

Design Document

a general overview:

A general overview of your system with a small user guide: We use the data from [City of Edmonton's Open Data initiative](#) and get some criminal information. Our program is to implement SQLite search in a host programming language, python. This program can allow users to enter the number of function to finish certain search as they want. We import pandas, sqlite3, Folium and matplotlib to help us analyze data easier.

a detailed design:

At the beginning, the program will required user to type in the database name (includes the .db Tail segment), then the program will work on this specific database.

In the initial interface, there are five choices: first four choices are functions from 1 to 4 and the last choice is exit. Unless you choose to exit, it will always go back to the initial interface after finishing each function.

Function1:

User input the start year, end year and crime type. Our program will give the user the month-wise total count of the given crime type(in a bar plot). We use left outer join to make sure when some months have no criminal events, the plot will show count zero. Otherwise, there would be no this month on the plot. For the sample database, there is a row have the same value of field name, the typeof(month1) = "integer" is to avoid this problem.

Function2:

User input a number N. the program will give the user N-most populous and N-least populous neighborhoods on a map. We first select the population and location and rank them in descending order and take the first N rows as rows. Then rank them in ascending order and take the first N rows as rows2. After that, we find all other rows have the same population of Nth row in rows or rows2. Finally, draw all of the circle on the map.

Function3:

User input a range of years includes start year, end year, a crime type, and an integer N. It will show (in a map) the Top-N neighborhoods and their crime count where the given crime type occurred most within the given range.

Function4:

User input a range of years and an integer N, it will show (in a map) the Top-N neighborhoods with the highest crimes to population ratio within the provided range. Also, it shows the most frequent crime type in each of these neighborhoods.

At first, we use a SQL to find the radio and the top neighbors then use a subquery to find the most frequent crime type and if there is more than one crime type we will select all of them. Then draw the circle in the map.

Testing strategy

Function1: Type in different range of years and crime type to check if the month from 1 to 12 is shown.

Function2: first test the function without tie in rank. The function runs normally. Then find a tie in 34th and 35th of data, try enter 34 as input, the map also show the 35th row's location.

Function3: Type in different range and number to test the function. The function runs normally. Then I find a tie when the input is 2009,2018,Homicide,4. There will be several ties. The map show all 6 points correctly.

Function4: do the normal test as describe above. No ties for the ratio in the sampled database. But the way to implement tie is the same of function2 and function3.

group work break-down strategy:

Chenge Liu and Tianyu Lang are mainly responsible for functions(1-2) and function(3-4) respectively.

Wang is mainly responsible for testing and rewriting some queries.

Each group member works about 5 hours.

In order to keep the project on track, we have a meeting first and share ideas. Then, we finish our own part. Finally, Wang helps us to test and solve some problems.