

## **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. 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:

-- 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.

### **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>

ID	CASE_NUMBER	DATE	BLOCK	ILUCR	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC	BATTERY	DISRICT	WARD	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
```

Demographic Data:

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

The data labels are:

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

Below is a snapshot of the data from the table.

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

S c h o o l I D	N A M E_ O F_ S C H O O L	E l e m e n t a r y , M i d d l e , o r H i g h S c h o o l	S t r e t A d d r e s s	C i t y	S t a t e	Z I P C o d e	P h o n e N u m b e r	L i n k	N e t w o r k M a n a g e r	C o l l a b o r a t i v e N a m e	A d e q u a t e Y e a r l y P r o g r e s s M a d e ?	T r a c k S c h e d u l e	C P S P e r f o r m a n c e P o l i c y S t a t u s	C P S P e r f o r m a n c e P o l i c y L e v e l	H E A L T H Y_ S C H O O L_ C E R T I F I E D	S a f e t y I c o n	S A F E T Y_ S C O R E
--------------------------------------	--	--	--	------------------	-----------------------	---------------------------------	---	------------------	--	---	---	---	--	---	--	--	--

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 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. 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 for the solution:**

- Analyze the crime data to identify neighborhoods with the least occurrences of crime.
- For these neighborhoods identify the the top 3 with high income, and population
- For these neighborhoods, run the Four Square APIs, to get the venues
- From these venues identify restaurants
- Verify of the concentration of QSRs
- Do specific checks of presence of “Chipotle” or “Subway” restaurants.
- Map these to identify the best place to procure the restaurant.