

---

# Neighborhood analysis of the City in Stuttgart, Germany

---

Daniel Fröhlich

**Date:** May 30, 2021

# 1 Introduction

## 1.1 Background

All over the world there are a lot of Cafes with different styles and different drinks and food. And all over the world, people love to go to Cafes for different reasons. E.g. they might meet with someone or they want to be alone to relax or to work on a project. There are different reasons why people love to visit one Cafe while another one is completely empty. However one major key to success of a Cafe is its place. Not only the country and city, but also the neighborhood.

## 1.2 Problem

A friend of mine lives in Stuttgart and he wants to open a Cafe in Stuttgart. So he is facing the question which district and which neighborhood is the most suitable for opening a Cafe.

Since the success of a Cafe depends on the number of visitors, one might want to open a Cafe where there are a lot of people around in the neighborhood.

Regarding the neighborhood important factors could be:

- Are there shopping possibilities in the neighborhood? Some people might want to relax and drink a coffee after shopping.
- Are there restaurants in the neighborhood? After a lunch of dinner, some people might want to have dessert.
- Are there offices in the neighborhood?

This project aims to analyze the different districts and neighborhoods in Stuttgart to find out, which place is the most suitable for opening a Cafe.

## 1.3 Interest

Not only my friend, but anybody who wants to open a Cafe in Stuttgart would be interested in the analysis of the different neighborhoods and districts of Stuttgart. Also tourists and visitors in Stuttgart might be interested in the analysis.

## 2 Data acquisition

### 2.1 Data sources

For the analysis of the different districts and neighborhoods in Stuttgart, Wikipedia<sup>1</sup> will be used to find out the names and postal codes of the districts in Stuttgart.

With the postal codes, the Python library *geopy*<sup>2</sup> will be used to convert the address into coordinates (latitude and longitude).

The Foursquare API<sup>3</sup> will be used to get the venues around the given coordinates.

### 2.2 Use the data to solve the problem

We get the names of the districts in Stuttgart via web scraping wikipedia with the python library *BeautifulSoup*<sup>4</sup>. The result is the following list: Stuttgart-Mitte, Stuttgart-Nord, Stuttgart-Ost, Stuttgart-Süd, Stuttgart-West, Bad-Cannstatt, Birkach, Botnang, Degerloch, Feuerbach, Hedelfingen, Möhringen, Mühlhausen, Münster, Obertürkheim, Plieningen, Sillenbuch, Stammheim, Untertürkheim, Vaihingen, Wangen, Weilimdorf, Zuffenhausen.

We can go through the list of district names and web scrape the specific website of each district by appending the district name to the following string: *https://de.wikipedia.org/wiki/*. In some cases we need to be careful, because the name of the website is not the district

---

1 [https://de.wikipedia.org/wiki/Liste\\_der\\_Stadtbezirke\\_und\\_Stadtteile\\_von\\_Stuttgart](https://de.wikipedia.org/wiki/Liste_der_Stadtbezirke_und_Stadtteile_von_Stuttgart)

2 <https://geopy.readthedocs.io/en/stable>

3 <https://foursquare.com>

4 <https://www.crummy.com/software/BeautifulSoup/bs4/doc>

name only. The reason is that there are other places in Germany called like the district. So we need to change the name in the following way: the text *Birkach* becomes *Birkach\_(Stuttgart)*.

The following list of postal codes is the result of web scraping the district names:

Stuttgart-Mitte: 70173, 70174, 70178, 70182

Stuttgart-Nord: 70174, 70191, 70192

Stuttgart-Ost: 70148-70190

Stuttgart-Süd: 70178, 70180, 70184, 70199, 70569

Stuttgart-West: 70176, 70178, 70193, 70197

Bad\_Cannstatt: 70331-70378

Birkach\_(Stuttgart): 70599

Botnang: 70195

Degerloch: 70597

Feuerbach\_(Stuttgart): 70469, 70499

Hedelfingen: 70329

Möhringen\_(Stuttgart): 70567, 70565

Mühlhausen\_(Stuttgart): 70378, 70437

Münster\_(Stuttgart): 70376

Obertürkheim: 70329

Plieningen: 70599

Sillenbuch: 70619

Stammheim\_(Stuttgart): 70439

Untertürkheim: 70327

Vaihingen\_(Stuttgart): 70563, 70565, 70569

Wangen\_(Stuttgart): 70327

Weilimdorf: 70499

Zuffenhausen: 70435, 70437, 70439

In case of Stuttgart-Ost and Bad\_Cannstatt there are a lot of postal codes, which can be identified by the hyphen in the returned postal codes: 70148-70190 and 70331-70378. In the following, we will concentrate on the 5 postal codes with the biggest population. For Stuttgart-Ost those are 70184, 70186, 70188, 70190 and 70327. For Bad\_Cannstatt: 70332, 70372, 70374, 70376 and 70378.

To avoid duplicates in the resulting *pandas* Dataframe, we will append the postal code to the district name, in cases of districts with more than one postal code. One example would

be Stuttgart-Mitte. Here we have the postal codes 70173, 70174, 70178 and 70182. So the resulting names would be Stuttgart-Mitte\_70173, Stuttgart-Mitte\_70174, Stuttgart-Mitte\_70178 and Stuttgart-Mitte\_70182.

Some districts share the same postal code in Stuttgart. We will chain the district names and the shared postal code, e.g. Stuttgart-Ost-Untertürkheim-Wangen\_70327 will be the name of the postal code 70327.

The first 5 rows of resulting dataframe look as follows:

	District	PostalCode
0	Stuttgart-Mitte_70173	70173
1	Stuttgart-Mitte_70182	70182
2	Stuttgart-Nord_70191	70191
3	Stuttgart-Nord_70192	70192
4	Stuttgart-Ost_70186	70186

The next step is to add the coordinates (latitude and longitude) to the postal codes. We will do this by using the python library *geopy*. We will loop through our dataframe and send a geocode request using the Nominatim package for each row (e.g. *70195, Stuttgart, Germany*) and store the result in a separate dataframe. The first 5 rows of the coordinates dataframe:

	PostalCode	Latitude	Longitude
0	70173	48.777808	9.178420
1	70182	48.774376	9.184793
2	70191	48.798102	9.182559
3	70192	48.795235	9.165006
4	70186	48.775952	9.207278

Next we will join the two dataframes to get the final dataframe which will be used in the analysis of the Stuttgart districts. Here is the head of the final dataframe:

	District	PostalCode	Latitude	Longitude
0	Stuttgart-Mitte_70173	70173	48.777808	9.178420
1	Stuttgart-Mitte_70182	70182	48.774376	9.184793
2	Stuttgart-Nord_70191	70191	48.798102	9.182559
3	Stuttgart-Nord_70192	70192	48.795235	9.165006
4	Stuttgart-Ost_70186	70186	48.775952	9.207278

With the given postal codes, we will get all the venues in the radius of 500 meters around the coordinates and try to cluster the districts of Stuttgart into clusters with different venues. Then we will analyze those clusters and see if the problem described in chapter *1.2 Problem* can be solved with this clustering.