

Analyzing The Coffee Shops Business in Downtown Chicago and Surroundings

Elie Feghali
19/8/2021

1. Introduction

Business Problem:

Our client is a European coffee shops chain planning to expand its business and open several shops in major US cities.

For many years our client has dominated the coffee shop industry in some European cities like Paris, Bruxelles and Milano to name few. But their revenue has been stable in the last two years and their management is under pressure by the shareholders.

So the client is looking for new markets to increase revenues and grow the business.

Scope of work:

As a start, we have been tasked to analyze the coffee shops industry in the downtown Chicago area to determine useful insights like:

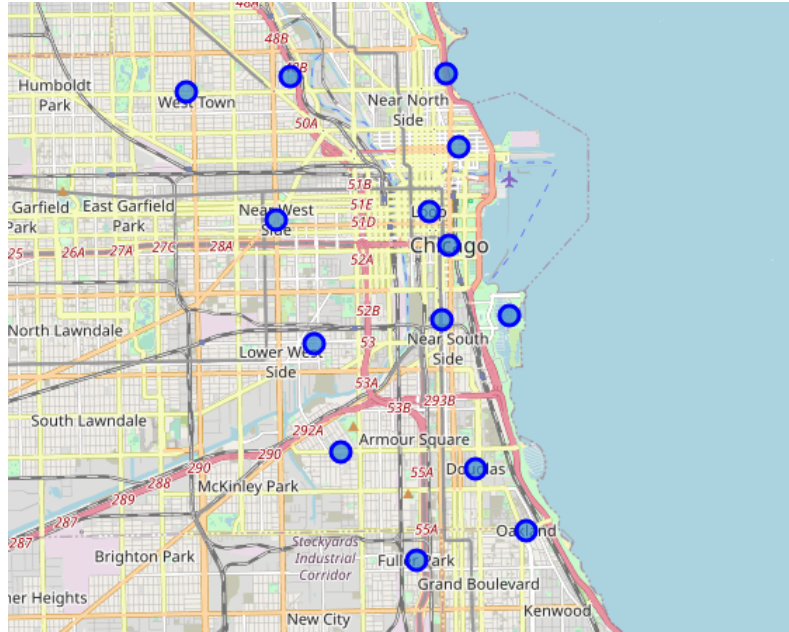
1. The location of the coffee shops in downtown Chicago neighborhoods.
2. The major coffee chains that should be considered as competitors.
3. Coffee shops venues that have the best ratings and are preferred by Chicago coffee drinkers.
4. Exploration of Starbucks neighboring venues knowing that it is the top US chain which implements a proven real estate strategy.

2. The Data

Downtown Chicago Neighborhoods:

We will collect the coordinates of downtown Chicago and its surrounding neighborhoods using nominatim from geopy.geocoders library.

	district	lat	long
0	Loop Chicago, IL	41.881609	-87.629457
1	Gold Coast Chicago, IL	41.906699	-87.625331
2	Streeterville Chicago, IL	41.893402	-87.622001
3	South Loop Chicago, IL	41.861785	-87.626121
4	West Loop Chicago, IL	41.875562	-87.624421



Coffee Shops Data

To get the data about downtown Chicago coffee shops, we query the string "coffee shop" using the Foursquare API for each of the neighborhoods. Then, we transform the resulting JSON file into a pandas dataframe.

Now, we can explore our data which includes: the shop name, location, category and ID. Eventually, Starbucks has by far the greatest number of branches.

We can also extract from Foursquare API each coffee shop's rating which helps us understand the consumers' sentiments about the competition in Chicago.

	Venue	ID	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude
0	Intelligentsia Coffee	4a5bc165f964a520b9bb1fe3	Loop Chicago, IL	41.881609	-87.629457	41.884517	-87.625783
1	Dollop Coffee & Tea	5396301d498e2660782b078a	Loop Chicago, IL	41.881609	-87.629457	41.876925	-87.628864
2	Peet's Coffee & Tea	544a7b1e498e333e2f59ecc7	Loop Chicago, IL	41.881609	-87.629457	41.881823	-87.632870
3	Peet's Coffee & Tea	54060b96498ed2702c0632f7	Loop Chicago, IL	41.881609	-87.629457	41.882467	-87.624686
4	Intelligentsia Coffee	49d51ce3f964a520675c1fe3	Loop Chicago, IL	41.881609	-87.629457	41.878092	-87.629615

Starbucks Data

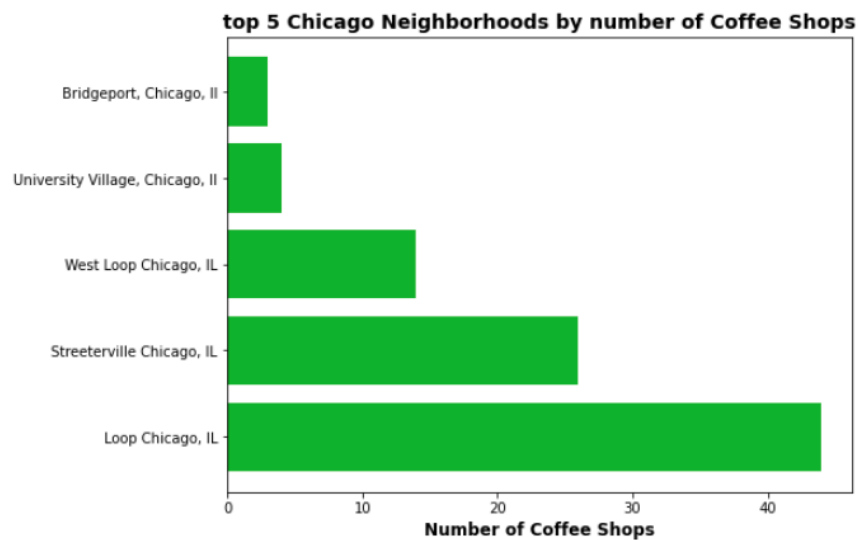
Since Starbucks is the coffee shops business leader, it is beneficial to explore its business in downtown Chicago. Starbucks is also known to have a successful real estate strategy as the decision makers invest largely in getting the best locations in any metropolitan area. With Foursquare API, we can get insights by gathering data like Starbucks locations, main categories of neighboring venues, branches ratings and more.

3. Methodology

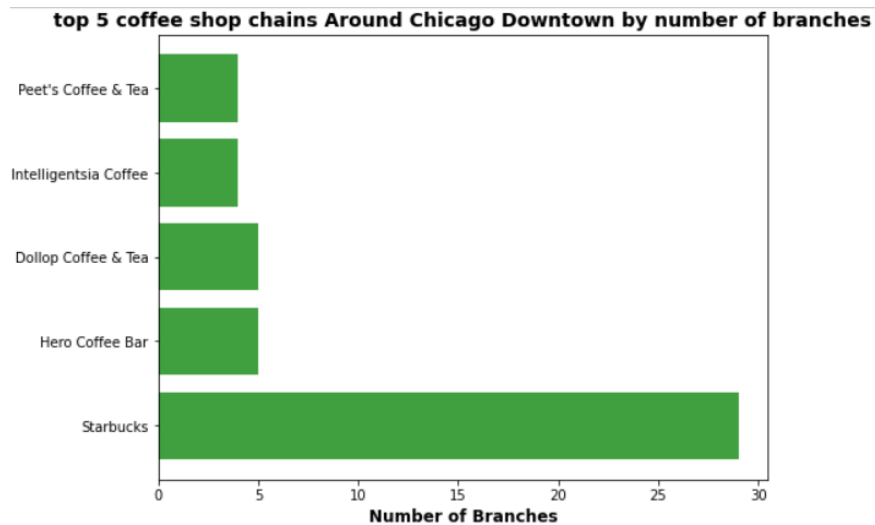
To analyze the coffee shops business in Downtown Chicago, we will explore the data we collected from Foursquare. We will also try to discover insights from Starbucks Chicago branches by collecting data about the venues surrounding them. K-means clustering model might also help us determine the type of neighborhood to open a coffee shop.

Data Exploration

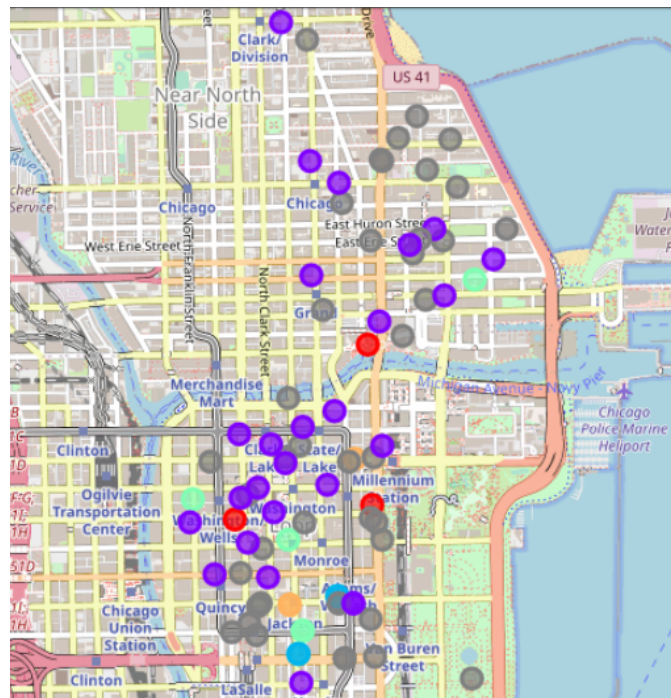
Most coffee shops are located in the Loop Chicago neighborhood. Streeterville Chicago comes next.



Starbucks has by far the most number of branches. It is followed by Hero Coffee Bar and Dollop Coffee & Tea



Here's the map showing the coffee shops distribution. Starbucks branches are shown in violet.

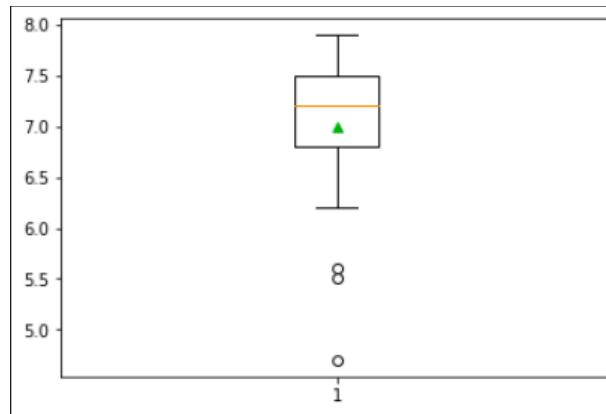


Coffee shops ratings

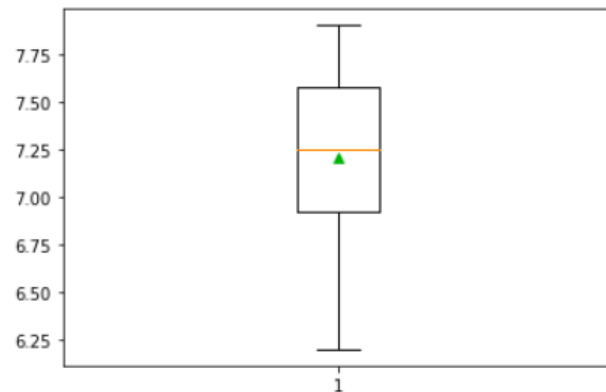
Next, we extract the rating of our coffee shops from Foursquare API to help us understand customer's sentiment and what makes a desirable coffee shop for Chicago citizens. We notice that Starbucks is not among the top 5 coffee shops and its branches have an average rating of 6.99. Please note that the result might not be too accurate due to lack of enough data.

	Venue	Ratings
0	La Colombe Coffee Roasters	9.20
1	Do-Rite Donuts & Chicken	9.00
2	Jackalope Coffee & Tea	8.90
3	Hotel Palomar	8.80
4	Cafecito	8.80
0	Starbucks	6.99

We investigate Starbucks ratings further using a boxplot. We notice that there are some outliers which draw the rating mean down.



So we remove the outliers and Starbucks, and the rating increases to almost 7.25/10. But Starbucks still does not make it to the top 5 coffee shops in terms of rating.



Exploring Starbucks Neighborhood

We will try to figure out the best locations in Chicago for our client to open its branches. We are trying to take advantage of Starbucks successful strategy in deciding where to locate their branches.

To achieve that, we will collect data about Starbucks' branches neighboring venues from Foursquare. Then, we will extract their categories (coffee shop, Italian restaurant, hotel, ...).

	Branch ID	Branch Latitude	Branch Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	4bec0a73a9900f4744781840	41.882818	-87.632493	Lotus Bahn Mi	41.883442	-87.633041	Vietnamese Restaurant
1	4bec0a73a9900f4744781840	41.882818	-87.632493	CorePower Yoga	41.883355	-87.632869	Yoga Studio
2	4bec0a73a9900f4744781840	41.882818	-87.632493	Starbucks	41.882818	-87.632493	Coffee Shop
3	4bec0a73a9900f4744781840	41.882818	-87.632493	Qdoba Mexican Grill	41.883395	-87.632670	Mexican Restaurant
4	4bec0a73a9900f4744781840	41.882818	-87.632493	California Pizza Kitchen	41.882839	-87.632896	Pizza Place

In some locations, two Starbucks branches are less than 500 meters away from each other's.

So, it is important to set the radius of the venue exploration API request by considering overlap between Starbucks branches and avoid multiple data collection of the same venue.

We then proceed by transforming the Venue Category column to dummy variable columns. This will help us group each Starbucks branch and sort the categories of neighboring venues using the frequency of their occurrence.

ID	American Restaurant	Antique Shop	Art Gallery	Asian Restaurant	BBQ Joint	Bakery	Bank	Bar	Building	Burger Joint	Burrito Place	Café	Camera Store	Chocolate Shop	Clothing Store	Cocktail Bar
49d81b0ff964a520aa5d11e3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4a6cf09ff964a5201fd211e3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4a75ed06f964a520afe111e3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4a88aa28f964a520540720e3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4a993362f964a5207f2d20e3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

ID	1st Venue	2nd Venue	3rd Venue	4th Venue	5th Venue	6th Venue	7th Venue	8th Venue	9th Venue	10th Venue
49d81b0ff964a520aa5d11e3	Dumpling Restaurant	Toy / Game Store	American Restaurant	Pizza Place	Sake Bar	Restaurant	Ramen Restaurant	Public Art	Pub	Portuguese Restaurant
4a6cf09ff964a5201fd211e3	Mexican Restaurant	American Restaurant	Pet Service	Sake Bar	Restaurant	Ramen Restaurant	Public Art	Pub	Portuguese Restaurant	Poke Place
4a75ed06f964a520afe111e3	Hotel Bar	American Restaurant	Pet Service	Sake Bar	Restaurant	Ramen Restaurant	Public Art	Pub	Portuguese Restaurant	Poke Place
4a88aa28f964a520540720e3	Italian Restaurant	American Restaurant	Pet Service	Sake Bar	Restaurant	Ramen Restaurant	Public Art	Pub	Portuguese Restaurant	Poke Place
4a993362f964a5207f2d20e3	Sandwich Place	American Restaurant	Pet Service	Sake Bar	Restaurant	Ramen Restaurant	Public Art	Pub	Portuguese Restaurant	Poke Place

Again, we notice repetition of some categories due to low venues data on Foursquare.

Now to understand the neighborhood types of each Starbucks branch, we can apply **K-means** clustering model to our dataframe.

In our case, we iterated the number of clusters from 3 to 5 and run the model and the results are almost the same. So, we proceed with 5 clusters and add the clusters list to the dataframe as a new column.

Venue	ID	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Latitude	Longitude	Venue Category	Ratings	Cluster Labels	1st Venue	2nd Venue	3rd Venue
0 Starbucks	4bec0a73a9900f4744781840	Loop Chicago, IL	41.881609	-87.629457	41.882818	-87.632493	Coffee Shop	7.7	0	Vietnamese Restaurant	American Restaurant	Pet Service
1 Starbucks	4b969aa8f964a52006d834e3	Loop Chicago, IL	41.881609	-87.629457	41.883350	-87.627362	Coffee Shop	7.9	0	Yoga Studio	Pet Service	Sake Bar
2 Starbucks	4b182fd6f964a52089ce23e3	Loop Chicago, IL	41.881609	-87.629457	41.882140	-87.630520	Coffee Shop	7.5	1	Coffee Shop	American Restaurant	Sculpture Garden
3 Starbucks	4a6cf09ff964a5201fd21fe3	Loop Chicago, IL	41.881609	-87.629457	41.880790	-87.632140	Coffee Shop	6.7	0	Mexican Restaurant	American Restaurant	Pet Service
4 Starbucks	4b10054af964a520946723e3	Loop Chicago, IL	41.881609	-87.629457	41.885902	-87.628860	Coffee Shop	7.3	0	Pizza Place	American Restaurant	Pet Service

We also visualize the clusters on Chicago map including each branch rating and cluster number in the label.



We notice that Starbucks branches that belong to cluster 3 and located in Streetsville, Chicago have the lowest customer ratings. These were the same outliers that we noticed earlier in our data exploration. This has to be investigated further.

4. Results

After collecting and analysing data about the Downtown Chicago coffee shops business and Starbucks in particular, we found the following results:

- 105 coffee shops are found on Foursquare in Downtown Chicago and its surroundings.
- 29 Starbucks branches were found.
- Starbucks is not among the 5 top rated coffee shops and its average customer rating is 7.25/10 after ignoring the outlier branches.

- Applying K-means clustering model on Starbucks branches considering their top neighboring venues, helped us identify the best type of locations to open a coffee shops. It is mainly at the Loop, Chicago.

5. Discussion

We discovered that Starbucks branches located at Streeterville, Chicago had an unusual low rating on Foursquare. Those branches were the outliers identified earlier and additional investigation might help recognize the reason behind the customers sentiment. So, K-means model with visualization of the cluster using Folium helped us reach this insight.

The same analysis will be applied to other major US cities to get a better understanding of the Foursquare data.

We should also note the lack of data for some of the venues like the low number of reviews and customers rating which could hinder our findings and cause biases.

6. Conclusion

Analyzing the coffee shops competition in the city of Chicago using Foursquare API proved to be very effective and would help our clients in their decision making. The insights would help avoid opening shops in some neighborhoods and further investigate the reasons behind the success and failure of some shops. Additional data that could be extracted from Foursquare with the application of other machine learning models can reveal more insights. We could explore customers sentiment by extracting their reviews.