

Determining what are the cities similar to Munich

Introduction

Background

As a software developer you have a lot of opportunities and can apply to companies located in different cities or countries, different to the one you currently live in. Normally you get attached of how a city is, the benefits it brings and so on.

Using data to compare the cities could help on taking a decision of where to move or where to apply for a job.

Problem

Imagine yourself in a situation where you want to move because of a job opportunity but you like the city you currently live in.

The aim of the project is to determine the cities you could apply for or could move because they are similar to Munich and it can be extrapolated to other cities.

Data acquisition and cleaning

Data sources

A group of cities has been selected just based in preferences, in this case mine:

1. **Munich, Germany**
2. Berlin, Germany
3. Hamburg, Germany
4. Dusseldorf, Germany
5. Vienna, Austria
6. Madrid, Spain
7. Barcelona, Spain
8. Montreal, Canada

9. Toronto, Canada
10. Ottawa, Canada
11. Vancouver, Canada
12. Amsterdam, Netherlands
13. San, United States
14. Medellin, Colombia
15. Zurich, Switzerland
16. Sydney, Australia
17. Wellington, New Zealand
18. Edinburgh, United Kingdom

For this project, the data to be used will be composed of three parts:

1. City venues: [foursquare api](#) will provide which venues are in each city center in a radius of 1km, those venues later will be categorized and used to compare each city.
2. Population of each city using this dataset from simple maps of World [cities population](#).
3. Cost of living, Rent and purchase power of each city, for this we are going to use the [dataset](#).
For more information please look in the URL
<https://www.numbeo.com/cost-of-living/rankings.jsp>

Data cleaning

Foursquare API

Foursquare provides a JSON REST api, really easy to be used, to query the API an API KEY and secret required.

Using the Venues endpoint for each city to retrieve which venues are in the city center limited to 1km radius and 200 entries and obtain the category of each one. Those categories later will be used to see what is the proportion of each category venue and later compare them to other cities.

Population Dataset

Using the Simple Maps world cities population dataset, we will obtain the population of the cities we want to compare. This is useful to see how many people live on the cities and to see how they compare to each other.

To extract them we just use pandas and read the csv, and select the data we are going to use, for example Rent Index, Cost of living index, etc.

Cost of living and Power of purchase

Numbeo.com contains a lot of information regarding the cost of living in city, those are provided in html format and can be easily extracted using web scraping.

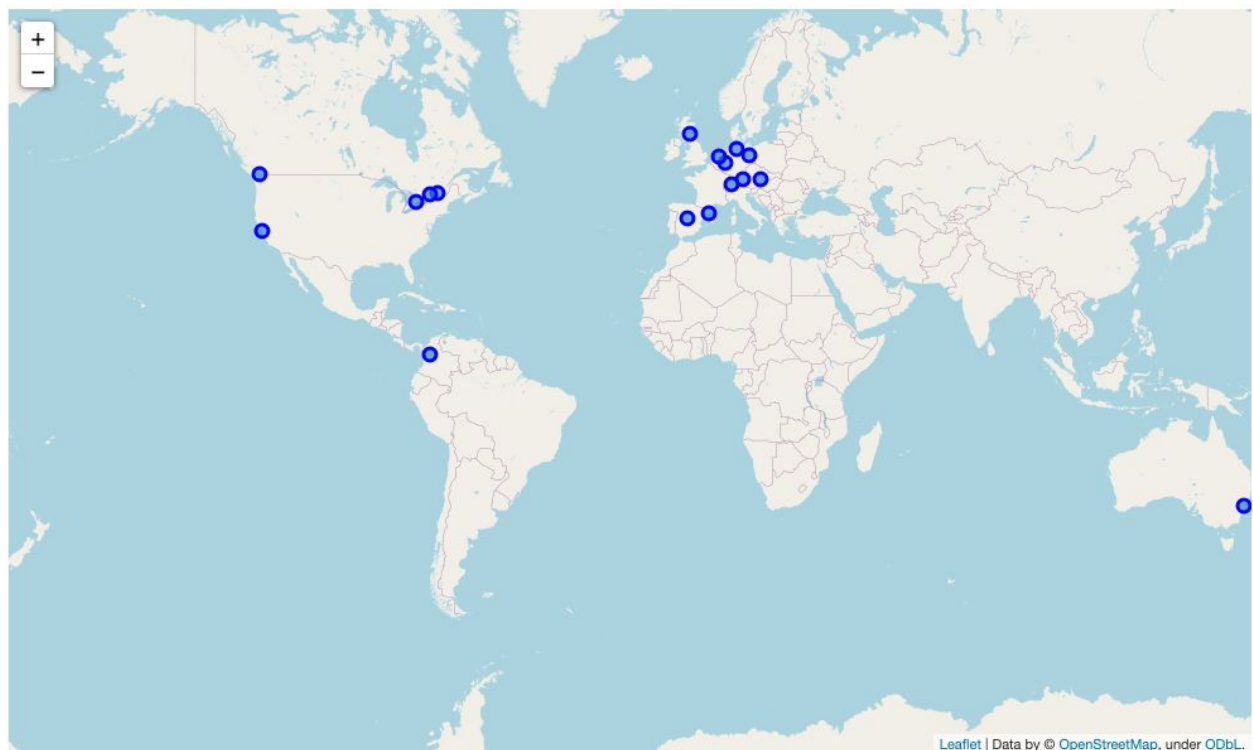
Pandas provided the option to read the HTML table and map it to a dataframe, the only required data cleaning was to remove the country name from the city name, to later be able to merge all the three datasets into one based on city name.

Exploratory Data Analysis

World map plot

This was used to show where the cities are located, most of them are located in Europe as shown in the map.

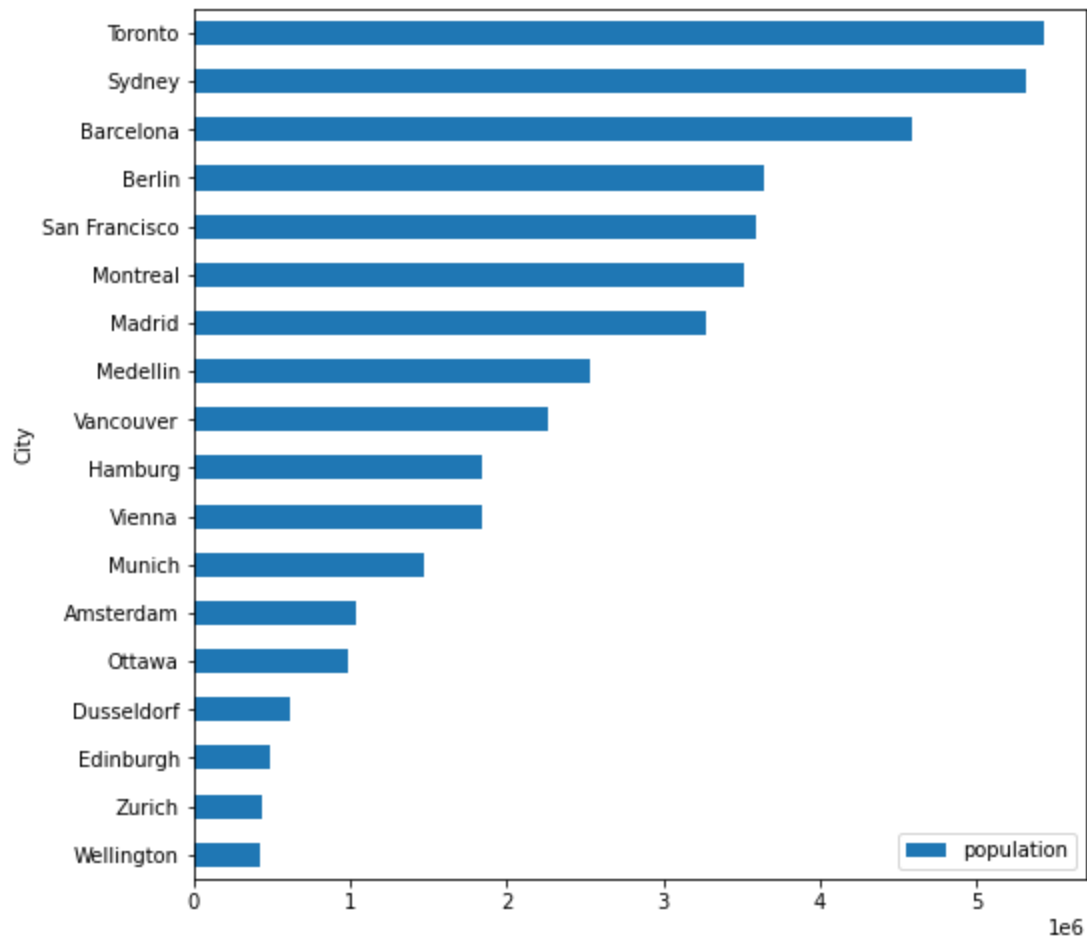
Those cities were selected based on my preferences and the job opportunities and companies located in those cities.



Population bar plot

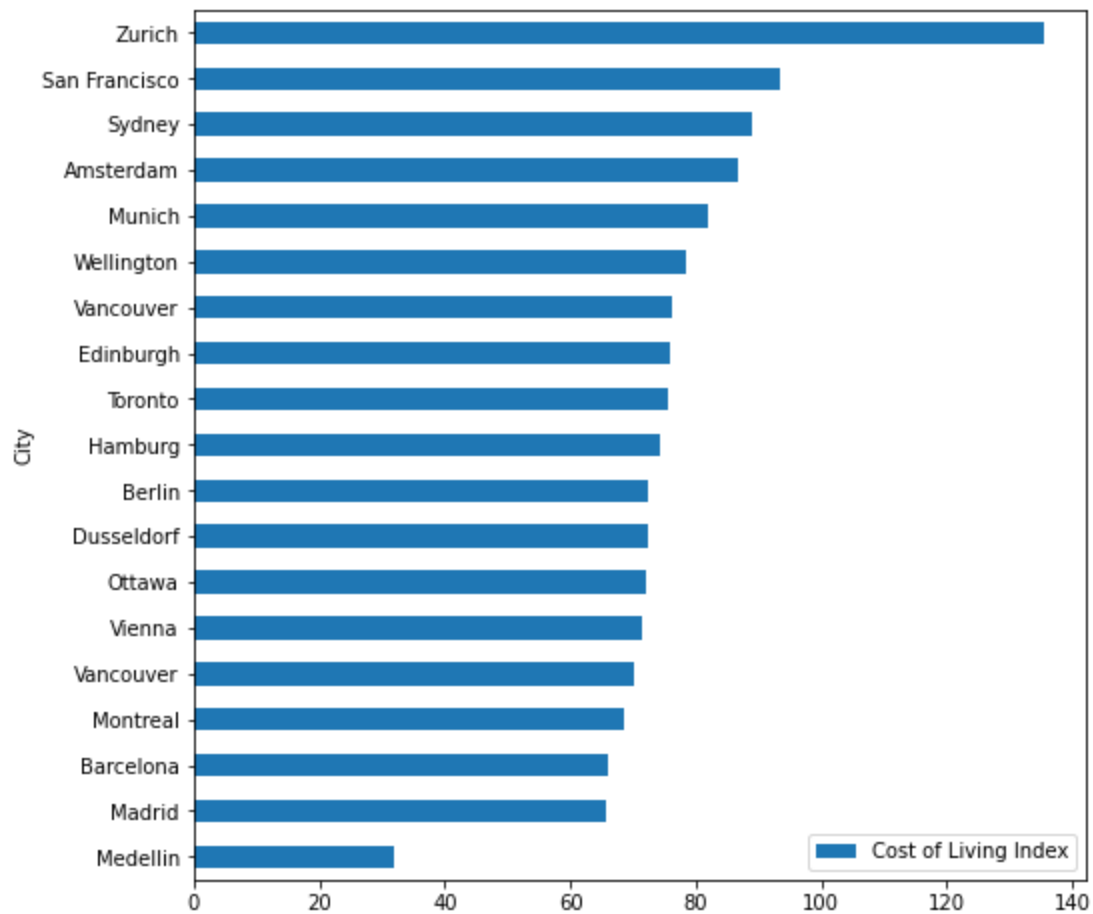
Having a bar plot to show how the city's population is distributed, helps to take a decision, as we can see Munich is a city with a small population and it's compared to Vienna, Amsterdam and Ottawa.

This information is useful to later choose which cities are comparable to munich



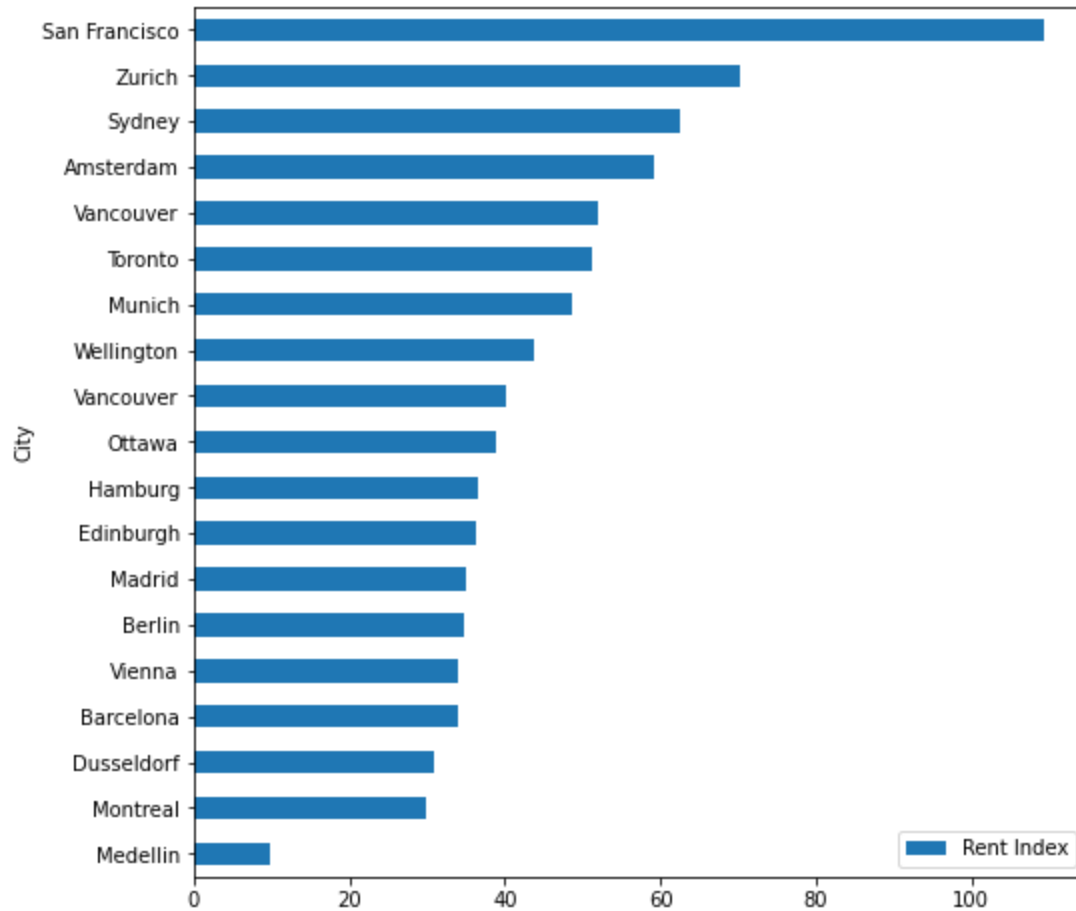
Cost of living plot

This plot shows how much living in the city costs. It can be observed that most of the cities are the same except for **Medellin** and **Zurich**.



Cost of rent per city

This plot shows the rent index by city, showing that Munich is comparable to Toronto, Vancouver and Wellington. Also San Francisco and Medellin are the exceptions.



Clustering the cities by similarities

Using K-clustering to cluster the cities we can obtain which cities are similar to each other using the next features:

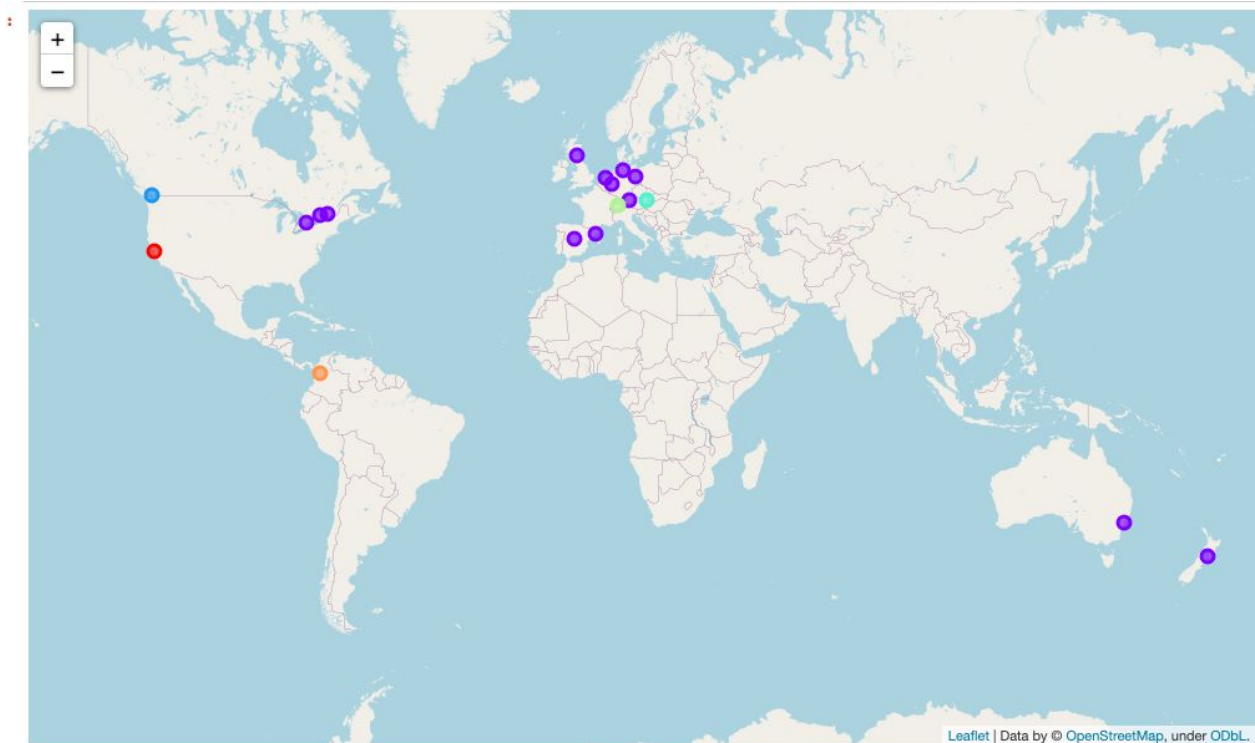
- Population.
- Cost of Living Index.
- Rent Index.
- Local Purchasing Power Index.
- All categories proportion of the venues.

Setting 6 clusters we can see the next result:

Cluster Labels		City
0	1	Munich
1	1	Berlin
2	1	Hamburg
3	1	Dusseldorf
4	3	Vienna
5	1	Madrid
6	1	Barcelona
7	1	Montreal
8	1	Toronto
9	1	Ottawa
10	2	Vancouver
11	1	Amsterdam
12	0	San Francisco
13	5	Medellin
14	4	Zurich
15	1	Sydney
16	1	Wellington
17	1	Edinburgh

It shows that a lot of cities are similar to Munich, showing that any of those cities can be chosen to move from munich.

It can be also seen in the next map, where the purple cluster is where Munich is contained.



Conclusions

In this study, I analyzed how Munich was compared to other cities in order to find which city I can move to if I have the opportunity.

We can see in the result that many cities are similar to Munich, probably adding more features like Job opportunities, salary and others can help to make the cluster smaller and easy to pick.

This study can be extended to many cities and can help other people with the same problem I tried to solve.