Recommeding Borough for moving people

Seungheui Lee

January 18, 2021

1 Introduction

1.1 Background

People move to new places for a variety of reasons, including work, study, and marriage. When moving to a new place, people consider not only the price of the house but also the various surroundings. Different people will have different considerations. For example, someone would want a park close and someone would prefer a market close. So how can we find a place according to our preferences? In this report, we try to solve this problem with a data science approach.

1.2 Problem

In this report, we will recommend suitable areas for the Lee's family who are moving to Hannover. Lee wants to have a supermarket near his place and he wish there is a park nearby, where he could take a walk with his family, and he likes restaurants where he can eat out, and a cafe with a nice atmosphere. Finally, he is thinking of a monthly rent of 10€/m^2.

1.3 Interest

This report is an analysis of the boroughs in Hannover customized to the needs of Lee. However, the approach is applicable to everyone with any specific needs. Based on venue information obtained from Foursquare, this recommendation approach could actually be used in real estate to recommend the desired place to customers.

2 Data

2.1 Description of the Data

We will use the following data to find the most promising borough for Lee.

1. Average price per m^2 of the apartments in Hannover : This information was collected on web page [2]. Some modifications need to be performed to transform the data frame in a format that can be used for further analysis.

2. Information about the venues in all boroughs of Hannover : This information was collected on web page [3]. The Geocoder Python package [1] is used to receive the latitude and longitude coordinate for all of the boroughs. The boroughs and their corresponding latitude and longitude are used as input for FourSquare to source information about the boroughs.

2.2 How data will be used to solve the problem

We will start with an exploratory data analysis, where we intend to understand the underlying data. The describe method provides valuable insights for the "average price per m^2 of the apartments in Hannover" investigations.

To get a first impression about the distribution of venues in Hannover, they are visualized using Folium map. The chosen color code will give immediate yet superficial insight, how Lee's favourite venues are distributed across Hannover.

One hot encoding and k-means will narrow the list of the most promising boroughs to three. Combining these with the pricing analysis we could recommend the best borough match for Lee.

2.3 Data Preparation

First, we created an Excel table of monthly rent prices and postal codes obtained from web page [2] and web page [3] for each boroughs of Hannover. The result of reading the excel file and creating a data frame is shown in figure 1. Looking at the column *Code*, there are rows with multiple values. We separated these values into each rows. The result is shown in figure 2.

	Borough	Price	Code
0	Ahlem	6,40	30453
1	Anderten	6,85	30559
2	Badenstedt	6,28	30453, 30455
3	Bemerode	6,72	30521, 30539, 30559
4	Bornum	5,80	30453

Figure 1: Data frame of prices and postal codes for each boroghs in Hannover

Next, we added the longitude and latitude information of all boroughs to the data frame using the Python Geocoder package [1].

Next, we collected the venues within the boroughs using the Foursquare API. We chose a radius of 500m and 50 as limit of number of venues returned. The result data frame is shown in figure 3.

	Borough	Price	PostalCode
0	Ahlem	6.40	30453
1	Anderten	6.85	30559
2	Badenstedt	6.28	30453
3	Badenstedt	6.28	30455
4	Bemerode	6.72	30521
5	Bemerode	6.72	30539
6	Bemerode	6.72	30559
7	Bornum	5.80	30453
8	Bothfeld	7.04	30655
9	Bothfeld	7.04	30657

Figure 2: Data frame of prices and postal codes for each boroghs in Hannover after separating rows

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ahlem	52.382135	9.663791	Lukullus	52.383930	9.664961	Greek Restaurant
1	Ahlem	52.382135	9.663791	Bäcker Göing	52.384274	9.664043	Bakery
2	Ahlem	52.382135	9.663791	Schüler Bäckerei	52.379869	9.663824	Bakery
3	Anderten	52.363262	9.855728	Palladion	52.360074	9.855496	Greek Restaurant
4	Anderten	52.363262	9.855728	Cortina Da Stella	52.363683	9.853782	Italian Restaurant

Figure 3: Data frame of venues in Hannover

3 Methodology

3.1 Exploratory Data Analysis

We started with an exploratory data analysis of the average price per square meter of apartments in all boroughs of Hannover which is shown in figure 4. Figure 4 shows the frequency of prices in Hannover. It is obvious that Lees target price is the highest compared to all other boroughs. Figure 5 shows the distribution of the prices across Hannover and the corresponding tables are specified in figure 6.

In a further exploitative analysis the distribution of venues in Hannover was investigated. The result is shown in figure 7. Parks, Cafes, Restaurants, Markets, and Others are marked in green, orange, yellow, blue and grey respectively.

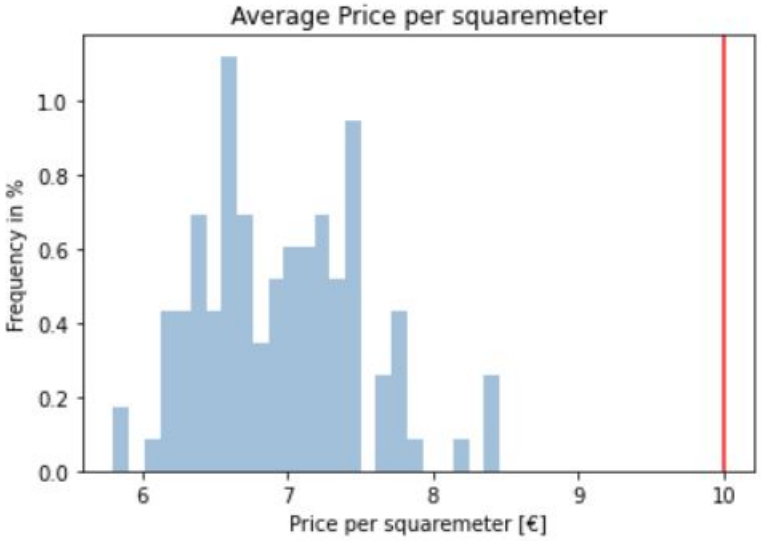

Figure 4: Frequency of price per m^2 in Hannover

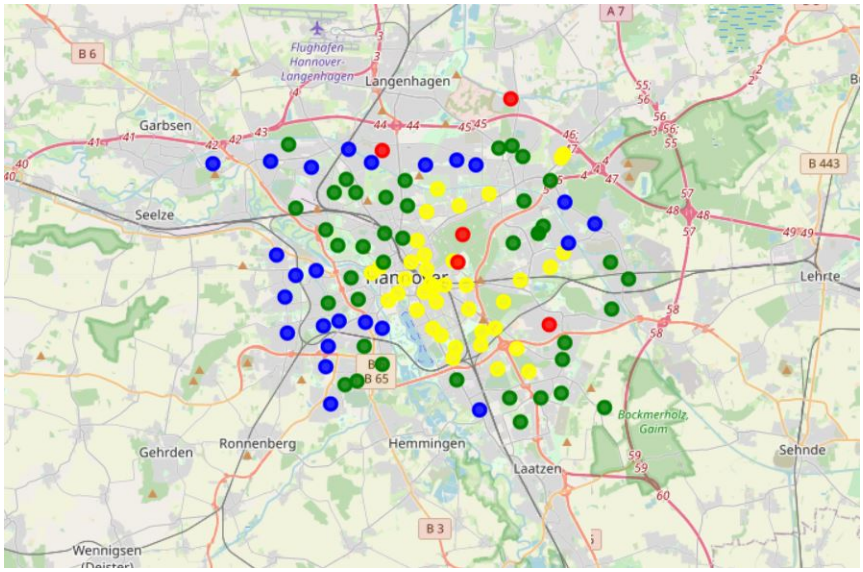

Figure 5: Price distribution across Hannover

3.2 One hot encoding

3.2.1 Lees favorite venues

Lee prefers to have parks, coffee shops, restaurants and markets nearby. In this analysis, we used one hot encoding to identify how many of the favourite venues

	Price
(6.465, 7.13]	41
(7.13, 7.795]	39
(5.797, 6.465]	24
(7.795, 8.46]	5

Figure 6: Price range and number of boroughs within the specific price range

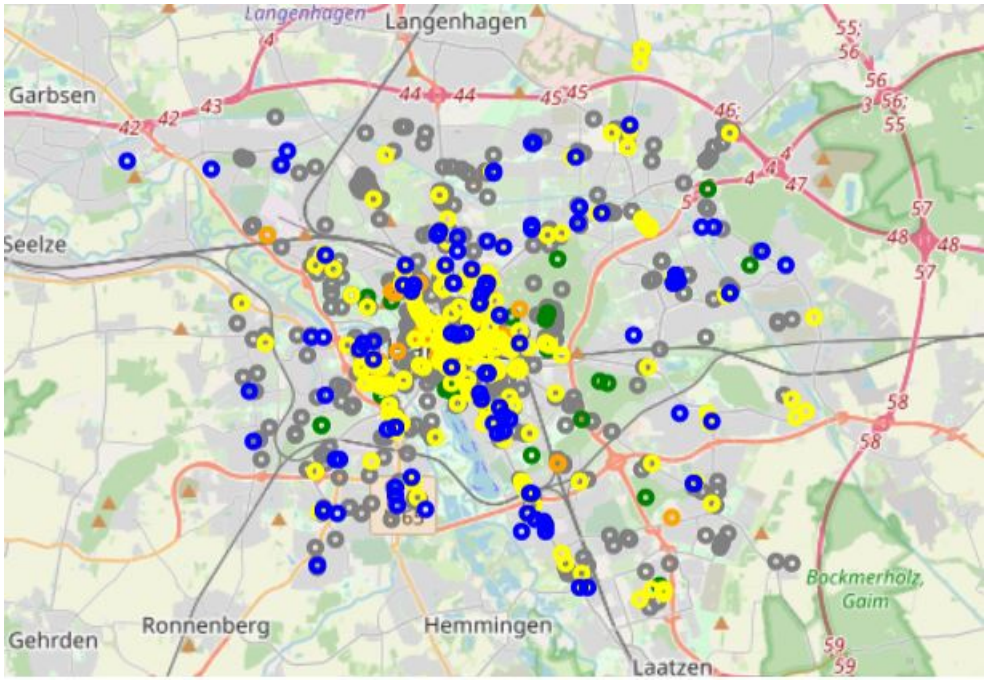

Figure 7: Venues across boroughs in Hannover

are located in the different neighborhoods. Looking at the data frame, that contains only Lees favourite venues we had several columns with sub-categories. As we were just interested in the category but not the specific subcategory, we combined all sub-categories in one column. We define a "Favorite score" as the normalized sum of favourite venues in the neighborhood:

$$\text{Favorite score } = \frac{\sum \text{Venues in this borough}}{\sum \text{All venues}}$$

The result is shown in figure 8. But looking at figure 8, there are several rows that have zero favorite venue. Therefore we dropped that rows. The result is

shown in figure 9.

	Neighborhood	Parks	Cafés	Markets	Restaurants	FavoriteScore
0	Mitte	2	20	7	61	0.196507
1	List	0	8	14	24	0.100437
2	Südstadt	3	3	8	29	0.093886
3	Oststadt	0	7	3	30	0.087336
4	Nordstadt	2	8	4	17	0.067686
5	Groß-Buchholz	1	0	9	11	0.045852
6	Linden-Nord	0	4	2	11	0.037118
7	Vahrenwald	0	0	7	9	0.034934
8	Linden-Mitte	1	2	1	10	0.030568
9	Bult	1	0	1	10	0.026201

Figure 8: One hot encoding and favourite score of all boroughs in Hannover

	Neighborhood	Parks	Cafés	Markets	Restaurants	FavoriteScore
27	Mitte	2	20	7	61	0.196507
37	Südstadt	3	3	8	29	0.093886
30	Nordstadt	2	8	4	17	0.067686
20	Linden-Mitte	1	2	1	10	0.030568

Figure 9: One hot encoding and favourite score of all boroughs in Hannover, that contain at least one of the venues of Lees interest

3.2.2 All venues in Hannover except Lees favorite venues

In this analysis we identified the ten most common venues in all boroughs of Hannover except the venues of Lees interest. Lee could use this list to explore the boroughs further before he makes a final decision. The result is shown in figure .

	Neighborhood	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	Ahlem	Bakery	Zoo	Friterie	Fountain	Food Court	Food & Drink Shop	Flower Shop	Field	Exhibit	Electronics Store
1	Anderten	Pool	Donut Shop	Fountain	Food Court	Food & Drink Shop	Flower Shop	Field	Exhibit	Electronics Store	Discount Store
2	Badenstedt	Tram Station	Bakery	Gastropub	Sports Club	Electronics Store	Fountain	Food Court	Food & Drink Shop	Flower Shop	Field
3	Bemerode	Tram Station	Insurance Office	Beer Store	Electronics Store	Shipping Store	IT Services	Metro Station	Pizza Place	Wine Shop	Deli / Bodega
4	Bornum	Big Box Store	Hotel	Arts & Crafts Store	Intersection	Zoo	Exhibit	Friterie	Fountain	Food Court	Food & Drink Shop
5	Bothfeld	Soccer Field	Discount Store	Tram Station	Steakhouse	Ice Cream Shop	Bus Stop	Hotel	Deli / Bodega	Creperie	Department Store
6	Brink-Hafen	Construction & Landscaping	Business Service	Hardware Store	Electronics Store	Friterie	Fountain	Food Court	Food & Drink Shop	Flower Shop	Field
7	Bult	Tram Station	Bakery	Sports Club	Miscellaneous Shop	Hotel	Concert Hall	Soccer Field	Gym / Fitness Center	Gas Station	Spa
8	Burg	Bus Stop	Zoo	Electronics Store	Friterie	Fountain	Food Court	Food & Drink Shop	Flower Shop	Field	Exhibit
9	Calenberger Neustadt	Plaza	Beer Garden	Metro Station	Pizza Place	Rock Club	Hotel	Tram Station	Bike Shop	Fruit & Vegetable Store	Pub

Figure 10: Except of most common venues in the boroughs of Hannover

4 Results

Different analysis have been performed to find the most suitable new neighborhood for Lee. Considering the price per m^2 Lee can move anywhere, as all of the boroughs have a lower price than target. The venues within the different neighborhoods have been divided in two different data frames: (a) Lees favourite venues and (b) All venues in Hannover except of Lees favourite venues. For the analysis where just Lees favourite venues have been considered (a), we defined a favourite score, taking into account, how often Lees favourite venues occurred in the specific neighborhoods. This analysis narrowed down the results to the following four neighborhoods:

- Mitte
- Südstadt
- Nordstadt
- Linden-Mitte

For the analysis, where all venues in Hannover except of Lees favourite venues have been considered (b), we identified the 10 most common venues within all of the boroughs which could be helpful for Lee, to further explore other neighborhoods.

5 Discussion

In this report, the analysis was performed by postal code, but if the analysis was performed by street, more detailed recommendations would be possible. Elaborate recommendations will be possible if a new measure is created by additionally considering the weight of the distance to the favorite venue including the location of available apartment.

6 Conclusion

The purpose of this project was to search new borough that has Lees favorite venues. We performed different analysis. Considering the average prices per m^2 for apartments in Hannover showed, that no matter were Lee is moving he will most likely find a cheaper apartment than target price. By just taking into account the venues that are important for Lee, we could find four boroughs. The k-means provided an insight into similar neighborhoods. After combining these results, we identified one single borough, that is most likely the best choice for Lees new area to live: Mitte. Of course, the price and points of his personal interest are not the only criteria how Lee should make his final decision, as additional factors like availability of apartments, noise, proximity to friends also matter. However, it serves as an orientation and good neighborhood to start searching.

References

[1] Geocoder. URL `https://geocoder.readthedocs.io/index.html`.

[2] Average price of the apartments in hannover. URL `https://mietspiegeltabelle.de/mietspiegel-hannover`.

[3] Bouroughs in hannover. URL `https://www.suche-postleitzahl.org/hannover-plz-30159-30669.45d1`