Clustering of the Island of Montreal

Noémie Rohel

February 13, 2021

1 Introduction

The neighborhoods of a city, especially the large ones, may be very different; the population, the real estate and the accessible venues can vary a lot even within the same area. However, most of the neighborhoods share a lot of characteristics with others. Therefore to have a quick and easy to understand description of a city, it would be interesting to identify the different kinds of neighborhoods and to associate all of them to a specific profile. This project aims to help people who want to move in to select the best place or people who want to start a business to select the best location. This study is addressing this issue for Montreal.

The project aims to cluster the Island of Montreal into similar neighborhoods based on the environment (density of restaurants, bars, cultural activities, schools, ...), the population (density of population, age, income, ...) and the housing (value, type of dwellings, ...). For each of these categories, a set of specific features is selected and a clustering method is used to gather the similar neighborhoods. Three clustering approaches are considered and implemented: K-Means, DBSCAN and agglomerative clustering, and the results are compared to identify the best method.

First, the data sources used for this project are presented, as well as the required data cleaning and preprocessing. An exploratory data analysis is conducted to highlight some preliminary observations about the collected data. Then for the three clustering approaches considered, a short description and a parameters selection is performed and the results are used to compare the methods. Finally, the clusters obtained with the best approach are described and analysed.