

Capstone Project Week 2

Find the safe hotels in New York neighborhood for consultants

Mal Thippana

Sep 1st 2019

Section 1

Introduction

Top consulting companies such as Ernest & Young, Deloitte, Pricewater coopers, Boston Consulting, KPMG hire consultants to work at their client locations. During the period of their assignment, they travel to these client locations from their homes and stay in hotels and eat out. All these boarding and lodging expenses are reimbursed by the consulting company. The consulting company interns would have accounted for all these expenses in finalizing the billing rate with the client. In the interest and safety of their consultants, the company allows their consultant to stay best and safe locations regardless of the daily hotel rate.

The problem

Since consultants have to stay near client location, it is good idea for the consulting companies to identify safe and nearby hotels and write a contract with the hotels. That way, the consulting company can get a better rate and pass those savings to the client to be competitive with other competing companies. It also provides a comfort to the consultants because they are staying at preferred hotels.

We are going to apply data science technology to identify best and safe hotels in New York

Section 2

Data and solving the problem

1. New York City Web site hosts a good number of free data feeds which among them is about the data containing Boroughs, Neighborhoods along with their latitude and longitude. This is used as a primary data source to identify the neighborhoods.

Source: https://cocl.us/new_york_dataset

2. NYPD collects and stores many feeds about crime and other data sets. For the current problem, I used YTD 2019 crime compliant data. This is a JSON feed which is then inputted into the algorithm. This dataset is used to understand the crime rates at different neighborhoods and select the ones that have low crime rates.

Source: <https://data.cityofnewyork.us/resource/5uac-w243.json>

3. Hotel data using Foursquare API <https://api.foursquare.com/v2/venues/search> is used to find venues which are in hotel category. For example.

- Hotel: 4bf58dd8d48988d1fa931735
- Bed & Breakfast: 4bf58dd8d48988d1f8931735
- Boarding House: 4f4530a74b9074f6e4fb0100
- Hostel: 4bf58dd8d48988d1ee931735
- Hotel Pool: 4bf58dd8d48988d132951735
- Inn: 5bae9231bedf3950379f89cb
- Motel: 4bf58dd8d48988d1fb931735
- Resort: 4bf58dd8d48988d12f951735

4. GeoSpace data to draw / visualize choropleth map. The identified neighborhoods (precincts are graphed on the map using this feeds)

<https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

<https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz>

5. Our goal is to find safe hotels among 306 neighborhoods . So, we used ranking of neighborhoods as relative comparison and selected the ones that are in top 50. The data needed for this is curated from two data sources

- Contains ranking: <https://www.dnainfo.com/crime-safety-report/ranking>
- Contains precincts: <https://www1.nyc.gov/site/nypd/bureaus/patrol/precincts-landing.page>

Data clean up

There are about 306 total neighborhoods in NY. After finding top 50 ranked neighborhoods, we removed any rows that contained NaN data. We followed the same for other data feeds. As required some of the data (for example: precincts value) needed to be converted into string from numeric to match with JSON data for geospatial mapping for choropleth.

All the data has been rechecked for accuracies to fit into data model

Section 3

Methodology used for solving the problems

Step 1: We have used Top 50 neighborhood ranking as relative safety comparison and selected only those that fell in that criteria.

Step 2: We then calculated mean crime count using YTD NYPD crime data and selected only the precincts that have less crime count than the mean.

Step 3: We then mapped data from steps 1 and step 2 to determine the neighborhoods that satisfy both the criteria.

Step 4: We then used results from Step 3 to run through the foursquare API using the latitudes & longitudes of the neighborhoods.

Step 5: We selected venues that fall in "Hotel" category and selected them for mapping on the geospatial map using choropleth map.

Section 4

Results

1. We eliminated neighborhoods that are less safe (selected top 50)
2. We eliminated neighborhoods based on recent crime rate (2019 YTD)
3. We were able to find 41 hotels from good/safe neighborhoods.
4. Off 306 Potential good neighborhood names are about 25

'Co-op City' 'City Island' 'Throgs Neck' 'Pelham Bay' 'Greenpoint' 'Williamsburg' 'Cobble Hill' 'Carroll Gardens' 'Red Hook' 'East Williamsburg' 'Upper West Side' 'Chelsea' 'Woodside' 'Sunnyside' 'South Ozone Park' 'Springfield Gardens' 'Rosedale' 'Rockaway Beach' 'Mariner's Harbor' 'Port Ivory' 'New Springville' 'Travis' 'Westerleigh' 'Graniteville' 'Arlington' 'Bloomfield' 'Bulls Head' 'Elm Park' 'Turtle Bay' 'Willowbrook'

Discussion and Future Direction

The model can be further enhanced by applying geo special K means algorithms DBSCAN to identify clusters where there is more crime, and exclude those areas for hotel selection. We can also extend this model using large volumes of data to fine tune the hotel identification

Section 5

Conclusion

It is very interesting to see how data science can apply in many areas. We can get a really good view into lots and lots of data and synthesize it into for meaningful conclusions and decisions. We can use many of the machine learning algorithms to study and solve complex problems that once thought require lot of effort. One should have curiosity into studying the data and it can do wonders and we will find things that we never would have thought.

In this case, if we don't study and apply these algorithms, we would not know what neighborhoods are safe to stay. We would probably decide by word of mouth or based on previous experience or simply by rates. But applying the data science algorithms, we were able to identify few neighborhoods that are safer to stay. Now consulting companies can go and work out a deal with the hotel management for getting a good deal on prices and it also provides comfort to the consultants whose safety depends on where they stay since they always live travelling.

Limitations

We used historical data to predict future stays. It doesn't guarantee that it is safer. Any potential crimes may take place in the identified neighborhoods. So, we can consider few more variables like unemployment rate, housing, drug usage and other predictive models to enhance this model

Annexure

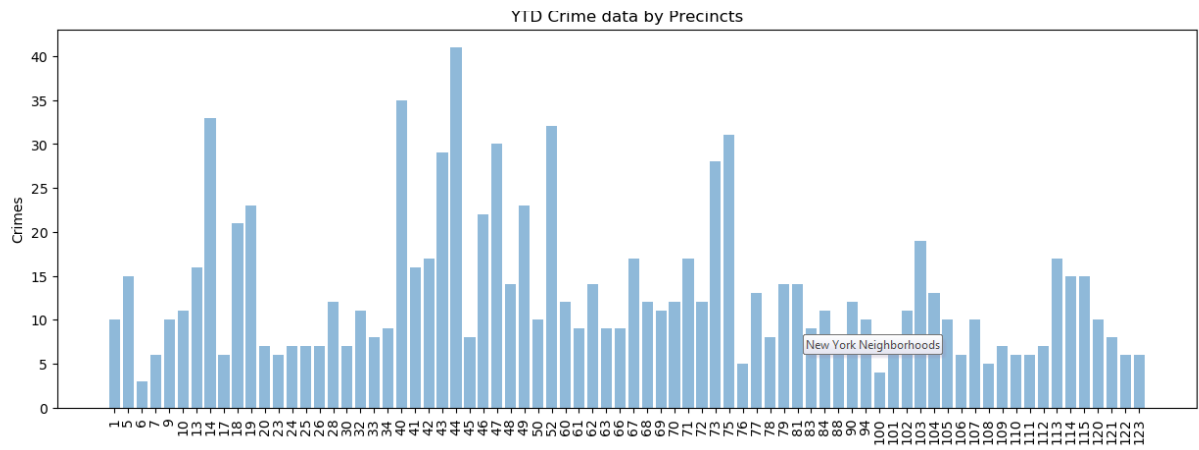
Neighborhood data

	A	B	C	D
1	Borough	Neighborhood	Latitude	Longitude
2	Bronx	Wakefield	40.89470518	-73.84720052
3	Bronx	Co-op City	40.87429419	-73.82993911
4	Bronx	Eastchester	40.88755568	-73.82780645
5	Bronx	Fieldston	40.89543743	-73.9056426
6	Bronx	Riverdale	40.89083449	-73.91258546
7	Bronx	Kingsbridge	40.88168737	-73.90281799
8	Manhattan	Marble Hill	40.87655078	-73.91065966
9	Bronx	Woodlawn	40.89827261	-73.86731497
10	Bronx	Norwood	40.87722416	-73.87939074
11	Bronx	Williamsbridge	40.88103888	-73.85744643
12	Bronx	Baychester	40.86685811	-73.8357976
13	Bronx	Pelham Parkway	40.8574135	-73.85475564
14	Bronx	City Island	40.8472467	-73.78648845
15	Bronx	Bedford Park	40.87018516	-73.88551218
16	Bronx	University Heights	40.85572708	-73.91041596
17	Bronx	Morris Heights	40.84789793	-73.91967159

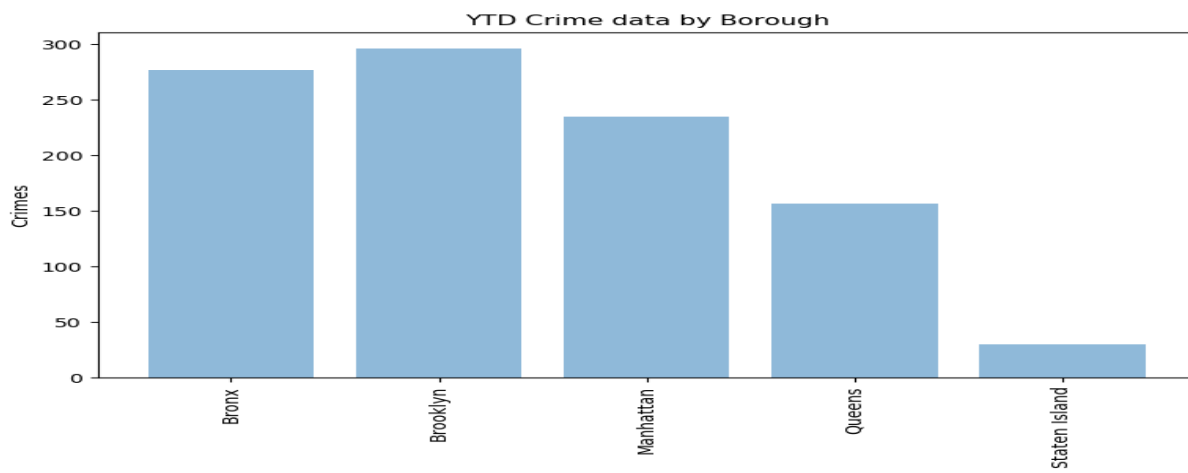
Top 50 safe neighborhoods

	A	B	C	D
1	Rank	Neighborhood	Borough	Precinct
2	35	Wakefield	Bronx	47
3	28	Co-op City	Bronx	45
4	35	Eastchester	Bronx	47
5	13	Fieldston	Bronx	50
6	13	Riverdale	Bronx	50
7	13	Kingsbridge	Bronx	50
8	13	Marble Hill	Manhattan	50
9	35	Woodlawn	Bronx	47
10	49	Norwood	Bronx	52
11	35	Williamsbridge	Bronx	47
12	43	Baychester	Bronx	49
13	43	Pelham Parkway	Bronx	49
14	28	City Island	Bronx	45
15	49	Bedford Park	Bronx	52

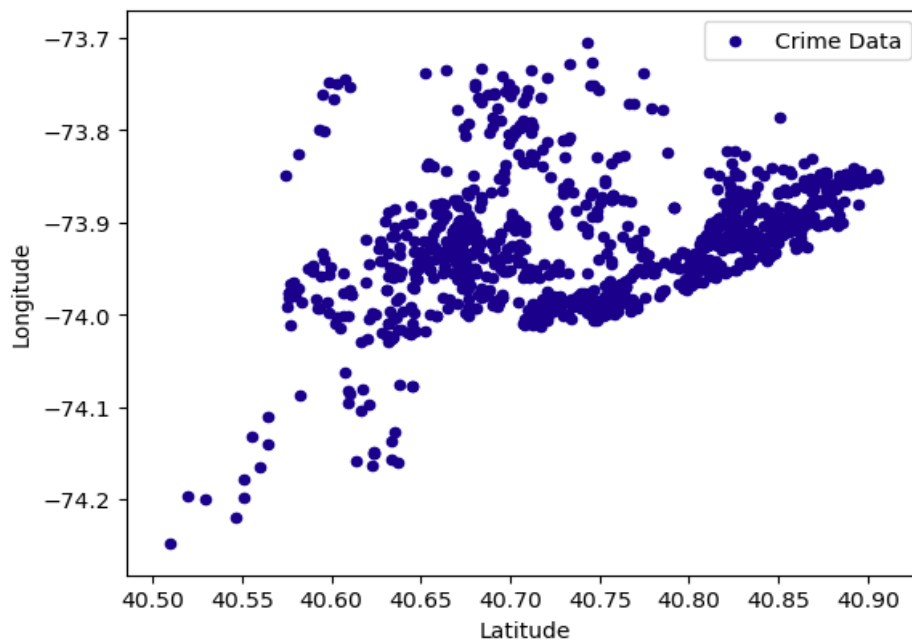
Crime count by Precinct



Crime count by Borough



Scatter plot of crime data



Identified hotels in choropleth map

