Final Project Phase I

Q1. Topic - 15 points

Please provide an overview of what your topic is going to be.

Q1.1 - 5 points

What topic have you chosen for your Final Project?

Answer: The topic I chose – Crime in Chicago

Q1.2 - 5 points

Why did you choose this specific topic and what are you looking to learn from the analysis?

Answer: I was particularly interested in different kinds of crimes and crime rates in the US. However, after doing some research to find data, I realized I should focus on a state and that would be easier to do some analysis in the future. I found some articles and different datasets for crime in Chicago. And this made me go with this topic.

Q1.3 - 5 points

Explain some of the concrete insights you expect to gather from your data and/or hypothesis you expect to answer.

Answer:

The insights I expect to gather from my data are:

- Overall idea of crime situation in Chicago compared to other states.
- Data about the type of crimes that are frequent there.
- Dangerous area in Chicago based on crime frequency.

Q2. Downloaded Dataset - 15 points

Please provided a brief overview of you downloaded dataset. This should demonstrated that you understand the data contained within the dataset.

Q2.1 - 2 points

Provide the link (url) to your downloaded dataset.

Answer:

https://data.cityofchicago.org/Public-Safety/Crimes-2022/9hwr-2zxp/data

Q2.2 - 3 points

What are the dimensions of your downloaded dataset in terms of rows x columns and file size? Ex. 50,000 rows x 20 columns and 5.4mb. If your file is a .json file, state the file size (mb, gb, etc.).

Answer:

`238985rows` x 22 and 56.9MB

Q2.3 - 5 points

Briefly discuss the structure of your dataset. For .csv or table type datasets list out the column titles and give examples of the data contained within. For json data map out the dictionary and give examples of the data contained within.

Answer:

Data: Crimes2022.csv

Column titles: ID, Case Number, Date, Block, IUCR, Primary Type, Description, Location Description Arrest, Domestic, Beat, District, Ward, Count, Community Area, FBI code, X coordinate, Y coordinate, Year, Updated on, Latitude, Longitude, Location

Q2.4 - 5 points

Please explain why you chose this specific dataset. How will this data be used in your analysis? Can insights be drawn from this data alone, or will it be combined with other data?

Answer: This CSV file contains crime incident cases reported during 2022 in Chicago. Even though the data is for one year, the file contains significant information like different types of crimes reported that year and places where they happened. This will help me get data of crime type, their frequency in different places in Chicago.

Q3. Web Requirement #1 (Web-scrape or HTML) - 15 points

Please provided a brief overview of you downloaded dataset. This should demonstrated that you understand the data contained within the dataset.

Q3.1 - 2 points

Provide the link (url) to your downloaded dataset.

https://www.way.com/blog/chicagos-crime-rates/#:~:text=Statistics%20say%20that%20the%20crime,1%2C985%20crimes%20per%20100%2C000%20people.

Answer:

Q3.2 - 3 points

Explain briefly how you plan to retrieve the data from this source.

Answer: The document has significant statistics about Chicago crime and Chicago crime-prone places which I will focus on extracting for my analysis.

Q3.3 - 5 points

Briefly discuss the structure of your dataset. For .csv or table type datasets list out the column titles and give examples of the data contained within. For json data map out the dictionary and give examples of the data contained within.

Answer:

This webpage contains valuable information about the overall crime situation in Chicago.

This webpage contains three tables with data of :different violent crime rates, dangerous places, and safest places in Chicago.

Q3.4 - 5 points

Please explain why you chose this specific dataset. How will this data be used in your analysis? Can insights be drawn from this data alone, or will it be combined with other data?

Answer:

I came across this document while researching about overall crime situation in Chicago. The data presented in this document are precise and I can verify my analysis from my other dataset (crimes2022, fatal&non-fatalShootingCases).

Q4. Web Requirement #2 (API or JSON) - 15 points

Please provided a brief overview of you downloaded dataset. This should demonstrated that you understand the data contained within the dataset.

Q4.1 - 2 points

Provide the link (url) to your downloaded dataset.

Answer:

https://catalog.data.gov/dataset/violence-reduction-victims-of-homicides-and-non-fatal-shootings

Q4.2 - 3 points

Explain briefly how you plan to retrieve the data from this source.

The json file presents data about individual-level homicide and non-fatal shooting cases from 1991 to present. However, the specific non-fatal cases stored there are from 2010 to the present. Therefore, I was planning on retrieving data from 2010 to present from this source to get unbiased analysis.

Answer:

Q4.3 - 5 points

Briefly discuss the structure of your dataset. For .csv or table type datasets list out the column titles and give examples of the data contained within. For json data map out the dictionary and give examples of the data contained within.

Answer: The structure of data set is like a table where each row represents different homicide or nonfatal shooting cases. Some key features of each cases are: case, crime place, timeline, shooting injury, information about victim (sex, age),

Q4.4 - 5 points

Please explain why you chose this specific dataset. How will this data be used in your analysis? Can insights be drawn from this data alone, or will it be combined with other data?

Answer: This source provide valuable data on specific crime in Chicago for a longer timeline. I can combine the data from my analysis of my CSV dataset about Chicago crime with this source.

Q5. Additional Datasets - 10 points (N/A)

Q6. Inconsistencies - 15 points

Please list at least 3 inconsistencies you have found in your datasets, and how you plan to address each of them.

Answer:

1) For My CSV dataset, there are several unknown and empty data present. If I plan on using those columns, I might want to skip those rows with unknown data.

2)For My Json dataset, there are also several unknown and empty data present. If I plan on using those data sections, I might want to skip those data with unknown/empty data. There are also some inconsistencies for some data (like FBI_code is sometimes string, sometimes number) for different cases, but I don't think I will utilize them for my analysis.

3) For json data, the non-fatal shooting cases are not stored prior to 2010 (even though entire timeline for all data is from 1991 to present). In order to address missing data, I would focus on data from 2010 to present where all homicide and not fatal shoring cases data are available to get unbiased analysis.

Q7. About Your Analysis - 10 points

Provide a BRIEF list of steps of how you plan on performing your analysis and the way you will gather/present your findings. (Non-technical, high-level overview)

Answer:

A high-level approach I am planning:

- I will retrieve different crime types and their frequency from my CSV and JSON sources.
- I will also retrieve the crime places in Chicago where the crime cases were reported.
- Once I have crime types and places, I will compare them with the statistics presented in my webpage source by the Chicago police department.
- Since my CSV, and JSON source some other additional information like victims, criminals, time, etc, I can extend my analysis to victims' age range, criminals' age range, and frequency of crime in different years.

Q8. About You - 5 points

Q8.1 - 2.5 points

List the names of each of the members of the group working on this project. If you are working alone, there should be one name listed.

Team Member 1: Maisa Basher

Team Member 2(If Applicable):

Q8.2 - 2.5 points

Each member of the group should initial below to indicate that you acknowledge this statement:

I affirm that all of the work in this project will be done my me/my team and is not duplicated from any other source. In addition, any references that I use or code that I choose to model after will be appropriately credited and referenced in my project.

Team Member 1 Initials: MB

Team Member 2 Initials (If Applicable):

Total - 100 points