CEN445 INTRODUCTION TO DATA VISUALIZATION COURSE - ASSIGNMENT (27.10.2025)

Submission Deadline: November 16, 2025, 23:59 (via Microsoft Form)
Submission Link:https://forms.office.com/r/j0BH7VuBVk?origin=lprLink

Exploratory Data Visualization Dashboard :
Develop an interactive dashboard using **Python** and **Streamlit** (an open-source Python framework) to explore and visualize a dataset to create meaningful, insightful, and well-designed visualizations.

1. Dataset Selection

- Choose a dataset containing **at least 2,000 rows and minimum of 7 columns,** including **multiple dimensions** and **varied data types** (e.g., numerical, categorical) that allows for a variety of visual analyses.
- You may use sources such as Kaggle, Data.gov, World Bank Open Data, or Google Dataset Search.

2. Data Preprocessing

- Clean and preprocess your dataset as needed. Handle missing values, detect outliers, and address any other data quality issues.

3. Dashboard Design and Functionality

- The dashboard must include **at least 9 distinct visualization methods (3 per team member)**, with **at least 6 advanced types** (e.g., treemap, parallel coordinates, sankey diagram, network diagram, geographical maps, etc.).
- Each visualization must present a **unique insight or analytical perspective** from the dataset.
- Each visualization must show examples of interactivity, including mouse hover effects, color highlighting, dragging, zooming, filtering, panning, tooltips, glyphs, etc.
- Include at least **three interactive components,** such as dropdown menus, sliders, or checkboxes, to ensure the layout is **well-organized**, **aesthetically clean**, and **user-friendly**.
- Label all charts appropriately (titles, axes, legends) and maintain visual consistency across the dashboard.
- You may also try including basic machine learning algorithms (e.g., clustering, regression, or classification) to enhance your analysis and provide deeper insights.

4. Visualization Techniques

- Create a total of 9 meaningful and relatively advanced visualizations using Python and relevant libraries (e.g., matplotlib, seaborn, plotly, altair).
- Use appropriate color schemes, legends, and annotations to enhance interpretability.
- Focus on communicating clear insights rather than just visual variety.

5. Code Submission

- Upload your full project to a **public GitHub repository** and make sure your GitHub repository is **accessible** and **well-documented**. Include the following files: app.py – your main Streamlit application, your data, and README.md – project description, dataset details, setup instructions, and dataset file or link to dataset source. You must clearly indicate each member's contribution in the report or README file. Briefly describe who did what, for example, who handled data preprocessing, visualization design, interactivity implementation, creating each chart, or report writing, etc.

6. Presentation

- You are required to present your visualization dashboard on the allocated date and time for your team. The presentation should clearly highlight your key findings, visual design choices, and any technical challenges you addressed. The presentation must be delivered in English, and each team member will be assessed individually. Failure to present your work will result in your project being graded out of a maximum of 50 points.

7. Documentation

- Prepare a **one-page** report briefly summarizing the dataset and its context or source, the main goals of your analysis, the visualization techniques used, and key insights and findings, and including the GitHub repository link. The report should be clear, concise, and written in English. Late submissions will incur a 5% penalty per day for the first 3 days, and a 10% penalty per day thereafter (for each team member). Submissions or presentation requests will not be accepted after one week of the deadline. Best of luck!