

Introduction or background

Crimes in Boston is the dataset I will work with during the project. This dataset contains information about crimes committed in the Boston area during the years 2015-2018. The city of Boston uploads data pertaining to crimes in their city yearly (starting August 2015). Although the dataset I will work with only contains data compiled from 2015 to 2018, the city of Boston has more data from future years.

Crimes in Boston is an interesting dataset due to its depth of information despite being quite simple in its data and data fields. The dataset has many fields which can be used to look for correlations.

During this project the following questions will be explored with the use of the data set:

- 1) What time of the year has the most crime incidents?
- 2) What is the most common offense committed?

Dataset

The dataset is named Crimes in Boston containing one .csv file with all the data of crimes committed in Boston from 2015-2018. The following data was compiled from yearly datasets provided by Boston Police Department (BPD). Within the Crimes in Boston dataset the following 17 fields are contained: incident id number, offense, district, reporting area, time of crime, and location.

In our dataset, there are 327821 entries, meaning there are 327821 different crime incidents reported during 2015-2018.

I think for a relational database, normalizing the table would be beneficial. For example, offense_code and offense_description is already in a separate table, RMS_OFFENSE_CODES provided by the Boston Police Department. Another is the location, lat(latitude), and long(longitude) fields, where the location field is simply a tuple of the lat and long fields.

Methods

This is what our table looks like in SQLite:

Database Structure Browse Data Edit Primitives Execute SQL															
Table: crime															
Filter in any column															
INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOUR	UCR_PART	STREET	Lat	Long
1182079933	3207	Property Found	PROPERTY - FOUND	ES	691	NULL	2018-10-03 14:30:00	2018	10	Wednesday	14	Part Three	CENTRE ST	42.28709355	-71.14822128
1182079803	1415	Vandalism	GRAFFITI	ES	711	NULL	2018-09-27 12:00:00	2018	9	Thursday	12	Part Two	WASHINGTON ST	42.26643568	-71.15135416
1182079735	3006	Medical Assistance	SIX/INJURED/MEDICAL - PERSON	ES	716	NULL	2018-10-02 20:35:00	2018	10	Tuesday	20	Part Three	CENTRE ST	42.27072413	-71.1619128
1182079654	3801	Motor Vehicle Accident Response	M/V ACCIDENT - OTHER	ES	736	NULL	2018-10-02 16:06:18	2018	10	Tuesday	16	Part Three	HASTINGS ST	42.28678123	-71.15741315
1182079605	520	Residential Burglary	BURGLARY - RESIDENTIAL - FORCE	ES	717	NULL	2018-10-02 13:50:00	2018	10	Tuesday	13	Part One	BROOK RD	42.27315669	-71.16132745
1182079597	3170	Other	INTIMIDATING WITNESS	ES	822	NULL	2018-10-01 12:00:00	2018	10	Monday	12	Part Three	W BOUNDARY RD	42.2647628	-71.14657354
1182079531	3301	Verbal Disputes	VERBAL DISPUTE	ES	822	NULL	2018-10-02 09:56:00	2018	10	Tuesday	9	Part Three	GEORGETOWNE DR	42.25661385	-71.14449037
1182079521	616	Larceny	LARCENY THEFT OF BICYCLE	ES	822	NULL	2018-10-01 22:30:00	2018	10	Monday	22	Part One	GEORGETOWNE DR	42.25943465	-71.14798064
1182079453	3006	Medical Assistance	SIX/INJURED/MEDICAL - PERSON	ES	563	NULL	2018-10-01 20:12:00	2018	10	Monday	20	Part Three	FAIRDALE RD	42.28906908	-71.12116753
1182079419	520	Residential Burglary	BURGLARY - RESIDENTIAL - FORCE	ES	555	NULL	2018-10-01 18:27:00	2018	10	Monday	18	Part One	WHITFORD ST	42.27797284	-71.12520155
1182079371	706	Auto Theft	AUTO THEFT - MOTORCYCLE / SCOOTER	ES	822	NULL	2018-09-29 14:00:00	2018	9	Saturday	14	Part One	WASHINGTON ST	42.2636235	-71.15452494
1182079236	802	Simple Assault	ASSAULT SIMPLE - BATTERY	ES	822	NULL	2018-09-30 13:00:00	2018	9	Sunday	13	Part Two	GEORGETOWNE DR	42.25943465	-71.14798064
1182079195	3802	Motor Vehicle Accident Response	M/V ACCIDENT - PROPERTY DAMAGE	ES	564	NULL	2018-10-01 07:00:00	2018	10	Monday	7	Part Three	WASHINGTON ST	42.29571922	-71.11827307
1182079086	2629	Harassment	HARASSMENT	ES	691	NULL	2018-09-30 17:30:00	2018	9	Sunday	17	Part Two	REDLANDS RD	42.28898462	-71.15224773
1182079066	1842	Drug Violation	DRUGS - POSS CLASS A - HEROIN, ETC.	ES	NULL	NULL	2018-09-30 15:51:00	2018	9	Sunday	15	Part Two	GROVE ST	NULL	NULL
1182079065	3207	Property Found	PROPERTY - FOUND	ES	691	NULL	2018-09-30 15:00:00	2018	9	Sunday	15	Part Three	CENTRE ST	42.28709355	-71.14822128
1182079023	3114	Investigate Property	INVESTIGATE PROPERTY	ES	691	NULL	2018-09-30 11:18:00	2018	9	Sunday	11	Part Three	REDLANDS RD	42.28898462	-71.15224773
1182079019	3831	Motor Vehicle Accident Response	M/V - LEAVING SCENE - PROPERTY DAMAGE	ES	729	NULL	2018-09-30 10:00:00	2018	9	Sunday	10	Part Three	VFW PKWY	42.28536992	-71.17244016
1182078926	3831	Motor Vehicle Accident Response	M/V - LEAVING SCENE - PROPERTY DAMAGE	ES	664	NULL	2018-09-30 01:10:00	2018	9	Sunday	1	Part Three	WASHINGTON ST	42.29251815	-71.12165832
1182078890	3115	Investigate Person	INVESTIGATE PERSON	ES	559	NULL	2018-09-29 20:06:00	2018	9	Saturday	20	Part Three	METROPOLITAN AVE	42.28075962	-71.13254052
1182078828	3001	Medical Assistance	DEATH INVESTIGATION	ES	561	NULL	2018-09-29 17:02:00	2018	9	Saturday	17	Part Three	CUMMINGS HWY	42.28342716	-71.12422635
1182078817	3410	Towed	TOWED MOTOR VEHICLE	ES	556	NULL	2018-09-29 14:33:00	2018	9	Saturday	14	Part Three	CUMMINGS HWY	42.28110635	-71.12133926
1182078788	3006	Medical Assistance	SIX/INJURED/MEDICAL - PERSON	ES	560	NULL	2018-09-29 12:49:00	2018	9	Saturday	12	Part Three	FLORENCE ST	42.28491345	-71.12749155
1182078780	3115	Investigate Person	INVESTIGATE PERSON	ES	687	NULL	2018-09-29 12:07:00	2018	9	Saturday	12	Part Three	BELGRADE AVE	42.28624079	-71.13763724
1182078730	1402	Vandalism	VANDALISM	ES	748	NULL	2018-09-28 19:30:00	2018	9	Friday	19	Part Two	BROADLAWN PARK	42.30299879	-71.14449506
1182078724	1831	Drug Violation	DRUGS - SICK ASSIST - OTHER NARCOTIC	ES	692	NULL	2018-09-29 08:02:00	2018	9	Saturday	8	Part Two	GREATON RD	42.29224995	-71.15272034
1182078696	3108	Fire Related Reports	FIRE REPORT - HOUSE, BUILDING, ETC.	ES	740	NULL	2018-09-29 02:53:00	2018	9	Saturday	2	Part Three	SHELLMAN ST	42.29121633	-71.16516532
1182078663	3115	Investigate Person	INVESTIGATE PERSON	ES	550	NULL	2018-09-29 01:30:00	2018	9	Saturday	1	Part Three	HIGHCREST RD	42.27442004	-71.13863966
1182078616	706	Auto Theft	AUTO THEFT - MOTORCYCLE / SCOOTER	ES	822	NULL	2018-09-28 22:19:00	2018	9	Friday	22	Part One	EDGECREST TER	42.2628035	-71.15407725
1182078604	3114	Investigate Property	INVESTIGATE PROPERTY	ES	688	NULL	2018-09-28 21:17:00	2018	9	Friday	21	Part Three	CLARKSON AVE	42.26100508	-71.15176700

For SQLite:

- 1) Create a new database
- 2) Under File, Import>Table from CSV
- 3) Import the crime.csv file

For mySQL

- 1) Task>Import Data (this opens the import wizard)
- 2) Choose data source “Flat File Source” and choose crime.csv
- 3) Choose Destination “SQL Server Native Client 11.0”
- 4) Finish import

Results

Queries:

SELECT month, count(indicent_number) as “Number of Crime”

FROM crime

GROUP BY month;

SQL 1

1 SELECT month, count(incident_number) as "Number of Crime" FROM crime GROUP BY MONTH

	MONTH	Number of Crime
1	1	23625
2	2	21661
3	3	24156
4	4	24108
5	5	26242
6	6	30622
7	7	34640
8	8	35137
9	9	34023
10	10	26437
11	11	23685
12	12	23484

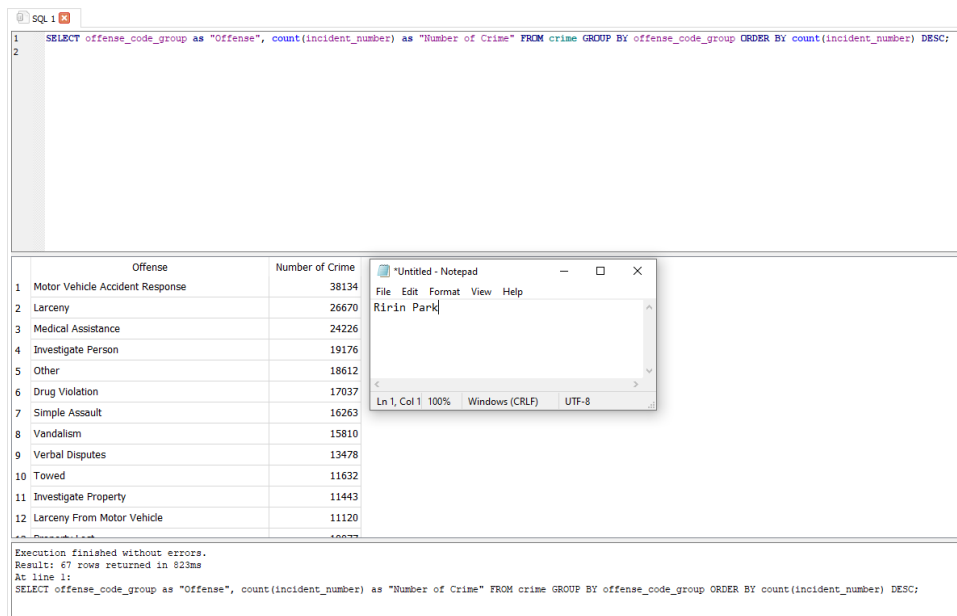
*Untitled - Notepad

File Edit Format View Help

Ririn Park

Ln 1, Col 1 100% Windows (CRLF) UTF-8

```
SELECT offense_code_group as "Offense", count(incident_number) as "Number of Crime"
FROM crime
GROUP BY offense_code_group
ORDEY BY count(incident_number) DESC;
```



When attempting to import the .csv flat file to the Microsoft SQL Server Management Studio, there occur some truncation errors. The table had all the fields, however none of the entities exist in the table.

After using MySQL through PHP from the IST webpage, the crime.csv file was too large for an import.



Discussion

Conceptually, SQLite and MySQL differ in how they store data. MySQL stores their data in servers while SQLite is embedded in the app.

As I was unable to find ways to fully import the .csv file to MySQL, from personal experience, SQLite had a much smoother experience with importing flat files such as .csv. However, my experience of the practicality of MySQL will not reflect how MySQL performs.