IBM Week 5 Capstone Project

Opening an Italian restaurant in Chicago, IL, USA

Introduction:

Restaurant is a place we choose not just for food but also to spend time with our loved ones. The restaurant business is one of the most competitive business which has flourished in huge numbers in the recent times. Specially in a metro city like Chicago, the Restaurant industry has a huge potential and there are established players in this business in the market. One of the most sought-after cuisines is the Italian cuisine. This project will focus on Italian chain of restaurants to find the most suitable neighborhood in Chicago, Il.

As per a recent study, there are total 92 cuisines in Chicago with 3.67 restaurants per square mile. Hence choosing the right neighborhood with minimum to NO competition is very important for the restaurant to make profit.

Problem Description:

Darden Restaurants Inc an American multi-brand restaurant operator headquartered in Orlando. The board has decided to open an Italian restaurant in Chicago area.

The objective of this project is to find the most accurate neighborhood based on location as a the feature, which will be profitable for Darden company to open an Italian restaurant and avoid conflict with any other restaurant of same cuisine.

Source of Data.

1. The list of neighborhoods is obtained from the https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago. There are 246 localities/neighborhoods in Chicago.
2. The latitude and longitude of each neighborhood is obtained using geocoder package.
3. The list of Italian restaurants in each neighborhood is obtained using Four Square API using venue endpoint.
4. The neighborhood data has 364 unique categories of venues obtained from the Four-Square API.

Methodology.

- Using web scrapping technique, the wiki page is read into a data frame using beautifulSoup package in python.
- Using geocode package, the latitude and longitude for each neighborhood is acquired.
- All the venues in each neighborhood are obtained using Four Square API.
- There are total 364 venue categories. The frequency of occurrences of each venue is then calculated.
- **Number of clusters is set to 3.** Dropping the neighborhood column from the dataframe, the dataFrame containing the frequency of occurrences of Italian restaurant for each venue, is fed into a new dataframe named **chicago_clustering to** train **the K-means** model named **kmeans**
- The **kmeans** model is now trained based on the clusters with 3 clusters.
- The Chicago_merged data frame is copied from chicago_restaurant , with labels added from kmeans, and then joined with original neighborhood dataframe on neighborhood column.
- Folium package is used to create a map from chicago_merged dataFrame

Results:

The results from the K means clustering shows that
Cluster 0: Has approx. 50 occurrences of Italian cuisine restaurant with 242 neighborhoods
Cluster 1: Has 1 neighborhood with 0 occurrences of Italian cuisines restaurant
Cluster 2: has 3 neighborhoods with 0 occurrence of Italian cuisines restaurant

Discussion/Suggestions/limitations:

The results show that the there is high concentration of Italian restaurants in Cluster 0. Whereas less concentration of to no Italian restaurants in cluster 2 with 3 neighborhoods. Cluster 1 also as no restaurants but has only 1 neighborhood based on the latitude and longitude information. **Hence, neighborhoods belonging to cluster 2 are more likely and recommended to open a new Italian restaurant to be profitable based on location.**

Please be advised that these recommendations are based on location data. there might be other factors like population, income of people, type of community in that neighborhood etc.

.

Conclusion

Cluster 1 and 2 has zero concertation of Italian restaurant in the Chicago neighborhood

Cluster 2 is the most accurate location with least or NO concentration of Italian restaurants in the city of Chicago Illinois in the US to open a new Italian restaurant in Chicago. This conclusion is based on location data only which meets the objective of the project. As mentioned there will be other features/dependencies like area population, income, eating preferences of people in that area etc.