Battle of Neighborhoods

WEEK -2

IBM Data Science
Authored by: Farzana Shaheen

1-Introduction & Business Problem :

Problem Background:

The City of New York is the financial and industrial hub of United States. It provides lot of business opportunities and business friendly environment. It is the center of fashion, and offers world-class creative talent, retail space in highly populated locations, best-in-class production companies and fashion and design schools. New York City is a global center of fashion, at the forefront of creativity, style, and innovation. This also means that the market is highly competitive for a new business venture.

To start a new clothing and accessories outlet where people already have multiple options ,the decision needs to be taken based on realistic logics . As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

2-Problem Description:

Fashion industry, multibillion-dollar global enterprise devoted to the business of making and selling clothes and NYC fashion industry makes millions annually in terms of revenue thus having a huge workforce working under umbrella of renown fashion brands.
A dress boutique is a business which designs and sells dresses according to the fashion trends for customers in return for money, either paid through credit /debit, cash on delivery ,cash payment after purchasing any item.
The City of New York is famous for its excellent fashion shows, fashion brands and street fashion. Its fashion culture includes an array of ethnic, western and modern designs, here are the fashion categories of brands designing items from high class to the budget friendly categories operating in in NYC.

1. **Haute Couture-** highest market level, sitting at the top of the fashion hierarchy Consisting of opulent bespoke garments which have been “constructed by hand from start to finish, made from high quality, expensive, often unusual fabric, sewn with extreme attention to detail and finished by the most experienced sewers, often using time consuming, hand-executed techniques” (Britta von Basedow. 2017),
2. **Luxury Fashion –** This level includes high quality designer brands, most of which belong to the three main designer conglomerates, LVMH, Kering and Richemont. Unlike Couture, the garments in Luxury Fashion are produced rather than hand-made, however they are not mass-produced, so they do still offer the desirability which comes with having limited availability and finite accessibility.
3. **Bridge Brands-** These include brand names which are considered to be placed below Luxury but above High Street, and they have been created to bridge the gap between the two .The brands within this level offer great quality clothing at a more adequate price point, so they are considered as being at the high end of affordable products. The typically entry price for a garment from a Bridge Brand would be around £50, and prices will increase depending on different items.

4. **Diffusion Lines –** Similar to Bridge Brands, these also build a bridge between high street and luxury, but the difference is that Diffusion Lines have actually been created by luxury labels as a more budget friendly secondary line.
5. **High Street-** Provide accessibility to quality fashion, with a longer life span than economy items at affordable prices.
6. **Fast Fashion-** Fast Fashion is a phenomenon in the Industry where the process of production is expedited in order to get new trends to the market as quickly and as cheaply as possible. It is a term used by fashion retailers to express that designs move from catwalk to stores extremely quick in order to capture the current fashion trends. The clothing collections created are usually based on latest fashion trends which have been presented at the most recent Fashion Weeks,
7. **Economy-** The lowest level of the fashion hierarchy. This is the home of mass production and it works efficiently at getting items from sweatshops to store. Creators of economy fashion will take on board key seasonal trends and they will attempt to turn them into affordable pieces in order to let their consumers look stylish at a much lower cost.

So, it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location and type of fashion category I will offer such as :

1. New York Population
2. New York City Demographics
3. Are there any Fabric wholesale Markets, Cheap labor availability ,professional artisans etc. nearby so that our product purchased and designed according to market trends and budget friendly too?
4. Are there any venues like Malls, Entertainment zones, Parks etc. nearby where floating population is high etc.
5. Who are the competitors in that location?
6. Fashion category of the competitors
7. Segmentation of the Borough

Therefore, my Boutique/outlet needs to open in the correct location to start my first venture. If this becomes successful, then I can replicate the same in other locations. First move is very important, thereby choice of location based on the factors above is very important.

Target Audience:

My objective is to locate and recommend the best location for my dress designing boutique to my team. I have to recommend about which neighborhood of New York city will be best choice. The recommendation I make should be based on facts and figures to convince my team .

Success Criteria:

The success criteria of the project will be a good recommendation of borough/Neighborhood choice based on Lack of such boutique in that location.

3-Data Set :

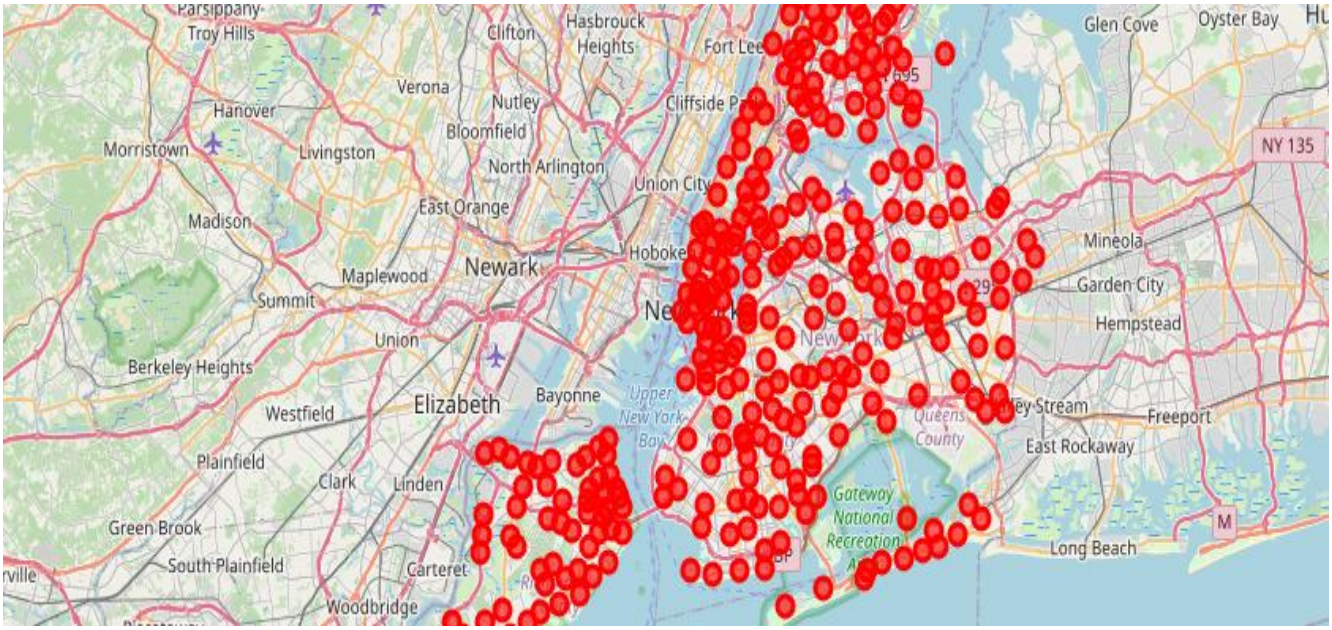
City to be analyzed in this project : New York City. We will be using the below datasets for analyzing New York city

Data 1 : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset https://geo.nyu.edu/catalog/nyu_2451_34572

Scrapping web data to analyze the neighborhoods of NYC and segmenting the neighborhoods according to population, racial index and boroughs Initially data was scraped to segment the neighborhoods and store the data in csv file. Total of 5 boroughs and 306 neighborhoods were segmented .

	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

Using Foursquare API latitude and longitude coordinated of NYC neighborhood were plotted.



Data 2 : For analysis of NYC dataset we will get data from Wikipedia as given below :

1. New York Population
2. New York City Demographics
   https://en.wikipedia.org/wiki/New_York_City
   https://en.wikipedia.org/wiki/Portal:New_York_City

To analyze the population of neighborhoods ,data was scraped from Wikipedia to segment population based on 5 boroughs.

```
54]: population_data = population_data.fillna('')
     population_data
```

54]:

	Borough	County	Estimate_2017	GrossDomesticProduct\n	square_miles	square_km	persons_sq_mi	squarekm	persons/sq.mi	persons/km2\n
0	The Bronx	Bronx	1,418,207	42.695\n	30,100	42.10	109.04			
1	Brooklyn	Kings	2,559,903	91.559\n	35,800	70.82	183.42			
2	Manhattan	New York	1,628,706	600.244\n	368,500	22.83	59.13			
3	Queens	Queens	2,253,858	93.310\n	41,400	108.53	281.09			
4	Staten Island	Richmond	476,143	14.514\n	30,500	58.37	151.18			
5		City of New York	8,336,817	101,000	842.343	302.64	783.83			
6		State of New York	19,453,561	89,000	1,731.910	47,214	122,284			
7		Sources:[14] and see individual borough articles								

The next step was to analyze the population data and segment it based on racial composition.

[92]:

	Racialcomposition	2010	1990	1970	1940
0	None	None	None	None	None
1	White	44.0%	52.3%	76.6%	93.6%
2	Non-Hispanic	33.3%	43.2%	62.9%	92.0%
3	Black or African American	25.5%	28.7%	21.1%	6.1%
4	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%
5	Asian	12.7%	7.0%	1.2%	–

[93]:
```
df.to_csv('nyc-demographics.csv',index=False)
```

Data 3 : Second data which will be fabric wholesale markets in NYC segmented into neighborhoods, we will cluster the data of each neighborhood and will explore the number of outlets.
https://www.yelp.com/search?find_desc=Wholesale+Fabric&find_loc=New+York%2C+NY

To analyze the business scope of clothing outlet ,I had to study the market trends and number of stores being opened from year 2015-2019.
Also, to identify the fashion category the brands data was analyzed based on number of outlets of brands opened in NYC boroughs.

]:
```
df_clothing
```

]:

	Retail Chain	2019	2018	2017	2016	2015	Brooklyn	Bronx	Queens	Manhattan	Staten Island
39	Children's Place	35.0	38.0	38.0	30.0	38.0	9.0	9.0	8.0	7.0	2.0
51	Gap (including GapKids & BabyGap)	28.0	29.0	33.0	30.0	31.0	5.0	3.0	5.0	13.0	2.0
54	V.I.M.	27.0	28.0	28.0	28.0	27.0	14.0	6.0	3.0	4.0	0.0
56	Jimmy Jazz	26.0	25.0	26.0	24.0	25.0	9.0	7.0	4.0	4.0	2.0
63	Aldo	24.0	24.0	23.0	23.0	24.0	4.0	1.0	3.0	14.0	2.0
...	...	...	...	...	...	...	...	...	...	...	...
306	Strawberry	1.0	1.0	2.0	6.0	7.0	0.0	0.0	0.0	1.0	0.0
307	True Religion	1.0	1.0	6.0	6.0	7.0	0.0	0.0	1.0	0.0	0.0
309	Avenue	0.0	6.0	6.0	6.0	7.0	0.0	0.0	0.0	0.0	0.0
310	Bebe	0.0	1.0	4.0	4.0	5.0	0.0	0.0	0.0	0.0	0.0
313	Gymboree	0.0	3.0	3.0	3.0	4.0	0.0	0.0	0.0	0.0	0.0

83 rows × 11 columns

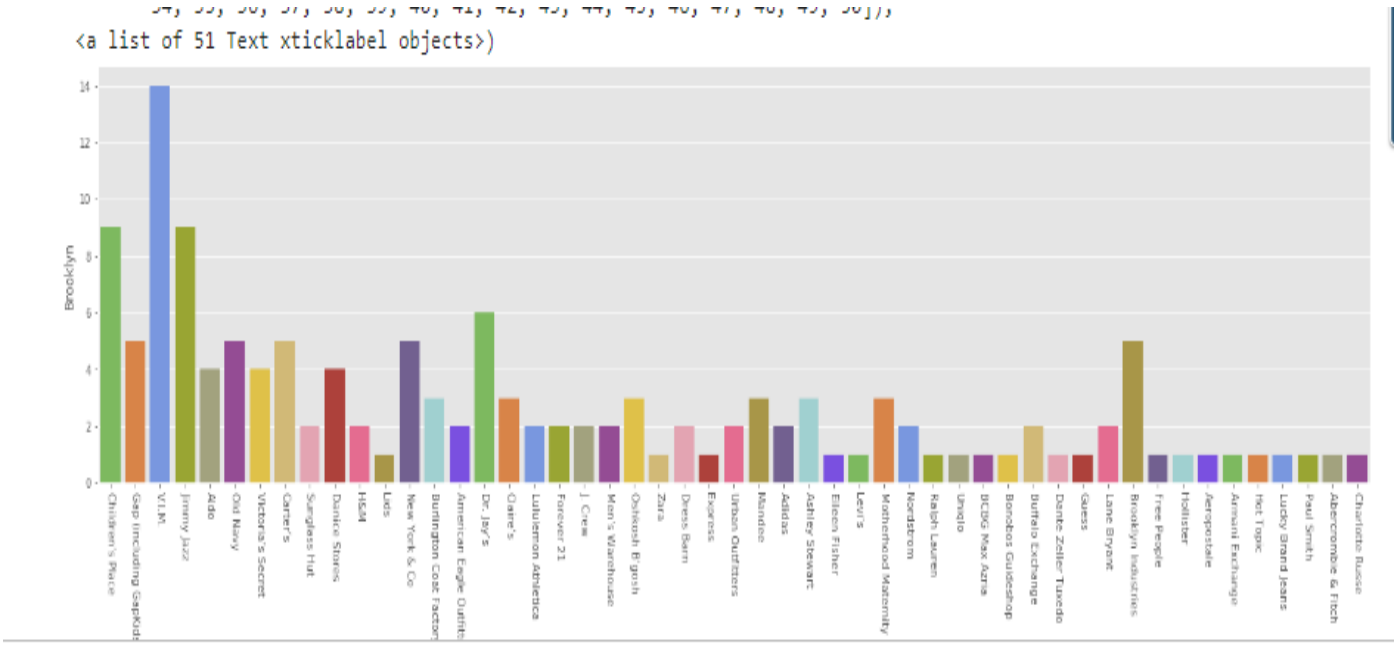
In next phase divided the data into segments of boroughs ,to identify how many stores/outlets were opened in that specific location and which brand is the most preferred or trendy one in that location. Initially analyzed the data of top 8 brands of NYC to get an idea about what sort of fashion category is preferred in the city. Here is the list of top 8 brands as per 2015-19 data and rest of the brands frequency was quite similar

1. Children's place
2. Gap
3. Old Navy
4. Jimmy Jazz

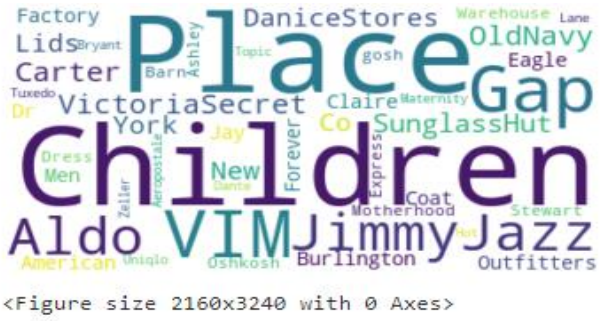
5. VIM
6. Aldo
7. Carter
8. Victoria Secret

<Figure size 7200x7200 with 0 Axes>

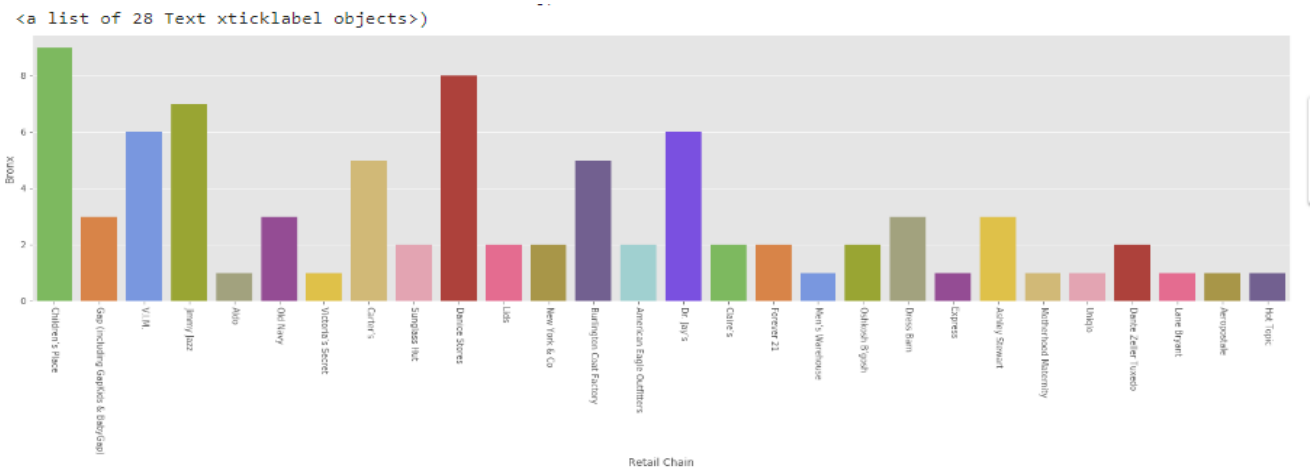
1. Total 139 clothing outlets were opened in Brooklyn in 5 years and here is the list of top brands opened in Brooklyn



2. In Bronx total 83 clothing stores were opened in 5 years and here is the list of to 5 most preferred brands in this neighborhood.

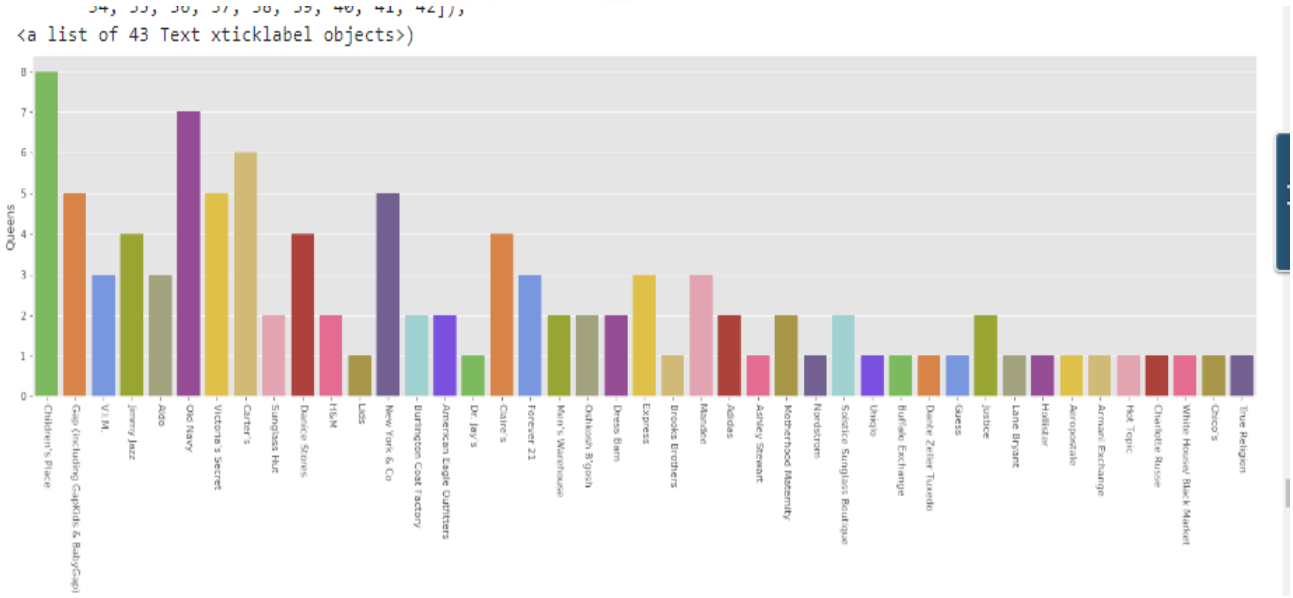


<Figure size 2160x3240 with 0 Axes>



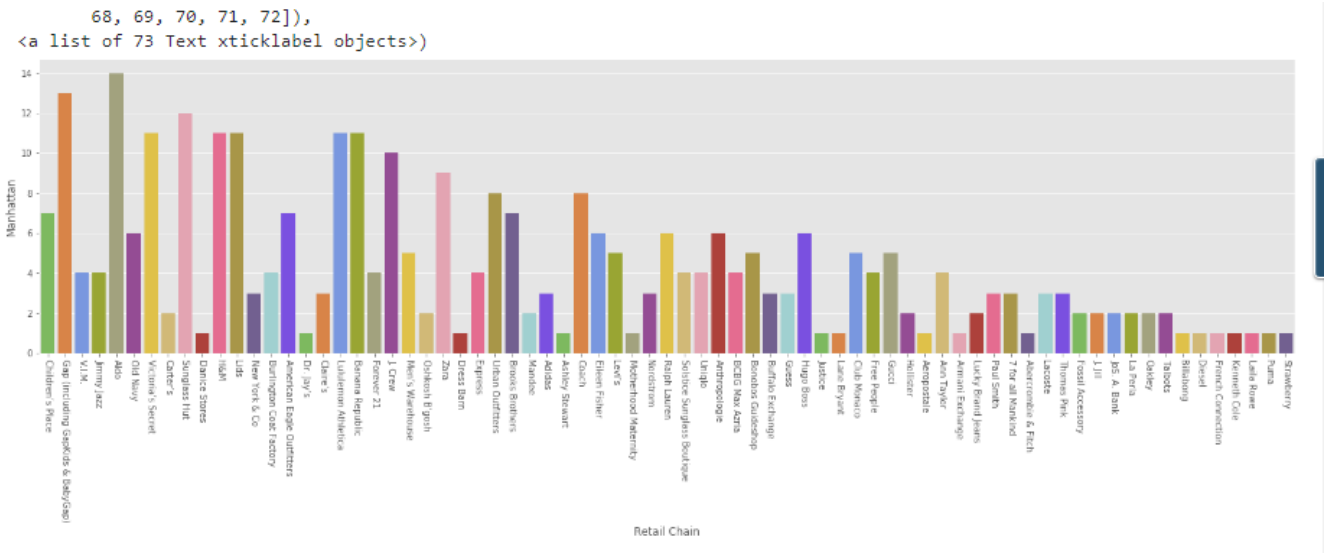
3. In Queens total 103 clothing brands were opened in 5 years and like most of the neighborhoods top 5 most preferred brands were similar.

<Figure size 2160x3240 with 0 Axes>

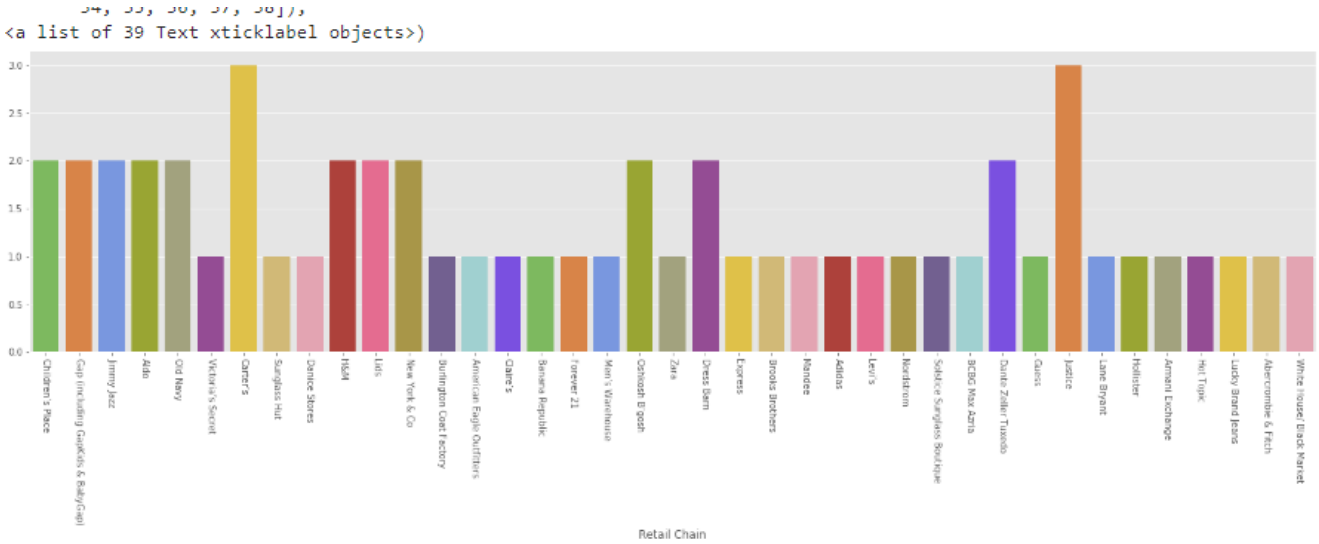


4. In Manhattan total 314 clothing outlets were opened in 5 years .Here is the word cloud of most preferred brands

<Figure size 2160x3240 with 0 Axes>



5. In Staten Island Only 54 clothing stores were opened .

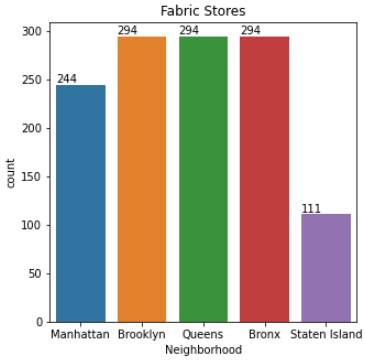


Data 4 : We will analyze the data of fashion/clothing stores opened in years ranging from 2015-2019 in New York city, we will segment the data in boroughs and will find out how many stores have been opened in neighborhoods and what is their frequency , what is the accumulative number of these fashion stores/outlets/boutiques in the recent years. This data will be scraped from the link below,
We will scrape data from page 10-24
Weblink : https://nycfuture.org/pdf/CUF_StateoftheChains_2019_7.pdf

To extract the data of wholesale fabric stores in neighborhoods of NYC I used Yelp API to scrape the pages of yelp business and I segmented the data borough wise to get an idea which neighborhood has most fabric stores .I got a dataset of around 1000 businesses to analyze in segmented into neighborhoods

[55]:

	Neighborhood
Bronx	294
Queens	294
Brooklyn	294
Manhattan	244
Staten Island	111



4-Exploring the Clusters of Queens & Bronx:

To explore the data of New York city , I created cluster of Bronx and Queens dataset .And extracted the venues Using Foursquare API accordingly.

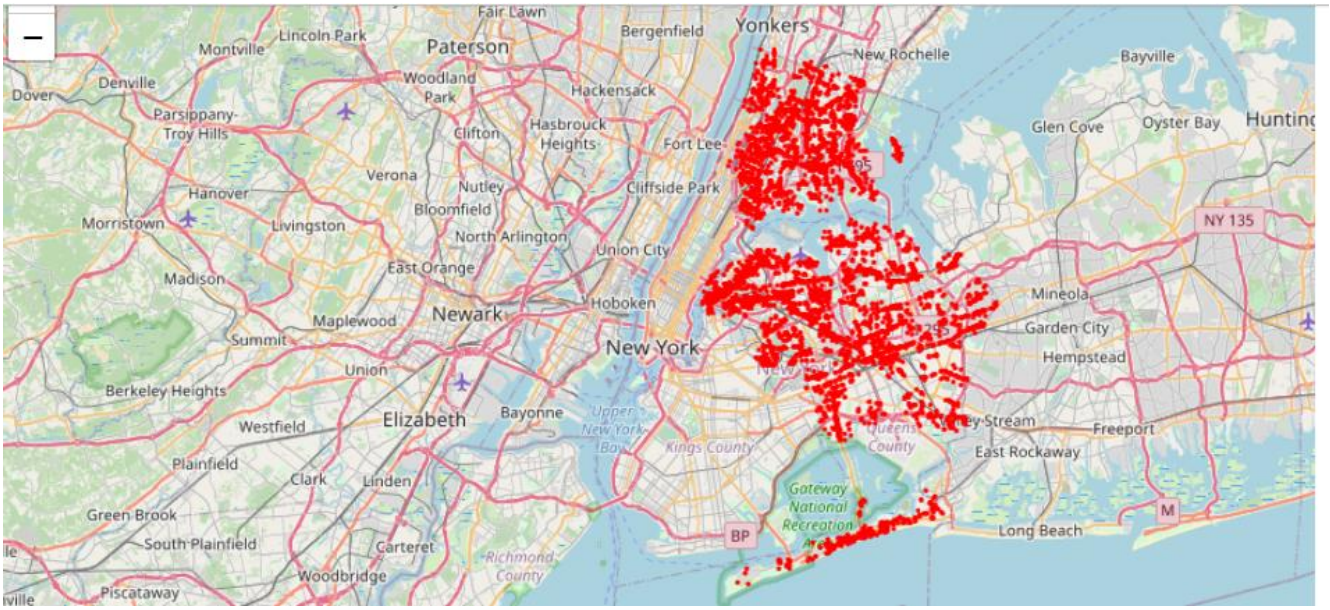
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category

I was successful in retrieving data of 371 venues and 133 neighborhoods in radius of 1000 meters. Afterwards the dataset was grouped based on type of venue category in each neighborhood.

[21]:

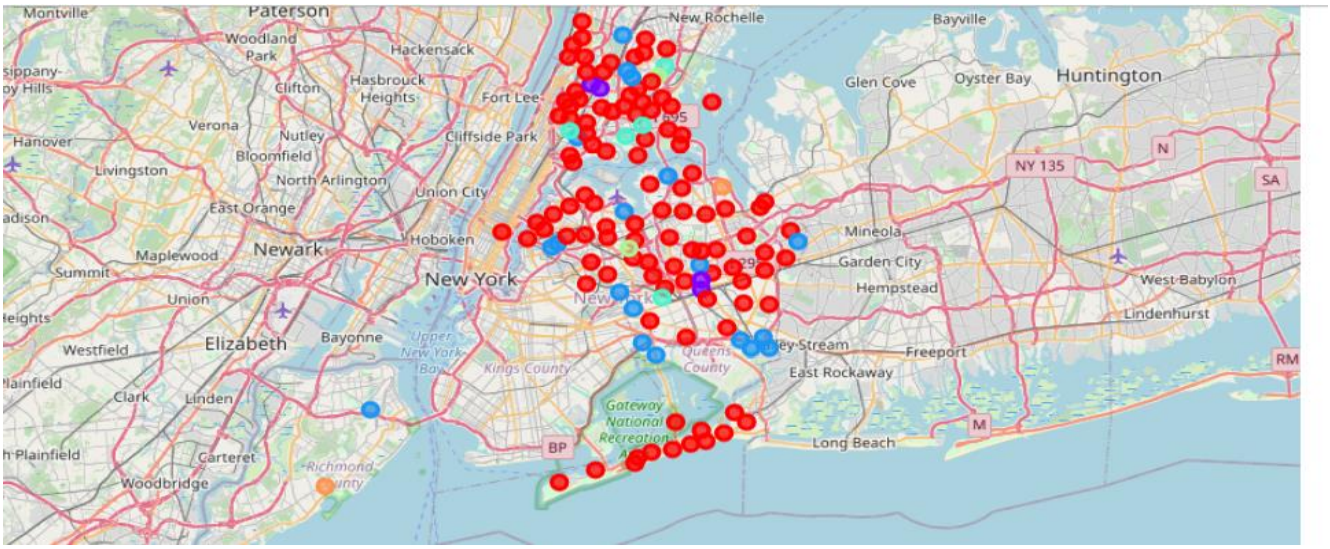
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.889062	-73.842993	Pharmacy

Queens and Bronx Venue Visualization: 371 venues and 133 neighborhoods



Clustering :

To explore the data of clothing business I refined data based on category “Clothing Store” to create clusters. Using K means created 6 clusters of datasets and plotted them accordingly.
Cluster 0 was unsaturated while cluster 1-5 were more saturated.



5-Conclusion:

This analysis was performed based on my own understanding and limited data, here are the few points I have concluded based on the analysis of data .

a) To open a clothing business market trends needs to be considered that what kind of product I am going to sell. After understanding of market trends and frequency of opened outlets in NYC one has to offer something like top 8 brands of NYC as a competitor, but this might require a lot of investment and capital.
b) After analyzing the population Bronx ,Brooklyn and Manhattan are top 3 populous boroughs which means that higher population higher consumer demand ,to target a lot of customers this can be one option to consider about starting business in these neighborhoods.
c) The number of top 3 outlets opened in NYC in last five years are higher in Manhattan and Queens which means there is already a lot of competition to start a new clothing line.
d) Along with clothing lines if wholesale fabric is considered as a competitor then Bronx, Brooklyn and Bronx has higher number of stores .
e) Based on all factors above we can consider the cluster 0 of untapped markets with less competitors and same practice can be performed rigorously for other boroughs too.
f) In the end to start a new business with less capital and less competitors Staten Island is the good option to experiment as it has lowest number of both wholesale fabric and clothing line stores.