Capstone Project - The Battle of Neighborhoods REPORT

Background

For the new entrepreneurs, New York offers a stable environment, a large economy, and access to one of the world's busiest regions. Small business owners in the state don't expect that to change and are largely optimistic about what the future holds. That means plenty of business opportunity, in New York City neighborhood. New Yorkers are, on average, wealthier than their national counterparts as well, meaning more money to spend on the goods and services small businesses have to offer. However, the heightened cost of living can prove difficult to manage. Still, entrepreneurs said that if they can overcome the steep expenses associated with payroll and rent, not to mention a tangled web of taxes and fees, operating in New York is an investment that pays off in the end. Starting a new business in New York City will be a tough work.

Problem

One of the biggest challenges for any New York City business is going to be the wide array of competition. In a city of eight million plus, there is going to be competition around every corner. The New York City market has become overly saturated in almost every single industry from doctor practices to restaurants. Businesses now have to go beyond just serving their customers and focus on marketing and reputation in order to improve customer acquisition and grow their businesses. With Queens having one of the most diverse places in the nation. Half of the neighborhood's residents speak Spanish. Others speak Chinese, Urdu, Hindi, Russian, Portuguese, Greek or Korean. Altogether, the neighborhood is said to be the home of 167 languages. It will be a challenge on what business to start and offer these cultures diverse in one of the five boroughs of New York City with it having the largest borough geographically.

Solving the Problem Using Data Science

To solve the existing problems, we will recommend to new business by collecting location data from Foursquare and apply data science techniques and tools. We are going to cluster New York neighborhoods in order to find existing business establishments and venues in order to scope out the competition and create a new trendy business

Data section

We extracted the ZIP Code Definitions of New York City Neighborhoods which is available from https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm. Created a csv file and uploaded to own server. To explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we will access data through FourSquare API interface and arrange them as a dataframe for visualization. By merging data on New York City Zip Codes by neighborhood and data on venues and essential facilities surrounding such properties from FourSquare API interface.

Methodology

1. Collect Data
2. Explore and Understand Data
3. Data Preparation and Preprocessing
4. Modeling

Results and Discussion

All clusters have 20 restaurants and food places on the area that caters to specific cuisines.

Cluster 1 has 1 neighborhood with 3 restaurants as common venues and 2 stores.
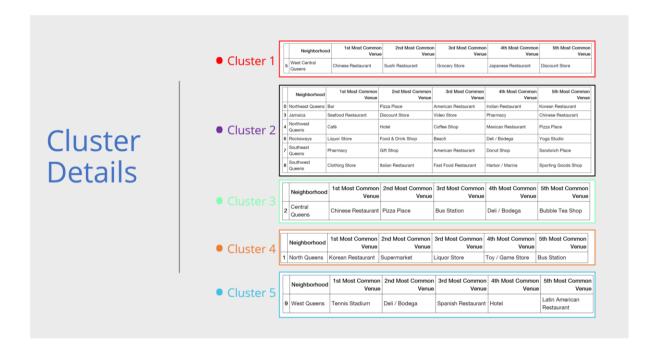Cluster 2 has 6 neighborhoods with 13 diverse food places and restaurants but has different types of stores.
Cluster 3 has 1 neighborhood with 2 restaurants and food places has a bus station
Cluster 4 has 1 neighborhood with 1 restaurant a supermarket and a bus station.
Cluster 5 has 1 neighborhood with 2 restaurants as common venues a hotel and a stadium.

The geographic location does give us a a decent view on where to open the new businesses as the thought is to begin with there are a lot of restaurants and cater to the diverse culture of the borough. It looks like the neighborhood of Cluster 2 are the greater ones and the vast majority of the neighborhood compared to the rest.

Cluster Details

Cluster 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	West Central Queens	Chinese Restaurant	Sushi Restaurant	Grocery Store	Japanese Restaurant	Discount Store

Cluster 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Northeast Queens	Bar	Pizza Place	American Restaurant	Indian Restaurant	Korean Restaurant
3	Jamaica	Seafood Restaurant	Discount Store	Video Store	Pharmacy	Chinese Restaurant
4	Northwest Queens	Café	Hotel	Coffee Shop	Mexican Restaurant	Pizza Place
6	Rockaways	Liquor Store	Food & Drink Shop	Beach	Deli / Bodega	Yoga Studio
7	Southeast Queens	Pharmacy	Gift Shop	American Restaurant	Donut Shop	Sandwich Place
8	Southwest Queens	Clothing Store	Italian Restaurant	Fast Food Restaurant	Harbor / Marina	Sporting Goods Shop

Cluster 3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
2	Central Queens	Chinese Restaurant	Pizza Place	Bus Station	Deli / Bodega	Bubble Tea Shop

Cluster 4

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	North Queens	Korean Restaurant	Supermarket	Liquor Store	Toy / Game Store	Bus Station

Cluster 5

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	West Queens	Tennis Stadium	Deli / Bodega	Spanish Restaurant	Hotel	Latin American Restaurant

Conclusion

The KMeans classification technique gives a new entrepreneur a good idea on what type of trendy business he/she can start with based on the type of venues identified using Foursquare api.

The cluster of venues identified within the neighborhood is predominantly restaurants and food shops, but there is no nearby coffee shop. Catching up with the trend of offering gourmet and specialty coffee from different parts of the world will definitely fit the diverse culture of this district.