

Raw Data Input into RStudio

Venukanan Subenthiran

2022-09-18

R Markdown

Importing Data

```
Link <- "C:/Users/Venu/OneDrive/Data Science - Ryerson Courses/CIND 820 - Big Data Analytics Project/Capestone Dataset/Credit Card Transactions Fraud Detection Dataset/fraudTrain.csv"
```

```
fraudTrain.db <- read.csv(file = Link, header = TRUE, sep = ",", na.strings = "")
```

```
Link2 <- "C:/Users/Venu/OneDrive/Data Science - Ryerson Courses/CIND 820 - Big Data Analytics Project/Capestone Dataset/Credit Card Transactions Fraud Detection Dataset/fraudTest.csv"
```

```
fraudTest.db <- read.csv(file = Link2, header = TRUE, sep = ",", na.strings = "")
```

```
fraudTotal.db <- rbind(fraudTrain.db, fraudTest.db)
```

```
summary(fraudTotal.db)
```

```

##      X      trans_date_trans_time      cc_num      merchant
## Min.   :      0      Length:1852394      Min.   :6.042e+10      Length:1852394
## 1st Qu.: 231549      Class :character      1st Qu.:1.800e+14      Class :character
## Median : 463098      Mode  :character      Median :3.521e+15      Mode  :character
## Mean   : 537193                                     Mean   :4.174e+17
## 3rd Qu.: 833576                                     3rd Qu.:4.642e+15
## Max.   :1296674                                     Max.   :4.992e+18
##      category      amt      first      last
## Length:1852394      Min.   :      1.00      Length:1852394      Length:1852394
## Class :character      1st Qu.:      9.64      Class :character      Class :character
## Mode  :character      Median :      47.45      Mode  :character      Mode  :character
##                                     Mean   :      70.06
##                                     3rd Qu.:      83.10
##                                     Max.   :28948.90
##      gender      street      city      state
## Length:1852394      Length:1852394      Length:1852394      Length:1852394
## Class :character      Class :character      Class :character      Class :character
## Mode  :character      Mode  :character      Mode  :character      Mode  :character
##
##
##
##      zip      lat      long      city_pop
## Min.   : 1257      Min.   :20.03      Min.   : -165.67      Min.   :      23
## 1st Qu.:26237      1st Qu.:34.67      1st Qu.: -96.80      1st Qu.:      741
## Median :48174      Median :39.35      Median : -87.48      Median :      2443
## Mean   :48813      Mean   :38.54      Mean   : -90.23      Mean   : 88644
## 3rd Qu.:72042      3rd Qu.:41.94      3rd Qu.: -80.16      3rd Qu.: 20328
## Max.   :99921      Max.   :66.69      Max.   : -67.95      Max.   :2906700
##      job      dob      trans_num      unix_time
## Length:1852394      Length:1852394      Length:1852394      Min.   :1.325e+09
## Class :character      Class :character      Class :character      1st Qu.:1.343e+09
## Mode  :character      Mode  :character      Mode  :character      Median :1.357e+09
##                                     Mean   :1.359e+09
##                                     3rd Qu.:1.375e+09
##                                     Max.   :1.389e+09
##      merch_lat      merch_long      is_fraud
## Min.   :19.03      Min.   : -166.67      Min.   :0.00000
## 1st Qu.:34.74      1st Qu.: -96.90      1st Qu.:0.00000
## Median :39.37      Median : -87.44      Median :0.00000
## Mean   :38.54      Mean   : -90.23      Mean   :0.00521
## 3rd Qu.:41.96      3rd Qu.: -80.25      3rd Qu.:0.00000
## Max.   :67.51      Max.   : -66.95      Max.   :1.00000

```

```
colnames(fraudTotal.db)
```

```
## [1] "X" "trans_date_trans_time" "cc_num"
## [4] "merchant" "category" "amt"
## [7] "first" "last" "gender"
## [10] "street" "city" "state"
## [13] "zip" "lat" "long"
## [16] "city_pop" "job" "dob"
## [19] "trans_num" "unix_time" "merch_lat"
## [22] "merch_long" "is_fraud"
```