Capstone Project
Clustering of Mobile Phone Shops

Adam Siemiątkowski

July 2020

1. Introduction

1.1. Background

Selection of a location for a Point Of Sale (POS) is a key decision for any retail or service business. The launch of a new POS requires usually significant resources in form of money and time spent on preparation of the location. Once this money spent it becomes sunk cost.

1.2. Problem

The purpose of this report is to verify whether clustering of existing Mobile Phone Shops (MPSes) based on information about surrounding venues available at Foursqare may reveal some patterns that could be helpful in identification of suitable locations for new MPSes.

1.3. Interest

The results of this report may be of interest of mobile operators who run their own POSes or independent entrepreneurs who invest and manage chains of Mobile Phone Shops.

2. Data

2.1. Data sources

The analysis is based on location data from Foursquare regarding venues around MPSes within radius of 500 m in Bonn, Germany area. The list of MPSes has been searched based on Foursquare's category 'Mobile Phone Shop'.

2.2. Data scrapping

In this section I have identified the existing MPSes and presented them on the map. I have started from identification of central latitude and longitude of Bonn, Germany. Then, I have leveraged the FourSquare API to obtain URL that leads to the raw data in JSON form. I have searched for Mobile Phone Shops within radius of 7 km from the center of Bonn. I have found in total 35 MPSes.

I have realized that names of MPSes are not unique so I have added to the list of MPSes a unique identifier.

Next I have searched for venues around identified MPSes within radius of 1 km. Again I have used Foursqare API to obtain raw data in JSON form. I have scraped the raw data in this JSON file in order to retrieve the following attributes of the venues surrounding MPSes: name, category, latitude and longitude.

Final list of first five MPS is as follows.

	name	categories	lat	lng	id	mps
0	Telekom Shop	Mobile Phone Shop	50.733911	7.100555	596d946a79f6c7178f5f5596	mps1
1	Vodafone Shop	Mobile Phone Shop	50.814173	7.159805	536766bf498e423eae106986	mps2
2	o2 Shop Bonn	Mobile Phone Shop	50.735623	7.098684	4c407f274a3e03bb03f56d0e	mps3
3	Vodafone Shop	Mobile Phone Shop	50.736130	7.098090	59a45dfd35d3fc3e2ecca93e	mps4
4	Vodafone Shop	Mobile Phone Shop	50.733718	7.099008	580e321238fa1f1ec7b8b010	mps5

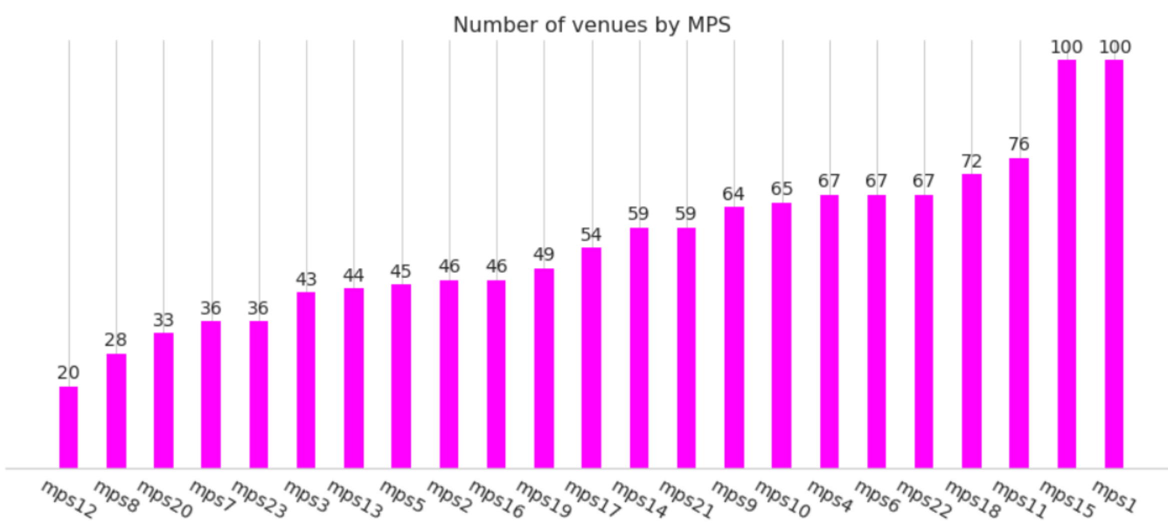
2.3. Data cleaning

First I have verified whether all results of search for MPSes are from category 'Mobile Phone Shop'. I have found in the results some venues with different category although I have used in the search the identifier of this 'MPS' category provided by Foursquare. I have dropped those results.

I have also decided to analyze further MPSes of selected mobile operators. I have filtered the list of MPSes so that it finally contains the ones operated by T-Mobile, o2 and Vodafone. Finally I have analyzed 23 unique MPSes.

Next I have verified the numbers of venues that have been searched around MPSes.

Chart 1 – Number of venues by MPSes



In total I have found 1.276 venues in 155 unique categories.

2.4. Data preparation

After cleaning the data still needed some more processing before it was suitable for clustering. First I have used one-hot encoding to get dummy variables.

Next I have grouped rows by MPS and calculated the mean of the frequency of occurrence of each venue category.

For further analysis I have used the top 10 venues for each MPS and put them into a new data frame.

The top 10 venues for a sample of MPS is presented below.

MPS	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
mps 1	Café	Pub	Plaza	Italian Rest’nt	German Rest’nt	Bakery	Hotel	Bar	Coffee Shop	Park
mps 10	Clothing Store	Café	Tram Station	Italian Rest’nt	Big Box Store	Mobile Phone Shop	Electronics Store	Farmacy	Shopping Mall	Shopping Plaza
mps 11	Ice Cream Shop	Falafel Rest’nt	Castle	Chinese Rest’nt	Shopping Mall	Clothing Store	Department Store	Rest’nt	Farmacy	Optical Shop
mps 12	Ice Cream Shop	Greek Rest’nt	Train Station	Gym	Food	Chinese Rest’nt	Soccer Field	Bus Stop	Pizza Place	Gourmet Shop
mps 13	Italian Rest’nt	Hotel	Farmacy	Super-market	Bakery	Middle Eastern Rest’nt	German Rest’nt	Rest’nt	Farmers Market	Metro Station

3. Methodology

In order to identify any patterns in location of MPSes I have used KMeans algorithm. This algorithm clusters data by trying to separate samples in n groups of equal variance, minimizing a criterion known as the inertia. Inertia is a within-cluster sum of squared distances between data cluster’s data points and its center called a centroid. KMeans algorithm requires that the number of clusters is specified. I have done this using an elbow method. This method identifies the number of clusters at which the gain in reducing distortion of the clustering result slows down.

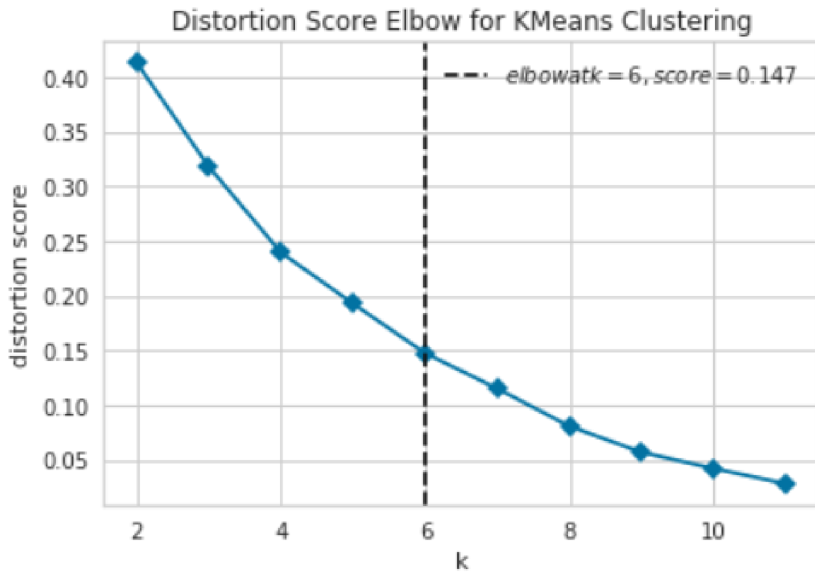
4. Results

The number of clusters has been set at 6 using an elbow method.

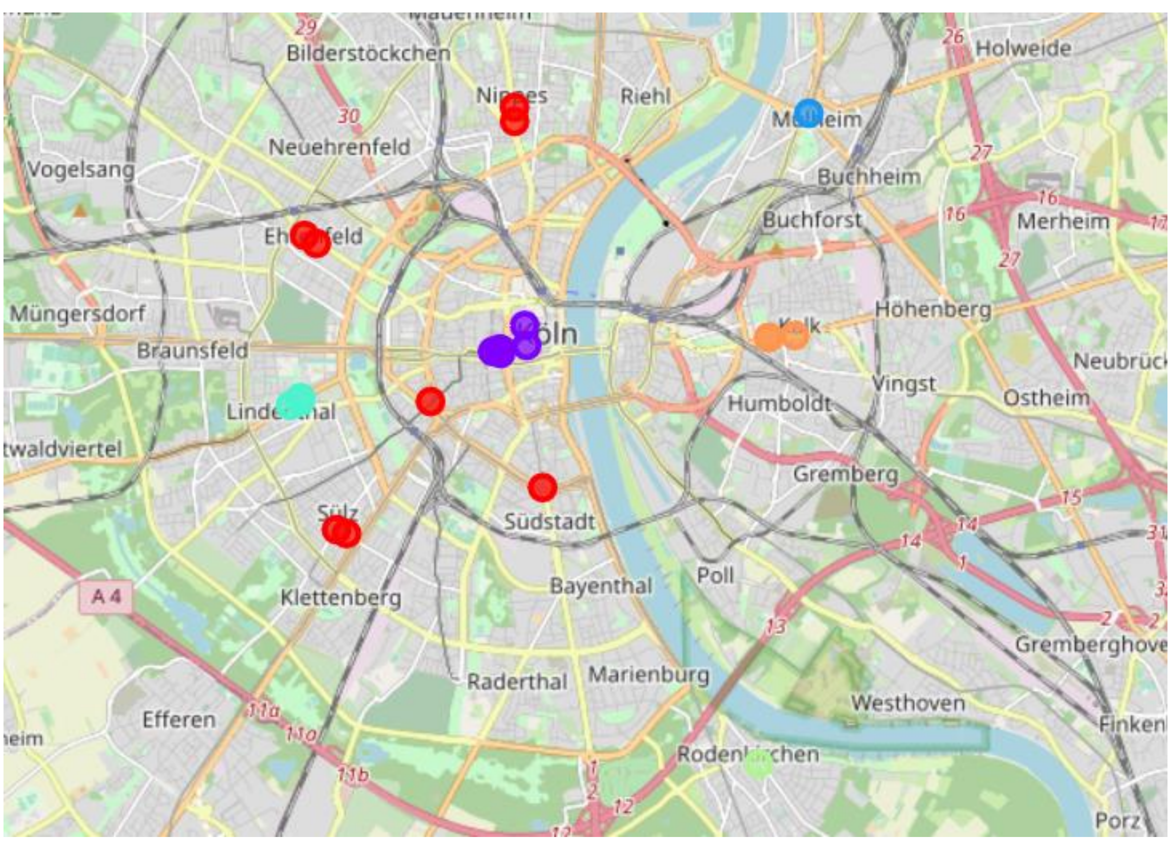
Capstone Project
Clustering of Mobile Phone Shops

Adam Siemiątkowski

July 2020



Based on the set number of clusters I have clustered MPSes using KMeans algorithm. Next I have presented the results of clustering on a map.



The data needed some additional transformation in order to analyze the composition of the clusters. I have decided to use polar plot as the most suitable plot for the data set with Clusters and venue categories.

Capstone Project
Clustering of Mobile Phone Shops

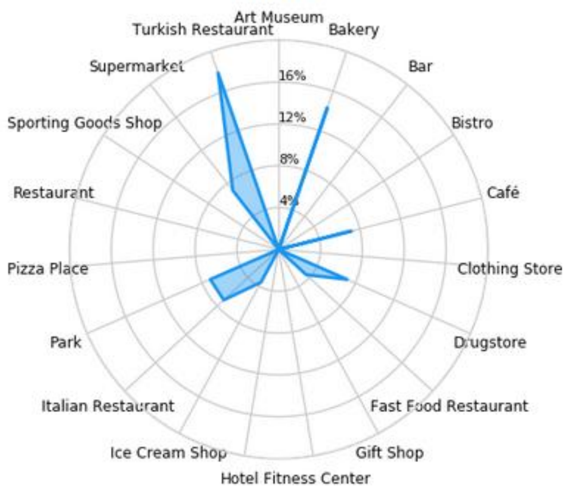
Adam Siemiątkowski

July 2020

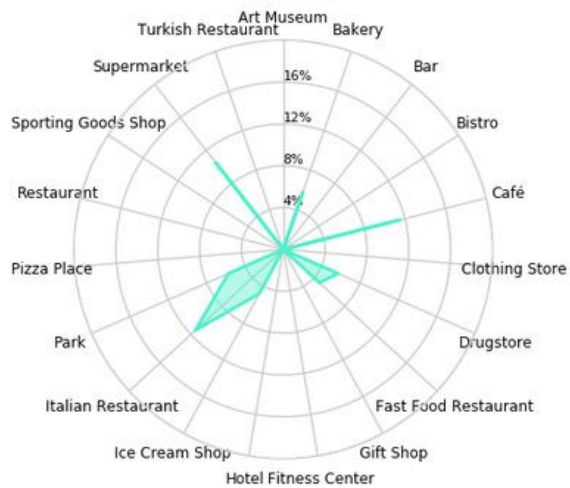
Cluster 0

Cluster 1

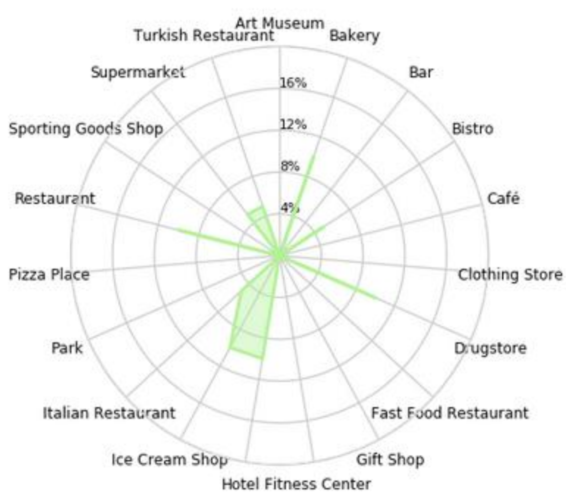
Cluster 2



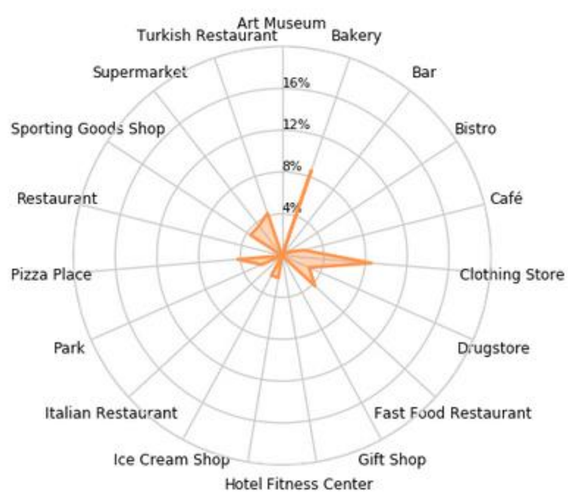
Cluster 3



Cluster 4



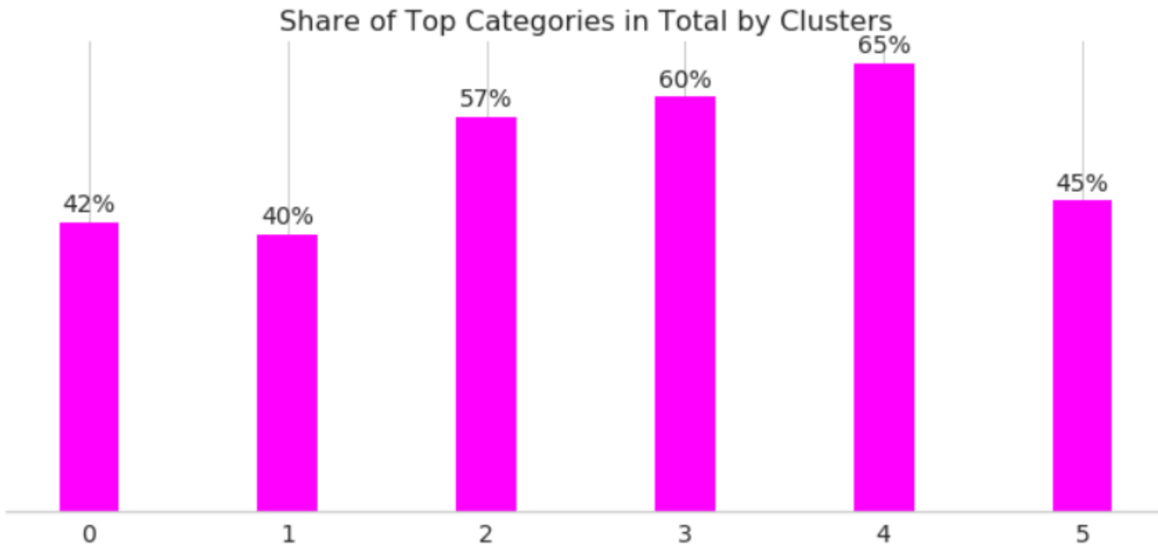
Cluster 5



Capstone Project
Clustering of Mobile Phone Shops

Adam Siemiątkowski
July 2020

I have reduced the number of the categories to be presented on plots from top 10 per cluster to top 6 in order to make the polar charts more readable. The plot below presents how much those top 6 categories per cluster cover the total frequency of all categories.



5. Discussion

Both charts show some clear patterns:

- some neighborhoods are more diversified (Clusters 2, 3 and 4) with more numerous venues in fewer categories and others (Clusters 0, 1 and 5) on the contrary,
- all Clusters include significant number of venues categorized as 'Bakery', 'Supermarket' and 'Drugstore',
- locations in the very center of the city (Cluster 1) are well diversified with significant numbers of venues categorized as 'Art Museum', 'Café', 'Gift Shop' and 'Italian Restaurant',
- locations around the city center (Cluster 0) are less diversified and 'Cafes' are dominant,
- other neighborhoods (Clusters 2-5) are least diversified and have a different combinations of a few dominant categories.

6. Conclusions

The analysis has showed some interesting patterns in locations of Points Of Sales. This is however only a first step to recommendation of a suitable place for new POSes. The following one should leverage the use of Machine Learning and help verify all potential locations based on the similarity of their neighborhoods to the ones with existing POSes.