Introduction

Growing technologies and advancement in business development is increasing at an exponential rate.Many tourist spots are sprouting and other factors which depend on them are also growing parallely like restaurants,cafes,parks,malls etc.Mainly the idea for choosing a new place to open a restaurant involves many things like the popularity of that place,other kind of similar restaurants in that locality, Number of tourist visits per month/year and so on. Above all that money plays a major role for the same. So a wise decision should be made.

Business problem

With the above discussed thing in mind, I would like to address the issue of choosing a right place for opening a restaurant so that it would be a win-win for both owners and tourists.

Target audience

This project would play a major role for owners who are willing to start a new restaurant and for tourists who are in trouble finding a good restaurant I.

Data

To solve the problem, we will need the following data:

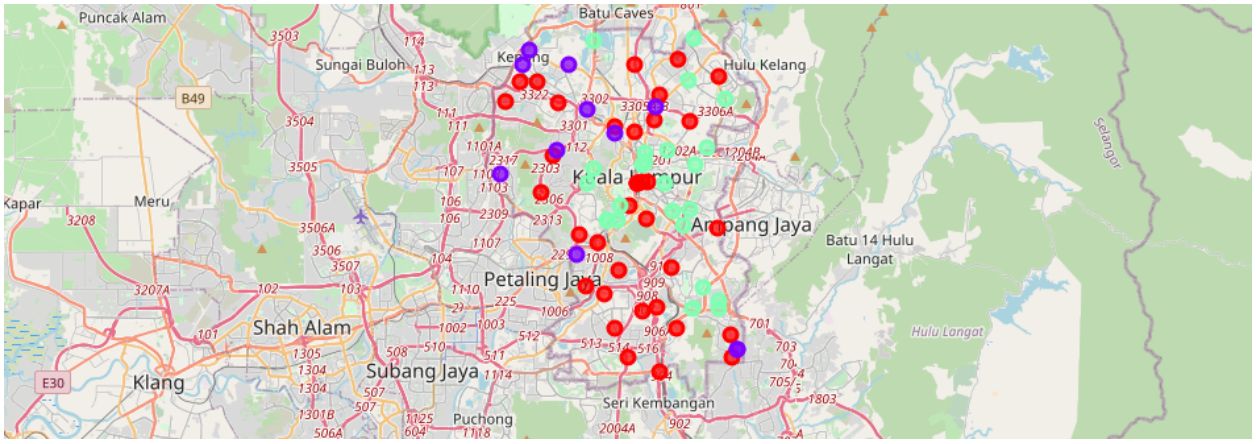
- List of neighborhoods in Kuala Lumpur. This defines the scope of this project which is confined to the city of Kuala Lumpur, the capital city of the country of Malaysia in SouthEast Asia.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

Methodology

Firstly, we need to get the list of neighborhoods in the city of Kuala Lumpur. We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data.We need to get the geographical
coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so,
we will use the Geocoder.After gathering the data, we will populate the data
into a pandas DataFrame and then visualize the neighbourhoods in a map using the Folium package.
Next, we will use Foursquare API. Foursquare will return the venue data in JSON format. Then, we will analyse each
neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category.we will perform clustering on the data by using k-means clustering. K-means clustering
algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.The results will allow us to identify which neighbourhoods have higher concentration of restaurants.

Results

The following image of the map shows the observed results.



Conclusion

So From the results we can infer that cluster 0 with color red has more number of places and it would be more suitable to open in any one of the places.