WEEK 2

THE BATTLE OF NEIGHBORHOODS

MARIÁNGELES SALAS

INTRODUCTION

The objective of this project, is give the people a tool that they will can explore their neighborhood with, it will can help people than want to know about the best site to live in Carabobo Venezuela.
Carabobo is an industrial state, whence a lot of people need to migrate to this state, for this reason it is important to have a tool that can help people to make better and smarter decisions.
In this project I'm going to create an analysis of features for people who are searching about the best places, doing a comparative analysis between neighborhoods.
It will help people to get best decisions about the sites that they visit, improving their satisfaction.

DATA

Foursquare API:
In the project I will use Forsquare API, It has a big database which provides locations and details about business in this locations

Clustering Approach:
To compare the neighborhood, I decided to explore neighborhoods, segment them and group into clusters to find simillarities.
To cluster the data I need a form of unsupervised machine learning (K-means clustering algorithm)

Libraries
Pandas: For dataframes.
Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
Scikit Learn: For importing k-means clustering.
Geocoder: To retrieve Location Data.
Beautiful Soup and Requests: To scrap the web pages.
Matplotlib: to make plots
Numpy: To manage the math data

METHODOLOGY

In this project I detected areas in Carabobo, Valencia, next to it I found the most important venues.
In first step we have collected the required data: location of zone in Carabobo. We have also identified the principals venues (according to Foursquare categorization).
Second step in my analysis wasthecalculation and exploration of ‘venuesdensity' across different areas of Carabobo
In third and final step I create clusters of locations. I present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point to people who wants to lives in Carabobo.

RESULTS AND DISCUSIONS

The analysis shows that although there is a great number of venues in Carabobo, Highest concentration of venues was detected in ciudad alianza, Aguas Calientes, Mariara, guacara, aguas calientes y bejuma, so I focused my attention in this areas.
Those location candidates were then clustered to create zones of interest which contain greatest number of venues. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.
This, of course, does not imply that those zones are actually optimal locations for a new people to live, Purpose of this analysis was to only provide info on areas information about the zone with more venues varieties, it is entirely possible that there is a very good starting point to look the zones what fix best with the people.
Recommended zones should therefore be considered only as a starting point for more detailed analysis which can integrate other information like prices of houses, schools etc.

CONCLUSION

Purpose of this project was to identify the zones with the more venues quantity, to give the people who wants to live in carabobo a best perspective about the zone. By calculating venues density distribution from Foursquare data we have first identified the zones with venues. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of venues).