ANALYSE AND VISUALISE THE FACILITIES OF NEW YORK'S BOROUGHS FOR F&B BUSINESS

Capstone Project - The Battle of Neighborhoods (Full Version)

Challenge 1 - Business Understanding and Data Science Methodology

Business Understanding

Business Scenarios
United States has been on top trending of paradise of new start up companies. There have been loads of immigrants to US with new business sections created and challenged. The trending generates more challenges for OPENING A BUSINESS (CAFE' AND RESTAURANT) in the Country. The City is diversed, competitive and captial of USA's Finance. The cosmopolitan in the city offers the City a lot of attractions to immigrants who wanna growth their Dream in the city.

Because of the aggressions of immigrants and high competition in New York City in terms of PLEASURING and CAFE' & RESTAURANT, an idea of creating a new business model that ENTER the market with a significant outstanding business in very short-time (6 months). There are 5 BOROUGHS in New York City which should be very potential to make the business idea happens. To firstly define strategic GO-TO-MARKET plan, we should analyse the city, 5 Boroughs and their Neighbourhoods to gain more insights from the city in terms of strategic location and population allocated. The ultimate of the project is to understand the market by strategical insights released based on Data Science in order to Business Owners make strategic decision for their CAFE' & RESTAURANT Business.

Business Problem Description
The perfect challenge is to ideate business model that provides insights to know which boroughs are ideal to open up a Cafe' & Restaurant based on the huge impact of population and neighbourhoods.

Big question is how we identify the perfect information of specifically idea BOROUGH and its NEIGHBOURHOODS in order to know the most impactful factor of business Enter.

With Data Science, we will steps by steps to demonstrate the Cluster of New York City neighbourhoods. By giving the visualation of Cluster, we can recommend potential BOROUGH and strategic locations for ONWNERS for their decision making.

Data Collection:

City - New York City.

Datasets to understanding the City.

Source 1 - Neighbourhood Data: https://geo.nyu.edu/catalog/nyu-2451-34572

Source 2 - New York Population: https://bit.ly/2StDkJJ (sources from Kaggle)

With multiple sources of data, we will access each of data to select target locations across different venues based on the accordance of filtering facilities and factors. To make a JOIN to data of properties and pricing and other factors, we will explore 4 sources and plus with Foursquare API interface, we will be able to recommend affordable accommodation

Measurables
1. Market Understanding of Strategic Locations
2. Ideal recommendation of accommodation in compared with Neighbourhood to Business Owners

Methodology
1. Data Understanding Data - CRISP-DM combines the stages of Data Requirements, Data Collection, and Data Understanding. -> Data Preparation
2. Data Preparation - Data Preprocessing - Transformed into a useable subset. Once a dataset is chosen, it must then be checked for questionable, missing, or ambiguous cases
3. Modeling - Select appropriate models, give meaningful insights, and hopefully new knowledge.

1. Data Collection

The first step is to collect data in various sources, format it in the column set with dataframe established. The showing of location's Lat and Long with information of Neighborhoods.

	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

The second step is to collect data from the other sources to input more data. The data of population is critical to define the top Boroughs to analyse. The combination of multiple tables into one by indexing and merging information with data type double checked to correct the data at side.

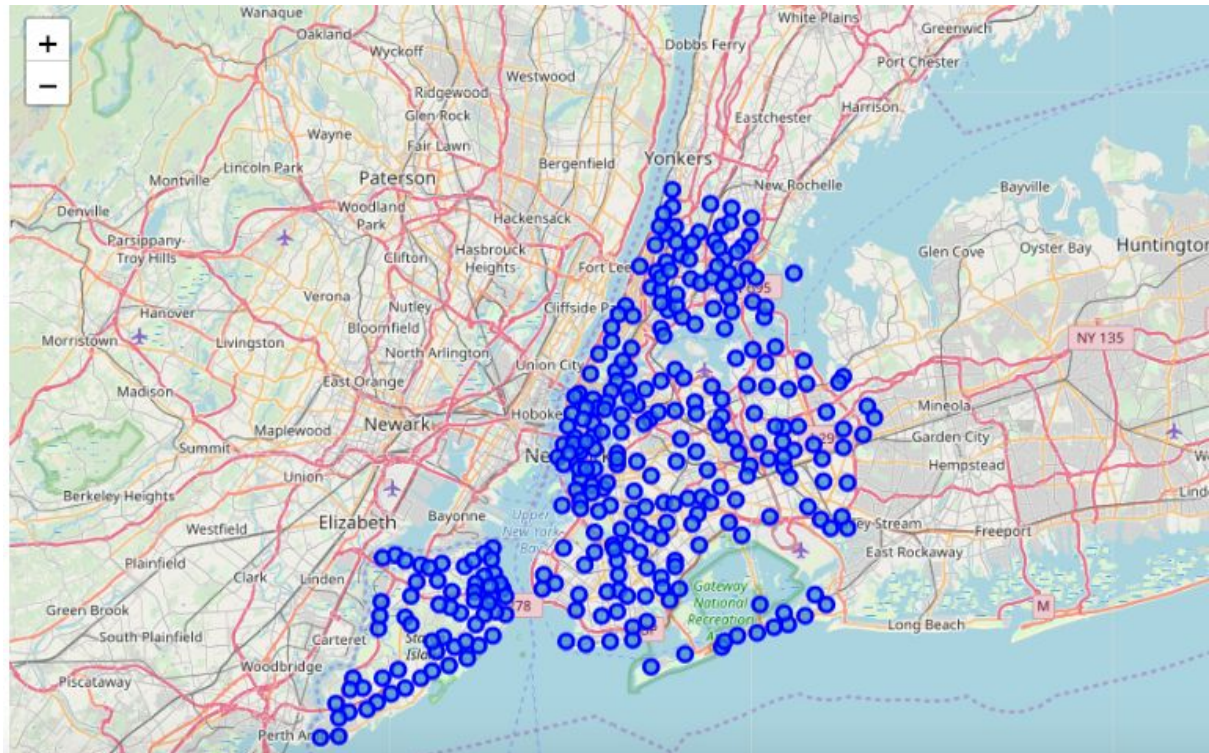|	Borough	Neighborhood	Latitude	Longitude	Population
0	Bronx	Wakefield	40.894705	-73.847201	91,497
1	Bronx	Wakefield	40.894705	-73.847201	52,246
2	Bronx	Wakefield	40.894705	-73.847201	79,762
3	Bronx	Wakefield	40.894705	-73.847201	146,441
4	Bronx	Wakefield	40.894705	-73.847201	128,200

Data Exploration.

To understand more the top boroughs with the highest populations, I did some codes to filter and select the values to analyse.

	Borough	Population
0	Bronx	91,49752,24679,762146,441128,20083,268139,2861...
1	Brooklyn	173,08399,617152,985112,634182,896104,709126,2...
2	Manhattan	60,97890,016163,277103,24551,673142,745209,084...
3	Queens	191,105113,200171,576172,598169,190113,257247,...
4	Staten Island	175,756132,003160,209175,756132,003160,209175,...

	Borough	Population
4	Staten Island	175,756132,003160,209175,756132,003160,209175,...
3	Queens	191,105113,200171,576172,598169,190113,257247,...
2	Manhattan	60,97890,016163,277103,24551,673142,745209,084...
1	Brooklyn	173,08399,617152,985112,634182,896104,709126,2...
0	Bronx	91,49752,24679,762146,441128,20083,268139,2861...

By understanding the critical values of analyse, I did pick the Staten Island to have more explorations in terms of categories and its neighborhoods

	Borough	Neighborhood	Latitude	Longitude
0	Staten Island	St. George	40.644982	-74.079353
1	Staten Island	New Brighton	40.640615	-74.087017
2	Staten Island	Stapleton	40.626928	-74.077902
3	Staten Island	Rosebank	40.615305	-74.069805
4	Staten Island	West Brighton	40.631879	-74.107182
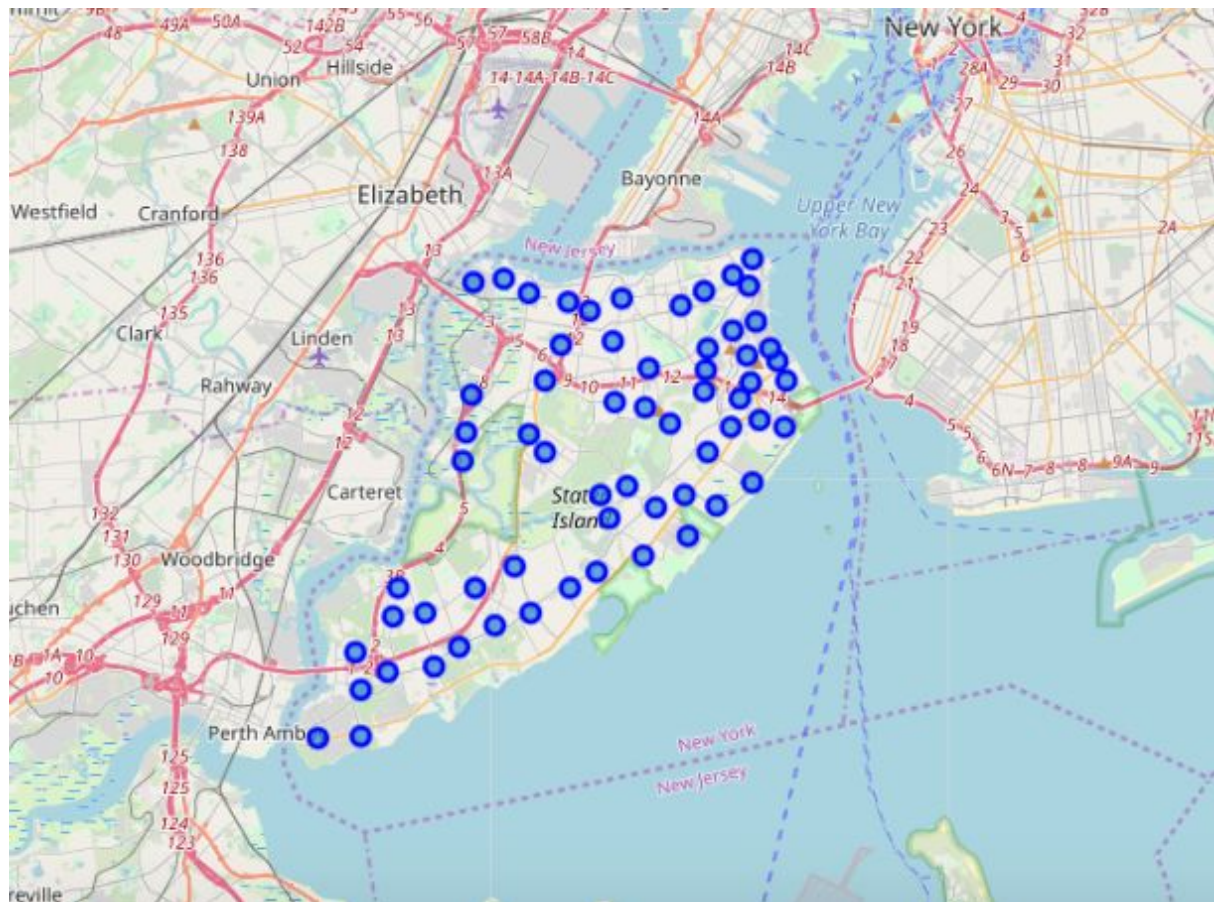
I decided to pick one of the Neighborhood of Staten Island to explore before to check total categories and neighborhoods.

	name	categories	lat	lng
0	A&S Pizzeria	Pizza Place	40.643940	-74.077626
1	Beso	Tapas Restaurant	40.643306	-74.076508
2	Richmond County Bank Ballpark	Baseball Stadium	40.645056	-74.076864
3	Staten Island September 11 Memorial	Monument / Landmark	40.646767	-74.076510
4	Nike Factory Store	Sporting Goods Shop	40.645753	-74.077702

at St. George area, there are including categories: Pizza, Tapas, Stadium, Landmark, Sport Goods

There are a lot of categories of St. George, however, I did select the head values to quick look. I realised some categories are listed such as Pizza, Restaurant, Stadium, Landmark, Sporting Goods Shop.

Finally, I checked up with total Nearby Venues of Staten Island to know all catogories to analyse.

```
Bulls Head
Richmond Town
Shore Acres
Clifton
Concord
Emerson Hill
Randall Manor
Howland Hook
Elm Park
Manor Heights
Willowbrook
Sandy Ground
Egbertville
Prince's Bay
Lighthouse Hill
Richmond Valley
Fox Hills

In [367]: print(StI_venues.shape)
          StI_venues.head()

          (829, 7)
```

There are 829 nearby Venues. Now, I continue to define each cate in the venues.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Annadale	Bakery	Dance Studio	Pizza Place	Sushi Restaurant	Park	Restaurant	Sports Bar	Diner	Train Station	Food & Drink Shop
1	Arden Heights	Pharmacy	Bus Stop	Coffee Shop	Pizza Place	Event Space	Food & Drink Shop	Food	Flower Shop	Fish & Chips Shop	Filipino Restaurant
2	Arlington	Deli / Bodega	Bus Stop	Boat or Ferry	Coffee Shop	Grocery Store	Fast Food Restaurant	Food Truck	Food & Drink Shop	Food	Flower Shop
3	Arrochar	Bus Stop	Deli / Bodega	Italian Restaurant	Food Truck	Polish Restaurant	Outdoors & Recreation	Supermarket	Middle Eastern Restaurant	Sandwich Place	Bagel Shop
4	Bay Terrace	Supermarket	Insurance Office	Italian Restaurant	Train Station	Salon / Barbershop	Donut Shop	Sushi Restaurant	Shipping Store	Wings Joint	Event Space

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Staten Island	St. George	40.644982	-74.079353	0.0	Clothing Store	Bar	Sporting Goods Shop	Italian Restaurant	Plaza	Scenic Lookout	Burger Joint	Farmers Market	Coffee Shop	Donut Shop
1	Staten Island	New Brighton	40.640615	-74.087017	3.0	Bus Stop	Park	Bowling Alley	Discount Store	Home Service	Playground	Fast Food Restaurant	Food Truck	Food & Drink Shop	Food
2	Staten Island	Stapleton	40.626928	-74.077902	0.0	Discount Store	Pizza Place	Mexican Restaurant	Bank	Sandwich Place	Residential Building (Apartment / Condo)	Skate Park	Motorcycle Shop	Fast Food Restaurant	Restaurant
3	Staten Island	Rosebank	40.615305	-74.069805	0.0	Pizza Place	Grocery Store	Italian Restaurant	Pharmacy	Cosmetics Shop	Discount Store	Eastern European Restaurant	Restaurant	Sandwich Place	Filipino Restaurant
4	Staten Island	West Brighton	40.631879	-74.107182	0.0	Coffee Shop	Sandwich Place	American Restaurant	Italian Restaurant	Music Store	Bank	Bar	Wings Joint	Burger Joint	Mexican Restaurant

As you can see, restaurants are the most Categories appeared in the Common Venues. Which different types of restaurants, Borough;: Staten Island and its neighborhoods are much more competitive are F&B.
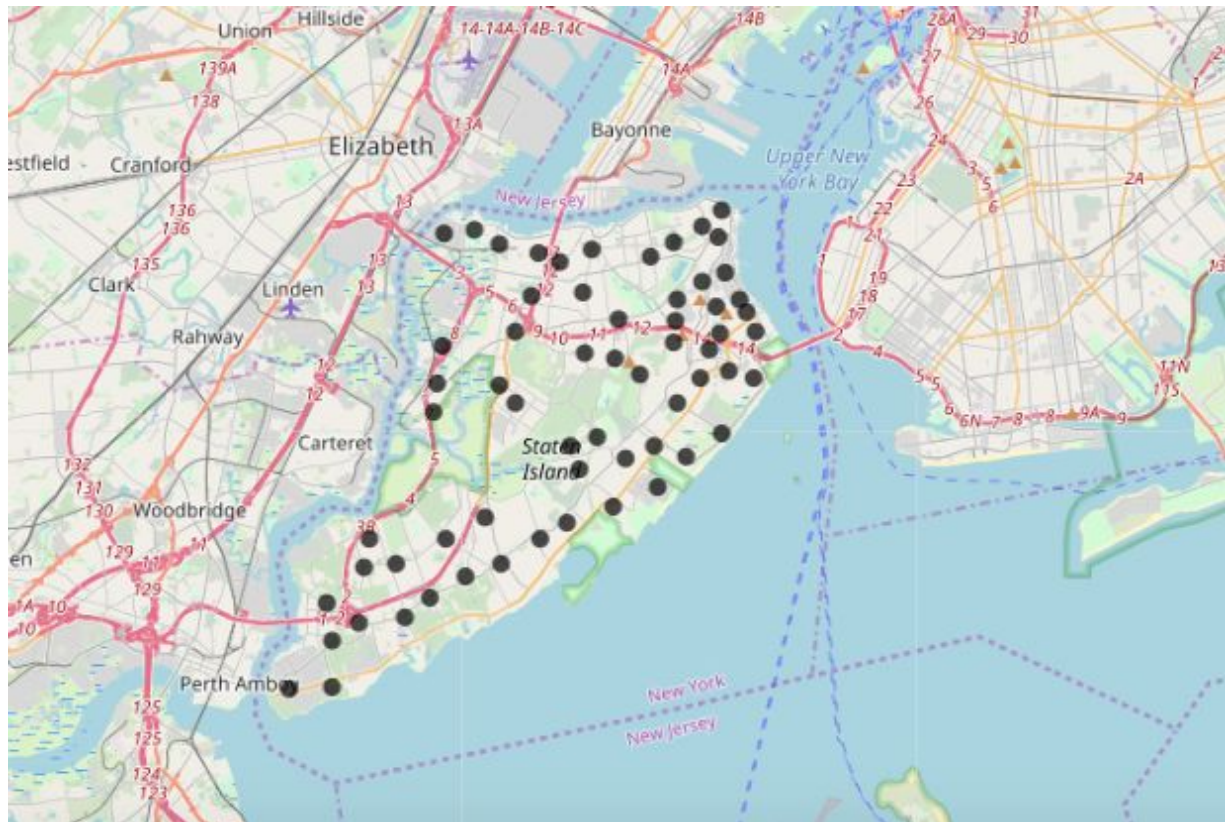
Conclusion

Staten Island is the potential Borough of New York to think of establishing business. However, it's cluster of categories and neighborhoods are the most concerns. While Restaurants are too diversified, some other categories are less competitive. This exercise to show out the strategic location of New York's Boroughs and its neighborhoods. Strategic location in line with population is Staten Island with high competition of F&B. Business Owner should consider more the Business Model and type of F&B to enter the market.