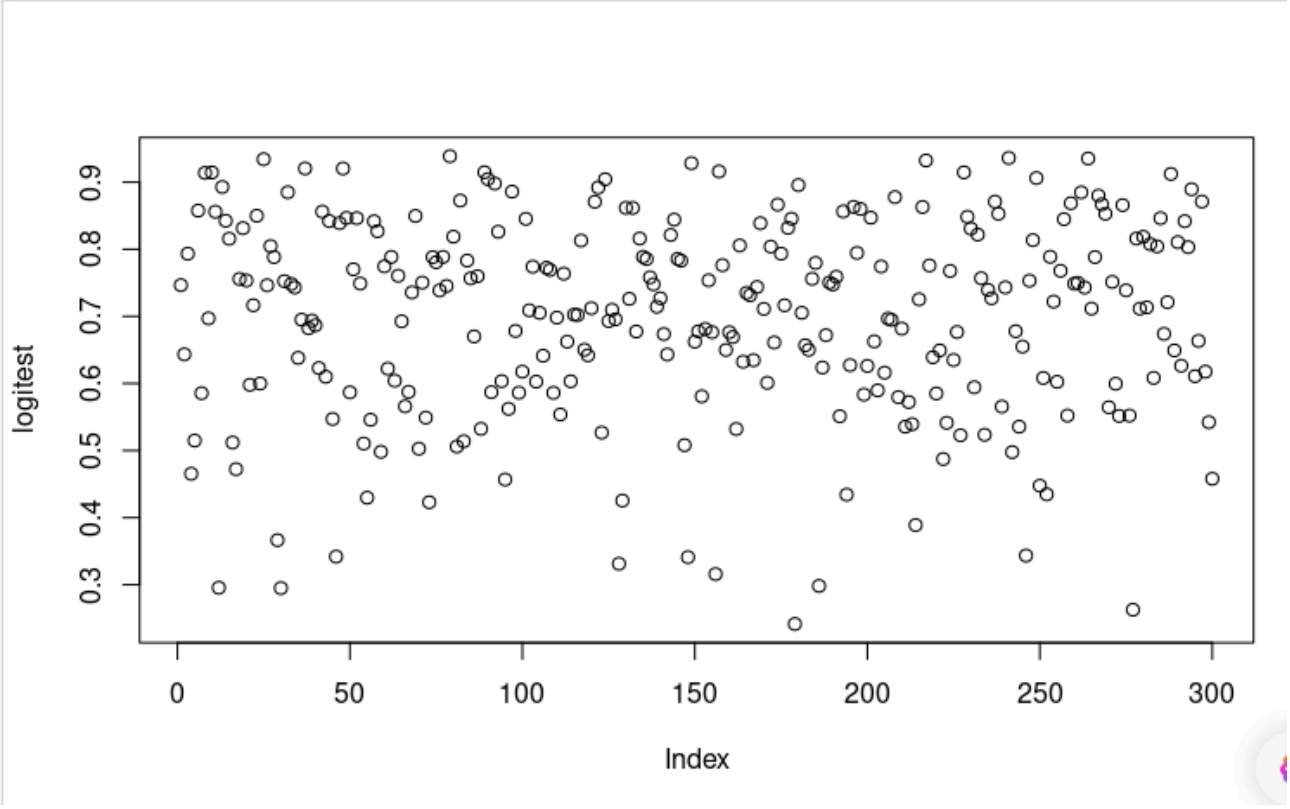
R CODE:

```
rm(list=ls())
data<-read.csv("CreditWorthiness_TRAIN.csv",stringsAsFactors = T)
logreg<-glm(formula = data$creditScore ~., family='binomial',data=data)
summary(logreg)
logitrain<-predict(logreg,type='response')
plot(logitrain)
tapply(logitrain,data$creditScore,mean)
TEST_data<-read.csv("CreditWorthiness_TEST.csv",stringsAsFactors = T)
logitest <- predict(logreg,newdata=TEST_data,type='response')
plot(logitest)
tapply(logitest,TEST_data$creditScore,mean)
TEST_data[logitest<=0.7,"LogiTest"]="bad"
TEST_data[logitest>0.7,"LogiTest"]="good"
install.packages("caret")
library(caret)
confusionMatrix(table(TEST_data[,5],TEST_data[,6]),positive='good')
```

OUTPUT:




logitrain
0.2
0.4
0.6
0.8
0
100
200
300
400
500
600
700
Index


logitest
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
50
100
150
200
250
300
Index

```
Console  Terminal  Background Jobs
R 4.3.3 . /cloud/project/
> rm(list=ls())
> data<-read.csv("CreditWorthiness_TRAIN.csv",stringsAsFactors  = T)
> logreg<-glm(formula = data$creditScore ~., family='binomial',data=data)
> summary(logreg)

Call:
glm(formula = data$creditScore ~ ., family = "binomial", data = data)

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)              8.320e-01  4.404e-01   1.889 0.058846 .
Cdur                    -3.240e-02  9.178e-03  -3.530 0.000416 ***
Cpurdomestic needs      -5.132e-01  7.925e-01  -0.648 0.517299
Cpureducation           -8.836e-01  4.318e-01  -2.046 0.040730 *
Cpurelectronics          3.648e-01  3.202e-01   1.139 0.254638
Cpurfurniture           -1.992e-01  3.305e-01  -0.603 0.546822
Cpurmiscellaneous       -2.014e-01  7.279e-01  -0.277 0.782048
Cpurnew vehicle          1.160e+00  4.524e-01   2.565 0.010324 *
Cpurrenovation          -7.147e-01  5.730e-01  -1.247 0.212308
Cpurretaining            5.438e-01  1.124e+00   0.484 0.628615
Cpursecond hand vehicle -5.999e-01  3.169e-01  -1.893 0.058367 .
Camt                    -3.647e-06  4.060e-06  -0.898 0.369109
age                      2.571e-02  8.527e-03   3.016 0.002565 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 860.23  on 699  degrees of freedom
Residual deviance: 793.13  on 687  degrees of freedom
AIC: 819.13

Number of Fisher Scoring iterations: 4

> logitrain<-predict(logreg,type='response')
> plot(logitrain)
> tapply(logitrain,data$creditScore,mean)
      bad      good
0.6302780 0.7243343
> TEST_data<-read.csv("CreditWorthiness_TEST.csv",stringsAsFactors = T)
> logitest <- predict(logreg,newdata=TEST_data,type='response')
> plot(logitest)
> tapply(logitest,TEST_data$creditScore,mean)
      bad      good
0.6293894 0.7302362
> TEST_data[logitest<=0.7,"LogiTest"]="bad"
```

```
Console   Terminal ×   Background Jobs ×
R 4.3.3 · /cloud/project/
> TEST_data[logitest>0.7,"LogiTest"]="good"
> install.packages("caret")
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
(as 'lib' is unspecified)
trying URL 'http://rspm/default/__linux__/focal/latest/src/contrib/caret_6.0-94.tar.gz'
Content type 'application/x-gzip' length 3573770 bytes (3.4 MB)
==================================================
downloaded 3.4 MB

* installing *binary* package 'caret' ...
* DONE (caret)

The downloaded source packages are in
        '/tmp/RtmpgadiwD/downloaded_packages'
> library(caret)
Loading required package: ggplot2
Loading required package: lattice
> confusionMatrix(table(TEST_data[,5],TEST_data[,6]),positive='good')
Confusion Matrix and Statistics


       bad good
  bad   53   34
  good  82  131

               Accuracy : 0.6133
                 95% CI : (0.5557, 0.6687)
    No Information Rate : 0.55
    P-Value [Acc > NIR] : 0.01553

                  Kappa : 0.1928

 Mcnemar's Test P-Value : 1.278e-05

            Sensitivity : 0.7939
            Specificity : 0.3926
         Pos Pred Value : 0.6150
         Neg Pred Value : 0.6092
             Prevalence : 0.5500
         Detection Rate : 0.4367
   Detection Prevalence : 0.7100
      Balanced Accuracy : 0.5933

       'Positive' Class : good

>
```