

The map displays the density of Airbnb stays across various New York City neighborhoods. The density is categorized by color: red for the highest density (over 100 stays), orange for high density (50-100 stays), yellow for medium density (20-50 stays), and green for low density (under 20 stays). Blue house icons indicate specific Airbnb listings. The map includes labels for major boroughs and surrounding areas: Teaneck, Englewood, Hackensack, Garfield, Lodi, Passaic, Wallington, Hasbrouck Heights, Ridgefield Park, Little Ferry, Palisade Park, Ridgefield, Cliffs Park, Fairview, Secaucus, Guttenberg, Union City, Weehawken, Hoboken, Jersey City, Manhattan, Bronx, Great Neck, Saddle Rock, Lake Success, New Hyde Park, Floral Park, and Roslyn.

THE BATTLE OF NEIGHBORHOODS

CHOICE OF NEIGHBORHOOD FOR AIRBNB STAY IN NEW YORK CITY

[illegible]

INTRODUCTION

- NEW YORK CITY
- DOMESTIC AND INTERNATIONAL TRAVELERS — PERSONAL SAFETY
- CHOICE OF ACCOMMODATION IN A NEW CITY

BUSINESS PROBLEM

- CHOICE OF SAFEST NEIGHBORHOOD TO STAY
 - BASING ON CRIME STATISTICS
 - VENUES/ PLACES OF INTEREST
 - AIRBNB PROPERTY LISTINGS IN THE CITY

DATA DESCRIPTION -1

- [HTTP://TOMSLLEE.NET/AIRBNB-DATA-COLLECTION-GET-THE-DATA](http://tomslee.net/airbnb-data-collection-get-the-data) - DETAILS OF AIRBNB LISTINGS (IN .CSV FORMAT) WITH THEIR LOCATION COORDINATES AND RATINGS INCLUDED FOR THE CITY OF NEW YORK WHERE ONLY DATA IN 2017 IS USED.

	borough	neighborhood
0	Queens	Jackson Heights
1	Brooklyn	Cypress Hills
2	Brooklyn	Sheepshead Bay
3	Manhattan	Hell's Kitchen
4	Manhattan	Upper East Side

DATA DESCRIPTION -2

- [HTTPS://DATA.CITYOFNEWYORK.US/PUBLIC-SAFETY/NYPD-COMPLAINT-DATA-CURRENT-YEAR-TO-DATE-/5UAC-W243/DATA](https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-CURRENT-YEAR-TO-DATE-/5UAC-W243/DATA) - New York City Crime Data acquired from City of New York website (.CSV file) which includes the type of crime, location of crime including the co-ordinates is used to overlay on the Map of New York using Folium Libraries in Python

	CMLNT_NUM	ADDR_PCT_CD	BORO_NM	CMLNT_FR_DT	CMLNT_FR_TM	CMLNT_TO_DT	CMLNT_TO_TM	CRM_ATPT_CPTD_CD	HAVEDELOPT	HOUSING_PSA	...	SUSP_SEX	TRANSIT_DISTRICT	VIC_AGE_GROUP	VIC_RACE	VIC_SEX	X_COORD_CD
0	651421035	41.0	BRONX	11/28/2018	00:00:00	11/28/2018	00:01:00	COMPLETED	NaN	NaN	...	U	NaN	25-44	BLACK	M	NaN
1	149013323	14.0	MANHATTAN	12/31/2018	23:40:00	12/31/2018	23:50:00	COMPLETED	NaN	NaN	...	NaN	NaN	45-64	WHITE	F	987866.0
2	642981531	73.0	BROOKLYN	12/31/2018	23:30:00	12/31/2018	23:40:00	COMPLETED	NaN	NaN	...	F	NaN	18-24	WHITE HISPANIC	F	1006995.0
3	429685363	67.0	BROOKLYN	12/31/2018	23:20:00	12/31/2018	23:30:00	COMPLETED	NaN	NaN	...	M	NaN	25-44	BLACK	F	999584.0
4	290330841	5.0	MANHATTAN	12/31/2018	23:15:00	12/31/2018	23:20:00	COMPLETED	NaN	NaN	...	U	NaN	25-44	WHITE	M	986164.0

DATA DESCRIPTION -3

- DATA SET FOR NEW YORK NEIGHBORHOOD DERIVED FROM EARLIER ASSIGNMENT - [HTTPS://IBM.BOX.COM/SHARED/STATIC/FBPWBOVAR7LF8P5SGDDM06CGIPA2RXPE.JSON](https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json) - .JSON FILE IS USED TO MAP THE NEIGHBORHOOD WITH ITS COORDINATES USING FOLIUM LIBRARIES

	Borough	Neighborhood	Latitude	Longitude
301	Manhattan	Hudson Yards	40.756658	-74.000111
302	Queens	Hammels	40.587338	-73.805530
303	Queens	Bayswater	40.611322	-73.765968
304	Queens	Queensbridge	40.756091	-73.945631
305	Staten Island	Fox Hills	40.617311	-74.081740

DATA DESCRIPTION -4

FOURSQUARE

- FOURSQUARE APIs – NEW YORK CITY GEOGRAPHICAL COORDINATES ARE USED AS INPUT, FOURSQUARE API CALLS ARE MADE TO ACQUIRE VENUES IN THE PRESCRIBED RADIUS

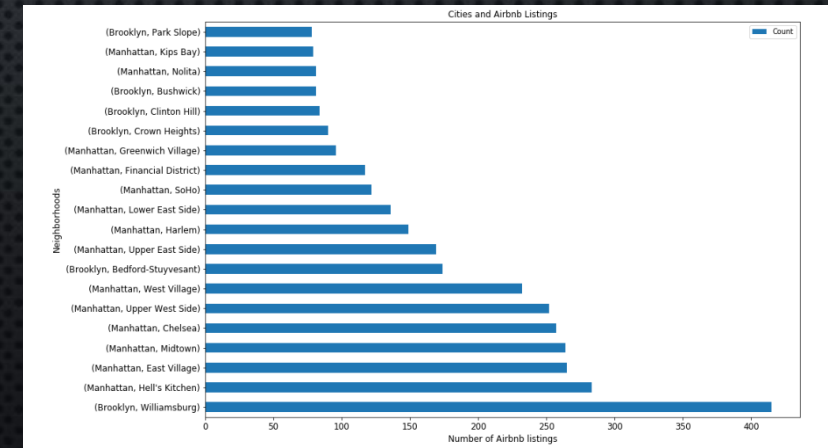
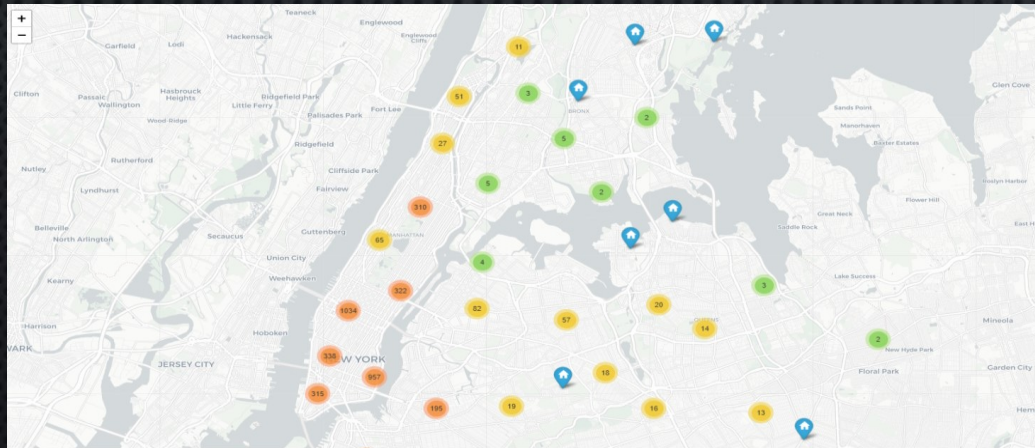
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.730862	-73.987156	Sake Bar Decibel	40.729416	-73.987745	Sake Bar
1	Wakefield	40.730862	-73.987156	Momofuku Ssäm Bar	40.731711	-73.985571	Asian Restaurant
2	Wakefield	40.730862	-73.987156	Han Dynasty	40.732130	-73.988090	Chinese Restaurant
3	Wakefield	40.730862	-73.987156	Hi-Collar - ハイカラ (Hi-Collar)	40.729449	-73.985918	Coffee Shop
4	Wakefield	40.730862	-73.987156	Casey Rubber Stamp	40.729962	-73.985098	Arts & Crafts Store

METHODOLOGY

EXPLORATORY DATA ANALYSIS

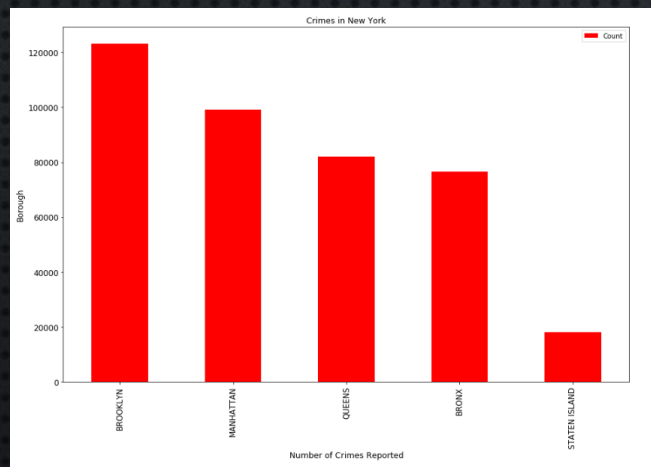
DATA 1

- AIRBNB DATA ACQUIRED FROM [HTTP://TOMSLEE.NET/AIRBNB-DATA-COLLECTION-GET-THE-DATA](http://tomslee.net/airbnb-data-collection-get-the-data)
- ONLY THE FIRST 5000 ROWS WERE LOADED
- TRANSFORMED THE ABOVE .CSV FILE INTO A DATAFRAME USING PANDAS
- DATA CLEANSING CARRIED OUT AND GROUPING PROPERTY DETAILS BY BOROUGH AND NEIGHBORHOOD TO ACQUIRE STATISTICS



DATA 2

- NEW YORK CITY CRIME DATA ACQUIRED FROM CITY OF NEW YORK WEBSITE (.CSV FILE) WHICH INCLUDES THE TYPE OF CRIME, LOCATION OF CRIME INCLUDING THE CO-ORDINATES
- [HTTPS://DATA.CITYOFNEWYORK.US/PUBLIC-SAFETY/NYPD-COMPLAINT-DATA-CURRENT-YEAR-TO-DATE-/5UAC-W243/DATA](https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243/data)
- THE DATASET CONTAINS 464035 ROWS WHICH WERE NOT CONDENSED EXCEPT FOR REMOVAL OF NA VALUES.
- BY CLEANING THE DATASET THE DATAFRAME IS GROUPED BY BOROUGH AND NEIGHBORHOOD COUNTING THE NUMBER OF CRIMES REPORTED IN EACH NEIGHBORHOOD.
- FOR VISUALIZATION AND STATISTICAL INFERENCING, TOTAL COUNT OF CRIMES IN EACH NEIGHBORHOOD IS CALCULATED AND PLOTTED ON A BAR GRAPH.



DATA 3

- DATA SET FOR NEW YORK NEIGHBORHOOD DERIVED FROM EARLIER ASSIGNMENT - [HTTPS://IBM.BOX.COM/SHARED/STATIC/FBPWBOVAR7LF8P5SGDDM06CGIPA2RXPE.JSON](https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json) - .JSON FILE IS USED TO MAP THE NEIGHBORHOOD WITH ITS COORDINATES USING FOLIUM LIBRARIES
- NY_NEIGHBORHOODS DATAFRAME CREATED WITH ASSIGNED COLUMNS
- POPULATED THE ABOVE DATAFRAME WITH THE IMPORTED JSON DATA AND SAVING IT AS .CSV FILE FOR FURTHER USE.
- BOROUGH NAMES WITH LOCATION COORDINATES DATAFRAME GENERATED WITH ABOVE DATA.

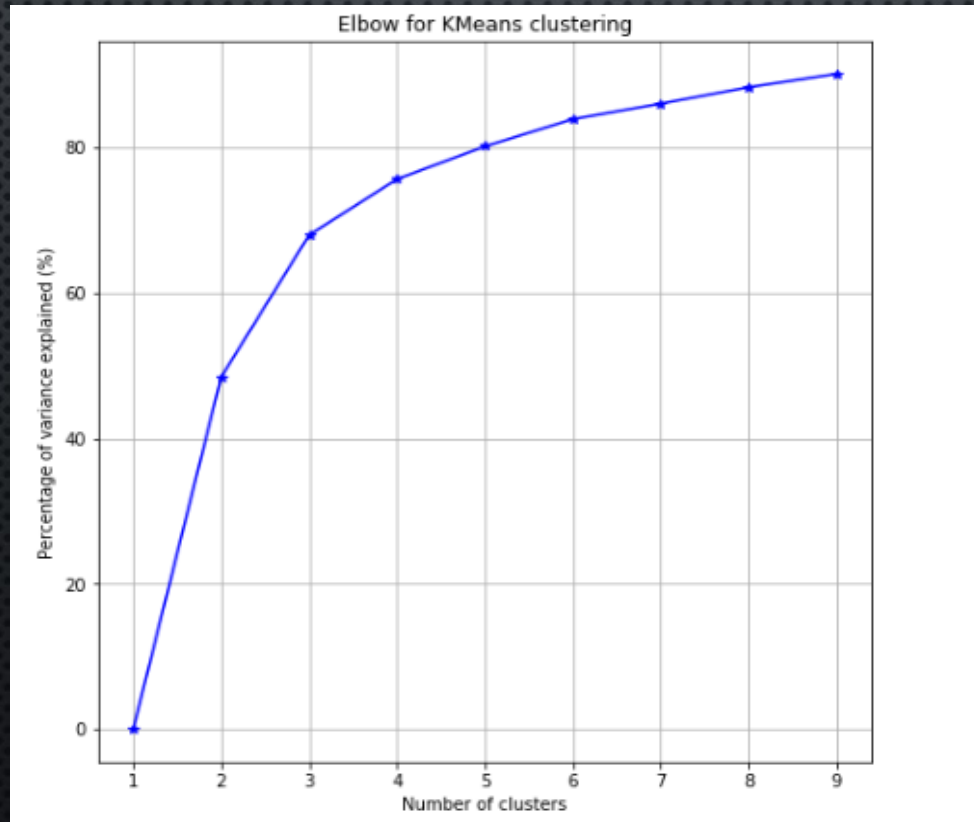
DATA 4

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.730862	-73.987156	Sake Bar Decibel	40.729416	-73.987745	Sake Bar
1	Wakefield	40.730862	-73.987156	Momofuku Ssäm Bar	40.731711	-73.985571	Asian Restaurant
2	Wakefield	40.730862	-73.987156	Han Dynasty	40.732130	-73.988090	Chinese Restaurant
3	Wakefield	40.730862	-73.987156	Hi-Collar - ハイカラ (Hi-Collar)	40.729449	-73.985918	Coffee Shop
4	Wakefield	40.730862	-73.987156	Casey Rubber Stamp	40.729962	-73.985098	Arts & Crafts Store

- FOURSQUARE APIs – NEW YORK CITY GEOGRAPHICAL COORDINATES ARE USED AS INPUT, FOURSQUARE API CALLS ARE MADE TO ACQUIRE VENUES IN THE PRESCRIBED RADIUS
- USING NOMINATIM, THE GEO COORDINATES OF NEW YORK WERE RETRIEVED.
- THE URL WAS DEFINED FOR PULLING VENUE DETAILS FROM FOURSQUARE API BY PASSING CREDENTIALS
- NEIGHBORHOODS AND NEARBY VENUE DETAILS WERE POPULATED IN THE DATAFRAME

METHODOLOGY – K-MEANS CLUSTERING

- THE RESTAURANTS IN THE VENUES LIST WERE TAKEN AND ONE HOT ENCODING WAS CARRIED OUT TO THE RESTAURANT DATAFRAME
- ELBOW METHOD WAS USED TO CALCULATE NUMBER OF CLUSTERS



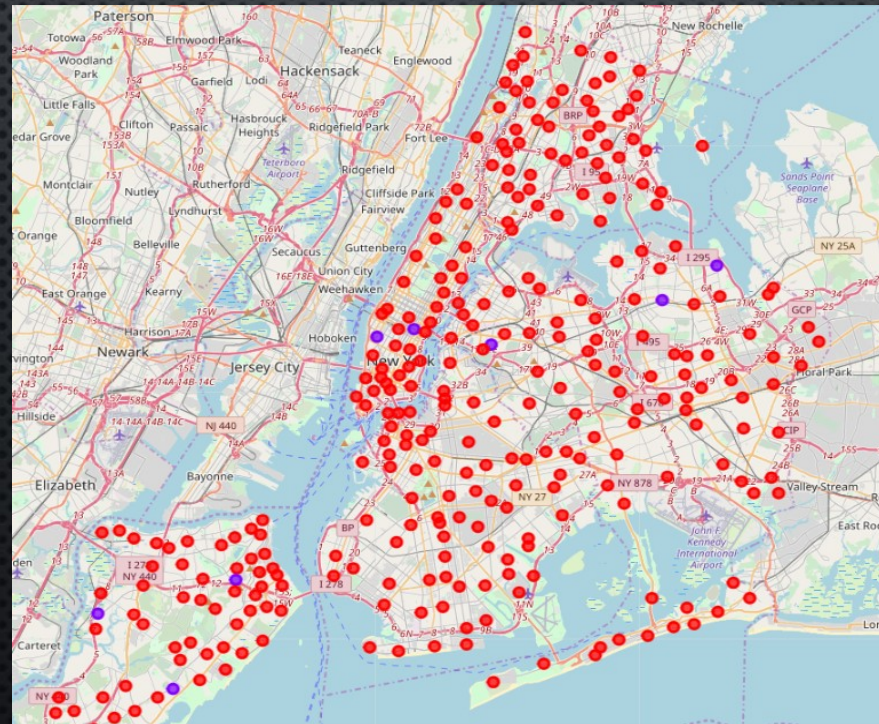
RESULTS

	American Restaurant	Arepa Restaurant	Asian Restaurant	Chinese Restaurant	Greek Restaurant	Italian Restaurant	Japanese Restaurant	Korean Restaurant	Mediterranean Restaurant	Mexican Restaurant	""	Soba Restaurant	Spanish Restaurant	Szechuan Restaurant	Tapas Restaurant	Udon Restaurant	Ukrainian Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Total	Total Sum
cluster0	1.0	1.0	1.0	3.0	1.0	1.0	3.0	2.0	1.0	2.0	...	1.0	1.0	1.0	1.0	1.0	1.0	3.0	2.0	33.0	66.0
cluster1	2.0	2.0	2.0	6.0	2.0	2.0	6.0	4.0	2.0	4.0	...	2.0	2.0	2.0	2.0	2.0	2.0	6.0	4.0	66.0	132.0

- **NEIGHBORHOOD K-MEANS CLUSTERING BASED ON MEAN OCCURRENCE OF VENUE CATEGORY**
- K-MEANS CLUSTERING ALGORITHM IS USED
- THE CLUSTER TOTAL AND CLUSTER SUM WERE ARRIVED AT AFTER CLUSTERING AS BELOW WITH CLUSTER 0 HAVING LESS TOTAL AND SUM THAN CLUSTER 1

RESULTS - CLUSTERS

- **CLUSTER 0:** THE TOTAL AND TOTAL SUM OF CLUSTER 0 IS SMALLER (SMALLEST OF 2 CLUSTERS) WHICH SHOWS THERE IS A SHORTAGE OF VENUES TO EXPLORE (IN COMPARISON WITH CLUSTER 1)
- **CLUSTER 1:** THE TOTAL AND TOTAL SUM OF CLUSTER 0 HAS THE HIGHEST VALUE WHICH SHOWS THE NUMBER OF RESTAURANTS ARE VERY HIGH



DISCUSSION

- SCOPE FOR IMPROVEMENT OF MODEL BY REFINING THE FOLLOWING:
 - POINTS OF INTEREST (E.G. CONCERT HALLS, MUSEUMS AND CARRYING OUT CLUSTERING) IN DEVELOPING THE MODEL
- CARRYING OUT CLUSTERING OF CRIME DATA
 - LINEAR REGRESSION RELATIONSHIP BETWEEN AIRBNB PROPERTY LOCATIONS AND VENUES
 - MULTILINEAR OR POLYNOMIAL REGRESSION RELATIONSHIP BETWEEN
 - AIRBNB PROPERTY LISTING
 - CRIME IN THE NEIGHBORHOOD
 - VENUES (POINTS OF INTEREST)

ASSUMPTIONS/LIMITATIONS

- DATASET OF AIRBNB LISTINGS WAS RESTRICTED TO 5000 ROWS WHICH MAY HAVE GIVEN US SKEWED RESULTS
- IT IS CONSIDERED THAT ALL AIRBNB PROPERTIES ARE AVAILABLE FOR OCCUPANCY FOR THIS PROJECT
- FOURSQUARE API RADIUS LIMITED TO 1000 AND RESULTS LIMITED TO 500
- DUE TO LARGE CRIME DATASET, RANDOM 500 ROWS WERE TAKEN FOR VISUALIZATION WHICH MAY HAVE SKEWED THE RESULT
- **MAPS FOR AIRBNB, CRIME LISTINGS AND VENUES CREATED SEPARATELY AS- DUE TO HIGH NUMBER OF DATA POINTS, THE FOLIUM MAPS ARE NOT RENDERING IN THE JUPYTER NOTEBOOK FOR OVERLAY OF THE ABOVE DATA POINTS IN A SINGLE MAP**

CONCLUSION

- THE ANALYSIS IS PERFORMED ON LIMITED/RESTRICTED DATA.
- BROOKLYN AND MANHATTAN WERE CLOSE IN THE STATISTICS OF AIRBNB LISTINGS ALTHOUGH MANHATTAN HAS HIGHER DENSITY OF LISTINGS PER SQ. MILE.
- THE MAXIMUM REPORTED CRIMES AS PER ANALYSIS WERE IN BROOKLYN FOLLOWED BY MANHATTAN. THE CLUSTERING OF RESTAURANTS HAS CLUSTER 1 SPREAD OVER BROOKLYN AND MANHATTAN ALMOST EVENLY.
- **LOWER MANHATTAN IS THE DESIRED LOCATION FOR STAYING AT AN AIRBNB WITH RELATIVELY LOWER CRIME STATISTICS AND MORE VENUES TO EXPLORE.**