**CS 5200 Homework 6**

Creating and manipulating a MySQL schema for crime data.

Many times, you are not given a description of the conceptual design but are provided a collection of data. Your task is to create a MySQL database given a terse textual description of the data along with a data dump.

**Domain Description**

The attached files contain information on the crimes reported in the city of Boston for the week of December 22nd to December 29th for the year 2022. The crime incidents are an excerpt from the [2022 City of Boston Crime Incident Report](https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system).

**Data File Description**

There are four different csv files: incident.csv, districts.csv, offense\_codes.csv and neighborhoods.csv. The incidents.csv contains a row for each crime committed in Boston for the specific time interval.. The file district.csv contains a row for each district in the city of Boston. The crimes.csv file contains the criminal codes for the potential crimes that can occur.

The **districts.csv** file contains 2 fields:

a district code : short code for all districts in the city of Boston

A district name - name of the district

You should import 13 tuples from this file.

The **neigborhoods.csv** file contains 2 fields:

a district code : short code for the district, found in the districts.csv file

a neighborhood name: a textual description of a neighborhood

There are multiple neighborhood names for 1 district code but only 1 district for a neighborhood.

You should import 22 tuples from this file.

The **offense\_codes.csv** file also contains 2 fields:

a crime code: a short code to represent a type of crime

a name: a longer textual description for a crime.

You should import 425 tuples from this file.

Each row within the **incident.csv** file represents a specific crime that was committed in Boston.

The data fields in the incidents.csv file are the following:

Incident number: Unique number to represent a crime incident

Offense Code: the code for the crime - values found in the offsense\_codes.csv file

District: The district where the crime was committed - legal values can be found in the districts.csv file

Reporting area: optional data, providing a more precise location in the district

Shooting: 1 if the incident involved a shooting, 0 if no shooting was involved during the crime incident

Year: year value for the crime, all are assigned 2022

Month: month value for the crime all are assigned 12

Day of week: day that the crime occurred, Values such as, Sunday, Monday, Tuesday, etc.

Hour: hour of the day the crime occurred , Values such as 0.. 23

Street: Street where the crime occurred, free formed text

Latitude: latitude of the crime

Longitude: longitude of the crime

Occurred\_on\_Date: the date the crime occurred (In YYYY-MM-DD HH:MM) format

(Note: In excel, if you open the csv, you might see a different format, but do not worry. After importing in MySQL you will see the above format)

You should import 992 tuples from this file.

A word of caution, the import panel sometimes **appears behind the main Workbench GUI**, so you may need to manipulate your windows to complete the Import process.

**Query creation:** When creating the SQL SELECT statements, make sure you do not design queries that are dependent on the data stored in the database or that only work for this specific instance of the schema. **Also, please provide the question number in comments before the solution.**

**Assignment Description**

1. Create a database for your schema named crime\_db2022*lastnamefirstnameinitial*. Create the database and the tables using the SQL CREATE command. (10 points)
2. Once you have created the database and the necessary tables, check the design by reverse engineering the crime database. Create a pdf or an image of the EER model. (5 points)
3. Import the .csv file into your tables using the Import tahble wizard or create INSERT commands to place the data in the tables. Make sure all rows are present in the database. (5 points)

Compose queries to answer the following questions:

1. For each calendar day in the database, generate the number of crime incidents that occurred on that day. The result should contain the calendar day and the count of the crimes. Rename the count of the crimes to num\_crimes. Return the results ordered by the incident date in ascending order. (5 points)
2. Which street had the most number of crime incidents? Return the street name and the number of incidents. (5 points)
3. What is the maximum number of crimes that could have occurred in the North End during the specific time period? Return the number (5 points)
4. How many crimes occurred in Hyde Park? Return the number. (5 points)
5. Report on all rapes that occurred during the time period. Return the crime code, the incident date and the district. Order the results by date, then by district. (5 points)
6. Determine the number of times each crime code occurred during the time period. Rename the count num\_occurrences. The results should contain the crime code, the crime description and the count num\_occurrences. Order the results in descending order using num\_occurences. Each crime code must appear in the result. If a crime code did not occur, then its num\_occurrences should be 0. (5 points)
7. For each district with crimes, determine the number of crimes that occurred for that district. The result should contain the district code, the district name, and the number of crimes for the district. Rename the number of crimes num\_crimes. Orders the results in descending order using num\_crimes. (5 points)
8. For each crime code, determine the number of districts it occurred in. If the crime did not occur, then the number of districts should be reported as 0. (5 points)
9. Generate a list of crimes that occurred between Christmas and December 28th, inclusively. Include the incident number, the name of the district, the description of the crime offense and the date of the crime. Order the results in ascending order by date of the crime (5 points)
10. Generate the top crime for each district. The result should contain the name of the district, the description of the crime offense and the number of incidents. (5 points)
11. For each crime code committed, create an aggregated list of district names where the crime was committed as well as the number of times the crime occurred. The result consists of the crime description, the aggregated field of the district names and the count of the number of times the crime code occurred. Rename the aggregated district names to districts and rename the count of crime instances to num\_crimes. Order the results in descending order using the num\_crimes field. (5 points)
12. What is the number of crimes that occur per hour of the day, for the hours between 6PM to 11:59 PM ? The results should contain the hour of the day and the number of crimes for that hour. Order the results in ascending order by the hour of the crime.(5 points)
13. What is the number of crimes that occur per day of the week? The results should contain the day of the week and the number of crimes for that day. Order the results by the days of the week starting with Monday and ending with Sunday.(5 points)
14. What is the average number of crimes that occur over the calendar days (use incident\_date) ? You must first determine the number of crimes for each calendar day, then take an average of these counts. The results should contain an average number. (5 points)
15. Which crimes were never committed during the database’s time period? The results should contain the crime code and the crime description. (5 points)
16. Generate a self-contained extract of your database to canvas using the ‘data export’ tool from the ‘Server’ menu. Make sure you include the create schema as well as other objects in the database. We must be able to import your schema so please ensure the extracted file works with import. (5 points)

**Assignment submission**

Please submit the following 3 files :

* hwk6crimeLogical*lastnamefi*.pdf (picture of the model)
* hwk6crimedump*lastnamefi*.sql (export of the created schema)
* Hwk6crimequeries*lastnamefi*.sql (SQL SELECT commands for the queries)

Submit the multiple files to canvas.