Assignment - 4

Date : / / 20

- Title:- write Python code that loads any dataset from & does som basic dat cleaning, Add Component on dataset

- Problem statement:- To python code for that loads any data set & plot the graph.

- pre-lab:- A basic understanding of computer Programming terminologies, A basic understanding of any of the Programming lang will help in understanding the python programming & datascience concepts.

- Theory:-

1. Step 1 Acquiring Data:-
Step one, acquiring data, the first step in the data science process is to acquire the data. you need to obtain the source material before analyzing or acting on it.

2. Step 2-A:- exploring Data.
- After you've put together the data that you need for you appln you might be tempted to immediatly build models to analyze the data.

3. Step 2-B pre-Prossing Data:-
The raw data that you get directly from your source are never in the format that you need to

Pooja

Date : / / 20

perform analysis on. There are two main focus in the data pre-processing step. The first is to clean the data to address data quality issues.

4. Step 3 :- Analysing Data :-

Analysing data Now that you have your data nicely prepared, the next step is to analyze the data, Data analysis involves building a model from your data, which is called input data The input data is used by the analysis techinique to build a model.

5. Step 4 : communicating Result :-

reporting insights the fourth step in our data science process is reporting the insights gained from our analysis this a very important step to communicate your insight & make a case for what actions should follow, it can change shape based on your audience and should not be taken lightly. So how do you get started? the first thing to do is to look all your analysis result & decide what to present or report as the biggest value or biggest set of values.

- Dropping columns in a 'Data frame

often you'll find not all the categories of data useful to you

Pooja

Date : / / 20

for e.g you might have a dataset containing student information. but want to focus on containing student inform^{n}. Pandas provided a handy way of removing unwanted column or rows from a Dataframe with the drop() fun. Let's look at a simple e.g. where we drop a no. of columns from a Data frame.
Let us now see how we can handle missing values using pandas.

- Conclusion :- Thus student can implement note book for (Perform steps (step-4 data science steps for any data (IPL-data) by using python tools like pandas, matplotlib, numpy etc.

Pooja