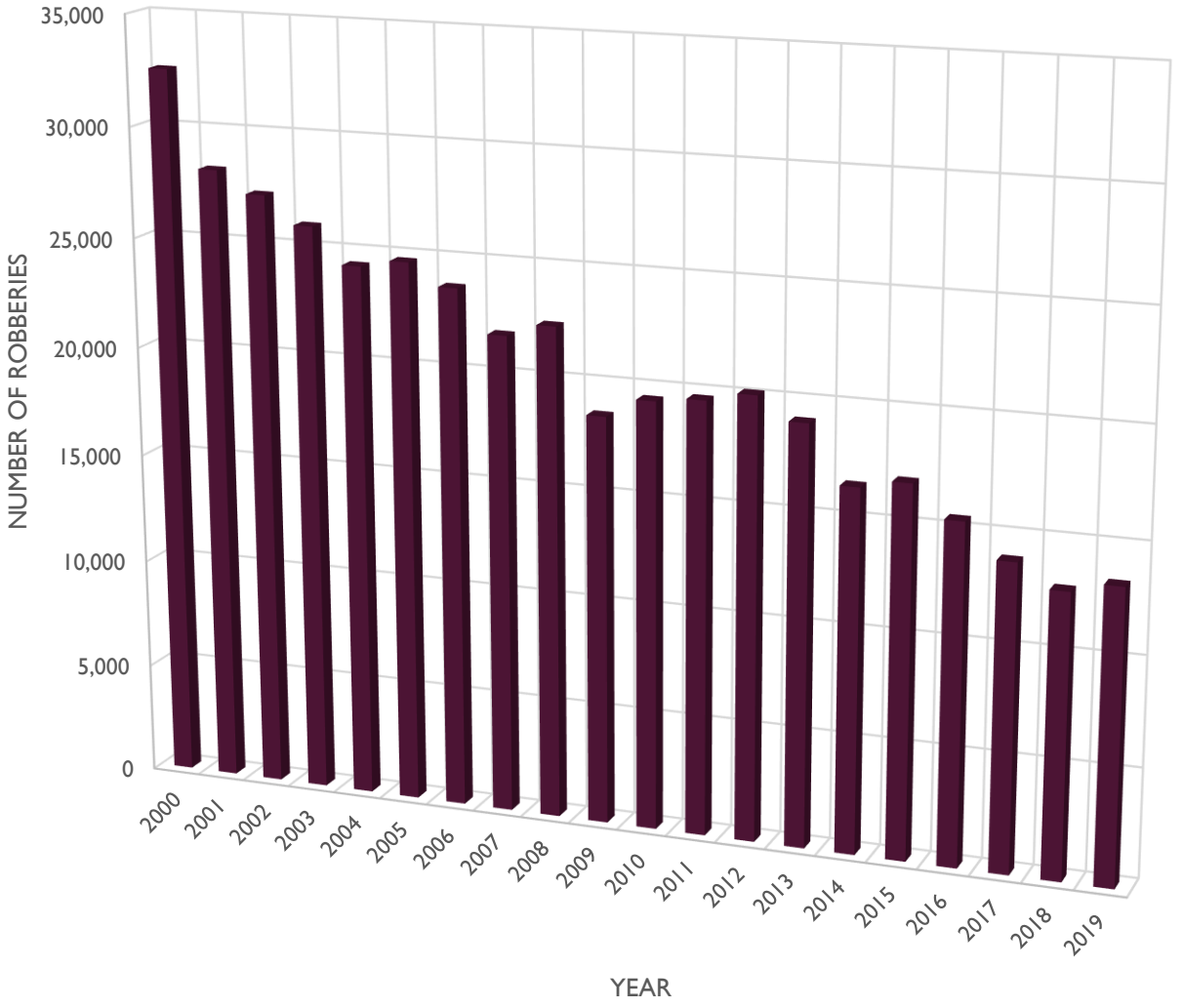
EXPLORING THE PREFERENCES OF ROBBERY CRIMES IN NEW YORK CITY

XINYU MA

BACKGROUND

- Robbery is one of the most common violent crimes in New York City.
- From 2000 to 2019, there are an average of about 20,000 robberies reported in NYC each year.



PROBLEM

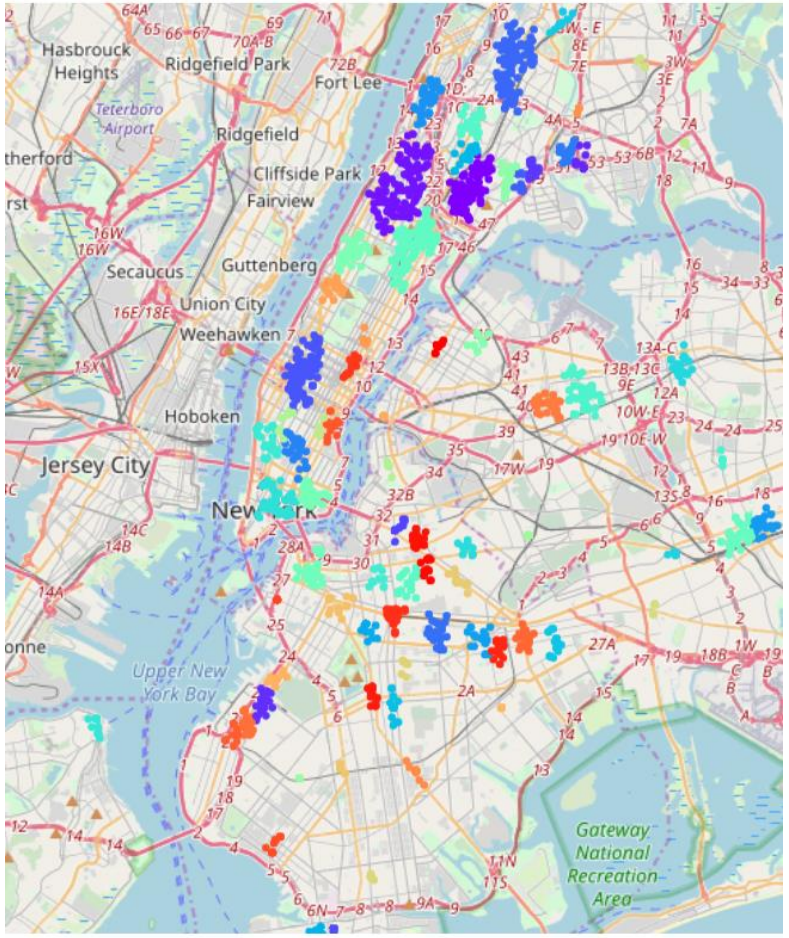
- The distribution of different categories of business and stores is likely to influence the risk of robbery in an area.
- Find a correlation between the robbery risk and distribution of business categories.
- Predict the robbery risks in the future, even for areas where the data of crime history is not available.

DATA ACQUIRING AND CLEANING

- Select all robbery records in 2018
- Leave latitudes, longitudes columns, drop other columns
- Calculate x, y coordinates

	ARREST_DATE	OFNS_DESC	Latitude	Longitude	x_coord	y_coord
0	2018-12-31	ROBBERY	40.673553	-73.866670	748832.285509	85499.716036
1	2018-12-31	ROBBERY	40.733927	-73.871582	748056.194523	92184.791880
2	2018-12-31	ROBBERY	40.851810	-73.909219	744175.993702	105110.962124
3	2018-12-31	ROBBERY	40.863086	-73.925693	742719.503585	106289.509355
4	2018-12-31	ROBBERY	40.816981	-73.921152	743375.630772	101187.687260

DBSCAN CLUSTERING OF ROBBERY RECORDS



- Left: clustered robberies are more densely packed in certain areas of the city
- Right: scattered robberies are more evenly spread all over the city with a lower density

DBSCAN CLUSTERING OF ROBBERY RECORDS

- Left: an example of convex hull of clusters of robbery records
- Right: some clusters do not have a convex hull due to all robberies were reported to occur at a single location (can only form a single dot) or two locations (can only form a line)

COMBINING VENUES INFORMATION WITH ROBBERY DENSITY DATA

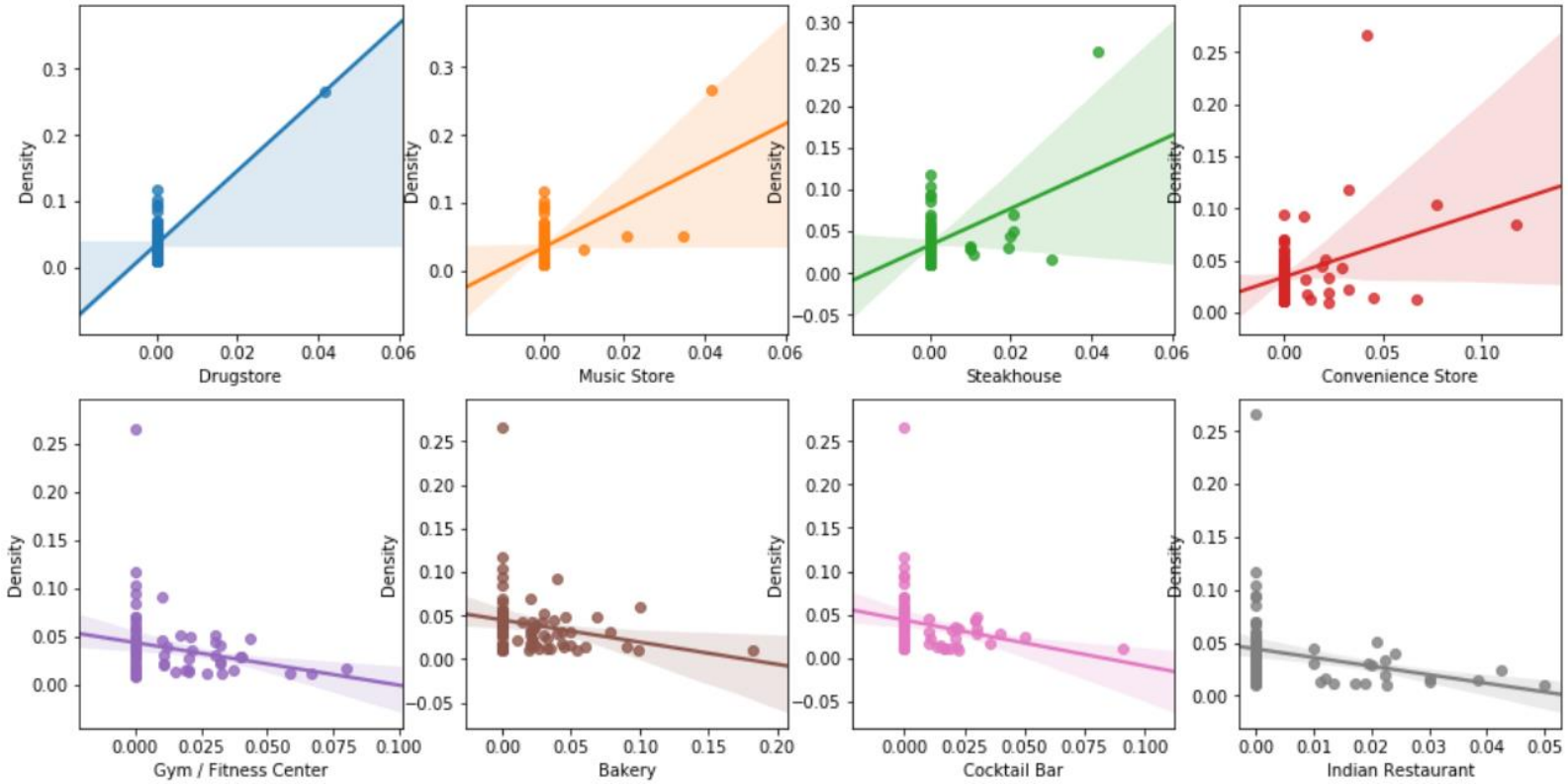
	Label	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	...	Video Store	Vietnamese Restaurant	Warehouse Store	Weight Loss Center
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0

	Label	Count	Center Latitude	Center Longitude	Area
0	0	309	40.814684	-73.920062	5979.788293
1	1	39	40.883147	-73.903487	802.784708
2	2	315	40.816792	-73.948283	8732.949515
4	4	118	40.823617	-73.870242	2133.276493
5	5	31	40.649825	-74.006771	2841.678924

- Merge the two DataFrames together

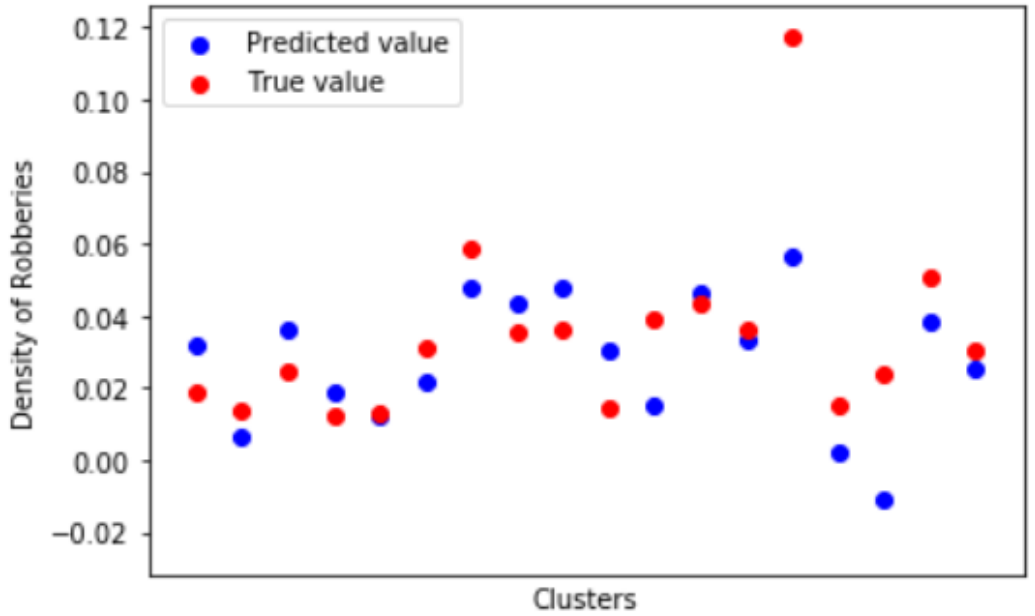
EXPLORATORY ANALYSIS

- Find out the features that influences the density of robberies the most
- Method: calculating the correlation coefficients
- Positively correlated:
  - Drugstore, music store, steakhouse, convenience store
- Negatively correlated:
  - Gym / fitness center, bakery, cocktail bar, Indian restaurant



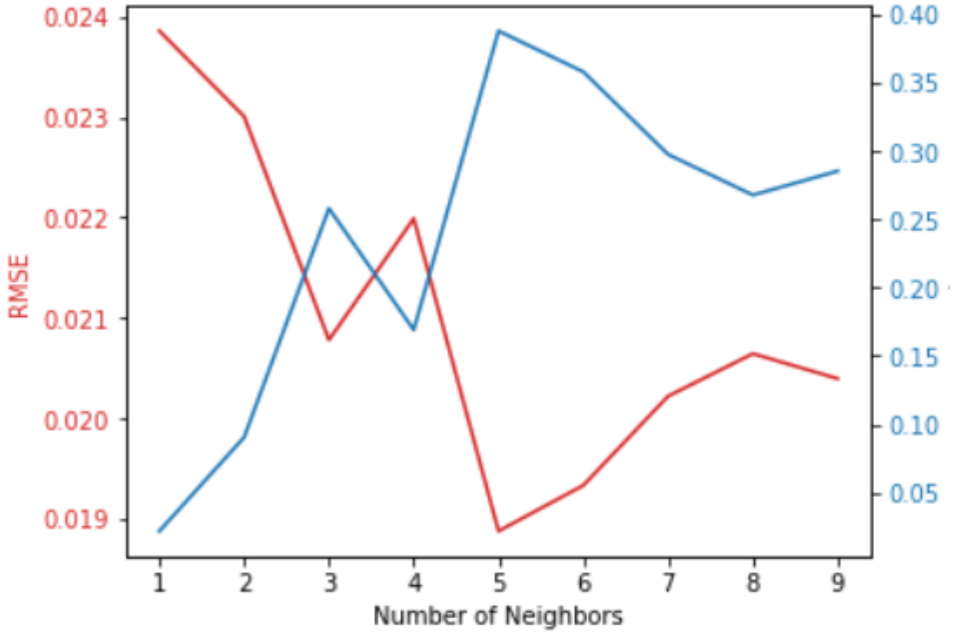
PREDICTIVE MODEL: LINEAR REGRESSION

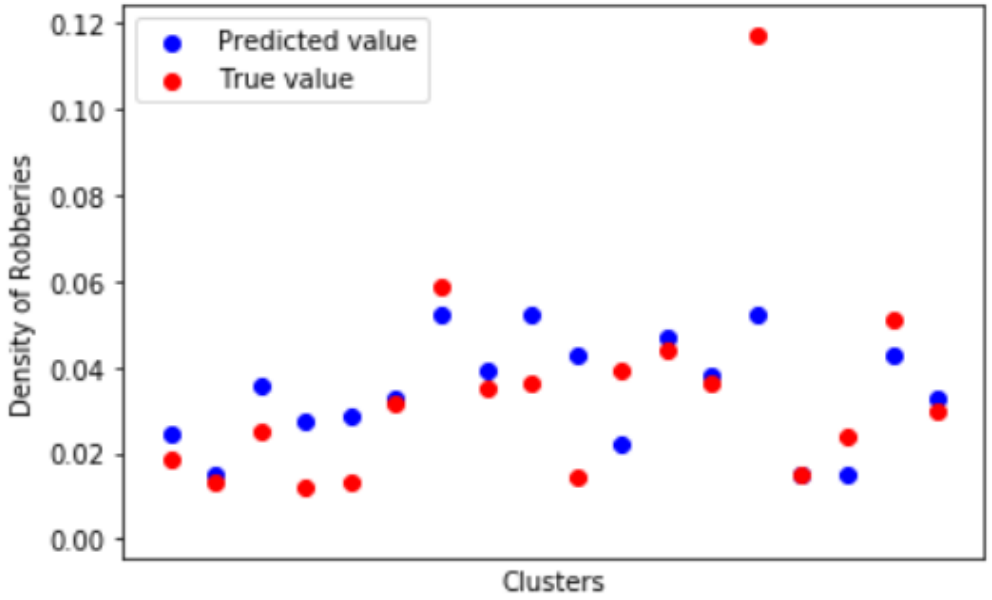
- RMSE = 0.0197
- R^2 SCORE = 0.336



PREDICTIVE MODEL: K-NEAREST NEIGHBORS REGRESSION

- Optimal k = 5
- RMSE = 0.0189
- R^2 SCORE = 0.388





CONCLUSION

- Positively correlated venues:
  - Drugstore, music store, steakhouse, convenience store
- Negatives correlated venues:
  - Gym / fitness center, bakery, cocktail bar, Indian restaurant
- Predictive modeling enables robbery risk prediction for any neighborhood with venues information available
- Future improvement:
  - More dataset fed to the models to increase accuracy of prediction