

Capstone Project:

Find the Most Optimal Area Zip Code for a Coffee Shop in Harris County

Caroline Li

May 29, 2020

- **Introduction: Business Problem**

This project is aim to recommend the neighborhoods that suit to open a coffee shop in Harris County in Houston, USA to the shop owner.

The location is better to be close to restaurants in case customers want to grab a coffee after meals. But the place should not be already crowded with other coffee shops including dessert shops. In general, we want to open the coffee shop in a place where we have the largest number of potential customers coming in throughout the year.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by shop owner.

- **Data**

Based on definition of our problem, factors that will influence our decision are:

- number of existing coffee shops in the neighborhood (any type of coffee shops including boba tea shops)
- number of and distance to the closeby restaurants in the neighborhood, if any
- number and type of venues in the neighborhood

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**
- number of existing coffee shops in the neighborhood in every neighborhood will be obtained using **Foursquare API**
- number of and distance to the closeby restaurants in the neighborhood will be obtained using **Foursquare API**

- coordinate of Houston center will be obtained using **Google Maps API geocoding** of a list of zipcodes from <http://zipatlas.com/us/tx/houston/zip-code-comparison/median-household-income.htm>

- **Methodology**

In this project we will direct our efforts on detecting areas of Houston that have restaurants in neighborhood, but low density of other coffee shop. We will limit our analysis to area ~500km around each zip code.

In first step I have collected the required data: location of all restaurants and cafe in each zip code neighborhood; location and type (category) of every venue within 350km from the center of each zip code.

Second step in our analysis will be filter out the areas without restaurant.

In third step, I will calculate and explore of 'coffee shop density' across different areas of Houston - we will use heatmaps to identify a few promising areas close to center with low number of coffee shops in general and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

- **Results and Discussion**

Our analysis shows that there are not many areas in Harris County, Houston that have both high density of restaurants and low density of cafes. Most of these areas are detected west from the Houston center. After we filtered out the areas with lowest density of restaurants within 350m from the area centers, we took a look at all the venues available in the districts. Those location candidates were then clustered into 4 groups to create zones of interest which contain greatest number of location candidates. Group 2 has the greatest number of location candidates and highest density of restaurants.

We recommend to choose from the areas with zip codes in group 2. There are 6 areas in group 2. 77010 has the greatest number of restaurants, but with the lowest number of population. Considering the most common venues, 77010 is probably in downtown area. 77063 and 77019 already have many coffee shops so they are not considered either. 77087 and 77027 do not have as many restaurants as 77098. At the same time, 77098 also has a large population. Overall, 77098 is the most recommended area to open a coffee shop in.

- **Conclusion**

Purpose of this project was to identify Harris County, Houston areas with high number of restaurants but low density of cafes in order to aid shop owner in narrowing down the search for optimal location for a new cafe. By calculating restaurant density distribution from Foursquare data we have first identified 10 zip code areas that justify further analysis, and then generated extensive collection of venues around those areas which satisfy some basic requirements regarding existing nearby restaurants. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations).

Final decision on optimal coffee shop location will be made by shop owner based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.