The Battle Of Neighbourhood Report(Week 4)

Introduction

In cities across the world there are neighborhoods that concentrate most of the activity and other neighborhoods that concentrate most of the housing space. These heterogeneities force people to move from their living places to the working and leisure centers, inducing road traffic that generates pollution and wastes people's time, especially in cities where public transport systems are not well adapted.

In this project I want to explore the differences between areas of a city, analyzing the kind of activities that create the most important leisure centers and exploring the changes, new businesses or activities, required to be created in some regions to make them more similar to the areas that concentrate the activity. The final aim would be for decision-makers to have a tool to design policies to reduce the differences between areas of a city, increasing the homogeneity and making a more livable city for the people.

I am going to focus the analysis on Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario's north western shore. Knowledge of field conditions is important for the development of this project, as some subjective information on the quality and development of the different areas of the city is required. The methodology, however, can be applied to any other city.

Data and Methods

The data used for this project will be obtained from Foursquare (https://es.foursquare.com/) , The postal codes were obtained through web scraping from (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) using beautyfulsoup package and the geospatial data of Toronto was obtained by (https://cocl.us/Geospatial_data)

All the data is combined into a dataframe that contains all the explicative variables for every section of the city. The information is then vectorized to include a count of the number of services offered in every section, that is, the total number of parking slots, total number of gardens, total number of parks is computed. This information is complemented with the information obtained from Foursquare, thus including

information about the number of cafeterias, restaurants, clothing shops and other businesses in every section. The so formed dataset will be the working dataset for the project.

A clustering algorithm will then be used to classify the different section into groups of similar sections. The total number of clusters will be selected in such a way that the clustering is compatible with the field experience, that is, ensuring that each cluster really represent a quantity and quality of activity that can be observed on the field. If discrepancies are observed between the clustering algorithm output and the classification based on experience, different machine learning techniques will be tested to map the section characteristics to the group assigned by experience and field work.

The objective of the previous task is to construct the decision boundaries that separate the different classes. Once these boundaries have been constructed, it will be easy to compute the minimal set of changes that should be implemented into one section of the city to transform it into a different type, that is, how to modify one section to change the overall level of activity and make it more similar to the objective that we may be pursuing as decision-makers.