Introduction

I used to work for a delivery company in Berkeley. There we used robots to complete the deliveries. We decided to create hubs where the robots would be parked. Those hubs would help the human couriers to find the robots easier and to load them.

I could improve the location of those hubs by using the location data of foursquare. The hub's location must be near the densest restaurant areas in order to reduce the meeting time between humans and robots. So, by using k-nearest neighborhoods, I can classify the Zip areas using the number of restaurants. It will help to place the hubs and to determine the number of robots that can be placed in each one.

The audience could be any logistic manager of a delivery company. This work can help them to locate useful resources in any city based on its restaurant location.

Data:

The data used for this project are the zip codes of Berkeley and the location data of the foursquare app. The zip codes will be retrieved from the zip-codes.com page. The location data will be retrieved using de foursquare API.

The zip data would be simply the zip codes and its latitude and longitude, for example, the code 94701 with latitude 37.8718 and longitude -122.2718. On the other hand, the foursquare API would give us the number of restaurants in the area, in this case, 17 restaurants.

Methodology:

After scraping the zip codes page, I was able to download each zip code and its latitude and longitude. With this information, I created a map, at first sight, the distribution of the zip codes is uneven.

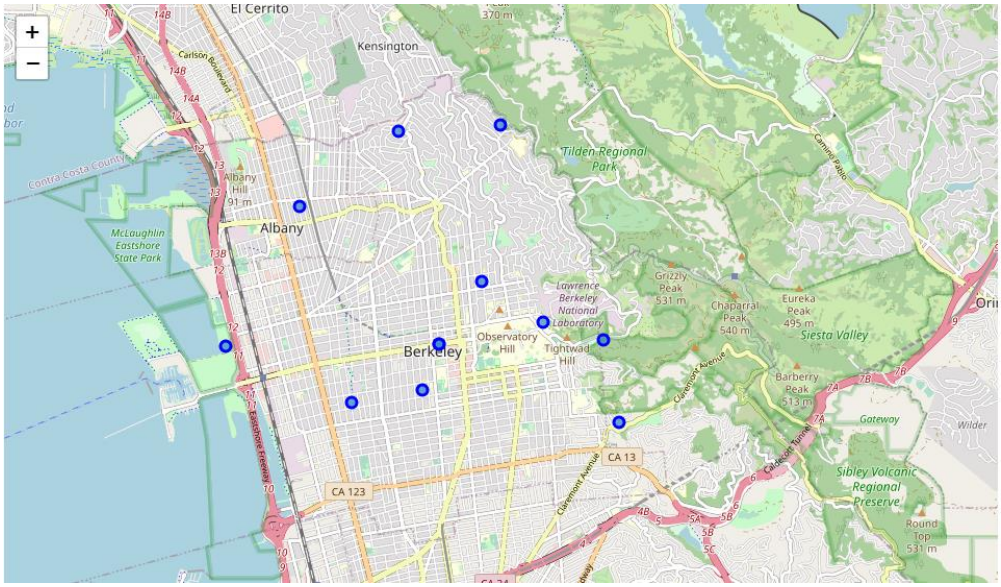

Figure 1 Zip Code distribution in Berkeley, CA.

Next, I access the 100 restaurants in each zip code. This search gave me a total of 199 restaurants in all the zip codes studied. With the latitude and longitude provided by foursquare, I created a map.

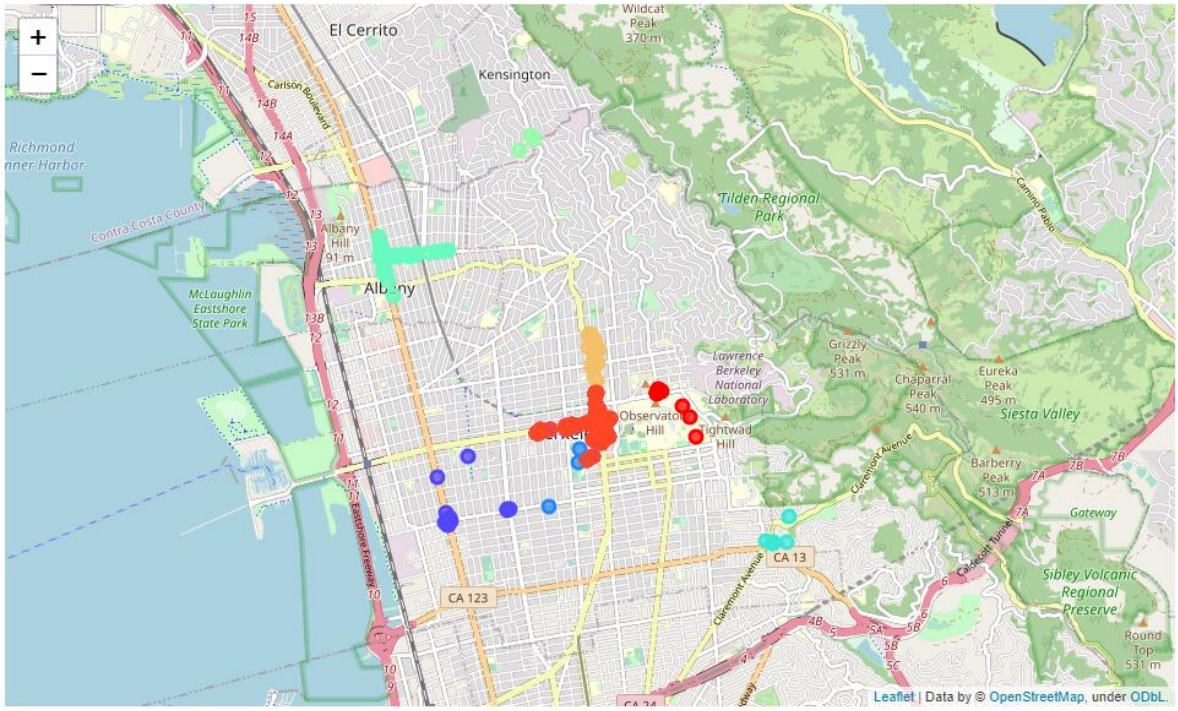

Figure 2 Venues distribution per Zip Code on Berkeley, CA

The map shows that there are Zip codes with a high density of restaurants, such as the 94712 (Red color) which has more than 30 restaurants on its area, while others don't have any, such as the 94704 that does not appear in the map.

Besides just analyzing just the map, I organized the zip codes for the number of restaurants in its area. As the map showed, the zip codes with the greatest number of restaurants are the 94712 and 94701.

	Restaurants
94701	43
94712	43
94706	39
94709	25
94720	20
94702	13
94705	7
94703	4
94707	3
94708	2

Downtown Berkeley's zip code is the 94712, while Albany's one is 94701, this explains the high number of restaurants in those areas. On the other hand, zip code 94704's location is near the California Memorial Stadium, in those upper part of Berkeley, the commerce quantity is low.

To cluster the zip code areas, I used the K-means algorithm. I choose this algorithm for its speed and the possibility to determine the number of clusters.

Results

The result clusters are the following:

	Cluster Labels	Zip_codes	Latitude	Longitude	Restaurants
0	0	94701	37.8718	-122.2718	43.0
1	2	94702	37.864164	-122.286234	13.0
2	1	94703	37.865733	-122.274618	4.0
3	1	94704	37.872228	-122.244743	0.0
4	1	94705	37.8616	-122.242051	7.0
5	0	94706	37.889622	-122.294909	39.0
6	1	94707	37.899439	-122.278492	3.0
7	1	94708	37.900138	-122.261764	2.0
8	2	94709	37.879862	-122.264766	25.0
9	1	94710	37.871464	-122.307095	0.0
10	0	94712	37.8718	-122.2718	43.0
11	2	94720	37.874602	-122.25467	20.0

By creating a map, it is easier to see clusters location:

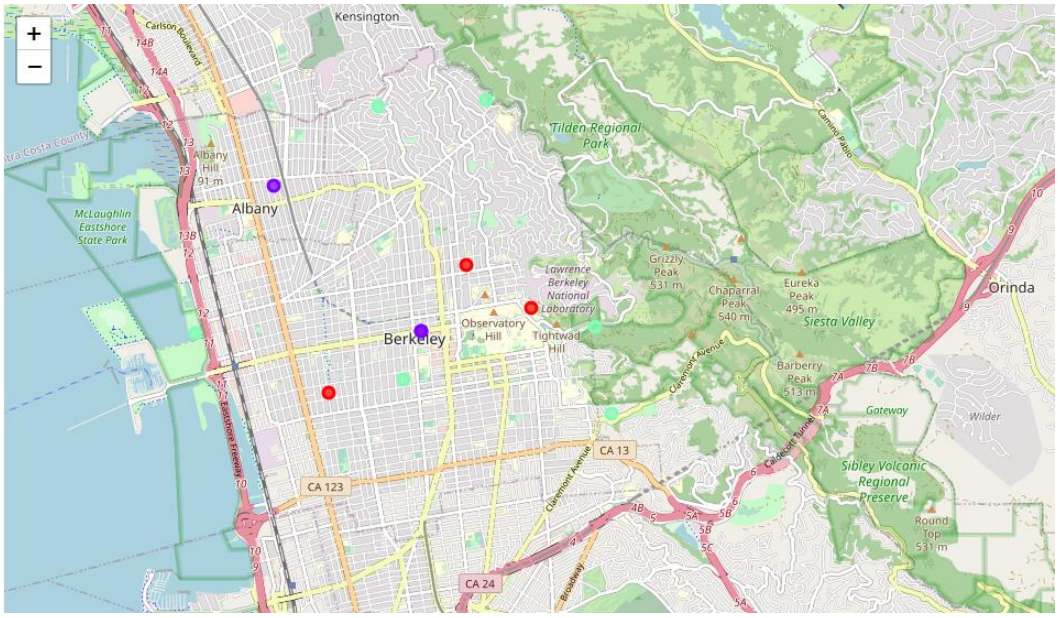

Figure 3 Zip Code clusters

The three clusters obtained can be described as follows:

- Zipcodes with a high number of restaurants, this one is composed of the Downtown Berkeley and Albany zip codes, each one with more than 40 restaurants (Labeled as 0).
- Zipcodes with a medium quantity of restaurants, this one is composed of three zip code areas, all of them with more than ten restaurants (Labeled as 2)
- Zipcodes with a low quantity of restaurants, the last one is composed of six zip code areas, all of them with less than ten restaurants (Labeled as 1)

Discussion:

Berkeley is a small city. The restaurant offer is limited as well as the demand for deliveries. It would be interesting to repeat this exercise in a bigger city with more restaurants and the possibility to use more clusters.

Also, it is crucial to improve this methodology using the demand for deliveries. This data could show new interesting sites to consider.

Conclusion:

The zip codes with a high number of restaurants are perfect spots for the hubs. But it is essential to validate this location with the demand data. It would also be valuable to explore the possibility to create sub-hubs in medium offer zip codes with a high demand for deliveries.