Discover Neighborhood with Your Favorite Restaurant

Introduction

Often time, you get tired of the restaurants in your neighborhood. You want to try other restaurants, but you don't have a certain restaurant to go to. Where should you start? In this report, neighborhoods in Jersey City were clustered based on the restaurant-style, this can give you a guide on which neighborhoods to go that most fit your style. Don't hesitate and just follow the guidance.

Data Requirement and Collection

The development of the work is based on two datasets. The first dataset is the information of all neighborhoods in Jersey City including postal code, name, latitude, and longitude. This dataset will be generated by scraping from Wikipedia. The second dataset is about the number of restaurants with different styles in the neighborhood. This dataset is mainly used for clustering. It will be generated using the Foursquare API.

Scrape information of Jersey City neighborhood from Wikipedia

The information we are scraping is specifically name of each neighborhood in the Jersey City. BeautifulSoup is used to scrape the information. The scraped information is organized in a dataframe with eventually three attributes: Neighborhood, Latitude and Longitude.

Add longitude and latitude of each neighborhood

Geopy package is used to obtain the longitude and latitude of each neighborhood. the data was then added to the dataframe df. Neighborhoods whose location were not found were directly removed from the dataframe.

Visualize neighborhoods on the map

The neighborhood is visualized on the map by using folium package. This provides us with a direct sense of where those neighborhoods locate. Neighborhood that are actually not in Jersey City is removed and also neighborhoods that actually share the same coordinates are joined.

A picture containing text, map

Description automatically generated

Figure 1. Map showing the neighborhoods

There are some neighborhoods that are apparently not in the jersey city should be removed, so a new attribute is added to the databased which is ‘Distance’. Specifically, it is the distance between the neighborhood and center of jersey city. The neighborhoods which are too far from Jersey City is removed, also it can be used to discover neighborhoods with the same location.

A screenshot of a cell phone

Description automatically generated

Table 1. Head of dataframe that shows neighborhoods in Jersey City.

All qualified neighborhoods are displayed in the map again.

A close up of a map

Description automatically generated

Figure 2. Map with all qualified neighborhoods

Get nearby restaurants from Foursquare

Foursquare API is used to get nearby restaurants, including their name, coordinates and categories. Explore endpoint is used and explore radius is set to be 300 meters. A new dataframe is generated to store the venues data.

A screenshot of a cell phone

Description automatically generated

Table 2. Head of dataframe that shows all venues

Cluster the neighborhoods based on restaurant type

Clustering is conducted based on the category attributes. Categorical attributes are transformed to numerical attributes using one-hot method for clustering. The optimal cluster number is found. To better visualize the result, the top 5 most common restaurant types are organized into a new dataframe with cluster label.

A screenshot of a cell phone

Description automatically generated

Figure 3. Plot of number of clusters vs. sum of squared distance from cluster center

There is an elbow at 8 which is not very clear, anyhow we will use 8 as the number of clusters.

K-means clustering method was used and the cluster label is integrated within table 1 plus top 5 most common categories for each neighborhood.

A screenshot of a cell phone

Description automatically generated

Table 3. head of dataframe with clusters and top 5 categories

Result and discussion

In this section, the clustered neighborhoods were displayed again in the map and examined, besides the food venues were further explored for each cluster.

A picture containing text, map

Description automatically generated

Figure 4. Map with clustered neighborhoods

A screenshot of a cell phone

Description automatically generated

Table 4. Sorted dataframe based on clusters.

As shown in the map and dataframe, cluster 1 and 7 have more than 1 neighborhood in the cluster, 6 and 5 respectively. All others only have one neighborhood. Besides, neighborhoods in cluster 1 are all located near the river in the downtown, which is also the most flourishing district. cluster 7 locates in the western area. Both neighborhoods in cluster 1 and cluster 7 are close to each other. For the other clusters who only has one neighborhood. they all located around the edge of Jersey City and each neighborhood is very far from each other. As for the top common food categories, restaurants in cluster 7 are more in aisan style including chinese restaurants, japanese restaurants and thai restaurants. Restaurants in cluster 1 have more food truck and American style.

Conclusion

This study is to explore and cluster the neighborhoods in Jersey City based on the restaurant categories. It can help people find the correct neighborhood for restaurants that best fit their tastes. There are 2 major conclusions:

1. Neighborhoods that are close to each other have similiar restaurant styles.
2. People who are interested in American food should go to neighborhoods near downtown, and people who are interested in asian food should go to western area of Jersey City.