

# CAPSTONE Project: Chicago QSR area selection

April 2020

## **BACKGROUND AND PROBLEM STATEMENT.**

Chicago may be windy eight months out of the year, but the city's reputation for fine dining and Midwestern comfort food more than makes up for its lakeside chill. The city has evolved beyond a meat-and-potatoes town into a sophisticated, innovative hub of culinary treasures.

There are approximately 2.7 million people and 1,194,337 households residing within 70 community areas within the city limits of Chicago. More than half the population of the state of Illinois lives in the Chicago metropolitan area[1]. Chicago is also very ethnic in its outlook and more than 100 languages / dialects are spoken everyday. Chicago has a diverse eating culture. Chicago has its share of non chain and QSR (quick service restaurants) in each area.

If a chain such as "Chipotle" or "Subway" wanted to open a new restaurant, how should they decide where to open the next franchise branch.

At first glance Chicago numbers are high (touted to be Murder capital of USA), but Chicago falls significantly lower when you look at its crime rates per 100,000 people. Since rates account for a city's population, it's often seen as a more accurate way to view a city's violence. Numbers and statistics do not define a city, but are merely a handful of variables cities can use to understand what problems they face and how to fix them.

Needless to say Chicago does have a high crime rate and any one wanting to open a business would want to take that into consideration. It would be good to know the numbers in a specific time frame to hone in on which areas are safer than the others.

For any QSR (Quick Service Restaurant) to open a branch it is important to get the lay of the land before wanting to put forward a proposal or start to scout property. There are many data points to look for:

# CAPSTONE Project: Chicago QSR area selection

April 2020

- Are there other franchises of their kind in the area?
- Are there similar types of restaurants in that area?
- What is the mean income of the neighborhood?
- How safe is the neighborhood?
- Some demographic data in terms of younger population between 18 and 64 with disposable income could also be a determining factor.

## **DATA AND SOLUTION.**

There is lot of data available from the city in terms of:

- number and type of crimes
- Demographic data of neighborhoods
- Chicago public school data
- Four Square Api's to identify the type of venues in the neighborhood
- Deep dive into restaurants and QSRs in the neighborhoods.

# CAPSTONE Project: Chicago QSR area selection

April 2020

## DATA.

Crime Data:

Below is a snapshot of the columns of Crime Data from  
**CHICAGO\_CRIME\_DATA:**

<https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv> [2]

INCIDENT_NUMBER	BLOCK	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC	DISORDERLY	WEAPONS	COMMUNITY_AREA_NUMBER	FBI_CODE	X_COORDINATE	Y_COORDINATE	YEAR	UPDATE_DATE	LATITUDE	LONGITUDE	LOCATION
-----------------	-------	--------------	-------------	----------------------	--------	----------	------------	---------	-----------------------	----------	--------------	--------------	------	-------------	----------	-----------	----------

Based on this data we can identify which neighborhoods have the least number of occurrences of crime. We can choose 2 to 3 neighborhoods to do additional analysis

The idea would be to get a count of crimes in each community area. There are a total of 70 community areas to choose from. We can do this by grouping on the community\_area\_number and getting a total count.

```
C_Crime['COMMUNITY_AREA_NUMBER'].nunique()
```

```
Out[78] 70
```

# CAPSTONE Project: Chicago QSR area selection

April 2020

Demographic Data:

From **CENSUS\_DATA**: <https://ibm.box.com/shared/static/05c3415cbfbtfnr2fx4atenb2sd361ze.csv>  
[2]

The data labels are:

<b>COMMUNITY_A REA_NUMBER</b>	<b>COMMUNITY_ AREA_NAME</b>	<b>PER CEN T OF HOU SING CRO WDE D</b>	<b>PERCE NT HOUSE HOLDS BELO W POVER TY</b>	<b>PERCE NT AGED 16+ UNEMP LOYED</b>	<b>PER CEN T AGE D 25+ WIT HOU T HIG H SCH OOL DIPLO MA</b>	<b>PER CEN T AGE D UND ER 18 OR OVE R 64</b>	<b>PER_CAPI TA_INCOM E</b>	<b>HARDSHI P_INDEX</b>
-----------------------------------	---------------------------------	--	---	--	---	--	------------------------------------	----------------------------

Below is a snapshot of the data from the table.

<b>0</b>	1.0	Rogers Park	7.7	23.6	8.7	18.2	27.5	23939	39.0
<b>1</b>	2.0	West Ridge	7.8	17.2	8.8	20.8	38.5	23040	46.0

# CAPSTONE Project: Chicago QSR area selection

April 2020

2	3.0	Uptown	3.8	24.0	8.9	11.8	22.2	35787	20.0
---	-----	--------	-----	------	-----	------	------	-------	------

We can identify demographic information such as Per\_Capita\_income and Percent of people within 18 to 64. This data will also be used to link the community area number (the key for the data in all other tables) to the community area names. The community area name and its coordinates are then used in the Four Square APIs.

Chicago Public School Data:

From **CHICAGO\_PUBLIC\_SCHOOLS**

<https://ibm.box.com/shared/static/f9gjvj1gjmxxzycdhplzt01qtz0s7ew7.csv> [2]

School ID	NAME_OF_SCHOOL	Elementary, Middle, or High School	Street Address	City	State	ZIP Code	Phone Number	Link	Network Manager	Collaborative Name	Adequate Yearly Progress Made?	Track Schedule	CPS Performance Policy Status	CPS Performance Policy Level	HEALTHY_SCHOOL_CERTIFIED	Safety Icon	SAFETY_SCORE
-----------	----------------	------------------------------------	----------------	------	-------	----------	--------------	------	-----------------	--------------------	--------------------------------	----------------	-------------------------------	------------------------------	--------------------------	-------------	--------------

The data columns above have been truncated to see only the relevant data points. The SAFETY\_SCORE is relevant here to verify the community area that we wish to identify as a place to open a QSR. There is a lot of information here

# CAPSTONE Project: Chicago QSR area selection

April 2020

but we are interested in fields shaded above. This can provide corroboration of the area chosen

Four Square Data points:

Foursquare data is very comprehensive and it powers location data for Apple, Uber etc. For this business problem, the Foursquare API for venues can be used to retrieve information about the popular spots [3]. The other relevant data is regarding the profile of the venues of the neighborhood. Further deep diving into the availability and concentration of QSRs in the specific areas. The popular spots returned depends on the highest foot traffic and thus it depends on the time when the call is made. So we may get different popular venues depending upon different times of the day. The call returns a JSON file and we need to turn that into a data-frame.

For this specific exercise we are choosing **Chipotle and Subway** as the QSRs looking to open a branch in a safe neighborhood.

## **METHODOLOGY OF THE SOLUTION.**

-- Analyze the crime data to identify neighborhoods with the least occurrences of crime. The below 6 Communities have the least count of crime occurrences.

COMMUNITY COUNT

33.0	1
75.0	1
47.0	1
55.0	1
12.0	1

# CAPSTONE Project: Chicago QSR area selection

April 2020

07.0

1

-- For these neighborhoods identify the top 2 with high income, and population. Below is the crime data mapped with the community demographic data. The 6 area names are below. Lincoln Park and Near South side are the 2 community areas which we can do a deep dive for venues.

6	7.0	Lincoln Park	0.8	12.3	5.1	3.6	21.5	71551	2.0
11	12.0	Forest Glen	1.1	7.5	6.8	4.9	40.5	44164	11.0
32	33.0	Near South Side	1.3	13.8	4.9	7.4	21.8	59077	7.0
46	47.0	Burnside	6.8	33.0	18.6	19.3	42.7	12515	79.0
54	55.0	Hegewisch	3.3	17.1	9.6	19.2	42.9	22677	44.0
74	75.0	Morgan Park	0.8	13.2	15.0	10.8	40.3	27149	30.0

-- For these neighborhoods, run the Four Square APIs, to get the venues

For the 2 community areas getting the coordinates using geocoders libraries.

0	7.0	Lincoln Park	(42.2505943, -83.1785361)
---	-----	--------------	------------------------------

# CAPSTONE Project: Chicago QSR area selection

April 2020

2	33.0	Near South Side	(41.8566999, -87.6247738)
---	------	-----------------	---------------------------

-- From these venues identify restaurants. The first community that was analysed was Near South Side. There were 51 venues identified by Four Square APIs.

nearby\_venues.shape

Out[59]: (51, 4)

The first 3 are listed below.

	name	categories	lat	lng
0	TeaPotBrew Bakery	Café	41.857845	-87.625431
1	Opart Thai House	Thai Restaurant	41.856084	-87.627154
2	La Cantina Grill	Mexican Restaurant	41.856310	-87.623778

-- Verify the concentration of QSRs. No QSRs were found in the entire neighborhood.

-- Do specific checks of presence of “Chipotle” or “Subway” restaurants.

Looking for “Chipotle” and “Subway” in the list of venues. We get a false across all the 53 rows.



# CAPSTONE Project: Chicago QSR area selection

April 2020

*# Count number of Chipotle's in the area*

```
nearby_venues['name'].str.contains('Chipotle').value_counts()
```

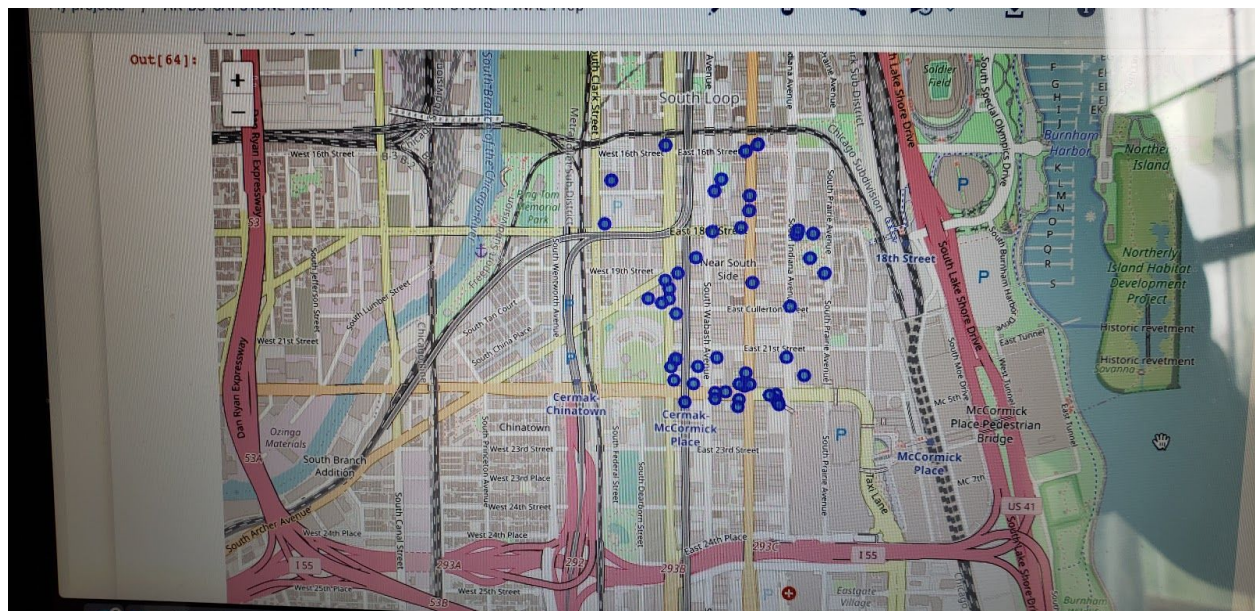
```
Out[46]: False    53 Name: name, dtype: int64
```

*# Count the number of Subways in the area.*

```
nearby_venues['name'].str.contains('Subway').value_counts()
```

```
Out[47]: False    53 Name: name, dtype: int64
```

-- Map these to identify the best place to procure the restaurant [4].



## RESULTS.

# CAPSTONE Project: Chicago QSR area selection

April 2020

The table below has the list of restaurants in the Near South Side community area. There are a lot of different cuisine based restaurants. There is only 1 fast food restaurant in the area. There are no sandwich shops, and there was only 1 non fast food mexican restaurant.

The data further substantiates that the Cermack / McCormick area seems to be the area of concentration of many venues including restaurants, hotels, parks, etc. This area could be the area to do further analysis of feasibility of opening a QSR (Chipotle or Subway).

Also, the other area near State Street and West Cullerton has a concentration of many venues including a park. This could also be a good location to do further analysis.

<b>categories</b>	<b>name</b>	<b>lat</b>	<b>lng</b>
<b>American Restaurant</b>	1	1	1
<b>Bistro</b>	1	1	1
<b>Burger Joint</b>	1	1	1
<b>Café</b>	2	2	2
<b>Caribbean Restaurant</b>	1	1	1
<b>Fast Food Restaurant</b>	1	1	1
<b>Food Court</b>	1	1	1
<b>Fried Chicken Joint</b>	1	1	1
<b>Italian Restaurant</b>	1	1	1
<b>Mexican Restaurant</b>	1	1	1
<b>New American Restaurant</b>	1	1	1
<b>Pizza Place</b>	2	2	2
<b>Restaurant</b>	1	1	1
<b>Seafood Restaurant</b>	1	1	1

# CAPSTONE Project: Chicago QSR area selection

April 2020

<b>Sushi Restaurant</b>	1	1	1
<b>Thai Restaurant</b>	1	1	1

## **OBSERVATIONS.**

Some observations of the Near South Side community area:

- -- Multi cuisine restaurants
- -- No QSR restaurants
- -- No Sandwich shops
- -- 1 Mexican restaurant but not fast food
- -- 2 potentially good locations 1 near McCormick place (Convention center) and 1 near Cullerton Park

A similar analysis of Lincoln park community area provides the following statistics:

- -- There were 18 venues
- -- No Chipotle nor Subway restaurants
- -- Near the Depaul University
- -- 2 QSRs Pizza Hut and White Castle
- -- 1 Mexican Bar and restaurant
- -- No Sandwich shops

## **CONCLUSION.**

We have got a small glimpse of how real life data-science projects look like. Some standard and frequently used python libraries have been used to scrape web-data, perform data manipulation, use Foursquare API to explore the community areas of Chicago and saw the results of recommendation of areas using the Folium leaflet map. Potential for this type of analysis in a real life business problem is discussed in great detail. Finally, since the analysis was mostly concentrated on the possibilities of opening a QSR restaurant targeting the safe, high income community areas of Chicago, I am sure more analysis can be done for other areas with different criteria. The data points to favorable observations regarding opening QSRs in the two community areas (Lincoln Park and Near South Side). Also, this is just a starting point, further analysis needs to be done on exact location, price analysis of locations, profitability and feasibility

# CAPSTONE Project: Chicago QSR area selection

April 2020

assessment of foot traffic, etc. Hopefully, this kind of analysis will provide some initial guidance to solve for more real-life challenges using data-science.

## **REFERENCES.**

- [1] Chicago — [Wikipedia](#)
- [2] IBM Data sources
- [3] [Foursquare API](#)
- [4] [Google Map](#)