

The Battle of Neighborhoods (Week 1)

Background

This project is for those who are planning to move in to a new apartment in the city of Vienna. It provides a suggestion on what would be the best district in Vienna to get an apartment which has already many good apartments. Vienna is Austria's most populous city, with about 1.9 million inhabitants (2.6 million within the metropolitan area, nearly one third of the country's population), and its cultural, economic, and political center. It is the 6th-largest city by population within city limits in the European Union.

Business Problem

This project focusses on the issue of finding an apartment to rent a city like Vienna which is the largest city in Austria. Potential stakeholders for this project would be people looking to rent an apartment. Business problem would be providing people with information like

- Which district has cheaper rent
- Choose to live in residential or commercial areas and can see for example which residential districts is best Or

If they already live in one of the 23 districts in Vienna, they will be able to see:

- If they are paying more than the average price for their apartment
- If there are similar districts to theirs with lower rents

Data & Usage

Two sources are identified as potential platforms of data we need for exploratory and descriptive analysis. One source is a popular apartment listings website 'willhaben.at' and the latter is 'Foursquare'.

We use web scraping to derive from the website and use Beautiful Soup library to parse and extract required field of interest like size of the apartment, number of rooms, address, price. Data is derived from Foursquare by making API calls to endpoints exposed by the data provider. Features of interest from this provider would be closet venues like supermarket, restaurant, theater, park etc.

After data wrangling procedures are performed on the collected data, it is taken to next step to cluster the districts into residential and commercial areas and visualize all the data on a single choropleth map using k-means clustering unsupervised machine learning algorithm.