Capstone Project

Kinfe Woldeyohannes

THE BATTLE OF NEIGHBORHOODS : Week 2

Coffee shop in NYC

Content

THE BATTLE OF NEIGHBORHOODS : Week 2

1. Introduction /Business Problem

- **New York City**
  - Densely populated
  - World's top financial center in 2019
  - Famous for food and culture and has many public and private schools, bars, restaurants, coffee shops, recreational parks

THE BATTLE OF NEIGHBORHOODS : Week 2

Business Problem

This report will explore which boroughs and/or neighborhood of New York City have the best coffee shop and where in the city would be ideal for the stakeholder to open coffee shop business.

2. Data Source

In order to analyze and explore 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.

Venues data: Once the neighborhoods data is preprocessed into a dataframe, 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.

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

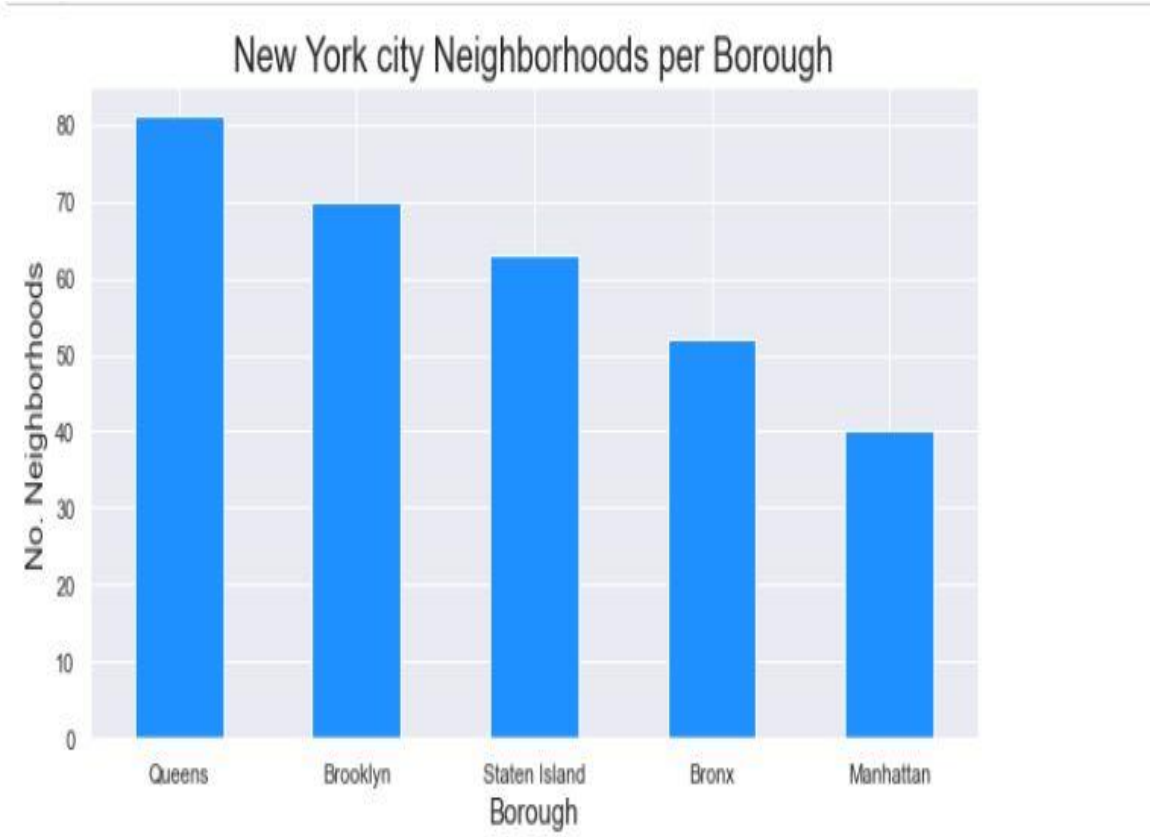
Target Audience

Tea and coffee are the first and second most consumed drinks in the world. Therefore, stakeholders and Individuals who are passionate about drinking coffee would 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 but also existing restaurants who are interested in adding coffee to their menu can make use of this project as well.

THE BATTLE OF NEIGHBORHOODS : Week 2

3. Methodology

- **New York City neighborhood data has 306 rows and 4 columns**
- **How many neighborhoods are in each borough**



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
```

THE BATTLE OF NEIGHBORHOODS : Week 2

Use Foursquare API to get venues

- **All the relevant data is in the *features* key**
- **Fetch coffee shops from the venues**
- **Get borough, neighborhood , latitude and Longitude**

In [13]:

```
# prepare neighborhood list that contains 'coffee shop'
column_names=['Borough', 'Neighborhood', 'ID','Name']
ny_coffee_shops=pd.DataFrame(columns=column_names)
count=1
for row in ny_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)

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

#     print('(',count,'/',len(ny_data),')','Coffee Shop in '+Neighborhood+', '+Borough+':'+str(len(italian_resturants)))
#     print(row)

    for cafe_shops_detail in coffee_shops.values.tolist():
        id, name , category=cafe_shops_detail
        ny_coffee_shops = ny_coffee_shops.append({'Borough': Borough,
                                                  'Neighborhood': Neighborhood,
                                                  'ID': id,
                                                  'Name' : name
                                                 }, ignore_index=True)
    count+=1
```

THE BATTLE OF NEIGHBORHOODS : Week 2

Use Foursquare API to get venues

- **All the relevant data is in the *features* key**
- **Get coffee shops from venues**
  - **Get Borough, Neighborhood , latitude and Longitude**
  - **The data has 306 rows and 4 columns**

In [13]:

```
# prepare neighborhood list that contains 'coffee shop'
column_names=['Borough', 'Neighborhood', 'ID','Name']
ny_coffee_shops=pd.DataFrame(columns=column_names)
count=1
for row in ny_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)

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

#     print('(',count,'/',len(ny_data),')','Coffee Shop in '+Neighborhood+', '+Borough+':'+str(len(italian_resturants)))
#     print(row)

    for cafe_shops_detail in coffee_shops.values.tolist():
        id, name , category=cafe_shops_detail
        ny_coffee_shops = ny_coffee_shops.append({'Borough': Borough,
                                                  'Neighborhood': Neighborhood,
                                                  'ID': id,
                                                  'Name' : name
                                                 }, ignore_index=True)
    count+=1
```

THE BATTLE OF NEIGHBORHOODS : Week 2

Use Foursquare API to get venues

- **Saved Foursquare info to a .csv for due premium calls limit**
- **Coffee shops per Borough**
  - **Financial District and North Side have more coffee shop businesses**

```
In [19]: 1 ny_coffee_shops.groupby('Neighborhood')['Name'].count().sort_values(ascending=False).to_frame()
```

Out[19]:

Neighborhood	Name
Financial District	12
North Side	11
Murray Hill	8
Long Island City	8
Carnegie Hill	8
Chelsea	8
Turtle Bay	7
Carroll Gardens	7
Civic Center	7
Park Slope	6
Greenpoint	6

```
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
```

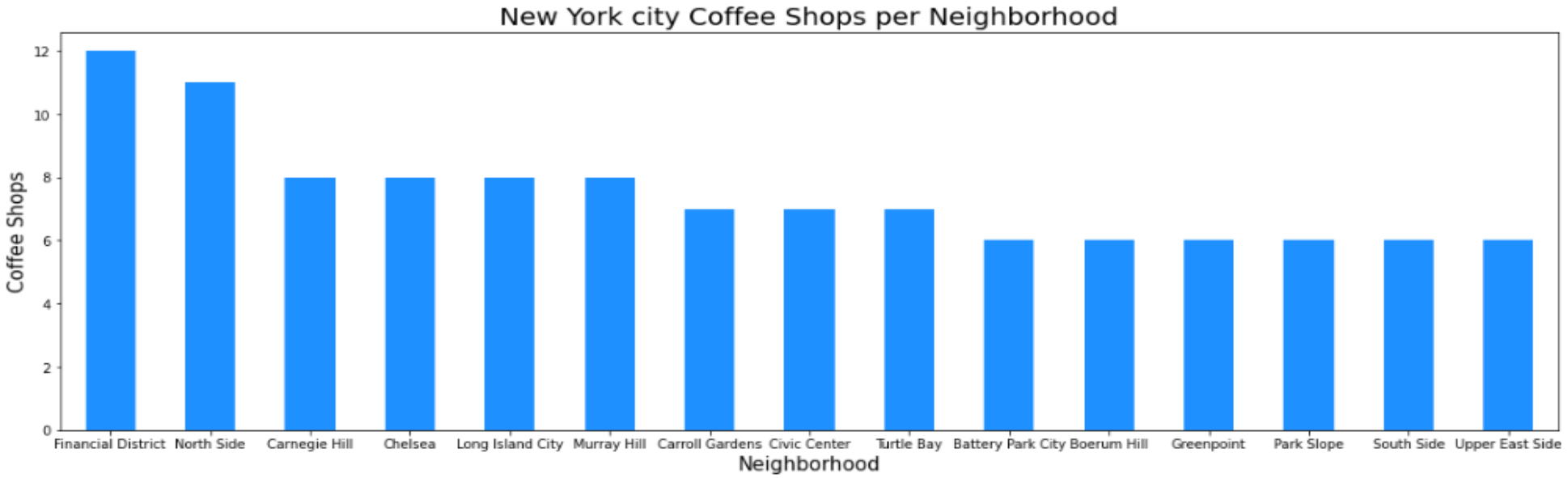
4. Result /Discussion

- **More coffee shops are located in Financial District**
- **The top 10 neighborhoods**

Lets visualize the top 10 New York city Coffee Shops

In [21]:

```
Num_of_neigh = 15 # top number for graphing all the same past 15
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()
```



THE BATTLE OF NEIGHBORHOODS : Week 2

Use Foursquare API to get Likes , Rating , Tips using Foursquare API

- **Saved Foursquare info to a .csv for due limited premium calls**
- **Clean the data by removing**
  - **'Unnamed: 0**
  - **ID = 0**

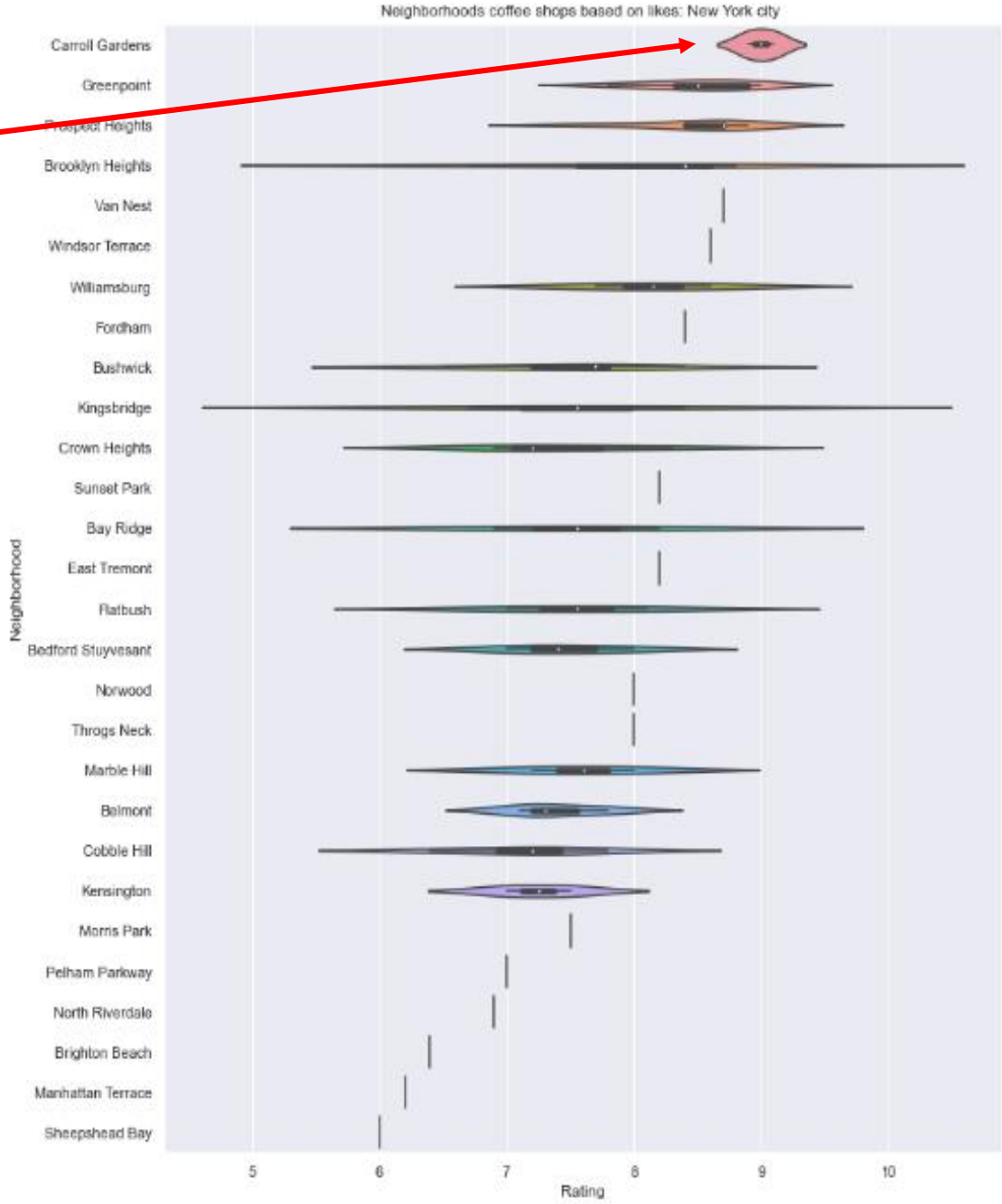
In [30]:
```
ny_coffee_shop_venue = pd.read_csv('ny_coffee_shop_venue_to_csv.csv', index_col=False)
ny_coffee_shop_venue
```

Out[30]:

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

THE BATTLE OF NEIGHBORHOODS : Week 2

- **Carroll Garden, Brooklyn, has the highest rating**
  - **Plot the neighborhood ratings**



In [60]:

```
## Coffee shop with maximum Ratings
ny_coffee_shop_venue_filt.sort_values(by='Rating',ascending = False).head(10)
df = ny_coffee_shop_venue_filt.copy()
df.head(10)
```

Out[60]:

	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
29	Brooklyn	Prospect Heights	5a6cc216f5e9d763b7a37c8f	Gran Caffe De Martini	27.0	8.7	7.0
28	Brooklyn	Prospect Heights	4a2407e6f964a520ef7d1fe3	Sit & Wonder	305.0	8.7	117.0
7	Bronx	Van Nest	4c1c5630e9c4ef3b4ccd45aa	Conti's Pastry Shoppe	47.0	8.7	15.0

THE BATTLE OF NEIGHBORHOODS : Week 2

Explore ratings, likes and tips Use Foursquare API

- **Brooklyn Heights in Brooklyn has the highest like**
- **Carroll Garden also in Brooklyn has the highest rating**

In [61]:
```
# Coffee shop with maximum Likes
ny_coffee_shop_venue_filt.sort_values(by='Likes',ascending = False).head()
```

Out[61]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
42	Brooklyn	Brooklyn Heights	4de26e0b814df7ebdc14af8b	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

In [60]:
```
## Coffee shop with maximum Ratings
ny_coffee_shop_venue_filt.sort_values(by='Rating',ascending = False).head(10)
df = ny_coffee_shop_venue_filt.copy()
df.head(10)
```

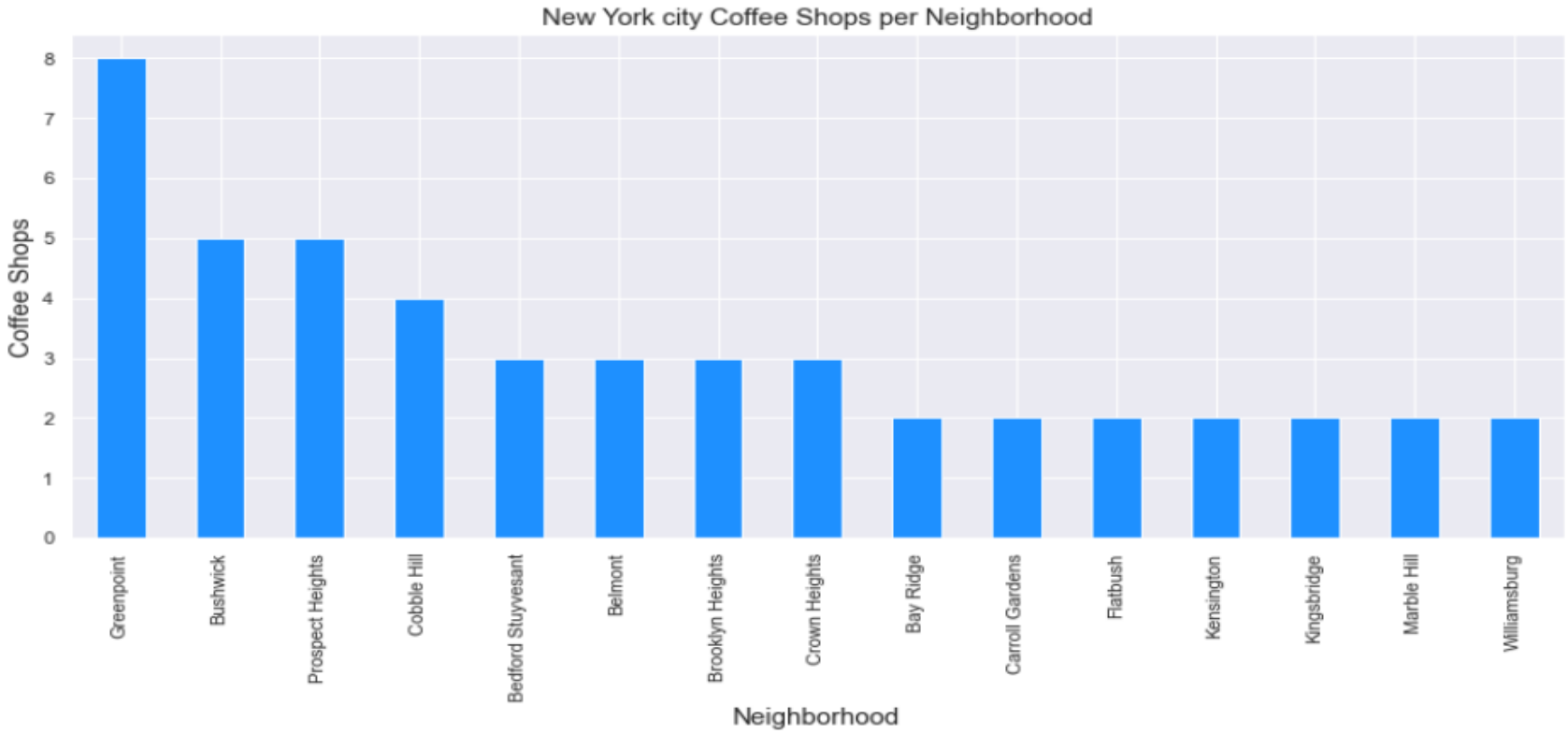
Out[60]:

	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
29	Brooklyn	Prospect Heights	5a6cc216f5e9d763b7a37c8f	Gran Caffe De Martini	27.0	8.7	7.0
28	Brooklyn	Prospect Heights	4a2407e6f964a520ef7d1fe3	Sit & Wonder	305.0	8.7	117.0
7	Bronx	Van Nest	4c1c5630e9c4ef3b4ccd45aa	Conti's Pastry Shoppe	47.0	8.7	15.0

THE BATTLE OF NEIGHBORHOODS : Week 2

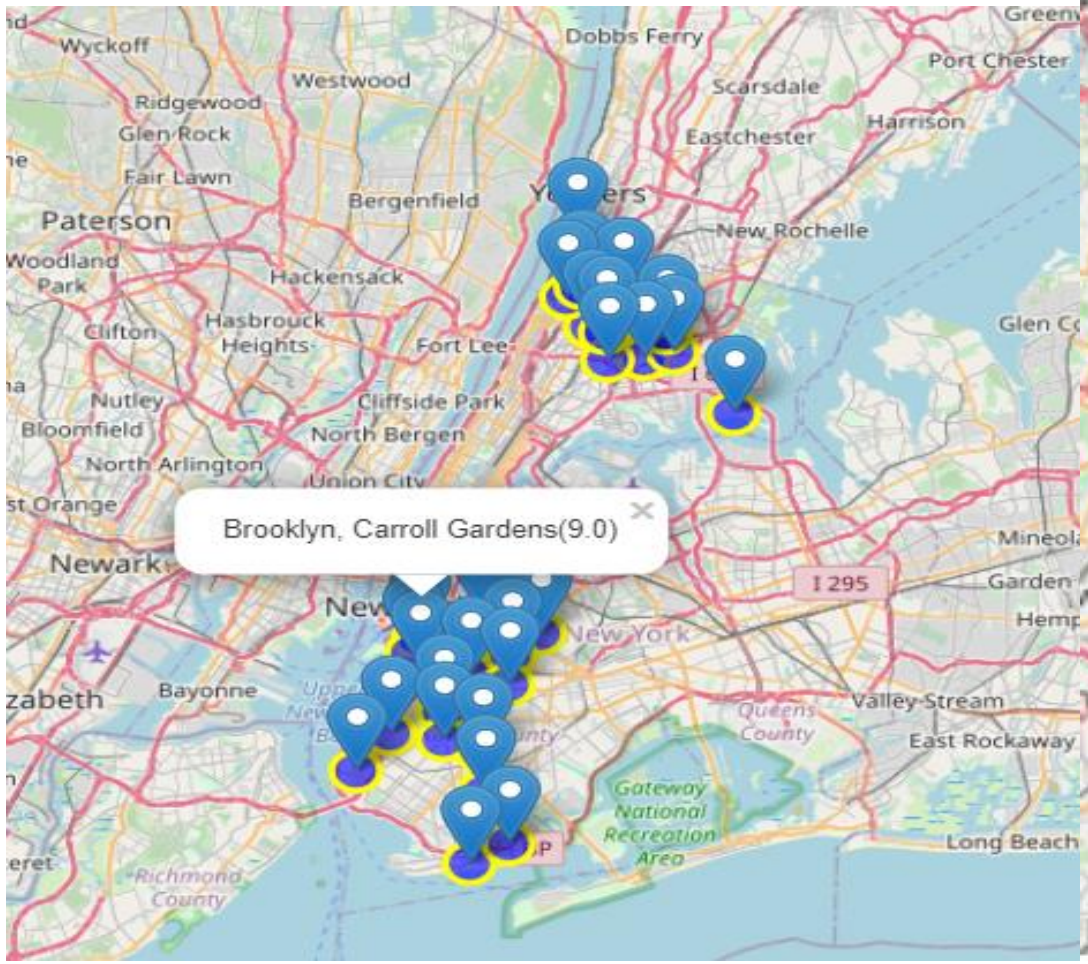
- **Carroll Garden in Brooklyn has the highest rating**
- **Brooklyn Heights which is also in Brooklyn has the highest likes**

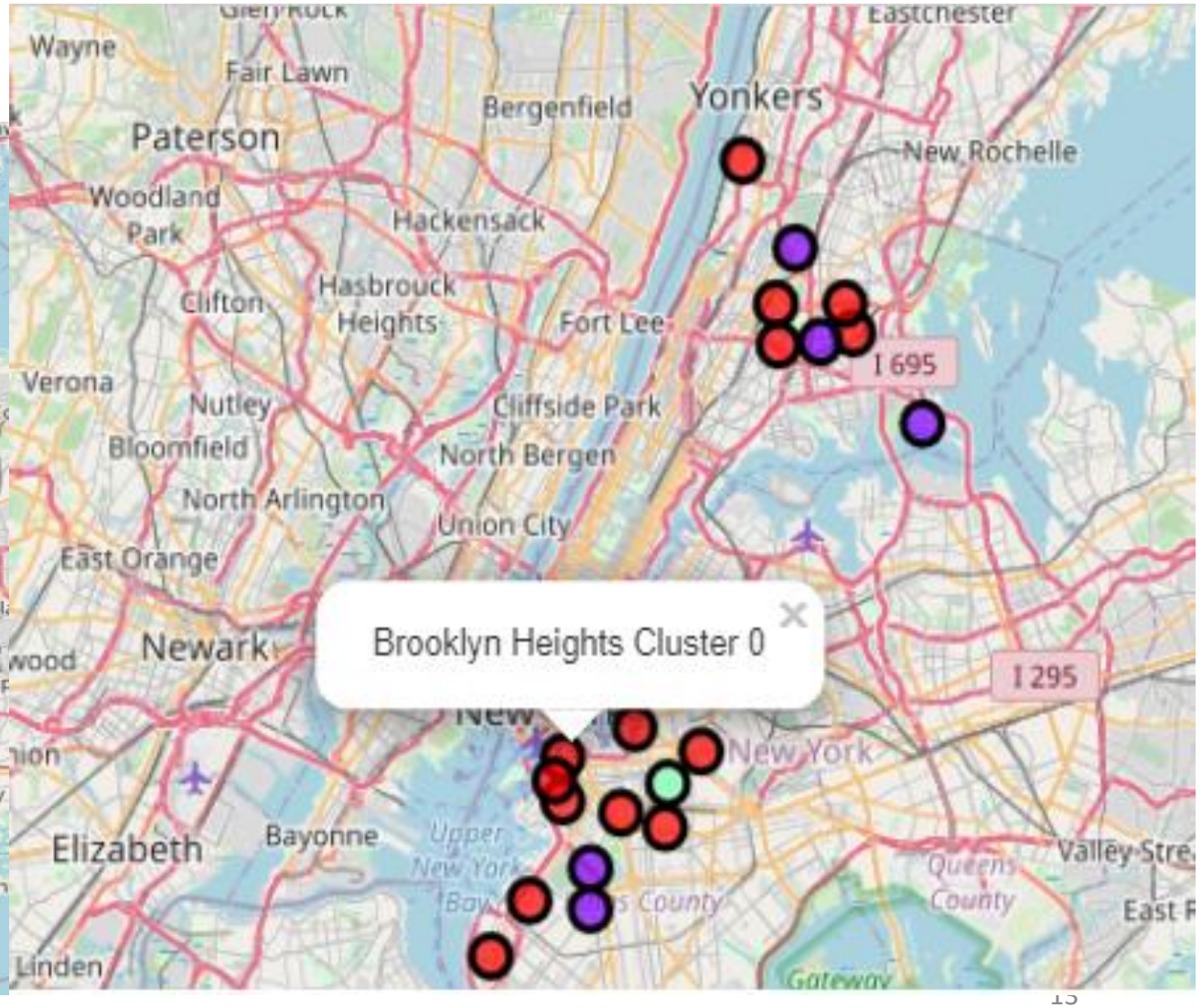
```
In [88]:
1 Num_of_neigh = 15 # top number for graphing all the same past 15
2 ny_coffee_shop_venue_filt.groupby('Neighborhood')['Rating'].count().nlargest(Num_of_neigh).plot.bar(figsize=(15,6), colo
3 plt.title('New York city Coffee Shops per Neighborhood', fontsize = 15)
4 plt.xlabel('Neighborhood',fontsize = 15)
5 plt.ylabel('Coffee Shops',fontsize = 15)
6 plt.show()
```



THE BATTLE OF NEIGHBORHOODS : Week 2

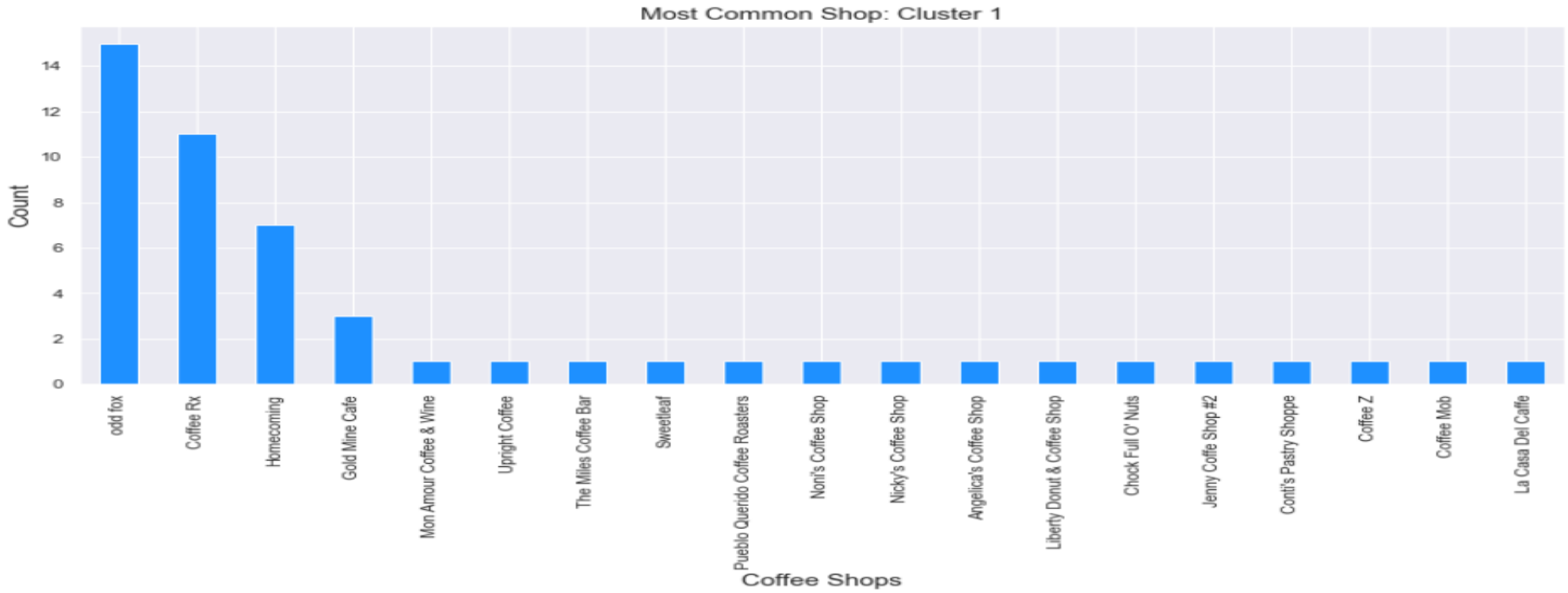
- **Show map**
  - **Carroll Garden rating and**
  - **Brooklyn Heights cluster**





THE BATTLE OF NEIGHBORHOODS : Week 2

- **Most common shops from cluster 1**
  - **Odd Fox and Coffee RX**



Conclusion

- East One Coffee Roasters in Carroll Garden has best rating.
- Foursquare API shows Green Point has more coffee shop businesses
- Finding a good business location is very challenging. It requires a reliable balanced data set and Comprehensive analysis.