**Data Visualization Track Requirements (75 points)**

**Data and Delivery (20 points)**

The dataset contains at least 100 unique records. (5 points)

**Clarification on Datasets Used in the Project**

**Starting Datasets**

1. **Crime Data**
   1. **Source:** <https://catalog.data.gov/dataset/crime-incidents-in-2020>
   2. **Description:** A dataset containing detailed records of crimes committed in Washington, DC, during 2020, categorized by offense type, census tract, and geographic location.
   3. **Initial Number of Rows:** 27,932
2. **Census Data**
   1. **Source:** <https://www.census.gov/data/tables/2020/demo/educational-attainment/cps-detailed-tables.html>
   2. **Description:** A dataset with demographic and educational attainment data for Washington, DC, organized by census tract. It includes key statistics such as population by race, educational levels, and total population figures.
   3. **Initial Number of Rows:** 3,915

After cleaning, normalizing, and merging the starting datasets, we generated three final refined datasets, each focused on different population segments:

1. **Male Population Dataset**
   1. **Description:** Contains detailed educational attainment and demographic data for the male population in Washington, DC. This dataset was normalized and aligned with the corresponding census tracts for analysis.
   2. **Number of Rows:** 207
2. **Female Population Dataset**
   1. **Description:** Includes the same educational and demographic data as the male dataset but specific to the female population. This was also normalized and aligned with the census tracts for consistency.
   2. **Number of Rows:** 207
3. **Total Population Dataset**
   1. **Description:** Represents the aggregate of male and female datasets, providing a comprehensive view of the entire population’s demographics and educational attainment in each census tract.
   2. **Number of Rows:** 207

File Path: Project 3 -> Data -> Merged -> Complete -> Merged\_data.csv

* A database is used to house the data (SQL, MongoDB, SQLite, etc.). (5 points)
  + Project 3 -> Data -> Merged -> Merging Code -> data\_frame\_merging.sql
* The GitHub repo has a README.md that includes the following: (10 points)
  + An overview of the project and its purpose
    - README.md
  + Instructions on how to use and interact with the project
    - Project 3 -> Flask\_app -> interactive map application instructions
  + At least one paragraph summarizing efforts for ethical considerations made in the project
    - README.md
    - Project 3 -> Project Description
  + References for the data source(s)
    - References found in code where needed
  + References for any code used that is not your own

**Visualizations (25 points)**

* A minimum of three unique views present the data. (10 points)
* **Visualizations Overview**
* Our project includes a total of four visualizations, demonstrating our effort to exceed the requirements and provide comprehensive insights.

1. **Interactive Map**:

• Built using JavaScript and integrated with Leaflet, this dynamic map allows users to explore data at the census tract level.

• Users can select from three data options: **Crime Data**, **Educational Attainment**, and **Census Data**. Each selection updates the map with relevant information displayed in a detailed popup for the selected tract, offering an engaging and user-friendly interface.

2. **Matplotlib Visualizations**:

• We created three separate plots using Matplotlib to analyze the relationship between educational attainment levels and crime rates:

• **No Educational Attainment vs. Crime Rate**: Highlights the impact of areas with low educational attainment on crime rates.

• **High School Educational Attainment vs. Crime Rate**: Examines how attaining a high school education correlates with crime levels.

• **College Educational Attainment vs. Crime Rate**: Investigates the connection between college-level education and crime.

* The visualizations are presented in a clear, digestible manner. (5 points)
* The data story is easy to interpret for users of all levels. (10 points)

**Usability (30 points)**

* The script, notebook, or webpage created to showcase data visualizations runs without error. (10 points)
* A Python or JavaScript library not shown in class is used in the project. (10 points)
* The project includes some level of user-driven interaction, conforming to one of the following designs: (10 points)
* ^^^ Project 3 -> Flask\_app -> app.py
* ^^^Project3 -> Flask\_app -> static -> js -> tracts\_map.js
  + HTML menus, dropdowns, and/or textboxes to display JavaScript-powered visualizations
  + Flask backend with interactive API routes that serve back Python or JavaScript created plots
  + Visualizations created from user-selected filtered data

**Data Engineering Track Requirements (75 points)**

**Database Design (40 points)**

* The project uses ETL workflows to ingest data into the database. (10 points)
  + Project\_3 -> Data -> Cleaning -> census\_data\_cleaning.ipynb
  + Project\_3 -> Data -> Cleaning -> crime\_data\_cleaned.ipynb
* The original dataset(s) are transformed prior to storing it in the database. (5 points)
  + Project\_3 -> Data -> Cleaning/Mapping/Grouped Data/Merged
* A database is used to house the data (SQL, MongoDB, SQLite, etc.). (5 points)
  + Project\_3 -> Data -> Cleaning/Mapping/Grouped Data/Merged
* The database has at least two tables (SQL) or collections (NoSQL). (5 points)
  + Project\_3 -> Data -> Cleaning/Mapping/Grouped Data/Merged
* The project documents the choice of the database used and why. (5 points)
  + We used an SQL database to merge and store the datasets for crime, census, and educational attainment data. SQL was chosen because it is efficient for managing structured data, joining datasets, and retrieving information dynamically for the app.
* The project includes documentation of the ETL workflow with diagrams or ERD. (10 points)
  + Refer to the structured folders in repository

**Data and Delivery (35 points)**

* The database contains at least 100 unique records. (5 points)
  + Project 3 -> Data -> Merged -> Complete -> Merged\_data.csv
* The project uses one additional library not covered in class related to data engineering. (10 points)
  + Project\_3 -> flask\_app -> app.py
* The project includes a method for reading data from the database and displaying it for future use, such as: (10 points)
  + Pandas DataFrame
    - Project\_3 -> Correlation -> correlation\_analysis.ipynb
  + Flask API with JSON output
    - Project\_3 -> flask\_app -> app.py
* The GitHub repo has a README.md that includes the following: (10 points)
  + An overview of the project and its purpose
  + Instructions on how to use and interact with the project
  + At least one paragraph summarizing efforts for ethical considerations made in the project
  + References for the data source(s)
  + References for any code used that is not your own

**Both Track Requirements**

**Group Presentation (25 points)**

* All group members speak during the presentation. (5 points)
* The content is relevant to the project. (5 points)
* The presentation maintains audience interest. (5 points)
* Content, transitions, and conclusions flow smoothly within any time restrictions. (10 points)