APPLIED DATA SCIENCE

CAPSTONE PROJECT
THE BATTLE OF NEIGHBORHOODS/BOROUGHS

Relocation to Munich based on crime, housing prices and venues of Munichs boroughs

Presentation by Alexander Fuchs | 25th May, 2020

https://upload.wikimedia.org/wikipedia/commons/4/40/Vista_panorámica_desde_Olympiapark%2C_Múnich%2C_Alemania_2012-04-28%2C_DD_03.JPG

Introduction

(https://upload.wikimedia.org/wikipedia/commons/1/17/Muenchen_Kleines_Stadtwappen.sv g

- Background
  - *Munich is vastly growing, very expensive city in southern part of Germany.*
  - *Finding a adequate housing is a major concern when relocating to Munich.*
- Business Problem
  - *Using data on safety (crime statistics), housing prices and venues the 'best-fit' borough for a customer should be determined.*
- Interest
  - *Building a business for, especially international, customers who want/need to relocate to Munich and want a safe, affordable housing with venues of interest nearby.*

Data Acquisition

- Crime statistics: The city of Munich and its police department publish data for crimes only on a borough level, **not** on a neighborhood level. Therefore, all following investigations are performed on a borough level since no date on a smaller scale (neighborhood) is available [Data ported from a pdf-file of Muenchen.de into an csv- file. csv-file used for further investigations].
- General population data: General population data as population per borough and/or neighborhood, population density per borough, area of the borough, percentage of foreigners per borough to find a diverse borough are needed [Data imported from Wikipedia.org].
- Housing prices: Housing prices are taken and imported from one of the leading and most prominent German real estate sites Immobilienscout24.de. Since this site has a vast number of apartments offers in all boroughs and also its prices, it is a good bases for housing prices in Munich. The prices are Euro per square meter, when buying a three-room apartment with about 80 square meters of area [Data imported from Immobilienscout24.de].
- Venues: Venues are taken/imported from FourSquare.com using the API of FourSquare.

Data
Preview, pre-processing, and cleaning

- Data set from data acquisition is pre-processed, translated into an csv-file, and imported into a Pandas data frame and merged
  - *Geographical (latitude and longitude) data is added*
  - *'Total crimes' is added as additional derivative parameter*
  - *'Total crimes per capita' is added as additional derivative parameter based on 'Total crimes' and 'Population'*
  - *Housing prices are merged with the data frame after translation to a borough level of detail*
  - *Venues are computed and added using FourSquare and their API*

Methodology
Overview of ‘Total crimes’

- The data set is sorted in descending order and the parameter ‘Total crimes’ is shown here to give an overview.

- The boroughs of Munich are vastly different when it comes to crimes committed on absolute basis.

Methodology
’Top 5’ of ‘Total crimes’

- Sorting the crime data set descending for the ‘Total crimes’ leads to the ‘Top 5’ boroughs of Munich with the highest absolute number of crimes committed in 2018.
- The borough of ‘Ludiwigsvorstadt-Isarvorstadt’ shows to be the borough of Munich with the highest number of total crimes

Methodology
’Bottom 5’ of ‘Total crimes’

- Sorting the crime data set ascending for the ‘Total crimes’ leads to the ‘Bottom 5’ boroughs of Munich with the lowest absolute number of crimes committed in 2018.
- The borough of ‘Allach-Untermenzing’ shows to be the borough of Munich with the lowest number of total crimes.

Methodology
Overview of ‘Total crimes per capita’

- ‘*Total crimes per capita*’, is computed, using the previously discussed ‘*Total crimes*’ data for each borough, this parameter is put into relation to the population size of each borough.
- By dividing the ‘*Total crimes*’ by the ‘*Population*’ and afterwards multiplying it with 100’000 a parameter ‘*Total crimes per capita*’ (here per 100’000 citizens) is computed and shown here to get an overview of this parameter.

Methodology
’Bottom 5’ of ‘Total crimes per capita’

- Analog to the absolute values of ‘*Total crimes*’ the ‘Bottom 5’ boroughs of Munich with the lowest number of ‘*Total crimes per capita*’ are concluded by sorting this parameter in an ascending order.
- The former safest borough of Munich, ‘Allach-Untermenzing’ is still in the ‘Bottom 5’ of safest boroughs, but not the safest borough of them all. This spot takes now the borough of ‘Hadern’.

Methodology
’Bottom 1’ of ‘Total crimes per capita’

- The majority of crimes committed in 2018 in the borough of ‘Hadern’ are of a minor scale.
- Only very few are major crimes (e. g. against life and against sexual self-determination) were committed.

Methodology
Housing prices – neighborhood scale

- Housing prices are taken and imported from one of the leading and most prominent German real estate sites Immobilienscout24.de. As mentioned in the data section, the housing prices are published on a neighborhood level of detail and range from 5335 Euro per square meter on the outskirts of Munich up to 13095 Euro per square meter in the city center.
- Unfortunately, the date for housing prices is given based on neighborhoods of Munich not based on boroughs.

Methodology
Housing prices – borough scale

- A transition from neighborhood level of housing prices to a borough level of housing prices has to be performed.
- This is done by averaging the housing prices in a borough by the values of the neighborhoods of this borough.
- As additional information and side note also the increase of housing price in the neighborhoods/boroughs is provide and given in per cent; but not further taken into account.

Methodology
‘Best combination’ of borough

Since there is no simple minimum for the two-factor problem of lowest total crimes per capita and lowest housing prices, a factor has to be computed. The methodology to get this 'best combination' factor is as follows:

1. Normalize 'Total crimes per capita' by the parameters maximum
2. Normalize 'Price' by the parameters maximum
3. Add both normalized values to an 'Optimum of crimes and prices'
4. Search for the minimum of this newly generated '*Optimum of crimes and prices*'

- Following this methodology the ‘best’ borough in Munich with the lowest ‘*Optimum of crimes and prices*’ is ‘Aubing-Lochhausen-Langwied’.

Borough	Latitude	Longitude	Neighborhoods	Number of neighborhoods	Total major crimes	Crimes against life	Crimes against sexual self-determination	Criminal offence	Simple theft	Aggravated theft	Crimes against property	Total other crimes	Total crimes	Area	Population	Population density	Foreigners	Total crimes per capita	Price	Increase	Total crimes per capita normalized	Price normalized	Optimum of crimes and price
Aubing-Lochhausen-Langwied	48.158437	11.414066	Aubing | Lochhausen | Langwied	3	1620	1	21	330	268	230	216	1604	3224	34.06	47813.0	1404.0	28.4	6742.936022	7061.0	5.566667	0.086955	0.539213	0.626169

Methodology
Venues with FourSquare API

- Using the FourSquare API and the previously added geographical data for each borough 100 venues are crawled, based on the boroughs ‘center-position’ and a radius of 500 meters around this position. The FourSquare API returns a JSON file which is then ported to a Pandas data frame for further data handling and manipulation.
- The venues are than grouped by borough, thereafter an investigation regarding the number of unique venues is carried out with result of 176 unique venues in Munich.
- Using a ‘one-hot’ encoding the venues are processed further, sorted and normalized by the mean value. Afterwards the 10 most common venues of each borough are computed by further sorting and manipulating the underlying Pandas data frame by a user-defined function.

	Borough	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
0	Allach-Untermenzing	Bavarian Restaurant	Gift Shop	Bakery	Italian Restaurant	Trattoria/Osteria	Tennis Court	Fruit & Vegetable Store	Fried Chicken Joint	French Restaurant	Fountain
1	Altstadt-Lehel	Café	Bavarian Restaurant	German Restaurant	Hotel	Plaza	Department Store	Church	Bar	Restaurant	Smoke Shop
2	Au-Haidhausen	Italian Restaurant	French Restaurant	German Restaurant	Concert Hall	Pool	Pizza Place	Gourmet Shop	Beach	Beer Garden	Bistro
3	Aubing-Lochhausen-Langwied	German Restaurant	Light Rail Station	Bistro	Supermarket	Greek Restaurant	Gym / Fitness Center	Bakery	Soccer Field	Hotel	Food & Drink Shop
4	Berg am Laim	Supermarket	Discount Store	Fast Food Restaurant	Metro Station	Café	Gastropub	Drugstore	Hotel	Food Court	Fruit & Vegetable Store

Methodology
Machine learning – k-means-algorithm

- Using the k-means- algorithm the boroughs of Munich are clustered into five different clusters based on their most common venues.
- There are three clusters (cluster 1, 3, and 4) with only one borough in it. The other two clusters contain the majority of boroughs of Munich with 5 (cluster 2) and 17 boroughs (cluster 5) respectively.

Results
‘Best option’ for each cluster (1, 3, 4)

- Using the cluster and the optimum parameter of crimes and housing prices, the 'best option' in each cluster, regarding the total crimes per capita and borough as well as housing prices in this borough can be concluded.
- For the three clusters with each only one borough in it (cluster 1, 3, and 4) nothing changes. They’re the only borough in their cluster and therefore the ‘best-option’ within this cluster.
- For cluster 1 this is the borough of ‘Allach-Untermenzing’, for cluster 3 the borough of ‘Berg am Laim’, and for cluster 4 the borough of ‘Bogenhausen’.

Results
‘Best option’ for each cluster (2)

- In cluster 2, which holds 5 boroughs, the ‘best-option’ is the borough of ‘Feldmoching-Hasenbergl’, because it has the lowest value for ‘Optimum of crimes and prices’.

Borough Nr.	Borough	Latitude	Longitude	Neighborhoods	Number of neighborhoods	Total major crimes	Crimes against life	Crimes against sexual self-determination	Criminal offence	Simple theft	Aggravated theft	Crimes against property	Total other crimes	Total crimes	Area	Population	Population density	Foreigners
24	Feldmoching-Hasenbergl	48.213804	11.541275	Feldmoching | Hasenbergl	2	2035	0	25	351	344	316	304	2019	4054	28.94	61774.0	2135.0	32.4

Results
‘Best option’ for each cluster (5)

- Cluster 5, which holds the majority of boroughs with a number of 17, the same methodology applies. Using it analogous, the borough of ‘Aubing-Lochhausen-Langwied’ proves to be the ‘best-option’ within this cluster.

Borough Nr.	Borough	Latitude	Longitude	Neighborhoods	Number of neighborhoods	Total major crimes	Crimes against life	Crimes against sexual self-determination	Criminal offence	Simple theft	Aggravated theft	Crimes against property	Total other crimes	Total crimes	Area	Population	Population density	Foreigners
22	Aubing-Lochhausen-Langwied	48.158437	11.414066	Aubing | Lochhausen | Langwied	3	1620	1	21	330	268	230	216	1604	3224	34.06	47813.0	1404.0	28.4

Discussion

- Based on key features and certain number of clusters people who want to relocate to Munich can be advised which borough to choose.
- Since the used data changes over time, and the links to the data keeps changing as well, one could invest more work into making this notebook more 'general' for future changes in data, links, cities, more factors to be considered, etc.
- Especially the generalization to other cities is a major task, since this would grand additional business opportunities, and therefore additional customers.
- Furthermore, the integration of more and more detailed key features seems to be a reasonable and valuable task. This would grant the opportunity to support the customers even more and give them a deeper and more customized place to relocate to.

Conclusion

- Based on pre-defined main features and certain number of clusters customers can choose from a borough which fits their needs best regarding crime statistics, housing prices and venues.

(https://upload.wikimedia.org/wikipedia/commons/1/17/Muenchen_Kleines_Stadtwappen.svg

https://upload.wikimedia.org/wikipedia/commons/4/40/Vista_panorámica_desde_Olympiapark%2C_Múnich%2C_Alemania_2012-04-28%2C_DD_03.JPG

References

- [1] Pictures of Munich and general information from Wikipedia.org: https://en.wikipedia.org/wiki/Munich
- [2] Most livable cities from DW.com: https://www.dw.com/en/munich-tops-list-of-worlds-most-livable-cities-berlin-and-hamburg-also-in-top-10/a-44478511
- [3] Most expensive cities from DW.com: https://www.dw.com/en/stuttgart-unseats-munich-as-germanys-most-expensive-city-for-renters/a-51374468
- [4] Crime statistics form Muenchen.de: https://www.muenchen.de/rathaus/dam/jcr:6291ac42-463d-4267-b436-c4b1a3313454/jt190904.pdf
- [5] General population data from Wikipedia.org: https://de.wikipedia.org/wiki/Stadtbezirke_Münchens)
- [6] Housing prices and picture with housing prices in boroughs from Immobilienscout24.de: https://www.immobilienscout24.de/neubau/ratgeber/aktuelle-neubau-themen/neubau-kauf-map-wohnungen-muenchen-2019.html
- [7] Venues from FourSquare.com: https://de.foursquare.com