Coursera Capstone Project:

Optimal Houston Districts for Restaurant Entrepreneurship

Nicholas Isakson

July 29, 2020

**Introduction:**

The purpose of this is to determine and recommend which district in Houston is most favorable for opening a restaurant. This requires the use of Foursquare API to determine competitors in each area of the city, as well as the success of other similar businesses. An area with many competitors is less favorable for opening a restaurant, while an area where restaurants are very popular is more favorable. Thus, the most ideal district will be the district where there are few restaurants, but the restaurants that are present are very popular.

The target audience for this project is restaurant entrepreneurs. If someone is looking to open a restaurant in Houston, this project will be helpful to them for deciding where the most optimal place to set their business up would be. They would care about this problem because most small businesses are unsuccessful. Having an advantage, knowing where a service is most likely to succeed will increase the probability of success for their restaurant.

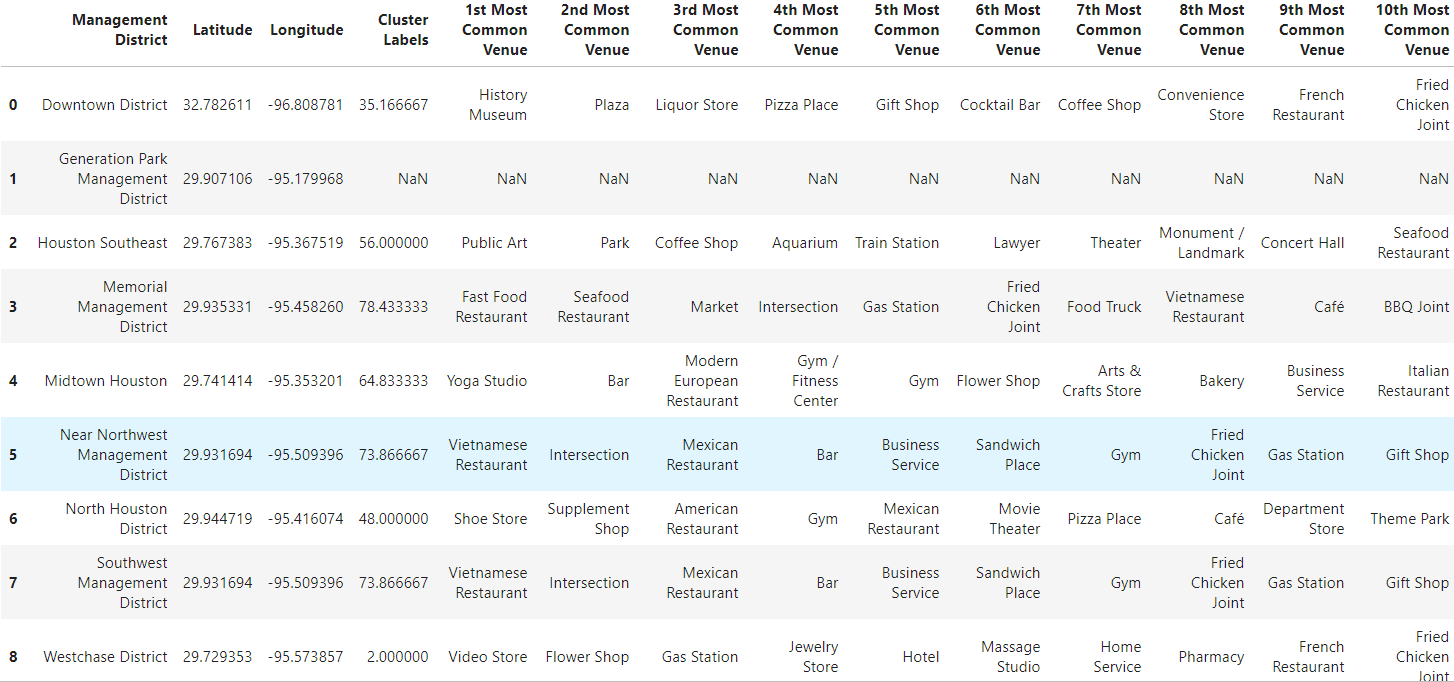
**Data:**

The data that will be used for this project will be a combination of the Foursquare API geolocator, venue finder, and wikipedia article containing a list of districts in Houston. The wikipedia article will be accessed first to find all the districts in Houston. Next, the location of each district will be determined using Python’s geolocation functionality. Finally, the venues near each district will be determined using the venue finding functions used in the previous lab. The top ten venues near each district will be recorded for analysis of a particular district. This is useful for analysis because the number of competitors and the popularity of competitors can be quantified from this data set.

The wikipedia article that will be used to determine the districts of Houston is: <https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods>. The table being used for this project is the lower table containing Houston management districts. The main problem with this dataset is that it includes data which is not necessary for this project. Because of this, the encompassed neighborhoods and boundaries columns are removed before further analysis.

When searching for the latitude and longitude of each district using a geolocator, it was found that some of the districts did not have positional data available. As a result, these districts were removed from consideration and were taken out of the analyzed data set. For a future project, one could determine accurate positional data for these districts and do the same venue analysis on them.

When generating data for each venue, more data than was needed is generated. The only relevant column is the venue category. This is because it is restaurant competitors that are being searched for. Venue category allows the determination of whether a particular venue could be a competitor to a restaurant or not. In the end, the top ten venues for each district are shown below in table 1.



**Table 1:** Top 10 most common venues for each district in Houston. Note that Generation Park Management District does not include venue data. It is taken out for the final analysis. The cluster labels column contains the restaurant score for each district. This is explained more in the methodology section.

**Methodology:**

The first thing necessary in this project was the determination of a city where the data necessary exists. For this project to be possible, city districts, district positional data, and venue lists must be available. The first city considered was Austin, however this was removed from consideration due to city district data not being tabulated in an accessible format. Next, Houston was considered which fit all the necessary criteria.

Once the required data was retrieved, data analysis was able to be performed on the data. The main process necessary for this was a quantitative method of determining a districts ‘restaurant score’. This score needed to be a measure of the favorability of opening a restaurant in a district. A district with a greater restaurant score is more favorable for opening a restaurant while a district with a lower restaurant score is less favorable. The formula for this determination needed to be decreased if a district had many restaurants, and increased proportionally to the popularity of restaurants which do exist in the district. First, the number of restaurants in the top 10 venues in a district was counted for each district. This was determined by searching each venue type for words related to restaurants or food preparation. These could be summed together to determine the number of competitors in a district. Then, a popularity score was developed using the following formula:

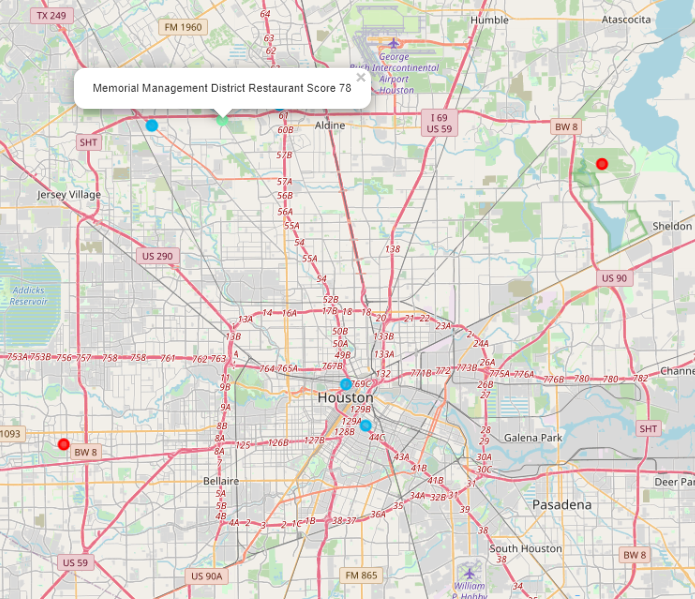
This equation summed the popularity ranking of each competitor in the district. For example, if a district included two restaurants, one being the fourth most popular venue in a district and one being the seventh most popular, its number of restaurants would be 2 and its district restaurant popularity would be 11. These values would then be used to generate a district restaurant score:

Using the example from before, with a district restaurant popularity of 11 and a number of restaurants of 2, the restaurant score would be 5.5. This formula satisfied the necessary requirements of a restaurant score. If the number of competitor restaurants increases in a district, its restaurant score goes down, but if the restaurants that do exist are more popular, then the score will increase. These calculations were repeated for each district to generate a list of scores.

After the restaurant scores were generated for each district, the most favorable district was determined by finding the district with the greatest restaurant score. The scores were then displayed on a color coded map, with brighter colors representing more favorable districts.

**Results:**

The restaurant scores for each district are shown above in table 1. From this data, it was found that the district with the greatest restaurant score was Memorial Management District. This district had a restaurant score of 78. This was only slightly greater than the Near Northwest Management District and Southwest Management district. The difference between restaurant scores between these areas is less than 10, so all three districts are favorable for opening a restaurant in Houston.



**Figure 1:** Map of Houston districts with restaurant scores color coded. Brighter colors are more favorable for opening a new restaurant. This map is also included in the Python notebook in an interactable form.

The map shown above was generated using the folium functionality of Python. This map shows each district of Houston with a color associated with its restaurant score. Brighter dots represent greater restaurant scores and darker, or red dots represent poor restaurant scores. This map supports the conclusion that Memorial Management District is the most favorable district in Houston for opening a restaurant.

**Discussion:**

Based on the results, it seems clear that the Memorial Management District is the best district to open a new restaurant. The nearby area, in Northwest Houston also seems favorable, with both adjacent districts having restaurant scores above 70. However, it is possible for other districts in Houston to be more favorable. There were several districts that were omitted due to lack of positional or venue data. These districts may be more favorable for restaurants, but more data needs to be tabulated and analyzed to determine if this is the case. For a future project, one could find positional and venue data for each of these districts and perform the same analysis that was done in this project to present district recommendations for the entire city. Another observation that should be noted is that determining the number of competitors was done by a search on venue type. It is possible that venues that are competitors which do not contain these words were omitted from analysis.

**Conclusion:**

Memorial Management District in Houston Texas is the most favorable district for opening a new restaurant. That is the recommendation supported by the data of this project. While there were some sources of error, including omitting districts and a possible disproportional method of calculating restaurant score, this project is fundamentally sound. As such, the trend of the results is expected to represent reality, with the Northwest region of Houston being the most favorable region for opening a new restaurant.