Capstone Project- The Battle of Neighborhoods

Opening a café chain in a new city

Introduction

Business Problem

A foreign café chain company wants to expand its business into new cities and therefore plans to introduce its brand to Cologne. They seek to follow a very aggressive expansion strategy and plan to open several new locations in short succession. To achieve this, they need a list with possible locations with a high suitability for their new cafe locations. Their plan is to be highly visible to customers by being as near as they could get to them to attract their attention.

The idea to realize this is to open the cafes in the very near vicinity of tram stations. They exclude bus stops since busses are usually not carrying so much passengers than trams. Hence, identifying tram stations in the city with a high suitability for opening a new café nearby is the main goal.

They assume that the customer density per café location/tram station will increase with every tram line stopping at the station and with every university or school site in the vicinity. They further expect, that their potential share on the customer base will decrease with every additional competitor in the vicinity of the station. Based on this assumption they plan to calculate a "suitability index" indicating if a station us suitable to open one of their cafes or not.

Audience of this analysis

Relevant persons for making the decision on how to further expand into the city of Cologne and help them with the business problem at hand.

Data

Open Street Map

http://nominatim.openstreetmap.org

Determine the center of Cologne as the starting point by its longitudinal and latitudinal coordinates.

Open data archive of the public transportation company KVB from Cologne

https://online-service.kvb-koeln.de

Data set

https://online-service.kvb-koeln.de/geoserver/OPENDATA/ows?service=WFS&version=1.0.0&request=GetFeature&typeName=ODENDATA:haltestellen&outputFormat=csv

The dataset has several hundred of entries of all kinds of public transport stations in and around Cologne which are serviced by KVB. The data first needs to be filtered for the tram stations we are

interested in. Afterwards the data needs further refinement to get the needed information like station names, coordinates of the stations or numbers of tram lines servicing it.

	FID	ASS	Name	Kurzname	Haltestellenbereich	Koordinate	Betriebsbereich	Linien
0	haltestellen.fid-5764443a_16b5a8f4c71_756	111	Heumarkt	HMG	1	POINT (6.9604570988 50.9357563921)	STRAB	1 7 9
1	haltestellen.fid-5764443a_16b5a8f4c71_757	112	Heumarkt	HMG	1	POINT (6.9595737676 50.9357045307)	STRAB	1 7 9
2	haltestellen.fid-5764443a_16b5a8f4c71_758	113	Heumarkt	HMG	1	POINT (6.9574528147 50.9351045185)	STRAB	5
3	haltestellen.fid-5764443a_16b5a8f4c71_759	114	Heumarkt	HMG	1	POINT (6.9579496548 50.9355407748)	STRAB	5
4	haltestellen.fid-5764443a_16b5a8f4c71_75a	121	Heumarkt	HMG	1	POINT (6.9603999897 50.9355099544)	BUS	106 132 133

Figure 1: Exemplary data extract

Foursquare

https://api.foursquare.com

From Foursquare the data on possible competition around tram stations is obtained. As relevant competition can be seen bakeries and other cafes. Further we fetch data on university and school buildings around the station (called potential high customer sites) to determine if there is an additional potential customer base added to the occasional customers searching for a café to rest and enjoy a good coffee.

From this data the “suitability index” for opening a new café near this particular station will be derived and used as a decision base.

All data will then be connected to obtain the most promising stations with a high suitability for our new cafes – meaning a big customer base without much competition. These refined results will then be displayed on a map of Cologne to give an easy to understand and handle tool for the decision makers to select the best locations for their expansion.

Methodology

After collecting the data from the different sources, it needs to be processed to transform it into a format it can be easily handled to achieve the goal to determine the best locations for the cafes.

The café chain brand plans to enter the city Cologne, so in a first step the coordinates for this city are acquired from Open Street Map (OSM). After extracting the information from the OSM response the coordinates are in a processable format and can be use throughout the notebook.

The data from the open data archive from the Kölner Verkehrs-Betriebe AG (KVB) is of a different kind. As it can be seen in Figure 1, the information we need is present in the data but not in an easily processable format. The data needs some refinement. The information wanted is

- Station name
  Column: Name; Information in a clear format.
- Coordinates
  Column: Koordinate; Information in an obfuscated format; Unneeded strings and characters need to be removed and the remainder split into separated values for longitude and latitude.

- Number of lines stopping at the station
  Column: Linien; Information in unsuitable format: The line numbers stopping at the station are mentioned, they need to be counted.
- Further, with the extracted coordinates of the stations their distance to the city center is calculated.
- Only information on tram stations, bus stations should be removed.

The refined information is stored into a new data frame shown in Figure 2.

	Name	Long	Lat	Lines count	Distance
0	Heumarkt	6.9604570988	50.9357563921	3	0.291943
1	Neumarkt	6.9481136124	50.935762763	3	0.895969
2	Poststr.	6.9500805214	50.9316916582	4	1.025377
3	Mauritiuskirche	6.9450683908	50.9346873217	1	1.141494
4	Rathaus	6.9595586023	50.9379333709	1	0.056102

Figure 2: Refined data on tram stations

To check the information itself and the whole refinement process for correctness the station data should be displayed for a visual inspection of the data.

In advance it needs to be defined how the shown data should look like to specify some passing or failing criterions for the data for the next processing steps. The criterions are:

- many stations in the city itself
- fewer ones outside the city
- there should be some kind of line pattern discernable

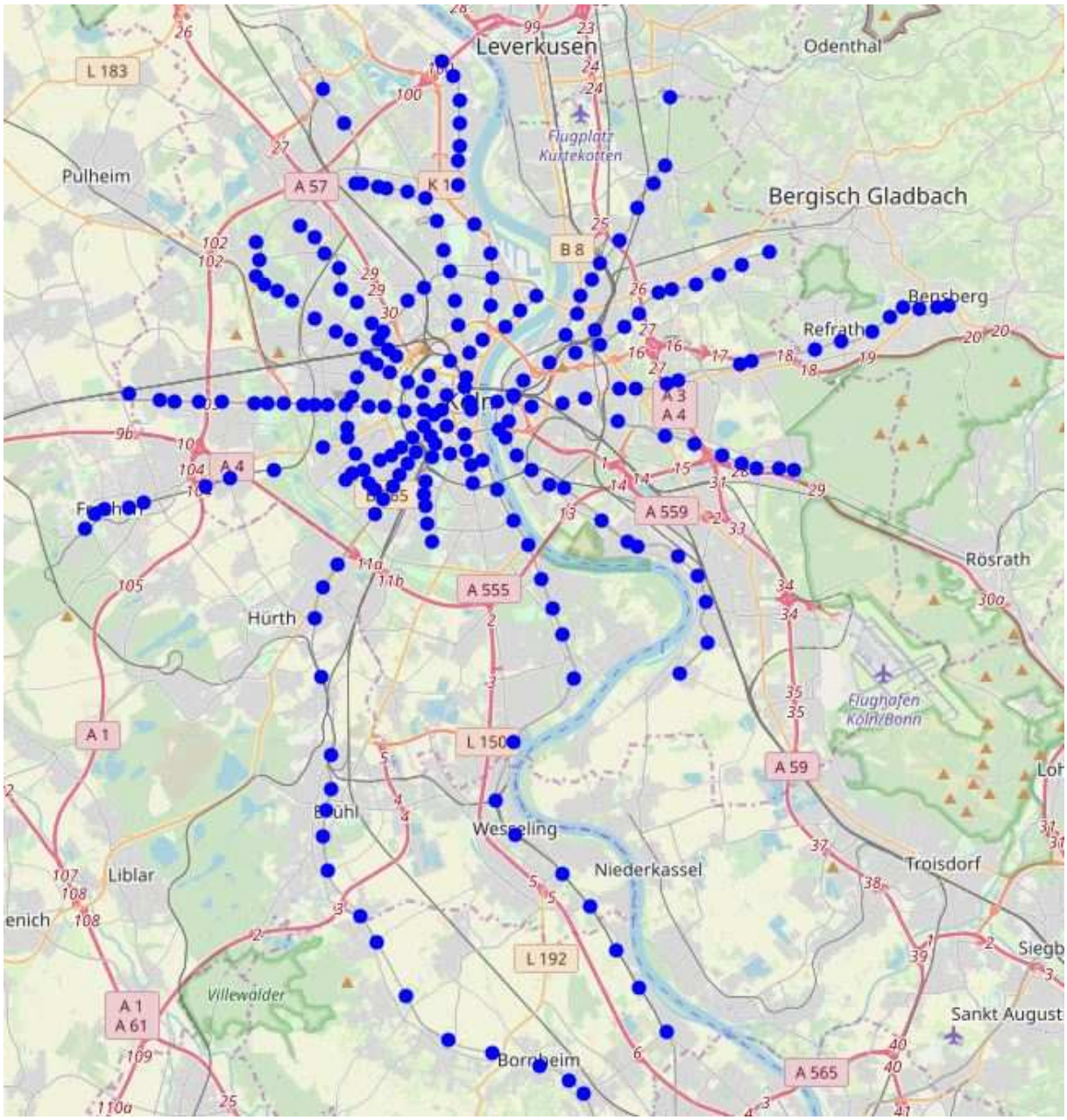

Figure 3: Map of Cologne depicting all KVB tram stations

The visualization in Figure 3 complies with the defined criterions and should be used further on.

The business plan foresees that new chain stores should be opened in the vicinity of tram stations. In this vicinity existing bakeries and cafes are seen as competition lowering the suitability of the station for opening one of the brands stores. Whereas sites with a high count of potential customers are seen as a plus and increase the suitability of the station.

To acquire the information to determine the suitability of the station, in the next data acquisition step location data from Foursquare is collected. As competition are define:

- bakery : 4bf58dd8d48988d16a941735
- café : 4bf58dd8d48988d16d941735

and as potential high customer sites

- university : 4d4b7105d754a06372d81259
- school : 4bf58dd8d48988d13b941735

In the map below they are shown as red markers for competition and green markers for high customer sites.

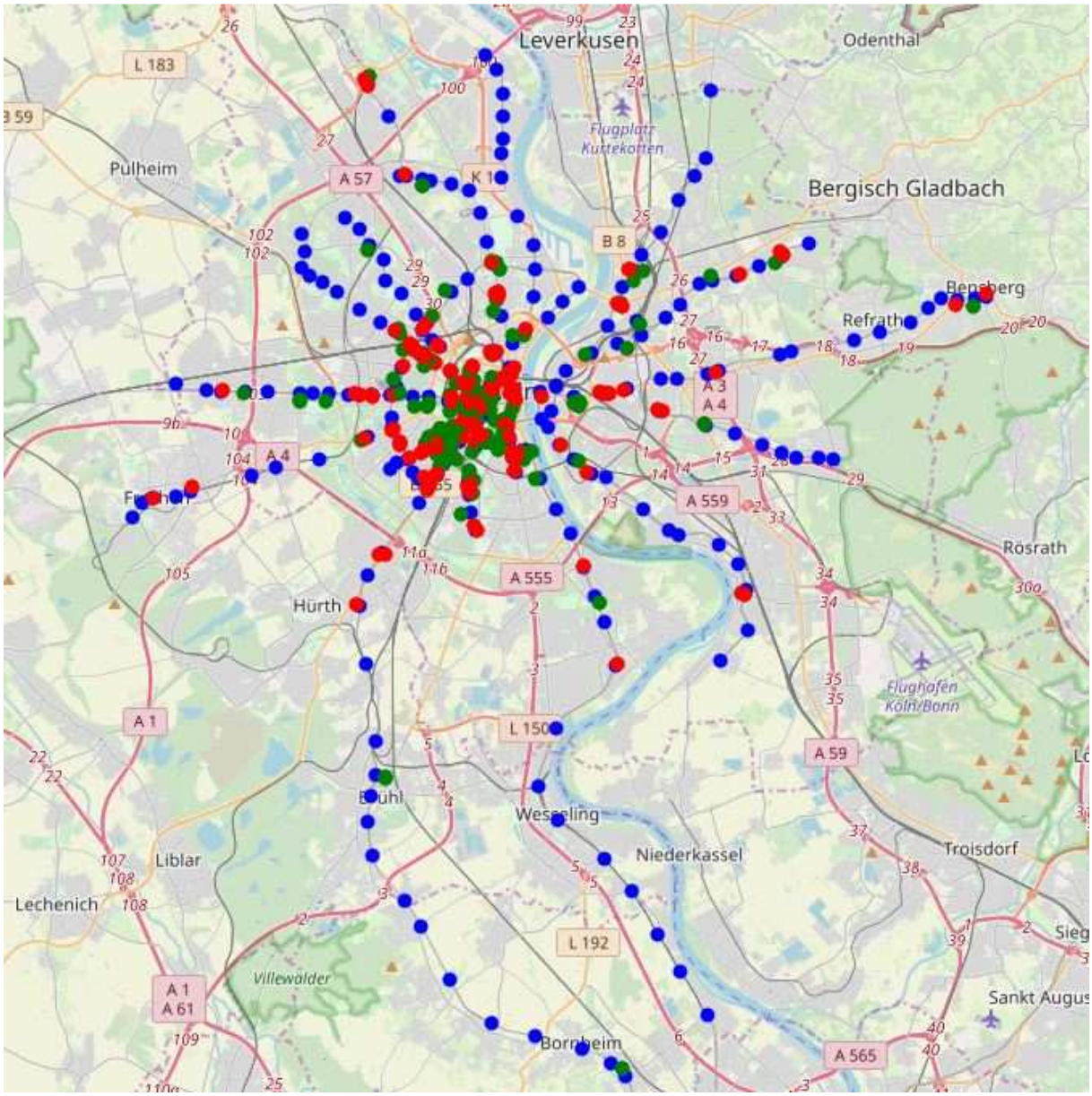

Figure 4: Map of Cologne depicting all KVB tram stations and the surrounding competition and hich customer sites.

The information density makes it impossible to read anything useful on preferable locations for our business from the map. So we use an easy metric to calculate a suitability index for a potential cafes of ours for each stations.

Metric definition:

(Number of tram lines stopping at the station + number of high customer sites) / number of competitors

	Name	Long	Lat	Lines count	Distance	Comp	Cust	Suitability
54	Universität	6.9312003766	50.9259899857	1	2.475556	0.0	22.0	23.0
11	Ulrepforte	6.951437134	50.9243235691	2	1.677367	0.0	8.0	10.0
31	Deutz Technische Hochschule	6.9859264567	50.9365857842	2	1.866090	0.0	6.0	8.0

Figure 5: Table depicting the suitability index in the last column, calculated by the defined metric.

Results

Looking at Figure 4 it can be assumed that the most promising stations for new cafes are in the city itself and the near suburbs. Therefor we limit the location search to stations within this area.

Utilizing the calculated suitability index, it is now possible to highlight promising stations on the map considering the occurring competition and the high customer sites. This can be done by using the index to specify the radius of the circle depicting a station on the map. Further an indicator is introduced to mark the most promising stations. The most promising ten stations are shown in green the second most promising ten stations are shown in orange and the remaining uninteresting stations will be colored red.

The support the visualization the map color is set to dark, so that the markers appear more visible.

Figure 6: Map with the stations colorized by the suitability index.

The ten most promising stations for opening a new café are:

	Name	Long	Lat	Lines count	Distance	Comp	Cust	Suitability
54	Universität	6.9312003766	50.9259899857	1	2.475556	0.0	22.0	23.0
11	Ulrepforte	6.951437134	50.9243235691	2	1.677367	0.0	8.0	10.0
31	Deutz Technische Hochschule	6.9859264567	50.9365857842	2	1.866090	0.0	6.0	8.0
2	Poststr.	6.9500805214	50.9316916582	4	1.025377	0.0	2.0	6.0
18	Eifelwall	6.937176778	50.924873049	1	2.215972	1.0	9.0	5.0
49	Universitätsstr.	6.9242751929	50.936653862	2	2.559527	0.0	3.0	5.0
52	Arnulfstr.	6.9305212496	50.918530254	1	3.050616	0.0	4.0	5.0
51	Weißhausstr.	6.9337951463	50.921716127	1	2.633557	1.0	6.0	3.5
104	Lohsestr.	6.9551338824	50.9585057929	2	2.268666	0.0	1.0	3.0
25	Ebertplatz	6.9595370283	50.9510131543	2	1.408531	0.0	1.0	3.0

The second most promising ten stations are:

	Name	Long	Lat	Lines count	Distance	Comp	Cust	Suitability
145	Ostheim	7.042401237	50.9291592467	1	5.981872	0.0	2.0	3.0
160	Bf Mülheim	7.0128560514	50.9569283058	2	4.308951	0.0	1.0	3.0
44	Siegstr.	6.99488326	50.8830097	2	6.647034	0.0	1.0	3.0
156	Stegerwaldsiedlung	6.9933994166	50.9485947332	2	2.647463	0.0	1.0	3.0
72	Oskar-Jäger-Str./Gürtel	6.9105605258	50.9393562796	1	3.534800	0.0	2.0	3.0
208	Brühl Nord	6.9018038124	50.8351066781	1	12.221681	0.0	2.0	3.0
41	Schönhauser Str.	6.9717962172	50.9145492339	2	2.781786	0.0	1.0	3.0
50	Melaten	6.917048536	50.936734511	2	3.074505	0.0	1.0	3.0
36	Pohligstr.	6.9415800881	50.9166833148	1	2.747891	0.0	2.0	3.0
102	Neusser Str./Gürtel	6.9513443152	50.972558269	2	3.855845	1.0	4.0	3.0

Discussion

According to the Foursquare data and the definition of the suitability index – depending on high customer sites and existing competition near the station - the most promising station for opening a café near by would be the station “Universität” – university.

The data for this station shows that it has one line stopping there, it has 22 sites of potentially high customers and zero competition.

This finding should be challenged thoroughly. It can be expected, that a university site belongs to the ones with the most potential customer available at all. But it must be questioned of this justifies it to weight it with a factor of 22. It must be assumed that within these findings are several doubles and that most entries are not defining the university itself but the single faculties or even buildings. Further it must be questioned of there is really no competition at this station momentarily. It must be checked if the Foursquare location data is sufficient and perhaps if there is a actually the

possibility to open a café. It could be, that the station is within a park area and no buildings are allowed in the vicinity.

A good amount of handywork seems therefore appropriate to sort through the results to determine their overall quality. Acquiring and using additional data on the sum of potential customers at the different sites would certainly help to improve the results substantially.

Conclusion

The project has shown, that Foursqure data can be used to determine interesting locations for opening new stores for a café chain.

However, it is important to verify your data source diligently. In the case of the station data it can be easily done be formulating some expectations on how the data should look like and then print it on a map for inspection. In the case of Foursquare data this can be done by comparing the company's data with a second source – for example like Google Maps. A very quick comparison of the two data sources has then clearly shown, that the Google data on cafes and bakeries in Cologne is much more extensive than the Foursquare data. But the reason for this deviation is not part of this paper and must be determined in another paper.