

Ideal Business Office Sites in Zurich, Switzerland

Capstone Project

Toward the IBM Data Science Professional Certificate

Applied Data Science Capstone by IBM/Coursera

Submitted by
Ting Jin

Zürich, on 12.02.2020

Table of Contents

TABLE OF CONTENTS II

1 INTRODUCTION..... 1

1.1 Problem Description 1

1.2 Background Discussion 1

2 METHODOLOGY 1

3 DATA PREPARATION AND TOOLS EXPLANATION 2

3.1 GeoNames Geographical Database Exploration 2

3.2 Data Acquisition with Foursquare API..... 3

3.3 Map Visualization 3

4 ANALYSIS AND DISCUSSION 4

4.1 General Geographical Analysis..... 4

4.2 Detailed Analysis of each Factor..... 5

4.2.1 Number of existing restaurants nearby 5

4.2.2 Existence of gym or sports center nearby..... 6

4.2.3 Convenience to take public transportation 8

4.3 Aggregated Analysis..... 9

5 RESULTS11

6 CONCLUSION12

LIST OF REFERENCES.....12

1 Introduction

1.1 Problem Description

In this project we will try to find an ideal location for startups who want to locate at Zurich, Switzerland. The target stakeholders are companies/contractors who are interested in starting their own business in Zurich, Switzerland with the special consideration to attract qualified employees.

1.2 Background Discussion

The selection of a company's location is complex in nature since it is influenced by many factors, such capital investment, government policies/regulations, labor, transportation, and influence of competitors. This project will focus on geographical factors which influences qualified job seekers' choices while applying jobs. For instance, number of restaurant/coffee-bar/supermarket, distance to public transportation.

In this project, we are going to use location data to explore Zurich and generate a few most promising neighborhoods for the target stockholders to select an ideal business office. The reason why those neighborhoods are the ideal location will be clarified within the data analysis process.

2 Methodology

Based on the problem definition of this project, factors that will influence the business office selection are:

1. Number of existing restaurants/coffee-bar/supermarkets nearby
2. If there is a gym or sport center nearby
3. Is it convenient to take public transportation (By comparing the number of public transport stations nearby)

According to the GeoNames Geographical Database (The GeoNames Geographical Database 2019), basic geographical data can be obtained and transformed to data frame by using the library pandas. Folium will be used to visualize maps and the latitude and longitude values will be acquired by the using of geopy library. Of course, the Foursquare API will be used to explore the Zurich region and segment them. In the end, the ideal business office sites in Zurich for stockholders will be presented both descriptively and graphically.

3 Data Preparation and Tools Explanation

In order to compare the number of existing food providers such as restaurant and coffee shop, to check if there is a gym or sport center nearby, to find out if it is convenient to take public transportation such as bus, tram, and train by comparing the total number of transport stations, all these data have to be acquired and prepared so that one can analyze and present the results to stakeholders. Based on the methodology, which is described in the previous chapter, the data preparation processes are composed by several steps with the use of various data science tools.

3.1 GeoNames Geographical Database Exploration

According to the GeoNames Geographical Database, GeoNames of Switzerland, namely, basic geographical data can be obtained and transformed to data frame by using the library pandas. Since the geographical data for Switzerland in this data base includes data from all the cantons, we must select only the canton Zurich first as this is the target market of this project. Secondly, except the necessary features such as Borough, Neighborhood, Latitude, Longitude, the extracted data also contains country code, city code and other information. We select only the necessary data which is useful for further analysis and visualization. Third, when check the data closely, some duplicated information exists with some features. Based on the author's understanding and deep checking, some duplications are removed, and some are assigned with discriminate names.

	0	1	2	3	4	5	6	7	8	9	10	11
0	CH	5000	Aarau	Kanton Aargau	AG	Bezirk Aarau	1901	Aarau	4001	47.3925	8.0442	4.0
1	CH	5001	Aarau 1	Kanton Aargau	AG	Bezirk Aarau	1901	Aarau	4001	47.3888	8.0483	NaN
2	CH	5004	Aarau	Kanton Aargau	AG	Bezirk Aarau	1901	Aarau	4001	47.3925	8.0442	4.0
3	CH	5017	Barmelweid	Kanton Aargau	AG	Bezirk Aarau	1901	Erlinsbach (AG)	4005	47.4159	7.9764	4.0
4	CH	5018	Erlinsbach	Kanton Aargau	AG	Bezirk Aarau	1901	Erlinsbach (AG)	4005	47.4052	8.0151	4.0

Table 1: Original Geographical Data in Switzerland

After the above data cleaning, one can see that the original data obtained from the GeoNames database in Table 1 with more than 4000 rows changed to the relevant data in Table 2 with less than 300 rows for our analysis in this project. The relevant data contains only information that is important for the analysis in this project and focuses on the Zurich canton. Map visualization and further using with Foursquare API to explore Zurich can be done with the combination of this data to obtain restaurant, gym, and public transportation related data.

	PostCode	Borough	Neighborhood	Latitude	Longitude
258	8902	Bezirk Dietikon	Urdorf2	47.3851	8.4258
259	8903	Bezirk Dietikon	Birmensdorf ZH	47.3552	8.4426
260	8904	Bezirk Dietikon	Aesch ZH	47.3367	8.4410
261	8951	Bezirk Dietikon	Fahrweid	47.4117	8.4156
262	8952	Bezirk Dietikon	Schlieren	47.3967	8.4476
263	8953	Bezirk Dietikon	Dietikon	47.4017	8.4001
264	8954	Bezirk Dietikon	Geroldswil	47.4221	8.4108
265	8955	Bezirk Dietikon	Oetwil an der Limmat	47.4283	8.3949
266	8000	Bezirk Zürich	Zürich5	47.3828	8.5307
267	8001	Bezirk Zürich	Zürich6	47.3667	8.5500

Table 2: Cleaned Geographical Data in Zurich

3.2 Data Acquisition with Foursquare API

Based on the data gathered from GeoNames database and cleaned in Table 2, we can apply it with the Foursquare API to find all the other venue-related data close to various neighborhoods. After defining the Foursquare credentials and version, we use it to obtain the top 100 venues that are in all the neighborhoods in Zurich within a radius of 1000 meters. The reasons for this based on three considerations. First of all, we need restaurants/coffee bars, gyms/sports centers, public transportation-related data around each neighborhood. Second, this project analyzes the ideal business office location, which indicates that the closer the venues to the ideal place the better. Thus, we select the radius 1000 around the neighborhood, which means that all the venues can be reached with 15 minutes by food and we only considered this as convenient. This distance means attractive to employees in this project. Third, considering that we have a limit to use the Foursquare API and to accelerate the speed, we select only the top 100 venues of each neighborhood to analyze.

After performing the above-mentioned method, over 2000 rows of venues are found around all the neighborhoods in Zurich under 261 venue categories. Since the table is too large and cannot show here easily, this report does not show the original venues gathered for each neighborhood. One can find detailed data information in the relevant notebook. In the next chapter of detailed analysis aims at each influence factor, we will find the top location choices for each element separately and aggregately with clustering analysis.

3.3 Map Visualization

In order to show the results visually and let stakeholders have a clear idea which is easy to understand, in addition, that the data analyzed in this project is geographically data, all the

4 Analysis and Discussion

4.1 General Geographical Analysis

[illegible]

Figure 2: Map of Canton Zurich Presented by Neighborhoods

4.2 Detailed Analysis of each Factor

After gaining a general geographical understanding of canton Zurich, the following analysis will focus on each of the factors, namely, the number of existing restaurants nearby, if there is a gym or sports center nearby, and is it convenient to take public transportation to find out the ideal choices while considering those factors separately and comprehensively.

Based on data acquisition with Foursquare API in the previous chapter, we found that there are around 260 kinds of venue categories with more than 2240 venues in total that are found for all the neighborhoods in Zurich. Afterward, rows are grouped by neighborhood and by taking the total frequency of each category occurrence since the location selection uses neighborhood as a unit and the total number of different venues determines the selection. After doing this, we target 259 neighborhoods with 261 venue categories. In the following analysis, we have to select all the venues relevant to each factor and find the total occurrence of all the venues for each factor for further selection and clustering.

4.2.1 Number of existing restaurants nearby

Let's first analyze the factor number of existing restaurants nearby. Since our focus is what people care about while looking for jobs such as the number of restaurant/supermarket-related venues close to the ideal business office, which companies also care in order to attract talented employees. Therefore, finding out how many relevant categories can be captured from all the returned venues for each neighborhood which relevant to a restaurant that employees can solve their breakfast, lunch, or rest needs is important.

	PostCode	Neighborhood	Latitude	Longitude	Food
0	8000	Zürich5	47.3828	8.5307	44
1	8001	Zürich6	47.3667	8.5500	40
2	8400	Winterthur	47.5056	8.7241	37
3	8401	Winterthur2	47.4967	8.7342	35
4	8152	Glattbrugg	47.4313	8.5627	24
5	8152	Glattpark (Opfikon)	47.4315	8.5693	22
6	8304	Wallisellen	47.4150	8.5967	22
7	8305	Dietlikon	47.4183	8.6188	15
8	8058	Zürich	47.4580	8.5853	13
9	8800	Thalwil	47.2918	8.5635	13

Table 3: Top 10 Neighborhoods with the Most Restaurants Nearby

While checking all the venues clearly, one can see that all the categories contain words such as restaurant, breakfast, cafeteria, café or coffee, and supermarket are all relevant to

To cluster neighborhoods based on the number of restaurants nearby, this project first runs k-means to cluster the neighborhood into 4 clusters. The map visualization of the resulting clusters can be found in Figure 2. As it is shown in the figure that the red colors are the worst choices in terms of attracting qualified laborers with sufficient food choices nearby compared to all the others. The purple points are better than others. This visualization map shows the selection of business office locations based on the number of restaurants nearby at a glance and can be presented to stakeholders for a clear understanding.

When it comes to if there is gym or sports center nearby, we find that all the 'Dance Studio', 'Sports Club', 'Tennis Court', 'Tennis Stadium', 'Yoga Studio', 'Rock Climbing Spot' categories are relevant to this factor. Since there are also sports shops which contain the word

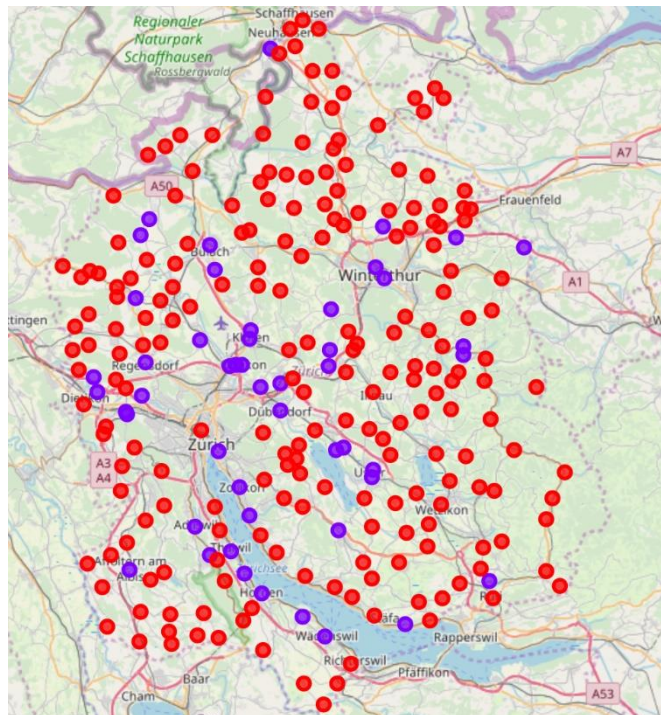
sport but not related to the factor, we must select all the columns with all the categories but not filter categories which include sport. One special case is that all the categories contain 'gym' is relevant to this influence factor. After filtering and selecting all categories related to gym or sports center and merging all the gym-related categories with the geographical data in Zurich such as latitude, longitude based on the neighborhood value, 49 neighborhoods are found with at least one sport place nearby. The top 10 neighborhoods with most gyms close to them can be found in the following Table 4.

	PostCode	Neighborhood	Latitude	Longitude	Gym
0	8613	Uster 3	47.3532	8.7216	3
1	8400	Winterthur	47.5056	8.7241	3
2	8800	Thalwil	47.2918	8.5635	2
3	8304	Wallisellen	47.4150	8.5967	2
4	8010	Zürich3	47.3950	8.4488	2
5	8952	Schlieren	47.3967	8.4476	2
6	8058	Zürich	47.4580	8.5853	2
7	8152	Glattbrugg	47.4313	8.5627	2
8	8184	Bachenbülach	47.5032	8.5456	2
9	8152	Glattpark (Opfikon)	47.4315	8.5693	2

Table 4: Example of 10 Neighborhoods with Gyms Nearby

Figure 3: Clustering and Visualizing (at least one gym nearby)

Figure 3 on the right shows the clustering and visualizing based on if there is at least one gym or sports center near each neighborhood. All the purple points represented neighborhoods with at least one gym nearby. Namely, all the red points are not good choices for stakeholders to select a business office according to this factor since there is no gym close to these neighborhoods. While comparing to Figure 2, the clustering based on the number of restaurants nearby, the distribution of recommended neighborhoods for stakeholders is quite different.



4.2.3 Convenience to take public transportation

Finally, the analysis comes to the last influence factor, namely the convenience to take public transportation, in this project for stakeholders to find an ideal business office location in canton Zurich. This project measures the factor by checking if there is a public transport station within 1000 meters radius of the location. If there is a public station nearby, we consider it convenient. Of course, the more stations nearby, the more convenient for employees to go and back from office as an attractive factor.

Comparing to the previous two factors, the number of restaurants nearby and the existence of a gym close to, it is much easier to find public station related venue categories. There are only 4 categories 'Bus Station', 'Light Rail Station', 'Train Station', 'Tram Station' relevant. Similarly, after selecting and merging the sum of transportation-related data with the geographical data in Zurich, this project finds that there are 133 neighborhoods with at least one public transportation nearby and 47 neighborhoods with at least two public transportations nearby. Moreover, the top 10 neighborhoods with the most public transportation nearby are shown below in Table 5.

	PostCode	Neighborhood	Latitude	Longitude	Transport
0	8152	Glattpark (Opfikon)	47.4315	8.5693	7
1	8152	Glattbrugg	47.4313	8.5627	6
2	8302	Kloten	47.4515	8.5849	4
3	8607	Aathal-Seegräben	47.3351	8.7700	3
4	8121	Benglen	47.3608	8.6369	3
5	8135	Langnau am Albis	47.2889	8.5411	3
6	8624	Grüt (Gossau ZH)	47.3115	8.7834	3
7	8166	Niederweningen	47.5061	8.3771	3
8	8127	Forch	47.3252	8.6434	3
9	8803	Rüschlikon	47.3069	8.5514	3

Table 5: Top 10 Neighborhoods with most Public Transportation Nearby

For a clear understanding, we group the neighborhoods into two clusters and the visualization map in Figure 4 shows the clustering results. The purple points represent the 47 neighborhoods that have at least 2 public transport stations within 1000 meters radio and the red points represent all the others. Therefore, it is clear that the purple neighborhoods are the recommended neighborhoods that should be shown to stakeholders who care about this factor.

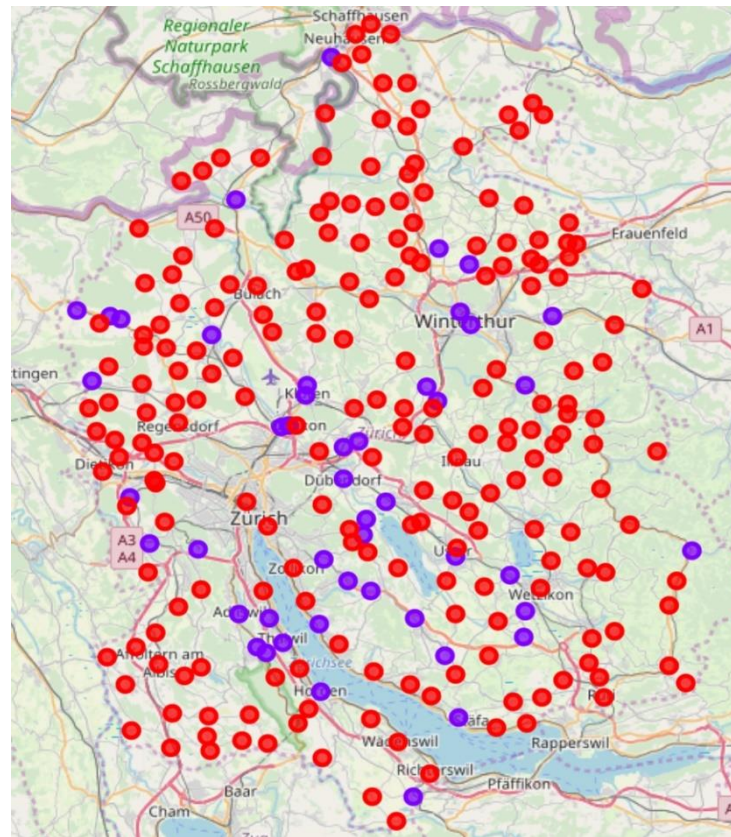


Figure 4: Visualization of 2 Clusters based on Public Transportation

While comparing the three visualization maps clustered for each factor, one can see that most of the preferable recommendations for stakeholders are different based on different factors. However, this project considers all three factors. How should we decide which neighborhoods should be presented to stakeholders as the results? In the next section, the aggregated analysis will be conducted with all influence factors together. Afterwards, the results will be found.

4.3 Aggregated Analysis

Based on the above analysis aim at each of the three criteria that we used at this project to select the ideal office location for business, the three tables show the top 10 choices of each consideration can be presented to stakeholders if they consider only one of the factors. Accordingly depends on stakeholders' preference, the aggregated analysis of the three factors can be adjusted according to stakeholders' specific demands. In this project, let's first analyze all neighborhoods that have at least one restaurant, one gym, and one public transport station nearby to satisfy the basic demands without the consideration of the amount. After merging and selecting the data to satisfy the basic demands, we get the following Table 6 as an example with actually 31 rows/neighborhoods for further analysis.

	PostCode	Neighborhood	Latitude	Longitude	Food	Gym	Transport
0	8134	Adliswil	47.3100	8.5246	5	1	2
1	8910	Affoltern am Albis	47.2774	8.4513	4	1	1
2	8804	Au ZH	47.2418	8.6441	4	1	1
3	8180	Bülach	47.5220	8.5405	5	1	1
4	8157	Dielsdorf	47.4815	8.4585	4	1	1

Table 6: Aggregated Data for Further Analysis

The 31 neighborhoods we obtained so far all satisfy the basic needs of this project. However, since some factors like restaurants are the more the better, which also applies to the other two factors just not as important as the first one. In order to narrow down the selection from 31 neighborhoods to a smaller number, this project groups the 31 choices to 3 clusters. The results show that 3 neighborhoods are considered as the best choices while considering the three factors in this project together for stakeholders who want to find a business office in Zurich. Another 14 neighborhoods are considered as equally second good. The details of those neighborhoods can be found below in Table 6. The value of the 'Cluster Labels' equals 1 means the best neighborhood and equals 2 means the second-best neighborhood.

	PostCode	Neighborhood	Latitude	Longitude	Food	Gym	Transport	Cluster Labels
0	8001	Zürich6	47.3667	8.5500	40	1	1	1
1	8400	Winterthur	47.5056	8.7241	37	3	2	1
2	8401	Winterthur2	47.4967	8.7342	35	2	2	1
3	8152	Glattbrugg	47.4313	8.5627	24	2	6	2
4	8152	Glattpark (Opfikon)	47.4315	8.5693	22	2	7	2
5	8304	Wallisellen	47.4150	8.5967	22	2	1	2
6	8305	Dietlikon	47.4183	8.6188	15	1	3	2
7	8800	Thalwil	47.2918	8.5635	13	2	2	2
8	8058	Zürich	47.4580	8.5853	13	2	2	2
9	8952	Schlieren	47.3967	8.4476	12	2	1	2
10	8010	Zürich3	47.3950	8.4488	12	2	1	2
11	8302	Kloten	47.4515	8.5849	11	1	4	2
12	8610	Uster	47.3471	8.7209	11	1	2	2
13	8600	Dübendorf	47.3972	8.6187	10	1	2	2
14	8613	Uster 3	47.3532	8.7216	10	3	1	2
15	8105	Regensdorf	47.4341	8.4687	10	1	1	2
16	8810	Horgen	47.2598	8.5978	9	2	3	2

Table 6: Results for Stakeholders

The following map visualization of the results shows the three clusters within all the possible 31 neighborhoods. One can see from the map clearly that the neighborhoods presented by the purple points are the most favorable ones. The turquoise points indicate the second

good choices. The red neighborhoods are still good choices. However, compared to the others, those ones may not take into primary consideration for this project.

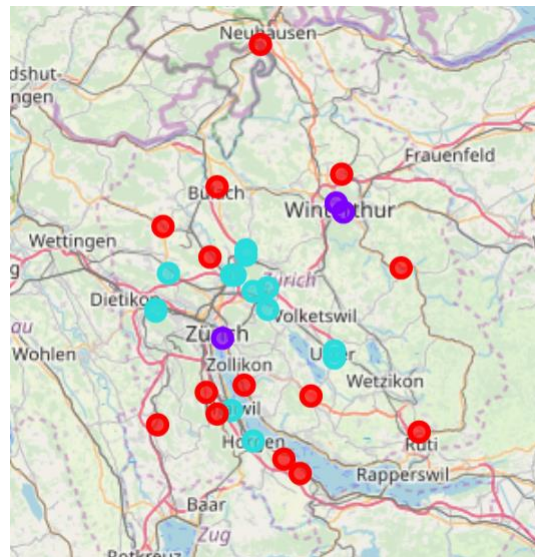


Figure 5: Map Visualization of the Final Results

5 Results

This project shows that there are more than 300 neighborhoods in Zurich can be selected by stakeholders as a business office location. However, many of them are less suitable for companies who want to select their business office close to a place where has many restaurants, at least one sports place and public transport station nearby as preconditions to attract qualified employees. After analyzing each of the three factors which indicated at the beginning of this project, the results show that the preferably business office locations are quite different when considering each factor separately or aggregated together.

The findings in this project can be presented to stakeholders with different important consideration of the three factors. The adjustment can be made accordingly based on specific stakeholders' demands.

In summary, the top ten neighborhoods of each factor can be presented, and the relevant cluster maps can be found in this project. By selecting only neighborhoods that contain at least one restaurant, one gym, and one public transportation station within a 1km radius, this project found 31 neighborhoods in Zurich can be selected to attract employees. Moreover, after taking the total amount those venues into consideration by clustering the 31 neighborhoods into 3 clusters, this project further targets 17 neighborhoods and 3 of them can be considered as priorities when other conditions are equal.

The purpose of this project is to provide statistical information for stakeholders who want to find an ideal business office location in Zurich where is attractive to employees. The Table 6 can be shown to stakeholders as final results. Of course, those recommended neighborhoods should be considered as a starting point for more detailed analysis for stakeholders who take the three factors contained in this project as a priority and many other factors need to be considered.

6 Conclusion

Based on the use of GeoNames Geographical Database, Foursquare API and some machine learning techniques with various libraries such as Pandas, Folium, Scikit-learn, we analyzed the Ideal Business Office Sites in Zurich based on three factors. Namely, the number of existing restaurants/coffee-bar/supermarkets nearby, if there is a gym or sports center nearby, and is it convenient to take public transportation. According to the analysis processes, we narrowed down the ideal business location in Zurich from the original almost 300 to around 30 neighborhoods. Moreover, by clustering all the possible ones, we further target a smaller portion of 3 of the best possible business office locations and 14 second-good ones to present to the relevant stakeholders.

The final decision on optimal business office locations can be made by stakeholders based on other specific considerations of neighborhoods in every recommended zone, taking into additional factors such as fixed costs like office renting price, closeness to park, levels of noise, and other social as well as economic considerations. It is important to point out that this project is based on the three factors we analyzed. There are also many other factors and ways to analysis and the results may vary based on distinct criteria and analyzed at different times. The market is dynamic and the importance of factors to various stakeholders is different.

List of References

The GeoNames Geographical Database (2019): GeoNames webservice and data download. Available online at <https://www.geonames.org/export/>, updated on 5/19/2019, checked on 2/4/2020.