Week 1 Neighborhood Battle L.V.B.

Introduction of business problem:

Madrid is the capital city of Spain, one of the famous characteristics of Madrid is its diverse type of restaurant. Furthermore, the density of restaurant in Madrid is significantly high, people can always find what they want to eat around them. It is very convenient; however, it is also a big issue for entrepreneur who wants to open a restaurant in the city Centre due to costs and conglomeration of competence. In this analysis we are going to see which place is ideal to open a new 'Japanese' restaurant. We will combine our analysis with machine learning technique and the map data of Foursquare API and Wikipedia.

Data requirement:

1. Madrid dsitrict data from Wikipedia.

2. Latitude and Longtitude from Geocoder package.

3. Venue related data from Foursquare API.

Methodology

I decided to use K-means since its very straightforward and easy to achieve, it is also sensitive to noise and for this situation I think K-means is a good fit.

Results

Cluster 0 seems to be the most promising cluster where to start seeking for the ideal district

Insights and discussion

Once we in there, we see that district of *Salamanca*, along with *Chamberí* show attractive characteristics (not too much presence of this food restaurants, and high diversity within this business). Also, they show high population and high GDPpc, as well as great connections with the rest of the districts throughout metro stations, easy to find carparks, and city entrances and exits nearby.

Conclusion

We took into consideration different aspects that might influence the entrepreneur's decision of opening a new not-Spanish restaurant. Thus, we looked at the diversity to the district, as well as the presence of other gastronomies coming from Asia (Chinese restaurants, Cantonese restaurants, noodle Houses, Korean Restaurants, etc.)

Also, we looked at their population, and the near by district's population as they form potential customers. This is the reason why we discarded Barajas district since the beginning, even though it had a high presence of the previous points.

Finally, we discovered our ideal cluster (cluster 0) and digging deeper to select the final district. The final decision was highly influenced by domain knowledge of the geolocations of these districts in the city of Madrid; Salamanca's district is a highly active area in the city (day and night) and one of the highest GDPpc in the city, the same happens with Chamberí district. All in all, let's give a try…