Tianyi Lin
Yang Cao

Database Phase 1

1.Members: Tianyi Lin, Yang Cao
2.Target domain: crimes in Baltimore area (and other cities) from 2001 to present
3.

1) What is the crime rate for a specific neighborhood?
2) What areas have higher crime rate at night and what areas have higher crime rate at day?
3) What kind of crime (e.g. murder, rape, robbery, b & e) is the most frequent around certain area?
4) What areas have higher accident rate (e.g. bicycle crash)?
5) What kind of accident is the most frequent around certain area?
6) What areas have higher accident rate at night and what areas have higher accident rate at day?
7) What neighborhood have caught more gun offenders?
8) Does weather correlate with accident rate?
9) Does weather correlate with bicycle crash rate?
10) Does light condition correlate with bicycle crash rate?
11) Does road condition correlated with bicycle crash rate?
12) What is the rate of car accidents that are caused by driver exceeding speed limit?
13) For bicycle crashes, how many of the Drivers have Alcohol Detected?
14) For people who committed crimes, how many of them are also gun offenders?
15) What age range (10 years, for example) has highest number of gun offenders?

4.
Crime_data

accident_id	States	Crime_type	Time	Police_station	People_involved	First name of People commited to crime	Last name of People commited to crime
123	MD	Robbery	19	Baltimore	Stan Lee, Eddie Brook	Eddie	Brook

Gun_offenders

caseNumber	created_date	modified_date	lastName	firstName	Date_Of_Birth	sex	full_address	district	neighborhood	Police_station	casualties	race
123	11/15/1028	11/18/2018	Fake	Name	1/2/2000	M	Room 123, Street 456.	Bad County	Hopkins	Baltimore Police Station	1 injured	Black

Crime_mapping

accident_id	crime_category	district	map_reference	location_category	lat	lon	location
123	Robbery	Bad County	123	Highway	111.222	333.444	123 Street 345

Bike_crash

City	Crash Date	Crash Location	Crash Time	Crash Severity	accident_id	Ambulance Response	Light Condition	Number of Lanes	Road Characteristics/Class/Condition/Configuration	Road Defects/Features
baltiomre	1/1/2018	Highway	11:23 pm	2 injured	123123	Yes	OK	3	No Special Feature	None

Bike_crash_people

accident_id	Bike/Pedestrian Age Group	Bike/Pedestrian Sex	Driver Age Group	Driver Estimated Speed	Speed Limit	Driver Alcohol Detected	Driver Injury	Crash Type
123	20	M	30	27	20	No	No	Motorist Overtaking - Bicyclist Swerved

5.

SQL statments for 15 questinos

1) What is the crime rate for a specific neighborhood?
    - select neighborhood with the highest count of crime for a neighborhood
2) What areas have higher crime rate at night and what areas have higher crime rate at day?
    - night: select top 3 neighborhoods that have the highest count of crime at night
    - day: select top 3 neighborhoods that have the highest count of crime at day
3) What kind of crime (e.g. murder, rape, robbery, b & e) is the most frequent around certain area?
    - group by crime type (from all tuples with the input neighborhood) and select the crime type that has max count
4) What areas have higher accident rate (e.g. bicycle crash)?
    - Combine Bike_crash with Crime_mapping to get the neighborhood in which a bike crash happens. Group by neighborhood and select the one with max count of bike crash
5) Rate of severe injury bike crashing people accident?
    - select count of severe injury accident / count of all accident from Bike_crash_people
6) What areas have higher accident rate at night and what areas have higher accident rate at day?
    - Night: group by neighborhood, select the neighborhood from Bike_crash and Crime_mapping that has max count of night accidents
    - Day: group by neighborhood, select the neighborhood from Bike_crash and Crime_mapping that has max count of day accidents
7) What neighborhood have caught more gun offenders?
    - Group by neighorbood, select max count of tuples from Gun_offenders

8) Does weather correlate with accident rate? (bad weather accident : all accident)
    - select count of accidents happening under bad weather / total count of accidents

9) Does weather correlate with bicycle crash rate? (bad weather bike crash : all crashes)
    - select count of bike crashes under bad weather / total count of bike crashes
10) Does light condition correlate with bicycle crash rate?
    - select count of bike crashes under bad light condition / total count of bike crashes
11) Does road condition correlated with bicycle crash rate?
    - select count of bike crashes under bad road condition / total count of bike crashes
12) What is the rate of car accidents that are caused by driver exceeding speed limit?
    - select exceed_count / totalCount from (select count of bike crashes where drive_speed > speed limit) as exceed_count, (select count of accidents) as totalCount
13) For bicycle crashes, how many of the Drivers have Alcohol Detected?
    - select count(*) where Driver_Alcohol_Detected = 'Yes'
14) For people who committed crimes, how many of them are also gun offenders?
    - select count(*) where Crime_data.First name of People commited to crime = Gun_offenders.FirstName AND Crime_data.Last name of People commited to crime = Gun_offenders.LastName

15) What age range (10 years, for example) has highest number of gun offenders?
    - select max from (select count(*) from gun_offenders group by age)

6. how to load database
    - We first put all the csv files into Json objects and put it in a json file. Then, along with the raw data that is originally in Json format, we use JDBC to handle the object and establish connection with a heroku server. Then within the java file, we populate the data into the database.

7.
    - We expect the output to be table of data, and our work includes combine multiple rows of data and present it to the user in a user friendly view. Possibly some design on the table format to make the information easier to read.

8.
    - Since we are in section 315, we choose to minorly focus on complex data extraction. So far, our raw data consists of both csv and json files, and we also plan to do some statistical analysis/calculation on these files before we push all of the datasets to the database.
    - We also plan to touch on JDBC to handle database connections and Json objects.