

Segmenting and Clustering Neighborhoods in Houston, Texas

IBM Applied Data Science Capstone



Introduction

Background

Houston is the most populous city in the U.S. state of Texas, fourth-most populous city in the United States, and sixth-most populous city in North America. With a high rate of growth, a lot of workers are moving to Houston to take advantage of employment opportunities there. As a cosmopolitan destination, Houston is also filled with world-class dining, arts, hotels, shopping and nightlife. To encourage business investments, attract talents and benefit citizens, open datasets provided by the city of Houston are available online for everyone to examine and develop. Thus, I conducted this location analysis to explore, analyze and cluster neighborhoods in the city of Houston to provide some ideas for new residents and new businesses.

Data Source

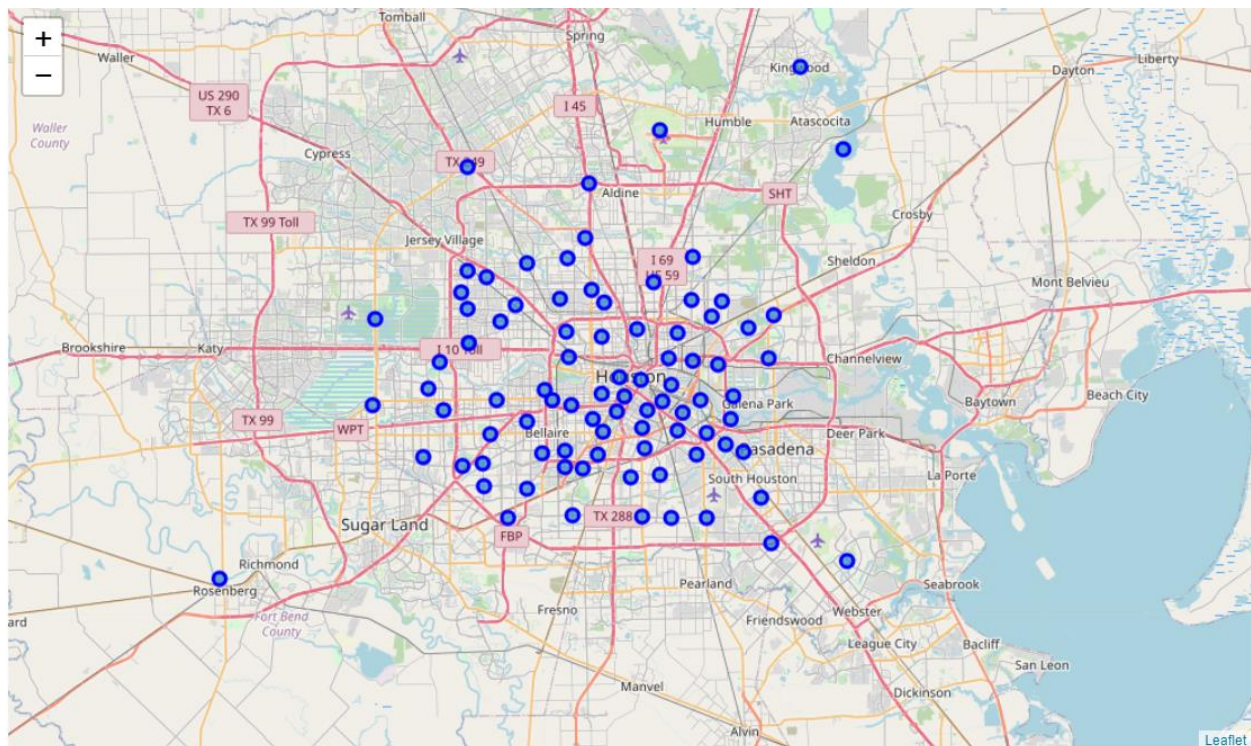
All the data that will be used in my analysis are shown below. The first dataset on Wikipedia includes neighborhoods and location information for the city of Houston. The second dataset on Kaggle lists latitude and longitude information for the Houston neighborhoods. Last, I will use the Foursquare API to explore venues and cluster neighborhoods of Houston. See more information on my Github: <https://github.com/AbigailYuHaixin>.

1. List of Houston
Neighborhoods: https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods
2. Latitude and Longitude of Houston
Neighborhoods: <https://www.kaggle.com/mrchristolpher/houston-texas-neighborhoods-lat-long-list>
3. Foursquare Developers Access to Location
Data: <https://developer.foursquare.com/developer/>

Analysis and Results

Data Preprocessing

All analyses were conducted using Python version 3.8.5. A total of 88 neighborhoods in Houston were included. The original dataset from Wikipedia has 3 variables. The original dataset from Kaggle has 6 variables. I merged the two datasets on Neighborhood, then dropped Boundaries and Full Address. Thus, there are 88 observations of 6 variables in our dataset. The six variables are: Neighborhood, Location, City, State, Latitude, Longitude. The eighty-eight neighborhoods are shown in the map below.

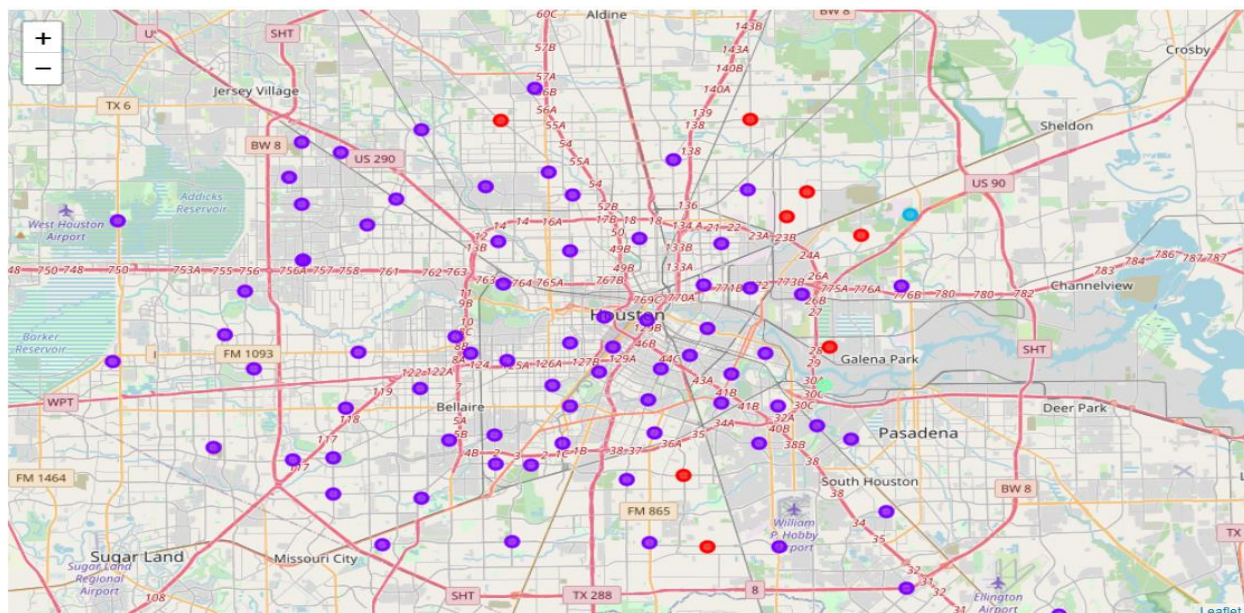


Explore and Segment Neighborhoods Using Foursquare API

Based on the location information provided in our dataset, I used the Foursquare API to explore and segment the neighborhoods in Houston. My explorations showed that the number of venue category in each neighborhood within a radius of 1500 meters (less than 1 mile) varied from 2 to 100.

Minnetex has the fewest variety of venues: Discount Store and IT Services. There are 15 neighborhoods have the largest number of venue categories: Afton Oaks/River Oaks, Downtown, Fourth Ward, Greater Heights, Greater Uptown, Greenway/Upper Kirby, Medical Center, Mid-West, Midtown, Museum Park, Neartown/Montrose, Spring Branch East, Spring Branch West, University Place, and Willowbrook. Of all the returned venues, there is a total of 325 unique categories. To know each neighborhood more deeply, I checked the frequency of occurrence for each category by neighborhood. I also put the top 10 venues for each neighborhood into a new dataframe. For instance, the top 10 venues for the neighborhood Acres Home are Construction & Landscaping, Food, Chinese Restaurant, Discount Store, Locksmith, Bus Station, Dumpling Restaurant, Eastern European Restaurant, Electronics Store, and Empanada Restaurant.

K-means clustering was used to segment the neighborhoods into 5 clusters. By visualizing the resulting clusters in the map below, we can conclude that Houston is a well balanced city. Most of the neighborhoods are in the same cluster. The first cluster has 8 neighborhoods: Acres Home, Clinton Park/Tri-Community, East Houston, East Little York/Homestead, EL Dorado/Oates Prairie, Minnetex, Settegast, and South Park. All the eight neighborhoods have dry cleaners. Six of them have donut shops, construction & landscaping, and discount stores. Four have duty-free shops and dog run. And three have gas stations and doctor's offices. It all sounds like a great place to live. But they are relatively far from downtown Houston and scattered along the northeastern and southern edges of the city. The second cluster includes 90% of the neighborhoods in Houston. Of these 77 neighborhoods, more than half have Mexican Restaurants and fast food restaurants. Other common venues are sandwich place, pizza place, fried chicken joint, and coffee shops. This indicates that Houston is a fast-paced but very convenient city. And because Texas is very close to Mexico, there are a lot of Mexican restaurants in Houston. The third cluster, fourth cluster, and fifth cluster all have one neighborhood. They are Hunterwood, Harrisburg/Manchester and Lake Houston. Hunterwood has auto shop, business service, and home service. It is like a traditional business area. Harrisburg/Manchester has shops, restaurants, zoo exhibit, and marine terminal. Obviously, it is near the sea and must be a fun place to go. Lake Houston also has shops, restaurants, and zoo exhibit. And it has a flea market.



Conclusion

I conducted this location analysis to explore, analyze and cluster neighborhoods in Houston to provide ideas for new residents and new businesses. Overall, Houston is a well developed city. Most of the neighborhoods are in the same cluster. If you are a young resident, cluster 2 probably has everything you want. If you are a middle-aged resident with child or an elderly person, then cluster 1 may be a good choice. You can have a relatively slow pace of life, but all your daily needs are around, such as dry cleaners, discount stores, doctor's offices, and gas stations. But if you depend on the sea for your life, the other three would be good. Especially for fishery industry, Lake Houston is an excellent place. Then for city planners, figuring out how to customize the development of these neighborhoods is your top priority. Diversity can bring more investments and talents to a city. But the good news is any neighborhood in Houston won't be too bad for businesses. And I think that's one of the reasons a lot of companies are moving to Houston and Texas.