Applied Data Science Capstone Project:
Battle of the Neighbourhoods

by

Bradley Fourie

Technical Report

March 2021

Contents

Chapter 1

Introduction

1.1 Background

When visiting a new destination for the first time, tourists are faced with a big decision, that is the task of finding accommodation. Oftentimes these tourists fall for tourist traps and are charged ridiculous prices for their accommodation. Conversely, these tourists might not do adequate research and book accommodation that is far away from the points of interest that they are yet to discover. This decreases the quality of their holiday experience, and increases the cost of travelling, which could otherwise have been spent on fun activities.

1.2 Problem Statement

Jonathan is a student who is planning on doing a semester abroad at Reutlingen University in Germany. During his stay abroad he is planning on visiting Stuttgart, however, he is unsure of where he should be looking for accommodation. His main interest in the city lies with its flourishing nightlife, and he would like to find accommodation that is central to the neighbourhoods with these types of venues. This project aims to use data science and visualisation techniques to find the best nightlife related neighbourhoods, offering accommodation, that are optimal for tourists interested in the nightlife of Stuttgart.

Chapter 2

Data Sources

2.1 Moving to Germany Website

As no datasets that store the suburbs/neighbourhoods in Stuttgart could be found online an alternate approach had to be taken. Here, the Moving to Germany website was used to provide information about the names of all of the suburbs in this city. Using the `BeautifulSoup` library the list containing the names of these suburbs was scraped from the website. Using the `geopy` library the coordinates of each of these suburbs was also found.

2.2 Foursquare API

Furthermore, the use of the `Foursquare` API played an important role in finding all of the nightlife venues, as well as the cheap hotels in Stuttgart. This geo-spatial data that we are interested in includes the name of the neighbourhood, the latitude and longitude of the nightlife venue, and the type of venue. We also repeat this process of data collection in finding the most suitable hotel.

This data is obtained from the *Foursquare* API in the following manner:

- Find all of the most popular nightlife venues listed for each neighbourhood.
- Cluster the neighbourhoods based on the similarity of their nightlife venues.
- Find all of the cheap hotels in each neighbourhood
- Plot the location pins of the cheap hotels on top of the map clusters
- Identify which cluster type suits the user's needs.
- Select the hotel that is most central to these related clusters.

Chapter 3

Methodology

This section provides a brief overview of the techniques used in this report, a more detailed analysis and examples of these are discussed in the following two sections.

3.1 Exploratory Data Analysis

The first step in this methodology includes performing an exploratory data analysis on the results of the `Foursquare` data. This is required to determine how the data should be manipulated, and which features have missing values. For this exploratory data analysis all data are reported on using pandas dataframes to easily view the datatypes and values of the features. Here, the data about the nightlife is analysed by using histograms to obtain a value count for each type of venue.

3.2 Machine Learning

The second step in this methodology includes using unsupervised machine learning techniques. Here, the k-means algorithm is used to cluster the neighbourhoods based on their similarity with regards to the types of nightlife venues that they contain. Using this similarity measure the different parts of the city (based on the type of nightlife venues) can be illustrated, and the different experiences a tourist can expect can be quantified.

Chapter 4

Results

4.1 Neighbourhoods and Nightlife Venues

The first step in this process includes scraping the data from the Moving to Germany website. Here, it was found that there are a total of 23 unique suburbs using the `shape` function of the `pandas` library.

Next we proceed to obtain the coordinates of these neighbourhoods using the `geopy` library. These coordinates will be useful later when we will visualise the clusters for each neighbourhood. Now we iterate through the list of this of neighbourhood names and we use the `Foursquare` API to obtain the venues near these neighbourhoods. Here, we limit our API call radius to 1000, and we provide '*nightlife*' as `query`, which is an additional parameter. The result of this query returns a table of the form shown in Figure 4.1, where only the first five rows of the table are shown.

Next, we proceed to investigate the total number of venues in each suburb, as well as look at how many of each type of venue exists in total. These descriptive statistical results are shown in Figure 4.2a as well as Figure 4.2b. After using the `shape` function of the `pandas` library we find that there are a total of 212 venues, belonging to 39 unique categories, in the city of Stuttgart. To visualise how these

	Suburb	Suburb Latitude	Suburb Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Stuttgart-Mitte	48.7759	9.1798	Bix Jazzclub	48.773178	9.179495	Jazz Club
1	Stuttgart-Mitte	48.7759	9.1798	Eduard's	48.775537	9.179935	Cocktail Bar
2	Stuttgart-Mitte	48.7759	9.1798	TATTI	48.774353	9.177209	Café
3	Stuttgart-Mitte	48.7759	9.1798	Mata Hari	48.773798	9.177704	Bar
4	Stuttgart-Mitte	48.7759	9.1798	Brenner Drinks, Food etc.	48.774059	9.182335	Bar

Figure 4.1: result of `df.head()` function

```
Stuttgart-Mitte     87
Münster             38
Bad Cannstatt       12
Stuttgart-West      10
Möhringen            7
Feuerbach            6
Stuttgart-Ost        5
Stuttgart-Nord       5
Mühlhausen           5
Weilimdorf           5
Zuffenhausen         4
Untertürkheim        4
Stuttgart-Süd        3
Vaihingen            3
Stammheim            3
Degerloch            3
Botnang              3
Obertürkheim         3
Hedelfingen          2
Sillenbuch           2
Birkach              1
Wangen               1
Name: Suburb, dtype: int64
```

(a) Number of venues in each suburb of Stuttgart

```
Bar                              50
Nightclub                        20
Pub                              19
Gastropub                        16
Lounge                           14
Cocktail Bar                     10
Beer Garden                       9
German Restaurant                 8
Café                              8
Hookah Bar                        7
Sports Bar                        6
Burger Joint                      4
Restaurant                        3
Hotel                             3
Hotel Bar                         3
Brewery                           3
Wine Bar                          3
Gay Bar                           2
Jazz Club                         2
Wine Shop                         2
Beach Bar                         2
Spanish Restaurant                1
Speakeasy                         1
Sushi Restaurant                  1
Mexican Restaurant                1
Asian Restaurant                  1
Whisky Bar                        1
Bowling Alley                     1
Pizza Place                       1
Bakery                            1
Deli / Bodega                     1
Event Space                       1
Modern European Restaurant        1
Music Venue                       1
Boat or Ferry                     1
Breakfast Spot                    1
Bistro                            1
Piano Bar                         1
Rock Club                         1
Name: Venue Category, dtype: int64
```

(b) Total number of nightlife venues

Figure 4.2: Totals of the number of venues per suburb as well as the number of each type of venue category

are distributed we plot the result on a bar chart, as shown in Figure 4.3.

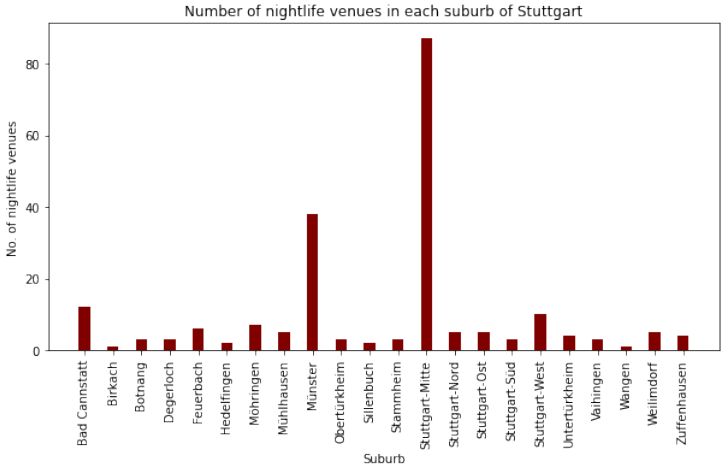
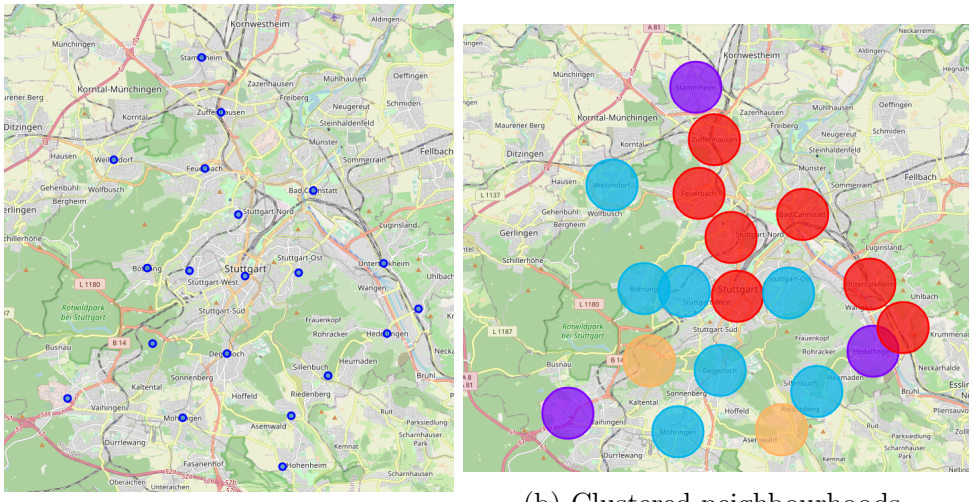


Figure 4.3: Bar chart depicting how the nightlife venues are distributed in Stuttgart



(a) Unclustered neighbourhoods.

(b) Clustered neighbourhoods.

Figure 4.4: Visual representation of Stuttgart

4.2 Clustering Neighbourhoods

Next, we proceed to cluster the neighbourhoods in Stuttgart based on the types of nightlife venues that they contain. This will allow us to get a 'feel' of the city. A visual representation of Stuttgart is shown in Figure 4.4a, where the `Folium` library was used to generate the map. We use the k-means clustering algorithm with k equals to 5 to cluster the neighbourhoods in the city, as shown in Figure 4.4b.

From Figure 4.4b we can see that the northern suburbs are very similar to Stuttgart-Mitte, and the southern suburbs are more similar to the suburb of Stuttgart-Ost. Now, to decide which cluster of suburbs best suit Jonathan's interests we have to investigate the type of nightlife venue in each suburb. Figure 4.5-4.9 displays this top ten category of venues in each suburb.

	Suburb	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	Stuttgart-Mitte	Nightclub	Cocktail Bar	Lounge	German Restaurant	Café	Wine Bar	Jazz Club	Burger Joint	Sports Bar
1	Stuttgart-Nord	Burger Joint	Hotel Bar	Beer Garden	Nightclub	Gastropub	Event Space	Deli / Bodega	Cocktail Bar	Café
6	Bad Cannstatt	Café	Beer Garden	Gastropub	Sports Bar	Hotel Bar	Boat or Ferry	Lounge	Wine Shop	Event Space
10	Untertürkheim	Bar	Gastropub	Hookah Bar	Burger Joint	Gay Bar	Event Space	Deli / Bodega	Cocktail Bar	Café
12	Zuffenhausen	Hookah Bar	Gastropub	Nightclub	Gay Bar	Event Space	Deli / Bodega	Cocktail Bar	Café	Burger Joint
13	Feuerbach	Bar	Gastropub	Restaurant	Pub	German Restaurant	Burger Joint	Gay Bar	Event Space	Deli / Bodega
17	Mühlhausen	Bar	Sports Bar	Beer Garden	Cocktail Bar	Burger Joint	Gay Bar	Gastropub	Event Space	Deli / Bodega
19	Obertürkheim	Lounge	Pub	Brewery	Gay Bar	Gastropub	Event Space	Deli / Bodega	Cocktail Bar	Café

Figure 4.5: Cluster 1

	Suburb	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
9	Stammheim	Hookah Bar	Wine Shop	Burger Joint	Gay Bar	Gastropub	Event Space	Deli / Bodega	Cocktail Bar	Café
11	Vaihingen	Deli / Bodega	Wine Shop	Brewery	Gay Bar	Gastropub	Event Space	Cocktail Bar	Café	Burger Joint
16	Hedelfingen	Lounge	Wine Shop	Burger Joint	Gay Bar	Gastropub	Event Space	Deli / Bodega	Cocktail Bar	Café

Figure 4.6: Cluster 2

	Suburb	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
2	Stuttgart-Ost	Beer Garden	Lounge	Pub	Wine Shop	Burger Joint	Gastropub	Event Space	Deli / Bodega	Cocktail Bar
4	Stuttgart-West	Pub	Gastropub	Sports Bar	German Restaurant	Wine Shop	Brewery	Event Space	Deli / Bodega	Cocktail Bar
6	Degerloch	Bar	Wine Shop	Brewery	Gay Bar	Gastropub	Event Space	Deli / Bodega	Cocktail Bar	Café
7	Möhringen	Gastropub	Beach Bar	Pub	Wine Shop	Burger Joint	Gay Bar	Event Space	Deli / Bodega	Cocktail Bar
15	Botnang	Gastropub	Pub	Wine Shop	Brewery	Gay Bar	Event Space	Deli / Bodega	Cocktail Bar	Café
18	Münster	Pub	Nightclub	Café	Gastropub	Lounge	Sports Bar	Spanish Restaurant	German Restaurant	Whisky Bar
20	Sillenbuch	Pub	Wine Shop	Brewery	Gay Bar	Gastropub	Event Space	Deli / Bodega	Cocktail Bar	Café
22	Weilimdorf	Bar	Beer Garden	Brewery	Wine Shop	Burger Joint	Gay Bar	Gastropub	Event Space	Deli / Bodega

Figure 4.7: Cluster 3

	Suburb	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
21	Wangen	Wine Shop	Hookah Bar	Gay Bar	Gastropub	Event Space	Deli / Bodega	Cocktail Bar	Café	Burger Joint

Figure 4.8: Cluster 4

	Suburb	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
3	Stuttgart-Süd	Lounge	Wine Shop	Hookah Bar	Gay Bar	Event Space	Deli / Bodega	Cocktail Bar	Café	Burger Joint
14	Birkach	Wine Shop	Brewery	Gay Bar	Event Space	Deli / Bodega	Cocktail Bar	Café	Burger Joint	Breakfast Spot

Figure 4.9: Cluster 5

From Figure **??** we can see that this cluster contains Stuttgart-Mitte, which is in the centre of the city. Here, we can expect to find the most nightclubs, cocktail bars as well lounges in the city. These are the types of nightlife venues that we are interested in. This is also central to all of the other suburbs, which is desirable as a tourist.

4.3 Displaying Cheap Accommodation

Finally, we proceed to look for cheap accommodation in the city of Stuttgart. Now we again iterate through the list of all neighbourhood names and we use the `Foursquare` API to obtain the cheap accommodation near these neighbourhoods. Here, we limit our API call radius to 1000, and we provide '*cheap hotels*' as `query`, which is an additional parameter. The result of this query, overlayed over the clustered neighbourhoods are shown in Figure 4.10.

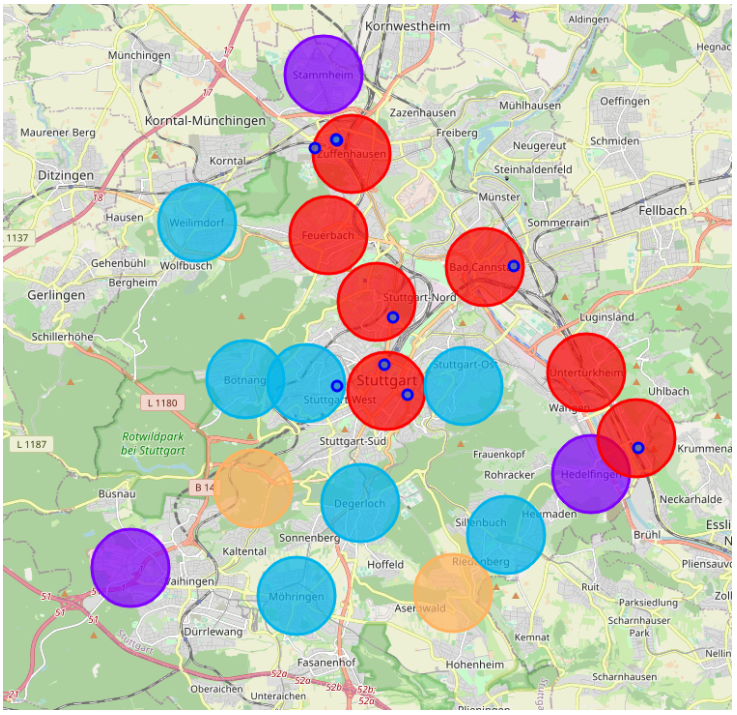


Figure 4.10: Cluster 5

Chapter 5

Discussion

5.1 Observations

Some of the most important observations that can be made about the nightlife venues in Stuttgart is that the majority of all nightlife venues can be found in the neighbourhood of Stuttgart-Mitte, which immediately raises our attention to the fact that this might be the best suited neighbourhood for tourists.

From our clustered map data we see that cluster 1 define the neighbourhoods with the most nightclubs, cocktail bars as well as nightlife lounges. This cluster provides a different experience than cluster 3 which contains neighbourhoods with pubs, bars and sports bars. Both of these clusters are in contrast to cluster 2, 4, and 5 which contain nightlife venues like wine shops and burger joints.

Next we proceed to plot the location of cheap hotels in the city of Stuttgart, these cheap hotels are mainly located in the far northern suburbs, or in the neighbourhood of Stuttgart-Mitte.

5.2 Recommendations

From our machine learning analysis of the neighbourhoods of Stuttgart we can see that the most suitable neighbourhood for a student who is interested in nightlife venues is Stuttgart-Mitte. This neighbourhood contains the most nightclubs, and has in total the most nightlife venues when compared to the other neighbourhoods.

From our analysis of the cheap accomodation in Stuttgart we see that once again Stuttgart-Mitte is a good option. To the north of Stuttgart-Mitte lies the neighbourhoods which are most similar, and thus accommodation near this area will be

the best option. Due to this the recommended accommodation would be at the Motel-One in Stuttgart-Mitte, as shown in Figure 5.1.

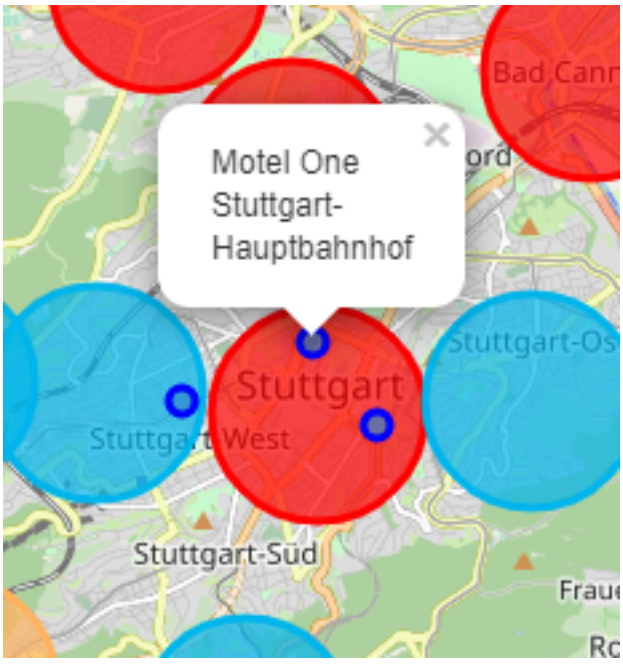


Figure 5.1: The location of the best cheap accommodation for tourists looking to enjoy the nightlife of Stuttgart

Chapter 6

Conclusion

In this report our goal was to find the neighbourhoods in the city of Stuttgart that best suit tourists looking to enjoy the city's nightlife. This was accomplished by first scraping the data about the neighbourhoods of Stuttgart from Moving to Germany. Using this data alongside with the `geopy` library and the `Foursquare` API all nightlife venues in these neighbourhoods were found. We completed an exploratory data analysis on this data using `pandas` and `matplotlib` and chose our features for our k-means clustering algorithm that was used to segment the neighbourhoods.

Using this clustered data we are able to get a 'feel' of the city, and which neighbourhoods are most similar with regards to nightlife venues. From our data it was found the city centre, as well as the northern-suburbs are most suitable in this regard. Finally, we proceeded to find the best cheap accommodation for our target market. Here, the Motel-One to the north of Stuttgart-Mitte was chosen as this is the closest to all of the nightlife venues, as well as to the other suburbs most similar to Stuttgart-Mitte.