

Coursera Capstone: The battle of Prague Neighborhoods

Table of Contents

Introduction: Business Problem:	2
Data Requirement:	2
Methodology:	2
Results:	5
Discussion:	6
Conclusion:	6
Limitations:	6
References:	6

Introduction: Business Problem:

Every day millions of people use the services of cafes and restaurants. For example, the restaurant and food service industry is a \$660 billion industry in the United States. This industry is divided into two groups: those that prepare and serve food and those that produce and distribute food, equipment, and services needed by food providers. The most common example of the first group is restaurants. The second group includes equipment suppliers and food and beverage suppliers.

Imagine that your company distributes food, equipment, and services and your main customers are restaurants. Nowadays the number of restaurants has decreased. And still you want to expand your business. The best way to do this is to find another location for it.

You select Prague the capital of Czech Republic. You need to know what areas are in the city of Prague and the number of restaurants of a certain type in different districts. In addition, the necessary information for solving logistics problems is the density of restaurants per square kilometer.

Data Requirement:

Based on the problem mentioned above, factors to be considered are:

- Number and area of neighborhoods in Prague
- Number of restaurants and their categories in each neighborhood

The following data sources will be required to extract relevant data:

- A list of neighborhoods in Prague from open data
- A geocoder library for extracting latitude and longitude coordinates
- Foursquare API for searching for restaurants within these neighborhoods and providing their various latitudes and longitude coordinates

Methodology:

First, let's look at the administrative division of Prague at https://en.wikipedia.org/wiki/Districts_of_Prague [1].

Administrative district	Municipal district
Prague 1	Prague 1
Prague 2	Prague 2
Prague 3	Prague 3
Prague 4	Prague 4, Kunratice
Prague 11 (part)	Prague 11, Šeberov, Újezd u Průhonic
Prague 12	Prague 12, Libuš
Prague 5	Prague 5, Slivenec
Prague 13	Prague 13, Řeporyje
Prague 16	Prague 16 (formerly Radotín), Lipence, Lochkov, Velká Chuchle, Zbraslav
Prague 17 (part)	Zličín
Prague 6	Prague 6, Lysolaje, Nebušice, Přední Kopanina, Suchdol
Prague 17 (part)	Prague 17 (formerly Řepy)

Prague 7	Prague 7, Troja (has been a separate municipal district since 1 January 1992)
Prague 8	Prague 8, Březiněves, Dolní Chabry, Ďáblice
Prague 9	Prague 9
Prague 14	Prague 14, Dolní Počernice
Prague 18	Prague 18 (formerly Letňany), Čakovice (has been part of administrative district 18 since 2007)
Prague 19	Prague 19 (formerly Kbely), Satalice, Vinoř,
Prague 20	Prague 20 (formerly Horní Počernice)
Prague 21	Prague 21 (formerly Újezd nad Lesy), Běchovice, Klánovice, Koloděje
Prague 10	Prague 10
Prague 11 (part)	Křeslice
Prague 15	Prague 15, Dolní Měcholupy, Dubeč, Petrovice, Štěrboholý
Prague 22	Prague 22 (formerly Uhříněves), Benice, Kolovraty, Královice, Nedvězí

I found geo info at open data server [2]. You can download data or use downloaded TMMESTSKECASTI_ansi.json.

Let's look at Area of districts. We need to use Shape Area field. I used Python to transfer json to Pandas dataframe. The biggest neighborhoods are:

Neighborhood Area

Praha 6 41.561204
Praha 5 27.510820
Praha 4 24.200364
Praha 12 23.317909
Praha 8 21.79410

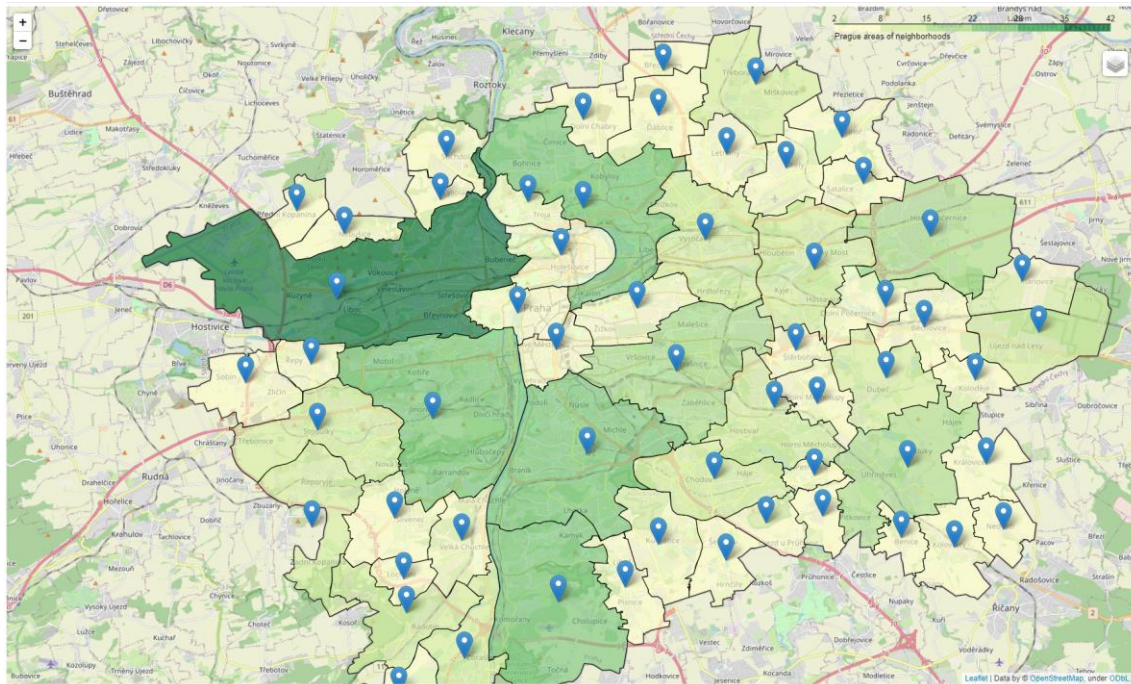
The smallest neighborhoods are:

Neighborhood Area

Petrovice 1.786429
Lysolaje 2.474802
Lochkov 2.716217
Benice 2.773783
Sterboholý 2.968908

The size of districts differs by more than 20 times. We can draw that information on a choropleth style map. These neighborhoods and markers were plotted with the help of Folium library. The dark green zone is the largest Prague 6 district on the left bank of the Vltava River. I used Google geocode API to find the coordinates of Prague.[5]

Unfortunately, folium maps are not correctly displayed through the pdf, so they can be seen in this report, although, not interactively. An interactive version of the map is available in Notebook at https://github.com/eyashin/Coursera_Capstone.



Next, we use Foursquare API to get 100 restaurants that are within a radius of 1500 meters from the center of each district. Then we make API calls to Foursquare in a Python loop. Foursquare returns the data in the JSON format and we extract the information from the following fields: neighborhood, Id, restaurant's name, restaurant's category, restaurant's latitude and longitude.

We loaded 2028 items for analysis.

Then we calculate the number of restaurants and their density for every district.

```
reports.reset_index().head()
```

	Neighborhood	Area	Location	Quantity	Density
0	Bechovice	6.834474	[50.081200160000066, 14.613675383500038]	7	1.024219
1	Benice	2.773783	[50.01399462650005, 14.602080507500062]	19	6.849852
2	Brezineves	3.380682	[50.16293176350007, 14.483752229500055]	6	1.774790
3	Cakovice	10.183716	[50.15851407000005, 14.529929156000037]	13	1.276548
4	Dablice	7.379376	[50.14854627900007, 14.481207592500056]	32	4.336410

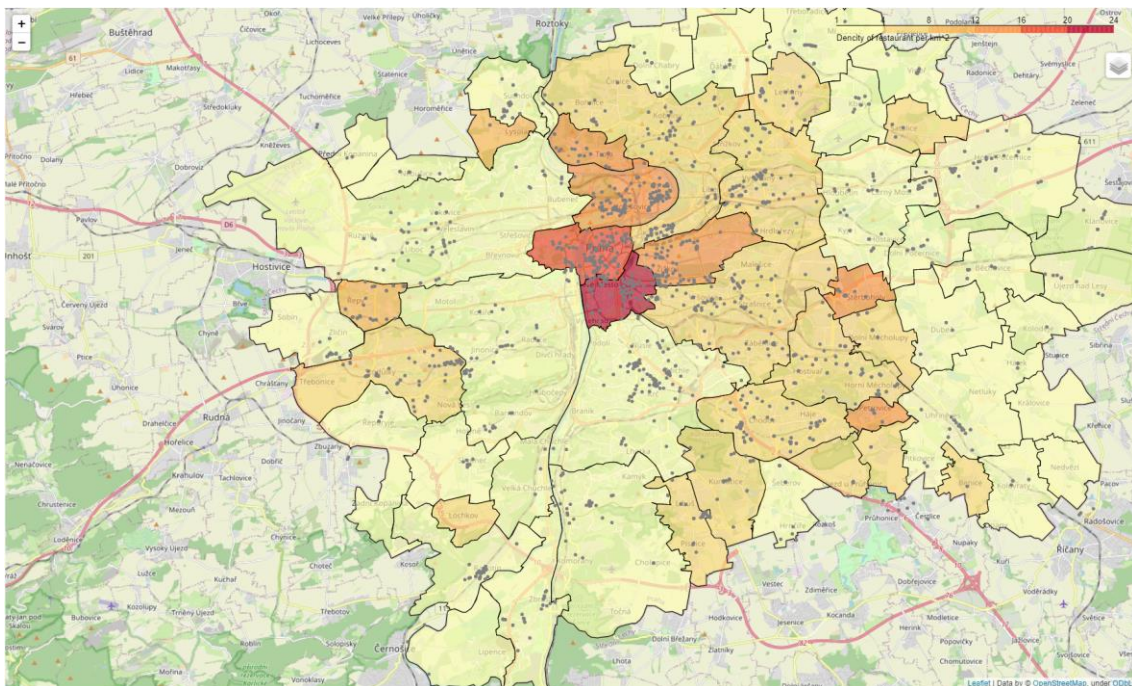
Than show top data by density and quantity

```
print('Top 5 district by Dencity')
print(reports.sort_values(by=['Dencity'], ascending = False).drop(columns=['Location']).head(10))
```

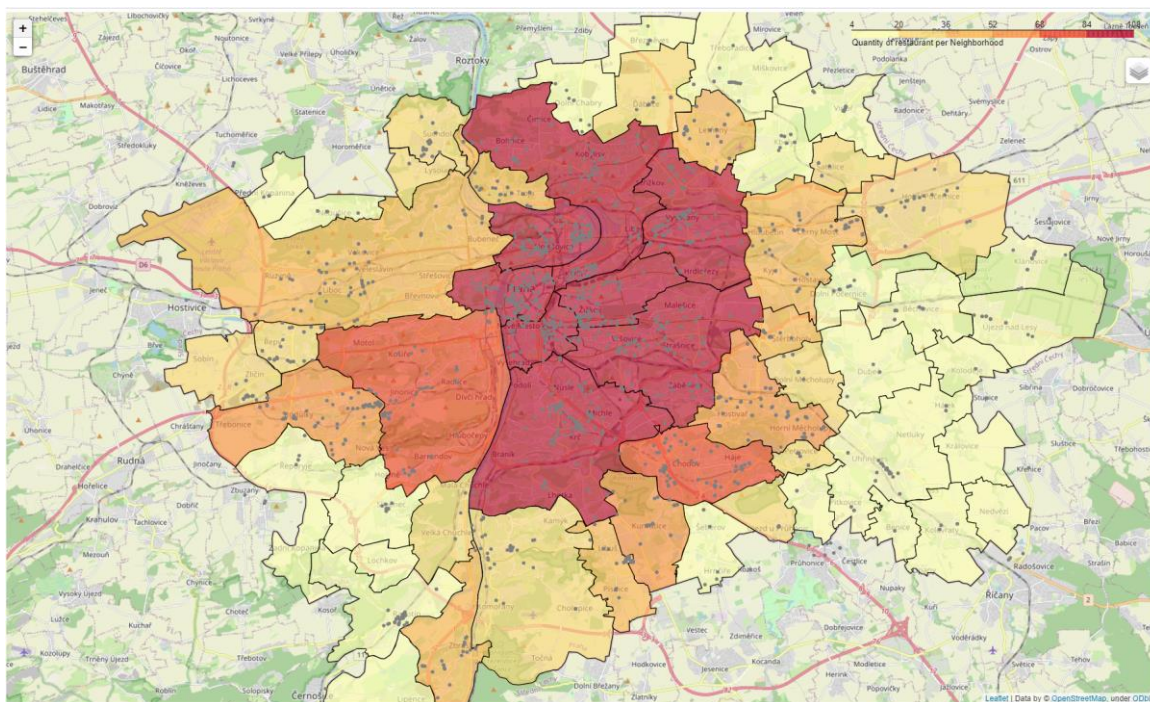
Top 5 district by Dencity			
	Area	Quantity	Density
Neighborhood			
Praha 2	4.184938	100	23.895217
Praha 1	5.538444	100	18.055613
Praha 3	6.482055	100	15.427207
Petrovice	1.786429	26	14.554173
Praha 7	7.094987	100	14.094458
Troja	3.371441	47	13.940626
Sterboholý	2.968908	37	12.462496
Lysolaje	2.474802	26	10.505890
Praha 17	3.253142	29	8.914458
Praha 18	5.606559	46	8.204676

Results:

One of the goals of my work was to visualize the Number of restaurants and their density per square kilometer with choropleth style map.



We can see that districts on the right bank of the Vltava River are more interesting for restaurant business.



The areas with the largest number of restaurants are clearly visible without clustering.

Discussion:

Our analysis shows that there is a great number of restaurants in Prague. If we take into consideration the quantity and density of restaurants the most interesting districts for business are:

- **Praha 1**
- **Praha 2**
- **Praha 3**
- **Praha 7**

Conclusion:

The purpose of this project is to identify Prague districts for distributors of food, equipment, and services. Using data analysis I have prepared several recommendations. Of course, final decision on expanding business will be made by stakeholders and will be based on the level of competition, prices and other economic factors in every recommended district.

Limitations:

In this project I used free Account of Foursquare API that came with limitations. The number of restaurants in one district is limited to 100.

References:

- [1] [Prague — Wikipedia](#)
- [2] [Prague Open Data](#)
- [3] [Forsquare API](#)
- [5] [Google Map API](#)