IE6400 Foundations of Data Analytics Engineering
Fall Semester 2025

Group Projects

Project 1:

Topic: Cleaning and Analyzing Crime Data

Objective:
In this project, you'll work with a real-world dataset containing crime data from 2020. Your goal is to clean and prepare the dataset for analysis, perform exploratory data analysis (EDA), and answer specific questions related to crime trends, patterns, and factors influencing crime rates.

Dataset:
You will use the crime dataset available at Crime Data from 2020 to Present.

Tasks:

1. **Data Acquisition:** Download the dataset from the provided link and load it into your preferred data analysis tool
2. **Data Inspection:**
    - Display the first few rows of the dataset.
    - Check the data types of each column.
    - Review column names and descriptions, if available.
3. **Data Cleaning:**
    - Identify and handle missing data appropriately.
    - Check for and remove duplicate rows.
    - Convert data types if needed (e.g., dates to date format, numerical values to appropriate numeric types).
    - Deal with outliers if relevant to your analysis.
    - Standardize or normalize numerical data as necessary.
    - Encode categorical data if present.
4. **Exploratory Data Analysis (EDA):**
    - Visualize overall crime trends from 2020 to the present year.
    - Analyze and visualize seasonal patterns in crime data.
    - Identify the most common type of crime and its trends over time.
    - Investigate if there are any notable differences in crime rates between regions or cities.
    - Explore correlations between economic factors (if available) and crime rates.
    - Analyze the relationship between the day of the week and the frequency of certain types of crimes.
    - Investigate any impact of significant events or policy changes on crime rates.

5. **Advanced Analysis (Optional):**
    - Use predictive modeling techniques (e.g., time series forecasting) to predict future crime trends.
    - Explore additional questions or hypotheses related to the dataset.
6. **Report:**
    - Prepare a report summarizing your data cleaning and analysis process.
    - Include visualizations, tables, and descriptive statistics to support your findings.
    - Present your findings to the class, highlighting key insights and any interesting patterns or trends you discover.

Guidelines for Analysis

Note: The following key questions are guidelines to direct your analysis. If the dataset does not provide sufficient information to answer a question, acknowledge this in your analysis and explain why. It is okay if not all questions can be answered due to data limitations.

1. **Overall Crime Trends:**
    - Calculate and plot the total number of crimes per year to visualize the trends.
2. **Seasonal Patterns:**
    - Group the data by month and analyze the average number of monthly crimes over the years.
3. **Most Common Crime Type:**
    - Count the occurrences of each crime type and identify the one with the highest frequency.
4. **Regional Differences:**
    - Group the data by region or city and compare crime rates using descriptive statistics or visualizations.
5. **Correlation with Economic Factors:**
    - Collect economic data for the same time frame and use statistical methods, such as correlation analysis, to assess the relationship between economic factors and crime rates.
6. **Day of the Week Analysis:**
    - Group the data by day of the week and analyze crime frequencies for each day.
7. **Impact of Major Events:**
    - Identify significant events or policy changes during the dataset period and analyze crime rate changes before and after these events.
8. **Outliers and Anomalies:**
    - Use statistical methods or data visualization techniques to identify dataset outliers and investigate unusual patterns.
9. **Demographic Factors:**
    - Analyze the dataset to identify any patterns or correlations between demographic factors (e.g., age, gender) and specific types of crimes.
10. **Predicting Future Trends:**
    - Employ time series forecasting methods, such as ARIMA or Prophet, to predict future crime trends based on historical data. Consider incorporating relevant external factors into your models.

Students would need to perform data cleaning, exploration, visualization, and statistical analysis as necessary for each of these tasks.

Deliverables:

- A Jupyter Notebook containing your code, data cleaning steps, exploratory data analysis, and any additional analysis.
- A project report in PDF format documenting your process and findings.

Grading Criteria:

Your project will be assessed based on the following criteria:

- Completeness and correctness of data cleaning.
- Thoroughness and relevance of exploratory data analysis.
- Clarity and organization of the project report.
- Creativity and depth of analysis (for advanced analysis, if attempted).

Important Dates:

- Project Submission Deadline: **October 15, 2025**

Please reach out if you have any questions or need clarification on the tasks. Good luck with your project!