IBM Applied Data Science Capstone

Investigating the Optimal New Café Location in Zurich

Hakki Karaman

22.06.2020

Introduction: Business Problem

This project aims to find the best location to open a new cafe at the center of Zurich (District 1-6). **The business problem is finding the optimal district for the prospective café by finding out the district with the lowest number of cafes per capita.**

Many similar studies were done in past to find out the optimal location for a new business, usually the frequency of the cafes/ restaurants/ shops and the distance to the center were taken into consideration however the population of the districts are neglected. The frequency of restaurants gives idea of the supply and the number of people living in the district is essential to find out the demand. **For each district, the key parameter restaurant per capita can be calculated with the number of cafes in the district divided by the population of the district.**

Data

To be compare the number of restaurant per capita for each district we need to main source of data:

- **Population of each district:** District populations of Zurich can be find in Wikipedia page [1]

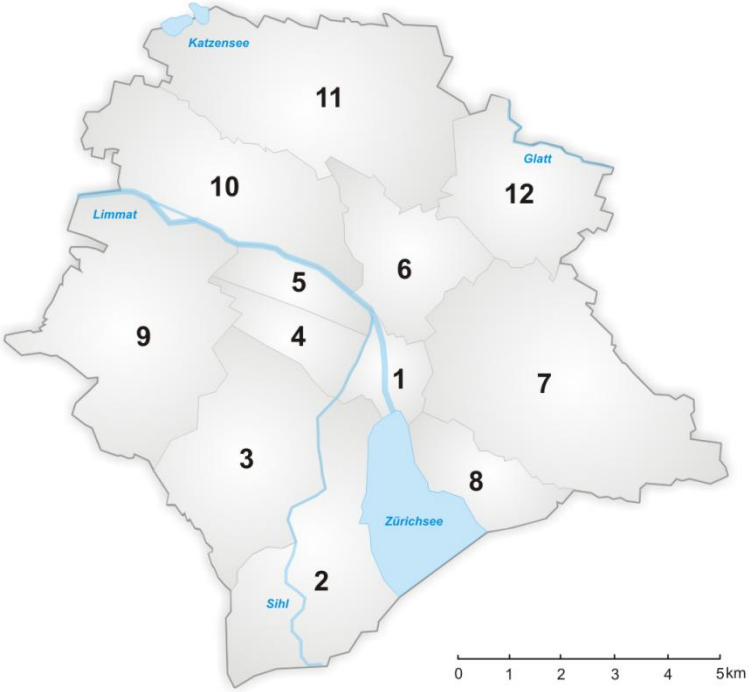

Figure 1: The map of the districts of Zurich [1]

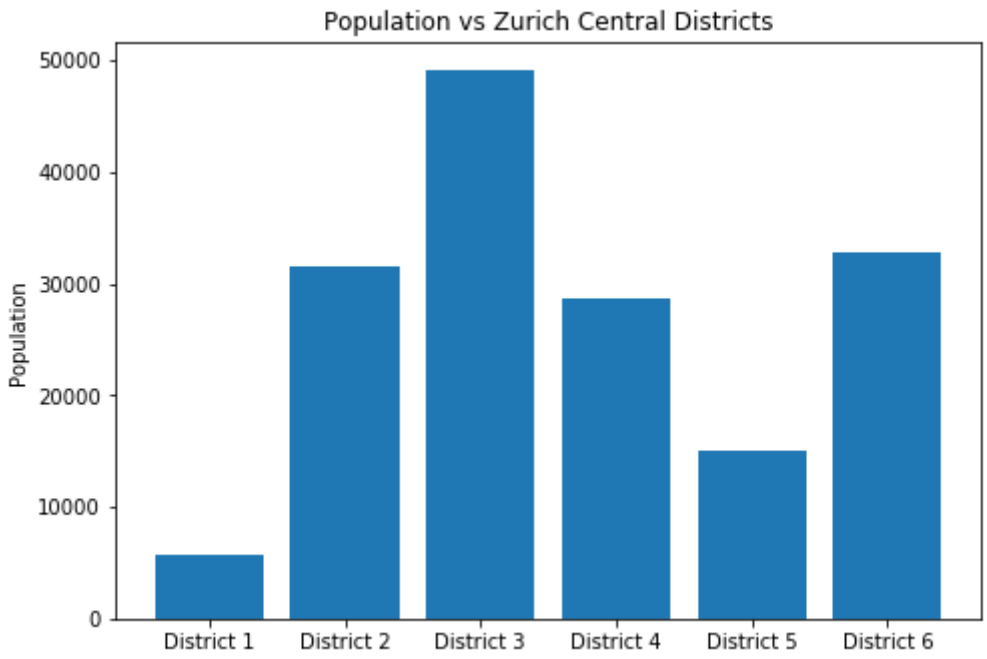

Figure 2: Population of Zurich District 1-6

- **All the location information about the cafes in Zurich can be found in Foursquare.** Foursquare developer account [2] has the data of cafes, it will be necessary to process the data and to figure out how many cafes there are in each district: The Foursquare data have 19 features. Let's have a look the features related to the location.

Table I: The Foursquare Relevant Data about the cafes in Zurich (the first 10 raw)

	name	lat	lng	distance	postalCode	formattedAddress	id
0	Café Oscar	47.377776	8.539513	37	8001	[Bahnhofplatz 15 (Südtrakt), 8001 Zürich, Schw...	53a57c62498e34ef9a4f6d69
1	Gran Café Motta	47.372346	8.542754	689	8001	[Limmatquai 66 (Rathaus), 8001 Zürich, Schweiz]	4be416f1cf200f472e72113c
2	Café Schober	47.371400	8.544149	828	8001	[Napfgasse 4, 8001 Zürich, Schweiz]	4b058888f964a52009cc22e3
3	Café Wühre	47.370761	8.542010	841	NaN	[Wuhre 11, Zürich, Schweiz]	4b9cac05f964a520997536e3
4	Café du Bonheur	47.379697	8.516199	1755	8004	[Zypressenstrasse 115, 8004 Zürich, Schweiz]	51af8be6498ea309db82eb84
5	Café Gourmet	47.377094	8.539425	112	8001	[Bahnhofplatz 7, 8001 Zürich, Schweiz]	4b8e149ff964a520c81733e3
6	A Confectioner's Café	47.378153	8.539920	42	8001	[Migros, Bahnhofstrasse 21. 8001 Zurich, Switz...	572c8b3bcd10a820361648f8
7	Le Cafe	47.377857	8.540275	73	8001	[Hauptbahnhof, 8001 Zürich, Schweiz]	4bc8acc5ab62d13a90c317d4
8	Restaurant Cafe Boy	47.377926	8.516215	1745	8004	[Kochstr. 2, 8004 Zürich, Schweiz]	4bb3353c4019a5935c9e37b8
9	Grande Café & Bar	47.375479	8.543395	421	8001	[Limmatquai 118, 8001 Zürich, Schweiz]	504a0c0f19a9a1bc79aae89a

In Figure 3, the locations of cafes which were found by using Foursquare developer account are visalized over Zurich map thanks to Folium library of Python.

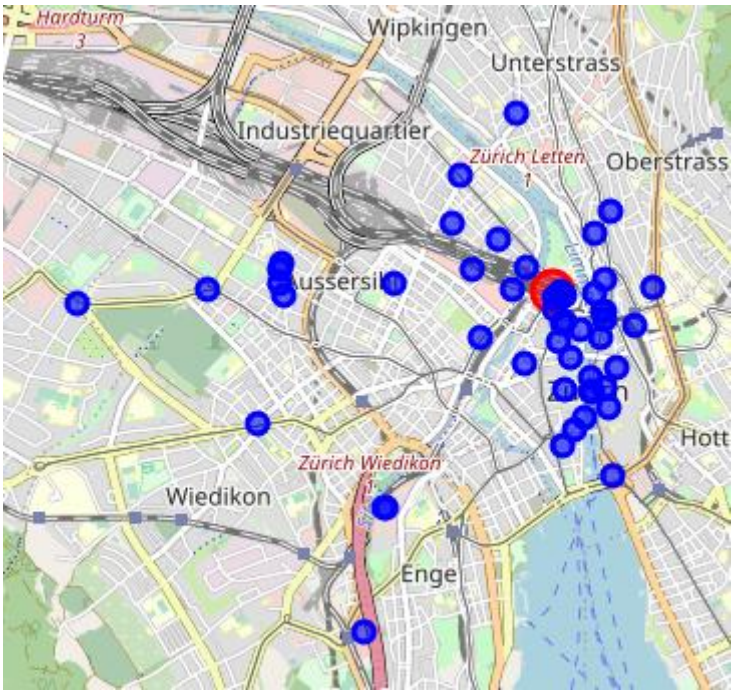

Figure 3: Visualization of Cafes in Zurich

Methodology

For each district (*district_i* where *i= {1,2,3,4,5,6}*) the number of cafes per capita can be found by following equation:

$$\#cafes\ per\ capita\ for\ district\_i = \frac{\#cafe\ of\ the\ district\_i}{\#population\ of\ the\ district\_i}$$

The minimum of cafes per capita will be indicate the optimal district for a café.

$$\text{The optimal district for cafe} = \underset{dist\_i}{\text{argmin}}(\#cafes\ per\ capita\ for\ district\_i)$$

As we have already had the population information for each district, now we need to find out the number of cafes for each district. The first idea would be finding out the districts based on latitude & longitude information of the cafes and finding out the districts thanks to GeoPy and Folium libraries of Python. However, there is an easier way for someone who is familiar with the postal code of Zurich. **The last 2 digits of Zurich postal code represents the district, for example if the postal code is 8005, it is the 5th district.**

The number of cafes in each district can be easily counted by using Pandas library of Python, can be seen in Table II and visualized in Figure 4.

Table II: Classification and counting of cafes with the postal code

postalCode	
8001	16
8002	2
8003	3
8004	4
8005	3
8006	2

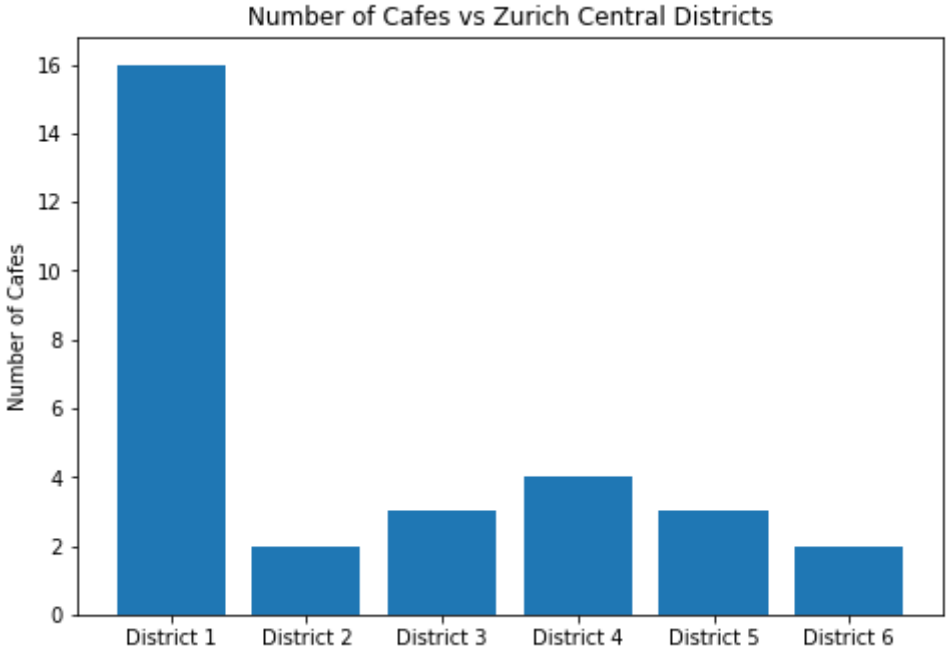

Figure 4: Number of cafes in Zurich District 1-6

District 1 has much more cafes compare to the other districts. The details will be discussed in the next section.

Results and Discussion

The next task is to calculate the number of cafes per capita. For each district the number of cafes is presented in Figure 4 and the population can be seen in Figure 2. Their ratio is easily calculated and can be seen in Figure 5. In Figure 5, it can be observed District 1 has order of magnitude higher value. Although, District 1 is smallest one, it contains both the historical center and the business center of the city. So, there are a lot of offices and stores for business, tourism, shopping and not many residential buildings. So, its population lowest but it has the highest number of cafes.

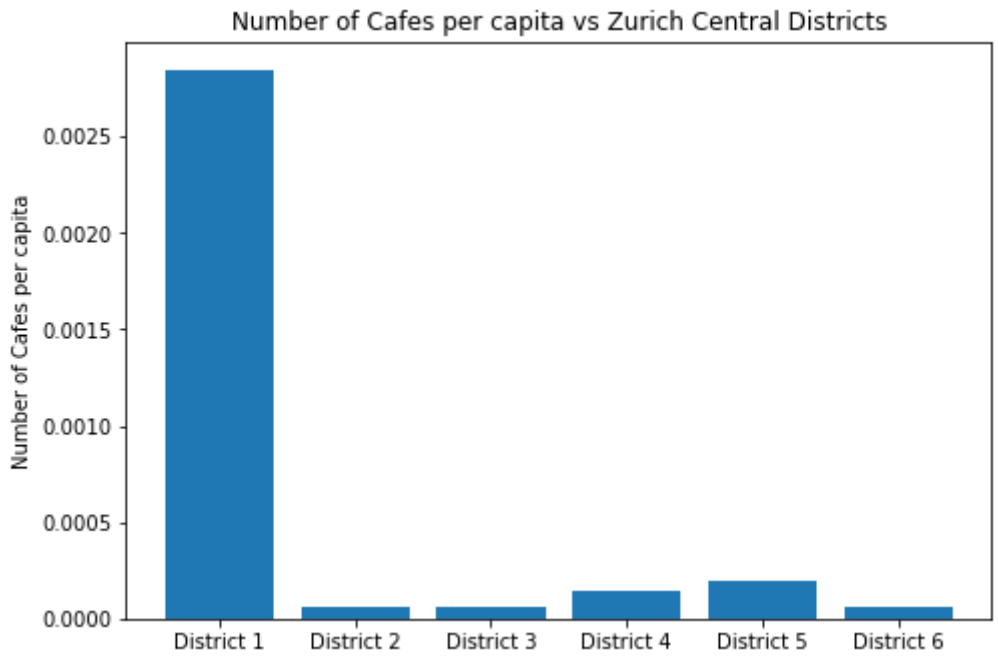

Figure 5: Number of cafes per capita in Zurich District 1-6

We are looking for the district with the minimum value of number of cafes per capita. **District 6 has the minimum numerical value so it is the optimal district to open the café.** District 2 and District 3 are also very close to minimum.

Conclusion

This study was done to find out the optimal district to open up a new café by using available data. The characteristic parameter is defined as number of cafes per number of residents in the district. The population of the districts was found in Wikipedia and the number of cafes per district was found by practically using the postal code of the cafes from Foursquare data.

According to the analysis, **District 6 is the optimal place for a new café as it has the minimum value for number of cafes per capita.** It is important to mention that District 2 and District 3 are also very close to the minimum value and the difference is not significant. As a future work one can consider to develop more sophisticated model to compare District 2-3-6 with further data such as the rent prices, the available locations with their location of the district, the flow of non-resident people etc.

References

1. https://en.wikipedia.org/wiki/Z%C3%BCrich#Population
2. https://developer.foursquare.com/