 **Context**:

You work for ChangeCo Finance, a national credit card processing company that handles credit card data for Visa and Mastercard accounts issued through McMallory Bank.

Yesterday, you received a call from the FBI that HouseMart, America’s largest hardware store empire, suffered a hacking attack to its point-of-sale systems, and that millions of credit card numbers were stolen.  The FBI is now asking you to review your credit card processing data of the past three weeks to determine if any fraud occurred on any of your managed accounts, and submit your results to the Department of the Treasury Financial Crimes Enforcement Network (FinCEN).

Your data file is attached:  [IT633\_Session\_12b\_CreditCard.csv](https://blackboard.valpo.edu/bbcswebdav/pid-440750-dt-content-rid-412258_1/xid-412258_1) .

The FinCEN reporting template has the following sections:

1. % of card accounts, most likely compromised
2. Store locations most likely attacked
3. Average fraudulent charges per compromised card
4. Predicted % of compromised cards in the next 3 weeks.
5. Data Mining methods and algorithms used to determine fraudulent activity.

NOTE that FinCEN awards $100,000 for use of Ensemble Learning methods in an effort to increase the reliability of analysis results.

Materials and Applications you may use:

•MS Excel, Weka, R, Python

•Hadoop/ Hortonworks, h2o

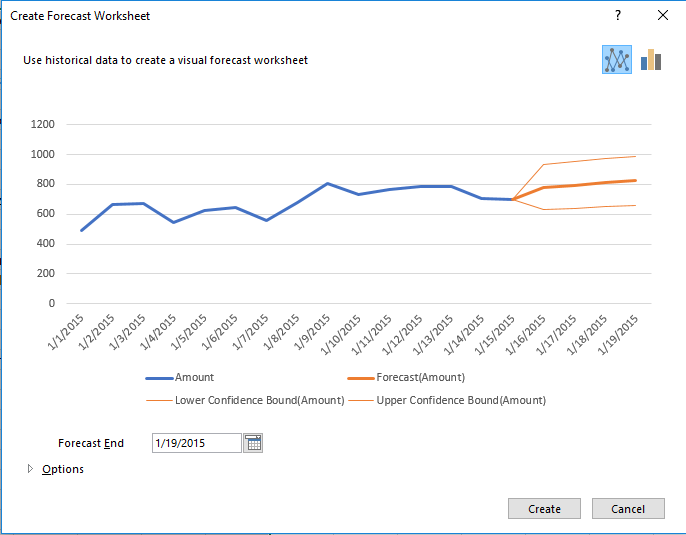
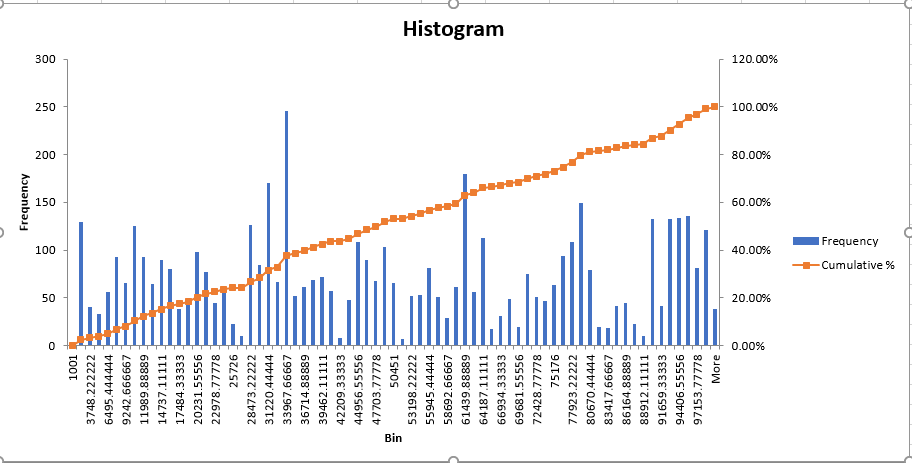
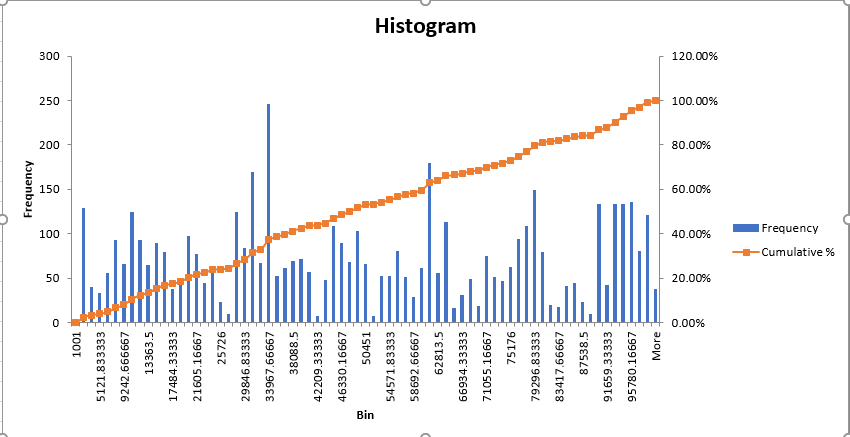
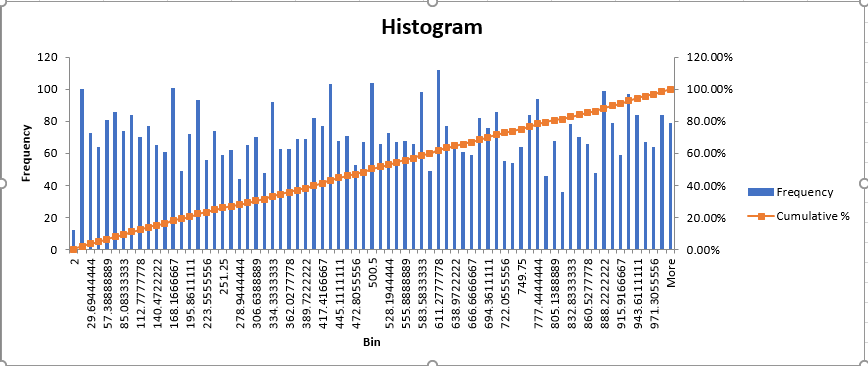
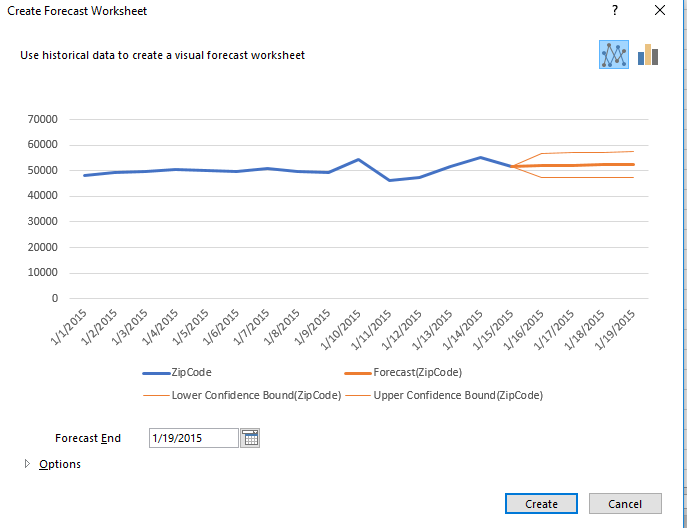
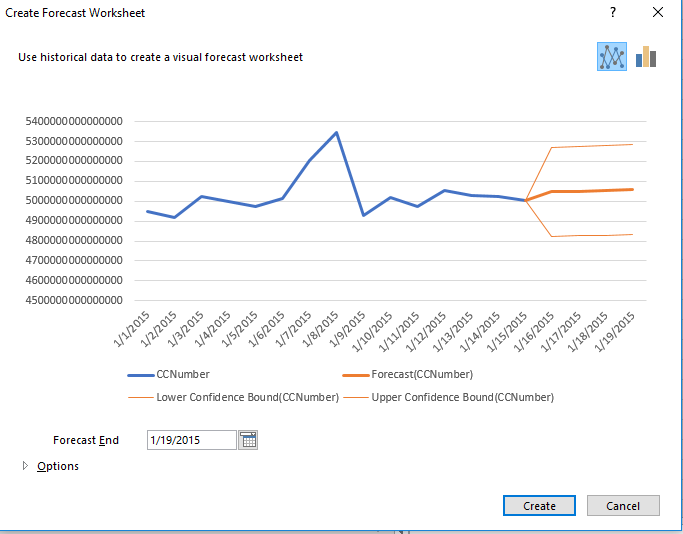
The Dataset provided given was regarding the credit card fraudulent and it was been given to identify frequent patterns of cards used and identified and make a re-utilizing of the card with the regular patterns, and for this identifying the regular patterns with using histogram and legitimate transaction using regular pattern.   
  
  
Here are the below screenshots of identifying the analysis of the provided CSV file of making an interpretation of what best values can be taken in consideration of regular pattern and then identifying the forecast. In order to identify the fradualent transaction the best method used is to make a **neural networks** identification to make a identification of best related attributes and make a pattern identification by creating a cluster.   
  
  


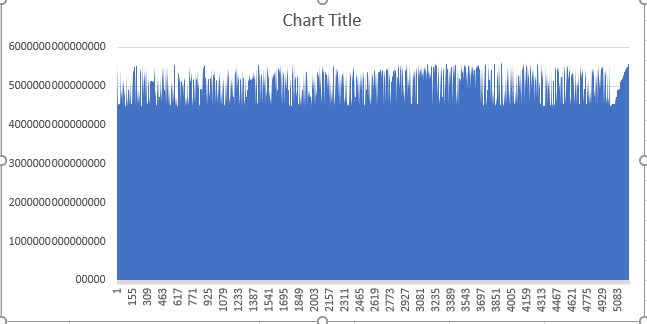
Fig1: Histogram and frequency

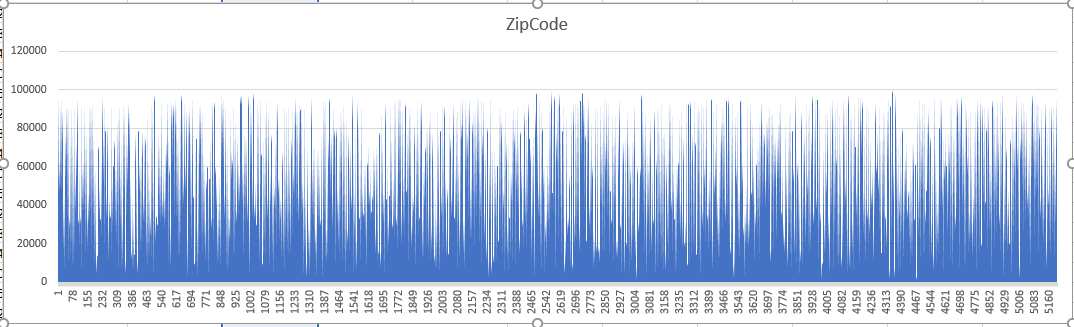


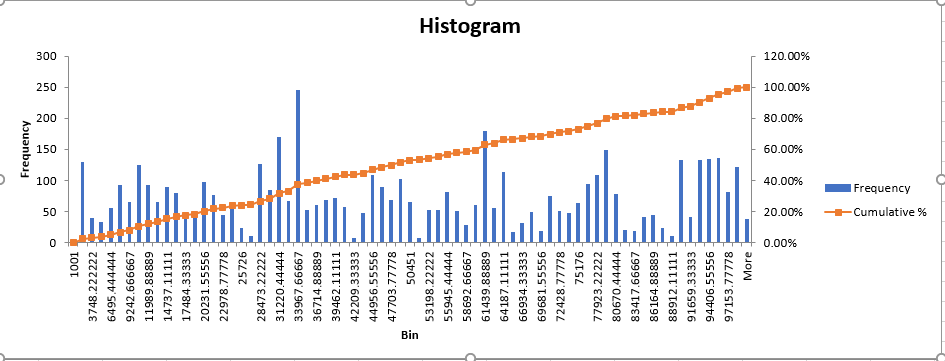




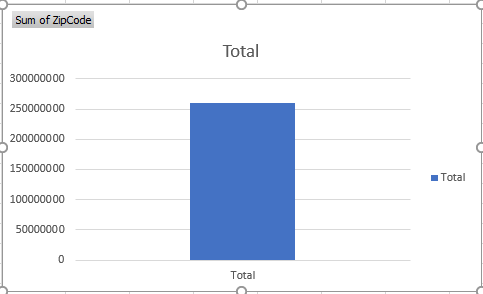


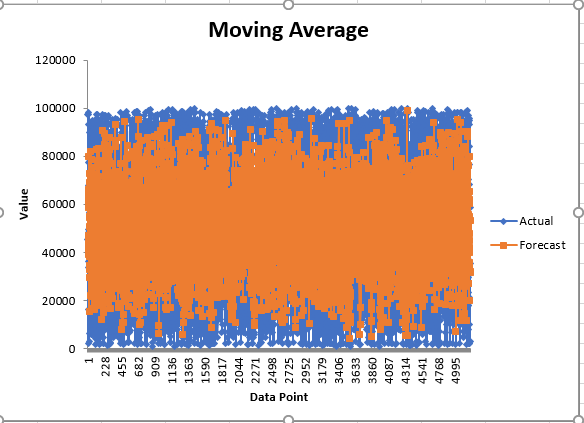






2.



  
  
  
  
  
K- Means is used for making the regular pattern of the respective credit card used and we can identify the regular pattern identified using the value of   
  
  
  
R Code:   
  
  
library(readr)

Goteti <- read\_csv("C:/DM SEM 2/week 14/IT633\_Session\_12b\_CreditCard.csv")

View(Goteti)

summary(Goteti)

install.packages("tree")

install.packages("party")

install.packages("rpart")

install.packages("car")

install.packages("mlbench")

install.packages("mboost")

install.packages("textir")

install.packages("class")

install.packages("e1071")

install.packages("randomForest")

library(textir)

library(MASS)

str(Goteti)

plot(Goteti)

na.rm=TRUE

na.na=TRUE

par(mfrow=c(3,3), mai=c(.3,.6,.1,.1))

plot(CCType ~ ., data=Goteti, col=c(grey(.2),2:6))

plot(CVV2 ~ ., data=Goteti, col=c(grey(.2),2:6))

n=length(Goteti$ZipCode)

nt=500

set.seed(1) ## to make the calculations reproducible in repeated runs

train <- sample(1:n,nt)

x=Goteti[,c(4,1)]

x[,1]=(x[,1]-mean(x[,1]))/sd(x[,1])

x[,2]=(x[,2]-mean(x[,2]))/sd(x[,2])

x[1:3,]

set.seed(1234)

model <- lm(formula=trainSet$ZipCode ~ . , data=Goteti)

training\_indices <- sample(seq\_len(nrow(Goteti$ZipCode)),

size=trainSize)

trainSet <- ZipCode[training\_indices, ]

testSet <- ZipCode[-training\_indices, ]

maxs <- apply(Goteti[,2:18], 2, max)

mins <- apply(Goteti[,2:18], 2, min)

library(class)

na.rm=TRUE

na.na=TRUE

nearest1 <- knn(train=x[train,],test=x[-train,],cl=Goteti$ZipCode[train],k=1)

nearest5 <- knn(train=x[train,],test=x[-train,],cl=Goteti$ZipCode[train],k=5)

data.frame(Goteti$ZipCode[-train],nearest1,nearest5)  
  
  
  
  
  
Output:  
  
> library(readr)

> library(readr)

> Goteti <- read\_csv("C:/DM SEM 2/week 14/IT633\_Session\_12b\_CreditCard.csv")

Parsed with column specification:

cols(

Date = col\_character(),

Time = col\_time(format = ""),

CCNumber = col\_double(),

Amount = col\_integer(),

CCType = col\_character(),

CCExpires = col\_character(),

CVV2 = col\_integer(),

GivenName = col\_character(),

MiddleInitial = col\_character(),

Surname = col\_character(),

ZipCode = col\_integer(),

TelephoneNumber = col\_character(),

MerchantID = col\_character(),

TransactionID = col\_double()

)

> View(IT633\_Session\_12b\_CreditCard)

Error in View : object 'IT633\_Session\_12b\_CreditCard' not found

> summary(Goteti)

Date Time CCNumber Amount CCType CCExpires CVV2 GivenName

Length:5203 Length:5203 Min. :4.485e+15 Min. : 1.0 Length:5203 Length:5203 Min. : 2.0 Length:5203

Class :character Class1:hms 1st Qu.:4.716e+15 1st Qu.: 318.5 Class :character Class :character 1st Qu.:240.0 Class :character

Mode :character Class2:difftime Median :5.157e+15 Median : 627.0 Mode :character Mode :character Median :499.0 Mode :character

Mode :numeric Mean :5.054e+15 Mean : 660.7 Mean :497.7

3rd Qu.:5.385e+15 3rd Qu.: 965.0 3rd Qu.:750.0

Max. :5.600e+15 Max. :2000.0 Max. :999.0

MiddleInitial Surname ZipCode TelephoneNumber MerchantID TransactionID

Length:5203 Length:5203 Min. : 1001 Length:5203 Length:5203 Min. :2.243e+07

Class :character Class :character 1st Qu.:27606 Class :character Class :character 1st Qu.:3.809e+10

Mode :character Mode :character Median :47906 Mode :character Mode :character Median :7.659e+10

Mean :50109 Mean :7.652e+10

3rd Qu.:76701 3rd Qu.:1.145e+11

Max. :99901 Max. :1.536e+11

> View(IT633\_Session\_12b\_CreditCard)

Error in View : object 'IT633\_Session\_12b\_CreditCard' not found

> library(readr)

> IT633\_Session\_12b\_CreditCard <- read\_csv("C:/DM SEM 2/week 14/IT633\_Session\_12b\_CreditCard.csv")

Parsed with column specification:

cols(

Date = col\_character(),

Time = col\_time(format = ""),

CCNumber = col\_double(),

Amount = col\_integer(),

CCType = col\_character(),

CCExpires = col\_character(),

CVV2 = col\_integer(),

GivenName = col\_character(),

MiddleInitial = col\_character(),

Surname = col\_character(),

ZipCode = col\_integer(),

TelephoneNumber = col\_character(),

MerchantID = col\_character(),

TransactionID = col\_double()

)

> View(IT633\_Session\_12b\_CreditCard)

> summary(Goteti)

Date Time CCNumber Amount CCType CCExpires CVV2

Length:5203 Length:5203 Min. :4.485e+15 Min. : 1.0 Length:5203 Length:5203 Min. : 2.0

Class :character Class1:hms 1st Qu.:4.716e+15 1st Qu.: 318.5 Class :character Class :character 1st Qu.:240.0

Mode :character Class2:difftime Median :5.157e+15 Median : 627.0 Mode :character Mode :character Median :499.0

Mode :numeric Mean :5.054e+15 Mean : 660.7 Mean :497.7

3rd Qu.:5.385e+15 3rd Qu.: 965.0 3rd Qu.:750.0

Max. :5.600e+15 Max. :2000.0 Max. :999.0

GivenName MiddleInitial Surname ZipCode TelephoneNumber MerchantID TransactionID

Length:5203 Length:5203 Length:5203 Min. : 1001 Length:5203 Length:5203 Min. :2.243e+07

Class :character Class :character Class :character 1st Qu.:27606 Class :character Class :character 1st Qu.:3.809e+10

Mode :character Mode :character Mode :character Median :47906 Mode :character Mode :character Median :7.659e+10

Mean :50109 Mean :7.652e+10

3rd Qu.:76701 3rd Qu.:1.145e+11

Max. :99901 Max. :1.536e+11

> View(Goteti)

> nstall.packages("tree")

Error: could not find function "nstall.packages"

> install.packages("tree")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/tree\_1.0-37.zip'

Content type 'application/zip' length 121726 bytes (118 KB)

downloaded 118 KB

package ‘tree’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("party")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/party\_1.2-3.zip'

Content type 'application/zip' length 715135 bytes (698 KB)

downloaded 698 KB

package ‘party’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("rpart")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/rpart\_4.1-11.zip'

Content type 'application/zip' length 925146 bytes (903 KB)

downloaded 903 KB

package ‘rpart’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("car")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/car\_2.1-4.zip'

Content type 'application/zip' length 1483848 bytes (1.4 MB)

downloaded 1.4 MB

package ‘car’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("mlbench")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/mlbench\_2.1-1.zip'

Content type 'application/zip' length 1033476 bytes (1009 KB)

downloaded 1009 KB

package ‘mlbench’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("mboost")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/mboost\_2.7-0.zip'

Content type 'application/zip' length 2155156 bytes (2.1 MB)

downloaded 2.1 MB

package ‘mboost’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("textir")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/textir\_2.0-4.zip'

Content type 'application/zip' length 239898 bytes (234 KB)

downloaded 234 KB

package ‘textir’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("class")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/class\_7.3-14.zip'

Content type 'application/zip' length 101236 bytes (98 KB)

downloaded 98 KB

package ‘class’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("e1071")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/e1071\_1.6-8.zip'

Content type 'application/zip' length 894886 bytes (873 KB)

downloaded 873 KB

package ‘e1071’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> install.packages("randomForest")

Installing package into ‘C:/Users/gotet/OneDrive/Documents/R/win-library/3.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.3/randomForest\_4.6-12.zip'

Content type 'application/zip' length 179017 bytes (174 KB)

downloaded 174 KB

package ‘randomForest’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\gotet\AppData\Local\Temp\RtmpInRgs6\downloaded\_packages

> library(textir)

Loading required package: distrom

Loading required package: Matrix

Loading required package: gamlr

Loading required package: parallel

Warning message:

package ‘textir’ was built under R version 3.3.3

> library(MASS)

> str(Goteti)

Classes ‘tbl\_df’, ‘tbl’ and 'data.frame': 5203 obs. of 14 variables:

$ Date : chr "1/1/2015" "1/1/2015" "1/1/2015" "1/1/2015" ...

$ Time :Classes 'hms', 'difftime' atomic [1:5203] 19390 19390 19390 19390 19390 ...

.. ..- attr(\*, "units")= chr "secs"

$ CCNumber : num 5.27e+15 5.33e+15 5.34e+15 5.47e+15 5.58e+15 ...

$ Amount : int 44 469 677 245 319 1 1 1 1 776 ...

$ CCType : chr "MasterCard" "MasterCard" "MasterCard" "MasterCard" ...

$ CCExpires : chr "20-May" "16-Nov" "19-May" "18-Jul" ...

$ CVV2 : int 59 233 955 754 842 824 263 89 521 908 ...

$ GivenName : chr "Sabine" "Emeterio" "Astolfo" "William" ...

$ MiddleInitial : chr "S" "E" "A" "S" ...

$ Surname : chr "Oster" "Tijerina" "Villegas" "Mason" ...

$ ZipCode : int 27801 45223 97230 33830 68039 16101 46625 61820 77480 2141 ...

$ TelephoneNumber: chr "252-500-7984" "513-542-6185" "503-260-0447" "813-267-5229" ...

$ MerchantID : chr "HouseMart\_2191" "HouseMart\_2191" "HouseMart\_2191" "BOFAATM\_NW2301" ...

$ TransactionID : num 1.78e+08 1.84e+08 1.94e+08 2.72e+08 2.95e+08 ...

- attr(\*, "spec")=List of 2

..$ cols :List of 14

.. ..$ Date : list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ Time :List of 1

.. .. ..$ format: chr ""

.. .. ..- attr(\*, "class")= chr "collector\_time" "collector"

.. ..$ CCNumber : list()

.. .. ..- attr(\*, "class")= chr "collector\_double" "collector"

.. ..$ Amount : list()

.. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"

.. ..$ CCType : list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ CCExpires : list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ CVV2 : list()

.. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"

.. ..$ GivenName : list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ MiddleInitial : list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ Surname : list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ ZipCode : list()

.. .. ..- attr(\*, "class")= chr "collector\_integer" "collector"

.. ..$ TelephoneNumber: list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ MerchantID : list()

.. .. ..- attr(\*, "class")= chr "collector\_character" "collector"

.. ..$ TransactionID : list()

.. .. ..- attr(\*, "class")= chr "collector\_double" "collector"

..$ default: list()

.. ..- attr(\*, "class")= chr "collector\_guess" "collector"

..- attr(\*, "class")= chr "col\_spec"

> plot(Goteti)

Error in plot.window(...) : need finite 'xlim' values

In addition: There were 12 warnings (use warnings() to see them)

> par(mfrow=c(3,3), mai=c(.3,.6,.1,.1))

> plot(Answer1 ~ ., data=Goteti, col=c(grey(.2),2:6))

Error in eval(expr, envir, enclos) : object 'Answer1' not found

> View(Goteti)

> plot(CCType ~ ., data=Goteti, col=c(grey(.2),2:6))

Hit <Return> to see next plot:

Error in plot.window(...) : need finite 'xlim' values

In addition: Warning messages:

1: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

2: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

3: In min(x) : no non-missing arguments to min; returning Inf

4: In max(x) : no non-missing arguments to max; returning -Inf

5: In min(x) : no non-missing arguments to min; returning Inf

6: In max(x) : no non-missing arguments to max; returning -Inf

> plot(CVV2 ~ ., data=Week8\_cheating\_1\_, col=c(grey(.2),2:6))

Error in eval(expr, envir, enclos) : object 'Week8\_cheating\_1\_' not found

> plot(CVV2 ~ ., data=Goteti, col=c(grey(.2),2:6))

Error in plot.window(...) : need finite 'xlim' values

In addition: Warning messages:

1: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

2: In min(x) : no non-missing arguments to min; returning Inf

3: In max(x) : no non-missing arguments to max; returning -Inf

> na.rm=TRUE

> par(mfrow=c(3,3), mai=c(.3,.6,.1,.1))

> plot(CCType ~ ., data=Goteti, col=c(grey(.2),2:6))

Hit <Return> to see next plot:

Error in plot.window(...) : need finite 'xlim' values

In addition: Warning messages:

1: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

2: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

3: In min(x) : no non-missing arguments to min; returning Inf

4: In max(x) : no non-missing arguments to max; returning -Inf

5: In min(x) : no non-missing arguments to min; returning Inf

6: In max(x) : no non-missing arguments to max; returning -Inf

> plot(CVV2 ~ ., data=Goteti, col=c(grey(.2),2:6))

Error in plot.window(...) : need finite 'xlim' values

In addition: Warning messages:

1: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

2: In min(x) : no non-missing arguments to min; returning Inf

3: In max(x) : no non-missing arguments to max; returning -Inf

> na.na=TRUE

> par(mfrow=c(3,3), mai=c(.3,.6,.1,.1))

> plot(CCType ~ ., data=Goteti, col=c(grey(.2),2:6))

Hit <Return> to see next plot:

Error in plot.window(...) : need finite 'xlim' values

In addition: Warning messages:

1: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

2: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

3: In min(x) : no non-missing arguments to min; returning Inf

4: In max(x) : no non-missing arguments to max; returning -Inf

5: In min(x) : no non-missing arguments to min; returning Inf

6: In max(x) : no non-missing arguments to max; returning -Inf

> plot(CVV2 ~ ., data=Goteti, col=c(grey(.2),2:6))

Error in plot.window(...) : need finite 'xlim' values

In addition: Warning messages:

1: In xy.coords(x, y, xlabel, ylabel, log) : NAs introduced by coercion

2: In min(x) : no non-missing arguments to min; returning Inf

3: In max(x) : no non-missing arguments to max; returning -Inf

> n=length(Goteti$type)

Warning message:

Unknown column 'type'

> n=length(Goteti$ZipCode)

> nt=500

> set.seed(1) ## to make the calculations reproducible in repeated runs

> Goteti1=Goteti[,c(4,1)]

> x=Goteti[,c(4,1)]

> x[,1]=(x[,1]-mean(x[,1]))/sd(x[,1])

Using the absolute value frequency we can identify the output of the for identifying the frequency of the data.  
  
R Code:   
  
mydata[,"CCNumber"] <- apply( mydata,1, function(x) { paste0( sum(freq\_matrix[x]) ,"/", length(x) )})

Mydata  
  
  
  
  
  
After that we can use Model based reasoning to make a effective and required logical analysis and making an interpretation on this.   
  
  
  
R Code:   
  
  
f <- stl(Goteti$Amount, "periodic", robust=TRUE)

print(f)

(outliers <- which(f$CCNumber<1e-8))

## set layout

op1 <- par(mar=c(0, 4, 0, 3), oma=c(5, 0, 4, 0), mfcol=c(4, 1))

plot(f, set.pars=NULL)

sts <- f$time.series

## plot outliers

points(time(sts)[outliers], 0.8\*sts[,"CCNumber"][outliers], pch="x", col="red")

par(op) # reset layout

Now identifying the average of Amount of card utilized and make it run utlizied.   
  
  
R Code:   
  
  
mean(Goteti$Amount)  
660.6958

And also identifying the summary is   
  
  
> summary(Goteti)

Date Time CCNumber Amount CCType CCExpires CVV2

Length:5203 Length:5203 Min. :4.485e+15 Min. : 1.0 Length:5203 Length:5203 Min. : 2.0

Class :character Class1:hms 1st Qu.:4.716e+15 1st Qu.: 318.5 Class :character Class :character 1st Qu.:240.0

Mode :character Class2:difftime Median :5.157e+15 Median : 627.0 Mode :character Mode :character Median :499.0

Mode :numeric Mean :5.054e+15 Mean : 660.7 Mean :497.7

3rd Qu.:5.385e+15 3rd Qu.: 965.0 3rd Qu.:750.0

Max. :5.600e+15 Max. :2000.0 Max. :999.0

GivenName MiddleInitial Surname ZipCode TelephoneNumber MerchantID TransactionID

Length:5203 Length:5203 Length:5203 Min. : 1001 Length:5203 Length:5203 Min. :2.243e+07

Class :character Class :character Class :character 1st Qu.:27606 Class :character Class :character 1st Qu.:3.809e+10

Mode :character Mode :character Mode :character Median :47906 Mode :character Mode :character Median :7.659e+10

Mean :50109 Mean :7.652e+10

3rd Qu.:76701 3rd Qu.:1.145e+11

Max. :99901 Max. :1.536e+11

Now let see which card fradulalized used by master card are 2932 and using visa card are 2272 by using this we can understand this the similar pattern using a classic identification it happens that the possibility of master card getting in Fraduality and make with using outliers identified with regular pattern of card used in regular zip code is possible.