

Report on Possibilities to open an Amazon Food Restaurant in Leipzig, Germany

Carlos Granados

May the 13th, 2020

1 Introduction

In the last years Leipzig has become a quite attractive and vibrant city in the Sachsen state in Germany. The past of the city and its cultural scene offers many new possibilities. In the city it is nowadays relatively easy to find a quite wide variety of restaurant options, but once one looks closer, it becomes clear that such offers are not so radically different from other metropolis in Germany in particular or in Europe in general.

For such reason it is quite interesting to explore the possibility to open new restaurants with really new offers, particularly with new cuisine styles, as the one that we are proposing here: A restaurant that gives its customers the possibility to taste the flavors from the Amazonas forest!

2 Data Acquisition and Cleaning

To start we will analyze different areas of Leipzig, in order to find the sectors more visited in different days of the week and to identify those with a higher density of restaurants. The focus must remain mainly on the commercial areas. In order to do so, we will use the Foursquare location data of different points of interest of the city, to classify them according to the number of visits and then on the number of restaurants and bars nearby. Different clustering modeling of the data will be crucial for such purpose.

Several options for the clustering can be explored, as in the type of restaurants or in the public interest on the sectors (visits frequency and times of the day with more traffic).

Once such sectors are identified, the next step the areas where alternative restaurants are, in order to explore the possibility to open the restaurant in one of such sectors. The clustering modeling obtained will help in order to determine if commercial or touristic areas are more suited to the kind of new restaurant.

2.1 Data Sources

We consider different points of interest in Leipzig, as well some commercial or touristic neighborhoods. Such points were be transformed to geographical coordinates and with the help of Foursquare location data they were be classified according to the number of visits to the sector in different days of the week and the number/kind of restaurants/bars nearby.

As as source the [Wikipedia](#) article about Leipzig was used. Different points of the city as recognized buildings, parks, museums, churches and touristic landmarks were extracted using the `bs4` (`BeautifulSoup`) package in `Python`. In order to do so, all the hyperlinks of the page were saved in a list, and then a search for different key names were was performed, such as “Main sights” or “Museums and arts”. For each search only a given number of points were saved, typically between 15 and 20 for each cathegory.

In this initial recollection of data, we register a total of 70 names.

2.2 Data Cleaning

Since we needed a list of different places in the city, we starting cleaning the data set with dropping duplicates and different personality names associated with the city, as Johann Sebastian Bach or Johann Wolfgang von Göthe, or different historical names as DDR or East Germany.

Once those names were dropped, we started to get the geographical coordinates for each point, using the `geocoder` package. If the status of the call was `OK`, the coordinates were saved; on the other hand if the status was `ZERO_RESULTS`, meaning that the name is not a geographical point, the coordinates were set to a `NaN`. Once all the coordinates were processed, we proceed to drop all the `NaNs` values. After all this cleaning process we finish with 51 different names.

We considered that no more cleaning procedures were needed in this stage of the project. The resulting list was transformed to a `pandas` data frame.

2.3 Additional Features

Once we had our set of names with its corresponding set of coordinates, we start to collect the names of different venues, collecting data through Foursquare data bases. Initially we wanted to analyze the trending venues for different sectors for different times, but unfortunately Foursquare does not provides any trending information for Leipzig. That forced us to use only the exploring option. Then, following different examples, we saved the top ten of venues around each saved point, and merged it with our original data frame.

-----Bach Archive-----			-----Battle of Leipzig-----		
	Venue	Freq.		Venue	Freq.
0	Sushi Restaurant	0.10	0	Tram Station	0.4
1	Theater	0.10	1	Historic Site	0.2
2	Hotel	0.05	2	Business Service	0.2
3	Lounge	0.05	3	Monument / Landmark	0.2
4	History Museum	0.05	4	African Restaurant	0.0
5	Gym / Fitness Center	0.05	5	Museum	0.0
6	German Restaurant	0.05	6	Plaza	0.0
7	Gastropub	0.05	7	Playground	0.0
8	Plaza	0.05	8	Pet Store	0.0
9	Deli / Bodega	0.05	9	Park	0.0

-----Leipzig Hauptbahnhof-----			-----Leipzig Zoological Garden-----		
	Venue	Freq.		Venue	Freq.
0	Hotel	0.25	0	Zoo Exhibit	0.25
1	Modern European Restaurant	0.05	1	Hotel	0.10
2	Restaurant	0.05	2	Hotel Bar	0.05
3	Irish Pub	0.05	3	Greek Restaurant	0.05
4	Coffee Shop	0.05	4	Zoo	0.05
5	Cocktail Bar	0.05	5	Trattoria/Osteria	0.05
6	Shopping Mall	0.05	6	Theme Restaurant	0.05
7	Sports Bar	0.05	7	Science Museum	0.05
8	Steakhouse	0.05	8	Plaza	0.05
9	Newsstand	0.05	9	Playground	0.05

Figure 1: Obtained venues top ten and its frequency, using Foursquare, for four different points of interest in Leipzig.

3 Exploratory Data Analysis

We started looking the obtained data frame, analyzing the venues near each selected point in the city. Some of the results are shown in Fig. 1. It is clear that hotels are a quite frequent venue around several points, in particular in the central train station (Hauptbahnhof); on the other hand, we see that a light train stop (Tram station) is quite frequent around the Battle of Leipzig, a known monument outside of the city. In Fig. 1 we see also different restaurants around each point. Such venues are used as key points to classify all considered points in the data frame.

4 Modelling

In order to classify the data we used the K-Means clustering method. To do so we loaded the `scikit learn` package. We considered different options, but one of the most defined clustering was obtained for $k = 10$. In Figs. 2 to 8 we show some snapshots of the different clusters. The first five clusters are the most populated ones, composed mainly by points of interest located at the inner part of the city. Some of the clusters can be visualized in the map shown in Fig. 9.

The first cluster is characterized by having a relative large number of hotels, restaurants, bars, coffee shops and other main sights nearby (see Fig. 2), while the second cluster (see Fig. 3) have several tram stops, super-

Cluster 1:

```
L_merged.loc[L_merged['Cluster Labels'] == 0, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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	Auerbachs Keller	Hotel	Gym / Fitness Center	Ice Cream Shop	Boutique	Bistro	Spanish Restaurant	Café	Italian Restaurant	Sushi Restaurant	Plaza
1	Augusteum (Leipzig)	Hotel	Plaza	Church	Bed & Breakfast	Concert Hall	Museum	Café	Scenic Lookout	Deli / Bodega	Lounge
2	Augustusplatz	Hotel	Plaza	Bistro	Church	Coffee Shop	Nightclub	Concert Hall	Café	Museum	Scenic Lookout
8	City-Hochhaus Leipzig	Hotel	Plaza	Church	Bed & Breakfast	Concert Hall	Gym / Fitness Center	Café	Museum	Scenic Lookout	Deli / Bodega
9	Europahaus	Hotel	Museum	Plaza	Nightclub	Burger Joint	Drugstore	Bed & Breakfast	Bar	Café	Scenic Lookout
18	Hotel The Westin Leipzig	Hotel	Zoo Exhibit	Steakhouse	Art Museum	Bagel Shop	Coffee Shop	Hotel Bar	Modern European Restaurant	Restaurant	Shopping Mall
21	Leipzig Hauptbahnhof	Hotel	Sports Bar	Irish Pub	Newsstand	Modern European Restaurant	Restaurant	Donut Shop	Bistro	Shopping Mall	Japanese Restaurant
29	MDR-Hochhaus	Hotel	Plaza	Church	Bed & Breakfast	Concert Hall	Gym / Fitness Center	Café	Museum	Scenic Lookout	Deli / Bodega
31	Museum der bildenden Künste	Hotel	Bistro	Bagel Shop	Steakhouse	Plaza	Sushi Restaurant	Indie Movie Theater	Ice Cream Shop	Lounge	Italian Restaurant
32	Museum of Antiquities of the University of Lei...	Hotel	Bistro	Plaza	Concert Hall	Deli / Bodega	Lounge	Opera House	Café	Italian Restaurant	Church
38	Paulinum (University of Leipzig)	Hotel	Plaza	Church	Drugstore	Concert Hall	Gym / Fitness Center	Café	Scenic Lookout	Deli / Bodega	Boutique
42	St. Nicholas Church, Leipzig	Hotel	Plaza	Bistro	Church	Lounge	Opera House	Café	Italian Restaurant	Deli / Bodega	Concert Hall
49	University of Leipzig	Hotel	Plaza	Church	Drugstore	Concert Hall	Gym / Fitness Center	Café	Scenic Lookout	Deli / Bodega	Boutique
50	Wintergartenhochhaus	Hotel	Bistro	Opera House	Bagel Shop	Coffee Shop	Steakhouse	Bar	Donut Shop	Bakery	Plaza

Figure 2: First cluster obtained using the K-Means clustering method in Python, for different points of interest in Leipzig. This cluster is characterized by having a relative large number of hotels and restaurants nearby.

markets, bars, plazas and diverse shops. In the third cluster (see Fig. 4) we see several hotels, museums, supermarkets, coffees shops and plazas; in the fourth one (see Fig. 5) the venues with a higher frequency are coffee shops, different type of restaurant, bars and night clubs. In the fifth cluster (see Fig. 6) we see restaurants, hotels, exhibitions, bars and theaters. The other clusters (see Figs. 7 to 8) are composed by one or two points, where most of them are located outside of the city center. They can be considered as the outliers of our original data set.

From the results that we obtained with the clustering classification, we can say the the clusters one, three, four and five are good initial candidates to locate our new restaurant. The largest cluster is the first one, and is conformed by several important landmarks in the city center. The other clusters contain several museums and other main points.

Cluster 2:

```
L_merged.loc[L_merged['Cluster Labels'] == 1, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
4	Battle of Leipzig	Tram Station	Historic Site	Business Service	Monument / Landmark	Event Space	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega
6	Center Torgauer Platz	Tram Station	Supermarket	Plaza	Pet Store	Grocery Store	Donut Shop	Clothing Store	Cocktail Bar	Coffee Shop	Concert Hall
14	German Museum of Books and Writing	Tram Station	Automotive Shop	Plaza	Event Space	Gym	Nightclub	Bar	Concert Hall	Construction & Landscaping	Cupcake Shop
15	German National Library	Tram Station	Plaza	Automotive Shop	Nightclub	Mediterranean Restaurant	Bar	Furniture / Home Store	Supermarket	Hockey Rink	Gym
30	Monument to the Battle of the Nations	Tram Station	Historic Site	Business Service	Monument / Landmark	Event Space	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega
47	Südfriedhof (Leipzig)	Tram Station	Historic Site	Monument / Landmark	Event Space	Coffee Shop	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega

Figure 3: Same as Fig. 2, bur for the second cluster. It is characterized mainly by tram stops. Most of them are located in the surroundings of the city.

Cluster 3:

```
L_merged.loc[L_merged['Cluster Labels'] == 2, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
16	Grassi Museum	Hotel	Museum	Supermarket	Coffee Shop	Plaza	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Gym	Drugstore
22	Leipzig Museum of Applied Arts	Hotel	Museum	Supermarket	Plaza	Karaoke Bar	Coffee Shop	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Exhibit
23	Leipzig Museum of Ethnography	Hotel	Museum	Supermarket	Coffee Shop	Plaza	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Gym	Drugstore
33	Museum of Musical Instruments of the Universit...	Hotel	Museum	Supermarket	Plaza	Karaoke Bar	Coffee Shop	Gym / Fitness Center	Art Gallery	Trattoria/Osteria	Exhibit

Figure 4: Same as Fig. 2, bur for the third cluster. It is characterized to have hotels, museums, supermarkets and several coffee shops nearby. The cluster itself is composed by museums.

Cluster 4:

```
L_merged.loc[L_merged['Cluster Labels'] == 3, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
10	Federal Administrative Court of Germany	Café	Chinese Restaurant	Bar	Bank	Sushi Restaurant	Bakery	Beer Store	Bagel Shop	Tech Startup	Soup Place
19	Leipzig Bayerischer Bahnhof	Tram Station	Hookah Bar	Bike Shop	German Restaurant	Italian Restaurant	Drugstore	Dessert Shop	Nightclub	Chinese Restaurant	Café
37	New Town Hall (Leipzig)	Café	Bar	Nightclub	Clothing Store	Park	Chinese Restaurant	Gym / Fitness Center	Plaza	Cupcake Shop	Deli / Bodega
40	Propsteikirche, Leipzig	Café	Beer Store	Grocery Store	German Restaurant	Dessert Shop	Deli / Bodega	Cupcake Shop	Coffee Shop	Nightclub	Park
41	Reichsgericht	Café	Tech Startup	Soup Place	Beer Store	Bar	Sushi Restaurant	Bakery	Italian Restaurant	Bagel Shop	Park
43	St. Peter, Leipzig	Chinese Restaurant	Café	Sushi Restaurant	Hotel	German Restaurant	Drugstore	Dessert Shop	Cupcake Shop	Organic Grocery	Cocktail Bar
44	St. Peters'	Chinese Restaurant	Café	Sushi Restaurant	Hotel	German Restaurant	Drugstore	Dessert Shop	Cupcake Shop	Organic Grocery	Cocktail Bar
48	Tower of New Town Hall	Café	Bar	Nightclub	Clothing Store	Park	Chinese Restaurant	Gym / Fitness Center	Plaza	Cupcake Shop	Deli / Bodega

Figure 5: Same as Fig. 2, bur for the fourth cluster. It is characterized mainly by having coffee shops, restaurants and bars as the most frequent venues nearby.

5 Conclusions

From the results that we obtained with the clustering classification, we can say the the clusters one, three, four and five are good initial candidates to locate our new restaurant. The largest cluster is the first one, and is conformed by several important landmarks in the city center; this make it quite interesting, but it also indicates that the competence can be quite high. The second cluster can be discarded, since it contains several landmarks that are not relatively close to other bars or restaurants in the city center. The third and fifth clusters are considered the most attractive, since there are several hotels and museums around them, where the probability to find the correct consumers for the restaurant is considerably higher. The fourth cluster contain several coffee shops and bars, making it a second option, but another factors must be considered.

With the analysis performed here we make a more educated choice of the sectors of Leipzig where it is possible to open a new restaurant. For a second stage of the project more points and maybe another restaurants with a similar profile can be included in the data set. The city regulations must be taken into account too, as also a profile on the availability of space and the rent prices.

Cluster 5:

```
L_merged.loc[L_merged['Cluster Labels'] == 4, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
3	Bach Archive	Sushi Restaurant	Theater	Gym / Fitness Center	Coffee Shop	Plaza	Deli / Bodega	Burger Joint	Restaurant	Lounge	Shopping Mall
5	Cantor (church)	Sushi Restaurant	Theater	Clothing Store	Restaurant	Deli / Bodega	Shopping Mall	Plaza	Spanish Restaurant	Indie Movie Theater	Ice Cream Shop
12	Funkturn Leipzig	Exhibit	Supermarket	Restaurant	Auto Dealership	Coffee Shop	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega
13	G2 Kunsthalle	Sushi Restaurant	Spanish Restaurant	Tapas Restaurant	Lounge	Plaza	Burger Joint	Restaurant	Indie Movie Theater	Coffee Shop	Bar
17	Hochhaus Löhr's Carree	Hotel	Zoo Exhibit	Supermarket	Art Museum	Currywurst Joint	Hotel Bar	Indie Movie Theater	Italian Restaurant	Japanese Restaurant	Modern European Restaurant
20	Leipzig Debate	Hotel	Sushi Restaurant	Restaurant	Deli / Bodega	Shopping Mall	Plaza	Spanish Restaurant	Coffee Shop	Indie Movie Theater	Ice Cream Shop
25	Leipzig Synagogue	Hotel	Restaurant	Sports Bar	Steakhouse	Supermarket	Sushi Restaurant	Italian Restaurant	Bagel Shop	Currywurst Joint	Ice Cream Shop
27	Leipzig Zoological Garden	Zoo Exhibit	Hotel	Science Museum	Greek Restaurant	Zoo	Hotel Bar	Italian Restaurant	Latin American Restaurant	Modern European Restaurant	Park
34	Naturkundemuseum Leipzig	Hotel	Zoo Exhibit	Hotel Bar	Currywurst Joint	Pub	Burger Joint	Science Museum	Italian Restaurant	Indie Movie Theater	Ice Cream Shop
39	Primate	Hotel	Sushi Restaurant	Restaurant	Deli / Bodega	Shopping Mall	Plaza	Spanish Restaurant	Coffee Shop	Indie Movie Theater	Ice Cream Shop
45	St. Thomas Church, Leipzig	Sushi Restaurant	Theater	Clothing Store	Restaurant	Deli / Bodega	Shopping Mall	Plaza	Spanish Restaurant	Indie Movie Theater	Ice Cream Shop
46	Stasi	Bar	Hotel	Ice Cream Shop	Thai Restaurant	Plaza	Burger Joint	Indie Movie Theater	Clothing Store	Supermarket	Sushi Restaurant

Figure 6: Same as Fig. 2, bur for the fifth cluster. It is characterized by venues as restaurants, theaters, exhibitions, bars and plazas.

Cluster 6:

```
L_merged.loc[L_merged['Cluster Labels'] == 5, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
36	New Leipzig School	Market	American Restaurant	Construction & Landscaping	Fast Food Restaurant	Concert Hall	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop	Donut Shop

Cluster 7:

```
L_merged.loc[L_merged['Cluster Labels'] == 6, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
7	Chimney of Stahl- und Hartgusswerk Bösdorf GmbH	Train Station	Zoo Exhibit	Exhibit	Coffee Shop	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop

Cluster 8:

```
L_merged.loc[L_merged['Cluster Labels'] == 7, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
24	Leipzig Panometer	Exhibit	Concert Hall	Cafeteria	Bus Stop	Cocktail Bar	Construction & Landscaping	Cupcake Shop	Currywurst Joint	Deli / Bodega	Dessert Shop

Figure 7: Same as Fig. 2, bur for the sixth, seventh and eighth clusters. All of them are composed by one city point, making them a part of the outliers from the original data set.

Cluster 9:

```
L_merged.loc[L_merged['Cluster Labels'] == 8, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
26	Leipzig Trade Fair	Fast Food Restaurant	Hardware Store	Tram Station	Bus Stop	Fountain	Event Space	Concert Hall	Construction & Landscaping	Cupcake Shop	Currywurst Joint

Cluster 10:

```
L_merged.loc[L_merged['Cluster Labels'] == 9, L_merged.columns[[0] + list(range(4, L_merged.shape[1]))]]
```

	Place Name	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
28	Leipziger Baumwollspinnerei	Art Gallery	Bistro	Arts & Crafts Store	Pub	Rental Car Location	Café	Music Venue	Park	Zoo Exhibit	Concert Hall
35	Neo Rauch	Art Gallery	Park	Arts & Crafts Store	Café	Music Venue	Bistro	Event Space	Construction & Landscaping	Cupcake Shop	Currywurst Joint

Figure 8: Same as Fig. 2, bur for the ninth and tenth clusters. Similar to the case for the clusters presented in Fig. 7.

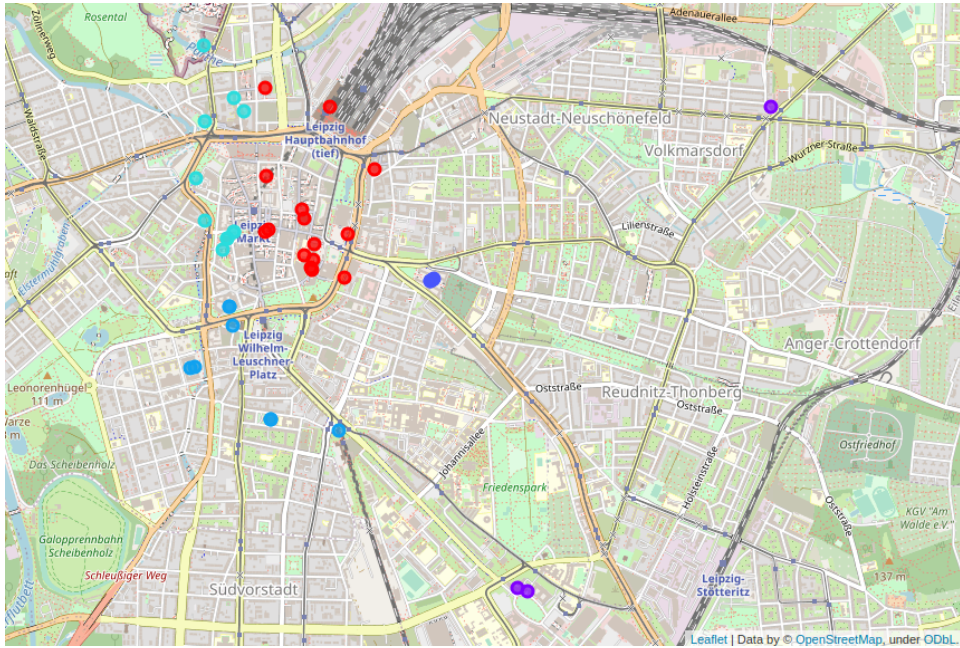


Figure 9: Map of Leipzig, Germany, showing some of the clusters obtained using the K-Means method. The first cluster (see Fig. 2) is in red, the second (Fig. 3) in purple, the third (Fig. 4) in dark blue, the fourth (Fig. 5) in light blue and the fifth (Fig. 6) is in cyan.