Selecting place for a fastfood in Brno

Mychajlo Chodorev

January 17, 2020

# **Introduction**

## **Background**

Last 20 years many large companies have moved or established their integrated service centers to Central Europe (Czech Republic, Poland, Hungary etc). Reasons for that are lower operational costs and willingness of local authorities to establish strong background for business conducting thus providing mutually beneficial conditions. This caused opening lots of new positions new business parks – locations with lot of companies’ offices and concentrated large amount of people in a limited location.

Concentrating of many people in limited locations has in its turn created a business opportunity for catering business – specifically creation of new fastfood restaurants.

One of cities quickly where large companies presentation growth rapidly is a Brno, Czech Republic. So the potential restaurant business might consider Brno to create a new fastfood restaurant.

## **Problem**

In order to create new fastfood restaurant it is necessary to identify most perspective area to target in terms of estimated customers’ amount. Therefore, this project is aimed to answer the question: what location in Brno is most suitable to create a new fastfood restaurant aimed to integrated service centers employees?

## **Interest**

The topic should pose an interest primarily to fastfood networks of any scale – international down to local ones. Another interested persons can be investors considering investing to fastfood business.

# **Methodology**

The main idea about estimating the most optimal location for new fastfood restaurant is to compare amount and size of large companies in specific location with amount of fastfood restaurants in the same location. In this way, we create an index of area saturation with fastfood restaurants.

Additionally we compare saturation index with one of neighbor cities that are considered to be developed. Such cities have settled business parks and they are likely saturated with fastfood restaurants. As such a reference we’ll use Berlin, Germany. Another reference city considered was Vienna, Austria. However during the discover it turned out there is not enough data for Vienna to make any conclusions.

# **Data acquisition and cleaning**

## **Data sources**

For saturation index we’ll need following data: large companies and fastfood venues.

## **Jobs**

We need size and location of companies offices.  
To get list of companies and their size I refer to Glassdoor, which is a popular vacancies resource. Using its API I’ll get list of open vacancies in defined city. The idea behind that is that company size (in terms of employees amount) is correlated with amount of open positions as turnover rate is relatively stable [[1](#REF_1)]

Then using several geocoding services I get coordinates of the company. Not all companies could be found. That’s the main reason why Vienna wasn’t used as a reference city – location of too few companies was obtained. I’ve made an assumption that companies without identified location aren’t significant enough and thus discarded them from the analysis.

Jobs are saved then to CSV file to save API calls, which is limited.

## **Fastfood restaurants**

I have obtained list of fastfood restaurants along with their coordinates from Foursquare using its API.   
First though I’ve defined manually a list of venue categories to search for, which can be considered as a fastfood restaurant. Categories are obtained from Foursquare documentation [2]:



Second filter to obtain venues is a location of venues. Location is identified by two parameters: coordinates of the location and radius in meters. Therefore venues obtaining is performed after companies clusters are identified. Second parameter is radius of the area. 500 meters is identifies as optimal radius as derived from optimal distance to go to the lunch so it is not too far. The example of Berlin venues are as following:



Once venues are obtained from Foursquare they are stored in CSV file to save API calls.

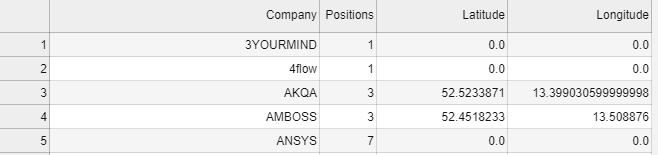
Data sources used:

* Glassdoor (<http://api.glassdoor.com/api/api.htm>)
* Foursquare (<https://api.foursquare.com/v2/venues/explore>)
* Geocoding:
  + Bing
  + Nominatim
  + Mapy.cz (<http://api.mapy.cz/geocode>)

## **Data cleaning**

## **Jobs**

Sample of open job positions obtained looks like this:



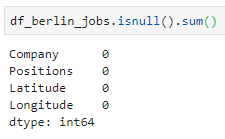
Total amount of companies found in Berlin is **268**:

Jobs with unidentified locations are removed from the dataset:

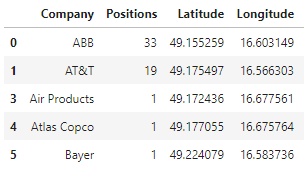


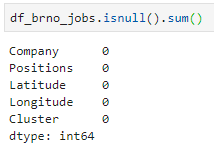
After removing companies with unidentified location there are **72** companies remained.

Checking whether dataframe has any null data:



Same steps are performed for Brno data:



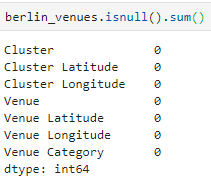


Total amount of companies in Brno is **78.**

Total amount of companies in Brno with identifies locations is **35**

## **Fastfood venues**

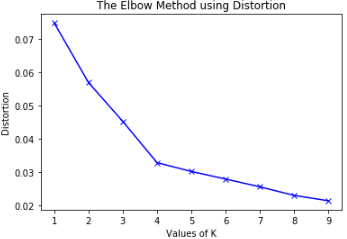
Fastfood venues were obtained using API calls to Foursquare and only relevant data were stored. Therefore data cleaning isn’t necessary. I only check that data has no empty values:



# **Exploratory data analysis**

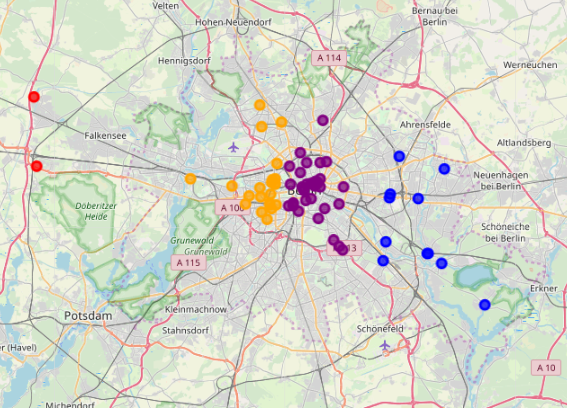
## **Choosing areas**

First approach to define areas of companies is to create clusters using sklearn.KMeans algorithm using companies coordinates as input. In order to identify optimal number of clusters an elbow method was used with clusters number range from 1 to 9.



According to evaluation result the optimal number is **4**.

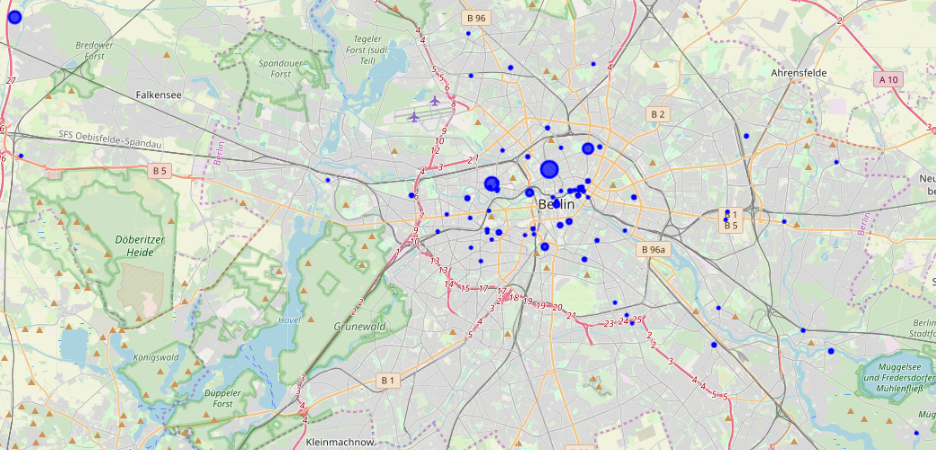
If we put all companies colored by cluster to the Berlin map we can see that clustering doesn’t take into account distance limit from each point to cluster center. Thus some companies are located much farther than predefined 500 meters from each other:



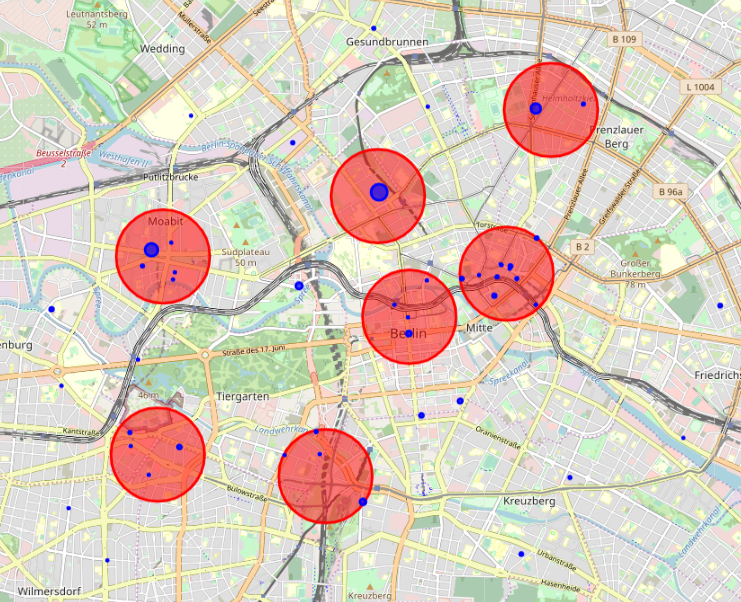
Another drawback of such approach is that size of the company isn’t taken into consideration.

Having this in mind, I’ve decided to identify clusters manually based on cluster radius (500 meters).

First all companies are put on the Berlin map. Point size is proportional to square root of the company size in order to show company size on the map and at the same time keep points size in reasonable boundaries:



After several iterations, following clusters are defined and put onto the map:



This way I have defined 7 clusters in the city center. Companies that are outside of clusters are ignored, as our primary goal is to identify clusters and thus areas for potential fastfood restaurant creation.

## **Finding fastfood venues**

Once areas are defined let us find fastfood venues in those ones using Foursquare API for venues search. Following filters are used:

ll={},{}&radius={}&limit={}&categoryId={}

where:

* ll – latitude and longitude of the area center
* radius – radius of the area (500 m in our case)
* limit – maximum amount of venues returned (free Foursquare API level allows max 100 venues to be returned)
* categoryId – list of categories venue should belong to. List of categories is defined as in 3.1.2

Short outcome of venues search in Berlin is as following:

Cluster 1... found 25 venues.

Cluster 2... found 87 venues.

Cluster 3... found 17 venues.

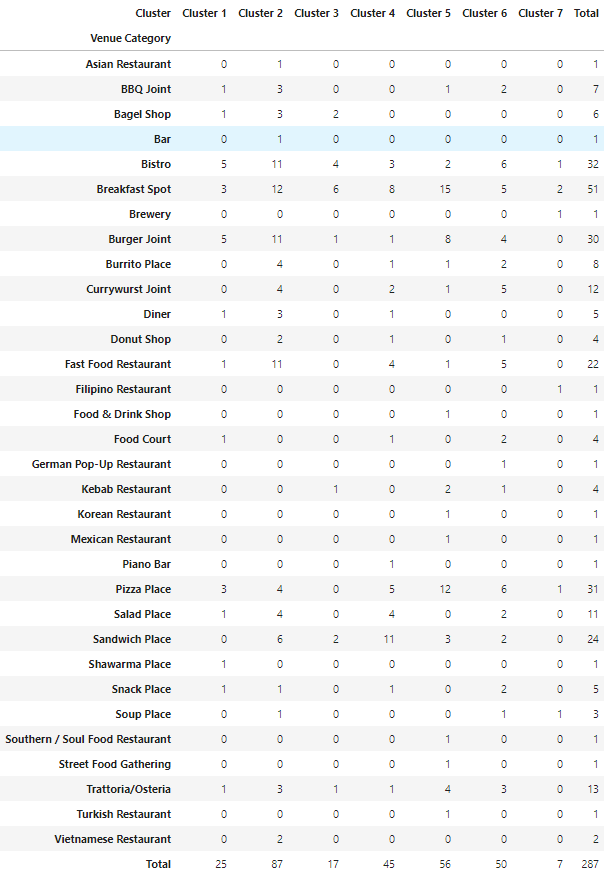
Cluster 4... found 45 venues.

Cluster 5... found 56 venues.

Cluster 6... found 50 venues.

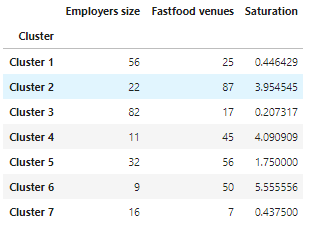
Cluster 7... found 7 venues.

Let us group venues by categories and clusters to see how different venues categories are represented in each cluster:



## **Defining saturation index**

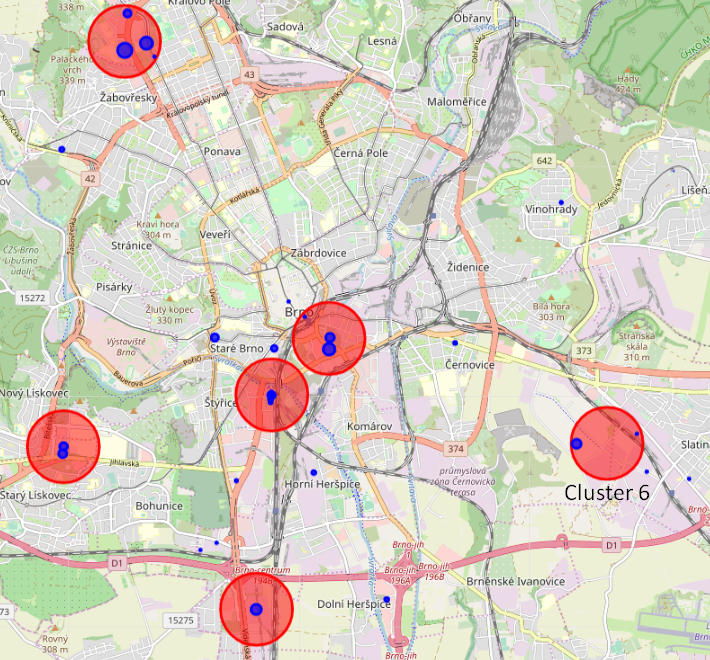
Once we have amount of venues in each cluster we can calculate a saturation index for each cluster and mean saturation index for Berlin:



Having this data the average saturation index is **2.3488936779442002**

## **Exploring Brno**

For Brno, I use same technique of clusters defining as for Berlin: first putting companies onto the map with points reflecting its size:



This way I have defined 6 clusters all over the city.

## **Looking for Brno venues**

Having Brno companies clusters defined let’s look for fastfood venues there. List of filters is same as for Berlin with Brno area locations used instead of Berlin ones:

Cluster 1... found 4 venues.

Cluster 2... found 34 venues.

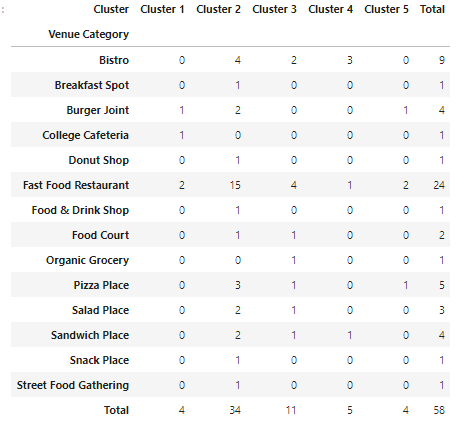
Cluster 3... found 11 venues.

Cluster 4... found 5 venues.

Cluster 5... found 4 venues.

Cluster 6... found 0 venues.

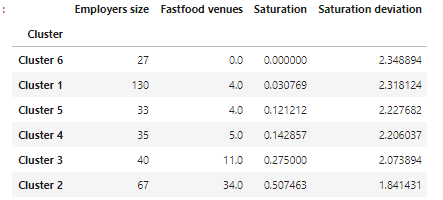
Having venues grouped by areas and categories, we’ve got following picture:



As we can see cluster 6 has no fastfood venues thus isn’t presented in the table above.

## **Defining saturation indices for each Brno area**

Having Brno employers’ data and fastfood venues we get following saturation indices for each area along with deviation from saturation index of reference city – Berlin:



From the table above it seems that cluster 6 is most preferable location to create a fastfood restaurant, as it has none so far (according to Foursquare). However, while having a look at this area it is obvious it is quite empty and its value in terms of employers’ size is formed mainly by single company. Therefore it doesn’t seem most perspective area for fastfood business.

Cluster 1 on the other hand has total largest companies representation and second smallest fastfood venues representation. Its saturation index is just by 0.03 larger than cluster 6’s one. And if we compare it with reference saturation index we can see it’s far from saturation. Therefore this area look quite perspective in terms of creation a fastfood restaurant.

# **Discussion**

During this research, several assumptions were made that affect general outcome and might be revised.

## **Companies size**

Correlation between employees turnover and company size is established. However, on the referenced study it is visible it changes over time and might be different for different cities and countries. Therefore identifying more accurate correlation between company size and employees turnover might be a subject for separate research.

Whilst Glassdoor contains large amount of open positions published local vacancies resources might contain even more comprehensive amount of ones and thus show more refined picture of companies turnover and hence their size.

For more accurate research companies size in terms of employees in target cities might be obtained with higher accuracy without need to estimate it based on indirect indicators such as turnover.

## **Companies location**

Many companies could not be located by all geocoding services I used. In order to get larger representation of the companies in the research and thus higher accuracy it is necessary to come up with different method of obtaining companies location. It definitely should not be manual looking for companies’ location, as it is a cumbersome and time-consuming process for larger cities such as Berlin where 268 companies were found.

## **Fastfood venues obtained from Foursquare.**

Foursquare is popular in countries of North America and Western Europe. In Central Europe however it’s not that popular yet and therefore many restaurants especially small ones and serving fastfood lunches, which is a target for my research aren’t present in Foursquare to full extent. Therefore it worth to consider different data sources to obtain more accurate and actual list of fastfood venues such as Google Maps or Tripadvisor.

# **Conclusion**

In this research, I have identified location clusters of large companies and established correlation between clusters’ companies’ size and amount of fastfood venues presented there.

I have identified that most perspective location to create a fastfood restaurant in Brno is a cluster 1.

For the purpose of the research, I have established a methodology, which can be reused for other target city with similar lunch behavior pattern.

# **Appendix A: References**

[1] - [2016 Turnover Rates by Industry](https://www.compensationforce.com/2017/04/2016-turnover-rates-by-industry.html)

[2] – [Foursquare Venue Categories](https://developer.foursquare.com/docs/resources/categories)