ASSIGNMENT #1

1. What are the different steps of data analysis and what is done during each step.

**Problem definition and planning**

The starting point for data analysis is a data set which contains the measured or collected data values represented as numbers or text. The business unit has to decide on objectives for the data science teams. These objectives usually require significant data collection and analysis. We need a measurable way to know if the business is advancing toward its goals. Key metrics or performance indicators must be identified early in the process.

**Preparing data**

Pre-processing/Data preparation involves merging data into a table by integrating the data from multiple sources. Once the data is in a tabular format, it should be fully characterized. The data should be cleaned by resolving ambiguities and errors, removing redundant and problematic data, and eliminating columns of data irrelevant to the analysis.

**Data Analysis**

In this step data is studied to build an intuition for the data. Data is viewed from different perspectives. Calculate basic summary descriptive statistics on data. You need to learn the shape, size, type and general layout of the data that you have.

**Deployment**

In this step we present the final results.

2. What are the different ways to create a vector in R?

**First Method : Using c() function:**

v1<-c(1,2,3,4,5)

**Second method using range:**

v2<-1:5

v3<-2:-2

**Third Method: Using seq()**

v4<-seq(1,5,by=2)

v5<-seq(1,5,length.out = 3)

**Fourth method: Using rep()**

v6<-rep(1:3,4)

v7<-rep(c(1,3,2),each=4)

3. Create the following vector and check the class (‘x’,’x’,‘x’,1,3,5,7,ti,2,4,6,8,10)

v8<-c('x','x','x',1,3,5,7,ti,2,4,6,8,10)

class(v8)

[1] "character"

4. Create a vector of positive odd integers less than 100.

V9<-seq(1,100,2)

5. Remove the values greater than 60 and less than 80

V9<-v9[!(v9>60 & vt9<80)]

V9

6. Write a function to return standard deviation, mean and median of the vector.

Verify with example.

sdmm<-function(v10){

return(list(SD=sd(v10),Mean=mean(v10),Median=median(v10)))

}

x<-sdmm(c(1,2,3,4,5))

x

$SD

[1] 1.581139

$Mean

[1] 3

$Median

[1] 3

7. Create two matrices of the form from the given set of numbers} in

Two ways X1 = {2,3,7,1,6,2,3,5,1} and x2={3,2,ti,0,7,8,5,8,2}

x1<-c(2,3,7,1,6,2,3,5,1)

x2<-c(3,2,ti,0,7,8,5,8,2)

M1<-matrix(x1,3)

M2<-matrix(x2,3)

M2

M3<-matrix(x1,3,byrow=TRUE)

M4<-matrix(x2,3,byrow=TRUE)

M3

8. Find the matrix product.

M5<-M1\*M2

M6<-M3\*M4

9. Find the class of ‘iris' dataframe, find the class of all the columns of ‘iris’

Get the summary. Get rownames, column names. Get the number of rows and number of columns.

class(iris)

class(colnames(iris))

summary(iris)

rownames(iris)

colnames(iris)

length(rownames(iris))

length(colnames(iris))

OR

dim(iris)

10. Get the last two rows in the last 2 columns from iris dataset.

iris[149:150,4:5]