Ideal Venues in NYC

By: Vatsal Patel

Introduction:

Background

A small ramen shop owner located in central Manhattan, New York has just saved a large sum of money from selling ramen. This realizes the success he received in this area of the city was because of the authenticity, quality, price and overall experience that the customers have when dining at his shop. Due to these factors and being located in a highly dense place such as Manhattan and maintaining a strong reputation for the quality of the food being served, he believes he can achieve similar levels of success in other locations in New York. He notices that his sales hit their peak during lunch time due to a large amount of businesses located near his shop. Furthermore, his shop is located near a subway station which makes it readily accessible to individuals who are staying / working further away. With his savings now large enough to open another shop, he is looking to expand his business.

Description of the Problem

With factors such as taste, quality of food, authenticity and price of the food remaining the same, the shop owner wants to at the very least achieve similar levels of sale (ideally the goal would be to outsell his current shop). With this being said what location is ideal for him to open up shop at? And what factors would deem a location ideal for a ramen business. The shop owner wants to better understand the demographics of the customers visiting his shop. Doing this he can create an environment that is best suitable for the demographic. In doing so, he will achieve similar levels of sales, improve restaurant ratings and create a loyal customer base.

Approach

For the restaurant owner to maintain / improve his sales, we can use the Four Square API to analyze location based data in New York. From the dataset we would want to identify the frequency and time of visits from the area's inhabitants

Based on this factors we can identify a location which is frequently visited, in an enjoyable neighborhood and be able to create an environment that is ideal to the age group of the possible customers visiting the area. Creating this positive and engaging environment at the restaurant should encourage the customers to write positive reviews and aid in essentially marketing the restaurant.

Interests

The solution to this problem would be useful to individuals who are trying to determine locations that is ideal for them to open a restaurant in. The concepts and approach used here can be applicable to various locations but the final results will be based in NYC.

Data Collection & Cleaning

Collection

The data can be easily access through the Four Square API. The data will focus on the latitude and longitude of subway station in New York and the various restaurants and businesses located near the station. The subway location dataset can be accessed from here: https://data.cityofnewyork.us/api/views/kk4q-3rt2/rows.csv?accessType=DOWNLOAD which can be retrieved from http://web.mta.info/nyct/service/.

Cleaning

The dataset will be segmented into sectors for the various subway stations in NYC. Following that I was able to identify the names, latitude and longitude of the locations. Following that, from importing the Four Square API for NYC we can identify and visualize the subway location and name based on the latitude and longitudes. Then we are able to identify various restaurants and businesses in a small radius of the subway station. Finally we are able to see the unique locations after removing repeated venues that were in the same radius of multiple subway stations.

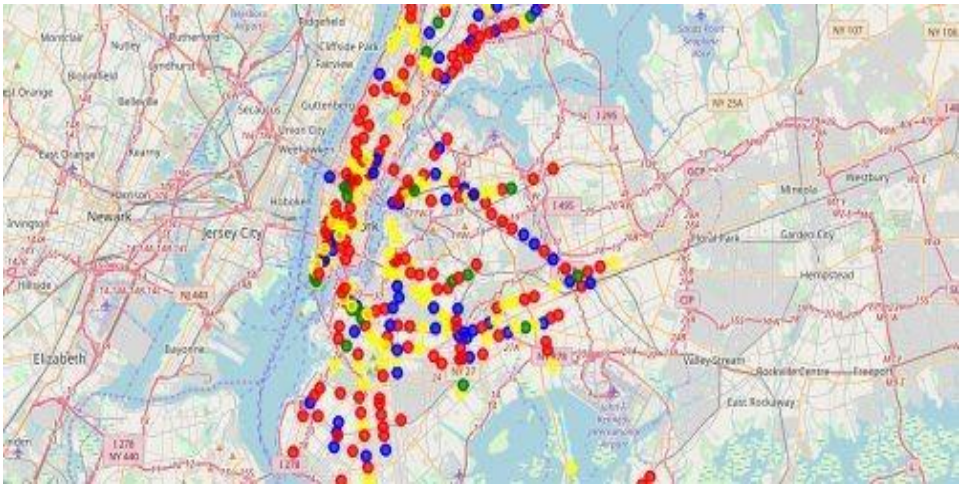
Methodology

We will use a clustering algorithm called K-means. Clusters are a group of objects similar to other objects in the cluster and dissimilar to data points in other clusters. We can partition data sets into groups with similar characteristics. Partition based clustering is also known as K-means clustering. This type of partition based clustering divides the data into non-overlapping subsets. The purpose of K-means is to minimize the intra-cluster distances and maximize the inter-cluster distances.

Using the K-means clustering algorithm we can create clusters of various venues which have high frequency visits near subway stations in New York. Then we are able to identify the venues and the frequency of customers based on the clusters created.

Results & Discussion

The analysis shows that there are various spots around New York that would generate similar levels of revenues from the frequency of the visits by the customers. A high concentration of locations are near pizza shops, Italian restaurants, parks, bars near the same location of downtown Manhattan. The model also showed a large clustering frequency of data points in the Brooklyn vicinity. The results of the K-means algorithm displayed various locations throughout the city that has high frequency of visitors. Based on this, a restaurant owner can see ample locations where they would want to open shop and receive a large amount of customers. From the first cluster we see that Italian Restaurants have high frequency of 13, second cluster shows that Pizza Places have high clusters of 27, the third cluster shows that Bars have a large cluster of 23 and finally the fourth cluster



shows that Theaters have a large cluster. Following from that they can customize their services and environment in the restaurant that is best suitable for the customer to maximize customer satisfaction. From this they are able to build a loyal and regular customer base which can market their brand and quality of food for them.

Conclusion

Based on the K-means clustering algorithm we can see that there are various locations around NYC where there is a high frequency of visits by consumers. Based on that, opening shop around those locations and creating an environment suitable to those locations would be ideal for the restaurant owner to maximize customer satisfaction.