Database System Fall 2017 Final Project

Liu, Wei
Wu, Yingyi

1. Project Description

This project builds an application which can access PostgreSQL and MongoDB, and allow user to query the databases. Language used to build up this project was Python 3.6.

2. Datasets:

Initially, two datasets are being imported to the databases.
They are both about the public safty in New York City.
The first one is *NYPD Complaint Data Historic.* User can access and export the dataset from the website of NYPD open data:
https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i

The second one is *NYPD Motor Vehicle Collisions.* User can access and export the dataset also from the website of NYPD open data:
https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

These two datasets should be exported as .csv files and be stored in the same directory with the python script.

3. Set up the databases:

Make sure you store the .csv files in the same directory as this script, and don't change the name of the .csv file, which means the names of the .csv files should be *"NYPD_Motor_Vehicle_Collisions.csv"* and *"NYPD_Complaint_Data_Historic.csv"*.

This script assumes the existence of a Postgres database running on localhost that's configured to allow remote connections (which should be the default), as well as the existence of the *postgres* database with a super user named *postgres* (default setting of Postgres database)

It also assumes the existence of a MongoDB running on localhost with the port number *27017* (supposed to be a default setting). MongoDB should be successfully connected ***before*** running the script.

To set up the databases, please run the script. We provided this function in our application.

1. It would ask you to enter the password of the user *postgres* in your local database. Make sure that you enter the correct password.
2. Please choose the function named *"initialize databases"*.

3. The whole process (create tables, load data, normalize relations) will take about 5 mins due to the scale of the datasets.
4. When the process finish, it will output *"loading complete!"*.

Before you start to set up, please read the fourth part about how to build up the application!

4. Application and how to use it

The packages this python 3 script need as follow:
os
json
psycopg2
pymongo
pandas
gmplot
psycopg2.extensions

On terminal, cd to the directory where the script in.
Run command *"python3 runner.py"* (command vary according to OS, make sure that the interpreter is python 3).

When the application start, follow the menu to explore the datadases.

Basic functions:

1. initialize databases:
Initialize the databases, you should run this function ***before*** you explore the databases if the databases has not been setted up yet.

2. Summary of the public safrty by borough:
Ask user to input a year (from 2012-2017). It will then display a table shows the total number of crimes classified by level of offense as well as the total number of motor collision of each borough in that year.

3. Effect of collision contributor:
This function will provide a table of collision contributors, and then ask user to choose one of them. The function will show the effect of the contributor in year 2012 - 2017.

4. Gap between time of occurrence of crime and time it was to police:

This function ask user to choose a level of offense, and then display the gap between time of occurrence of crime, which belongs to this offense level, and the time it was report to police.

5. Top 5 offenses in a specific time period:
This function asks user to input a time period, and then shows the top 5 most frequently happened types of crimes.

6. Heatmap of crimes:
This function ask user to choose one of the level of offense (same as the 4th function), and then display a table top 15 most frequently happened crimes, which belong to the level user chose. User can then choose the type of crimes in the table, and then this function will plot a heatmap of those crimes.
The heatmap will be stored as a html file named “heat_map.html” in the same directory as this script. User should open it manually for viewing.