The Battle of Neighborhoods

Capstone Project - Coursera

Business Problem

It became very challenging for the stakeholder or for the new business to decide in which area they should start their business in order to obtain the highest revenue with the least possible competition.

This project tries to solve the problem above by suggesting to the Target Audience which is the best place to open new restaurants and obtain the maximum profits in Toronto.

Data

The following data sources will be needed to extract / generate the necessary information:

- candidate area centers will be generated algorithmically and approximate addresses of centers in these areas will be obtained using https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- number of restaurants and their type and location in each neighborhood will be obtained using the Foursquare API.

Methodology

To solve the problem, I will use the "K-Means clustering algorithm". K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (that is, data without defined categories or groups). The purpose of this algorithm is to find groups in the data, with the number of groups represented by the K variable. The algorithm works iteratively to assign each data point to one of the K groups based on the characteristics that are provided. The data points are grouped based on the similarity characteristic.

Analyse

Three steps:

1. Data identification, capture and cleaning
2. Explore Toronto neighborhoods
3. Clustering

Results and discussion

Based on our initial assumption that the cluster with the maximum number of restaurants will have the best chance of having a new restaurant due to the need of the area. Based on the resulting clusters, it appears that Cluster 1 and Cluster 5 have a greater number of restaurants than the rest of the clusters.

It is quite possible that there is a very good reason for the small number of restaurants in any of these areas, reasons that would make them unsuitable for a new restaurant, regardless of the lack of competition in the area. The recommended zones should therefore only be considered as a starting point for a more detailed analysis that could eventually result in a location that not only has no competition nearby, but also other factors taken into account and all other relevant conditions met.

Conclusion

The final decision on the ideal location of the restaurant will be made by the stakeholders based on the specific characteristics of the neighborhoods and locations in each recommended area, taking into account additional factors such as the attractiveness of each location (proximity to the park or water), noise levels / proximity to main roads, availability of properties, prices, social and economic dynamics of each neighborhood, etc.