Capstone Project

Kinfe Woldeyohannes

THE BATTLE OF NEIGHBORHOODS: Coffee shop in NYC

1. Introduction/Business Problem

This study will explore the best coffee shop locations throughout the city of New York. New York City is a city that needs no introduction. It is located in New York state with a population of over 8.34 million (census.gov, 2020) which makes it the most densely populated city in United States. The global financial center index ranked the city as the world's top financial center in 2019 (World Atlas). The city is famous for food and culture and has many public and private schools, bars, restaurants, coffee shops, recreational parks, etc.

Business Problem

You probably have heard how hectic life is in the city and overwhelming at times so a nice cup of coffee can help us unwind. But, where do we find the best coffee shops or cafe in the city to

do just that? Also, let's stallholders from a west coast wants to expand their coffee business to east coast and have chosen New York City due to population and diversity. Where would be the ideal place for the stakeholders to launch coffee shop in the city? This report will explore which boroughs and/or neighborhood of New York City have the best coffee shop and where in the city should the stakeholder open coffee shop business.

2. Data

In order to analyze and solve the above business problems, data on New York City neighborhoods, boroughs, neighborhood boundaries, latitude, longitude and coffee shops along their ratings and likes are required.

Neighborhoods data: Data provided by the coordinators of "IBM Data Science Professional Certificate" will be utilized to process New York City neighborhoods, boroughs, latitude and longitude. This dataset will be processed, cleaned, and then read it into a pandas datagram so that it is in a structured format.

Venues data: Once the neighborhoods data is preprocessed into a data frame, all data that describes the top venues like restaurants, coffee shops, ratings and likes will be retrieved from Foursquare website via the Request library in Python. The best coffee shops will be then analyzed using Panda libraries.

Neighborhood boundaries: New York city data containing neighborhood boundaries will be obtained from NYC open data source:

https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm.

Finally I will analyze further and see how many coffee shops there are in each neighborhood and borough and then graph and present the results to the stallholders to make decision.

Target Audience:
Tea and coffee are the first and second most consumed drinks in the world, respectively. Therefore, Individual and stakeholders who are passionate about drinking coffee will benefit from this study. The project is also for business owners and stakeholders who want to expand their businesses and wonder how data science could be applied to the questions at hand. Not only stakeholders and business owners could benefit from this but also existing restaurants who are interested in adding coffee to their menu can make use of this project as well.

3. Methodology:

We'll start by loading the necessary libraries to import the New York City neighborhood data. All the relevant data is in the *features* key, which is a list of the neighborhoods for each boroughs.

Neighborhood Analysis

Load New York City neighborhood data.

We have NY City neighborhoods along thier geopraphical coordinates.

```
In [9]:
1 display(ny_data.shape)
2 ny_data.head()
```

```
(306, 4)
```

Out[9]:

	Borough	Neighborhood	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
4	Bronx	Riverdale	40.890834	-73.912585

How many neightborhood are there in each borough?

```
In [10]:
1 ny_data.groupby('Borough')['Neighborhood'].count()
```

```
Out[10]: Borough
Bronx          52
Brooklyn       70
Manhattan      40
Queens         81
Staten Island  63
Name: Neighborhood, dtype: int64
```

Fig.1 New York City neighborhood and coordinates

Queens has the most neighborhoods. Let's plot the data.

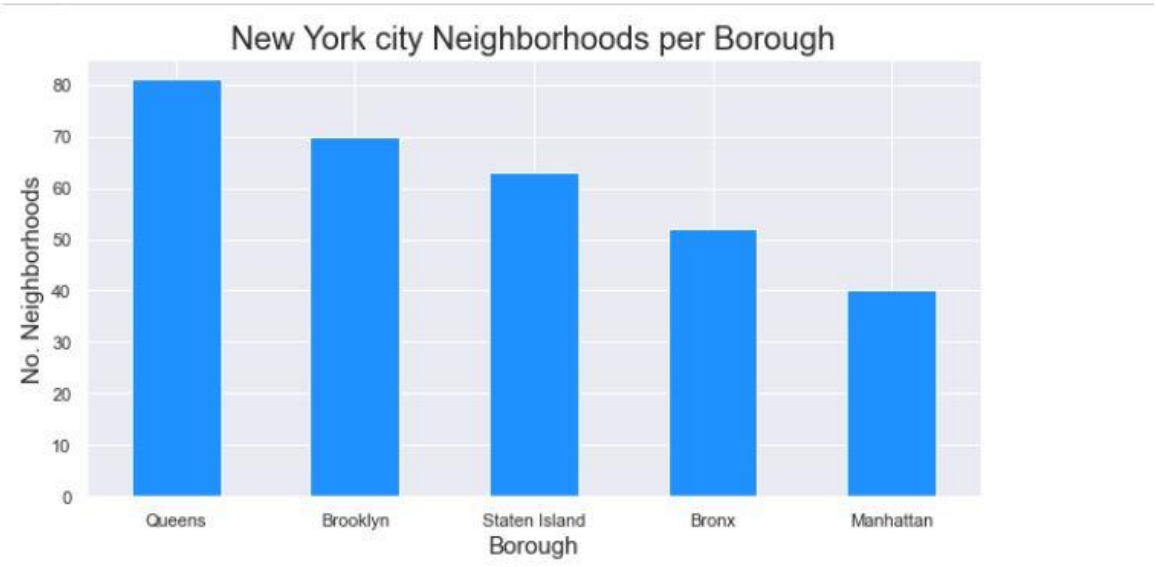

Fig.2 New York City neighborhood per each borough

Let's continue to analyze and see how many coffee shops there are in each neighborhood and borough using the Foursquare API. First we define a function to interact with Foursquare API to get the top 100 venues within a radius of 500. The function will return borough, Neighborhood, ID and Name of the category. Since the study is about coffee shops in New York City, the category will focus on coffee shops only.

#set variables
LIMIT=100
# define radius
radius = 500

coffee_shops = venues[venues['Category']=='Coffee Shop']

The Foursquare API retrieved more than 300 coffee shops in the in New York City.

```
In [16]:
1 ny_coffee_shops.tail()
2
```

Out[16]:

	Borough	Neighborhood	ID	Name
313	Bronx	Kingsbridge Heights	4b71c776f964a520fb5b2de3	Perista
314	Bronx	Kingsbridge Heights	4c483d70c047be9a430c3379	Starbucks
315	Bronx	Kingsbridge Heights	5156f35be4b0dc046e1af723	Perista Coffee Shop
316	Manhattan	Hudson Yards	5c683caaf96b2c00397a8609	Oslo Coffee Roasters
317	Manhattan	Hudson Yards	526d9114498ec8efda93fcce	Romeo and Juliet Coffee

How many coffee shops are there in New York City?

```
In [17]:
1 display(ny_coffee_shops.shape)
2 ny_coffee_shops.head()
3
```

```
(318, 4)
```

Fig.3 New York City coffee shops data frame

Find out how many coffee shops are in each borough and neighborhood?

Total number of coffee shops per borough

```
In [18]:
1 ny_coffee_shops.groupby('Borough')['Name'].count().sort_values(ascending=False)
```

```
Out[18]: Borough
Manhattan        150
Brooklyn          96
Queens            38
Staten Island     17
Bronx             17
Name: Name, dtype: int64
```

Fig.4 New York City coffee shops per borough

Even though Manhattan has the fewest neighborhood, half of the coffee shops are located in Manhattan. State Island and Bronx have the least number of coffee shops.

```
In [21]:
Num_of_neigh = 15 # top number for graphing all the same past 13
ny_coffee_shops.groupby('Neighborhood')['Name'].count().nlargest(Num_of_neigh).plot.bar(figsize=(20,6), color=clr)
plt.title('New York city Coffee Shops per Neighborhood', fontsize = 20)
plt.xlabel('Neighborhood', fontsize = 15)
plt.ylabel('Coffee Shops', fontsize=15)
plt.xticks(rotation = 'horizontal')
plt.show()
```

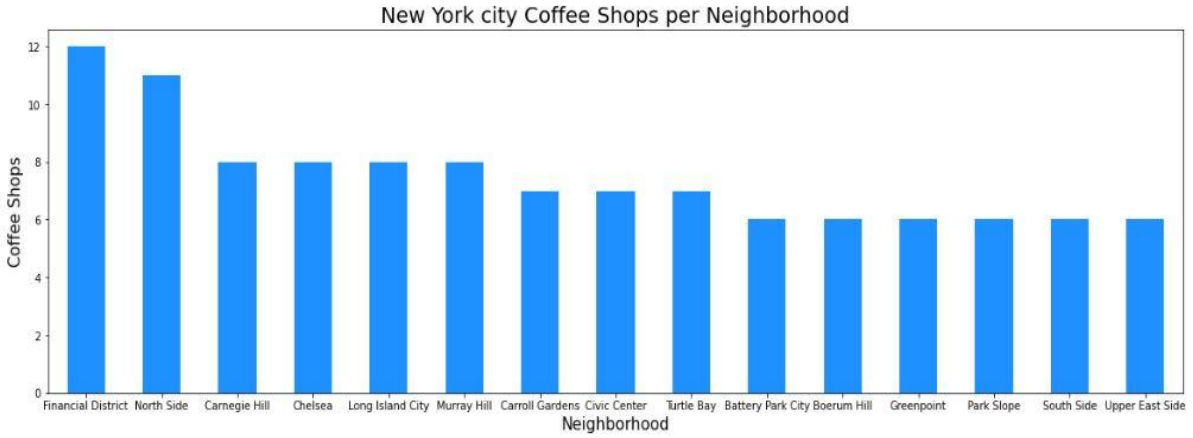

Fig.5 Top 10 neighborhoods that have the highest number of coffee shops.

As we can see from the above plot, Financial District which is located in Manhattan borough does have the highest number of coffee shop business.

Let's use Foursquare API to get the likes, rating and tips.

```
In [33]:
ny_coffee_shop_venue
```

Out[33]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Kingsbridge	5660c06b498e4003dba169a5	Mon Amour Coffee & Wine	22	8.4	7
1	Manhattan	Marble Hill	55f81cd2498ee903149fcc64	Starbucks	24	8.0	2
2	Manhattan	Marble Hill	57655be738faa66160da7527	Starbucks	7	7.2	0
3	Bronx	Norwood	4c6d5ce5e6b7b1f7af76a98e	Nicky's Coffee Shop	12	8.0	10
4	Bronx	Pelham Parkway	4bb62b326edc76b05d80301c	Liberty Donut & Coffee Shop	14	7.0	8
5	Bronx	West Farms	0	0	0	0.0	0
6	Bronx	Throgs Neck	58b1b3e6076be15eeba26a3f	The Miles Coffee Bar	13	8.0	6
7	Bronx	Van Nest	4c1c5630e9c4ef3b4ccd45aa	Conti's Pastry Shoppe	47	8.7	15
8	Bronx	Morris Park	4c743793b474a1cd3bf5b5bf	La Casa Del Caffe	16	7.5	8
9	Bronx	Belmont	53a0be61498eeec38e5112e0	Starbucks	68	7.3	11
10	Bronx	Belmont	5cb4b7c5fdb9a7002c70843d	Starbucks	0	7.1	0
11	Bronx	North Riverdale	4b8d1887f964a520aae732e3	Noni's Coffee Shop	8	6.9	5

Fig.6 New York City coffee shops rating and likes data frame

As you can see from the fig 6 above, these are Foursquare API returned venues that have 0 as their ID. Let's remove coffee shops that have 0 IDs.

```
In [35]:
1 ny_coffee_shop_venue_filt
2 ny_coffee_shop_venue_filt = ny_coffee_shop_venue_filt[ny_coffee_shop_venue_filt["ID"] !='0']
3 ny_coffee_shop_venue_filt.sort_values(by ='Rating', ascending = False, inplace = True)
4 ny_coffee_shop_venue_filt.head()
```

Out[35]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
48	Brooklyn	Carroll Gardens	58d933702f91cb026f478e38	East One Coffee Roasters	244	9.1	36
76	Brooklyn	Greenpoint	54d43863498e653d2ab8343f	Early	129	9.0	30
18	Brooklyn	Greenpoint	576eb1f5cd10a371033f7ad6	Maman	179	8.9	36
15	Brooklyn	Greenpoint	518cf9ec498e8c38bda0268d	Homecoming	217	8.9	45
49	Brooklyn	Carroll Gardens	5d7ce692b4fc9c0008394021	Hungry Ghost	20	8.9	0

```
In [36]:
1 ny_coffee_shop_venue_filt.groupby('Borough')['Name'].count()
```

```
Out[36]: Borough
Bronx        13
Brooklyn     46
Manhattan     2
Name: Name, dtype: int64
```

Fig.7 New York City coffee shops after cleaning

Due to limited premium calls, the Foursquare API retrieved 13 coffee shops from Bronx, 46 from Brooklyn and 2 from Manhattan after few trials.
Let's work on these data.

```
4 g.set(xlabel= 'Coffee Shops' , ylabel = 'Neighborhood' )
```

```
Out[37]: [Text(0.5, 0, 'Coffee Shops'), Text(0, 0.5, 'Neighborhood')]
```

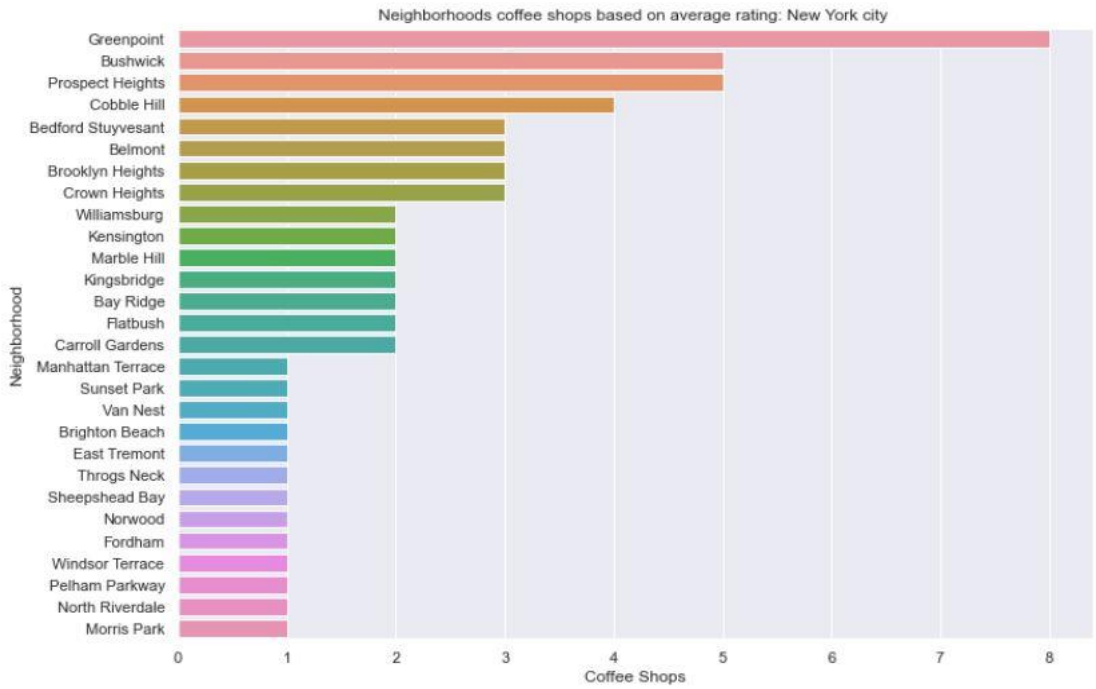

Fig 8. Coffee shops by Neighborhood

Since we have the coffee shops ratings, let's find out which coffee shop has the best rating

```
In [47]:
## Coffee shop with maximum Ratings
ny_coffee_shop_venue_filt.sort_values(by='Rating',ascending = False).head(10)
```

Out[47]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
48	Brooklyn	Carroll Gardens	58d933702f91cb026f478e38	East One Coffee Roasters	244.0	9.1	36.0
76	Brooklyn	Greenpoint	54d43863498e653d2ab8343f	Early	129.0	9.0	30.0
18	Brooklyn	Greenpoint	576eb1f5cd10a371033f7ad6	Maman	179.0	8.9	36.0
15	Brooklyn	Greenpoint	518cf9ec498e8c38bda0268d	Homecoming	217.0	8.9	45.0
49	Brooklyn	Carroll Gardens	5d7ce692b4fc9c0008394021	Hungry Ghost	20.0	8.9	0.0
99	Brooklyn	Prospect Heights	5d920abe452d400008d5dc69	Ciao, Gloria	40.0	8.9	9.0
41	Brooklyn	Brooklyn Heights	59d7872e5a2c911745c2e8ca	Joe Coffee	95.0	8.8	13.0
7	Bronx	Van Nest	4c1c5630e9c4ef3b4ccd45aa	Conti's Pastry Shoppe	47.0	8.7	15.0
28	Brooklyn	Prospect Heights	4a2407e6f964a520ef7d1fe3	Sit & Wonder	305.0	8.7	117.0
29	Brooklyn	Prospect Heights	5a6cc216f5e9d763b7a37c8f	Gran Caffe De Martini	27.0	8.7	7.0

Fig 9. Coffee shops list based on rating

Groupby the coffee shops based on average ratings. Carroll Gardens in Brooklyn is the winner here.

```
In [49]:
ny_neighborhood_coffee_shops=ny_coffee_shop_venue_filt.groupby('Neighborhood',as_index=False).mean()[['Neighborhood','Ra
ny_neighborhood_coffee_shops.columns=['Neighborhood','Average_Rating']
ny_neighborhood_coffee_shops.sort_values(['Average_Rating'],ascending=False).head()
```

Out[49]:

	Neighborhood	Average_Rating
6	Carroll Gardens	9.0000
25	Van Nest	8.7000
27	Windsor Terrace	8.6000
12	Greenpoint	8.5125
21	Prospect Heights	8.4600

Fig 10. The top 5 Coffee shops based on customer rating

```
In [46]:
# Coffee shop with maximum Likes
ny_coffee_shop_venue_filt.sort_values(by='Likes',ascending = False).head(10)
```

Out[46]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
42	Brooklyn	Brooklyn Heights	4de26e06814df7ebdc14af8b	Vineapple Cafe	465.0	8.4	151.0
28	Brooklyn	Prospect Heights	4a2407e6f964a520ef7d1fe3	Sit & Wonder	305.0	8.7	117.0
45	Brooklyn	Cobble Hill	55423f82498e5156ac21cfb6	Swallow Cafe	253.0	7.8	47.0
48	Brooklyn	Carroll Gardens	58d933702f91cb026f478e38	East One Coffee Roasters	244.0	9.1	36.0
15	Brooklyn	Greenpoint	518cf9ec498e8c38bda0268d	Homecoming	217.0	8.9	45.0
18	Brooklyn	Greenpoint	576eb1f5cd10a371033f7ad6	Maman	179.0	8.9	36.0
43	Brooklyn	Brooklyn Heights	4acf7fb1f964a5201dd420e3	Starbucks	170.0	6.7	49.0
20	Brooklyn	Greenpoint	55cf6244498e24c4ebe3ae0b	Sweetleaf	170.0	8.5	24.0
19	Brooklyn	Greenpoint	4d8f7d48d265236a02c72117	Upright Coffee	159.0	7.8	49.0
21	Brooklyn	Brighton Beach	4b59aedcf964a520999228e3	Starbucks	129.0	6.4	37.0

Fig 11. Coffee shops list based on likes

Plot each neighborhoods based on average rating

```
In [50]:
ny_borough_coffee_shops=ny_coffee_shop_venue_filt.groupby('Borough',as_index=False).mean()[['Borough','Rating']]
ny_borough_coffee_shops.columns=['Borough','Average_Rating']
ny_borough_coffee_shops.sort_values(['Average_Rating'],ascending=False).head()
```

Out[50]:

	Borough	Average_Rating
1	Brooklyn	7.819565
0	Bronx	7.692308
2	Manhattan	7.600000

Fig 12. Borough average rating

Plot each neighborhoods based on average rating

```
In [56]:
g = sns.catplot(data = ny_neighborhood_coffee_shops_stats, y='Neighborhood', x='Average_Rating', kind='bar',aspect=2.4)
g.set(title='Neighborhoods coffee shops based on average rating: New York city ')
```

Out[56]: <seaborn.axisgrid.FacetGrid at 0x1abd98ee588>

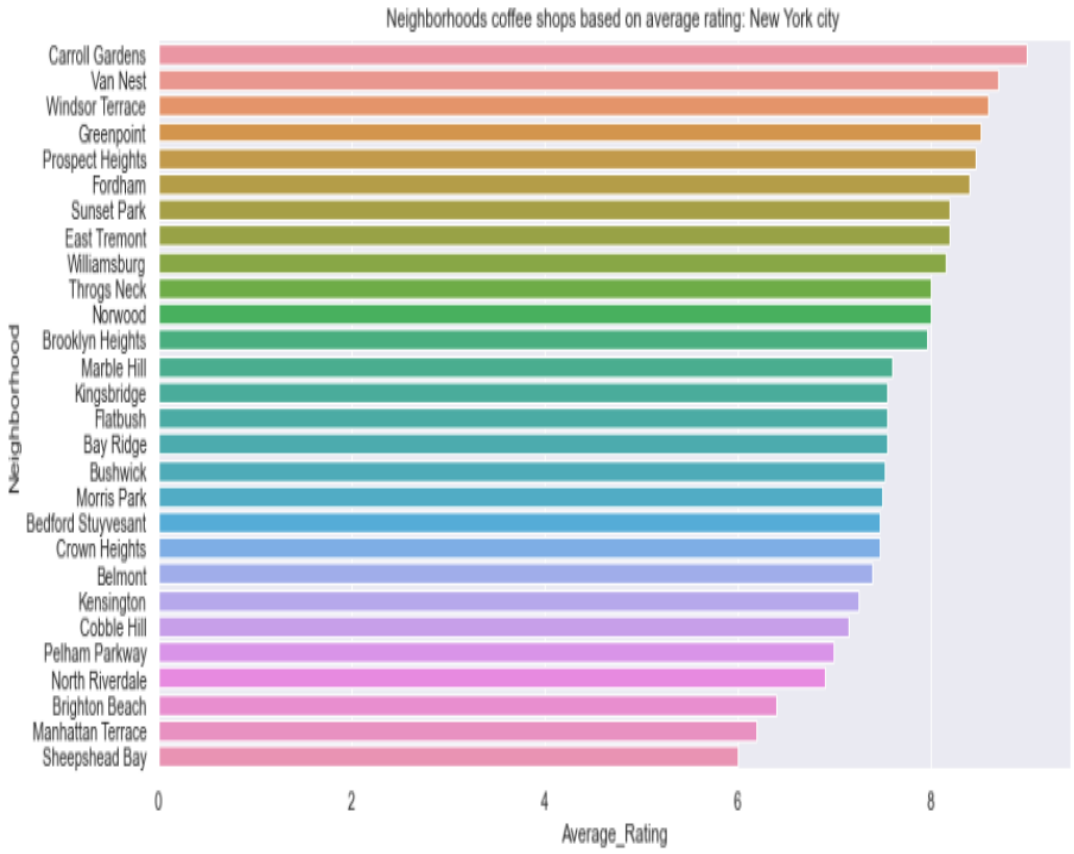

Fig 13. Neighborhood average rating

Let’s locate the coffee shop, Carroll Garden on the New York City map

```
In [63]:
1 ny_neighborhood_coffee_shops_stats['Label']=ny_neighborhood_coffee_shops_stats['Borough']+', '+ny_neighborhood_coffee_sh
2
3 # add pop-up text to each marker on the map
4 for lat, lng, label in ny_neighborhood_coffee_shops_stats[['Latitude','Longitude','Label']].values:
5     folium.Marker([lat, lng], popup=label).add_to(ny_map)
6 # add incidents to map
7 ny_map.add_child(incidents)
```

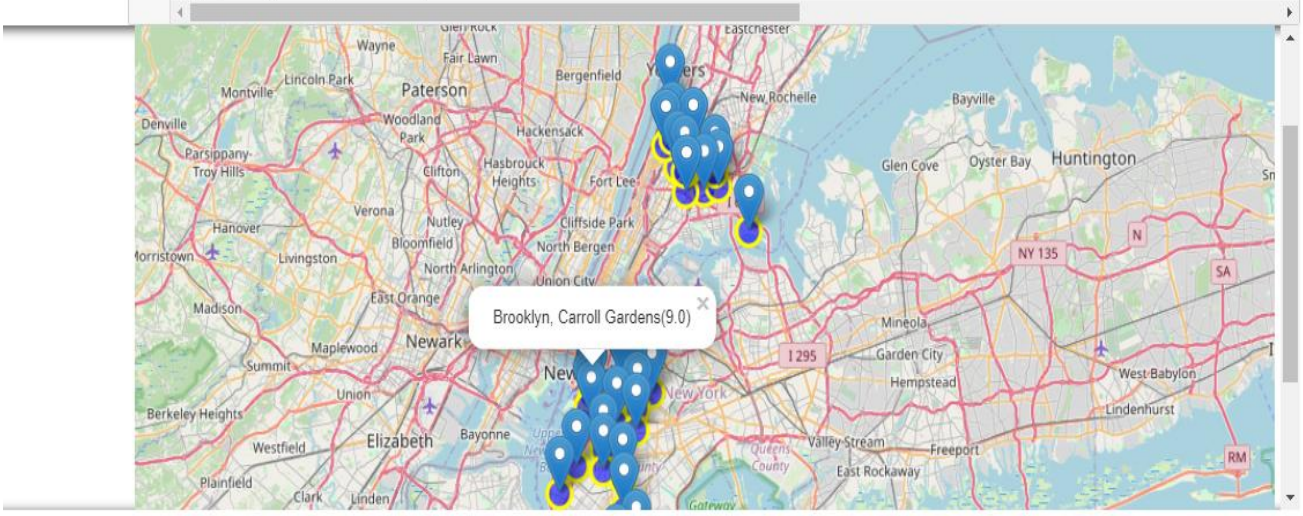

Fig 14. Neighborhoods map

Analyze Clusters

Using KMeans algorithm, let’s segment and cluster the coffee shops business and see how similar the neighborhoods are.

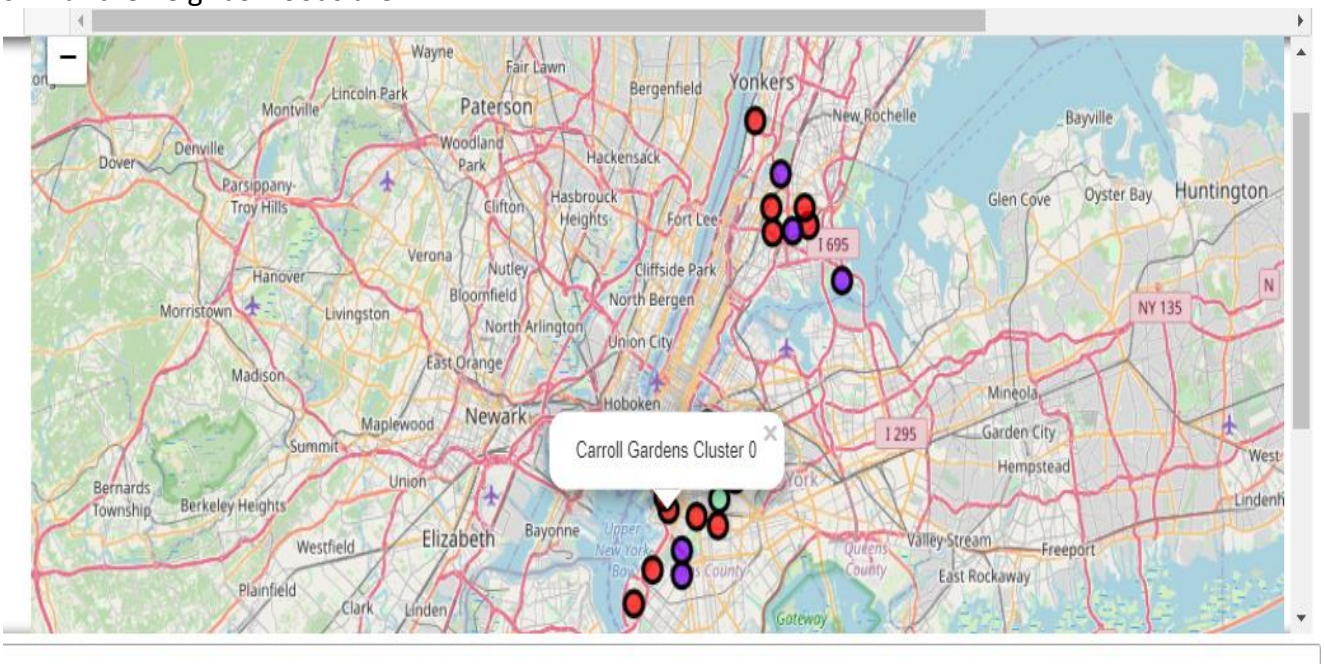

Fig 15. Show Carroll Garden cluster

Odd fox and Coffee RX are the most common coffee shops

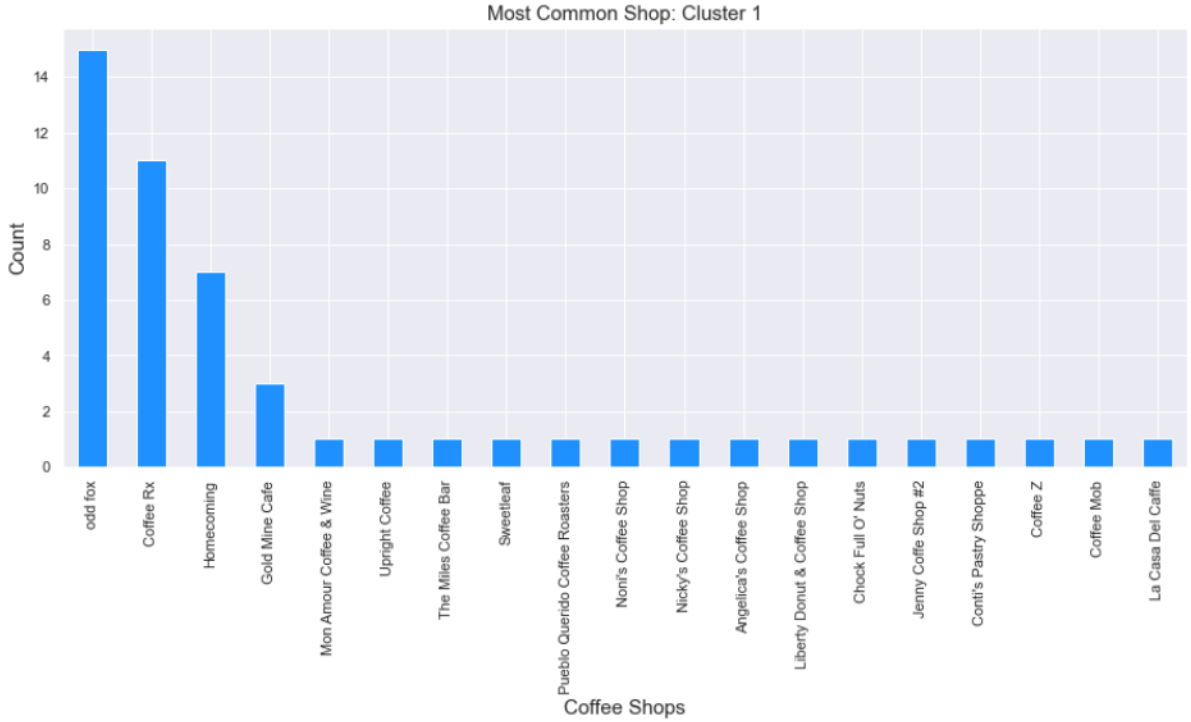

Fig 16. Most common coffee shop

4. Results / Discussion

According to the Foursquare API, Manhattan has 150 out of 318 coffee shops and Brooklyn has about 96 coffee shops. Thus, 77.35 % of all NYC coffee shops are located in Manhattan and Brooklyn alone making Bronx, Staten Island and Queens less competitors for coffee business.

Even though Queens has most neighborhoods in all of NYC, no data was retrieved from Queens and Staten Island due to foursquare limit premium calls. Based on the available customer average rating, Brooklyn and Bronx have the best rated coffee shops in New York City. These coffee shops are East One Coffee Roasters in Carrol Gardens and Early coffee shop in Green point. Over all, Green point has better rated coffee shops.

5. Conclusion

Finding the best coffee shop is relatively easy and less risky. However, finding the best location to start a business is very challenging. At times quite frustrating due to many known and unknown circumstances. Beside the quality of business's product, location has a big impact on the success of a business and should be looked at carefully.

I hope you can quickly gain meaningful insights from York City and its neighborhoods based on today's available data. I also hope that you can make use of this insight to explore other venues. Finally, as for my coffee, I would go with East One Coffee Roasters in Carroll Garden, Brooklyn since it has the best customer rating of all the NYC coffee shops.

6. Bibliography

- Neighborhood boundaries:

  https://data.cityofnewyork.us/City-Government/2020-population/t8c6-3i7b

- World Atlas :

  https://www.worldatlas.com/articles/the-world-s-top-financial-cities.html

- Census.gov:

  https://www.census.gov/quickfacts/newyorkcitynewyork