roject 3

Start Assignment

* **Due** Mar 4 by 11:59pm

* **Points** 100

* **Submitting** a text entry box or a website url

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

**Data and Delivery (20 points)**

* The dataset contains at least 100 unique records. (5 points)
* A database is used to house the data (SQL, MongoDB, SQLite, etc.). (5 points)
* 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

**Visualizations (25 points)**

* A minimum of three unique views present the data. (10 points)
* 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)
  + 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)
* The original dataset(s) are transformed prior to storing it in the database. (5 points)
* A database is used to house the data (SQL, MongoDB, SQLite, etc.). (5 points)
* The database has at least two tables (SQL) or collections (NoSQL). (5 points)
* The project documents the choice of the database used and why. (5 points)
* The project includes documentation of the ETL workflow with diagrams or ERD. (10 points)

**Data and Delivery (35 points)**

* The database contains at least 100 unique records. (5 points)
* The project uses one additional library not covered in class related to data engineering. (10 points)
* The project includes a method for reading data from the database and displaying it for future use, such as: (10 points)
  + Pandas DataFrame
  + Flask API with JSON output
* 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)

**Requirements for Projects that Began Before 12/15/2023**

**Data and Delivery (25 points)**

* Data components used in the project are clearly documented. (5 points)
* The dataset contains at least 100 unique records. (5 points)
* A database is used to house the data (SQL, MongoDB, SQLite, etc.). (5 points)
* The project is powered by a Python Flask API and includes HTML/CSS, JavaScript, and the chosen database. (10 points)

**Back End (25 points)**

* The page created to showcase data visualizations runs without error. (7.5 points)
* A JavaScript library not shown in class is used in the project. (7.5 points)
* The project conforms to one of the following designs: (10 points)
  + A Leaflet or Plotly chart built from data gathered through web scraping
  + A dashboard page with multiple charts that all reference the same data

**Visualizations (25 points)**

* A minimum of three unique views present the data. (5 points)
* Multiple user-driven interactions (such as dropdowns, filters, or a zoom feature) are included on the final page. (5 points)
* The final page displays visualizations in a clear, digestible manner. (5 points)
* The data story is easy to interpret for users of all levels. (10 points)

**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)