Real Estate Pricing Analysis in Germany

Introduction-Business problem

Germany is one of the biggest and most financially powerful countries in Europe . The population of the country is radically increasing annually due to its sustainable economic and industrial development. Since 2012 more than 1.000.000 people every year immigrate to Germany, mostly due to their need for financial growth.
The biggest percentage of the immigrants decide to move to big german cities such as Berlin, Hamburg, Munich etc. , cities that are more multicultural and offer more job opportunities to foreigners. This situation leads to an enormous increase concerning the real estate demand, and thus a research in that field becomes very interesting.

Real estate companies, Industrial companies, even individuals become more and more interested in housing price, since it turns out to be a quite sufficient life investment. A closer look to the current housing pricie situation in Germany appears to be very appealing to individuals and companies which have a tendency to economic growth.
This survey is a Machine Learning problem and more precisely a Clustering problem. The machine learning method which will be used is the K- Means method .For acquiring information concerning the venues located close to each area the data platform Foursquare will be utilized.
The purpose of the project is to create an appropriate number of clusters based on the price, the quantity and the kind of venues located in each german state (Bundesland).
 In the end of this project we will be able to see the difference between the german states according to the housing price and what kind of attractive destinations each area has to offer. Additionally, we will acquire a general view of the most visitable venues in each Bundesland.

Data

For the purpose of this survey a dataset from kaggle.com (Germany Real Estate Housing Price) will be utilized. It contains data from April 2020 concerning the average housing price in different German States (Bundesländer) , the respective postal code and the current conditions of each house – appartment (number of rooms, year of construction etc.)
The postal code in this case is very important so that we know in which area of the relevant german state the house is located.
Except for the postal code the benefits of Foursquare will be utilized, since the quantity and the kind of venues that are closely located have a strong impact in the configuration of the prices. Venues such as Hotels, Museums, Restaurants, Bars as we already know , are features that are capable to transform an area to a very attractive one.