

Assignment - 10

1. Import dataset from the following link:
<https://archive.ics.uci.edu/ml/machine-learning-databases/00360/>
Perform the below written operations:
 - a. Read the file in Zip format and get it into R

```
airquality <- read.csv("AirQuality.csv")  
View(airquality)
```
 - b. Create Univariate for all the columns.

```
summary(airquality)  
describe(airquality)
```
 - c. Check for missing values in all columns.

```
summary(airquality)
```
 - d. Impute the missing values using appropriate methods

```
library(mice)  
  
mean(newair$Solar.R)  
mean(newair$Solar.R, na.rm = T)  
#impute my mean  
newair$Solar.R[is.na(newair$Solar.R)] <- mean(newair$Solar.R, na.rm = T)
```
 - e. Create bi-variate analysis for all relationships

```
library(psych)  
pairs.panels(airquality[,c(1,2,3,4,5,6)],  
             method = "pearson", # correlation method  
             hist.col = "red",  
             density = TRUE, # show density plots  
             ellipses = TRUE, # show correlation ellipses  
             lm=TRUE,  
             main = "Bivariate Scatter plots with Pearson Correlation &  
Histogram"  
             )
```
 - f. Test relevant hypothesis for valid relations
Some of the tests are
Ho: Mean of first variable - Mean of 2 variable is equal to 0
Ha: Mean of first variable - Mean of 2 variable is not equal to 0

```
t.test(x=airquality$Ozone, y=airquality$Solar.R, alternative =  
"two.sided", mu=0, paired = TRUE)  
t.test(x=airquality$Temp, y=airquality$Wind, alternative =  
"two.sided", mu=0, paired = TRUE)  
t.test(x=airquality$Ozone, y=airquality$Temp, alternative =  
"two.sided", mu=0, paired = TRUE)  
t.test(x=airquality$Day, y=airquality$Solar.R, alternative =  
"two.sided", mu=0, paired = TRUE)
```
 - g. Create cross tabulations with derived variables

```
mytable <- xtabs(~x+y, data = airquality)  
#crosstabulate  
  
library(gmodels)  
CrossTable(x, y)
```
 - h. check for trends and patterns in time series
 - i. Find out the most polluted time of the day and the name of the chemical compound.

