CAPSTONE Project

Final Assignment for 'IBM Data Science Certification'

by

Sven Viehrig

March 2020

Title

Preparing for a Move by Comparing the Old and New Neighborhood

Table of Content

Introduction

This work is the final assignment of a Coursera course called “IBM Data Science Certification” and will prepare young people for a move to a new city. Hopefully it will provide them with more confidence for the new era in their lives in a new neighborhood.

Management Summary

The data was gathered through scraping of a website and the Foursquare API. Some binning was done to group rental prices . The data was sliced in various way to obtain a direct comparison of the old and new home via a radar chart. Also a WordCloud was used.
The new neighbourhood is similar the old neighbourhood in terms of venue offerings but offers more cafes (10) , burger places, nightclubs, clubs and stores. Both locations have plenty of restaurants.
There should be no regrets moving to the new location.

Problem Statement and Purpose

When moving cities for work purposes , especially for young people, finding affordable rental housing/apartments in Germany these days is very difficult. Very often there are hundreds of applicants applying for the same house or apartment with queues forming outside at an open house to meet the real estate agent or landlord.
Often the applicants do not have the choice of their new neighbourhood and are happy just to find and be accepted for an affordable place within the city overall.

This report is intended to prepare them for their move by providing them with more confidence about their new home location (exact address) with a direct comparison to their previous home location in terms of venues in the vicinity through a Radar chart. The report will also map out the average rental price per suburb in the new city and compare the neighborhoods in the new city.
It will end with proving the top trending places in their close proximity of the new address.

The goal is also to make this report reusable for others with very little effort by just changing the addresses. This particular report will be for a move from Berlin to Munich in Germany.

Data Section

The data for the rental price per square meter in the new city will be obtained by scraping the website
https://de.statista.com/statistik/daten/studie/260438/umfrage/mietpreise-in-muenchen-nach-bezirken/

The latitudes and longitudes of the various addresses in Berlin and Munich with its suburbs will be obtained through the geopy / geolocator package .

The detailed data for venues and categories will be obtained by using the FourSquare API. Within FourSquare we will obtain a list a venues, venue category and top trending places within a certain radius of a longitude/latitude marker.

The main component of the report will be a direct comparison of the number of venues within the various venue categories between the old and new address. It will be shown with a radar charts with an overlay for easy comparison.

Methodology

As a first step I scraped the website listed in the “Data” section, which provided all the districts within the centre of Munich as well as the rental price for each district per square meter rental space for residential living.

The file had to be cleaned since the rental price was stated in cents and not in Euros. In addition the trailing zero was dropped, which had to be reinstated by multiplying the price by 10 if rental was less than 1001.

The next step was to obtain the geographical coordinates using the library geopy. In this step I also added the address of two additional 'districts' that I wanted to add to the dataframe, called “New Home” (in Munich) and “Old Home” in Berlin. Obtaining the respective latitudes and longitudes worked for all districts correctly with the exception of one. 'Milbertshofen-Am Hart' was not recognized by geopy, but dropping everything after the hyphen resolved the problem.

I then did some binning so that the rental price could be grouped into 3 categories that I called LOW, Medium and HIGH.

The rental ranges from 14.54€ per m² to 22.65€ per m². The NumPy package defined the groups as follows :

Low :	14.54€ – 17.20€
Medium :	17.21€ – 19.00€
High :	19.01€ – 22.65€

This then provided me with the following dataframe for follow on work :

	District	Rental	Latitude	Longitude	Euro	rental_bin
0	Altstadt-Lehel	2265	48.1378	11.5746	22.65	high
1	Maxvorstadt	2154	48.1466	11.5714	21.54	high
2	Ludwigsvorstadt-Isarvorstadt	2066	48.1318	11.5558	20.66	high
3	Schwabing, Schwabing-West	2047	48.1599	11.5762	20.47	high
4	Au-Haidhausen	1893	48.1288	11.5905	18.93	medium

In the next step I then used the Folium package to create an interactive map of Munich with a marker for each district reflecting the different categories by color.
Red symbolizes category High, yellow would be medium and green reflects the low category. Clicking on the marker would display the district name and rental price.

As a last step for the rental price I created a horizontal bar chart with districts in the y-axis and rental price in the x-axis (see next Results section for diagram) .

I then proceeded on to getting all the venues for each district using the FourSquare API. I selected various values for the radius and eventually settled on 750m for a representative selection useful for portraying. The number of venues per district is shown in a horizontal bar chart.

Using the “One-Hot Encoding” methodology was used to create a dataframe representing the number of venues per venue category for each district. It returned 216 venue categories :

```
Size is (30, 216)
```

Out[162]:

	District	Afghan Restaurant	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Austrian Restaurant	Auto Dealership
0	Allach-Untermenzing	0	0	0	0	0	0	0	0	0

I then tried to group certain categories , for example all categories with the word “restaurant” or “shop” etc. to try and reduce the number of categories. This then resulted in only 109 categories, which is more meaningful for displayable results.

```
Size is (30, 109)
```

	District	Art Gallery	Auto	BBQ Joint	Bakery	Bank	Bar	Beach	Beer	Bike Rental / Bike Share	...	Supermarket	Taverna	Tea Room	T
0	Allach-Untermenzing	0	0	0	1	0	0	0	0	0	...	2	0	0	1

From this point forward I focused only on 3 categories, namely the center and most expensive district, the New Home location and the Old Home location. Since I was not interested in the other districts there was also no need for clustering with K-means .
The venue categories were then sorted by the number of venues showing the 15 most common categories per district in order.

A radar chart was used to overly the “Old Home” district with the 'New Home” location for the most common venues. Here the Restaurants were so much more than any other category that I removed Restaurants in order to have a viewable chart for all the other categories.

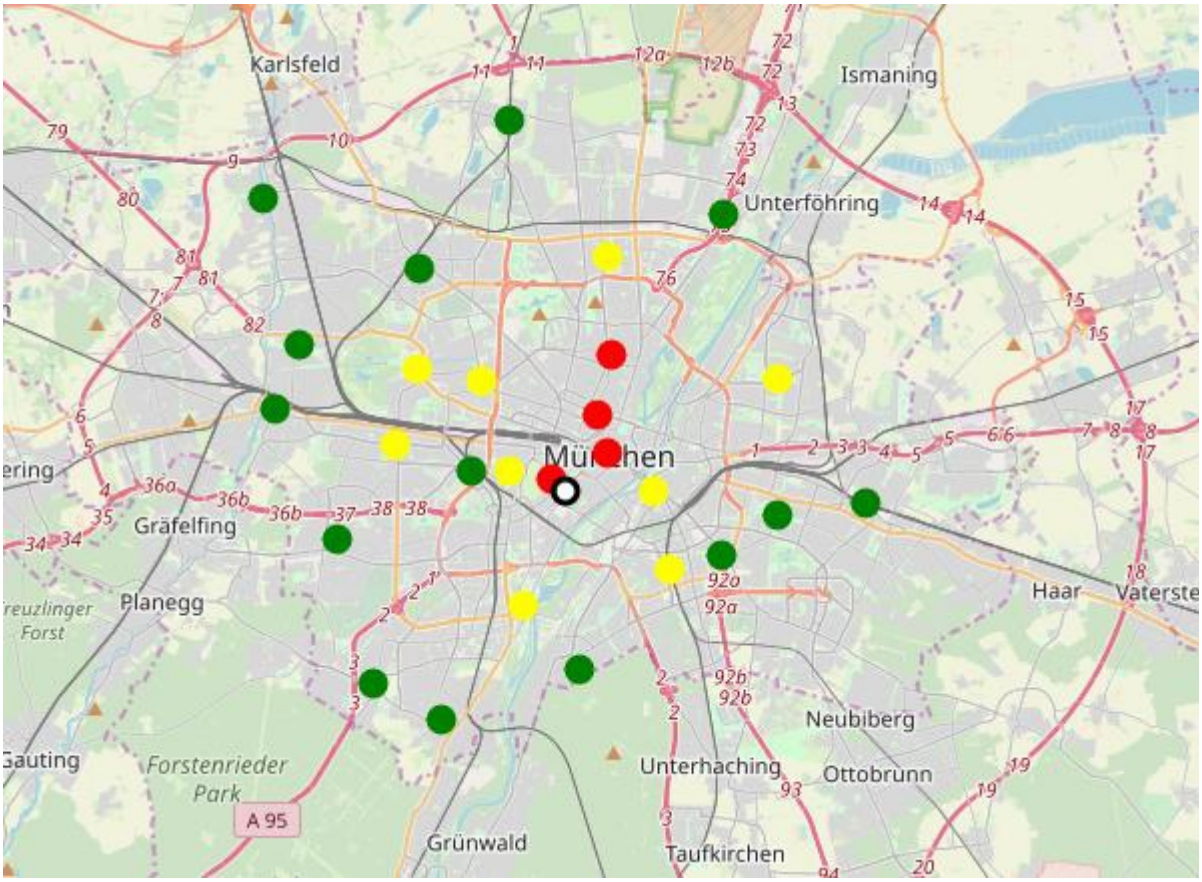
Lastly a “WordCloud” chart is used to represent the 15 most common venue categories for the most expensive district.

I also tries to find a trending spot, based on foot traffic , but it came back that there are no trend spots, which is understandable during the corona crisis.

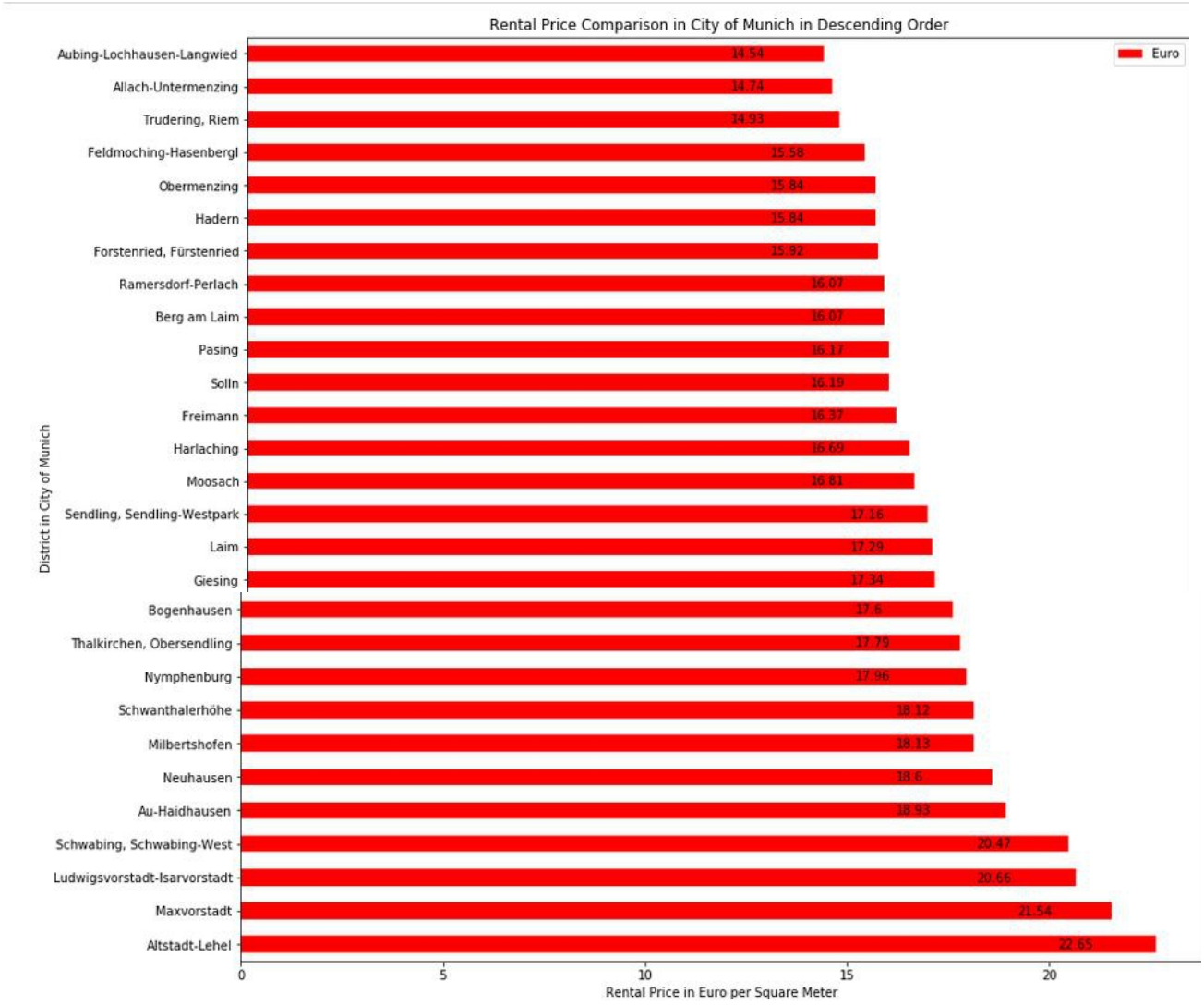
Results

The price per m² rental space is typical in the sense that the closer you are to the centre of Munich the more expensive it gets .

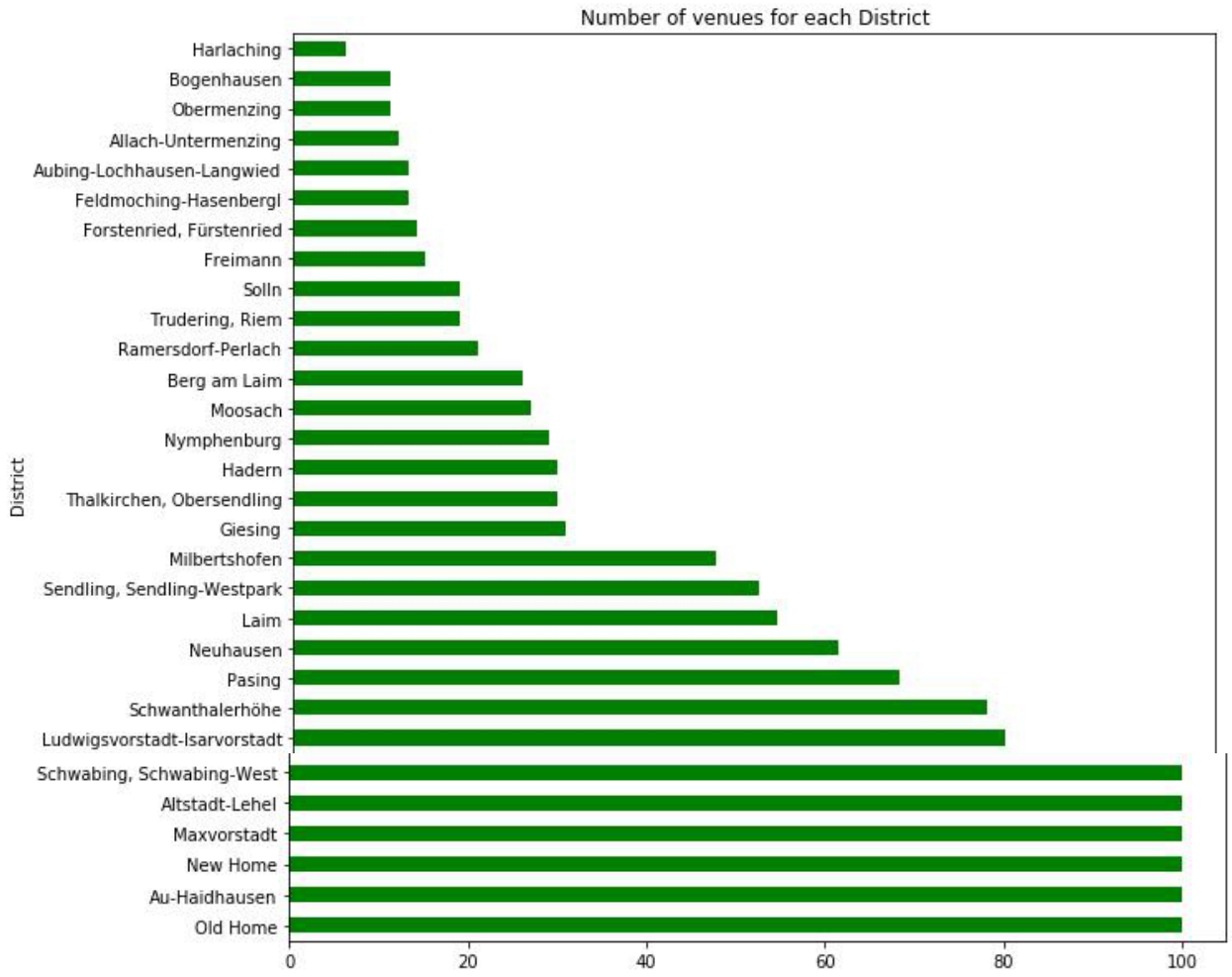
Red = High, Yellow=Medium, Green=Low, White=New Home



A different representation is the bar chart:



Following is a barchart with the number of venues per district.



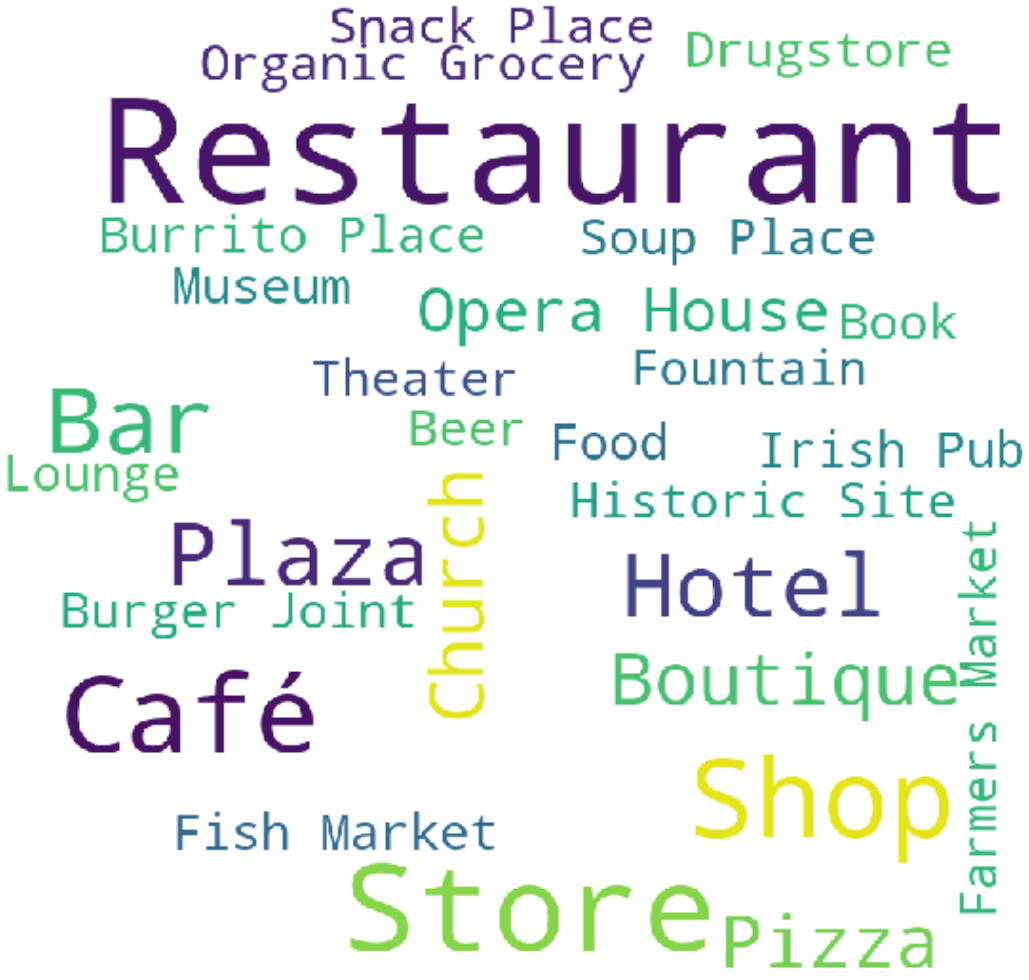
The 15 most common venue categories for the 3 districts are as follows:

District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Altstadt-Lehel	Restaurant	Store	Shop	Café	Bar	Hotel	Plaza	Boutique	Church	Pizza
New Home	Restaurant	Café	Bar	Store	Shop	Club	Burger Joint	Hotel	Nightclub	Sandwich Place
Old Home	Restaurant	Shop	Bar	Café	Beer	Pizza	Plaza	Breakfast Spot	Hostel	Store

District	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue
Altstadt-Lehel	Pizza	Opera House	Museum	Fish Market	Organic Grocery	Farmers Market
New Home	Sandwich Place	Pizza	Bistro	Pub	Plaza	Noodle House
Old Home	Store	Hotel	Bakery	Boutique	Church	Nightclub

The Radar Chart is a direct comparison between the Old Home and New home for the union of the most common categories of the 2 places, but ignoring restaurants. Both places had so many more restaurants that this chart was unreadable for the other categories.

The Word Cloud for the top district “Altstadt-Lehel” in Munich looks as follows:



Discussion

The new home location is in an area where rental is in the top 4 most expensive districts in Munich. On the other hand this district also offers a lot more venues in the immediate vicinity (750m) . Compared to the districts with lower rentals the number of venues is significantly higher.
In a direct comparison to the old home location, the new home location also has lots to offer with similar places . There is no shortage of restaurants in either location.
The new locations has double the number of cafes (10) in the 750m vicinity . Has more burger places, nightclubs, clubs and stores but fewer shops and bars.
Surprisingly Berlin had more beer places than this district in Munich, but I am sure the next beer garden in Munich is not far away.

Conclusion

The new location is no worse in terms of attractions and the top district is right next door. Hence one can move to the new location with no regrets.

The interactive maps and Python code can be viewed at :

https://nbviewer.jupyter.org/github/SvenVieh/Coursera---Capstone/blob/master/Capstone%20-%20Final.ipynb