Urban audit for students

Capstone project of Anika Cornelius – 6/5/2019

1. Business Problem

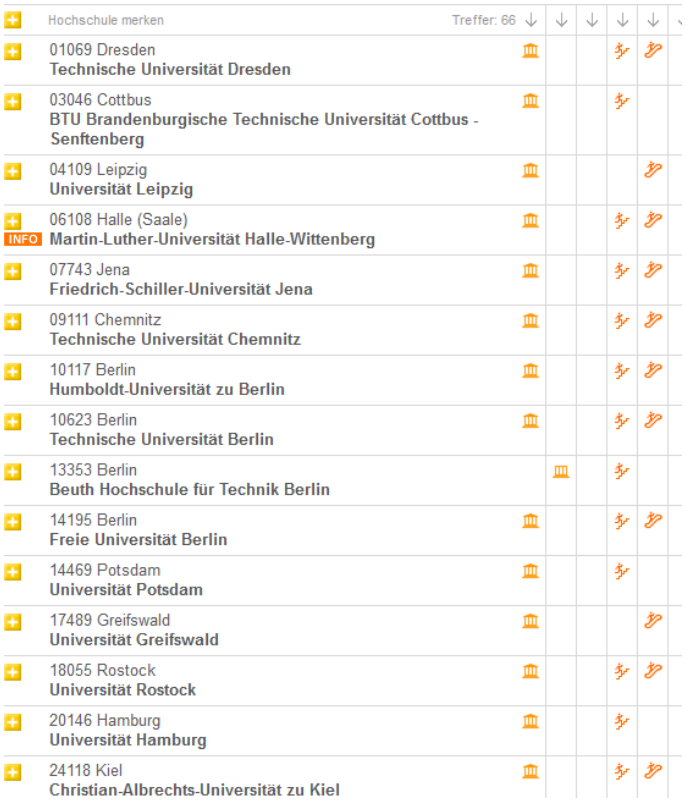
If you study in Germany and you would like to switch university but you enjoy the venues and leisure time possibilities in your actual city, you would love to find similar cities with the same faculty. I'd like to give a hint to students which cities in Germany could also be interesting for them. Because of my background, I'll focus on students which are enrolled to mathematical courses.

Can we find clusters of comparable university cities with similar venues?
Which city/cities could we recommend to a student who studies in Bremen?

2. Data

At first, I need an overview of German cities with mathematical faculties. I found it on the German website: www.studieren.de (studieren = to study). I'll get the necessary information with the beautiful soup package.

The page looks like this:



Hochschule merken	Treffer: 66
01069 Dresden **Technische Universität Dresden**	
03046 Cottbus **BTU Brandenburgische Technische Universität Cottbus - Senftenberg**	
04109 Leipzig **Universität Leipzig**	
INFO 06108 Halle (Saale) **Martin-Luther-Universität Halle-Wittenberg**	
07743 Jena **Friedrich-Schiller-Universität Jena**	
09111 Chemnitz **Technische Universität Chemnitz**	
10117 Berlin **Humboldt-Universität zu Berlin**	
10623 Berlin **Technische Universität Berlin**	
13353 Berlin **Beuth Hochschule für Technik Berlin**	
14195 Berlin **Freie Universität Berlin**	
14469 Potsdam **Universität Potsdam**	
17489 Greifswald **Universität Greifswald**	
18055 Rostock **Universität Rostock**	
20146 Hamburg **Universität Hamburg**	
24118 Kiel **Christian-Albrechts-Universität zu Kiel**	

Then I'll use the geopy.geocoders package from python to convert an address into latitude and longitude values.

For example:

```
address = 'Dresden'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of {} are {}, {}.'.format(address, latitude, longitude))
```

```
The geograpical coordinate of Dresden are 51.0493286, 13.7381437.
```

With these information I'll extract the top 100 venues of each city with the Foursquare API:

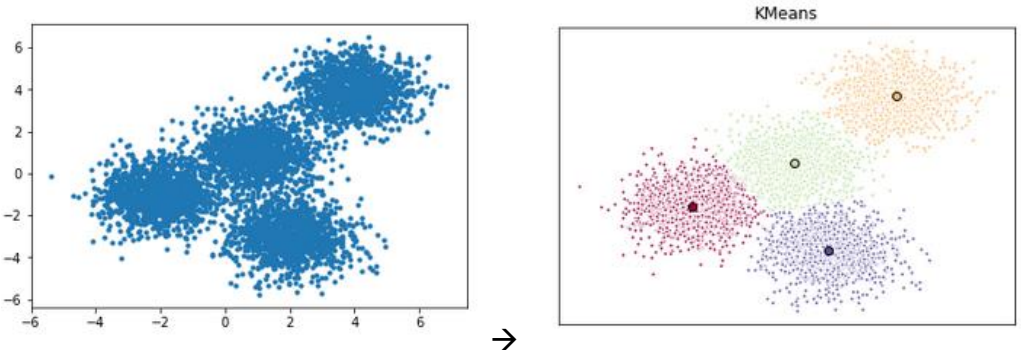
	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state	id
0	Altmarkt	Plaza	Altmarkt	DE	Dresden	Deutschland	NaN	34	[Altmarkt, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.04959612524969...	51.049596	13.737888	NaN	01067	Sachsen	4bb77c8b314e95212d2b479d
1	Watzkes Wurstküche	German Restaurant	Dr.-Külz-Ring 11-13	DE	Dresden	Deutschland	NaN	265	[Dr.-Külz-Ring 11-13, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.04756312824165...	51.047563	13.735591	NaN	01067	Sachsen	552a9a48498e37eed8b52891
2	Neumarkt	Plaza	Neumarkt	DE	Dresden	Deutschland	NaN	298	[Neumarkt, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.05142549103657...	51.051425	13.740793	NaN	01067	Sachsen	4c00d1bedf6c0f47fefa8b22
3	Apple Altmarkt-Galerie	Electronics Store	Altmarkt 25	DE	Dresden	Deutschland	NaN	193	[Altmarkt 25, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.0499639, 'lng'...	51.049964	13.735567	NaN	01067	Sachsen	4d3b2c1d97e0a1cd2288753e
4	REWE	Supermarket	Webergasse 1	DE	Dresden	Deutschland	NaN	204	[Webergasse 1, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.04823635936872...	51.048236	13.735800	NaN	01067	Sachsen	4c6a89a63bad2d7ffe98b2ee
5	Camondas Schokoladen-Kontor & Schokoladenmuseum	Chocolate Shop	Schloßstr. 22	DE	Dresden	Deutschland	NaN	358	[Schloßstr. 22, Dresden, Deutschland]	[{'label': 'display', 'lat': 51.05253728748261...	51.052537	13.737704	NaN	NaN	Sachsen	501bffb0e4b0a704df534a59
6	Hans im Glück - Burgergrill	Burger Joint	Altmarkt 24	DE	Dresden	Deutschland	NaN	86	[Altmarkt 24, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.04987221052044...	51.049872	13.737264	NaN	01067	Sachsen	56696c60498e387b4d3cbac0
7	Starbucks	Coffee Shop	Altmarkt 7	DE	Dresden	Deutschland	NaN	16	[Altmarkt 7, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.04926397345353...	51.049264	13.737925	NaN	01067	Sachsen	4c39cfb03849c92866d1c1b1
8	Historisches Grünes Gewölbe	Art Museum	Taschenberg 2	DE	Dresden	Deutschland	im Residenzschloss	351	[Taschenberg 2 (im Residenzschloss), 01067 Dre...	[{'label': 'display', 'lat': 51.05242395737157...	51.052424	13.737137	NaN	01067	Sachsen	4b9cdbf4f964a520197f36e3
9	LadenCafé aha	Vegetarian / Vegan Restaurant	Kreuzstr. 7	DE	Dresden	Deutschland	NaN	112	[Kreuzstr. 7, 01067 Dresden, Deutschland]	[{'label': 'display', 'lat': 51.04885644791869...	51.048856	13.739572	NaN	01067	Sachsen	4cf6a573c020b60caf4e2175

Afterwards I'll check the categories, count the frequencies of them, and cluster the cities with these features.

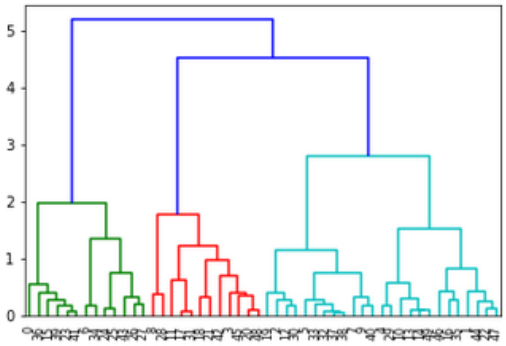
3. Methodology

Our aim is to cluster cities. There exist many models for clustering. I'll use two of them: k-means and hierarchical clustering.

K-means is the model that is considered one of the simplest models amongst clustering models. It's very popular and useful if you need insights from unlabeled data quickly.
The method partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.



The second method I'll use is hierarchical clustering. It's a method of cluster analysis, which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types: Agglomerative ("bottom-up") and Divisive ("top-down"). For the clustering of cities I used the agglomerative approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. I link cities together with the maximum distance between elements of each cluster. The results of hierarchical clustering are usually presented in a dendrogram. Each merge is represented by a horizontal line. By moving up from the bottom layer to the top node, a dendrogram allows us to reconstruct the history of merges that resulted in the depicted clustering.



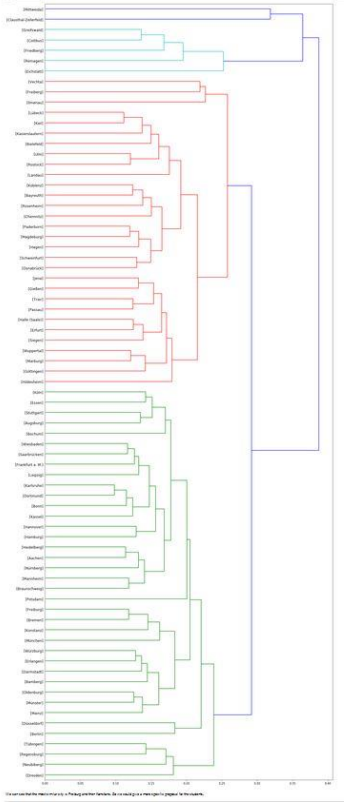
4. Results

Here is the map with the different colored (per cluster) university cities in Germany:

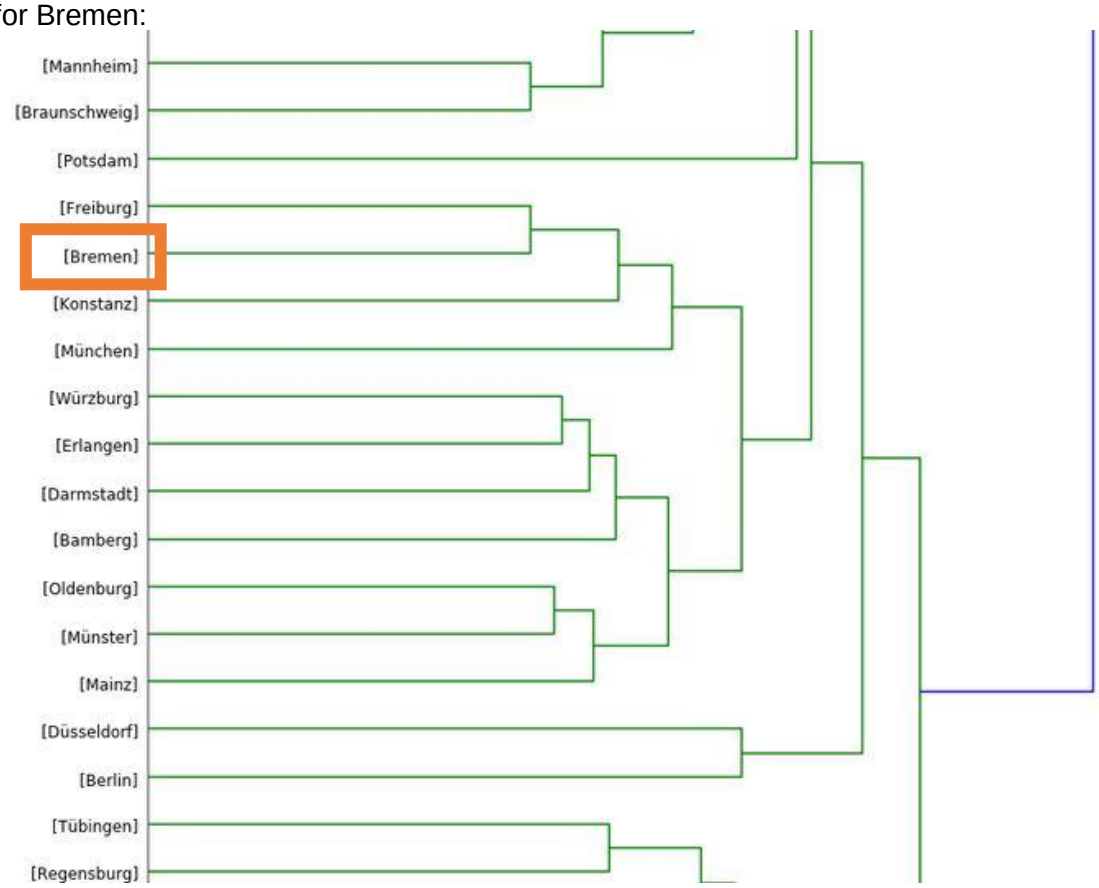
And the cities which are in the same cluster like Bremen: Aachen, Augsburg, Bielefeld, Bochum, Bonn, Braunschweig, Darmstadt, Dortmund, Düsseldorf, Essen, Frankfurt a. M., Heidelberg, Karlsruhe, Kassel, Kiel, Köln, Leipzig, Mainz, Mannheim, München, Münster, Nürnberg, Potsdam, Saarbrücken, Stuttgart, Wiesbaden

	City	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	Aachen	Café	Bar	Coffee Shop	Sushi Restaurant	Italian Restaurant	Bakery	Park	Supermarket	Plaza	Ice Cream Shop
1	Augsburg	Italian Restaurant	Café	Steakhouse	Bar	Beer Garden	German Restaurant	Hotel	Bakery	Pub	Turkish Restaurant
5	Bielefeld	Bar	Supermarket	Restaurant	Café	Burger Joint	Hotel	Italian Restaurant	Sushi Restaurant	Middle Eastern Restaurant	Park
6	Bochum	Bakery	Café	Supermarket	Park	Event Space	Sushi Restaurant	Bar	Ice Cream Shop	Market	Asian Restaurant
7	Bonn	Café	German Restaurant	Pub	Italian Restaurant	Plaza	Museum	Burger Joint	Movie Theater	Korean Restaurant	Coffee Shop
8	Braunschweig	Café	Italian Restaurant	Bar	Park	Turkish Restaurant	Plaza	Coffee Shop	Pizza Place	Steakhouse	Restaurant
9	Bremen	Café	Hotel	German Restaurant	Bar	Italian Restaurant	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Burger Joint	Market	Asian Restaurant
13	Darmstadt	Café	Italian Restaurant	German Restaurant	Middle Eastern Restaurant	Supermarket	Organic Grocery	Park	Beer Garden	Sushi Restaurant	Bar
14	Dortmund	Café	Italian Restaurant	German Restaurant	Coffee Shop	Pub	Turkish Restaurant	Trattoria/Osteria	Japanese Restaurant	Sushi Restaurant	Bar
16	Düsseldorf	Japanese Restaurant	Hotel	Coffee Shop	Brewery	Grocery Store	Plaza	Bar	Park	Dessert Shop	Art Museum
20	Essen	Café	Italian Restaurant	Pub	Restaurant	Ice Cream Shop	Nightclub	Falafel Restaurant	Park	Indie Movie Theater	Plaza
21	Frankfurt a. M.	Café	Bar	Hotel	German Restaurant	Art Museum	Restaurant	Japanese Restaurant	Plaza	Market	Burger Joint
32	Heidelberg	Café	Bar	Plaza	Coffee Shop	Hotel	Falafel Restaurant	Burger Joint	Italian Restaurant	Cocktail Bar	Ice Cream Shop
37	Karlsruhe	Café	Italian Restaurant	Pub	Coffee Shop	Turkish Restaurant	German Restaurant	Bar	Burger Joint	Ice Cream Shop	Thai Restaurant
38	Kassel	German Restaurant	Café	Cocktail Bar	Ice Cream Shop	Italian Restaurant	Drugstore	Pub	Art Museum	Art Gallery	Plaza
39	Kiel	Café	Coffee Shop	Hotel	Supermarket	Gas Station	Lounge	Japanese Restaurant	Italian Restaurant	Burger Joint	Restaurant
41	Köln	Italian Restaurant	Café	Hotel	Plaza	Bakery	Brewery	Art Museum	Scenic Lookout	Vietnamese Restaurant	Cocktail Bar
44	Leipzig	Hotel	Café	Zoo Exhibit	Coffee Shop	Bar	Irish Pub	Cocktail Bar	Park	Plaza	German Restaurant
47	Mainz	German Restaurant	Café	Ice Cream Shop	Bar	Coffee Shop	Plaza	Wine Bar	Restaurant	Pizza Place	Supermarket
48	Mannheim	Café	Bar	Italian Restaurant	Hotel	Burger Joint	Coffee Shop	Park	Turkish Restaurant	German Restaurant	Clothing Store
51	München	Café	Plaza	German Restaurant	Hotel	Ice Cream Shop	Cocktail Bar	Church	Italian Restaurant	Department Store	Coffee Shop
52	Münster	Café	Italian Restaurant	Coffee Shop	German Restaurant	Bar	Park	Supermarket	Hotel	Nightclub	Ice Cream Shop
54	Nürnberg	Café	Bar	Coffee Shop	Hotel	Burger Joint	Steakhouse	Franconian Restaurant	Asian Restaurant	Ice Cream Shop	Sushi Restaurant
59	Potsdam	Historic Site	Palace	Park	German Restaurant	Café	French Restaurant	Coffee Shop	Ice Cream Shop	Hotel	Italian Restaurant
64	Saarbrücken	Café	Bar	Italian Restaurant	German Restaurant	Coffee Shop	Turkish Restaurant	Ice Cream Shop	Sushi Restaurant	Nightclub	Hotel
67	Stuttgart	Italian Restaurant	Bar	German Restaurant	Café	Plaza	Park	Ice Cream Shop	Cocktail Bar	Sushi Restaurant	Movie Theater
72	Wiesbaden	German Restaurant	Café	Park	Plaza	Hotel	Thai Restaurant	Bar	Falafel Restaurant	Italian Restaurant	Burger Joint

The dendogramm of the hierarchical clustering:



And for Bremen:



5. Discussion

The kMeans clustering on the 10 most common venue-categories yields three big clusters and some smaller ones. The hierarchical clustering on the frequency of all venue-categories in the cities yields 4 big clusters but the most similar city can be spotted in the dendogram. The differences in the results are on the one hand because of the different data and on the other hand because of the different method. For the student who studies in Bremen we should use the hierarchical clustering because Bremen is a city in a quite big cluster, so we have 26 cities, which are quite similar. I would recommend a student from Bremen to have a look on the dendogram and have a closer look at the cities of Freiburg, Konstanz and Munich.

6. Conclusion

To get better results, I could add more data to the dataset. For example, it's also interesting analyzing available apartments, living costs, the total number of students and inhabitants, other faculties and many more data.

Once there is a good model, I could built an application for students. They have to put in their subject, the actual city and other necessary information and they'll receive a recommendation for some similar cities.