Joie Roque

Finding a Spot for Some Bubble Tea

Introduction

Some stakeholders want to open a bubble tea shop in New York City, and are looking for the most strategic area for one based on competition and local demographic. The goal is to spread the love of bubble tea beyond the Asian community; the intent is to find areas with relatively low direct and indirect competition but with a noticeable Asian population to have an initial customer base to grow from.

Data

Based on the definition of our problem, indicators for potential locations include: number of existing bubble tea shops in the neighborhood, percentage of venue population that is sweet refreshment-type shops in the neighborhood, and percentage of the neighborhood population that is Asian. To create these features, we will need to extract venue location and category information from Foursquare API, and New York City census data by neighborhood and borough.

Methodology

In this project we will try to identify optimal neighborhoods for a new bubble tea shop based on number of bubble tea shops (direct competition), percentage of refreshment-type shops out of total venues (indirect competition), and Asian demographic residing in the neighborhood.

Using the generated features, we looked at top 10 Asian population, and bottom 10 bubble tea shop count and percentage of refreshment shops. There are an abundance of neighborhoods without any competition, however, so we need to be slightly more sophisticated. Therefore, we use k-means clustering to assign a grouping to the neighborhoods we have complete data for to determine which areas we should be most interested in, and plot these clusters on a map.

Analysis

For clustering we simply use the rule of thumb $\sqrt{\frac{N_{samples}}{2}}$ since we simply want to find a most relevant cluster and have a small enough sample size to not impact calculation time. With 282 neighborhoods, this comes out to be about 12.

Results

Our brief analysis shows that a few well-known neighborhoods in Manhattan and Queens would be good places to look to start the bubble tea business, based on demographic and competition data.

To start, we used Foursquare in conjunction with census data to generate the following features: Asian population percentage, number of bubble tea shops, and percentage of refreshment-type shops out of total venues. To get a high-level overview, we looked at highest Asian population and lowest competition. However, many neighborhoods have no competition, so we're not able to glean much from this.

We took it a step further by using k-means clustering on the aforementioned variables to isolate groupings of neighborhoods that have both high Asian population and low competition. This resulted in discovery of a small cluster of 8 neighborhoods that have ≥50% Asian population, no other bubble tea shops, and venues comprising of ≤10% refreshment shops. Within this cluster are areas that make sense, including a few neighborhoods in Manhattan, like Upper East Side/East Harlem. This is informative, and a good indicator that the features we are using as metrics are working.

Below are the cluster's data and map in Figures 1 and 2, respectively.

Neighborhood	Percent Asian	Bubble Tea Shop Count	Percentage Refreshment Shops	Borough
Chelsea	64.5	0.0	5.77	Manhattan
Chelsea	64.5	0.0	5.77	Staten Island
East Harlem	58.8	0.0	0.00	Manhattan
Erasmus	57.3	0.0	10.00	Brooklyn
Maspeth	57.3	0.0	3.45	Queens
Mount Hope	54.4	0.0	7.69	Bronx
Queensboro Hill	63.6	0.0	4.00	Queens
Upper East Side	50.3	0.0	3.00	Manhattan

Figure 1. Cluster Data

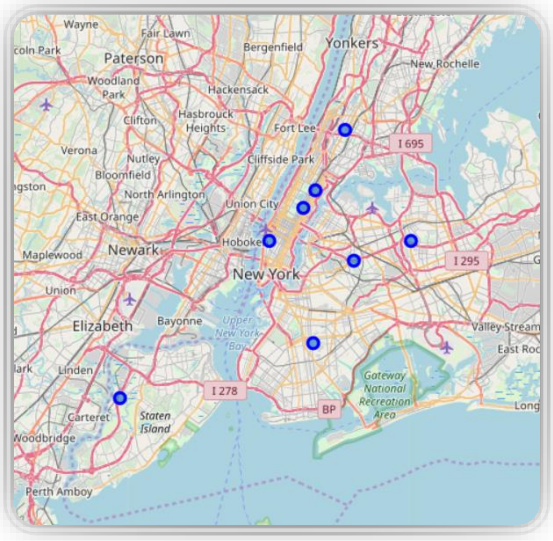

Figure 2. Cluster Map

Joie Roque

Discussion

There is more to be done here. As a next step, we should look at factors outside of market, like rent and crime rates. Factors like these can prohibit the starting of new businesses of this or any type. This would explain the abundance of low competition areas as well.

Also, we could check some of the other high Asian population clusters with slightly higher competition for completeness. Yet another improvement improvement that could be made is to thoroughly curate the census data; there were some issues with matching that were caused by inconsistent naming of neighborhoods.

Conclusion

The purpose of this analysis was to take a first step into finding a suitable neighborhood to open a bubble tea business. By quantifying competition and demographic data from Foursquare and the NYC census, respectively, we have narrowed down our search to 8 neighborhoods.

This information will be used by the stakeholders to proceed with further analysis, involving other metrics like crime rates and rent.