The Battle of Neighborhoods

Segmenting Sydney

Introduction

"Black Hide Steakhouse" is Brisbane's finest steakhouse restaurant that has over 60 years' experience in the steak dining scene, especially wagyu and angus steak cuts, all 100% halal certified. The owners of the restaurant want to expand their steak business in the country and they are very interested in opening a new restaurant in Sydney city.

Sydney is the most populous city in Australia and Oceania, with a metropolitan population of more than 5 million. Sydney was made up of 658 suburbs, 40 local government areas and 15 contiguous regions.

The Black Hide team has asked us to a study of different places in the city in order to get the best spot to open their second steakhouse that can guarantee them success.

The team has determined that we must pay special attention to tourists' hot spot and activities services, because they considered tourists as their primary customers. They also stresses that there are many steakhouses in the city, so we must pay attention to the steakhouses locations to avoid splitting of customers

The project will include the obtaining information of the neighborhood and make recommendation of the best places for the client's restaurant.

Data

The required data collected:

- Sydney neighborhood information, which is by the postal codes from Wikipedia and Corra website: http://www.corra.com.au/australian-postcode-location-data/
  - Sydney information is important for us to get the list of postal code thus acquire different neighborhood in Sydney.

- Location data and common venues of each postcode are collected using FOURSQUARE API.
  - By using Foursquare services, we can acquire most common venues in each neighborhood. We can then segment the neighborhood and see what each cluster consist of.

Methodology

The case development is done using a Jupyter Notebook. Data is processed and fed into the machine learning algorithm. Following steps are taken throughout the development:

1. **Build neighborhood list**
   The list of neighborhood is acquired from Wikipedia and Corra. As result, a dataframe called syd_df is created.

2. **Data wrangling**
   Data is processed and analysed to obtain initial insights.

3. **Venues compilation**
   Location data is obtained using Foursquare service. Location data contains 10 most common venues for each postal code. The location data is then sorted according to the commonness of the venues. New dataframe is created.

4. **Neighborhood segmentation**
   The dataframe is fed into clustering algorithm. For the project, kmeans clustering is chosen for its suitability. Elbow Curve is used to determine the best number of clusters to include in the algorithm. From the Elbow Curve, we determine the best number of clusters is 5. As a result of the cluster, a new dataframe is created consists of segmentation of neighborhood.

5. **Segment Analysis**
   Each cluster is analysed and studied to determine the best region to establish the steakhouse.

Results

Five clusters are defined as a result of segmentation.

- **Cluster 1**
  It is a big cluster that made up of cafes, Japanese restaurant, grocery shopping and shopping mall. This region is more likely a residential area which does not locate tourists' hotspots. Thus, the region is not suitable to open a steakhouse.

- **Cluster 2**
  This cluster consists of Middle Eastern restaurant, Thai restaurant, supermarket and Climbing gym. There is no interesting places for tourist hence it is not suitable.

- **Cluster 3**
  This cluster is made up variety of places such as airport, restaurants, hotels, cafes, market and pub. This region is located in the middle of the city where tourists converge. The region would be the best places to open a new restaurant.

- **Cluster 4**

It is small cluster made up of electronics store, Chinese restaurant and market. This region is more likely to be a residential area thus it is not suitable for a restaurant.

- **Cluster 5**
  Small cluster consisting playground and restaurants. Not suitable for the restaurant.

Conclusion

We have gathered the data and applied strong method to process the data. A region that would be the best region to establish the restaurant in achieved.