The battle of the neighborhoods

Finding best socio-economically suitable location for living in Jersey City, US.

SEPTEMBER 3

IBM COURSERA DATA SCIENCE
Authored by: Santosh Tharali

IBM

coursera

Contents

1. Business Problem

Jersey City is the second most populous city in the U.S. state of New Jersey, after Newark. Jersey City is raising its profile as an alternative to New York City, given its closeness to the Big Apple and more reasonable housing prices. Located just across the Hudson River from Manhattan's West Side, Jersey City is being touted by some as the latest alternative to New York City's sweltering real estate market. People who choose to remain close to New York but do not want to spend on luxury real estate prices of NYC, see Jersey City as a cheap suburb of New York City. If you are taking job in NYC or coming for study, you will very likely be going to stay in Jersey City.

You will look for neighborhood that has great amenities and essential venues such as fast food joints, pharmacies, parks and same time it is economically affordable. Some neighborhoods are less densely populated. If you like living in peaceful surrounding or love to be close to nature you would consider neighborhood having Parks or less buzzing. You do not find holistic snapshot of competitive difference between neighborhoods on apartment rental websites. Sometimes you do not want to go by information on the rental websites as the data is motivated to sell apartment. Visiting government statistics websites, rental websites and venue APP to gather and crunch information is overwhelming for you and take plenty of time. Despite tremendous effort, you still be not sure you got right deal, perhaps you still be missing on insights that are not easily visible and could enable best deal. Data science provide deep insight into massive and complex data and enable you to make informed decision. Our research paper finds neighborhoods with similar characteristics in terms of socio-economic indicator and venues, and offer you vital parameters to enable you to decide on best neighborhood suitable for you for living in the city.

1.1. Target Audience

Different groups of people would benefit from our project's results, namely:

1. New immigrants moving to NYC for job
2. Students coming for study in NYC or Jersey City colleges
3. Local want to move to better neighborhood in the city
4. Even resident of NYC looking for cheaper areas for longer stays. One market study shows the high price of NYC real estate drives those who want to have families to the suburbs or cheaper areas like the Jersey City housing market.
5. Real Estate Investor who want to invest in potential growing areas and gain from rising rental prices.

2. Data

The data to be used for this project comes from four different locations:

- Foursquare. It is a local search-and-discovery service which provides information on different types of entertainment, drinking and dining venues. Foursquare has an API that can be used to query their database and find information related to the venues, such as location, overall category, reviews and tips.

- New Jersey Neighborhood Names, Median Household income and Population - Available on https://statisticalatlas.com/neighborhood/New-Jersey/Jersey-City/, this is used to obtain the neighborhood names and socio-economic information of the city.

- Median Rent Price – This data was obtained from multiple websites. All except 4 neighborhoods, data was take from https://www.rentjungle.com/average-rent-in-jersey-city-rent-trends/

  Hackensack Riverfront, West Side and Liberty Park - https://www.padmapper.com

  Lincoln Park - https://www.trulia.com/

- Geographic coordinates of Neighborhoods - Data available on https://www.distancesto.com

Below are the details on how we will use each data source during this project.

2.1. Foursquare API data

For this project we will use the Foursquare Places API. One of the features of this API is to provide a list of venues within a specific location, based on the Lat/Lon coordinates and a radius. In order to obtain a list of venues within a specified area, we use the "explore" endpoint from the API. By passing the proper parameters via an HTTP request to the explore endpoint, we get a JSON object with the information shown in the table below:

Field	Description
id	A unique string identifier for this venue.
name	The best known name for this venue.
location	An object containing none, some, or all of `address` (street address), `crossStreet`, `city`, `state`, `postalCode`, `country`, `lat`, `lng`, and `distance`. All fields are strings, except for `lat`, `lng`, and `distance`. Distance is measured in meters. Some venues have their locations intentionally hidden for privacy reasons (such as private residences). If this is the case, the parameter `isFuzzed` will be set to true, and the `lat`/`lng` parameters will have reduced precision.
categories	An array, possibly empty, of categories that have been applied to this venue. One of the categories will have a `primary` field indicating that it is the primary category for the venue. For the complete category tree, see categories.

Figure 1. Information contained in response to request towards "explore" endpoint

The *location* object contains the coordinates of each venue, which will be used to associate it with its respective neighborhood.

The *categories* array will be used to categorize the neighborhood. Basically, we will count how many venues from all available categories are found on each neighborhood, and then use that information to compare neighborhoods in Jersey City.

.

2.2. Jersey City Neighborhoods

Jersey City Neighborhoods data is not available in one dataset on any website. The data is taken from several websites and stored in a CSV file. Neighborhoods names are taken from https://statisticalatlas.com. Latitude and Longitude of Neighborhoods are manually retrieved from https://www.distancesto.com.

Latitude and Longitude will be used to do geographic visualizations of Jersey City Neighborhoods using Folium library. The map will be superimposed with Venues, Socioeconomic information like Median Household Income, Median Rent Price and population etc.

2.3. Median Household Income, Median Rent Price and Population

The neighborhood CSV file is assorted with household income, rent price and population from different data sources as described in the section above. This data will be used to cluster neighborhoods with similar economical strata, household density and amenities. We will create choropleth map with Household Income, Rent Price and superimpose it with cluster and venue information. The map will provide holistic view

and enable you to easily locate the city area resonating with your search for apartment.

3. Methodology
4. Results
5. Discussion
6. Conclusion