Trending Business Ventures in Zurich, Switzerland

Eldhose Poulose

June 10, 2021

1 Introduction

Switzerland is vibrant and international. Even though migrating to Switzerland is not an easy task, it's still every ones dream to live and explore different parts of Switzerland. The aim of this capstone project is to demonstrate how Foursquare API can be used to explore and compare cities in Canton Zurich, Switzerland. This project will help people in exploring better facilities around their neighborhood in Zurich. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Zurich, Switzerland.

1.1 Background

From the lectures, we have seen how Foursquare location data can be used to extract information. This project will apply the skills that are acquired from the lessons such as how to construct a URL to send a request to the API to search for a specific type of venues, to explore a particular venue, to explore a geographical location, and to get trending venues around a location. Also, this project covers the visualization library, Folium, to visualize the results.

1.2 Problem Statement

Every tourist who visit a new city always ask the same kind of questions like, "Where can I have the best food?", "What are the best venues to visit nearby?". The major purpose of this project, is to **recommend the user with top 5 Restaurants and top 10 venues** which are available nearby.

1.3 Interest of the Economy

Obviously, Tourism departments, Restaurants and other business people can use this model to increase their public audience and thereby improving their business. Also, this model provides an overview of the provinces (Bezirk) in Switzerland, which they can use for planning their business ventures. Note: due to the API request limitation, this project analyse only the City (Canton) Zurich.

2 Data

This section elaborates the dataset that is used in this project. There are lot of open datasets available which contains the information regarding the Postalcode, Latitude, Longitude and the City information of Switzerland. I used the dataset available from a Github repository which has all the information we need to explore and cluster the neighborhoods in Zurich [1]. Therefore we wrangle this data, clean it, and then filter it into a pandas dataframe so that it is in a structured format as per our requirement.

The Dataset consist of several features as explained below. We filter the entire dataset and take out only the Zurich City data, which is where state is represented as ZH. Later among the features we use the zipcode, state, Latitude, Longitude features for the analysis using Foursquare tools.

- country_code : CH means Switzerland
- zipcode : the Postal Code
- place : main city name

```
swiss_data = pd.read_csv('zipcodes.ch.csv')
#swiss_data.head()
print(swiss_data.shape)
swiss_data.sample(n=10)
```

```
(4356, 11)
```

[36]:

	country_code	zipcode	place	state	state_code	province	province_code	community	community_code	latitude	longitude
1614	CH	7075	Churwalden	Kanton Graubünden	GR	Region Plessur	1848	Churwalden	3911	46.7814	9.5438
1829	CH	2884	Montenol	Jura	JU	Porrentruy District	2603	Clos du Doubs	6808	47.3518	7.1487
1963	CH	6232	Geuensee	Kanton Luzern	LU	Sursee District	314	Geuensee	1085	47.1997	8.1069
818	CH	3654	Gunten	Canton de Berne	BE	Thun District	247	Sigriswil	938	46.7119	7.7024
1950	CH	6207	Nottwil	Kanton Luzern	LU	Sursee District	314	Nottwil	1094	47.1357	8.1371
3794	CH	1973	Nax	Canton du Valais	VS	Hérens District	2305	Mont-Noble	6090	46.2282	7.4310
3006	CH	6900	Massagno	Ticino	TI	Lugano District	2105	Massagno	5196	46.0126	8.9435
3703	CH	1976	Daillon	Canton du Valais	VS	Conthey District	2302	Conthey	6023	46.2574	7.3067
1447	CH	8865	Bilten	Kanton Glarus	GL	Glarus	800	Glarus Nord	1630	47.1500	9.0255
292	CH	9105	Schönengrund	Kanton Appenzell Ausserrhoden	AR	Bezirk Hinterland	1501	Schönengrund	3003	47.3259	9.2269

Figure 1: Structure of the Dataset

- state : State or Canton names
- state_code : Corresponding State Code
- province : District name
- province_code : Corresponding District code
- community : Local Community
- community_code : Corresponding Local Community Code
- latitude : Latitude Information
- longitude : Longitude Information

3 Methodology

This section explains the methods used in this project. To begin with, this project uses the Foursquare API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business. Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 50 and the radius parameter would be set to 500.

Furthermore, we will explore, segment, and cluster the neighborhoods in the city Zurich, Switzerland based on the postal code and Bezirk (Province) information. And also, we figure out the top 5 restaurants and top 10 venues around each province in Zurich.

3.1 Exploratory Data Analysis (EDA)

This section explains and shows the EDA strategy and plots. First of all, a histogram plot showing the number of postal codes were plotted. As a second part, the restaurants located within the radius range equal to 500 is extracted and their frequency is counted and plotted using a histogram. In canton Zurich it is clearly visible that the Zurich Town itself contains 38 restaurants followed by Winterthur town with 18 restaurants.

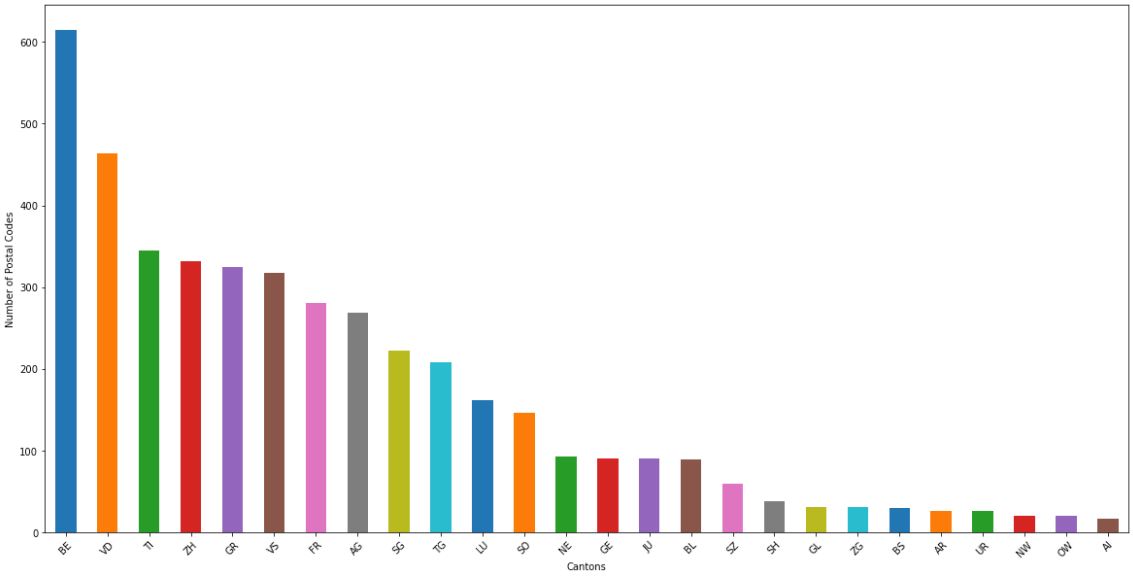

Figure 2: Distribution of Postal codes among each Canton

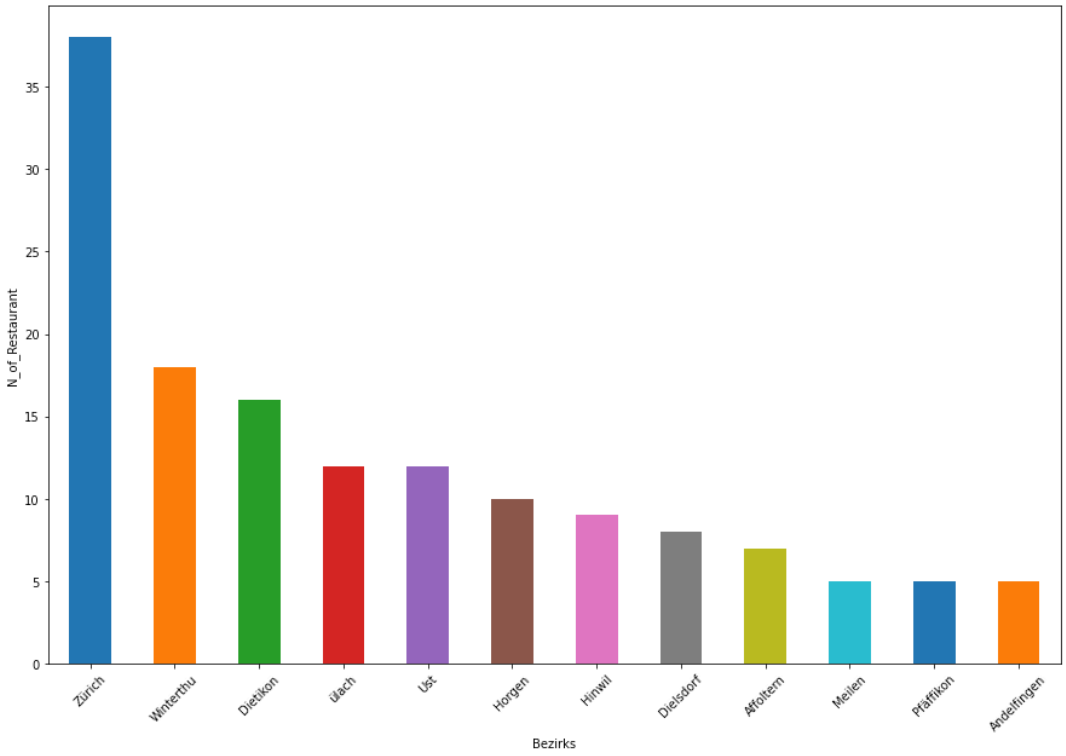

Figure 3: Distribution of Restaurants in Zurich within each Province (Bezirk) around 500m radius

3.2 Modelling

This section explains the modelling used in this project. Since the goal is to find group of ventures the K-means clustering is selected. K-Means is an unsupervised machine learning algorithm, which has been used to group similar data points together and discover underlying patterns. Since the number of data points were small a random value of k as 3 is selected.

4 Results

Based on the available data, the provinces/Bezirks in Zurich Canton is divided into 3 clusters. Among 12 Provinces Cluster 0 [See Fig: 10] contains 9 provinces where the aMost common venues are mainly Hotels. Next in the Cluster 1, [See Fig: 11] which contains only one Province which is Pfäffikon. The suggested most common venue is the Restaurant. Finally, the Cluster 2 contains 2 provinces such as Meilen and Dietikon. Interestingly the 1st most common venue for both is Supermarket. Also, these two provinces are geographically far apart. The reason for this cluster allocation is need to be studied further.

In addition to the clustering part we plotted the wordcloud which visually represent the most common venues. See Fig: 8, where the font size represent the frequency of occurrence of the corresponding venue.

Results shows that Restaurant (Swiss, Italian, Indian/Asian), Supermarket businesses are popular within the 500 m city radius in Canton Zurich.

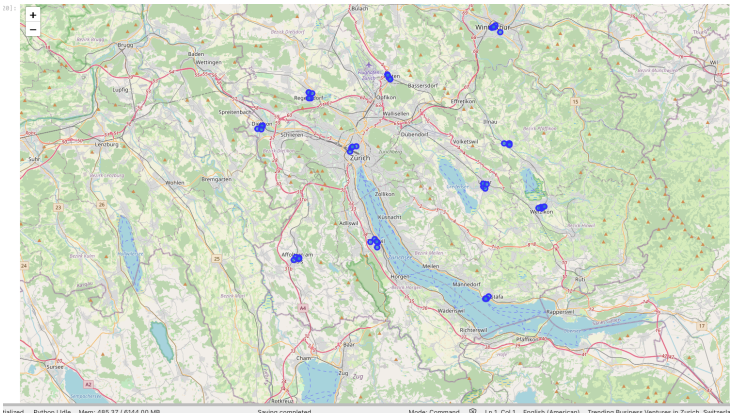

Figure 4: Map Showing the Canton Zurich with the Provinces marked in Blue dots. Within each Province Top 5 Restaurants are also marked.

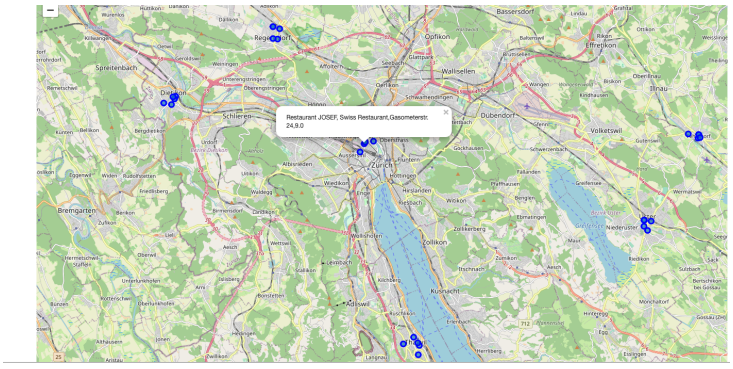

Figure 5: Each restaurant is marked with name, category, address and rating. Here for example: Restaurant JOSEF, Swiss Restaurant, Gasometerstrasse 24, 9.0

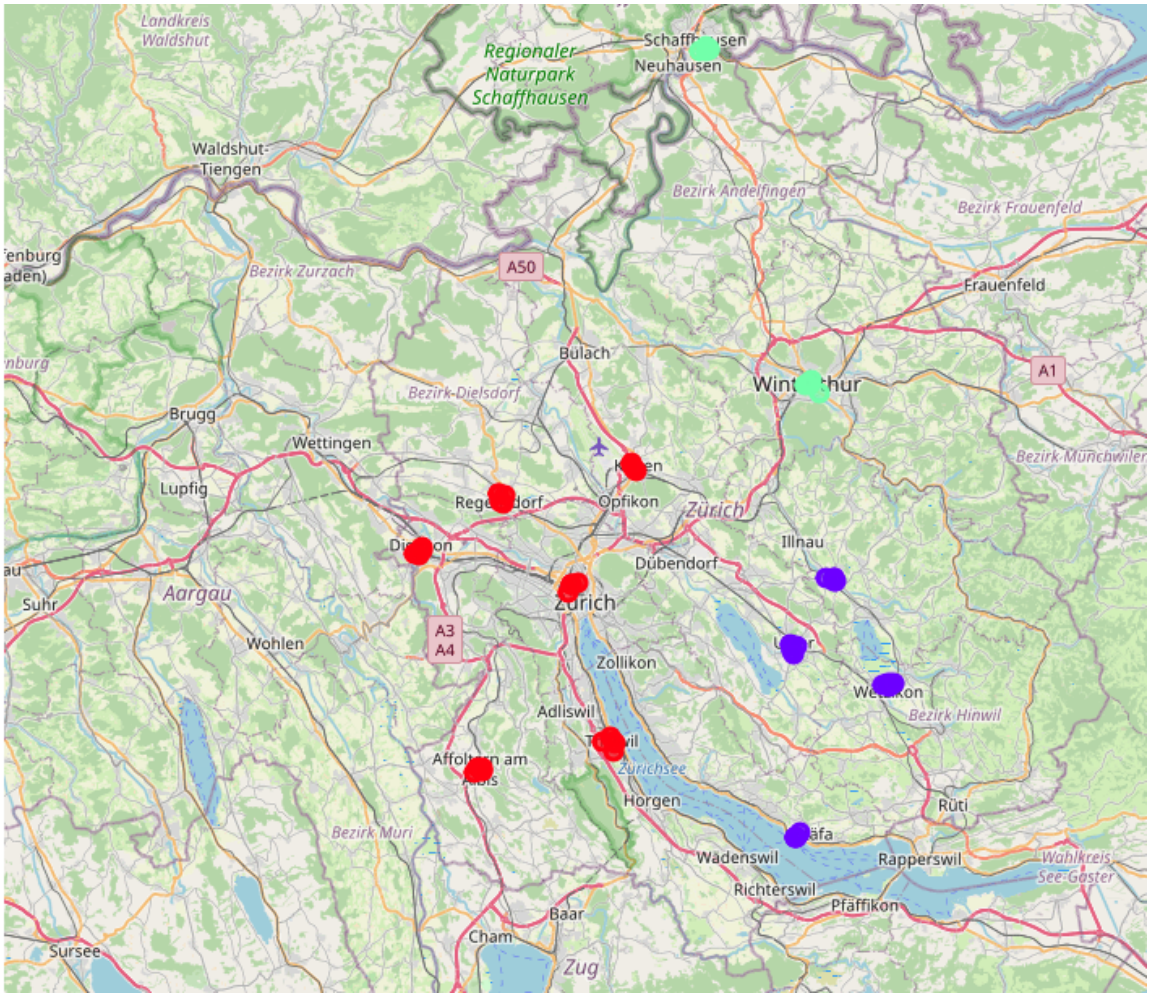

Figure 6: k-Means Clustering with k=3

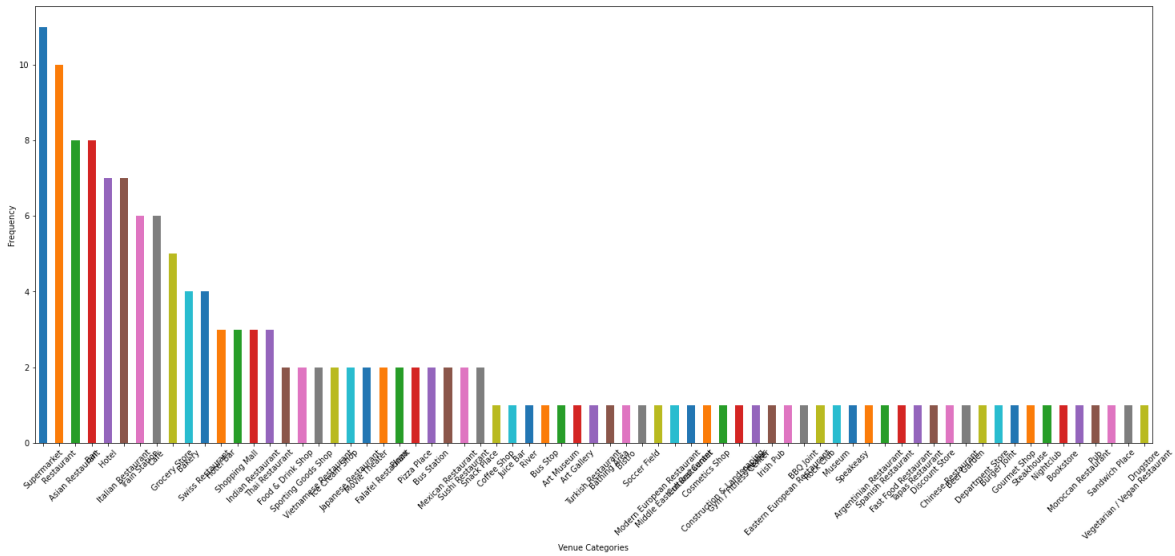

Figure 7: Analysis of each Province: The Distribution of the venue categories, LIMIT=50 for each Province and radius= 500

Figure 8: Analysis of each Province: Wordcloud showing popular venues in Zurich, LIMIT=50 for each Province and radius= 500

[22]:

	Neighborhood	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	Affoltern	Plaza	Train Station	Grocery Store	Tapas Restaurant	Supermarket	Restaurant	Eastern European Restaurant	Vietnamese Restaurant	Discount Store	Construction & Landscaping
1	Andelfingen	Train Station	Shopping Mall	Cosmetics Shop	River	Vietnamese Restaurant	Drugstore	Construction & Landscaping	Cultural Center	Department Store	Discount Store
2	Dielsdorf	Restaurant	Gym / Fitness Center	Supermarket	Hotel	Shopping Mall	Bookstore	Pizza Place	Sporting Goods Shop	Asian Restaurant	Train Station
3	Dietikon	Supermarket	Italian Restaurant	Hotel	Bus Stop	Sushi Restaurant	Drugstore	Falafel Restaurant	Cultural Center	Department Store	Discount Store
4	Hinwil	Irish Pub	Grocery Store	Shopping Mall	Movie Theater	Supermarket	Gourmet Shop	Food & Drink Shop	Fast Food Restaurant	Falafel Restaurant	Coffee Shop
5	Horgen	Hotel	Swiss Restaurant	Falafel Restaurant	Sporting Goods Shop	Bakery	Mexican Restaurant	Asian Restaurant	Grocery Store	Train Station	Gourmet Shop
6	Meilen	Supermarket	Restaurant	Train Station	Bus Station	Eastern European Restaurant	Cosmetics Shop	Cultural Center	Department Store	Discount Store	Drugstore
7	Pfäffikon	Restaurant	Food & Drink Shop	Soccer Field	Bakery	Construction & Landscaping	Drugstore	Cosmetics Shop	Cultural Center	Department Store	Discount Store
8	Ust	Snack Place	Argentinian Restaurant	Italian Restaurant	Japanese Restaurant	Grocery Store	Fast Food Restaurant	Chinese Restaurant	Ice Cream Shop	Theater	Asian Restaurant
9	Winterthu	Café	Hotel Bar	Bar	Nightclub	Restaurant	Hotel	Burger Joint	Plaza	Swiss Restaurant	Italian Restaurant
10	Zürich	Bar	Asian Restaurant	Café	Indian Restaurant	Italian Restaurant	Vietnamese Restaurant	Thai Restaurant	Museum	Moroccan Restaurant	Modern European Restaurant
11	ülach	Hotel	Restaurant	Supermarket	Steakhouse	Pizza Place	Pub	Department Store	Bus Station	Discount Store	Bakery

Figure 9: Top 10 Venues for each Province in Canton Zurich

[105]:
```
#Cluster 0
bezirk_merged.loc[bezirk_merged['Cluster Labels'] == 0, bezirk_merged.columns[[1] + list(range(2, bezirk_merged.shape[1]))]]
```

[105]:

	Latitude	Longitude	Bezirk	N_Restaurants	Cluster Labels	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	47.3828	8.5307	Zürich	38	0	Bar	Asian Restaurant	Café	Indian Restaurant	Italian Restaurant	Vietnamese Restaurant	Thai Restaurant	Museum	Moroccan Restaurant	Modern European Restaurant
1	47.4967	8.7342	Winterthu	18	0	Café	Hotel Bar	Bar	Nightclub	Restaurant	Hotel	Burger Joint	Plaza	Swiss Restaurant	Italian Restaurant
3	47.4515	8.5849	ülach	12	0	Hotel	Restaurant	Supermarket	Steakhouse	Pizza Place	Pub	Department Store	Bus Station	Discount Store	Bakery
4	47.3471	8.7209	Ust	12	0	Snack Place	Argentinian Restaurant	Italian Restaurant	Japanese Restaurant	Grocery Store	Fast Food Restaurant	Chinese Restaurant	Ice Cream Shop	Theater	Asian Restaurant
5	47.2918	8.5635	Horgen	10	0	Hotel	Swiss Restaurant	Falafel Restaurant	Sporting Goods Shop	Bakery	Mexican Restaurant	Asian Restaurant	Grocery Store	Train Station	Gourmet Shop
6	47.3264	8.7978	Hinwil	9	0	Irish Pub	Grocery Store	Shopping Mall	Movie Theater	Supermarket	Gourmet Shop	Food & Drink Shop	Fast Food Restaurant	Falafel Restaurant	Coffee Shop
7	47.4341	8.4687	Dielsdorf	8	0	Restaurant	Gym / Fitness Center	Supermarket	Hotel	Shopping Mall	Bookstore	Pizza Place	Sporting Goods Shop	Asian Restaurant	Train Station
8	47.2774	8.4513	Affoltern	7	0	Plaza	Train Station	Grocery Store	Tapas Restaurant	Supermarket	Restaurant	Eastern European Restaurant	Vietnamese Restaurant	Discount Store	Construction & Landscaping
11	47.6905	8.6436	Andelfingen	5	0	Train Station	Shopping Mall	Cosmetics Shop	River	Vietnamese Restaurant	Drugstore	Construction & Landscaping	Cultural Center	Department Store	Discount Store

Figure 10: Clustering based on the top 10 venues: Cluster 0 - Canton Zurich

[106]:
```
#Cluster 1
bezirk_merged.loc[bezirk_merged['Cluster Labels'] == 1, bezirk_merged.columns[[1] + list(range(2, bezirk_merged.shape[1]))]]
```

[106]:

	Latitude	Longitude	Bezirk	N_Restaurants	Cluster Labels	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	47.3878	8.7515	Pfäffikon	5	1	Restaurant	Food & Drink Shop	Soccer Field	Bakery	Construction & Landscaping	Drugstore	Cosmetics Shop	Cultural Center	Department Store	Discount Store

[107]:
```
#Cluster 2
bezirk_merged.loc[bezirk_merged['Cluster Labels'] == 2, bezirk_merged.columns[[1] + list(range(2, bezirk_merged.shape[1]))]]
```

[107]:

	Latitude	Longitude	Bezirk	N_Restaurants	Cluster Labels	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
2	47.4017	8.4001	Dietikon	16	2	Supermarket	Italian Restaurant	Hotel	Bus Stop	Sushi Restaurant	Drugstore	Falafel Restaurant	Cultural Center	Department Store	Discount Store
9	47.2425	8.7234	Meilen	5	2	Supermarket	Restaurant	Train Station	Bus Station	Eastern European Restaurant	Cosmetics Shop	Cultural Center	Department Store	Discount Store	Drugstore

Figure 11: Clustering based on the top 10 venues: Cluster 1 and 2 - Canton Zurich

5 Discussion and Conclusion

To summarise, this project identified a business problem, specified the data required, extracted and prepared the data, performed data analysis, and lastly provided recommendations to the stakeholders. During the project, different data science methods and tools are used to get the answer to the goal of the business problem: ”Where can I have the best food?”, ”What are the best venues to visit nearby?”. And the goal was to **recommend the user with top 5 Restaurants and top 10 venues** which are available near to the given Province. The model using K-Means clustering algorithm by combining Foursquare location data and the dataset describing the location details of Switzerland provided 3 different clusters. Other than this, the results were visualised using wordcloud. The wordcloud showed the trending venues around different cities within 500 meters. Even though the results shows that Restaurant (Swiss, Italian, Indian/Asian), Supermarket businesses are popular within the 500 m city radius in Canton Zurich a more detailed studies are suggested.

As a Future work, the datasets can be extended by adding different features from each province such as; Air Quality Index, Population Density, Area and migration data. Also, more area can be evaluated, different cantons can be considered and finally instead of Restaurant as a keyword the popular venue can be used as a keyword and apply the same analysis to it.

References

[1] Github-Account, “Reference to the dataset.” [Online]. Available: https://raw.githubusercontent.com/zauberware/postal-codes-json-xml-csv/master/data/CH/zipcodes.ch.csv