# Market analysis for a new café in Houston, TX

Joana Ázara

## 1. Introduction: The Business Idea

Discover the optimal location to open a new café in the city of Houston

### 1.1 Background of the entrepreneur

- A Brazilian that grow up in a country famous for its coffee (but is a tea lover!)
- An admirer of the art of baking, cooking and bringing a smile to people.
- A production engineer that has always loved cooking and knows that a successful business starts with good planning.

With the support of the love of her life (who has a brilliant business mind), it's time to put this planning of opening a café in motion!

A very important part of this business plan is the market analysis, for which I will use real data from Four Square and the US Census to support my hypothesis based on years of observation of cafés around the world. My hypothesis is laid out below.

### 1.2 Hypothesis

Although cafes are a common business and the industry has commoditized the rapid "caffeine fix" service, I see a trend toward a new generation of café. The new café must cater to coffee lovers' demand for a "coffee moment" through specialization with new flavors and brewing methods as well as offering customers variety like specialty teas.

I believe there is a potential market for this new generation café for which I must determine the best location for this business in the city of Houston.

So, let's start talking about coffee. Being born and bred in Brazil, this is not a new topic for me. You could dream of coffee at any time of the day and my grandma would have it ready and hot for you. But even in a country that had its history written by coffee trading, the behavior of coffee drinkers has been going through rapidtransformation. In the past if you wanted a coffee, there you had it, filtered black coffee. Okay, maybe not that simple. You could order a coffee with milk (Caffe au Lait) or a carioca (which is similar to an americano). It was not common knowledge that the taste from different coffee beans and roasts could bring such a diversity to the once commoditized drink.

Reality is different now. People are not only going for the coffee as a pure habit to get their caffeine dosage, but each day more and more people are learning to savor this drink. This behavior that is far from being restricted to Brazil. The moves of two big international chains – Starbucks and McDonald's – are proof of this transformation. Starbucks launched their blond roast and created the new brand called Starbucks Reserve that promises the rarest most extraordinary coffee Starbucks has to offer.

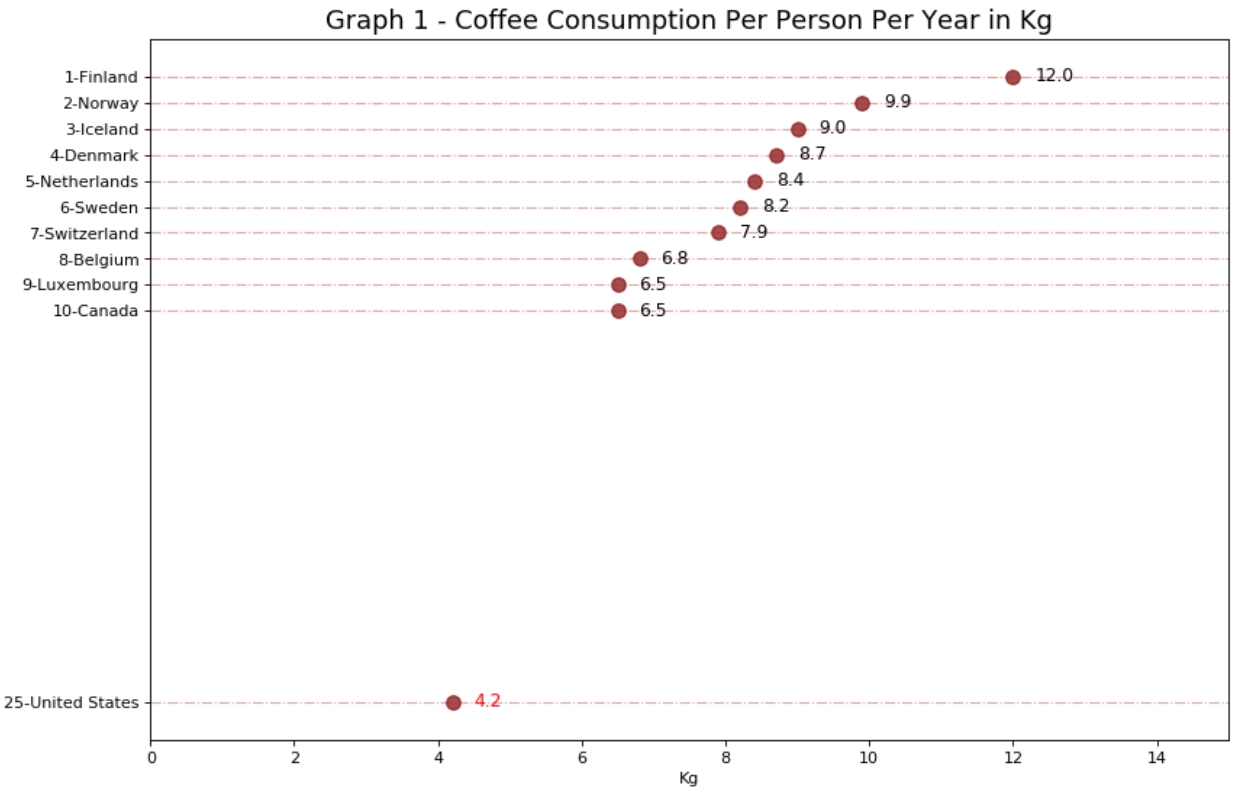
McDonald's, the seller of simplicity and low prices, realized it needed to step up and started introducing Barista Style coffees at some stores.

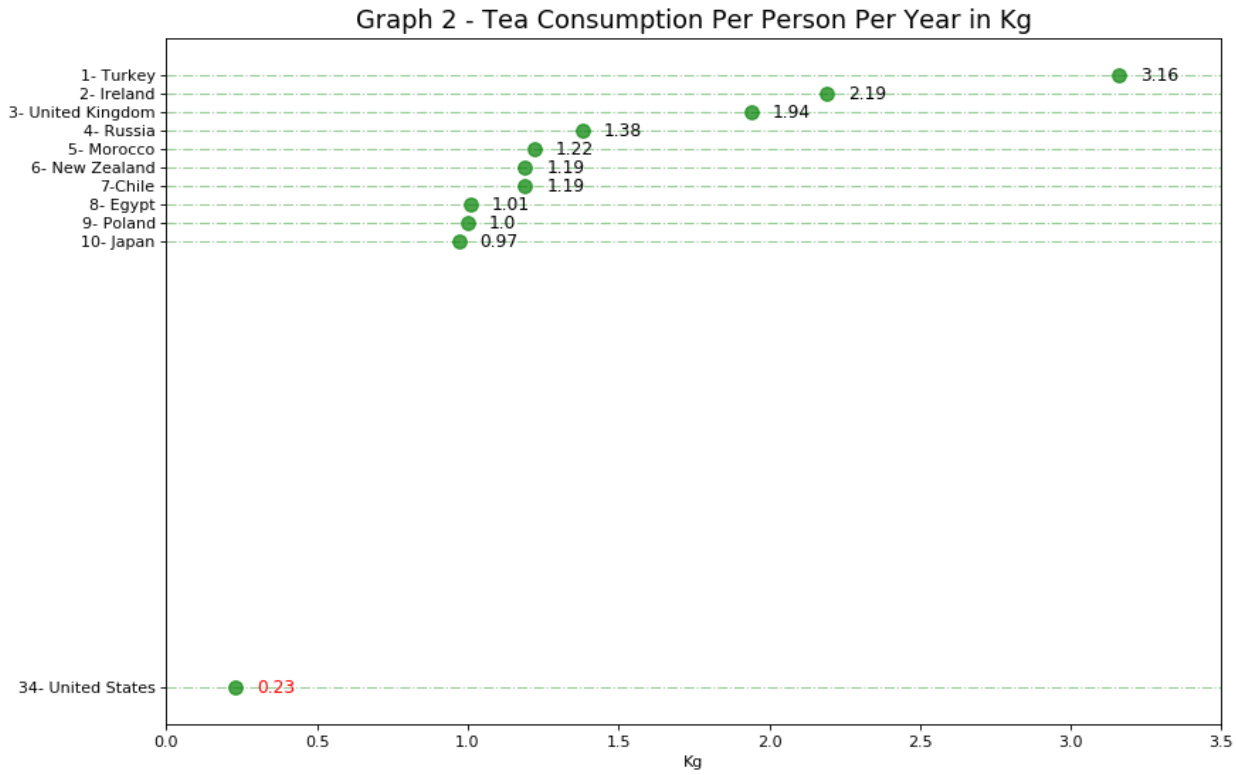
A similar wave is being observed at the tea market. According to the Food and Agriculture Organization (FAO) "Over the next decade, Western countries in general are expected to see lower consumption growth. In the UK, for instance, tea consumption is projected to decrease as black tea struggles to maintain consumers' interest amid increased competition from other beverages, including coffee." But the FAO's tea group believes tea companies can stall or even reverse these trends by diversifying into other segments, such as organic and specialty teas, and promoting health and wellbeing benefits.

The FAO also states, "The demand for tea has accelerated due to the ongoing retail revolution and the growing investment into tea education bringing new clientele to know more about tea, where it is sourced, the benefits of drinking tea, and how to properly brew it. Due to this, loose leaf tea has seen a new relevance in the US. Promoting tea culture-based market development and immersion in the cultural identity of societies across the world should be one of the strategies to sustain and expand consumption."

But even with the recent changes in consumption, the United States is far from being the top consumer of either coffee or tea. According to the World Atlas (Graph 1), the USA is only the 25th country in coffee consumption in the world, with a North American consuming an average of 4.2 kg per year, lagging behind Finish by 7.8 kg, and consuming 2.3 kg less than a Canadian. And when it comes to tea, the USA is only the 34th in the ranking of tea consumption (Graph 2).

This data indicates that there is room for growth of consumption with the right product and marketing strategies.






## 2. Data

In my data analysis I will be using the python programming language and sources of demographic and competitor data.

### 2.1 Population Data

The target market for the café are people between the ages of 25 to 44 with a minimum education of Associate's Degree and the data from the USA Census will be the key to answer which of the Houston neighborhoods the target consumers live.

The python package CensusData (https://jtleider.github.io/censusdata/api.html) will be used to gather data from the following tables of the American Community Survey (ACS) 5-year estimates (2013-2017):

1. B19326: Median Income in the Past 12 Months (in 2015 Inflation-Adjusted Dollars) by Sex by Work Experience in the Past 12 Months for the Population 15 Years and Over With Income
2. B15001: SEX BY AGE BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 18 YEARS AND OVER

The ACS-5 (2013-2017) is the most recent of the ACS that contains the data for the granularity needed for this analysis: by zip code.

### 2.2 Competitors' Data

The competitors' data will be gathered from the Foursquare, using its location data API.

### 2.3 Names and zipcodes of the main Neighborhoods in Houston

Names and zipcodes are going to be taken from the website http://www.mccannproperties.com/Houston-Neighborhood-ZipCodes.php, using BeautifulSoup to scrap the data from the website and converting into a Panda Dataframe. Those zipcodes will be used when getting the data from the American Community Survey.

### 2.4 Latitude and Longitude of each zipcode

In order to get the data of venues from FourSquare, latitude and longitude are needed. The package Geocoder will be used in order to get latitude and longitude for each of the neighborhoods.

## 3. Methodology

Below is a description of my data sources, how I will be formatting the data, and finally how I will analyze the data with the use of python.

### 3.1 Gathering Data

#### *3.1.1 Main neighborhoods of Houston*

The website http://www.mccannproperties.com/Houston-Neighborhood-ZipCodes.php was scraped using the BeautifulSoup package in order to get the data related to the Inner Loop and Close-in neighborhoods. After downloading the information into a *soup* object, the data was treated in order to get to a Panda DataFrame that consisted of one single zip code per line. Each zip code was then passed into the Geocoder package (geocoder.arcgis) in order to obtain their Latitude and Longitude center locations..

#### *3.1.2 Population data for the selected neighborhoods*

Using the package CensusData to access the data from the tables B19326 and B15001 from the ACS-5 (2013-2017) and feeding it with the zip codes obtained at topic **3.1.1**, it was possible to gather data of income and age profile of the population of Houston. The data from the zip code 77046 was not consistent, returning negative median income for the employed population, and because of that this neighborhood was dropped from the analysis.

### *3.1.3 Venues from Fousquare*

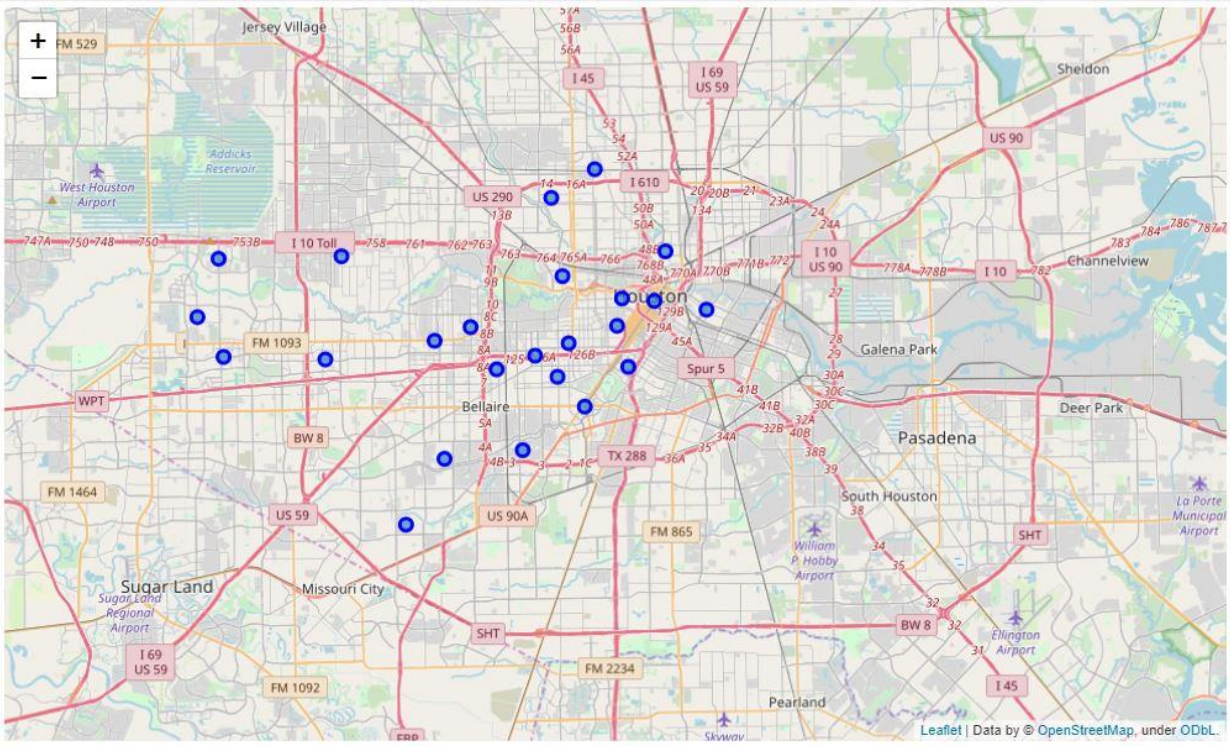
The latitude and longitude for the center of each neighborhood gathered at the topic **3.1.1** was passed into the Foursquare API to grab the data of the venues located in a 500 meter radius each of them. The API returned 818 different venues classified in 202 unique categories. The distribution of the venues per neighborhood can be seen at the chart below.

Those 202 unique categories of venues were simplified into the following 5 categories:

1) Cafe and Tea House: Direct competitors
2) High level of competition: Bakeries, cake shops and breakfast spots that constitute strong competitors to the proposed business
3) Medium level of competition: Restaurants and other food venues
4) Low level of competition: Bar and lounges
5) Promoters of foot traffic: retail spots, metro stations, offices and other non related food venues that can attract people to the region but does not constitute competitors

### *3.1.4 Consolidated dataframe*

The 3 dataframes built at topics **3.1.1**, **3.1.2** and **3.1.3** were consolidated in a single dataframe. A representation of the center of each of the neighborhoods in a map of Houston as well as a snapshot of the 5 first rows of the dataframe can be seen below.



| Neighborhood | zipcode | Latitude | Longitude | Median Income | Median Income - Employed | Population 18+ | Population 25-44 | Population 25-44 with Degree | Cafe and Tea House | High level of competition | Medium level of competition | Low level of competition | Promoters of foot traffic |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Montrose, Midtown | 77006 | 29.74 | -95.39 | 55950 | 142568 | 20194 | 10122 | 7835 | 4 | 3 | 24 | 14 | 2 |
| Midtown, | 77004 | 29.73 | -95.38 | 22499 | 103813 | 32793 | 11382 | 6323 | 4 | 1 | 12 | 4 | 16 |
| Downtown, | 77002 | 29.76 | -95.37 | 24254 | 107684 | 12203 | 6134 | 2081 | 6 | 3 | 54 | 11 | 24 |
| Memorial Park, Rice Military | 77007 | 29.77 | -95.42 | 73209 | 171943 | 31926 | 19434 | 17100 | 1 | 1 | 10 | 9 | 5 |
| Heights | 77008 | 29.81 | -95.42 | 54244 | 154166 | 28723 | 15403 | 11985 | 0 | 0 | 8 | 7 | 9 |

### 3.2 Analyzing the data

#### *3.2.1 Identifying the potential neighborhoods for opening the proposed business*

As described at the Data session, the target market for the proposed business lies in the population between 25 and 44 years old. Other factors analyzed that are important for the success of a business are: the purchasing powerof potential clients and attention from the public generated by foot traffic of non-competing businesses.
The median income was used as a reference for the power purchase. The number of venues located in the neighborhood but which business does not compete with a café were classified as promoters of foot traffic (session iii - Building a dataframe with venues from Fousquare) and will be used as a reference for the likelihood that the business will be exposed enough to public attention.

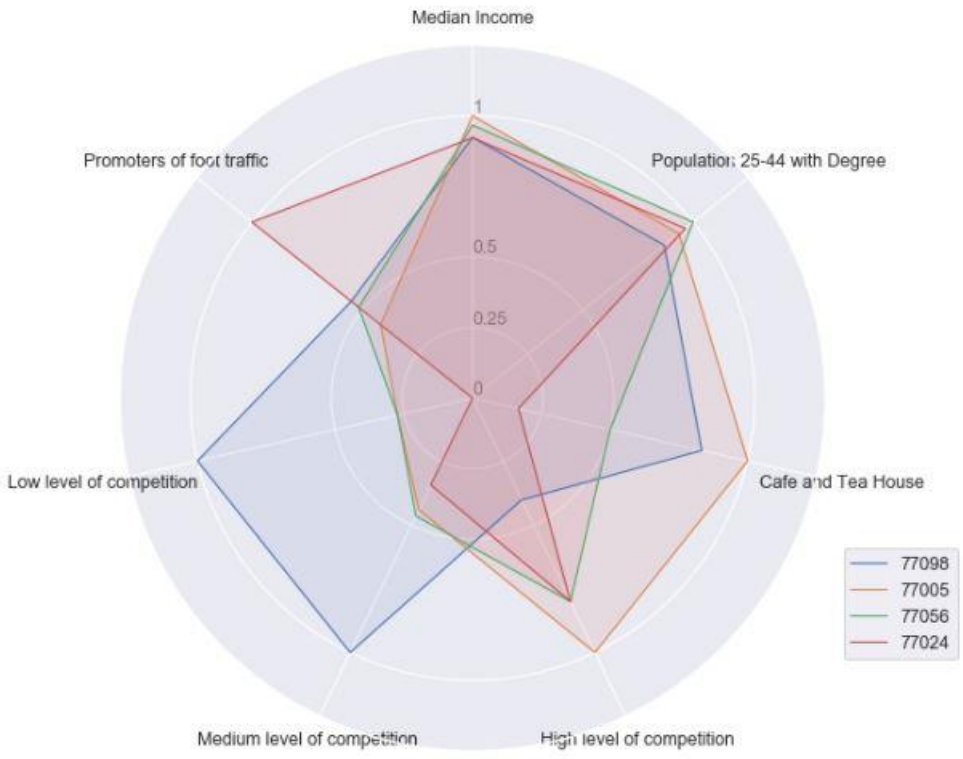
The following scatter plot was built in order to cross reference these 3 observations:
a) Axis x: Income
b) Axis y: Population with ages 24 to 45
c) Size of the bubbles: Number of venues that can potentially promote food traffic

From the chart it can be observed that the neighborhoods with zip code 77007 and 77008 have both high income and high population in the 25-44, but there are not a lot of venues in those zipcodes. This indicates that 77007 and 77008 are highly residential neighborhoods. Considering the characteristics of the Houston transportation system, in which most people commute by car, it was implied that a business in one of those areas would not be exposed to a lot of foot traffic.
The decision was then made to focus on the areas with high income and high concentration of potential promoters of foot traffic, which narrows down the analysis to the following areas:
a) 77005
b) 77024
c) 77056
d) 77098

### *3.2.2 Competition Analysis*

Now that the analysis was narrowed to 4 neighborhoods, let's verify how they compare to each for these 7 characteristics:

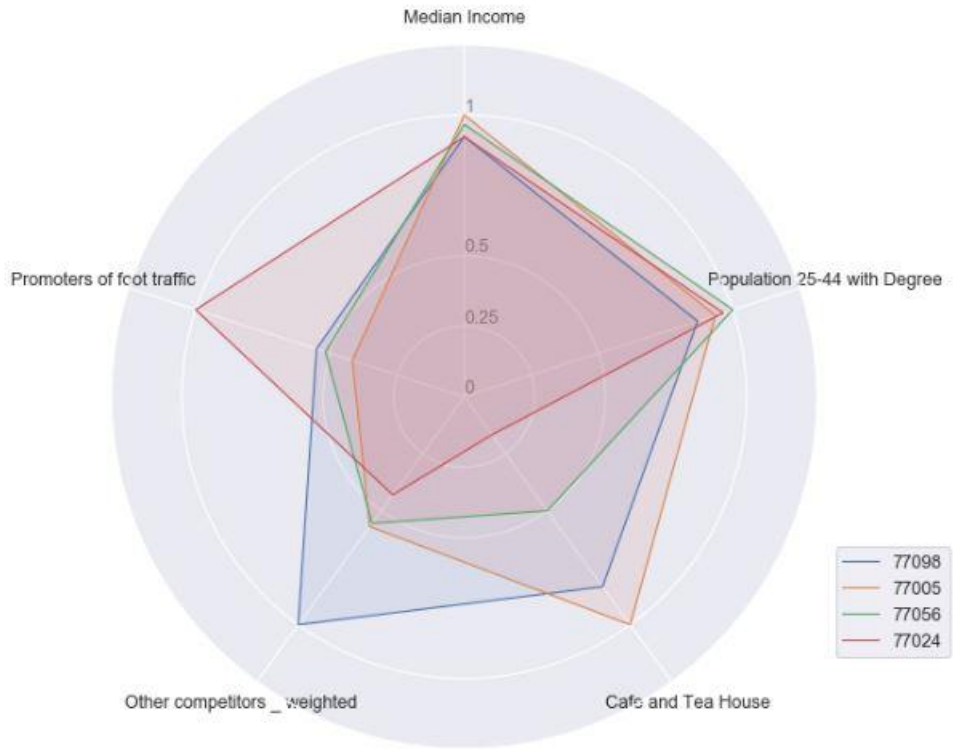
1) Median Income
2) Population 24-55 Years-Old with Associate's degree or +
3) Number of venues that can potentially promote food traffic
4) Number of cafés and tea houses (direct competition)
5) Number of venues with high level of competition
6) Number of venues with medium level of competition
7) Number of venues with low level of competition



The analysis of this radar chart puts the Memorial, Memorial Villages (zip code 77024) as a high candidate for the best neighborhood to establish a new café. But before coming to this conclusion, a second test will be done.
Let's aggregate all the venues from low to high level of competition at one single variable, that will be called "Other competitors _ weighted", in which venues received the following weights:

| Variable | Weight |
|---|---|
| Number of venues with high level of competition | 5 |
| Number of venues with medium level of competition | 3 |
| Number of venues with low level of competition | 1 |

The results obtained are below:




The second analysis lead to the same conclusion that the neighborhood Memorial, Memorial Villages (zip code 77024) still presents itself as the best option for opening a new café in Houston, having the following characteristics:

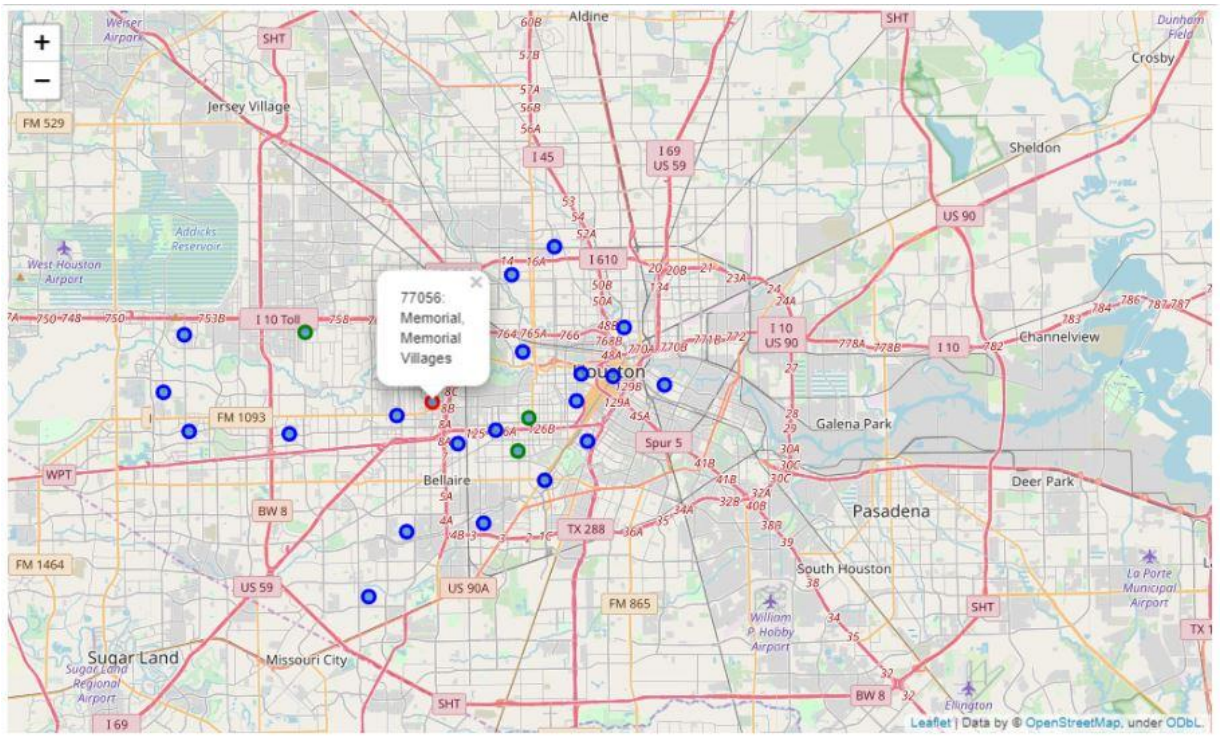
1) Similar Median Income as the other potential neighborhood >> **Neutral**
2) Similar population of target customer as the other potential neighborhood >> **Neutral**
3) Higher number of venues that can promote foot traffic but does not compete at the same market as the proposed business >> **Positive**
4) Smaller number of competitors >> **Positive**
5) Smaller number of cafés or tea houses >> **Positive**

## 4. Results

The analysis pointed the following 4 neighborhoods as the ones with the best potential for opening a new café in the city of Houston.

| | Neighborhood | zipcode | Latitude | Longitude | Median Income | Population 25-44 with Degree | Cafe and Tea House | High level of competition | Medium level of competition | Low level of competition | Promoters of foot traffic |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | Upper Kirby | 77098 | 29.74 | -95.41 | 61816 | 5714 | 5 | 2 | 41 | 11 | 33 |
| 9 | West University Place, Southside Place, Southa... | 77005 | 29.72 | -95.42 | 66926 | 6148 | 6 | 5 | 18 | 3 | 25 |
| 11 | Galleria, Uptown, | 77056 | 29.74 | -95.47 | 64714 | 6575 | 3 | 4 | 19 | 3 | 31 |
| 13 | Memorial, Memorial Villages | 77024 | 29.78 | -95.54 | 61733 | 6340 | 1 | 4 | 14 | 0 | 60 |

Those 4 neighborhoods can be better visualized at the following map, with the selected one (77024 – Memorial, Memorial Villages) shown in red.




## 5. Conclusion and Recommendations

The Memorial and Memorial Village neighborhood (zip code 77024) was identified as the recommended area for a new café. But this is just the first step for a full market analysis.

Having selected the location of the cafe, it is recommended to move from the macro analysis performed in this project to a micro analysis of the venues at the neighborhood level. In the micro analysis stage an in depth evaluation of the venues will be studied to understand the behavior and distribution of competitors.