



Coursera
IBM Data Science Specialization

Final Capstone Project Report

Clustering Neighborhoods in Munich

Submitted by:	Patrick Brus
Submission date:	10.04.2020
Course of studys:	IBM Data Science Professional Certificate

Final Capstone Project Report

1 Introduction

1.1 Bussiness Problem

An individual wants to move to munich or wants to find a new apartment in munich. Before he decides for an apartment, he first wants to get some information about the districts in munich and how similar they are. He wants to know, what kind of venues are in which district and, if he already was or lives in munich, wants to be able to move to a district, where the neighborhood is similar to the one he knows or were he lives in.

1.2 Solution

The KMeans clustering algorithm is used to cluster the neighborhoods of each district in munich according to its venues. This helps the user to identify similar districts and maybe to help him to find a district where he wants to live in.

2 Data

The postal code and district names of all districts of Munich are required to solve the task. The data published at <https://www.muenchen.de/int/en/living/postal-codes.html> is used in order to fetch the necessary data. An overview of the data from the page and therefore all districts and the according postal codes can be seen in table 1. As next step the latitude and longitude of all districts are required. They are then used for the foursquare API in order to get the venue data of all districts. So as first step the table 1 gets split, such that all postal codes have an own entry in the data table. Allach-Untermenzing has five different postal codes and therefore gets five entries in the new data table each with it's own postal code. This step is required, because each district can be very large and can therefore have a lot of different neighborhood clusters within it's district. The new table gets two additional columns. One containing the latitude and one the longitude. Table 2 contains the first five entries of this table. Afterwards this table gets used in order to get the venue data for each district by fetching the data using the foursquare API and the latitude and longitude values. In total 200 unique venue categories are available in Munich. Afterwards this data gets one hot encoded and grouped by each district and then a representative data set for each district is available for clustering.

	District	Postal Code
0	Allach-Untermenzing	80995, 80997, 80999, 81247, 81249
1	Altstadt-Lehel	80331, 80333, 80335, 80336, 80469, 80538, 80539
2	Au-Haidhausen	81541, 81543, 81667, 81669, 81671, 81675, 81677
3	Aubing-Lochhausen-Langwied	81243, 81245, 81249
4	Berg am Laim	81671, 81673, 81735, 81825
5	Bogenhausen	81675, 81677, 81679, 81925, 81927, 81929
6	Feldmoching-Hasenberg	80933, 80935, 80995
7	Hadern	80689, 81375, 81377
8	Laim	80686, 80687, 80689
9	Ludwigsvorstadt-Isarvorstadt	80335, 80336, 80337, 80469
10	Maxvorstadt	80333, 80335, 80539, 80636, 80797, 80798, 8079...
11	Milbertshofen-Am Hart	80807, 80809, 80937, 80939
12	Moosach	80637, 80638, 80992, 80993, 80997
13	Neuhausen-Nymphenburg	80634, 80636, 80637, 80638, 80639
14	Obergiesing	81539, 81541, 81547, 81549
15	Pasing-Obermenzing	80687, 80689, 81241, 81243, 81245, 81247
16	Ramersdorf-Perlach	81539, 81549, 81669, 81671, 81735, 81737, 81739
17	Schwabing-Freimann	80538, 80801, 80802, 80803, 80804, 80805, 8080...
18	Schwabing-West	80796, 80797, 80798, 80799, 80801, 80803, 8080...

FIGURE 1 – Overview of Postal Codes and District Names in Munich

	District	Postal Code	Latitude	Longitude
0	Allach-Untermenzing	80995	48.190034	11.468105
1	Allach-Untermenzing	80997	48.192790	11.484461
2	Allach-Untermenzing	80999	48.195994	11.457013
3	Allach-Untermenzing	81247	48.176884	11.476058
4	Allach-Untermenzing	81249	48.176884	11.476058

FIGURE 2 – Overview of Postal Codes and District Names in Munich