Assignment 2.2

Name : Y Vasudev

Batch : DA with R , Excel and Tableau

1. Read multiple JSON files into a directory to convert into a dataset.

I have files text1, text2, text3 in the directory JSON.

> library(jsonlite)

> install.packages("jsonlite")

> setwd( path of the directory of Json files )

> list.files()

> json.list=jsonfile.list(“path of the json files “)

> json\_data\_frame = as.data.frame(json.list)

> print (json\_data\_frame)

2. Parse the JSON into a data frame.

>library(jsonlite)

>json <- '[

>{"name" : NULL,

>"release\_date\_local" : NULL,

>"title" : 3(2011),

>"opening\_weekend\_take" : 1234,

>"year" : 2011,

>"release\_date\_wide" : [2011-09-16],

>"gross" : 59954

>}

>]'

>mydf <- fromJSON(json, simplifyVector = TRUE, simplifyDataFrame = >simplifyVector)

>mydf

3. Write a script for Variable Binning using R.

> df$cat = cut (df$v,c(-Inf,6,12,Inf))

> df = data.frame(v=sample(1:60,1000,TRUE))

> head(df)

v

1 10

2 16

3 29

4 5

5 22

6 52

> df$cat = cut (df$v,c(-Inf,6,12,Inf)) # this is splitting into diff’ categor

> head(df) # ies ( -infinity to 6 , 6 to 12 , 12

v cat # to +infinity )

1 10 (6,12]

2 16 (12, Inf]

3 29 (12, Inf]

4 5 (-Inf,6]

5 22 (12, Inf]

6 52 (12, Inf]

> table(df$cat)

(-Inf,6] (6,12] (12, Inf]

122 108 770

Also with diff more categories of including 50 ..

> df$cat = cut (df$v,c(-Inf,6,12,50,Inf))

> head(df)

v cat

1 10 (6,12]

2 16 (12,50]

3 29 (12,50]

4 5 (-Inf,6]

5 22 (12,50]

6 52 (50, Inf]

> table(df$cat)

(-Inf,6] (6,12] (12,50] (50, Inf]

122 108 595 175

Square bracket after a # means that # is included

Round bracket before a # means that # is excluded