Assignment 1 Outline

Dataset

We have chosen a dataset of United States crime data broken down by geographical area (county), population, and race data. The United States has one of the highest violent crime rates in the world and as a result, there are many places in the United States that have extremely high crime rates. Many existing crime maps do not give enough context and as a result, are often used as tools to promote fear. Being able to show this data with proper context (for example, median income, graduation rates, etc) can help show what problems need to be solved to reduce crime within these areas.

Description of the tech stack

We will be using several tools available within datasette and third parties. The most important tool we are using is the map. We had multiple options for our map; Map Box, OpenStreetMap, and Google Maps. MapBox is created by a small, private company. It has a free tier that we would have no problem using for the usage within this program. However, as this is a very small company, there is a risk of them changing their monetization model due to the current financial situation similar to what Heroku did with their free tier. Google Maps also has a free-tier allowing us to use $200 of usage per month which equates to 28,500 map loads. For google maps, we risk surpassing our monthly allotment leading to a surprise bill (such as the ones common with Amazon Web Services). Lastly, we considered open street map, an open source project with no required costs. As it is an open source project with no financial risk to us, we decided to use the open street map service for our frontend map.

For our data processing, importing, we will be using the python programming language to make manipulations. There are multiple programming languages we can use for this purpose including C, C++, Javascript, etc. We decided on using python as it is a relatively easy language to use and we all have experience with it.

To store our data, we will be using SQLite as this is a requirement for the datasette tools that we want to use.

Lastly, we want to create a frontend dashboard in the form of a website to showcase data to the end-user. There are several ways to create a website, HTML/CSS, Flutter, and React are the most common ways in the industry. All of us have some experience in using HTML/CSS and have found it to be difficult, creating many unnecessary steps compared to the alternatives. Flutter is an excellent alternative to React, allowing us not only to create a website, but mobile apps as well. However, since Flutter is not used as much as react, and the fact that it was made by google (leading to the risk of it getting killed off) led us to choose React.

Product Development Plan

Milestone 1
Milestone 1 is going to be mainly focused on our backend. We will want to have the following tasks completed.

- Importing data into our SQLite database
- Processing data for our database
- Connect our database to our datasette libraries that we are using
- Create tests for each function
- Create a react website with a dashboard
    - Just a plain template without any data on it

Milestone 2
Milestone 2 is mainly focused on creating a frontend with basic functionality and connecting that to our backend. Tasks that should be completed are

- Connect our backend to our front end
- Support different types of requests
- Have basic functionality on the frontend to display our data

Milestone 3
Milestone 3 will be our final milestone where we will complete our project. The tasks that will need to be completed in this milestone are

- Have different visualizations and have our data respond accordingly
- Have users be able to import similarly formatted datasets
- If time permits (stretch goals)
    - Import data table by URL
    - Format data from UI
    - View data stats from the UI
        - Similar to np.describe
    - Live datastream
    - Crime vs population visualizations
- Prepare our final presentation

Toy Application

For Assignment 1, we created a toy application to show the key components of the tech stack in action. Specifically, the user of this toy application can import a dataset of any kind to an SQLite database and launch it with Datasette to view in their web browser. The modifiable script currently works with the already included US Crime dataset that we cleaned and added to the repository, which has location coordinates for each data point so that we can use a Datasette Map Plugin to visualize the data. The toy application also has a test which checks to see if the SQLite database has the correct number of rows for the US Crime dataset.