Houston Neighborhoods for Coffee Lovers

IBM Data Science
Capstone Project

Polina Kaz

Introduction

Houston is the fourth-populous and one of the largest cities in the United States. This city is culturarly diverce, and year after year Houston is included into top ratings of the best food cities in America. Breakfast and brunch spots are popular nowadays, but is your breakfast considered to be complete without a nice cup of coffee?

The coffee market in Houston is growing tremendously. Houston has top-rated local coffee roasters that provide nicely roasted coffee beans to the local cafes and nationwide. Using machine learning techniques and data science methodology, the project objective is to evaluate top-rated coffee shops within the Houston City Limits, choose the best locations for opening a new local coffee venue that will meet both owner and customers' expectations.

Problem

Houston has 88 neighborhoods with more than 2.3 million people living within the City Limits. More than 5000 businesses and retailersare operating within the area. The current project's aim is to find the best developing neighborhoods that are going to be ready to accept the new coffee joints into their hoods.

Who will need this?

This project will help business owners or future startapers, or just those who can't live without a nice cup of coffee, to focus on the particular locations to establish the new location of grinding, brewing, and enjoying.

Data

To solve this problem the following data is being used:

- Neighborhood information within the Houston City Limits;
- Latitude and longitude coordinates of the neighborhoods in order to get the venue data;
- Foursquare API with coffee venues.

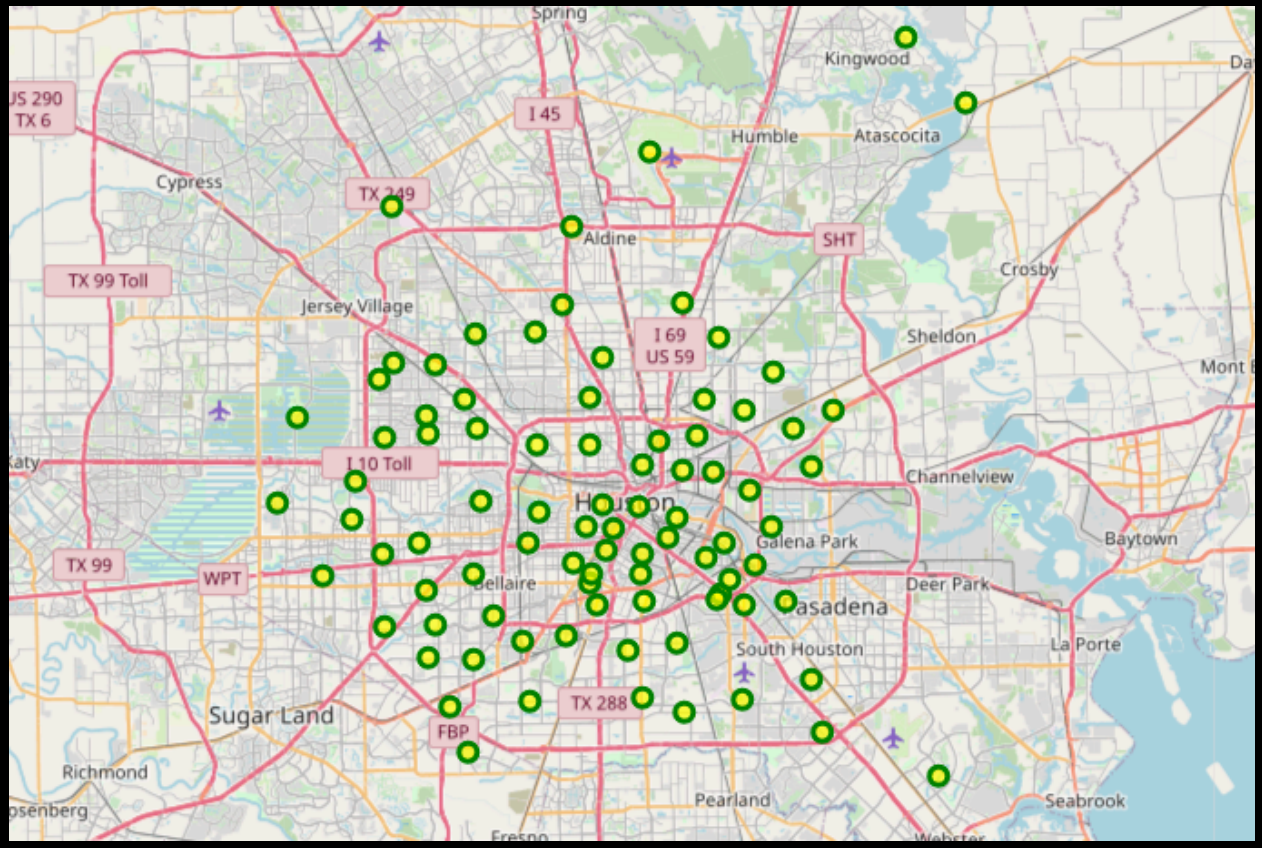
Methodology

The list of Houston neigborhoods was scraped from the Wikipedia page, the table is transformed into a *pandas* dataframe. Using *geocode* the longitude and latitude is found for each neighborhood and added to the neighborhood table. The information on the local coffee shops for each neighborhood is collected using the *Foursquare API* as well as their presence in each neighborhood and frequency.

Using the merged dataset, the data will be clustered and mapped. Each cluster will be examined in order to determine the best developing and not overcaffeniated area to open the new coffee shop.

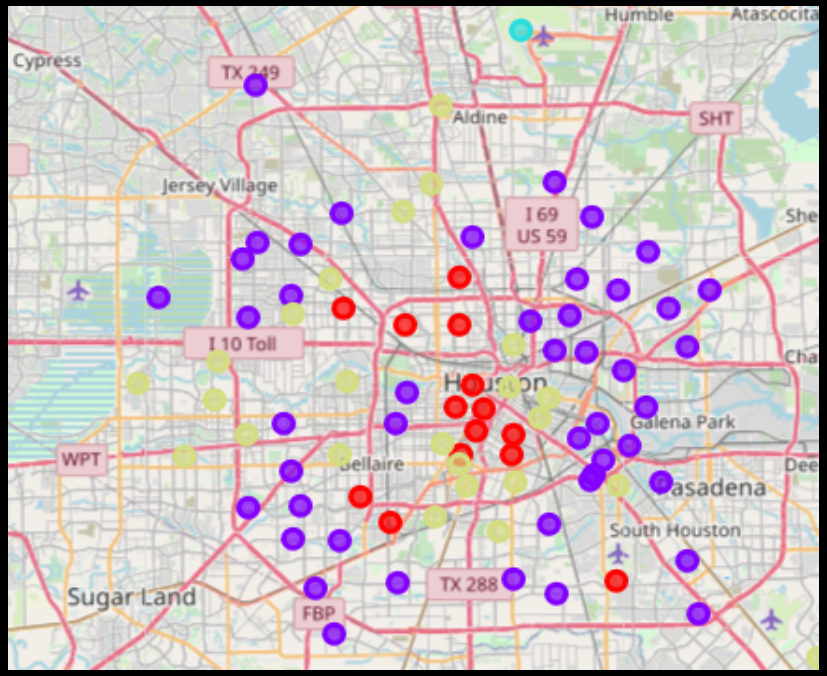
Results

Houston with all 88 neighborhoods gathered in one map. There is a total of 169 coffee shops within the area. What an opportunity to add one more coffee joint to get an even number!



Results (continued)

Here's a look on clustered neighborhoods by the frequency of coffee joints:

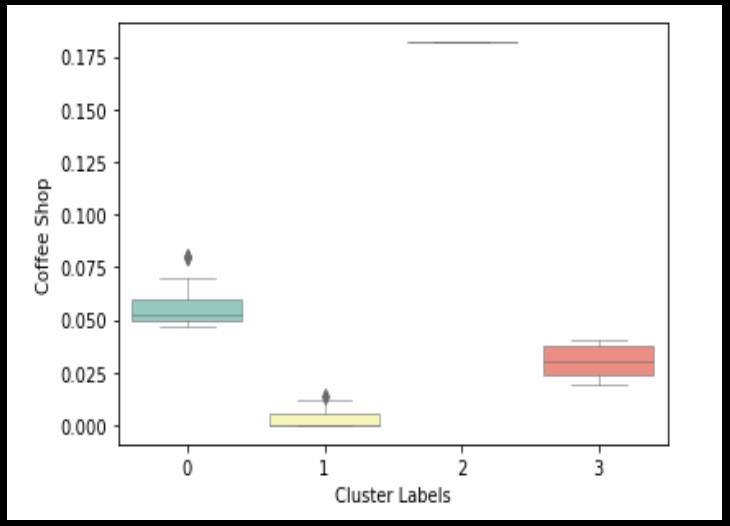


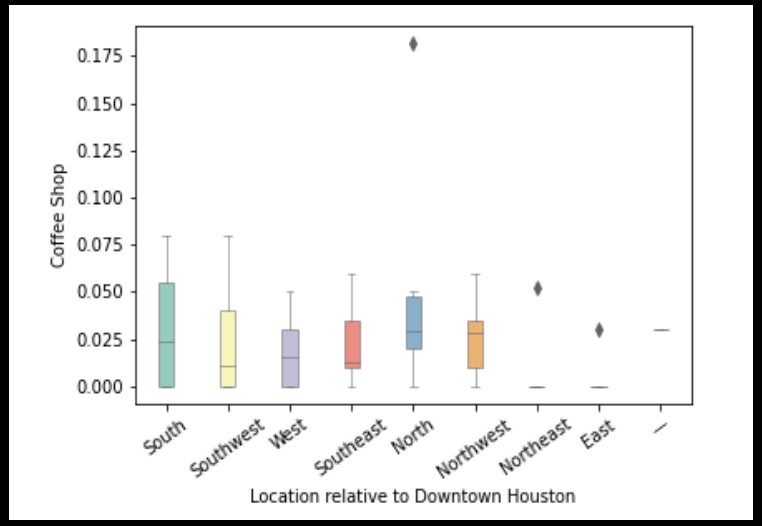
The neigborhoods were clustered into 4 groups of frequency of coffee shops:

- (•) Cluster #0: **high**
- (•) Cluster #1: **low**
- (•) Cluster #2: International Airport Houston.
- (•) Cluster #4: **moderate**

Results (continued)

These two boxplots show the frequency of coffee joint by clustered neighborhoods and more large areas relative to Downtown in Houston, TX.





Discussion

(•) Cluster #1. Has very low appearance of coffee shops. Neighborhoods from (•) Cluster #0 have the highest frequency of coffee shops in the area comparing to all others. It seems like Central and Northwest Houston drink coffee more than anyone in the city. (•) Cluster #4 had moderate frequency of coffee joints. West area needs definitely needs more coffee!

Conclusion

The goal of this project was to analyze the neighborhoods within the city limits of Houston, TX in order to determine the best location to start a new coffee business.

The neighborhood data was collected from Wikipedia page and converted to *pandas* dataframe. Then by using *geocode* the latitude and longitude data waa collected and added to the Houston neighborhood dataframe. To find all the coffee locations, the *Foursquare API* was used. All the datasets were merged and clustered for further analysis.

The overall conclusion states that Houston has relatively low frequency of coffee shops with the promising market sizes, which allow us to assume that new coffee joints might be successfull within the area.

This project was made using the basic *Foursquare API Developer Account* with limited numbers of API calls which results in lack-of-depth analysis. For more accurate results of chosing the right place to start a new coffee business more information should be taken into account such as population density, income level, and other demographic information.

Thank you!