



## **COURSERA CAPSTONE IBM Applied Data Science Capstone**

### **Opening a new coffee shop in Germany**

By MTruong  
March 2020

#### **1. Introduction**

Given the fact that Vietnam is the second largest coffee producer and has some of the best coffee beans in the world, many Vietnamese friends of mine are planning to establish coffee business in Germany. Obviously, having a successful business requires further measure and some serious consideration. One of the most important factors of not only coffee shops but also any Food & Beverages industry business is location. Particularly, apart from many other important factors such as the menu, customer service, etc., the ability to attract as much appropriate traffic as possible will determine whether the coffee shop will be a success or a failure.

#### **Business problem**

The objective of this capstone project is to analyse and select the best city locations in Germany to open a new coffee shop. Using data science methodology and machine learning such as segmenting & clustering, this project aims to answer the business questions: Among different areas in Germany, where should I open my coffee shop to maximize the appropriate traffic?

#### **Target Audience**

The target Audience is definitely anyone who is interested in opening a coffee shop. Furthermore, I would like to make the code as agile as possible, so that I can apply to different city areas/countries across Germany, to further measure in case of expansion.

#### **2. Data**

To solve the problem, we will need the following data:

- List of cities in Germany with population & gdp per capita
- Latitude & Longitude coordinates of those cities. This is to plot the map & also to get the venue data.
- Venue data from Foursquare relating to coffee shops, office building and universities. We will use these data to perform segmenting and clustering on the areas.

## Sources of data and methods to extract them

First thing first, I will import all the libraries needed for this exercise. The German cities data is a combination of 2 dataframes. I will load this csv file

<https://simplemaps.com/static/data/country-cities/de/de.csv> contains information of

German cities, in terms of states, population, latitude & longitude, into the first dataframe. Secondly, this Wikipedia page [https://en.wikipedia.org/wiki/List\\_of\\_German\\_cities\\_by\\_GDP](https://en.wikipedia.org/wiki/List_of_German_cities_by_GDP) contains a list of cities in Germany with gdp per capita. With support from Python requests & beautifulsoup packages, I will use web scraping techniques to extract the data from Wiki page. I will then do a lot of data cleaning and then merge the two tables to get all the information needed. After that, I will be using Foursquare API to get the venue data for those cities. The venue data will include the top 100 venues within the 5km radius of city center of all cities in Germany. From this, I calculate the frequency of coffee shops appearing in the most popular 100 venues of every city, which I later use to rank which cities have higher/lower competition compared to others.

### 3. Methodology

To choose the best city location to open my coffee business, we need to select cities with low competition, high population and high gdp per capita. I will use K-means clustering to classify the cities based on their population, gdp per capita & competition.

#### 1/ Cluster cities

Run k-means to cluster cities into 5 clusters based on their population, gdp per capita & competition

#### 2/ Map the cluster cities



### 4. Results Discussion & Conclusion

Cluster 1:

	city	coffee	population	gdp per capita	Cluster Labels	latitude	longitude
0	Augsburg	0.060000	259196	48824	0	48.371538	10.898514
2	Bielefeld	0.054795	291573	38588	0	52.028332	8.542002
3	Bonn	0.080000	313125	71222	0	50.734380	7.095485
4	Braunschweig	0.160000	235054	46928	0	52.265939	10.526726
6	Bremerhaven	0.041667	117446	34771	0	53.550209	8.576735
7	Coburg	0.106383	41901	83501	0	50.259366	10.963843
9	Cottbus	0.076923	84754	33067	0	51.757691	14.328875
14	Ernden	0.031250	48046	70448	0	53.367454	7.207776
15	Erfurt	0.110000	175476	38284	0	50.978695	11.032831
17	Flensburg	0.100000	85838	42827	0	54.784310	9.439611
18	Fürth	0.076923	112025	33415	0	49.481210	10.967620
19	Gera	0.000000	103055	27391	0	50.875507	12.083296
21	Heidelberg	0.130000	143345	53079	0	49.404123	8.683733
22	Ingolstadt	0.075758	120658	127523	0	48.765081	11.423725
23	Jena	0.123077	95171	40609	0	50.928782	11.589901
24	Karlsruhe	0.110000	283799	63147	0	49.009365	8.404440
25	Kassel	0.070000	194501	49937	0	51.313782	9.484645
26	Kiel	0.120000	232758	45821	0	54.312145	10.135526
27	Koblenz	0.030000	107319	69504	0	50.353573	7.578835
29	Lübeck	0.060000	212207	38078	0	53.866870	10.688367
30	Magdeburg	0.030000	224931	32978	0	52.127731	11.629163
31	Mainz	0.090000	184997	54696	0	49.984188	8.279096
32	Mannheim	0.160000	313174	64483	0	49.496706	8.479547
34	Münster	0.160000	237041	56612	0	51.981114	7.624239
35	Oldenburg	0.140351	159218	43934	0	53.141178	8.214674
36	Osnabrück	0.129032	166462	48732	0	52.272642	8.049799
37	Potsdam	0.080000	145292	39293	0	52.402247	13.057122
38	Regensburg	0.060000	129151	82967	0	49.034512	12.119234
39	Rosenheim	0.114754	60167	49671	0	47.857316	12.119255
40	Rostock	0.060000	198293	34910	0	54.086813	12.113453
41	Schwerin	0.057971	96641	36917	0	53.628373	11.409462
43	Ulm	0.060000	120451	75044	0	48.397053	9.968956
44	Wiesbaden	0.060000	272432	61913	0	50.082582	8.249322
45	Wuppertal	0.168675	349470	37186	0	51.270268	7.167553
46	Würzburg	0.080000	133731	59404	0	49.793912	9.951214

## Cluster 2:

	city	coffee	population	gdp per capita	Cluster Labels	latitude	longitude
20	Hamburg	0.14	1739117	62793	1	53.575323	10.015340
33	Munich	0.15	1260391	75186	1	48.150000	11.583333

## Cluster 3:

	city	coffee	population	gdp per capita	Cluster Labels	latitude	longitude
1	Berlin	0.07	3094014	36798	2	52.516667	13.4

## Cluster 4:

	city	coffee	population	gdp per capita	Cluster Labels	latitude	longitude
5	Bremen	0.090000	546501	50052	3	53.073789	8.826754
8	Cologne	0.110000	963395	59407	3	50.933333	6.950000
10	Dortmund	0.110000	588462	36781	3	51.514942	7.465997
11	Dresden	0.070000	486854	37993	3	51.048562	13.745794
12	Duisburg	0.093333	488005	33634	3	51.432469	6.765161
13	Düsseldorf	0.110000	592393	79619	3	51.228304	6.793849
16	Essen	0.100000	573468	41512	3	51.456570	7.012282
28	Leipzig	0.080000	504971	35123	3	51.344190	12.386504
42	Stuttgart	0.080000	606588	82397	3	48.782343	9.180819

Examining the data, we can conclude that cluster 2, Berlin specifically, is the best location city to open a new coffee shop in Germany. Berliners are coffee-underserved because the city has the biggest population across Germany while having a relatively small amount of coffee shops in the 5km radius of city center. On the other hand, in cluster 1, despite having a huge population and high gdp per capita, Munich & Hamburg are not ideal cities to kick off your coffee business. The two cities is the most populous areas in Germany, in terms of cafeteria, which results much more competition. Another cluster, which is also worth-mentioning, is cluster 3. Among cluster 3, the cities such as Bremen and Stuttgart are highly potential locations for coffee start up, given low competition, average population, and high gdp per capita.

If I have to use one word to describe the 4 clusters, they will be:

Cluster 1: 50/50 chance cities

Cluster 2: most competitive cities

Cluster 3: under-served cities

Cluster 4: potential cities