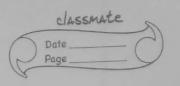
Name : Prajakta	Keer
ROU NO: 33231	



ASSIGNMENT 5

Ain: Perform different operations using R/Python.

Problem Statement: Perform the following operations using R/Python on the Amazon Book Review and Facebook metric data sets

- a) Create data subsets b) Merge Data c) Sort Data
- d) Transposing Data e) Melling data to long format
- f) casting data to wide format

Objectives :

· To learn R/Python programming

· To learn different data preprocessing techniques

Theory: R Programming Language: Developed by Ross
Ihaka and Robert Gentleman in 1993. R possess an
extensive catalog of statistical and graphical methods.
It include ML algorithms, linear regression, time series,
statistical reference, etc.

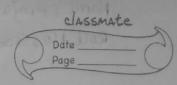
Dataframer: It is a table or a 2D array-like structure in which each column contains values of one variable and each now contains one set of variables from each column Characteristics of DF:) Column names should not be empty 2) Row names should be wright

3) The data stored can be of numeric, factor or character type. 4) Each col. should contain same no of data I tems.

Reading CSV files:

read · csv2 (fire, header = True, sep = ";", dec = ",", ...

- · file : path to the file containing data to be imported
- · sep: the field seperator character
 - dec: the character used in the file for decimal points.
 Eg: my Data = read. CSV2 ("basic.csv")



Subetting Data : i) Selecting variables

Sub1 = dara [c("v1", "v2", "v3")]

2) Using subset function

Sub2 = subset (data, age > = 20 | age < 10) Merge Data : 1) Adding Columns

total = merge (dataA, dataB, by = "TD").

2) Adding Rows

total = rbind (clata A, dataB) Sort Data: # sort by mpg (accending) and up (descending), newdata = care [order (mpg, -yl)] trasdata = t (data) Metting Data: We have to use the reshape package (g: library (reshape)

modata = melt (mydata, i'd = c ("i'd", "time")) Melting is done to organize the duta. Using melt (), dataframe is converted into long format and stretches the data frame. Casting Data & cout () is used to convert long format data back to some aggregated form.

eg: subfincans = cast (mdata, id~ variable, mean) Conclusion: I learnt various preprocessing techniques in R programming language and performed operations like sorting, merging, melling and casting on the datasets.