

ANALYZING HOUSTON NEIGHBORHOOD FOR RESTAURANT BUSINESS OPPORTUNITY

By Prashant Khatri for Data Science Project

BACKGROUND



Houston is a large city with 637.4 square miles.



It has 88 super neighborhoods and has diverse population distribution.



It is also an energy hub of US known mainly for oil and gas industries.



Plus, many tourist destination centers are present in Houston.



All of these makes Houston a perfect place to open a new restaurant.

PROBLEM STATEMENT

- Finding neighborhoods that are suitable for opening a restaurant based on the presence of other venues.
- Exploring the type of cuisine (restaurant) that has a high chance of getting success in those areas.

DATA DESCRIPTION AND SOURCE

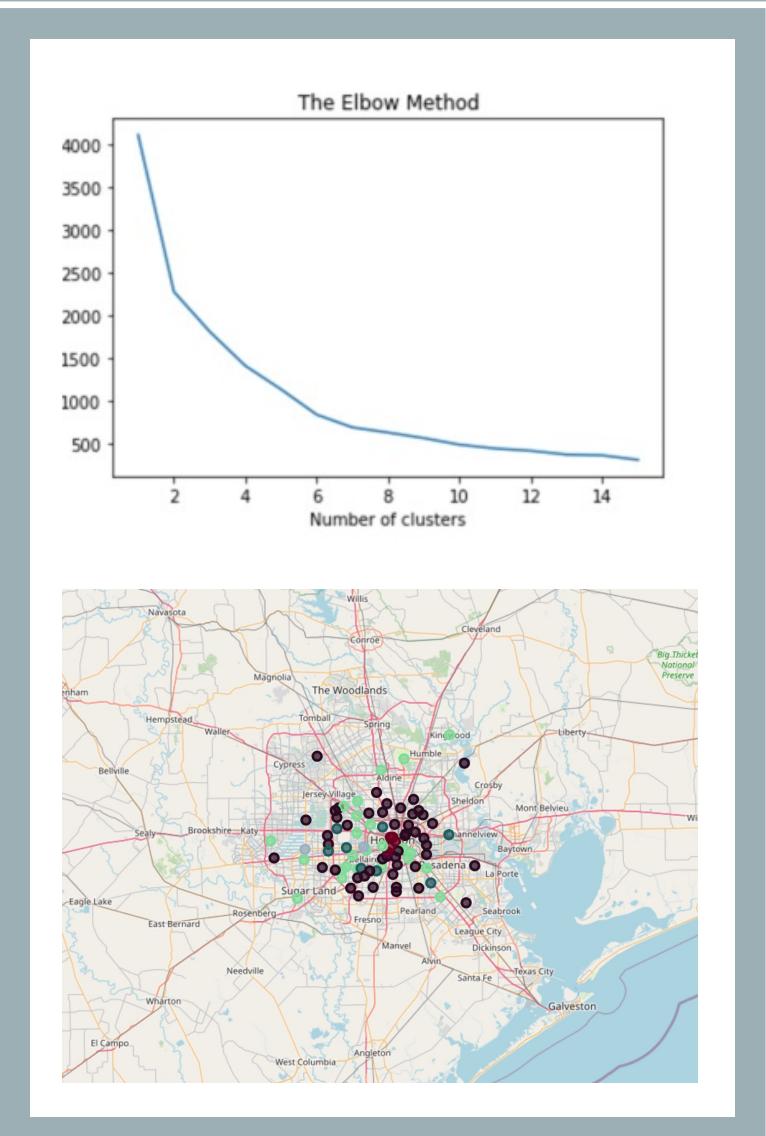
- Obtain names of neighborhood in Houston using Wikipedia.
- Acquire latitude and longitude of neighborhood using GeoPy geocoding. Nominatim is the preferred GeoPy package for this project.
- Manually input latitude and longitude for neighborhoods for which GeoPy could not get the correct location.
- Next find out the most common venues in every neighborhood using Foursquare API.

METHODOLOGY

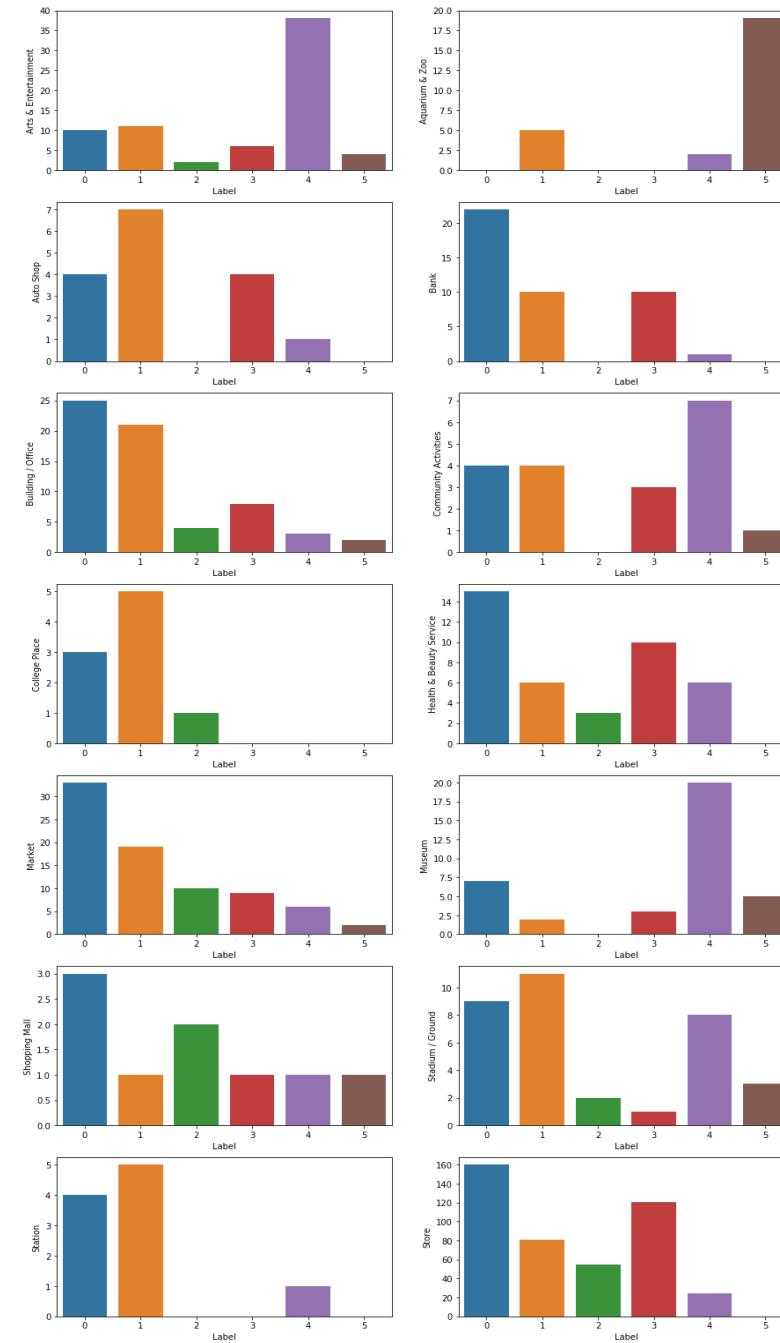
- Web Scraping – to get Houston neighborhoods name from Wikipedia.
- Nominatim GeoPy package and google search – to get geographical coordinates of Houston neighborhoods.
- FourSquare API - to get the information about venues in the neighborhoods.
- One Hot Encoding – performed on venues categories.
- K-Means Clustering – to divide neighborhoods into cluster based on similar characteristics.
- Exploratory Data Analysis – to analyze neighborhood and restaurant type in visual format.

RESULTS

- K-Means indicated neighborhood should be divided into 6 clusters.
- Neighborhoods that belongs to cluster 0 (light green as shown in figure) have many popular venues / sites that can bring more people in neighborhoods.

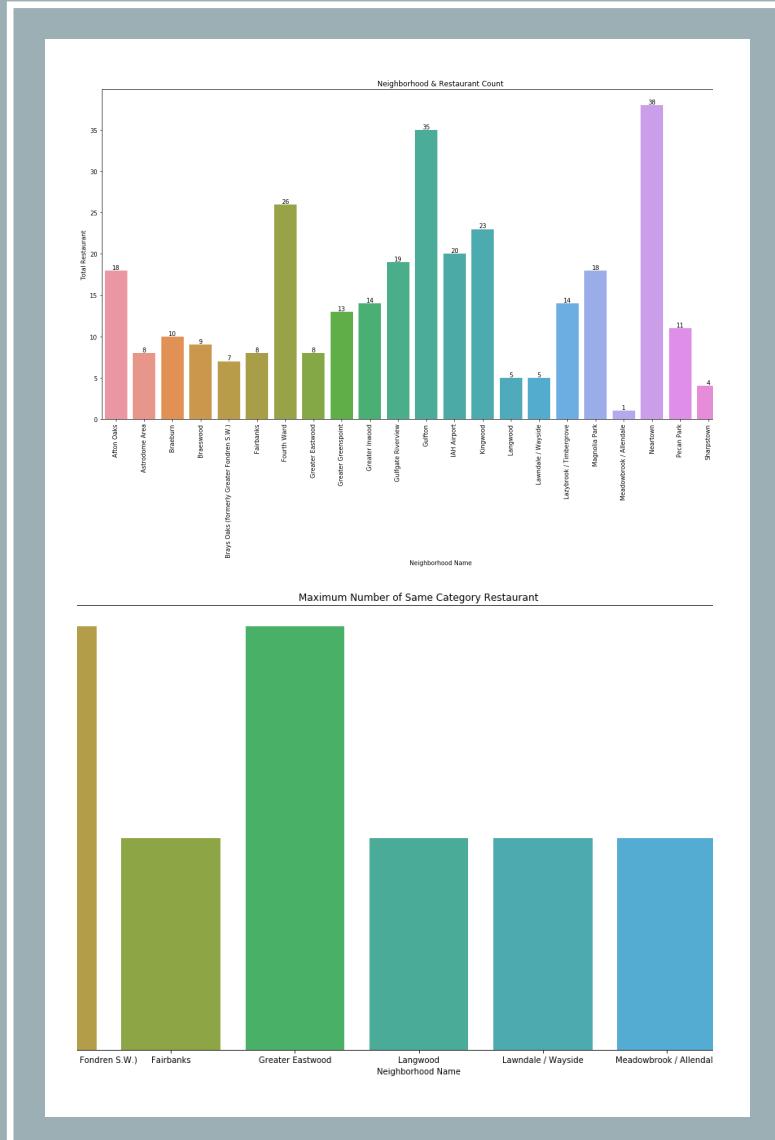


RESULTS (CONT)



RESULTS (CONT)

- 11 neighborhoods in cluster 0 have restaurants less than 10 and hence are probably the best location to open a new restaurant.
- The name of 11 neighborhoods are Astrodome Area, Braeswood, Brays Oaks (formerly Greater Fondren S.W.), Fairbanks, Greater Eastwood, Langwood, Lawndale / Wayside, Meadowbrook / Allendale, Sharpstown, South Belt / Ellington, and Spring Branch Central.
- Further analysis showed that every kind of cuisine should be able to do well in those locations.



DISCUSSION

- This study showed that 11 neighborhoods in Houston have a high chance for any new restaurant to do a good business.
- This study has certain limitations - the result is only based on selecting some popular venues categories in neighborhoods and population density is not taken into consideration while analyzing the data.
- A future study including population density in neighborhood along with other attributes might provide better prediction.

CONCLUSION

- Described the business problem for solving it using data.
- Performed necessary steps such as data pulling, data pre-processing, exploratory data analysis and utilizing machine learning techniques to find the answer.
- Answer to the business problem - 11 neighborhoods in cluster 0 seems to be a good location for opening any kind of restaurant (cuisine).
- A future study including neighborhood population density could provide even better prediction.

REFERENCES

- List of Houston neighborhoods. (2020, August 07). Retrieved August 23, 2020, from https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods
- FourSquare. (n.d.). Retrieved August 23, 2020, from <https://foursquare.com/about>
- One-hot encoding in Python. (n.d.). Retrieved August 23, 2020, from <https://www.educative.io/edpresso/one-hot-encoding-in-python>
- Pulkit Sharma. (2020, April 23). The Most Comprehensive Guide to K-Means Clustering You'll Ever Need. Retrieved August 23, 2020, from <https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/>