Professional data science program – IBM

Week 4 assignment

Affordable rental property in the Randstad area of the Netherlands

By Ali Soleymani

March 2020

Introduction

The Netherlands is one of the small countries in the Europe where finding a place to live is usually difficult and need spending a quite huge amount of time. This problem is more intense in the Randstad area of the Netherlands. The Randstad is a megalopolis in the central-western Netherlands consisting mainly four largest Dutch cities (Amsterdam, Rotterdam, The Hague and Utrecht) and their surrounding areas. Most of the business and research institutes located in this area. So, most of the expats live in this crowded area. In general, the Netherlands face housing shortage and it is estimated to get worse in the next few years.

Problem

It is important for residents of the area, job seekers and the government and policy makers to know more about the housing situation in each city of this area. More importantly, finding an affordable place to live in this area is important. So, here we will use data science techniques to classify and cluster the cities in this area by the number of available affordable price rental property. It can help the people to have a better decision and decrease their worries and stress of finding a place.

Data

For solving this problem and answer the questions we are addressing here, we need the following data.

- List of cities and their latitude and longitude.
- Average of housing prices and number of available properties

Source of data and extracting methods

Funda.nl is one of the main housing property providers in the Netherlands. It contains a list of available properties and their prices. I will use web scraping technologies to extract the data from this website using Python requests and beautifulsoup packages. Then we will use the geographical coordinates of different cities using Python Geocoder package.

Then, we will use foursquare API, data cleaning, data wrangling, K-mean clustering and map visualization (Folium) to solve our problem.

In the methodology section we will discuss the procedure in more details.