NATIONAL UNIVERSITY OF COMPUTER AND EMERGING
SCIENCES ISLAMABAD

Data Analysis and Visualization Fall-2022

Assignment No. 1
Exploratory Data Analysis using Visualizations

Due Date: 11:59 PM 19th, September 2022.   **Total Marks: 50**

Instructions:
• Zero marks will be awarded to the students involved in plagiarism.
• You must submit a single zip folder of code file, output, project report, data & data sources in Google Classroom. Folder name should be your name and Roll No.
• Be prepared for viva or anything else after the submission of assignment.
• **NO queries will be entertained before two days of submission.**

Description: (Data analysis on the flood in Pakistan)
Pakistan continues to endure the worst flooding in its history and many towns and villages are still being evacuated as the floodwaters move downstream through Sindh and eventually into the Arabian Sea. As of August 23rd, over 20 million people have been affected of whom more than four million are homeless, and more than eight million are in urgent need of humanitarian aid. Recovering from a natural disaster of this magnitude will take many years and billions of dollars.

Tasks to perform:
Exploratory Data Analysis (EDA), also known as Data Exploration, is a step in the Data Analysis Process, where several techniques are used to better understand the dataset being used. 'Understanding the dataset' can refer to several things including but not limited to.
  • Extracting important variables and leaving behind useless variables.
  • Identifying outliers, missing values, or human error.
  • Understanding the relationship(s), or lack of, between variables.
  • Ultimately, maximizing your insights of a dataset and minimizing potential error that may occur later in the process.
In short, EDA refers to performing visualizations & identifying significant patterns, such as correlated features, missing data, and outliers. EDAs are also essential for providing hypotheses for why these patterns occur

<div align="center">

Rubrics:

Data Collection	10
Plots' for Analysis	15
Reporting of Plotting	10
Analysis and Insights	15
Total	**50**

</div>

Instructions:
1. You are to collect the latest data on the flood situation in Pakistan. You cannot share your collected data with any other student otherwise his/her marks will be zero.
2. Do some data preprocessing to make your data more useful.
3. Do some processing on your data and make plots for better understanding and visualization.
4. Briefly describe your plots.
5. After the analysis you must write a report on what you understand from the analysis with the help of some hypotheses/questions.

Note:
1. You can use Python libraries like BeautifulSoup or Selenium to collect/scrape data.
2. You must provide the source from where you collected the data.
3. Repeating your analysis can be an ongoing activity to refine your plot and insights.
4. If you are able to do some good analysis with useful results, we can publish it on any article or news website.

Submission:

You are required to submit a short report with response to each of your analysis with supporting graphs, tables, and stats. Python code, output, analysis report, data & data sources for the experiment must be submitted. We shall run your code ourselves to verify your graphs and test results.

<div align="center">

</div>