# The Battle of Neighborhoods

Applied Data Science Capstone Project

Akash Apoorv Yadav

## Introduction

In this project, we will help people who are looking for renting an apartment in Vienna. If they are looking to move to Vienna they can see:

- Which district has cheaper rent or,
- They can choose to live in residential or commercial areas and can see for example which residential districts are 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

The data on apartments: size, number of rooms, address, and the price is collected by scraping a local website with apartment listings (willhaben.at). We clean up the values and calculate the price/$m^2$ by dividing the price by the size column. The data is pre-processed and we get our first data frame:

| | PostalCode | District | Size | Rooms | Price | Price/m2 |
|---|---|---|---|---|---|---|
| 0 | None | None | 130.0 | 4.0 | 2152.10 | 16.55 |
| 1 | 1010 | Innere Stadt | 104.0 | 3.0 | 2260.28 | 21.73 |
| 2 | 1220 | Donaustadt | 12.0 | 1.0 | 432.00 | 36.00 |
| 3 | 1120 | Meidling | 75.0 | 3.0 | 840.00 | 11.20 |
| 4 | 1120 | Meidling | 43.0 | 2.0 | 700.00 | 16.28 |

It is further cleaned by removing NA values and outliers, which results in a dataset of 8045 apartments. From this dataset, we extract the rows with a unique Postal Code, and then using Geopy we find the coordinates for each district.

Using Foursquare we collect the closest venues (supermarket, restaurant, park, etc.) and select the top 10 venues for each district. After the data collection, we can run k-means clustering to cluster the districts into residential and commercial areas and visualize all the data on a single choropleth map.