CSE583 - Big Data & Policing
HW 2 - Answer Key

Part 1

1. Plot frequency of crime incidents across all districts. Which district has the most number of incidents? Which has the least? (10)
   a. 5 marks for a plot, 5 marks for below inference
   b. South Eastern district has most number of incidents and North Eastern district has least incidents.

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)
   a. 5 marks for heatmap, 5 marks for below inference
   b. incidents don't occur with same frequency across the city. There are two main concentrated spots.

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)
   a. – On which day of the week does most crime occur?
      i. 5 marks for plot, 5 marks for Tuesday
   b. – Which crime is most likely to occur between 2000 Hrs and 2359 Hrs?
      i. 5 marks for plot, 5 marks for larceny from auto

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)
   a. 5 marks for each plot = 15 marks
   b. 5 marks for answering larceny is the most common crime
   c. 5 marks for answering it occurs most commonly around 1600 Hrs
   d. 5 marks for stating it isn't evenly spread out

Part 2

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)
   a. 5 for plot, 5 for 1970-1980

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)
   a. 2 plots = 10 marks
   b. 2014 has highest # of terror attacks, Middle East and North Africa most number of attacks => 10 marks

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)
    a. 2 word clouds = 10 marks
    b. 10 marks for any valid explanation
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)
    a. Total number of coordinated attacks are: 7335
    b. The largest set of coordinated terror attacks recorded: 80