Fundamentals of Big Data **Analytics**

Home Task - 01

Deadline: 10th March 2022 (23:59)

Lab Homework Instructions:

1. Carefully read and follow all instructions of both questions Q1 & Q2.

2. You can search the basics of python, concepts, and syntax online

3. No Submission would be done after Deadline, its already too much flexible.

4. Submission Pattern, just a Jupiter notebook renamed as follow

   rollnumber_firstname_h1.ipynb

   EG: Roll-11_Muhammad_h1.ipynb

5. Do keep in mind that different tools would be used for checking Plagiarism

6. Plagiarism will result straight F.

Q1. Groceries Data Manipulation

Task 1

Create a Data Frame sales that looks like below:

Month	Bananas	Apples
September	200	120
October	165	90

Task 2

Create a series variable M that looks like:

```
MIS      67
DBMS     93
A&F      56
CA       67
dtype: int64
```

Task 3

Read the following csv dataset of Groceries given to you in zip folder into a
DataFrame called grocer

Task 4

Select the itemdescription column from grocer and assign the result to the variable itemdesc.

Task 5

Select the first value from the member column of grocer and assign it to variable first_member.

Task 6

Select the first 20 row of data and save it to variable first_rows.

Task 7

Create a variable selected_grocer containing the values of date and item description column only for purchase made in 2015

Task 8

Create a variable containing the member and date columns of the first 500 records.

Q2 Automobile Data Cleaning

Task 1

Find the ? in given data and replace it with nan

Task 2

Count Missing values in each column and display the results.

Task 3

Calculate the mean value for the 'horsepower' column:

Task 4

Replace "NaN" by mean value:

Task 5

Find the car that have maximum highway mile per gallon

Task 6

Find all honda car details.

Task 7

Count total cars per company

Task 8

Find each company's highest price car