Choosing best location for a new Coffee shop in Chicago

Juan J. Belaustegui

February 2020

1. Introduction

1.1 Background

Americans consume 400 million cups of **coffee** per day, making USA the leading consumer of **coffee** in the world. To have an idea, that's mean an average of 250 Cups of espresso and coffee drinks are sold per day at almost any Coffee shop with a great visible location.

If we have in mind that market equals to $12 billions in annual sales[1], seems very interesting to analyze if there is space to open a new Coffee Shop.

Chicago is the 3rd major city in US and the statistics says their citizens are above the average of US coffee consumption per capita.

1.2 Problem

We have a customer that is in the Coffee shop business and it is interested to open a Cafeteria in Chicago. They have requested a Data Scient to analyze public data available and provide a recommendation with best option where to implement new Cafeteria.

2. Data acquisition and cleaning

2.1 Data sources
Getting some insights about coffee consumption, I found there are many cities that are plenty of coffee shops: Seattle, Manhattan, San Francisco and Pittsburg leading the ranking[1] in the US.

The next one in the ranking is Chicago, that shows it in the limit of the "National Saturation Line".
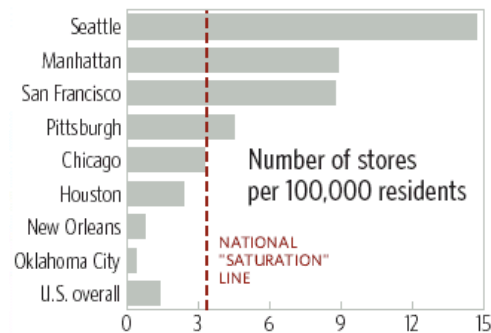

Fig.1 Coffee stores density US

This is an average, so probably we may find Areas with too many coffees and others with no one. So, it will be relevant to check every single Area and neighborhood from Chicago.

The data of the community areas and neighborhoods from Chicago has been taken from Wikipedia[2]

I will be using Geocoder library to get the coordinates from different Neighborhoods and using *Foursquare* API to get all the venues from each Neighborhood. Especial attention to all Coffee shops and Cafes that we can find, in order to exclude them and focus on Areas where is a lack of these stores.

Additionally, I'll get the neighborhoods population, so we can have another information on what to make a decision. This information has been gathered it from last census data[3]

2.2 Data cleaning
The data of the neighborhoods was read as a table from the web. After some cleanup, we got a Pandas dataset like table 1.

	index	Area	Neighbourhood
0	0	Albany Park	Albany Park,Mayfair,North Mayfair,Ravenswood M...
1	1	Archer Heights	Archer Heights
2	2	Armour Square	Armour Square,Chinatown,Wentworth Gardens
3	3	Ashburn	Ashburn,Ashburn Estates,Beverly View,Crestline...
4	4	Auburn Gresham	Auburn Gresham,Gresham
5	5	Austin	Galewood,The Island,North Austin,South Austin
6	6	Austin, Humboldt Park	West Humboldt Park
7	7	Avalon Park	Avalon Park,Marynook,Stony Island Park
8	8	Avondale	Avondale,Jackowo,Wacławowo
9	9	Avondale, Irving Park	Polish Village
10	10	Belmont Cragin	Belmont Central,Brickyard,Cragin,Hanson Park

Table 1: Community Areas and Neighborhoods from Chicago

The information related to neighborhood population was read from a pdf document. I used **Camelot** module to read the file. After some parameter's adjustment to have a good presentation of the data, I deleted and disregarded some data not necessary for the analysis, like population in other years and percentage of growing. I got another Pandas dataframe like table2.

	Neighborhood	Population
0	Rogers Park	54,991
1	West Ridge	71,942
2	Uptown	56,362
3	Lincoln Square	39,493
4	North Center	31,867
5	Lake View	94,368
6	Lincoln Park	64,116
7	Near North Side	80,484
8	Edison Park	11,187
9	Norwood Park	37,023

Table 2: Neighborhood Population. Census 2010

After that I'll be using Geocoder to get coordinates and incorporate everything within the same dataframe and plot them in a map.
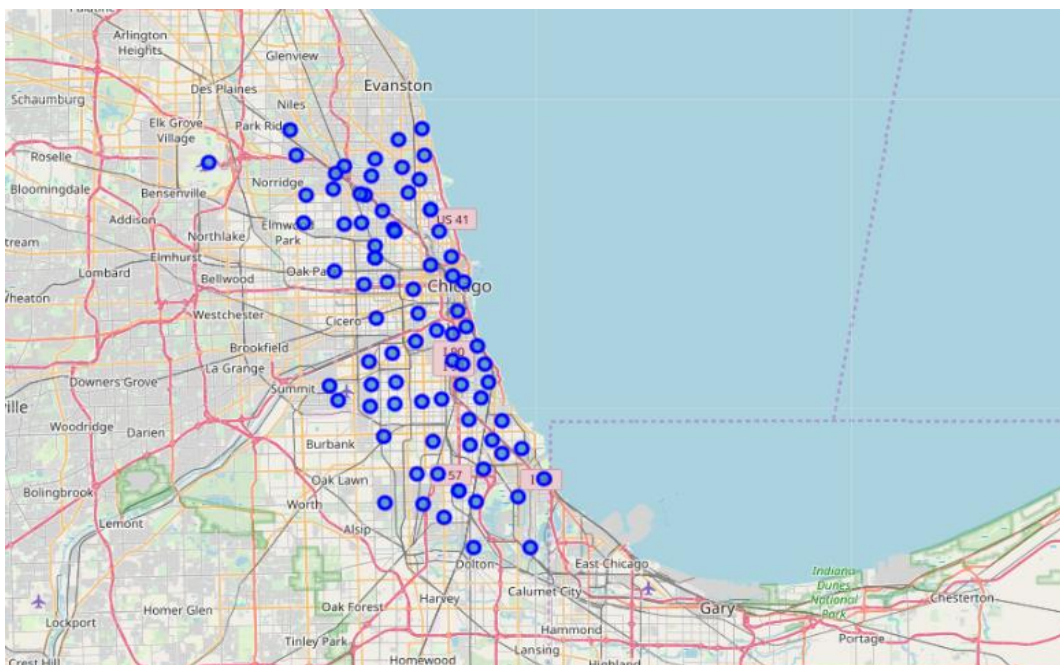

Fig 2: Neighborhoods from Chicago

2.3 Data selection

Once I have all the data into the same dataframe, I'll be using **Foursquare** API to explore the venues around the different areas. I'm going to exclude the neighborhoods where I find relevant presence of Coffee shops.

With a more concentrated data I'll be running a clustering with K means and will choose the one with presence of venues that can makes attractive for a coffee, taking in consideration of course, the amount of people that lives close in the Area, in order to increase the probability to reach a consumption of at least 250 cups of coffee per day.

3. Bibliography

1. http://www.e-importz.com/coffee-statistics.php
2. https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago
3. www.chicago.gov/content/dam/city/depts/zlup/Zoning_Main_Page/Publications/Census_2010_Community_Area_Profiles/Census_2010_and_2000_CA_Populations.pdf