IE6600 Computation and Visualization for Analytics
Summer 1 Semester 2024
Project 2

Project Assignment:
Advanced-Data Analysis and Visualization Using Seaborn

Objective:
This project is designed to enhance students' skills using the Seaborn Library for advanced data visualization and statistical data analysis. Groups will select distinct datasets from **data.gov**, exploring various sectors like health, environment, finance, or transportation. The primary focus is creating sophisticated static visualizations using Seaborn, saving these visualizations as image files, and effectively presenting the analysis through a report or a webpage. Additionally, students must compile a comprehensive report detailing their findings and methodologies.

Preparation:
- Select a unique dataset from **data.gov** and have it approved by the teaching assistant to ensure no duplication.
- Familiarize yourself with the Seaborn library for data visualization.

Tasks:
1. Dataset Selection and Confirmation:
    - Choose a dataset from data.gov.
    - Secure confirmation from the teaching assistant to verify its uniqueness.
2. Data Acquisition and Inspection:
    - Download the dataset and integrate it into your analysis environment.
    - Perform a preliminary review to understand its structure and content.
3. Data Cleaning and Preparation:
    - Address any issues of missing data, duplicates, and inconsistencies.
    - Perform necessary data type conversions and normalizations.
    - Encode categorical data where applicable.
4. Exploratory Data Analysis (EDA) Using Seaborn:
    - Utilize Seaborn for creating detailed and informative static visualizations.
    - Save all visualizations as image files for inclusion in the presentation and report.
    - Apply statistical analysis techniques to explore and interpret the data.
5. Reporting:
    - Compile a full report detailing your data cleaning, exploratory analysis, and any advanced methods used, along with your interpretations and conclusions.
    - Develop a webpage showcasing your visualizations and key findings. **(Optional)**
6. Advanced Analysis (Optional):
    - Employ more complex analytical methods or additional datasets for a deeper exploration.

Deliverables:

- A Jupyter Notebook contains all the code, data cleaning, EDA, and advanced analysis steps.
- All visualizations are saved as image files.
- A webpage showcasing your findings. (Optional)
- A comprehensive report in PDF format detailing your entire analytical process and findings.

Grading Criteria:

- Creativity and relevance of the selected dataset.
- Effectiveness and sophistication of Seaborn visualizations.
- Depth and accuracy of statistical analysis.
- Quality and thoroughness of the final report.
- Professionalism and clarity of the webpage. (Extra points)
- Innovation and analytical depth in advanced analysis (if attempted).

Due Date:

- June 5, 2024

Please contact us if you have any questions or need clarification. We look forward to seeing your analytical prowess and creative visualizations!