The Battle of Neighborhoods

Introduction

When you have to book a hotel in a city that you don´t know, it is hard to know which hotels are better in terms of location. The aim of this project, is to group the hotels of one city by, we can say, its "flavour". By flavour, we mean how the taste of the sorrounding area of each hotel is. I think that selecting only the venues of our interest when we are travelling to another city, bucketing the hotels sharing common venues, and finally using wordmaps to get the big picture, we can have a proxy of what would be the taste of the sorrounding area of each hotel. To test this project I choose the city of Boston, as I was travelling very frequently there for almost three years, and I really want to come back again someday. The venues that we are selecting are:

- Monument / Landmark
- Historic Site
- Restaurants
- Coffee Shop
- Food & Drink Shop
- Bakery
- Park
- Museums
- River
- Irish Pub
- Steakhouse
- Lake
- Breakfast Spot
- Rental Car Location
- Pharmacy
- Harbor / Marina
- Street Food Gathering
- Train Station
- Bus Line

Data

The data we are using within this project are:

- Foursquare Data
- Coordinates of five selected areas from Boston. This coordinates was manually obtained using google.

Methodology

The steps followed to get the final buckets and its linked wordmaps are:

- We choose five stragegic points of Boston to cover the area of interest. There is a map showing where the five points are.
- We make a query to Four Square for each of the previous points to get all the sorrounding hotels (radius = 10.000 meters).
- Then we have a list of Hotels with its coordinates. We drop duplicates and plot all the hotels in a new map.
- Next step is to make a query to four square getting the sorrounding venues of each hotel (radious = 500 meters).
- We filter only the desired venues and normalize the data.
- Using the normalized data we perform the K means algorithmn setting the number of buckets to 7.
- We plot the bucketted hotels in a new map.
- Finally we generate one wordmap for each bucket so we can see what the flavour of each bucket is.

Results

We perform the elbow method with no clear results, but it seems that around K = 10 the slope tends to decrease at a slower rate. We try several clusters mapping the results on the map and finally we choose K = 7 as the one with better results.
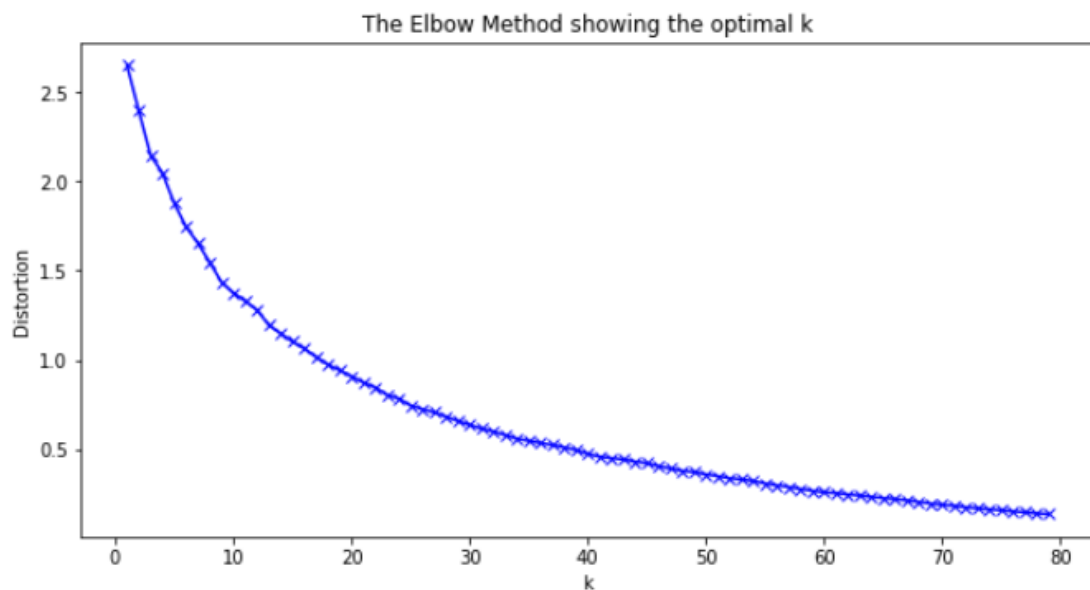
Running the code we get seven clusters with different tastes:
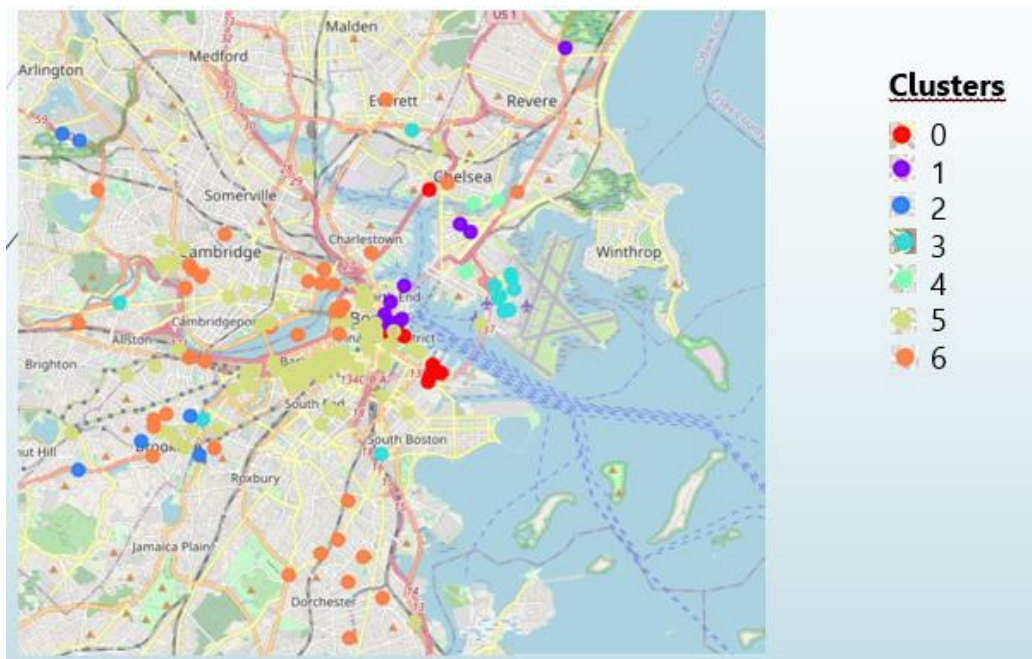
The idea is to use wordmaps to get the big picture of each cluster, so that we can choose the one we like the most:

Cluster 0

Cluster 1

Cluster 2

Cluster 3

Cluster 4

Cluster 5

Discussion

The city of Boston as a city close to the sea and therefore has a lot of sea food restaurants and also as it is a business city there are a lot of coffe shops. The venues are filtered by my own preferences but we can customize them to have different results. It is clear also that the restaurants are predominant in most of the clusters so a good idea would be to generate another sets of results dropping all the restaurants to have another vision of the city.

Conclusion

As this is only a first approach, using wordclouds and the K Means algorithm we can get the big picture of the sorrounding áreas of every hotel in a city.

This information can help us choosing one hotel when we have to travel to a new city.