IceCreamDataScienceReport

August 5, 2021

1 Ice Cream Shops Exploration

1.1 Business Problem

A young, international and fast growing chain of Ice Cream Shops would like to expand their business to Germany. The business owners are not familiar with current situation of Ice Cream Business in Germany and they would like to explore it better. So they would like to collect statistics about current presence of ice cream shops in German cities.

Also, the company would like to receive the list of best candidates cities to open new shops, based on few criteria: - the cities with the smallest density of existing coffee shops per 1000 person of population; - the cities with at least 50000 people of population

1.2 Data

In order to collect the required statistics about the Coffee Shops in German cities, we need: 1. The list of german cities with population data for each; 2. The information about current Ice Cream Shops for each city.

The list of german cities with corresponding population was found on the web page of Statistisches Bundesamt: https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/05-staedte.html). Here, the corresponding .csv file was downloaded and will be used from local repository. This file contains list of all german cities with following useful information: - City Name - City Postal Code - City Population - City Surface Area

Additionally, since the data file from Statistisches Bundesamt has only one Postal Code per city, but the large cities as Berlin or Hamburg have many Postal Codes, additional database was downloaded from https://github.com/zauberware/postal-codes-json-xml-csv for Germany. This database contains: - postal codes for every single neighborhood of each city - latitude and longitude coordinates

The information about current Ice Cream Shops will be retrieved from ForeSquares.com API based on coordinates of the cities from above databases.

1.3 Methodology

We start our analysis by collecting the data from database of Statistisches Bundesamt and merging it with data from Zauberware, in order to retrieve geographical coordinates for each city.

We filter out the cities list to only cities with more than 50 000 population as business problem requires. Also, we calculates radius for each city based on it's surface area, assuming that cities have

round shape, based on the formula A=PiRR. We will need the radius for latter calls of Foursquare API.

Then, we attempt to collect the Ice Cream Shops list from FourSquare API for each location using the city name and previously calculated radius. If we are not able to retrieve any results by city name (which is sometimes a case because of different versions of writing), we try again by coordinates.

Also, since FourSquare results are limited to only 50 Results per response, for some very big cities (like Berlin, Hamburg, Köln) it is impossible to get the list of all Ice Cream shops with only one call. For these cities, we go through all neighborhoods (from the database of Zauberware), make call to Foursquare API for them separately, and create the full list of unique ice cream shops out of these responses.

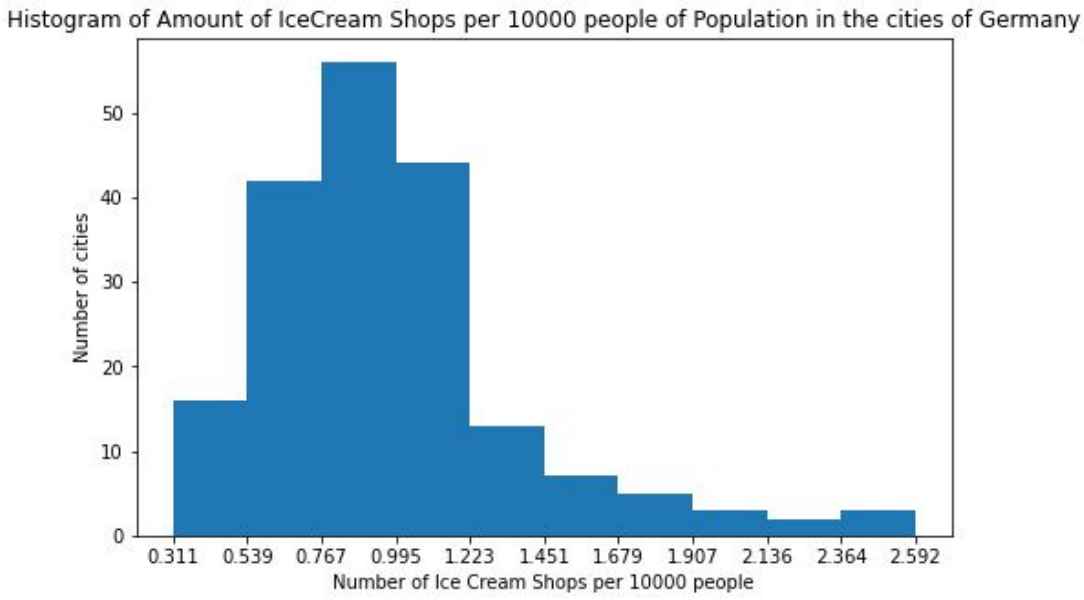
After we get in this way the amount of Ice Cream Shops for all cities, we can start with exploratory analysis. First, we calculate the density of Ice Cream Shops per 10000 People of population based on data about total population and amount of shops.

After that we look for the possible correlations between density (our target variable) and other variables. I.e. we examined the correlation between Density and Total Population, as well as between Density and Geographical Coordinates.

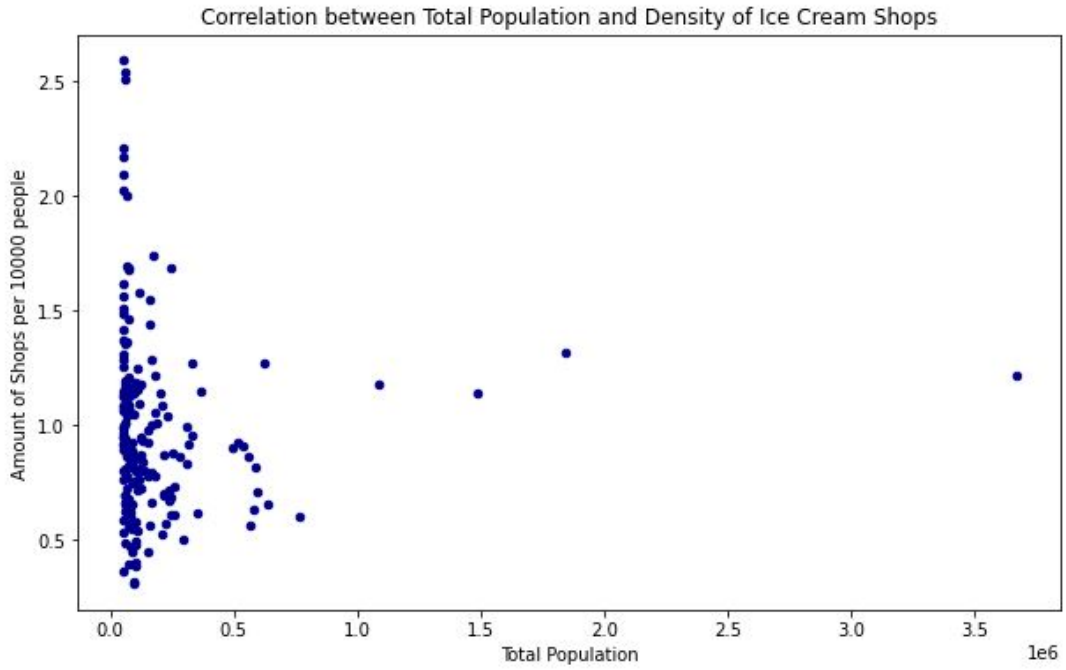
Finally, we run Clustering algorithm (k-means) on the dataset, in order to capture some tendencies and groupings in the data.

1.4 Results

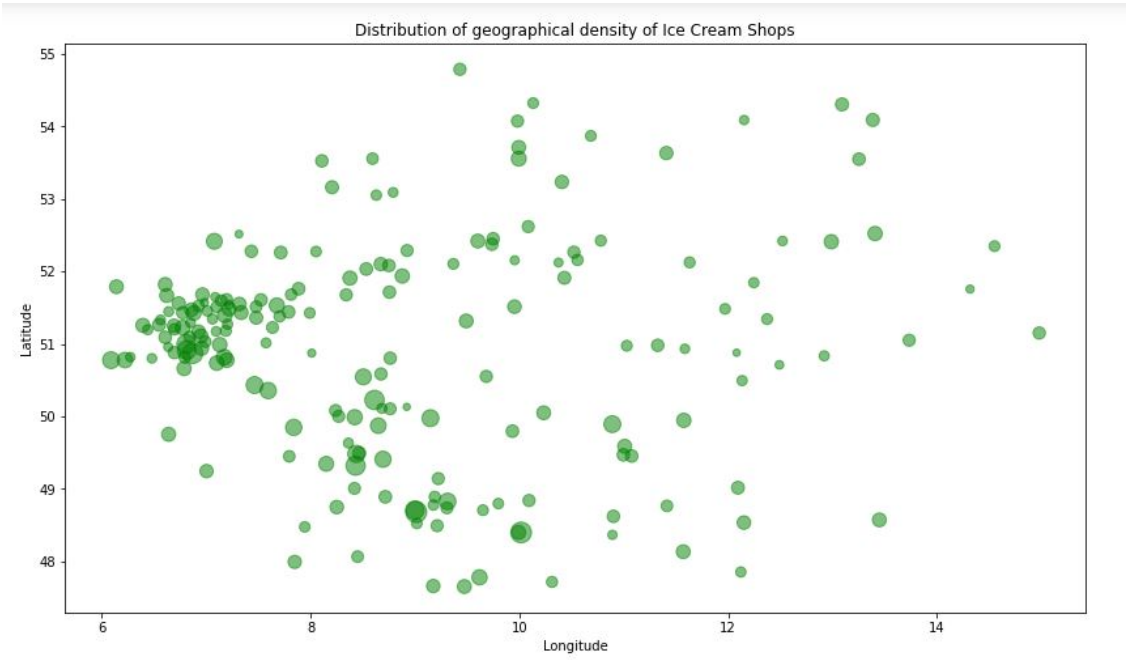
Here are the results of the exploratory Analysis:



The distribution of Density is rather smooth across cities with most frequent density around 0,8 shops per 10000 person, with rather long tail to the right.

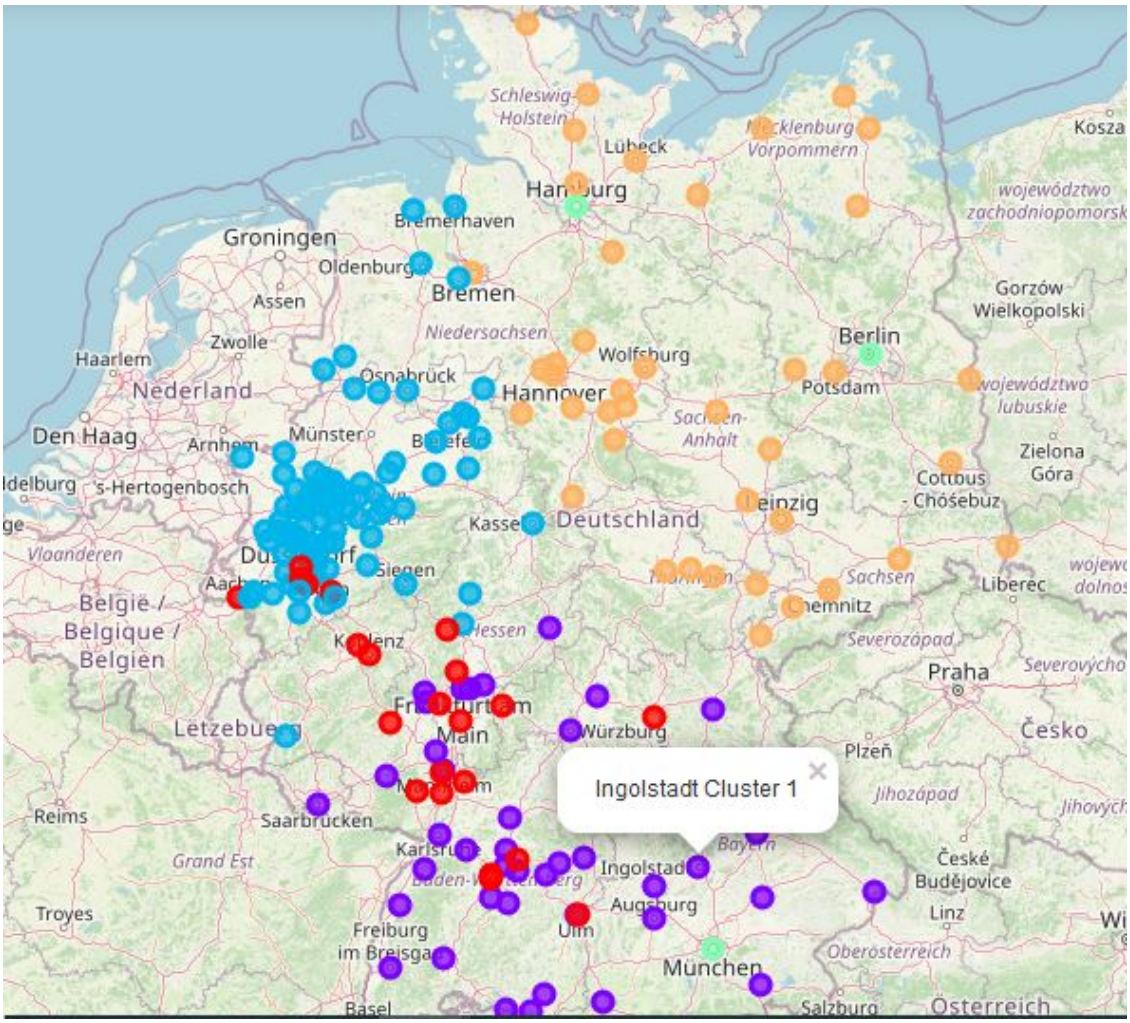


We can notice that bigger cities (with population above 1 Million) have pretty same density of Ice Cream Shops per 10 000 population: around 1,3 shops. The smallest cities have very different density (from close to 0 up to 2,5 Shops per 10000), however with growing of size of the city the density of shops has tendence to decline.



It looks like the cities in West and South Germany have tendence to have more shops per 10 000 Population as in East Germany.

Finally, here is the results of Clustering Algorithm. K-Means Algorithm was used, with following components for clustering: Population, Density of Ice Cream Shops, Latitude und Longitude.



1.5 Results

Both Cluster 0 and Cluster 1 are geographically distributed in South Germany (Red and Violet points). However, the most important difference is that Cluster 1 groups cities with significantly smaller density of Shops as Cluster 0. (Mean of 0,88 vs 1,82) and standard distribution of 0,21. Thus it can make sence to concentrate on cities of Cluster 1 to open new Ice Cream Shops.

Cluster 2 (blue points) has also not too high density of shops (mean 0,90), but the cities are placed more at north of Germany.

Cluster 3 contains just 3 cities, and these are the cities with very high total population (Berlin, München, Hamburg). The density of Ice cream shops in these 3 cities is pretty same and is rather high (mean of 1,22).

Cluster 4 has even smaller density of Ice Cream Shops per 10000 than Cluster 2 (mean of 0.78), and all these cities are geographically placed in East Germany.

Below is the top 10 list of the cities with lowest density for 3 geographical cluster as required output for the given data analysis. The cluster 3 and cluster 0 are excluded as they don't compete with other clusters for the density of ice cream shops.

	Best Cities of South Germany	Density	Best Cities of North Germany	Density	Best Cities of East Germany	Density
0	Hanau	0.310907	Lingen (Ems)	0.365577	Gera	0.322148
1	Augsburg	0.505762	Siegen	0.389219	Cottbus/Chó?ebuz	0.401292
2	Worms	0.598501	Gladbeck	0.396773	Zwickau	0.451009
3	Frankfurt am Main	0.602583	Herne	0.447430	Salzgitter	0.479428
4	Rosenheim	0.629416	Marl	0.475811	Hildesheim	0.491676
5	Schwäbisch Gmünd	0.654268	Bergheim	0.487005	Rostock	0.525835
6	Tübingen	0.655695	Eschweiler	0.531143	Jena	0.538875
7	Stuttgart	0.660470	Düren	0.548149	Brandenburg an der Havel	0.554139
8	Offenburg	0.667034	Solingen	0.565167	Bremen	0.563818
9	Göppingen	0.691886	Ratingen	0.571298	Chemnitz	0.608929

1.6 Conclusion

The geographical clustering of Ice Cream Shops would allow the efficient planing of regional ice cream shop sets based on location. Each cluster was examined on the most promising cities based on list of cities with smallest densities. However it should be noticed that this model doesn't consider the flow of tourists for each city. The given analysis is based purely on own population of the cities, i.e. local ice consumers. However big flows of tourists can impact the profitability of Ice Cream Shops significantly. Thus it would make sence to extend the analysis also to this aspect of cities profile, which however is not in scope of this project.

[ ]: