NEW YORK CITY LIVING

A look into the safest and most family friendly neighborhoods in NYC

JULY 8, 2020

MARK CHINTIS

I. Introduction

Suppose you receive a new job in New York City. Let's also suppose you have school aged children and you want to live somewhere safe. Who wouldn't, right? There are plenty of online resources that aim to help users find quality places to live. However, there are not a lot of resources that can give the specific results laid out in this project. It is also very difficult to find clear explanations or the data in which some of these qualities of life or quality of city ratings come from. This project aims to gather data from across the internet to provide a qualitative recommendation for the safest and most educational neighborhood in NYC. The data are collected from the New York City repository for crime, New York City's Department of Education repository for schools and programs, and real estate data from PropertyShark.com.

II. Data

Data for each of the NYC police precincts are scraped from the website https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz. Data from NYC Department of Education for the types and places of after school activities for children, and the "quality review" list of school that meet said quality criteria. The after-school activity dataset contains coordinate information, types of options available, among other unused information. The quality review data set contains only the names of the schools. These names are then cross-referenced with the Foursquare API. Other Foursquare API data is used for the collection of locations of afterschool activities in certain neighborhoods throughout New York City. The mean data of rental property costs were gathered from PropertyShark.com. These data are available at https://github.com/markchintis/NewYorkCityLiving.

III. Methodology

For this project, I assume that user is interested in finding the safest neighborhood to live in. For this reason, the 10 precincts with the lowest number of total crimes are shown in Figure 1.

	Precincts	Total Crimes
0	22	29.0
1	123	140.0
2	76	205.0
3	100	206.0
4	122	238.0
5	112	246.0
6	26	294.0
7	102	301.0
8	30	314.0
9	33	324.0

Figure 1 Ten lowest crime precincts

Ten neighborhoods with the highest total amount of after school activities are shown in Figure 2.

	Neighborhood	Count of After School Activities
0	Brownsville	33
1	East New York	26
2	Washington Heights South	24
3	East Harlem South	23
4	Central Harlem North-Polo Grounds	22
5	Crown Heights North	22
6	Bushwick South	21
7	Hunters Point-Sunnyside-West Maspeth	21
8	Chinatown	21
9	Prospect Lefferts Gardens-Wingate	20

Figure 2 Top 10 neighborhoods with highest amount of after school activities

Cross-referencing these neighborhoods with police precincts with Python's shapely module using polygons and points, Washington Height South (number 3 most after school activities) neighborhood is inside the 33rd precinct (the 10th safest precinct). Nearby schools are found in the Foursquare API at a one-kilometer radius from the centroid of the Washington Heights neighborhood (shapely module). From these schools, one school was found from cross-referencing the quality schools list. This school is P.S. 128 Audubon. Now that we are sure there are plenty of after school activities, a quality school in the neighborhood, housing costs need to be considered.

From the dataset collected from PropertyShark.com, median rent from 2006 to 2018 is available. To estimate the median rent in 2020, simple linear regression is employed with single predictor Year and outcome variable, Price. The model is given by

$$Y_i = b_0 + b_1 x + \epsilon_i$$

where $\epsilon \in iidN(0, \sigma^2), i = 1,2,3, \dots ,13$, and β_0, β_1 and ϵ are unknown parameters. K-fold cross-validation with 10 splits is used (although there is such small data) and a simple ensemble of the average of model intercepts and coefficients to optimize the prediction model. Since an ensemble method was used, Pearson's correlation inside Python's spicy.stats module was used to assess predictions versus actual prices. This model is finally used to predict the median rent cost in 2020.

IV. Results

The neighborhood within the top ten safest precincts, with the most amount of after school activities is Washington Heights South in Precinct 33. This neighborhood is in the Manhattan borough. It has 24 available locations with after school activities. The neighborhood with the most after school activities is Brownsville with 33. In 2019, the 33rd precinct had 324 total crimes recorded. The precinct with the lowest crimes in 2019 is the 22nd precinct (AKA Central Park Precinct), also in the Manhattan borough. There was one school, P.S. 128, in the Washington Heights South neighborhood on the NYC DOE quality schools list. Below is a choropleth map of the locations of after school activities and P.S. 128.

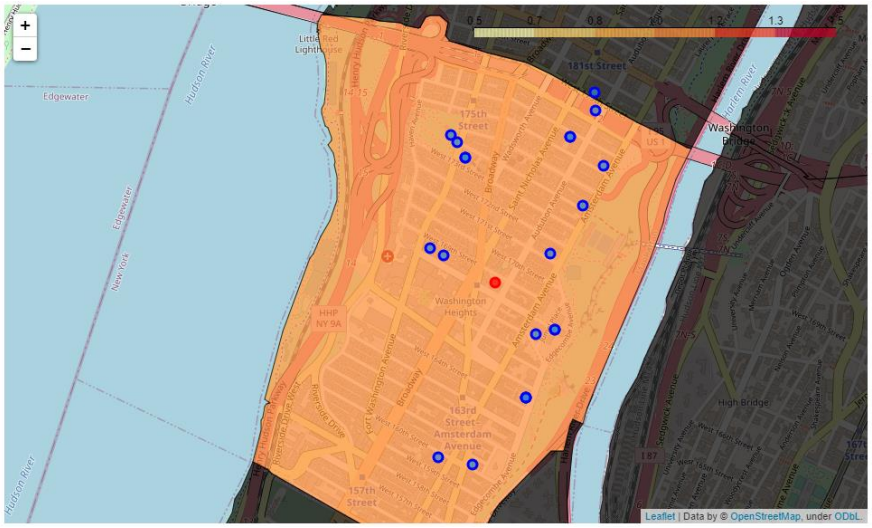

Figure 3 Choropleth map of Washington Heights after school activities(blue) and PS 128(red)

From the k-fold cross validated linear regression, the final model for median rent price is given by,

$$\hat{y}_i = -55566.44 + 28.22x_i + \epsilon_i$$

where $\hat{y}_i$ is the predicted value of median house rent price and x_i is a numerical variable for year. Using an ensemble k-fold validation models, a Pearson correlation coefficient of 0.966 with p-value 8.9 x 10^{-8}. Residual diagnostics testing (provided in the code in the appendix section https://github.com/markchintis/NewYorkCityLiving/blob/master/NewYorkCityLivingProject.ipynb) shows that the errors relatively normal and have uniform variance. Assuming the individual using this program will be moving to New York City in the year 2020, the estimated median rent price would be

$$\hat{y} = -55566.44 + 28.23(2020)$$

$$= \$1458.16.$$

V. Discussion

The analyses above displayed how 3 simple life-style choices (safety, after school activities, and quality schools) can affect one's choice in neighborhoods in which to life. This project assumed that cost was no issue and the deliverable of rent price was just an informative number. Provided with more time and data resources, the calculations of the prices of the individual neighborhoods could also be used as a determining factor for overall best neighborhood to live. What we can recommend however from this project is that a family moving to Washington Heights South in Manhattan can expect a very low crime rate compared to the rest of the city, being a top 10 neighborhood and the overall safest borough. The user can also enjoy the fact that there are 24 after school activities in the neighborhood to send their kids and a top school according to the NYC Department of Education.

VI. Conclusion

Although moving to a new city, especially a gigantic one like New York City, the use of algorithms such as the one used in this project can be very useful. Adjustment of simple parameters can help give users a much clearer idea of the place they are moving to than some of the current online resources.