CSE583 - Big Data & Policing

HW 2

Deadline - February 12, 2019
Total Marks - 140

Instructions

1. Preferably use Python3 for this assignment. You can use applications such as QGIS to generate heatmaps.
2. All datasets pertaining to this HW are available in this Google Drive folder
3. The final submission must consist of all your source code and a report containing all your findings.
4. In case of any doubts/issues, please post on the Moodle forum.
5. Do not copy. This could result in an 'F' grade in the course.

Part 1 (70)

- Use the `baltimore_crimes.csv` dataset to answer the following questions -
    1. Plot frequency of crime incidents across all districts. Which district has the most number of incidents? Which has the least? (10)
    2. Generate a heatmap of the crimes committed. Are the crimes spread evenly across the city, i.e., do incidents occur with same frequency all across the city? (Try out matplotlib, gmplot, mapsplotlib, and qgis). (10 marks)
    3. Generate two plots - Frequency of crime vs day of the week, and frequency of crime vs time of day. Using these, answer the following:- (20)
        - On which day of the week does most crime occur?
        - Which crime is most likely to occur between 2000 Hrs and 2359 Hrs?
    4. Using a frequency vs crime plot, extrapolate what crime is the most common? Using another plot - most_frequent_crime vs time - find approximately what time of the day does it occur most often? Plot a heatmap of the most popular crime. Is it evenly spread out across the city? (30)

Part 2 (70)

- Use the `global_terrorism_database.xlsx` dataset. Go through the dataset carefully, and answer the following -
    1. Plot a histogram of frequency of terrorist attacks in North America vs decade. Using this, find out in which decade did North America experience the most number of terrorist attacks? (10)
    2. Plot the frequency of terrorist attacks from 1970 to 2017. Which year had the highest number of terrorist attacks? In that year, using another plot, find which continent experienced the most number of attacks? (20)
    3. Generate word clouds (after removing dates and stop words) for the summary and motive for all the terrorist attacks. Explain the most popular reasons and descriptions for attacks using these two word clouds. (20)
    4. Construct an undirected graph in which two terrorist attacks have an edge between them if they are related. Using this graph, find the total number of coordinated terror attacks, i.e., the total number of strongly connected components. Also, find the largest set of coordinated terror attacks recorded. (20)