

IE6600 Computation and Visualization for Analytics
Spring 2026
Project 1

Project Assignment:

Data Analysis and Visualization with a Focus on Static Visualizations and Statistical Analysis

Objective:

This project aims to develop students' static visualization skills and apply probability and statistical analysis to real-world data. Groups will select unique datasets from **data.gov**, each focusing on a different sector, such as health, environment, finance, or transportation. The primary objectives are to use the matplotlib library for static visualization and to apply probability and statistics skills for in-depth analysis of the selected dataset. Groups are encouraged to use additional datasets for a more comprehensive analysis.

Preparation:

- Choose a dataset from **data.gov** and get it confirmed by the teaching assistant to ensure no overlap with other groups.
- Review the basics of matplotlib for static visualizations and revisit critical concepts in probability and statistics relevant to data analysis.

Tasks:

1. **Dataset Selection and Confirmation:**
 - Select a dataset from **data.gov**.
 - Obtain approval from the teaching assistant to ensure your dataset is unique among the groups.
2. **Data Acquisition and Inspection:**
 - Download and load the dataset into your analysis environment.
 - Conduct an initial inspection to understand the dataset's structure, data types, and documentation.
3. **Data Cleaning and Preparation:**
 - Address missing data, duplicates, and inconsistencies.
 - Convert and normalize data types as necessary.
 - Encode categorical variables if present.
4. **Exploratory Data Analysis (EDA) with a Focus on Static Visualization:**
 - Use matplotlib to create static visualizations that reveal patterns, trends, and relationships in the data.
 - Implement statistical methods to analyze the dataset, drawing on your knowledge of probability and statistics.
5. **Advanced Analysis (Optional):**
 - Apply advanced analytical methods, potentially using additional datasets for deeper insights.
 - Use statistical models or machine learning techniques as appropriate.

6. Report:

- Document your process and findings in a detailed report, emphasizing your use of matplotlib and statistical analysis.

Deliverables:

- A Jupyter Notebook with code, data cleaning steps, EDA, and advanced analysis.
- A detailed project report in PDF format.

Grading Criteria:

- Innovation in dataset selection.
- Effectiveness in using matplotlib for static visualizations.
- Application and understanding of probability and statistical methods in analysis.
- Clarity and depth of the project report.
- Complexity and insightfulness of advanced analysis (if attempted).

Due Date:

- February 22, 2026

If you have any questions or need further clarification, please contact me. Best of luck with your insightful analysis and exploration!