CSIT 359/553 Exploratory Data Analysis and Visualization

Project 1: Tabular Data Visualization

Instructions: In the project, you need to prepare an idea and a data set from real world. Convert them to Pandas and apply multiple techniques for data analysis.

Group work: Both individual and group work are allowed in this project. Each group can include at most **3** students. All the names of group members should be indicated in the project design report.

About the data set:

You could find the data by your self or select from the following resources:

Stanford Large Network Dataset Collection	https://snap.stanford.edu/data/
Dataverse Network	https://dataverse.org/
Reddit Open Data	https://www.reddit.com/r/opendata/
CDC Data	https://www.cdc.gov/nchs/tools/index.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fnchs%2Fdata_access%2Fdata_tools.htm
World Bank Catalog	https://datacatalog.worldbank.org/
Metor Boston Data Common	https://datacommon.mapc.org/
COVID-19 Data Repository by Johns Hopkins University	https://github.com/CSSEGISandData/COVID-19

Don'ts

- Don't use a standard machine learning dataset (Kaggle, UCI ML Repository). These are pre-processed and only suitable for analysis, not for the whole DS process
- Don't pick a dataset where structured data is hard to extract, E.g.,
  - text-only, relying on advanced NLP,
  - extracting data from collection of PDFs,
  - running your own survey (it's hard to run a good survey)

Project Requirements

The project **MUST** includes the following techniques:

1. Exploratory Data Analysis

- Data Loading
- Data Cleaning
- Data Analysis using Descriptive Statistics
- Select **at least one** of the following techniques in data manipulating:
  - Data Wrangling
  - Data Aggregation
  - Time Series

2. Data Visualization

- Visualization Design
  - At least three different designs are required in your project
  - You need to implement your visualization in Python

- Description of Data Visualization Design
  - Describe the questions your visualization is designed to answer.
  - Describe the visualization you created and how its design evolved. (What marks and channels are used?)
  - Describe how the visualization can be used to answer the questions.

Presentation Requirements

A presentation for each team is required. Each team will get approx. 10 min for presentation. Please plan your talk accordingly. Slides are required during the presentation with the following contents:

- The description of the project, including the project objectives and the description of the data set - should be with reference to the data
- Description of the exploratory data analysis
- Data visualization, including the plots and the description of the design
- Live demo/ demo snapshots of execution of your program
- Conclusion from your observation

Project Submission

A final submission should include all the source code, data set and slides for the presentation.

CSIT 359/553: Exploratory of Data Analysis and Visualization

Rubric of Project 1

Project Title: _____

Student Names: _____

	Points out of Total
1. Exploratory Data Analysis	_____/ 10
a. Data Loading (2 points)	_____
b. Data Cleaning (3 points)	_____
c. Data Analysis using descriptive statistics (2 points)	_____
d. Other techniques in data manipulating (3 points)	_____
2. Data Visualization	_____/ 20
a. Three different design of data visualization in Python (15 points)	_____
b. Description of the design (5 points)	_____
3. Live demo / demo snapshots of execution	_____/ 10
a. The program can be executed successfully (6 points)	_____
b. Students can answer the questions about the source code (4 points)	_____
3. Project Presentation	_____/ 10
a. Presenters are well-prepared (2 points)	_____
b. Slides should present material in an informative manner (2 points)	_____
c. Presentation is logically organized and presenters appear to be fluid (2 points)	_____
d. There is a balance between high-level motivational material & technical detail (2 points)	_____
e. Presenters should respond well to questions and critique (2 points)	_____
Total Score	_____/ 50

Graders Comments: