SELECTING THE BEST NEIGHBORHOOD TO LIVE IN NEW YORK

Author: Thahseen AG

INTRODUCTION

The Bronx, Brooklyn, Manhattan, Queens, and Staten Island are the five boroughs of New York, which are located where the Hudson River meets the Atlantic Ocean. For each of the five fundamental constituent parts of the newly consolidated city, the word borough was adopted to define a specific form of governmental administration. Manhattan, a densely populated borough that is part of the country, is at its heart.A densely populated borough that is home to some of the world's most significant commercial, environmental, and cultural hubs The Empire State Building and Central Park are two of the city's most prominent landmarks.A place where people live and communicate with each other is a neighborhood. Neighborhoods tend to have their own reputation based on the people who live there and the surrounding locations, or "feel". In terms of big cities, neighborhoods are typically listed, but suburban or rural areas often have neighborhoods.

BUSINESS PROBLEM

The objective is to select a neighborhood in New York with a low crime rate and having the amenities like Grocery Store,Restaurant,Gym,Pharmacy nearby.

DATA

Data that might contribute to selecting a neighborhood include the overall crime that has occured in the neighborhood for the past one year and an amenity score which determines desired amenities near to the neighborhood.

Data Sources

The crime reported in each borough will be taken from the NYPD Open Data available. To complement the above data, we will be using a Borough to Neighborhood mapping data set along with the Latitude and Longitude of each Neighborhood, available in

https://cocl.us/new_york_dataset. The above datas are combined with the Foursquare data to 'score' each neighborhood on the basis of the availability of desired amenities.

Data Wrangling

The Data downloaded from NYPD OPen Data had many undesired columns which were not required for further processing. Only the below features are used for crime processing and the columns are renamed to a more readable name format for easy understandability.The resulting data set has 1000 samples with 5 features to the data.

Column Name	Renamed Column Name	Description
BORO_NM	Borough	The name of the borough in which the incident occurred
CMPLNT_FR_DT	Date	Exact date of occurrence for the reported event (or starting date of occurrence, if CMPLNT_TO_DT exists)
OFNS_DESC	Offence	Description of offense corresponding with key code
Latitude	Latitude	Midblock Latitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)
Longitude	Longitude	Midblock Longitude coordinate for Global Coordinate System, WGS 1984, decimal degrees (EPSG 4326)

Table 1. New York Crime Data Schema

The data from the New York Data set include each borough with its neighborhoods and Latitude and Longitude.

METHODOLOGY

The data obtained from the NYPD is combined with the New York Data set to determine which neighborhood the crime occured.

	DATE	OFFENSE	BOROUGH	Latitude	Longitude
2020000	2019-04-12	CRIMINAL MISCHIEF & RELATED OF	MANHATTAN	40.764007	-73.996005
2020001	2019-04-12	DANGEROUS DRUGS	QUEENS	40.602195	-73.749104
2020002	2019-04-12	HARRASSMENT 2	QUEENS	40.751482	-73.822033
2020003	2019-04-12	PETIT LARCENY	MANHATTAN	40.718027	-73.999958
2020004	2019-04-12	PETIT LARCENY	BROOKLYN	40.630754	-73.977167

Fig 1. New York Crime Data Set

	Borough	Neighbourhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Fig 3. Neighborhood Latitude and Longitude Data Set

This is done with the Haversine formula, which determines the great-circle distance between two points on a sphere given their longitudes and latitudes.

$$d = 2r\ arcsin(\sqrt{sin^2(\varphi_2 - \varphi_1)/2\ + cos(\varphi_1)cos(\varphi_2)sin^2(\lambda_2 - \lambda_1/2)}\ )$$

where,
φ_1, φ_2 are the latitude of point 1 and latitude of point 2 (in radians),
λ_1, λ_2 are the longitude of point 1 and longitude of point 2 (in radians).

d is the distance between the two points along a great circle of the sphere (see spherical distance),

r is the radius of the sphere.

	DATE	OFFENSE	BOROUGH	Latitude	Longitude	Neighbourhood
2020000	2019-04-12	CRIMINAL MISCHIEF & RELATED OF	MANHATTAN	40.764007	-73.996005	Clinton
2020001	2019-04-12	DANGEROUS DRUGS	QUEENS	40.602195	-73.749104	Far Rockaway
2020002	2019-04-12	HARRASSMENT 2	QUEENS	40.751482	-73.822033	Queensboro Hill
2020003	2019-04-12	PETIT LARCENY	MANHATTAN	40.718027	-73.999958	Little Italy
2020004	2019-04-12	PETIT LARCENY	BROOKLYN	40.630754	-73.977167	Borough Park

Fig 2. Neighborhood Crime Details Data Set

Once the nearest neighborhood to each crime is obtained, the crime data set is merged with the neighborhood details for each crime. The 'score' for desired amenities is calculated by the weighted average.

Amenties	Score
Grocery Store	3
Restaurant	3
Gym	2
Pharmacy	2

Table 2. Amenities Score

Use the foursquare API to obtain the venues near the neighborhood.Each neighbourhood is enlisted different types of Venues under the venue category.Each of the category is then grouped with only the desired amenities.Each of the amenities is then scored with a weighted average score to determine desirability of the neighborhood. These Venue Dataset,along with crime data sets are used to determine the desirable neighborhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
4	Wakefield	40.894705	-73.847201	Subway	40.890468	-73.849152	Sandwich Place

Fig 4. Venue Details for Neighborhood

	Neighborhood	Gym	Grocery	Pharmacy	Restaurant	Score
0	Allerton	0.0	0.0	0.0	1.0	0.3
1	Annadale	0.0	0.0	0.0	2.0	0.6
2	Arden Heights	0.0	0.0	1.0	0.0	0.2
3	Arrochar	0.0	0.0	0.0	2.0	0.6
4	Astoria	1.0	0.0	0.0	2.0	0.8

Fig 5. Amenities Score using Average Weight