**Session -2- introduction to Working With R**

**Assignment2**

**Problem Statement:**

**Read multiple JSON files into a directory to convert into a dataset. I have files text1, text2, text3 in the directory JSON.**

>library(json)

>filenames <- list.files(pattern = “\*.json”)

#will give a character vector with each filename represented by an entry

Now to import all the JSON files into R as a dataframe

myJSON <- lapply(filenames,function(x) fromJSON(file=x)

#returns a list where each element is one of the JSON files.

1. Parse the following JSON into a data frame.

js<-'{

"name": null, "release\_date\_local": null, "title": "3 (2011)",

"opening\_weekend\_take": 1234, "year": 2011,

"release\_date\_wide": "2011-09-16", "gross": 59954

}'

The JSON file is read by R using the function **from JSON()**. It is stored as a list in R.

>library(‘rjson’)

>result <- fromJSON(file = “input.JSON”)

>print(result)

>json\_data\_frame <- as.data.frame(result)

>print(json\_data\_frame)

1. Write a script for Variable Binning using R.

**Binning** is the term used in scoring modeling for what is also known in Machine Learning as **Discretization**, the process of transforming a continuous characteristic into a finite number of intervals (the bins), which allows for a better understanding of its distribution and its relationship with a binary variable. The bins generated by the this process will eventually become the attributes of a predictive characteristic, the key component of a Scorecard.

dataset <- c(4,7,9,1,10,15,18,19,3,16,10,16,12,22,2,23,16,17)

>data.frame(dataset, bin=cut(dataset, c(1,4,9,17,23), include.lowest=TRUE))

I would like to create four dummy categories, in which I bin the continuous dataset by custom breaks . .. for example: 1:4, 5:9, 10:17, 18:23.

The output dummy categories would have the same length as the original continuous vector (18 in this case)

