Assignment 10.1

Name : Y Vasudev

Batch : DA with R , Excel and Tableau

Import dataset from the following link: AirQuality Data Set

Perform the following written operations:

1. Read the file in Zip format and get it into R.

forecasturl = paste('https://archive.ics.uci.edu/ml/machine-learning-databases/00360/', 'AirQualityUCI.zip', sep='')# create a temporary directorytd = tempdir()# create the placeholder filetf = tempfile(tmpdir=td, fileext=".zip")# download into the placeholder filedownload.file(forecasturl, tf)# get the name of the first file in the zip archivefname = unzip(tf, list=TRUE)$Name[1]fname# unzip the file to the temporary directoryunzip(tf, files=fname, exdir=td, overwrite=TRUE)# fpath is the full path to the extracted filefpath = file.path(td, fname)fpathd = read.csv(fpath,sep = ";")View(d)

2. Create Univariate for all the columns.

univariateTable(~date,time,data(airqu))

3. Check for missing values in all columns.

d1[is.na(PT08.S1.CO.)] = mean(PT08.S1.CO.)

d1[is.na(d1$PT08.S1.CO.)] = mean(d1$PT08.S1.CO.)

for(i in 1:ncol(airq)){

airq[is.na(airq[,i]),i] <- mean(airq[,i], na.rm = TRUE)

}

4. Impute the missing values using appropriate methods.

imp = impute(airq, classes = list(integer = imputeMean(), factor = imputeMode()),

dummy.classes = "integer")

5. Create bi-variate analysis for all relationships.

6. Test relevant hypothesis for valid relations.

7. Create cross tabulations with derived variables.

8. Check for trends and patterns in time series.

table(data$a,data$b)

plot.ts(datasetname)

souvenir\_decomp=decompose(souvenir\_ts)

plot(souvenir\_decomp)

9. Find out the most polluted time of the day and the name of the chemical compound.

**NOT YET COVERED IN THE SESSION S**