Visualization of Complex Data

DATS 6401

Final Term Project- Proposal

We are currently launching the first phase of the final term project (data selection & proposal). The selected dataset must satisfy the following criteria:

- Pick an interesting and real-world dataset from the industry. You need to justify why the dataset is interesting.
- It must be a multivariate dataset with at least 50K observations. If you have an interesting dataset with less than 50K samples, please come forward and talk to me.
- It must contain numerical & categorical data with at least four for each category. Please keep in mind that you can also perform feature engineering and develop new features.
- It must come from a non-classified (public) database.
- If two students select the same dataset, the dataset will be assigned to the first student, and the second student needs to pick another dataset. First come first serve.

Write a paragraph proposal (A4 size page) and address the following:

- List of features. Define whether they are categorical or numerical. Is there any need for feature engineering? [10pts]
- List of static plots for numerical features and list of static plots for categorical features. [10pts]
- Draft of interactive dashboard with multiple tabs. You need to create the dashboard draft using MS PowerPoint. List of items to be displayed on the front end : data cleaning [various methods], outlier detection and removal, dimensionality reduction[PCA,], normality tests[three tests will be covered], data transformation[normalization, standardization…], loading data, plots of numerical features, plots for categorical features, statistics. [20pts]

You can select the dataset from the following repository.

- https://www.kaggle.com
- https://archive.ics.uci.edu/ml/index.php
- https://datasetsearch.research.google.com
- https://analyticsindiamag.com/top-10-popular-publicly-available-datasets-deep-learning-research/

Submission Guidelines

1. The deadline for submitting the term project proposal and the selected dataset is **Friday 21/2/2025**.
2. Submit the pdf of the proposal before the deadline.
3. Upload the Excel, CSV, JSON… of the selected dataset through the course BB shell under proposal and dataset.
4. Fill out the following shared Excel sheet with the selected dataset before the deadline:

https://docs.google.com/spreadsheets/d/12HjdU2PHMoOprPgxkZ_J3ZQSaZRTH-_YrvdIxwoRGQY/edit?pli=1&gid=0#gid=0

Please note: Make sure to sign in with your GWU email address to access the sheet to add your dataset details.