PROJECT ON BANK ACCESSIBILITY IN ILLINOIS AND NEWYORK

Based on Foursquare Search

Abstract

This project tries to find top banks in Illinois and Chicago region in terms of number of branches in the state of Illinois/Chicago and New York State and regions in both states where banks are denser and sparser in terms of number of banks

Deepak Kumar Tiwari

deepakbhme.2011@gmail.com

Contents

1. Introduction

1.1 Background

New York and Illinois both are one of the biggest states in United States in terms of population and revenues. According to http://worldpopulationreview.com, New York is 4th most populous state and Illinois is 6th most populous in state with population of 19 millions and 12 millions respectively. Similarly, New York City and Chicago are one of the biggest mega cities not only in the United States but in whole world in terms industries, population diversities and architectures.

In this project we are trying to find the top banks in these two states in terms of accessibility to the population and what regions in these two states have the greatest number of banks and in what region population has very less to almost no accessibility to banks.

1.2 Business Problems

This project aims to answer few business and service questions which can be helpful for banks, governments as well as people.

1. How Illinois and Newyork are similar when compared in terms of banks density?
2. What are the top banks in terms of frequenncy in illinois and Newyork States?
3. If new bank want to open their branch or ATMs in either of State then which area should choose?
4. Who are the competitiors for the new aspirator bank or top bank operating in both States?
5. Which areas are having maximum banks and which banks dominates in those regions?
6. If an indivual want to open a bank account based on the bank nearest to him/her then what bank should he/she should be choosing?

1.3 Audience

Since this project tries to find top banks in terms of accessibility and density in Illinois and New York state, any bank who are operating a business or has aspirations to do so in these two states would be interested in knowing what are locations where there is need to open a bank or who are their top competitors so, that they can change their marketing or customer acquiring process accordingly in those regions.

People who wants to open a new bank account and doesn't want to waste their time in driving car to the bank or maybe doesn't have access to cars or public transportation would be interested in knowing what are the banks which are having maximum branches in the particular locations or which are closest to the their location.

Finally, government would be interested in knowing this result because there can be many places where there is not a single bank or is very sparsely present. If people are finding it hard to access the banks because of distance, then government must be willing to take steps to make private banks open their locations in those regions or simply open a government bank branch.

2. Data Collection and Usage

Data: To answer all these questions we would need following data:

1. Name of all neighborhoods and their location in Illinois and New York : Illinois neighborhood names have been obtained by scrapping the website (https://www.geonames.org/postal-codes/US/IL/illinois.html) using beautiful soup and requests library. New York neighborhood names and their locations has been obtained from the site https://geo.nyu.edu/catalog/nyu_2451_34572

2. Then using location data, I have used foursquare developer account to get list of nearest 10 banks within 500 meters range from the location if person tries to search banks. For that I have used search option and category id of banks as 4bf58dd8d48988d10a951735
3. I would be mainly dealing with the frequency of banks location or 10 search results for each location to answer all the above three questions.
4. We will analyze how many areas in both the states has how many banks within 500 meters range.
5. Which bank shows up as closest to the location and in how many areas. For, example if bank A is closest to the location o say 1st search in most of the region then it would be most accessible banks to the people. Similarly, I will find 5 topmost accessible banks
6. We would also try to find what are the regions in both states where there are maximum banks and where less or no banks.

3. Data Cleaning and Preparation

I scrapped the Illinois data from https://www.geonames.org/postal-codes/US/IL/illinois.html. It returns JSON file as

```
In [720]: df

Out[720]: '[{"0":null,"1":"Place","2":"Code","3":"Country","4":"Admin1","5":"Admin2","6":"Admin3"},{"0":1.0,"1":"Chicago","2":"6060
1","3":"United States","4":"Illinois","5":"Cook","6":null},{"0":null,"1":"41.886\\/-87.618","2":null,"3":null,"4":null,"5":nu
ll,"6":null},{"0":2.0,"1":"Chicago","2":"60603","3":"United States","4":"Illinois","5":"Cook","6":null},{"0":null,"1":"41.88
\\/-87.628","2":null,"3":null,"4":null,"5":null,"6":null},{"0":3.0,"1":"Chicago","2":"60605","3":"United States","4":"Illinoi
s","5":"Cook","6":null},{"0":null,"1":"41.871\\/-87.628","2":null,"3":null,"4":null,"5":null,"6":null},{"0":4.0,"1":"Chicag
o","2":"60606","3":"United States","4":"Illinois","5":"Cook","6":null},{"0":null,"1":"41.887\\/-87.639","2":null,"3":nul
```

This JSON data is then converted to datafram using pandas

Converting data into dataframe

```
3]: illinois_data=pd.DataFrame(data )
    illinois_data.head(10)
```

3]:

	0	1	2	3	4	5	6
0	NaN	Place	Code	Country	Admin1	Admin2	Admin3
1	1.0	Chicago	60601	United States	Illinois	Cook	None
2	NaN	41.886/-87.618	None	None	None	None	None

This dataframe after several data cleaning steps is turned into final pandas data frame as “

```
In [950]: Final_Data.head()
```

Out[950]:

	Place	postcode	county	latitude	longitude
0	Chicago	60601	Cook	41.886	-87.618
1	Chicago	60603	Cook	41.88	-87.628
2	Chicago	60605	Cook	41.871	-87.628
3	Chicago	60606	Cook	41.887	-87.639
4	Chicago	60607	Cook	41.872	-87.658

We have following variables that we would be using for FOURSQUARE search. Place name, Postal Code, County, Latitude and Longitude. For search we would be mainly using Place, latitude and longitude. We have 200 places names with different postal codes:

```
In [848]: Final_Data.shape
Out[848]: (200, 5)
```

Similarly Final New York data sets after data cleaning steps looks like this:

```
In [804]: Final_Data1.head(10)
Out[804]:
```

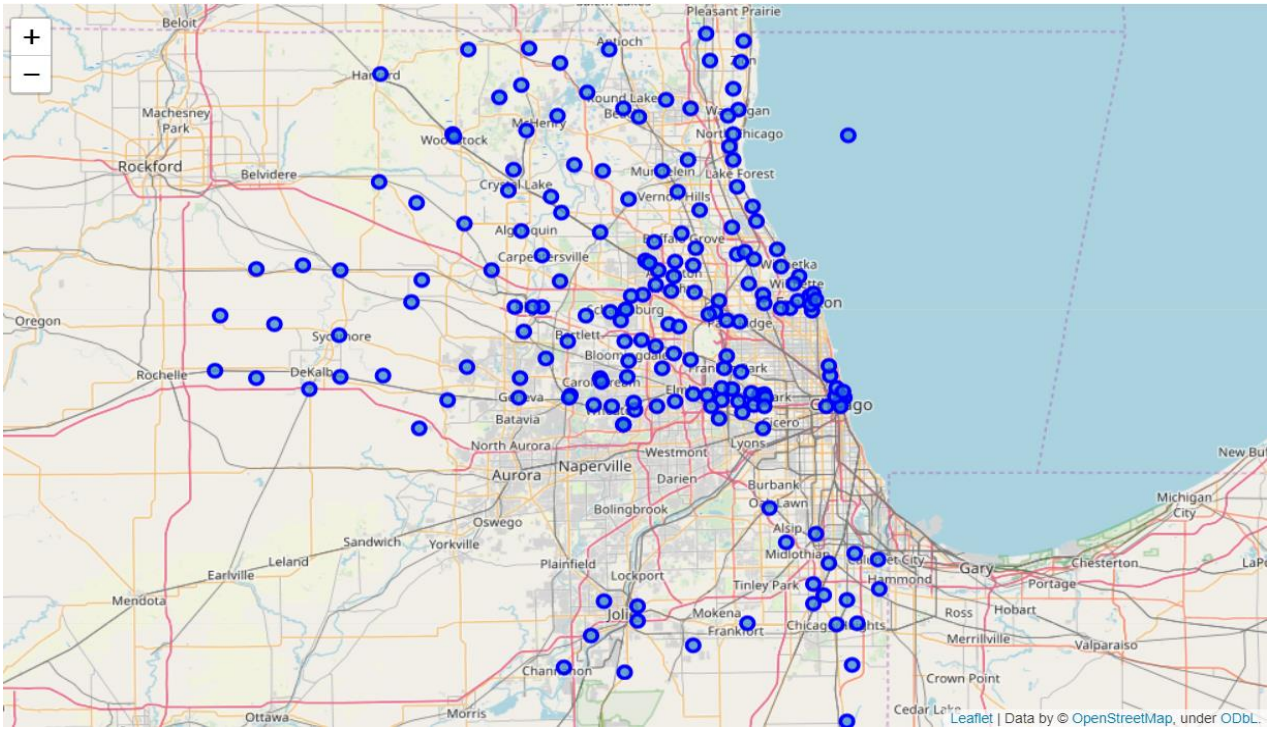
	County	Place	latitude	longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643

We have 306 places names with Place name and their latitudes an longitudes:

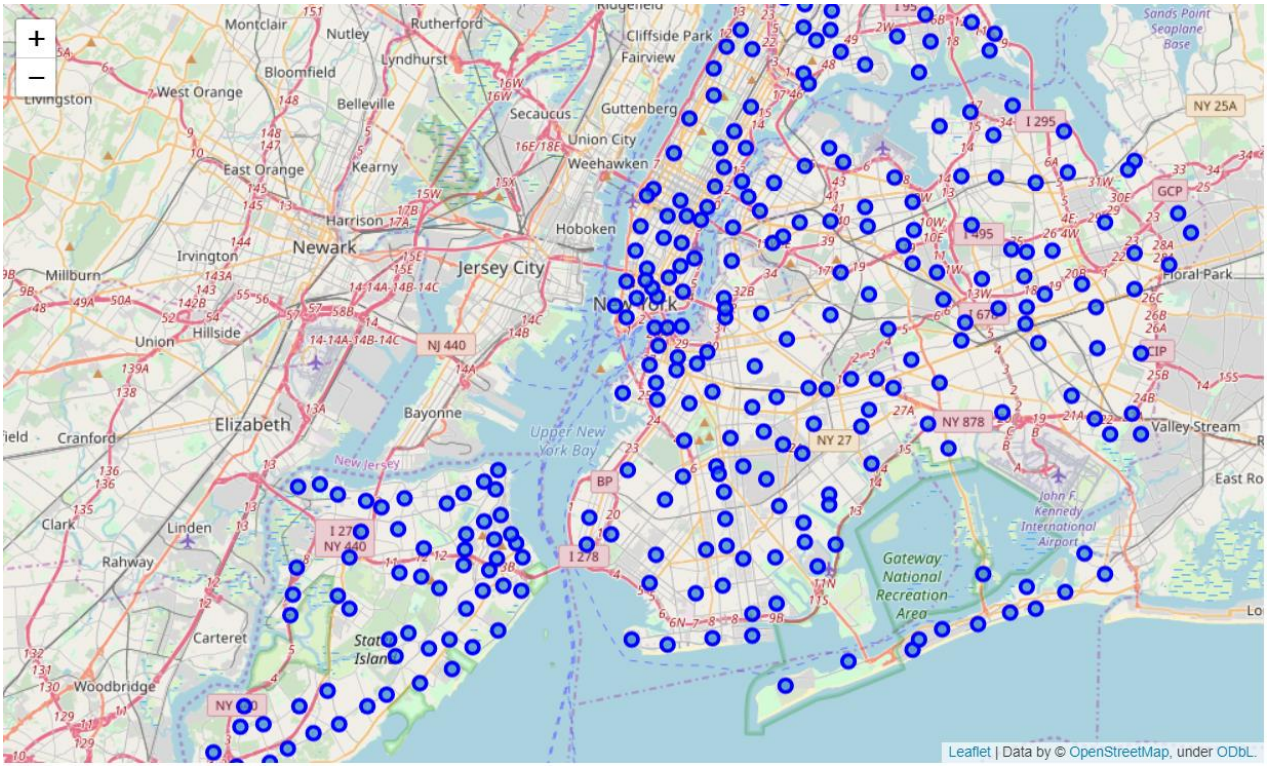
```
In [805]: Final_Data1.shape
Out[805]: (306, 4)
```

3. Location data exploration

Let’s see all the places on the map of Chicago and it’s neighborhood areas:



Similarly let’s explore the state of New York City:



`

4. Methodology

4.1. Foursquare Search

This project uses FOURSQUARE developer account to do the search. To do the search one needs to have developer account in turns the user gets CLIENT ID, CLIENT SECRET and VERSION.

Our aim for this project would be to search banks for all 200 places and 306 places in Illinois and New York respectively. For our search through Foursquare we have limited our distance to 500 meters or 0.310686 miles. I have assumed that the person wouldn't be willing to walk more than that distance. Also, number of searches has been limited to 10 for the same reason that the searcher would not look over results after 10 searches.Search is done using URL and enter the credentials and parameters: 'https://api.foursquare.com/v2/venues/search?categoryId=4bf58dd8d48988d10a951735&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'. Here “venues” is entered if we are searching places, category id# 4bf58dd8d48988d10a951735 is for banks.

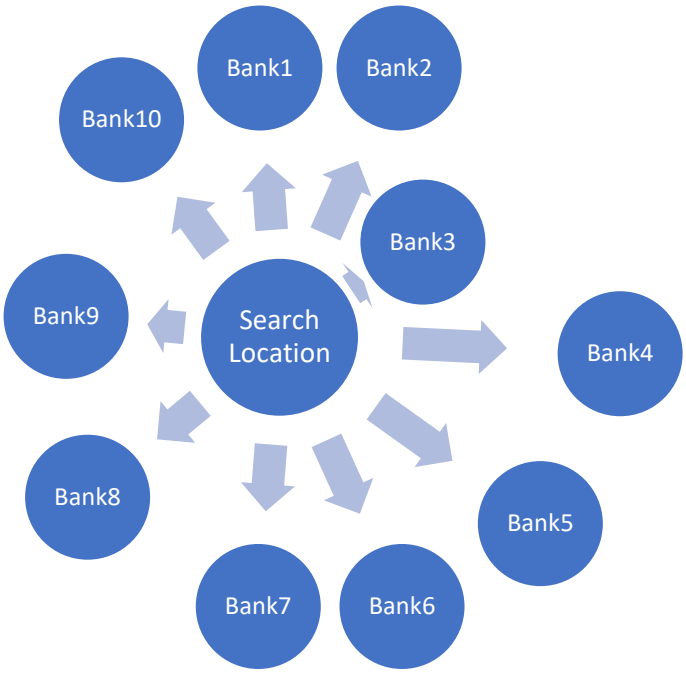


Fig 4.1

It returns top 10 closest banks as shown in the above figure. Foursquare returns JSON files as shown below. This is one complete bank for one location:

```
In [752]: result
Out[752]: {'meta': {'code': 200, 'requestId': '5d755ebed69ed0002cc1d739'},
           'response': {'venues': [{'id': '4fb533827beb65c72e11deea',
              'name': 'Chase Bank',
              'location': {'address': '3115 Ridge Rd',
               'lat': 41.56448,
               'lng': -87.541091,
               'labeledLatLngs': [{'label': 'display',
                 'lat': 41.56448,
                 'lng': -87.541091}],
               'distance': 366,
               'postalCode': '60438',
               'cc': 'US',
               'city': 'Lansing',
               'state': 'IL',
               'country': 'United States',
               'formattedAddress': ['3115 Ridge Rd',
                'Lansing, IL 60438',
                'United States']},
              'categories': [{'id': '4bf58dd8d48988d10a951735',
                'name': 'Bank',
                'pluralName': 'Banks',
                'shortName': 'Bank',
                'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/financial_',
                 'suffix': '.png'},
                'primary': True}],
              'referralId': 'v-1567973054',
              'hasPerk': False},
```

4.2 Interpreting Foursquare Result

I have stored all the list of the banks for each location in a list and later then transformed it to data frame to several analysis.

This is the dataframe containing closest 10 banks for each location in Illinois and New York.

```
In [763]: pd9
```

Out[763]:

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	place
0	Chase Bank	US Bank	Fifth Third Bank	Citibank	PNC Mortgage	PNC Mortgage	U.S. Bank Branch	PNC Bank	PNC Bank	Chase Bank	Chicago
1	Capital One Café	Chase Bank	Bank of America Building	BMO Harris Bank Headquarters	Northern Trust	Chase Bank	Federal Reserve Bank of Chicago	Fifth Third Bank & ATM	Northern Trust	Chase Bank	Chicago
2	Chase Bank	Byline Bank	New City Bank	Harris Bank	Coinstar	First American Bank	Fifth Third Bank & ATM	TCF Bank	BMO Harris Bank	PNC Mortgage	Chicago
3	Paypal / Venmo	Citibank	Ziegler	Chase Bank	MB Financial Bank	Huntington Bank	U.S. Bank Branch	Wells Fargo	Bank of America	Chase Bank	Chicago
4	Citibank	Chase Bank	Fifth Third	National Republic Bank	Lakeside Bank	Washington Federal Savings	ABC Bank	NaN	NaN	NaN	Chicago
5	Chase Bank	Gold Coast Bank	PNC Bank	5/3 Bank	Coinstar	TCF Bank	Chase Bank	Coinstar	North Community	Bank of	Chicago

Fig 4.2

In Fig 4.2, 1st is the bank which was topmost result by Foursquare for that location. Similarly, 2nd, 3rd and so on till 10th were results following 1st in the proper order of distance. 1st the closest bank in that location while 10th is the farthest banks in that location.

Similar methodology is for New York state bank search results too.

4.3 Clustering Methodology

This project clusters regions in terms of density of banks or number of banks in that region. Different methodologies have been used for both Illinois and New York State because of standard search result. Also, considering New York is large and has a greater number of places that is 306 as compared to 200 only in Illinois.

Illinois:

```
# set number of clusters
final_count_data["cluster"]=None
for index, row in final_count_data.iterrows():
    if row[5]>=50:
        final_count_data['cluster'].iloc[index]=1
    elif row[5]>=10:
        final_count_data['cluster'].iloc[index]=2
    elif row[5]>=5:
        final_count_data['cluster'].iloc[index]=3
    elif row[5]>=2:
        final_count_data['cluster'].iloc[index]=4
    elif row[5]>=0:
        final_count_data['cluster'].iloc[index]=5
```

If in any region, number of bank search results is more than 50 it is cluster 1 or highly dense area, if number is between 50 and 10 then mild dense area, if between 10 and 5 then mild low dense area, if between 5 and 2 then low dense region and if only 1 banks then very low dense regions. If in map there is no marker at all in Illinois, then there is probably no bank search results in those regions for that search locations within 500m range.

New York:

New York has a greater number of areas and is comparatively more spread out in terms of number of banks. If region has greater than 12 banks in 500m range then it's cluster 1 or highly dense region, if between 12 and 10 then mild dense region and falls in cluster 2. If bank's number is between 10 and 5 then the region would be in cluster 3 or mild low dense region. If bank's number in the region is between 5 and 2 in 500m range of the search location, then it is in cluster 4 or low dense region and if there is only 1 bank in 500m range of the search location then it's very low dense region. In New York state map, if there is no marker then there are no banks in 500m range of those search locations.

```
# set number of clusters
final_count_data1["cluster"]=None
for index, row in final_count_data1.iterrows():
    if row[4]>=12:
        final_count_data1['cluster'].iloc[index]=1
    elif row[4]>=10:
        final_count_data1['cluster'].iloc[index]=2
    elif row[4]>=5:
        final_count_data1['cluster'].iloc[index]=3
    elif row[4]>=2:
        final_count_data1['cluster'].iloc[index]=4
    elif row[4]>=0:
        final_count_data1['cluster'].iloc[index]=5
```

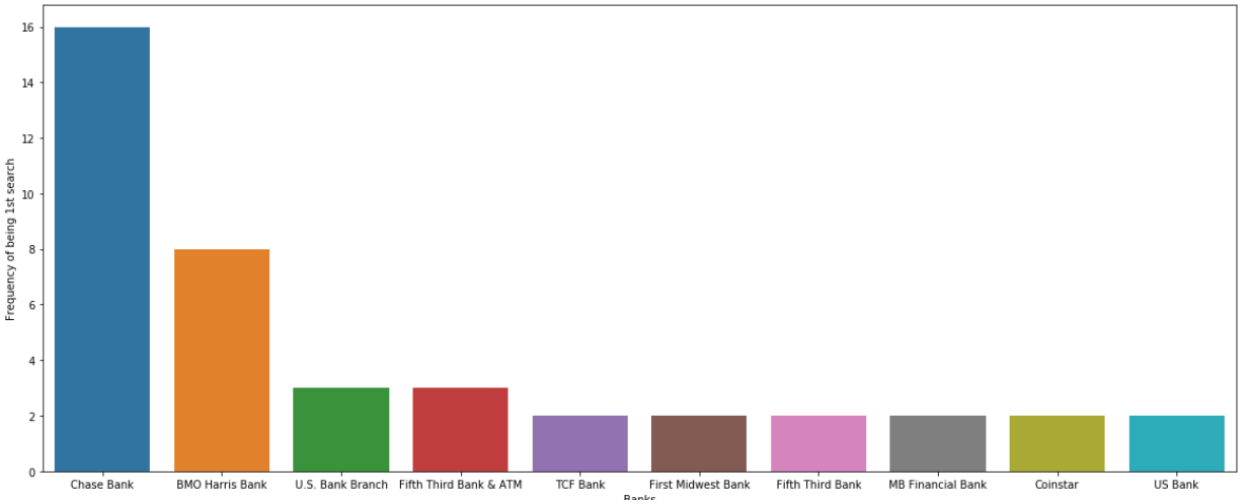
5. Results

There are several new and interesting results that we discovered in both New York and Illinois and particularly Chicago region. First, we will list down results for each state then compare it.

Illinois:

1. Out of 200 places in Illinois, only 81 places had bank in 500 meters range of the search location which is like only 40 percent areas has access to bank in 500 meters range.

2. Chase Bank is closest bank to the search location in 16 out of 81 places which has banks in 500m range which is like almost 20% of the region which had banks and 8 percent of the total 200 places in Illinois. Chase is followed by BMO Harris, US Bank and Fifth-Third Bank.

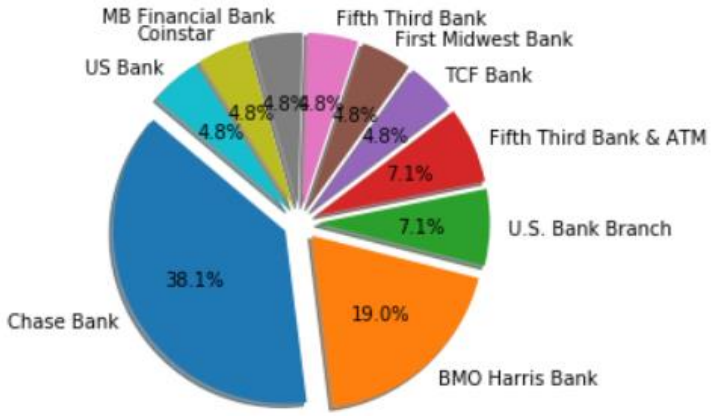


3. Out of 200 places in Illinois only 7 places had 9 or more banks in 500 m range. 10 places had 8 or more banks, 14 places had 7 or more banks, 16 places had 6 or more banks and 54 places had 2 or more banks.

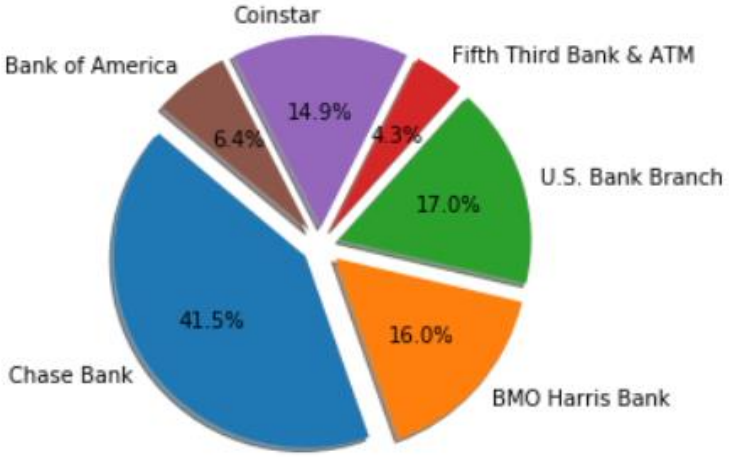
```
pd9.groupby(by='place').count().sum()
```

```
1st     81
2nd     54
3rd     42
4th     25
5th     21
6th     16
7th     14
8th     10
9th      7
10th     7
dtype: int64
```

4. If we consider only top 10 banks in Illinois then Chase bank dominates other banks by a huge margin with being a closest bank in 38% of the region.

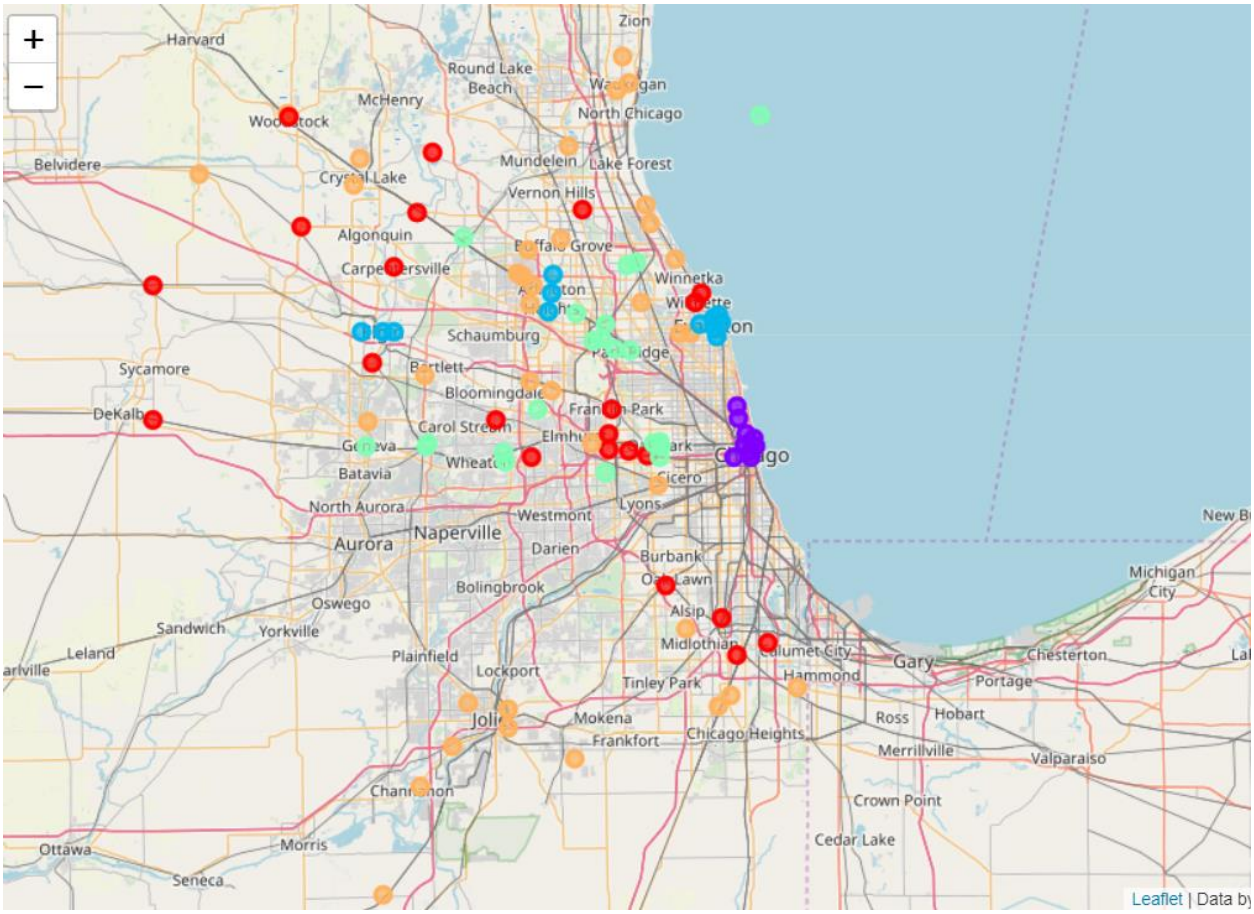


5. Chase was in top 10 closest banks in 39 out of 81 regions followed by US Bank, BMO Harris, Bank of America and Fifth Third. We have excluded Coinstar because of its status of not being a bank.



6. Overall Greater Chicago Area were much denser in terms of bank number than rest of Illinois followed by Elgin and Arlington Heights. This is the cluster of regions shown in the map:

Here Dark Blue is cluster 1 or say high dense banking areas, we can see all of them are in Chicago. Oceanic blue is mild high dense Banking areas. They are mainly Arlington Heights, Elgin and Evanston. Light Green is cluster 3 or mild low dense areas. They are Des Plaines, Mount Prospect, Geneva, Barrington and others. Light Brown are low dense bank presence areas and includes areas like Joliet, Lansing, Palatine and others. Red are very low dense areas and includes places like Cortland, Woodstock, Genoa, Lombard and other.

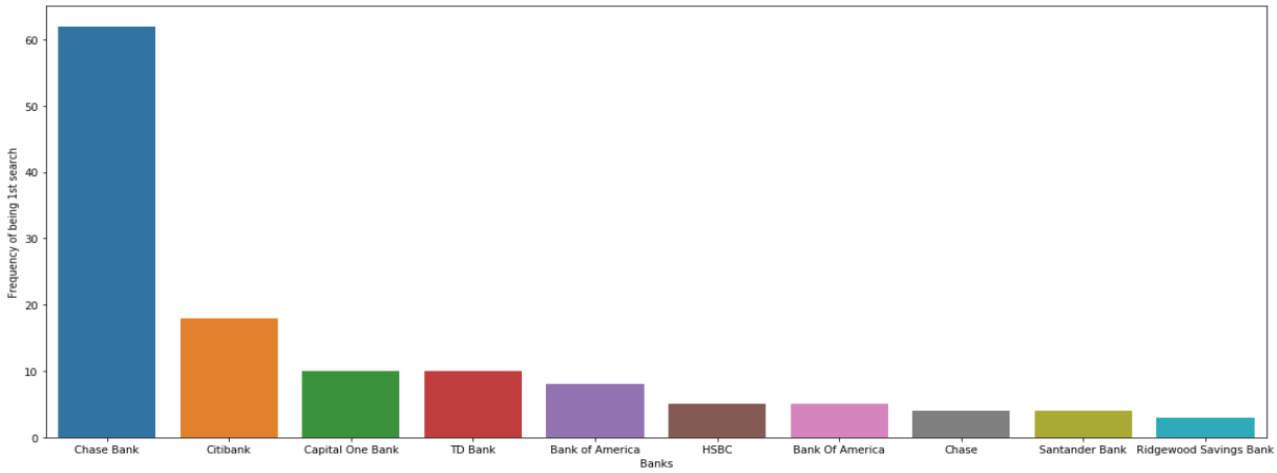


New York:

1. Out of 306 places in Illinois, only 172 places had bank in 500 meters range of the search location which is like only 56 percent areas has access to bank in 500 meters range.

2. Chase Bank is closest bank to the search location in 62 out of 172 places which has banks in 500m range which is like almost 36% of the region which had banks and 20 percent of the total 306 places in Illinois. Chase is followed by Citi Bank, Capital One Bank, TD Bank and Bank of America.

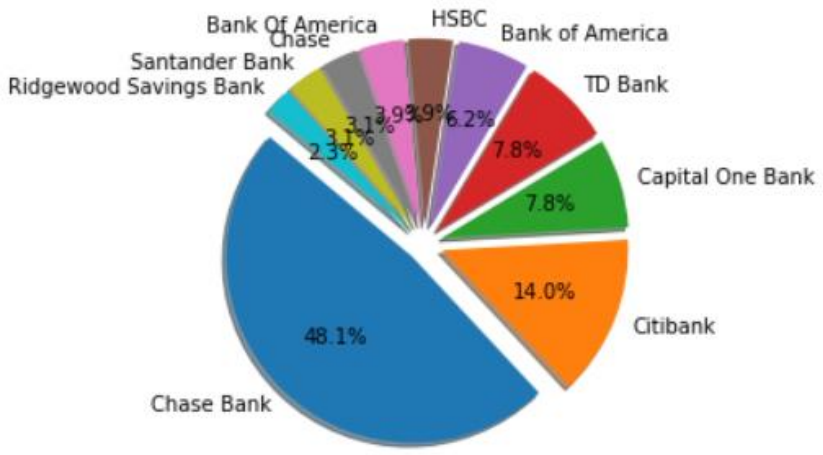
```
<function matplotlib.pyplot.show(*args, **kw)>
```



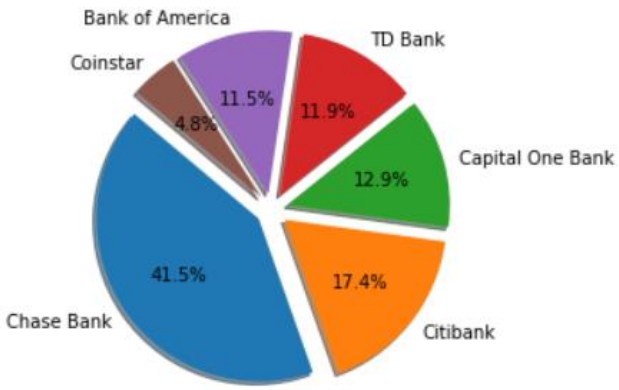
3. Out of the 172 locations in New York, we have 2 locations where we have 15 banks, 27 locations where there are 10 banks, 17 places having 10 banks and so on. There are 30 places where we have only one banks can be very good chances for banks to open their branch at those locations.

count	County	Place	latitude	longitude
1	30	30	30	30
2	18	18	18	18
3	22	22	22	22
4	14	14	14	14
5	18	18	18	18
6	8	8	8	8
7	9	9	9	9
8	10	10	10	10
9	17	17	17	17
10	27	27	27	27
15	2	2	2	2

4. If we consider only top 10 banks in New York then Chase bank dominates other banks by a huge margin with being a closest bank in 48% of the region. Chase is followed by Citi, Capital One, TD and Bank of America.

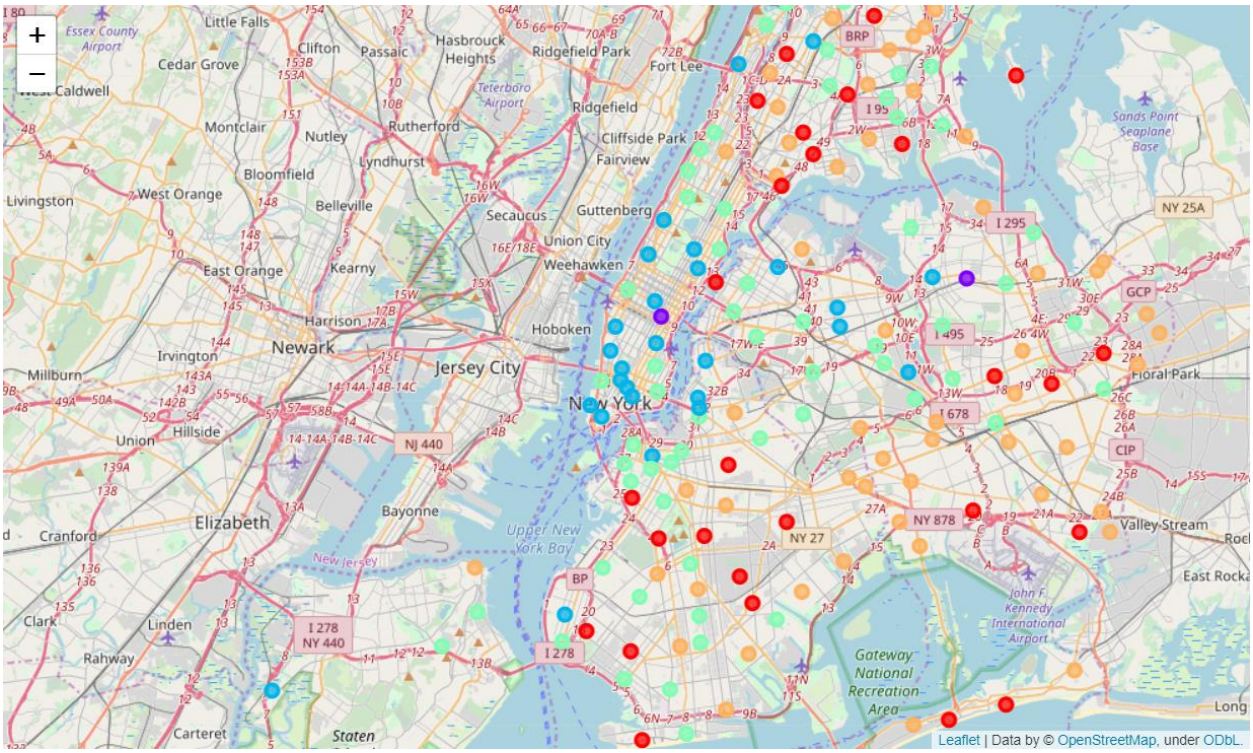


5. Chase was in top 10 closest banks in 172 out of 172 regions followed by Citi Bank, Capital One Bank, TD bank and Bank of America. We have excluded Coinstar because of its status of not being a bank.



6. Overall New York State has banks evenly spread out and not concentrated at any specific points. This is the cluster of regions shown in the map:

Here Dark Blue is cluster 1 or say high dense banking areas, we can see all of them are in Murray Hill. Oceanic blue is mild high dense Banking areas. They are areas in New York City like China Town, Lincoln Square. Light Green is cluster 3 or mild low dense areas. They are Manhattan Valley, Rego Park, Woodside and others. Light Brown are low dense bank presence areas and includes areas like Crown Height, Kensington and others. Red are very low dense areas and includes places Bensonhurst, Brookville and South Ozone Park.



6. Discussion

We have few limitations:

1. The search is done on foursquare. The result may have very high probability of being correct but not 100 percent correct
2. Search is done at specific location and doesn't represent any location completely
3. Same banks may operate with different names; I have considered names of banks as it is returned by the Foursquare
4. To decide if banks should be opened at any low denser banking or no banking locations depends on populations at that locations and analysis needs to be done on population.
5. A person might not choose a bank in terms of accessibility only. Other factors like interest rate on savings, service and staffs are also important factors.

Recommendations would be to do analysis on population and how many banks are present per miles square versus density of number of persons per miles square. Also, locations at every 500 m regions needs to be chosen to find correct study of banks presence in those regions.

7. Conclusions

Following are the conclusion of this project:

1. Chase Bank dominates in both Illinois and New York with little strong presence in New York
2. Bank of America is in top 5 banks in both Illinois and New York but is strongly present in either of the region. Maybe they need to make their presence better when compared to its competitors
3. There are many places in Illinois and New York where there were no banks in 500 meters region of the search location and might be a good option of location if any bank is looking to expand but proper study of population and demographics needs to be done.
4. If a person is in Illinois these are the best options: Chase, BMO Harris, US Bank, Fifth third Bank and Bank of America.
5. If a person is in New York, Chase, Citi Bank, Capital One Bank, TD Bank and Bank of America are the best options.

8. References

Final file having all banks name for all 81 regions in Illinois can be found here:

https://drive.google.com/file/d/1owaUuXVQ259W62zVNoYDEgxBP2plJeYL/view?usp=sharing

Final file having all banks for all 172 regions in New York can be found here.

https://drive.google.com/file/d/1eo_BKBkR7hbKN_1Ct3MQwzdDVmLE8NWg/view?usp=sharing

The analysis notebook can be found at following link:

https://nbviewer.jupyter.org/github/deepak2025/Coursera_Capstone/blob/master/Final_Capstone_Project.ipynb