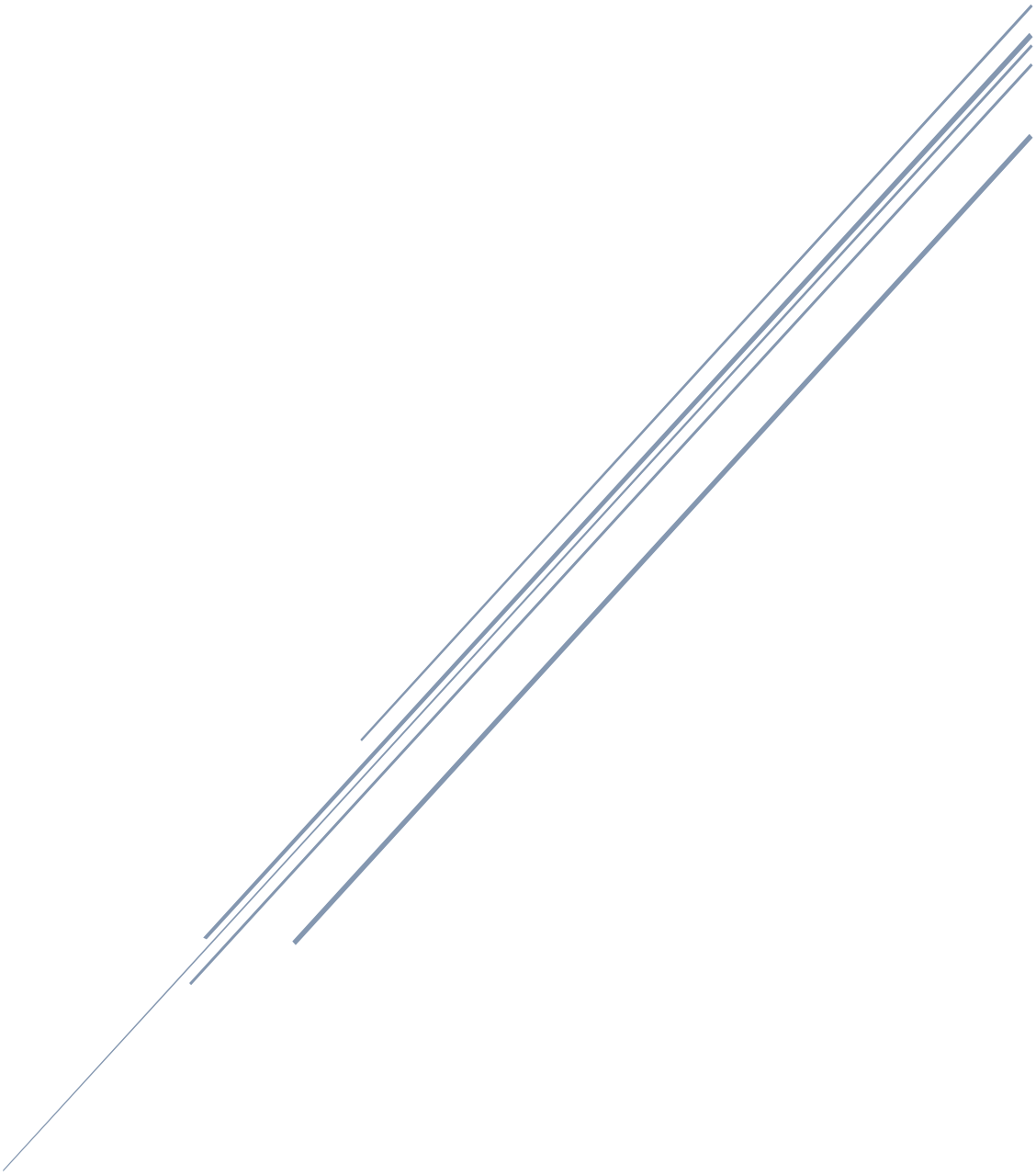
Peer-graded Assignment: Battle of Neighbourhoods – Final Report

Potential Analysis - Germany



Author: Dimitrios Ziliaskopoulos

16.11.2019

Table of contents

1.Introduction

1.1. Disclaimer

All details are given without guarantee.

All information mentioned here has been prepared for the course: "Capstone Project - The Battle of Neighborhoods" and does not serve the purpose of business use. Likewise, it is not a recommendation which is hereby pronounced.

1.2. Background

Currently there is no possibility to have all data in one place, so it is not visible and has to be evaluated again and again. Pytohn offers the possibility to evaluate the data based on the libraries.

1.3. Problem

A person who on the one hand works in a large company or can be a private individual needs to make a decision, where a new store is opened, usually comprehensive location information. This information can be information about the population in a city, the GDP per person or how many coffee's there are already in a city. On the basis of this information, as a person of a company not operating in Germany, you can get an overview of which areas within Germany have the potential to open a business.

In addition to the advantages mentioned above, this chart can also be used to illustrate real estate investment decisions.

1.4. Interest

The interests are to learn, as an outsider, how the German economy is organised. This is pointed out by the GDP per head. Furthermore it is represented by the Foursqare API how many Caffees are within the city centre. In combination of these two data, it can be balanced which city is suitable for the opening of a coffee bar.

2.Data acquisition and cleaning

2.1. Data sources

The project is based on various publicly available data.

This includes information from Wikipedia:

https://de.wikipedia.org/wiki/Liste_der_Gro%C3%9Fst%C3%A4dte_in_Deutschland

https://de.wikipedia.org/wiki/Liste_der_deutschen_St%C3%A4dte_nach_Bruttoinlandsprodukt

Furthermore, information was extracted from the Foursqare API.

2.2. Data cleaning

This is the result after downloading the data. In the first step, the largest German cities were merged into one table. And enriched with the region (State).

	City	Inhabitants_2018	Area km²(2016)	Inahb./km²(2018)	Region
0	Berlin	3.644.826	89168	4.088	Berlin
1	Hamburg	1.841.179	75522	2.438	Hamburg
2	München	1.471.508	31070	4.736	Bayern
3	Köln	1.085.664	40502	2.681	Nordrhein-Westfalen
4	Frankfurt am Main	753.056	24831	3.033	Hessen

In the next step, the GDP per head of the respective cities was determined over a further API.

	City	Bruttoinlandsproduktin Mio. €	Bruttoinlandsproduktpro Kopf in €
0	Berlin	130.537	36.798
1	Hamburg	112.959	62.793
2	München	109.571	75.186
3	Frankfurt am Main	66.917	91.099
4	Köln	63.463	59.407

In order to have a clear table, with all data, in the next step the two tables were merged.

	City	Inhabitants_2018	Area km²(2016)	Inahb./km²(2018)	Region	GDP in MIO	GDP per capita	Avearage GDP per capita (Region)
0	Berlin	3.644.826	89168	4.088	Berlin	130.537	36.798	40.568
1	Hamburg	1.841.179	75522	2.438	Hamburg	112.959	62.793	65.603
2	München	1.471.508	31070	4.736	Bayern	109.571	75.186	47.946
3	Nürnberg	518.365	18638	2.781	Bayern	28.130	55.071	47.946
4	Augsburg	295.135	14684	2.010	Bayern	14.060	48.824	47.946

In the last step of this table, the coordinates of the respective cities were added. The implementation is done via the library geopy. Based on this algorithm, all coordinates could be searched automatically. The coordinates also refer to the city centre, so that on the basis of further analyses concrete data can be made according to the emergence of different shops.

	City	Inhabitants_2018	Area km²(2016)	Inahb./km²(2018)	Region	GDP in MIO	GDP per capita	Avearage GDP per capita (Region)	Latitude	Longitude
0	Berlin	3.644.826	89168	4.088	Berlin	130.537	36.798	40.568	52.517037	13.388860
1	Hamburg	1.841.179	75522	2.438	Hamburg	112.959	62.793	65.603	53.550341	10.000654
2	München	1.471.508	31070	4.736	Bayern	109.571	75.186	47.946	48.137108	11.575382
3	Nürnberg	518.365	18638	2.781	Bayern	28.130	55.071	47.946	49.453872	11.077298
4	Augsburg	295.135	14684	2.010	Bayern	14.060	48.824	47.946	48.366804	10.898697

The data of the Foursqare API will be stored in a separate table. They are included in the dataframe shown here and can therefore be analysed according to different categories. Furthermore, on the basis of this table it can be visualized where most of the coffees are. Clustering is also possible to show outsiders where the center is.

	name	categories	address	city	distance	lat	lng	state

2.3. Feature selection

The most important features are the GDP per inhabitant of a city.

For the Foursqare data, it is the number of coffee bars in a city.

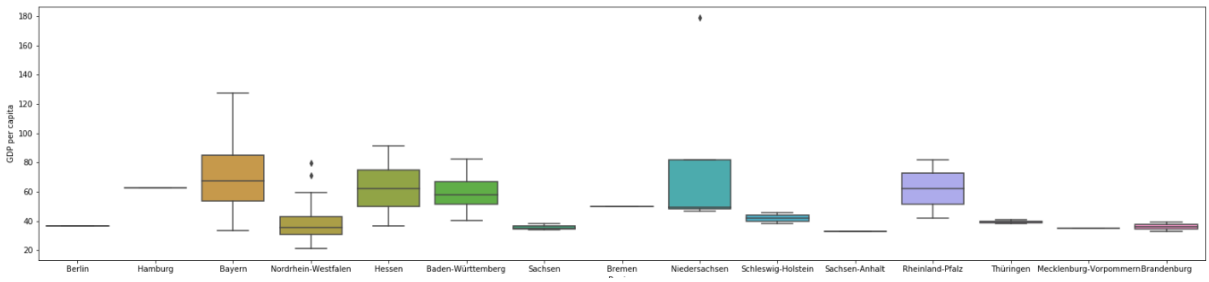
3.Exploratory Data Analysis

3.1. Overview of the Regions

The next figure shows which GDP per capita is available in the respective regions (federal states).

Bavaria has a very stable GDP per capita on average.

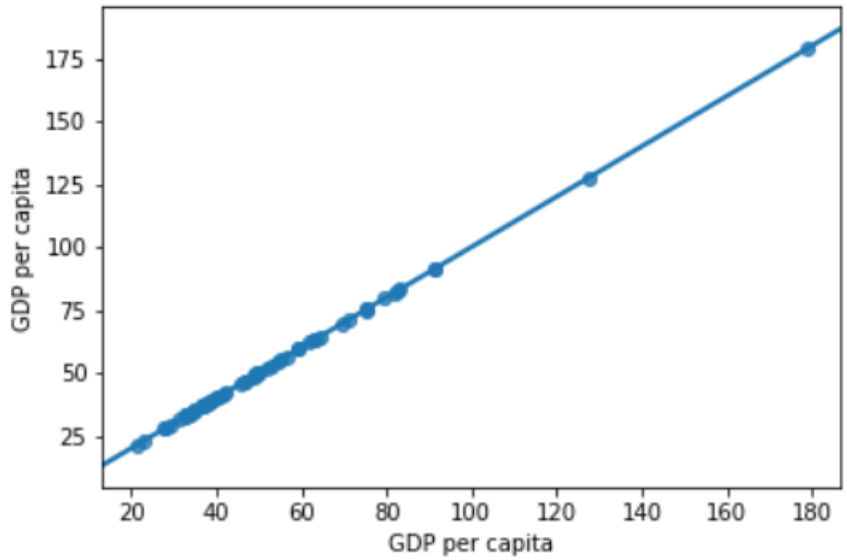
The outlier in Lower Saxony is striking. With a GDP per capita of 178,706 euros, Wolfsburg has the highest GDP per capita in Germany. The reason for this is that the worldwide positioned car manufacturer VW has its location there.



3.2. Distribution of the GDP per head

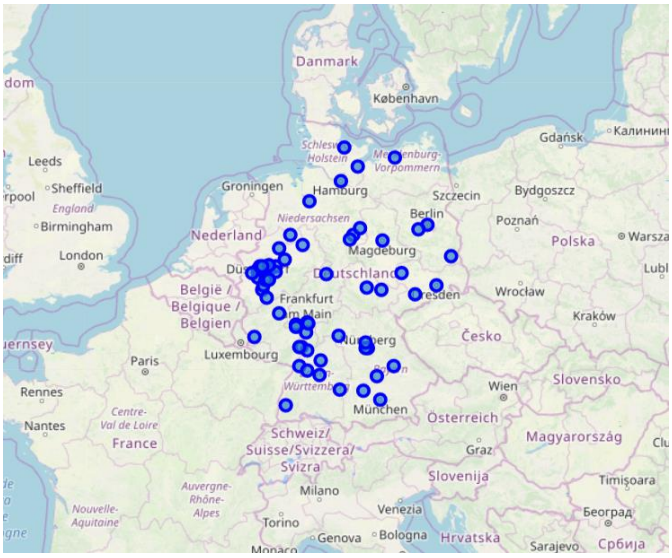
This graph shows how the distribution of the GDP's per head is in Germany.

Generally, it showed up that the GDP per head is on average with 50,000 euros and the lowest value is with 21,229 euros and the highest amount with 178,706 euros per head.



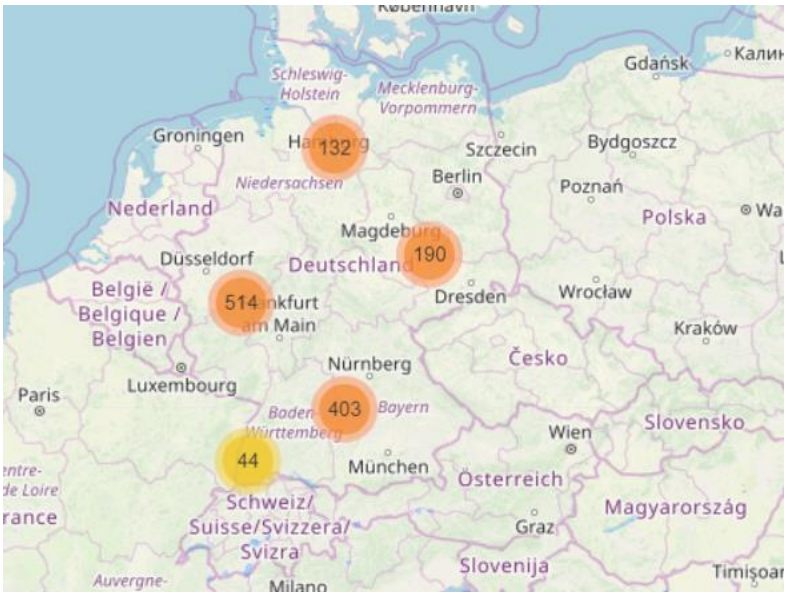
3.3. Map Germany

This map shows the biggest cities of Germany on one map.



3.4. Map Coffee Bar´s

This map shows the number of coffee bars within 500 meters of the city center.



4.Conclusions

Altogether it showed up that Germany has a very different structure of the GDP's. Thus, there are cities, which are small and due to the very strong economic resources an above average GDP have. This potential is to be found out in the next step in order to find a business case purposefully for this employment.

5.Future directions

The next steps could be to include more categories from the Foursquare API, so you could use this data not only for coffee bars, but also for other shops. There is, so to speak, no limit to which industries it can be. The only limit is the Foursquare data. Thus, for example, it can be evaluated for Italian restaurants where it is worth opening a restaurant.

Furthermore, it is interesting to find out which people are interested in shopping at this shop. This data can also be retrieved via the API.