IBM Data Science Professional Certificate

Opening a Restaurant in New York City

1. **Introduction**
   1.1. **Background:** New York City is a popular tourist destination. According to NYC & Company's Annual report, New York City receives over 65 million visitors a year on average. As people visit the City, they will be looking for places to eat with a focus on restaurants that are popular in the City and close to where they are staying.
   1.2. **Problem:** Opening a restaurant in New York City focused on tourism can be a profitable business. The key is to understand where to open a restaurant and what type of cuisine to serve. This report will look at restaurant trends in New York City close to hotel locations to identify potential restaurant ideas. This report will focus on the center of New York tourism, Time Square.
   1.3. **Interest:** This information can be useful to investors looking to open a new restaurant in New York City with detail on location and type of cuisine. In addition, it may help existing restaurant owners assess whether or not their current business strategy is appropriate.
2. **Data acquisition and cleaning**
   2.1. **Data Sources:** Data required for this analysis consists of hotel and restaurant counts, locations, restaurant foot traffic, and restaurant categories. This information will be retrieved using the Foursquare API.
   2.2. **Data Cleaning:** To acquire the hotels data set a "Search" call was made to the Foursquare API using the search term "Hotel" and the geographic coordinates for New York Time Square. The results were organized into a pandas dataframe and cleansed to keep name and venue data as well as clearly label each venues category. As the search results contained venues with a category other than hotel, they were dropped from the dataframe.
   For restaurant detail, an "Explore" call was made to the Foursquare API to identify trending venues around Time Square. The results were organized into a dataframe. Similar to the hotel process, name and location data was retained. Venues not classified as Restaurants were removed from the data.
   2.3. **Feature Selection:** For this analysis we needed two data sets. First, hotel information based on latitude and longitude. This information was used to identify areas around New York City that may be more dense with tourists. Second, a restaurant dataset, based on trending locations, cuisine type, and geographic coordinates.
3. **Methodology**
   3.1. Identify tourist areas: The first part of the analysis is to identify areas around New York City that will have high tourist traffic. To identify where there may be a high density of tourists, I used a Search call for Hotels centered around New York City Time Square

from the Foursquare API. After cleansing the data, I was left with 34 hotels, which were mapped in blue as shown below.

3.2. The 34 hotels were then grouped into three clusters utilizing K-Means clustering algorithm. K-Means clustering is a machine learning algorithm that groups items into clusters based on similar data points. To create the clusters, latitude and longitude were used. Based on visual inspection of the hotel locations, it was decided to group the hotels into three clusters. Each hotel was assigned to a cluster, resulting in the revised map below (Color designating cluster).

3.3. Using the output of K-Means, the centers of each cluster were also mapped. The center of each cluster will feed into later analysis.

3.4. Next, trending venues around New York Time Square were identified using an explore call. The initial call identified 100 trending venues. After isolating restaurants, 21 venues remained, which are mapped below. From these restaurants, restaurant type and location data will be used for further analysis.

3.5. Using the location data of each restaurant, it was classified into one of the three clusters created in the earlier analysis of hotels. To do this, the K Nearest Neighbors ("KNN") Classification algorithm was used. This is a machine learning algorithm that helped assign the restaurants into categories. This information was added to the dataframe. A map of the restaurants by cluster is show below.

3.6. Based on the above, the data was broken down to look at frequency of trending restaurants in each cluster and frequency of restaurant category across all trending restaurants.

categories	counts
American Restaurant	4
Italian Restaurant	3
Seafood Restaurant	2
French Restaurant	2
Japanese Restaurant	1

4. Results – See above for mapping of restaurants and hotels. This section will discuss the resulting output.

4.1. As a result of the hotel clustering analysis, the following was observed 19 hotels in Cluster 1, 13 hotels in Cluster 2 and 2 hotels in Cluster 3.

4.2. Using the KNN algorithm, trending restaurants were organized into the same clusters with 1 restaurant in Cluster 1, 17 restaurants in Cluster 2 and 3 restaurants in Cluster 3.

4.3. Of the trending restaurants in New York City around Time Square, American restaurants were the most popular, followed by Italian, Seafood and French.

5. Discussion

5.1. Based on the results above, it appears Cluster 1 is underserved when it comes to trending restaurants. The area has 19 hotels but only 1 trending restaurant. This would suggest possible opportunity for an investor looking to open a restaurant focused on serving tourists as competition is low but density of tourists is high.

5.2. When it comes to cuisine type, American cuisine appears to be most in demand around Times Square. An American restaurant located in Cluster 1 may be a good investment for a new restaurant.

6. Conclusion

6.1. The objective of this analysis was to identify a suitable location for a new restaurant focused on tourism in New York City. In addition, the analysis looked to gain an understanding of popular cuisine types in New York City.

6.2. Hotel data was used to identify clusters and assess tourism density. The K-Means Clustering algorithm sorted the hotels into three groups and provided a center point for each cluster.

6.3. Trending restaurant data was then obtained to find popular dining locations in the same general location as the hotels. The KNN Classification algorithm was used to sort the restaurants into the same clusters as the hotels. In addition, the category of each restaurant was used to assess dining trends.

6.4. Based on the results, Cluster 1 had a high volume of hotels, but a low volume of popular restaurants. An investor looking to open a new restaurant should investigate that location.

6.5. An American cuisine restaurant in that area may be the optimal choice as it is a popular cuisine choice in New York City.