**Project Overview: Analysis of New York City Neighborhoods for Aspiring Businesses**

**Introduction**

*Business Problem:*

A small, family-owned coffee shop business is planning its move to

New York City. Although the coffee shop has great beverages, food options, and unique ambiance, its main competition is large, established coffee shop chains. Where in New York City will this business be able to capture the attention of audiences, with limited competition?

*Scope:*

One of the most important aspects of starting a business is determining the optimal location for its implementation. The location of an aspiring business can be vital to its long term success, due to factors of competition and target audience market. The premise of this Capstone Project is to leverage the Foursquare API location data for the city of New York, to analyze the business composition of the city’s major neighborhoods. The boroughs in consideration for the scope of this project include Manhattan, Brooklyn, Queens, and the Bronx. Using the clustering method analysis, the optimal location for an aspiring business can be determined, based upon the factors of competition and potential market.

**Data**

*Data Set:*

* New York City Data: Data set describing the boroughs and neighborhoods that make up the city of New York. This data set was cleaned and appended to include latitude and longitude data.
* Foursquare API Data: Well-established and utilized API that provides crowdsourced business and location data.

*Data Analysis:*

* The Foursquare API will be used to gather, organize, and analyze data describing the business composition of the major neighborhoods in New York City. Through clustering method analysis, the neighborhoods will be divided according to their business composition, and visualized. Through this analysis, aspiring businesses can target locations within the city which provide a viable market for their growth, with limited competition. As a result, the business problem can be solved.

**Methodology**

*Assumption:* The business composition of neighborhoods is dictated by socioeconomic factors.

1. New York City data set was obtained, which described the neighborhoods of the boroughs of New York being analyzed (Manhattan, Brooklyn, Queens, Bronx).
2. New York City data set was appended to include longitude and latitude coordinate information.
3. Foursquare API was used to obtain information on the businesses present in each of the New York City neighborhoods, according to the New York City data set, and compiled.
4. In order to gain insight into the business composition of the neighborhoods, the category frequency of the top businesses was determined for each neighborhood. Additionally, this information was displayed by listing the top 10 most common business types, for each neighborhood.
5. The final data analysis technique used to analyze the business composition of New York City was KMeans. KMeans, a clustering method, was selected in order to organize the neighborhoods of each of the boroughs, according to their business composition. As a result, neighborhoods in the same cluster will have similar business compositions. KMeans was used because the assumption of this analysis is that socioeconomic factors dictate the business composition of a neighborhood. Due to this assumption, KMeans was selected as the clustering method because the clusters will accurately categorize neighborhoods.
6. The KMeans results for each of the boroughs and their neighborhoods were visualized onto a map of New York City. Using the results from the KMeans clustering analysis, it can be determined where the optimal location for a new coffee shop would be in each of the boroughs of New York City being considered.
7. The analysis takes into account 2 factors: competition and potential markets. By determining which of the neighborhood clusters contains the most restaurants but lacks coffee shops, both factors are incorporated in determining the potential location in each of the boroughs.

**Results**

|  |  |  |
| --- | --- | --- |
| Borough | Relevant Cluster | Number of Potential Neighborhoods |
| *Manhattan* | **#1, #2, #3** | **4** |
| *Brooklyn* | **#1, #4** | **14** |
| *Bronx* | **#3, #4** | **17** |
| *Queens* | **#3** | **80** |

*Manhattan:*

The results from the Manhattan clustering analysis showed that clusters 1, 2, and 3 were the most relevant in targeting the potential neighborhoods for a new coffee shop. In total, there were only 4 neighborhoods that contained a potential market and had limited competitors.

*Brooklyn:*

The results from the Brooklyn clustering analysis showed that clusters 1 and 4 were the most relevant in targeting the potential neighborhoods for a new coffee shop. In total, there were 14 neighborhoods that contained a potential market and had limited competitors.

*Bronx:*

The results from the Bronx clustering analysis showed that clusters 3 and 4 were the most relevant in targeting the potential neighborhoods for a new coffee shop. In total, there were 17 neighborhoods that contained a potential market and had limited competitors.

*Queens:*

The results from the Manhattan clustering analysis showed that cluster 3 was the only relevant cluster in targeting the potential neighborhoods for a new coffee shop. In total, there were 80 neighborhoods that contained a potential market and had limited competitors. As a result, Queens is the optimal location for establishing a new coffee shop business, with a variety of specific neighborhoods to choose from.

**Discussion**

An important observation that can be made from the KMeans analysis is that as the number of relevant clusters for each borough increases, so does the number of potential of neighborhoods. This observation speaks to the effectiveness of the KMeans analysis, in categorizing and organizing the data into clusters.

The results from the KMeans clustering analysis clearly shows that Queens is the best borough of New York City for establishing a new coffee shop business, with 80 neighborhoods to choose from. My final recommendation to the family-owned coffee business is to target the borough of Queens and specify their search for the neighborhood, based upon financial factors and logistics.

**Conclusion**

In order to determine the optimal location for a new coffee shop in New York City, a New York City dataset was created, which leveraged Foursquare API to compile data about the business composition of the boroughs of Manhattan, Brooklyn, Bronx, and Queens. Using KMeans clustering analysis, each of the neighborhoods in these 4 boroughs was organized into clusters which could identify the neighborhoods in which there was a potential market for a coffee shop, and where there was limited competition. The KMeans clustering analysis clearly showed that the Queens borough had the largest market and limited competition, with 80 potential neighborhoods to choose from.