Dorchester Crime Analysis

Rich Huebner

December 24, 2020

1. Introduction

1.1 Background

Dorchester, Massachusetts is a large suburb located directly south of Boston. It is considered one of Boston's most diverse neighborhoods. You can find Umass Boston, the JFK Presential Library, the Neponset River, and Franklin Park all within Dorchester. Exceptional schools also serve Dorchester, including Boston Latin School.

The Boston Police Department (BPD) keeps track of all crimes in Dorchester, and reports basic data about these crimes periodically. The citizens of Dorchester and surrounding towns should be made aware of these crimes and their locations. Data is reported down to the number and street address, but without knowing exactly where those places are, it's better to see those places directly on a map. This gives a better sense of where crimes are being committed and what types of crimes.

1.2 Problem

There is a gap in knowing how close these crimes are to places of interest within Dorchester. As a patron of several venues within Dorchester city limits, it is beneficial to know the frequency and proximity of those crimes. Further, those living in Dorchester should know.

1.3 Interest

City officials and citizens of Dorchester and surrounding communities can benefit from additional analysis and information about crime locations, especially when people plan to move into the different neighborhoods.

2. Data acquisition

2.1 Data sources

Data was downloaded from The Universal Hub Web site contains various data for crime in Dorchester. This specific data set contains variables:

- Date of occurrence
- Description of crime
- Type of crime

- Street where it occurred

The dataset can be found here: https://www.universalhub.com/crime/dorchester.html

This analysis will use the data contained within the table on the Universal web page and load it into a data frame for further processing. Data may also be grouped according to crime type as part of the analysis. I used a geocoding library (geopy) for getting the latitude and longitude of each of the crimes.

3. Methodology

The basic methodology is as follows.

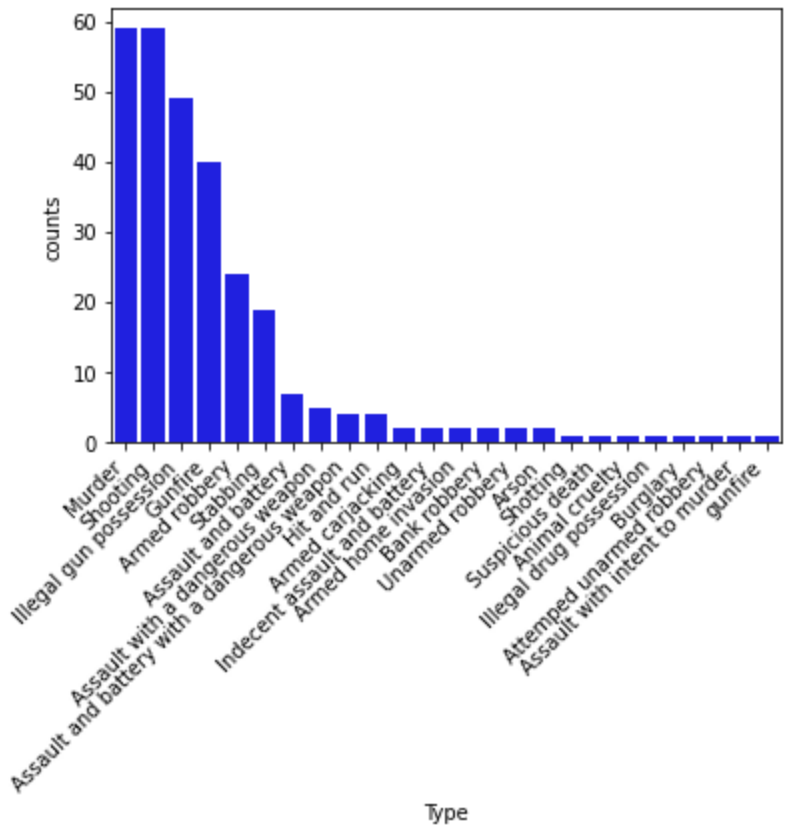
a. Load crime data from CSV file
b. Determine various categories of crime and provide basic exploratory analysis
c. Plot crime locations on a map
d. Find and plot venue data overlaid on the same above map
e. Conduct a cluster analysis on crime location data
f. Determine crimes that are very close (or at) certain venues

4. Results

The crime data from Dorchester covers 2018 through October 2020. A brief listing of the data is here:

	Date	Description	Type	Street	Year	Month
0	2020-12-04 03:15:00	Man shot repeatedly in Dorchester	Shooting	Victory Rd. and Adams St.	2020	12
1	2020-12-01 21:54:00	Police arrest man they say tried to rob two in...	Armed robbery	205 Adams St.	2020	12
2	2020-11-29 02:15:00	Police: Two guys were busy zipping through Dor...	Illegal gun possession	Columbia Rd. and Brinsley St.	2020	11
3	2020-11-26 13:40:00	Dorchester man charged as gunned-up sovereign ...	Illegal gun possession	53 McLellan St.	2020	11
4	2020-11-16 19:27:00	Six guns seized, three arrested in a park in D...	Illegal gun possession	Wainwright St. and Brent St.	2020	11

Next, I grouped the crimes by type to get the number of crimes reported within this data set.



Next, let's just look at the top five types of crimes that occurred in Dorchester during this period:

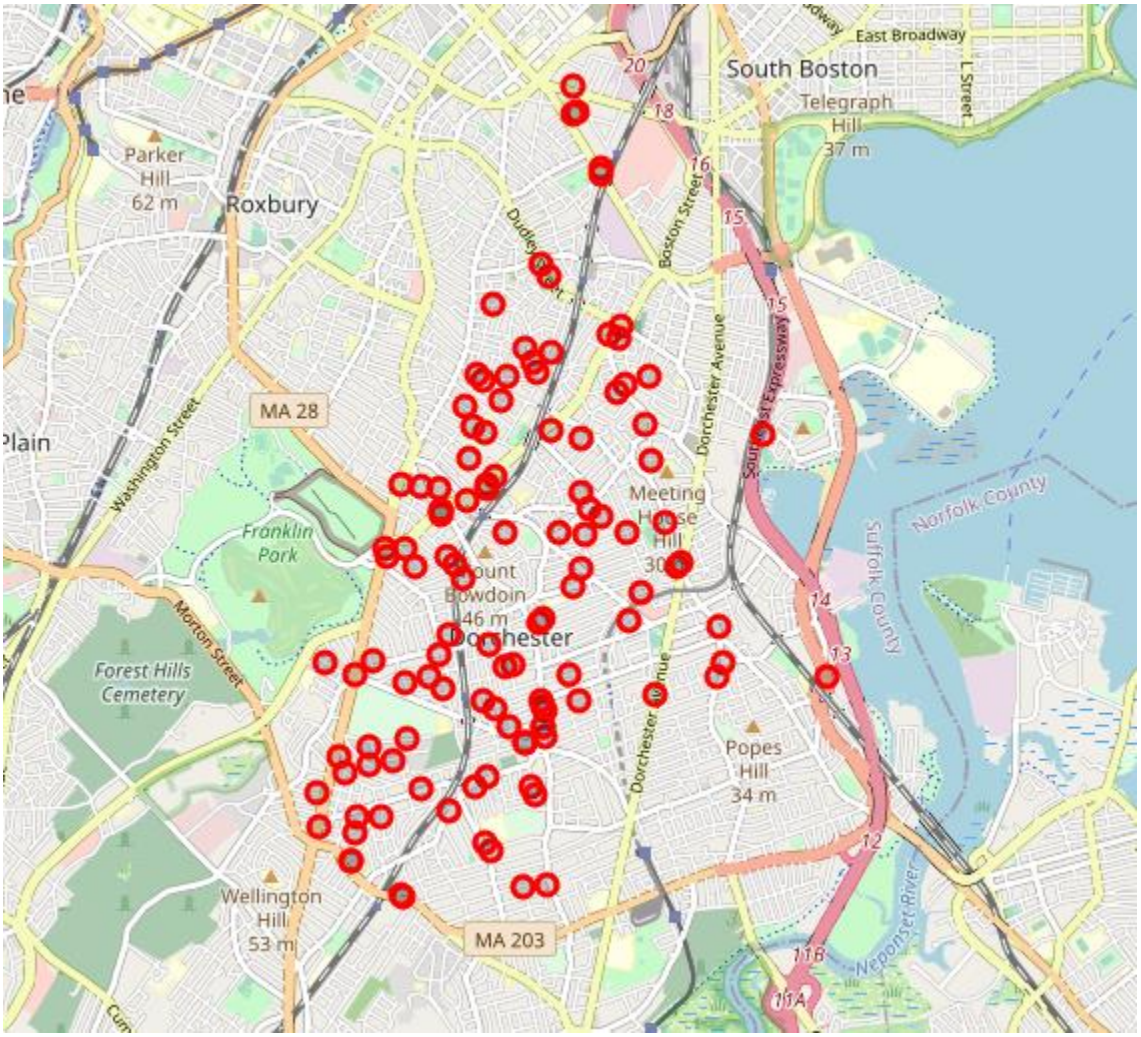
	Type	counts
17	Murder	59
18	Shooting	59
15	Illegal gun possession	49
12	Gunfire	40
3	Armed robbery	24

I also had to geocode each address where a crime occurred. I did this through using the Google Maps geocoding API, with these results, which are also saved back to a .csv file for later use.

	Date	Description	Type	Street	Year	Month	Latitude	Longitude
0	2020-12-04 03:15:00	Man shot repeatedly in Dorchester	Shooting	Victory Rd. and Adams St.	2020	12	42.295743	-71.056263
1	2020-12-01 21:54:00	Police arrest man they say tried to rob two in...	Armed robbery	205 Adams St.	2020	12	42.301834	-71.060200
2	2020-11-29 02:15:00	Police: Two guys were busy zipping through Dor...	Illegal gun possession	Columbia Rd. and Brinsley St.	2020	11	42.306013	-71.078422
3	2020-11-26 13:40:00	Dorchester man charged as gunned-up sovereign ...	Illegal gun possession	53 McLellan St.	2020	11	42.299677	-71.084191
4	2020-11-16 19:27:00	Six guns seized, three arrested in a park in D...	Illegal gun possession	Wainwright St. and Brent St.	2020	11	42.290027	-71.066875

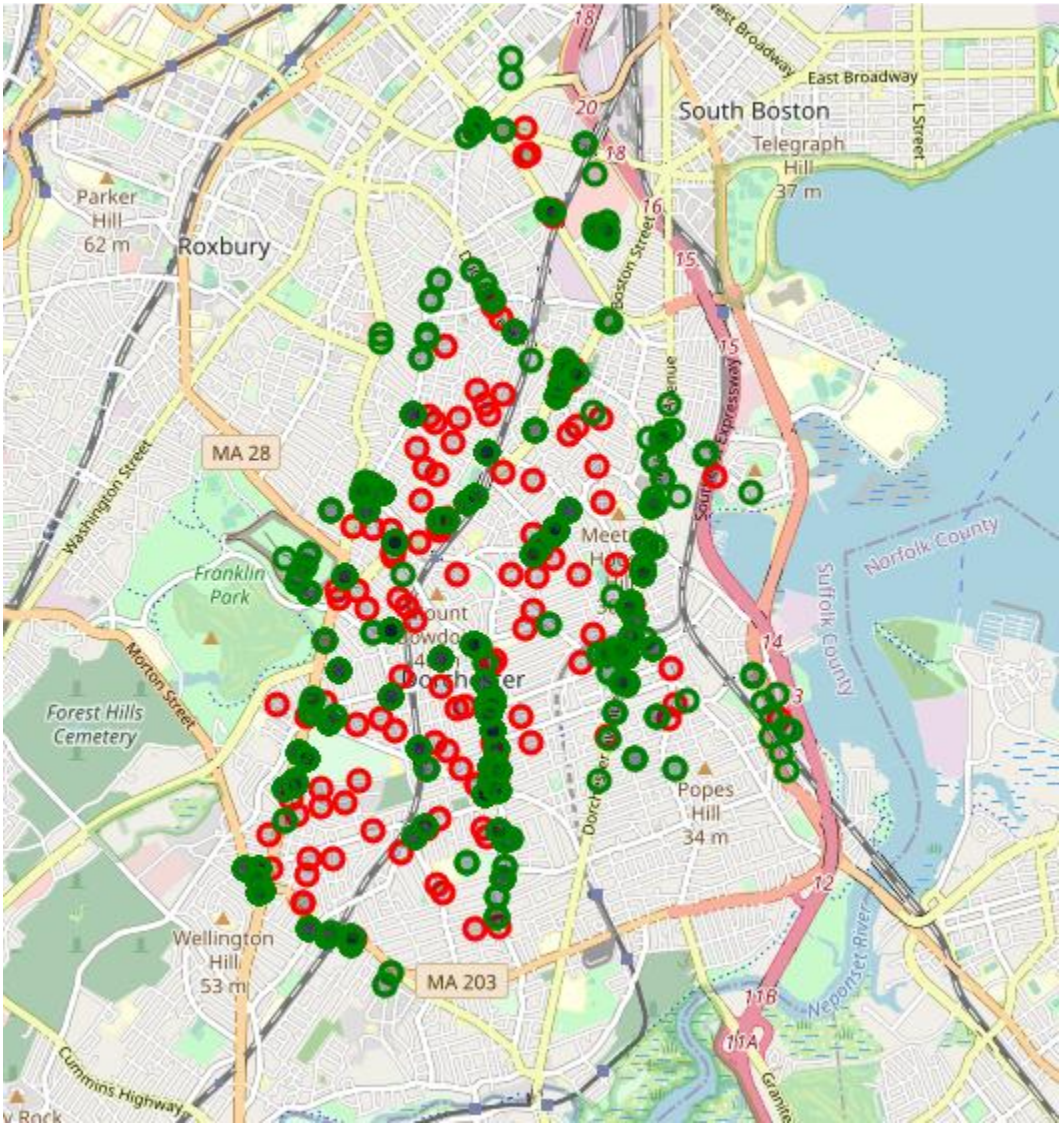
Next, for the purposes of the analysis and further mapping, I narrowed down the types of crimes to the two most frequently occurring crimes: murder and shooting. By subsetting the data, the analysis will become a bit cleaner.

The following map demonstrates where the shootings and murders took place in Dorchester during the time period. Murders and shootings are indicated in red.



The purpose of this analysis is to find venue and points of interest information that overlays the crime data. This can be useful for deciding whether to patronize area businesses or which you may want to stay away from. It is also helpful to understand if more crimes are occurring in certain neighborhoods of Dorchester – to find the relative safety of different parts of the city.

The next step of the analysis was to overlay the venue data. The following map shows the venues plotted in green.



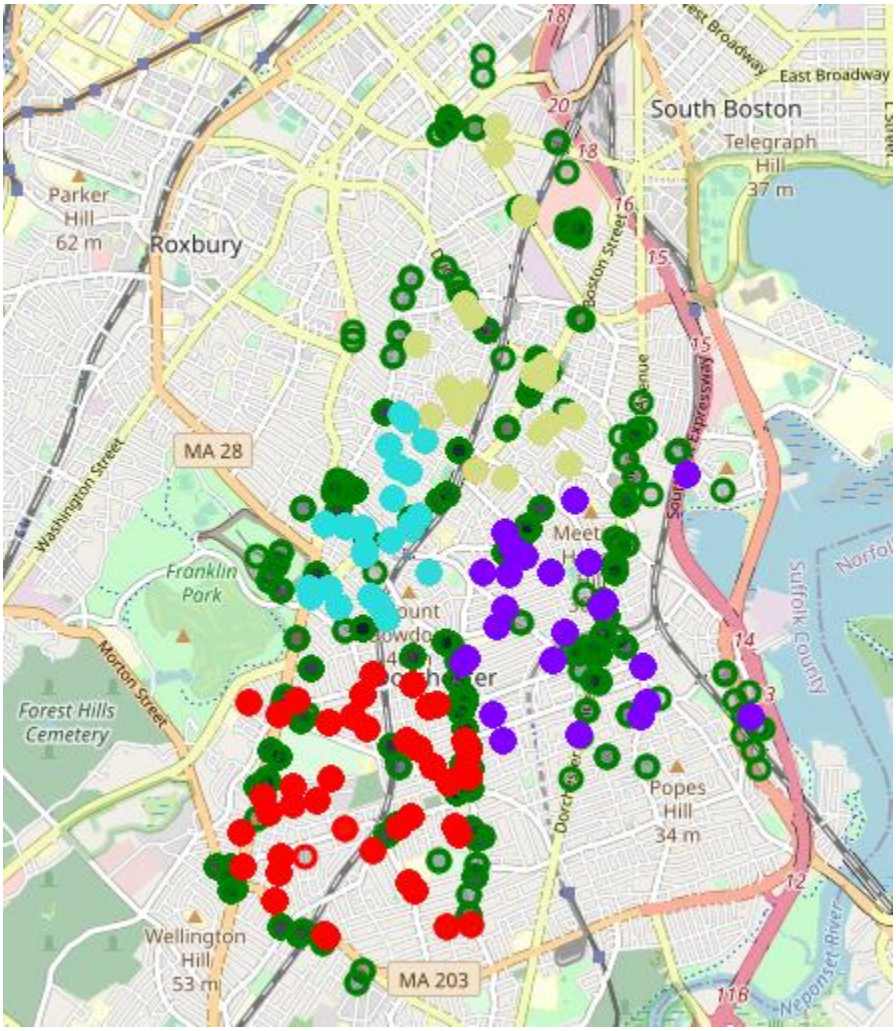
Here’s an example of the type of venue data retrieved from the FourSquare API.

	Address	Address Latitude	Address Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Distance
0	Victory Rd. and Adams St.	42.295743	-71.056263	Chill On Park	42.298746	-71.060477	Ice Cream Shop	481
1	Victory Rd. and Adams St.	42.295743	-71.056263	Anh Hong	42.299301	-71.058212	Vietnamese Restaurant	427
2	Victory Rd. and Adams St.	42.295743	-71.056263	SUBWAY	42.297147	-71.060495	Sandwich Place	381
3	Victory Rd. and Adams St.	42.295743	-71.056263	Slice O' Pie	42.294979	-71.057808	Pizza Place	153
4	Victory Rd. and Adams St.	42.295743	-71.056263	Dunkin'	42.296954	-71.060426	Donut Shop	368

You can see that there are several venues within the proximity of a crime that occurred on the intersection of Victory Rd. and Adams St.

Distances between the crime scenes and the venues were also calculated, also using the Google Maps geocoding API. These distances will be helpful in a moment to determine the venues closest to the crimes.

Cluster Analysis



The crimes were clustered based on location data. In this analysis, four clusters were used. There is a significant cluster of crimes that seem to be more concentrated in the southwest part of town, in red. These crimes occurred to the west of Washington St, between Gallivan Blvd and Harvard St. The crime scenes appeared to be more spread out the closer to Boston they occurred.

A significant number of crimes occurred between Columbia Ave and Dorchester Ave (the locals call it Dot Ave), but were not as tightly packed as the ones to the southwest.

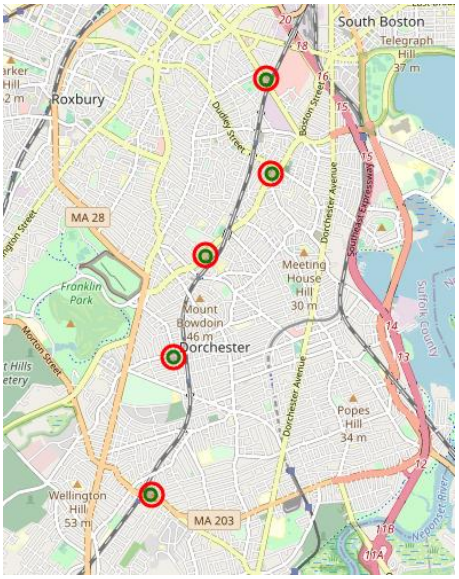
Interestingly enough, the Popes Hill area of town saw relatively few crimes compared to the other parts of town, with a distinct separation on Dot Ave.

5. Discussion

We can see from the above map that there are fairly clear cluster boundaries: a large cluster in the southwest part of Dorchester, as well as a cluster to the north part of town closest to Boston. This larger, southern-most cluster appears to occur around Washington St and to the west of Washington St. There are also two smaller clusters located in the east and west part of town. Interestingly enough, Dorchester Ave (known to locals as Dot Ave) marked a fairly clear boundary of crime, with a cluster of murders taking place over 2018-2020 in the Talbot Ave/Washington St area. The southwest part of Dorchester appears to have the most shootings and murders. There were 59 shootings and 59 murders during the 2018-2020 time frame.

Another area of concern is in and around Columbia Road. Cluster 1 and 2 seem to be located in that general area. I particularly got a chuckle from the FourSquare API which had venue information for a "Smelly Camel" at the Franklin Park Zoo. I think I'll pass, thanks!

Crimes that occurred either at or extremely close to the points of interest grabbed from the FourSquare API are as follows.



This northern-most crime occurred at Victoria’s Diner.

	Cluster	Address	Address Latitude	Address Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Distance
72	3	1024 Massachusetts Ave.	42.327016	-71.066816	Victoria's Diner	42.326974	-71.066808	Diner	4
300	0	144 Harvard St.	42.296203	-71.080767	Harvard and Bernard	42.296173	-71.080734	Liquor Store	4
201	2	Columbia Rd. and Devon St.	42.307488	-71.075953	Burritos Pizzeria	42.307429	-71.075947	Taco Place	6
833	0	130 Evans St.	42.281031	-71.083947	Chez Monique	42.280991	-71.084077	Cosmetics Shop	11
336	3	566 Columbia Rd.	42.316594	-71.066143	Upham's House of Pizza	42.316536	-71.065977	Pizza Place	15

There are 5 places where the analysis shows that the murder/shooting occurred at one of the venues reported by the FourSquare API. These include:

- Victoria's Diner on Mass Ave.
- Harvard & Bernard on Harvard St.
- Enterprise CarShare on Ames St.
- Upham's House of Pizza on Columbia Rd.
- Morton Seafood and Pizza on Morton St.

Victoria's Diner is a cute little place on Massachusetts Ave (we all call it Mass Ave) that offers almost 24/7 service on Friday and Saturday nights. Check it out here: https://www.victoriasdiner.com/

I have eaten at this cute little diner and was shocked to learn that a murder had taken place there. That is a little too close for comfort.

6. Conclusion

Overall, this analysis shows the use of grabbing web-based data, geocoding location data, and accessing venue/POI information via an API. Mapping both the crimes and POI data on a single map allows people to have more context about where crimes occurred in the city. Further analysis can be done by examining the clusters in more detail, and even perhaps bringing in other demographic data about Dorchester.

The analysis used Python libraries such as pandas, seaborn, folium, matplotlib, geopy, and other pertinent Python libraries, all done in a Jupyter notebook environment on an Intel i7-9700K 3.6GHz CPU, with 64GB of RAM and another 11GB of RAM on an nVidia GeForce RTX 2070