

# THE BATTLE OF THE NEIGHBORHOODS: ICE CREAM SHOP IN BUDAPEST, HUNGARY

Diogo Henrique da Silva Lenzi

July 09, 2020

## 1. INTRODUCTION

This report is part of the Applied Data Science Capstone course available at the Coursera Platform. This course is the ninth and final in the composition of the IBM Data Science Professional Certificate.

This project consists of two main parts that will be delivered separately:

1. Business problem and data preparation
2. Methodology, visualization, results and conclusion

### *1.1. Budapest*

Budapest is the capital and the most populous city of Hungary, and the ninth-largest city in the European Union by population within city limits. It is considered an Alpha City, in other words, a Global City with strengths in commerce, finance, media, art, fashion, research, technology, education, and entertainment. It is Hungary's financial centre and was ranked as the second fastest-developing urban economy in Europe. Budapest attracts around 12 million international tourists per year, making it a highly popular destination in Europe. The city was chosen as the Best European Destination of 2019, a major poll conducted by EBD, a tourism organisation partnering with the European Commission. It also topped the Best European Destinations 2020 list by Big7Media.

Source: [Wikipedia](#)

### *1.2. Business Problem*

Considering the great historical and tourist importance of Budapest and its economic influence, not only over Hungary but over the entire European Union, many investors and entrepreneurs have turned their attention to the region. To achieve the purposes of this project, this report is focused on presenting a study that identifies the best location for

opening a food venue in Budapest. More specifically, this study is aimed at stakeholders and investors who wish to open an Ice Cream Shop in the **busiest region of the city**.

Furthermore, a systematic solution to this problem should be presented with the fulfillment of three main requirements:

1. **Finding the busiest district in Budapest:** The interest of investors and stakeholders in identifying such location is justified considering that a busy spot brings greater movement and circulation to an establishment, therefore increases its chance of success, due brand awareness and a potential increase in clientele.
2. **Finding direct competitors in the area of the busiest district:** The goal of this requirement is to find the regions within the busiest district that are already saturated with direct competitors. This new Ice Cream Shop should be open in an area with a minimum possible competition. For this analysis, a "direct competitor" is defined as another venue that is specialized in ice cream.
3. **Finding clusters of food venues that are not direct competitors:** Most cities have regions with clusters of restaurants, bars, and pubs. These clusters are an optimal location for this Ice Cream Shop, considering that after a proper meal, people tend to look forward to a great dessert.

### **1.3. Interest**

Briefly, the Business Problem is finding a place with a great movement of people, surrounded by food venues, with a low incidence of direct competitors. Fulfilling the upper mentioned requirements is the goal of this study, which is beneficial to the stakeholders and investors of this new Ice Cream Venue. Assertiveness in the location choice of this venue is a keen factor to its future success.

## **2. DATA ACQUISITION AND PREPARATION**

Based on the business problem's definition, the main factors that will influence the decision-making of the stakeholders are:

1. The number of people per square kilometer (population density) of the neighborhood
2. The number of direct competitors (Ice Cream Shops) within the densest neighborhood
3. The existence of many other types of food providers within the densest neighborhood

## 2.1. Data Sources

### DEMOGRAPHIC DATA

Data on the population density of each district of Budapest can be found on Wikipedia:

- Article Name: [Budapest](#)
- Article Name: [List Of Districts In Budapest](#)

### GEODATA

- JSON and GeoJSON files for mapping purposes can be found in this [GitHub Repository](#)
- The [Foursquare API](#) will be providing the geospatial data of venues.

### BUSINESS DATA

- The [Foursquare API](#) also will be providing the data of venue categories, competitors, and food-related businesses.

## 2.2. Data Preparation

In this stage data was scraped from the sources upper mentioned to create a unified data frame with all the relevant information to this study.

	geometry	name	official_name	population	area_km2	density
0	MULTIPOLYGON (((19.02373 47.49557, 19.02374 47...	Budapest I. kerület	Vákerület	25196	3.41	7388.8
1	MULTIPOLYGON (((18.92617 47.54345, 18.92979 47...	Budapest II. kerület	Rózsadomb	89903	36.34	2473.9
2	MULTIPOLYGON (((18.97760 47.57392, 18.97721 47...	Budapest III. kerület	Óbuda-Békásmegyer	130415	39.70	3285.0
3	MULTIPOLYGON (((19.06720 47.56314, 19.06741 47...	Budapest IV. kerület	Újpest	101558	18.82	5396.2
4	MULTIPOLYGON (((19.04202 47.51471, 19.04283 47...	Budapest V. kerület	Belváros-Lipótvaros	26284	2.59	10148.2
5	MULTIPOLYGON (((19.05482 47.49917, 19.05483 47...	Budapest VI. kerület	Terézváros	38504	2.38	16178.1
6	MULTIPOLYGON (((19.05519 47.49754, 19.05621 47...	Budapest VII. kerület	Erzsébetváros	53381	2.09	25541.1
7	MULTIPOLYGON (((19.06000 47.49435, 19.06014 47...	Budapest VIII. kerület	Józsefváros	76811	6.85	11213.2
8	MULTIPOLYGON (((19.05481 47.48579, 19.05626 47...	Budapest IX. kerület	Ferencváros	59056	12.53	4713.1
9	MULTIPOLYGON (((19.09867 47.47562, 19.10196 47...	Budapest X. kerület	Kőbánya	78414	32.49	2413.4
10	MULTIPOLYGON (((18.97028 47.47607, 18.97029 47...	Budapest XI. kerület	Újbuda	151812	33.49	4533.0
11	MULTIPOLYGON (((18.93483 47.53359, 18.94673 47...	Budapest XII. kerület	Hegyvidék	58171	26.67	2181.1
12	MULTIPOLYGON (((19.04064 47.52093, 19.04202 47...	Budapest XIII. kerület	Angyalföld-Újlípótváros-Vizafogó	120256	13.43	8954.2
13	MULTIPOLYGON (((19.07240 47.51858, 19.07252 47...	Budapest XIV. kerület	Zugló	124841	18.13	6885.8
14	MULTIPOLYGON (((19.10273 47.54232, 19.10325 47...	Budapest XV. kerület	Rákospalota-Pestújhely-Újpalota	80573	26.94	2990.8
15	MULTIPOLYGON (((19.14546 47.52587, 19.15239 47...	Budapest XVI. kerület	Árpádföld-Cinkota-Mátyásföld-Sashalom-Rákossze...	73486	33.51	2192.9
16	MULTIPOLYGON (((19.19625 47.46272, 19.20272 47...	Budapest XVII. kerület	Rákosmente	87793	54.82	1601.4
17	MULTIPOLYGON (((19.15401 47.44476, 19.15636 47...	Budapest XVIII. kerület	Pestszentlőrinc-Pestszentimre	101738	38.60	2635.6
18	MULTIPOLYGON (((19.11394 47.45502, 19.11412 47...	Budapest XIX. kerület	Kispest	60731	9.38	6474.5
19	MULTIPOLYGON (((19.08789 47.43996, 19.08900 47...	Budapest XX. kerület	Pesterzsébet	65321	12.19	5385.5
20	MULTIPOLYGON (((19.01895 47.39010, 19.02805 47...	Budapest XXI. kerület	Csepel	76911	25.75	2985.8
21	MULTIPOLYGON (((18.95159 47.39394, 18.95232 47...	Budapest XXII. kerület	Budafok-Tétény	54611	34.25	1594.4
22	MULTIPOLYGON (((19.09132 47.42850, 19.09149 47...	Budapest XXIII. kerület	Soroksár	23641	40.77	579.8

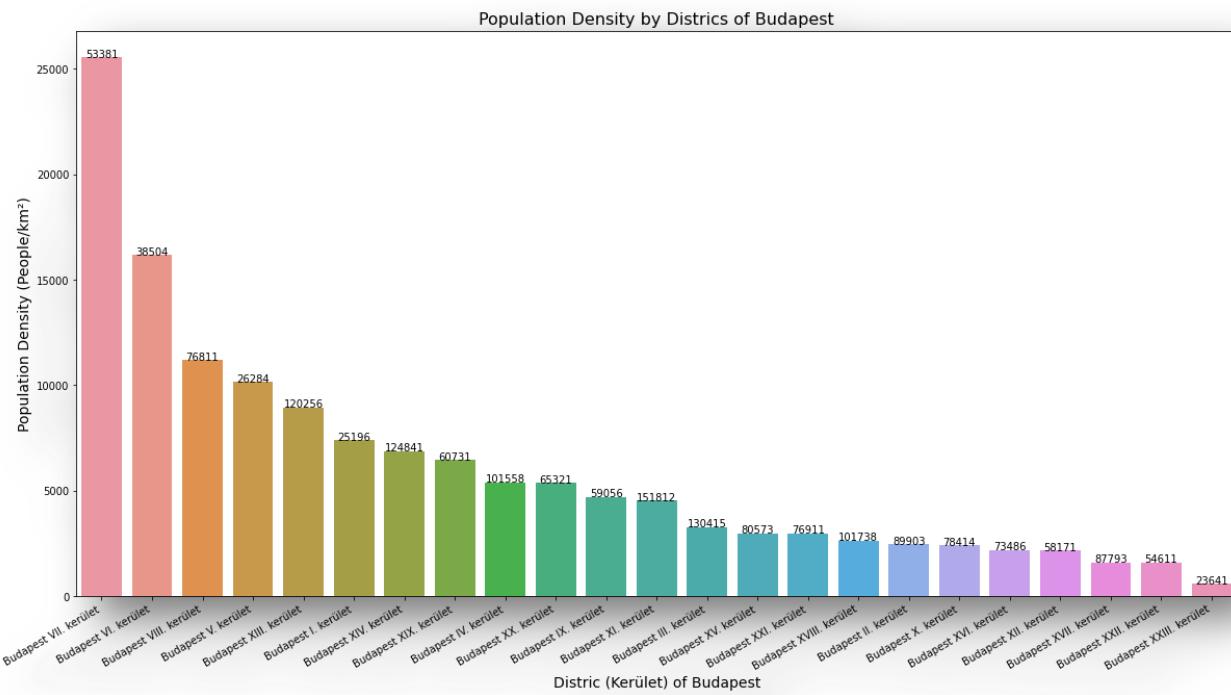
### 3. METHODOLOGY

One of the most important things to initiate an analysis is to understand the data at reach. An excellent way to do that is to visualize it throughout graphics and maps. The target is to fulfill the requirements presented previously.

#### 3.1. Finding The Busiest District In Budapest

In this stage, it will be considered the "busiest district" as the one that has more persons per kilometer square, in other words, the one that has the higher population density. To firstly identify visually which district was the most demographically dense, a Bar Plot was generated.

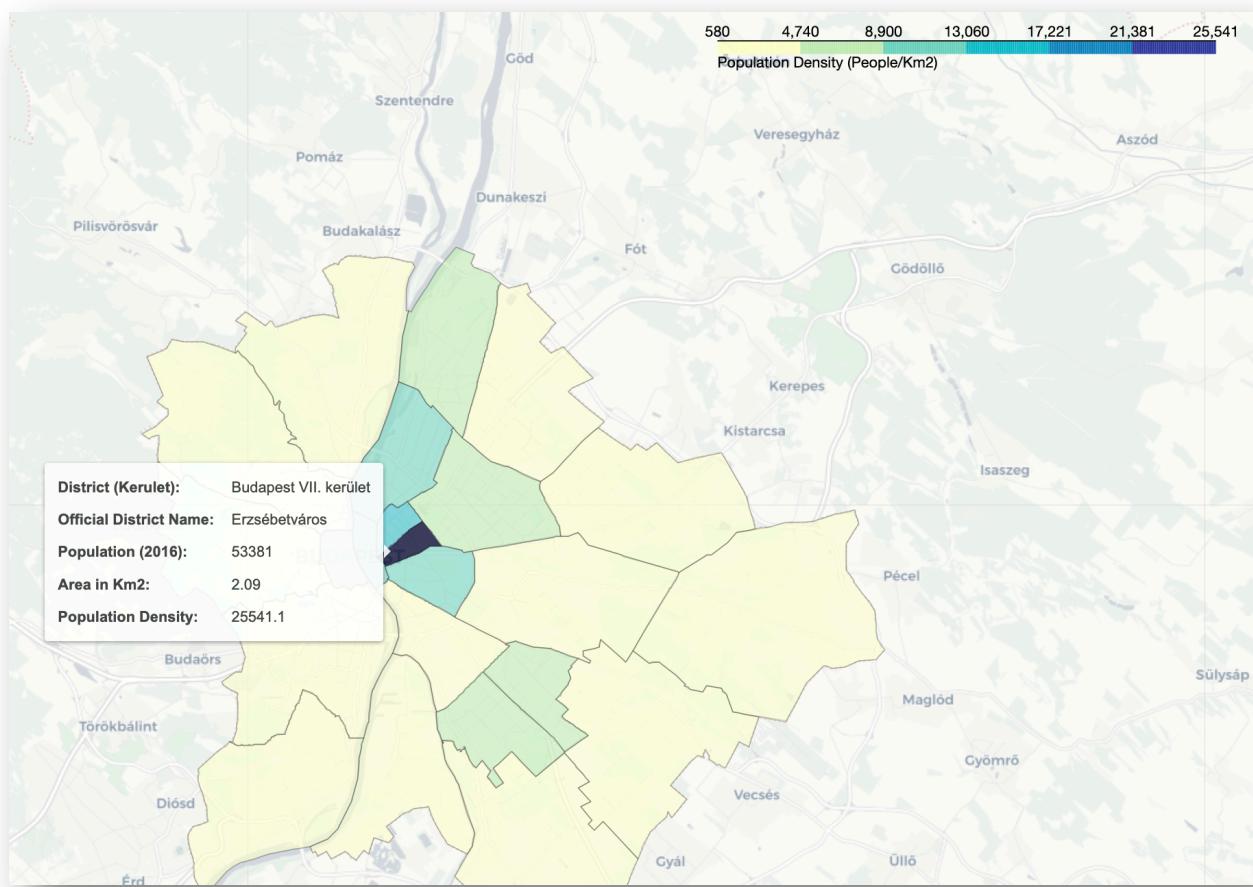
#### BAR PLOT



By analyzing this graphic, it is possible to visualize that the VII. District, also locally known as Erzsébetváros, is by far the densest district in the city of Budapest. Although this information is already sufficient to fulfill the first requirement of the solution, this form of visualization presents limitations on insights. Thus, a better form of presentation is now required.

Secondly, a more interactive tool was developed, choropleth map.

## CHOROPLETH MAP

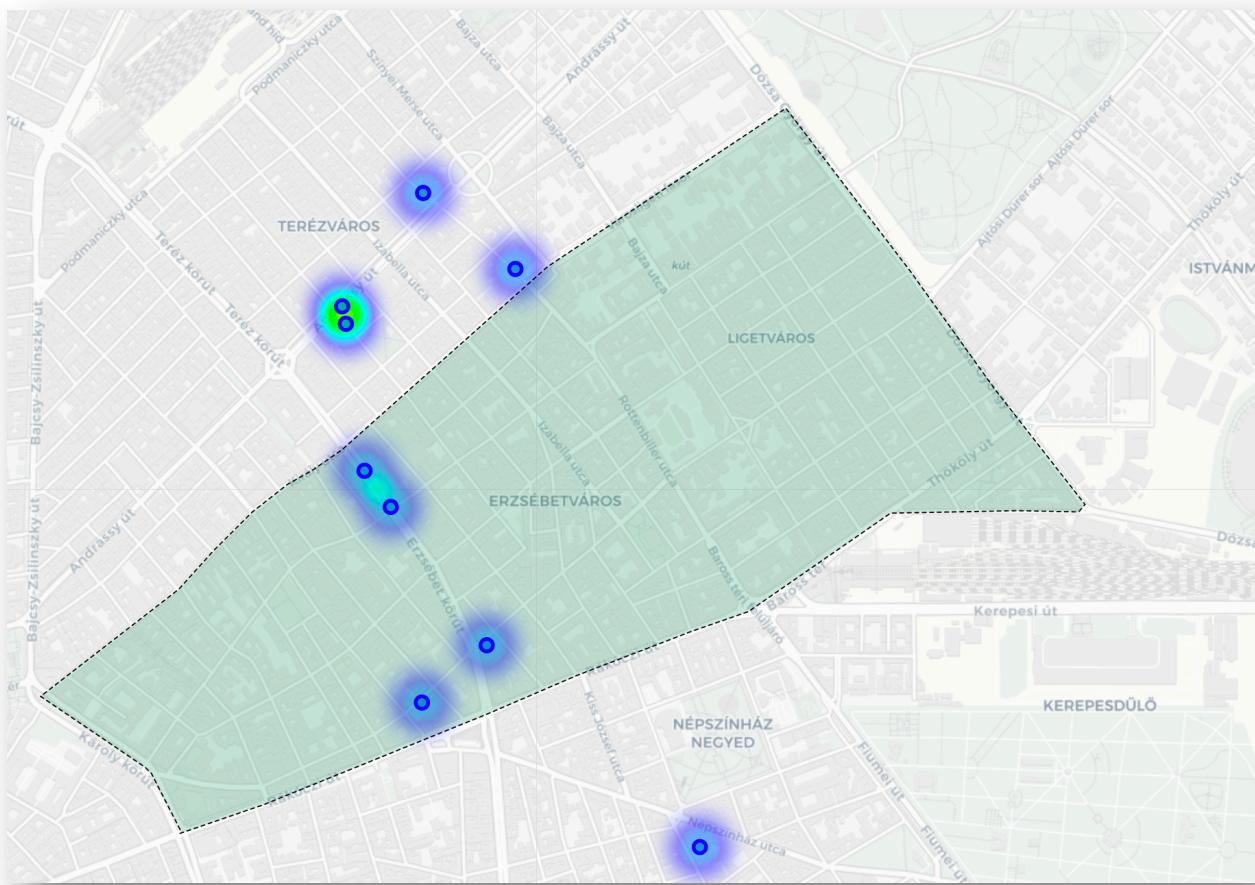


This category of map presents a much better and comprehensive visualization, on which the XII. District of Budapest is displayed as the neighborhood with the higher population density in the city. The map shows all the demographic data of each state when the viewer moves the cursor above it.

### 3.2. Finding Direct Competitors In The Area Of The Busiest District

Now that the busiest district was found, according to the definitions established, for this stage of the report, the usage of the Foursquare API is requested. As the XII. District is relatively small in size, with an area of only 2.09 km<sup>2</sup>, as seen in the previous map, the search parameters of the Foursquare API will be configured for a radius of 1 kilometer. In this way, the search for competitors will be limited to the surrounding neighborhood in focus.

## DIRECT COMPETITORS

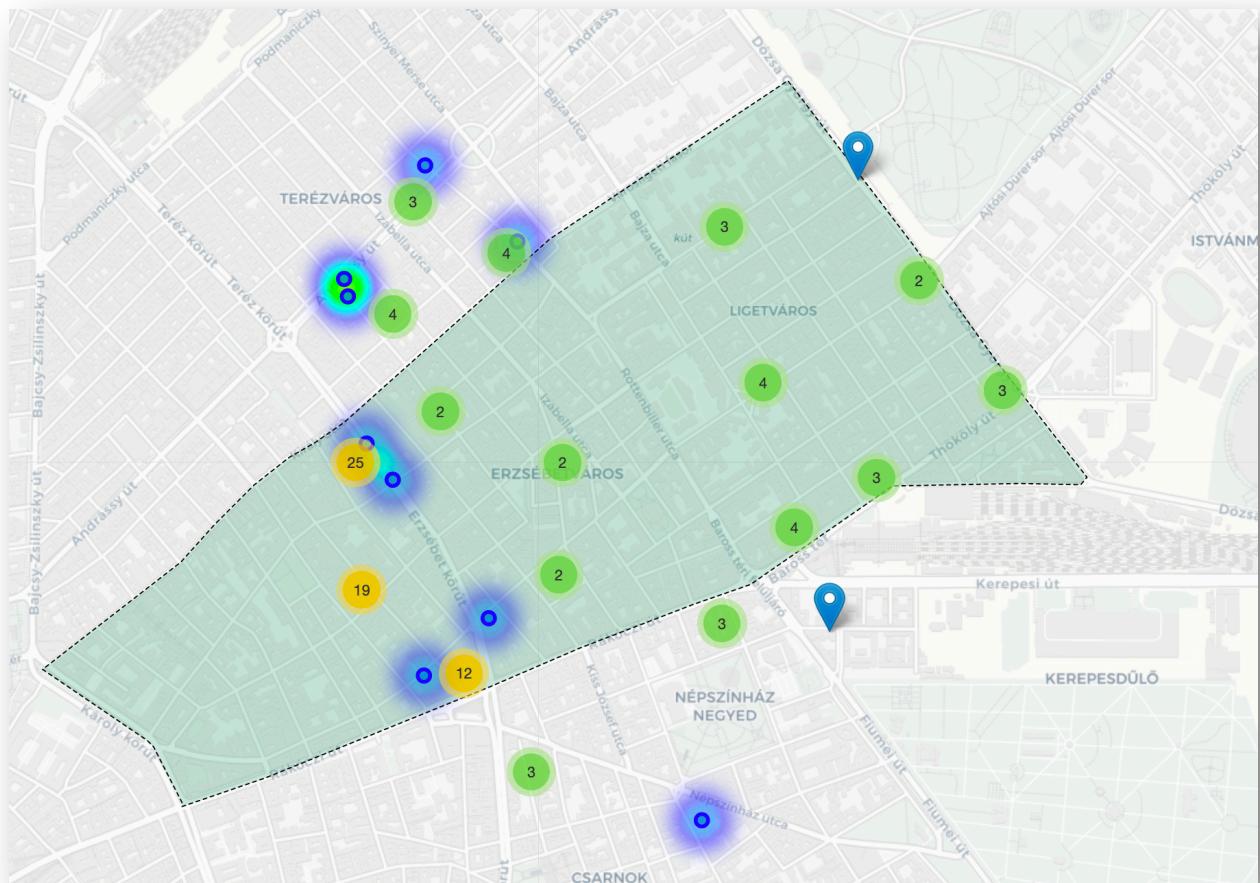


As can be seen, the map above shows all the Ice Cream venues that are within a one-kilometer radius from the centroid of the XII. District, successfully completing the second requirement of the solution needed. Now it is time to work towards the identification of clusters of food venues in this same area.

### 3.3. Finding Clusters Of Food Venues That Are Not Direct Competitors

Similarly to the previous section, in this analysis we will be using the Foursquare API, however this time with a broader parameter of category, aiming to return all food venues in the area, with of course the exception of the Ice Cream Venues already mapped.

## DIRECT COMPETITORS AND CLUSTERS OF OTHER FOOD VENUES

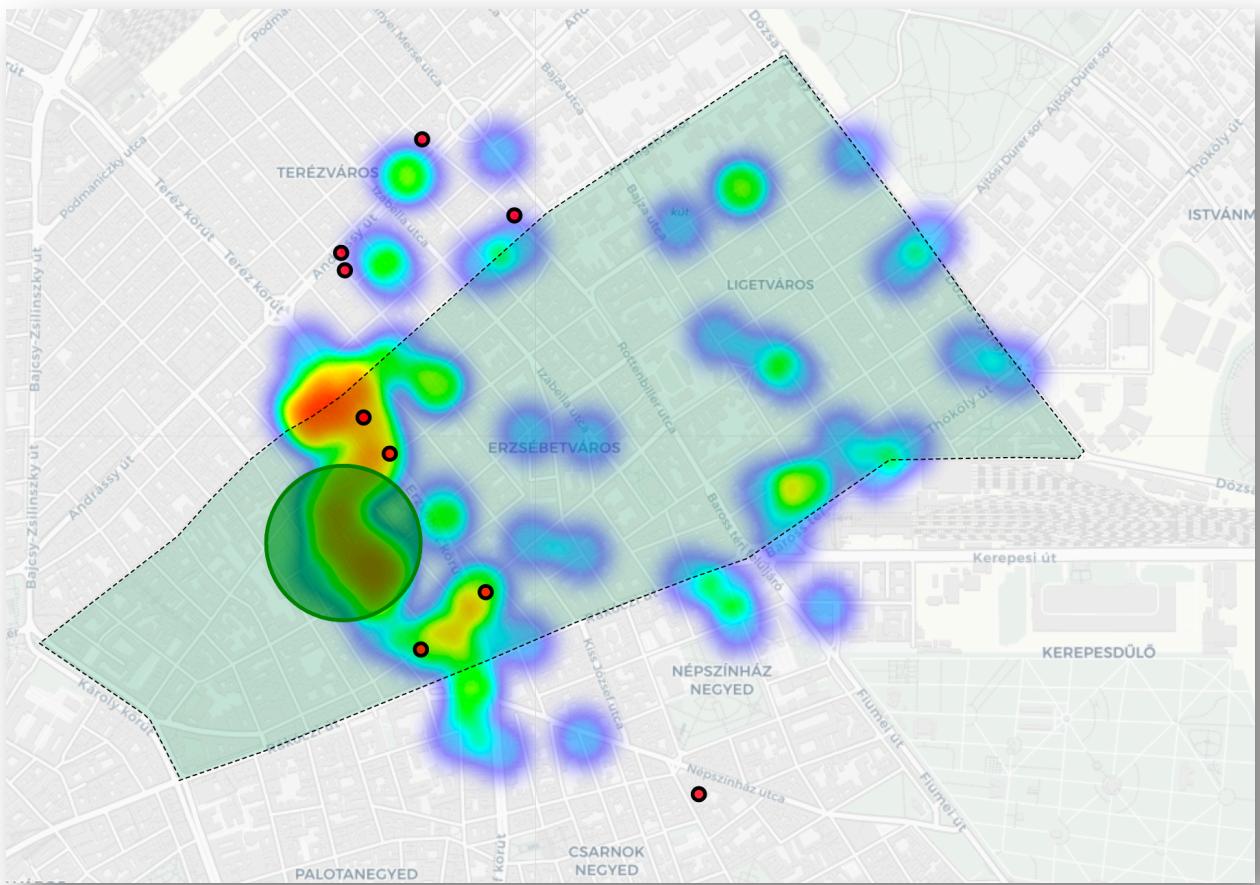


As shown above, the map with both Direct Competitors and Other Food Venues is complete. The viewer can investigate and choose layers on the map. At the default zoom, it is possible to identify three important clusters of other food venues.

## 4. RESULTS

Now that the analysis is completed and all the three requirements were successfully fulfilled, it is possible to create a final map with the optimal area for a new Ice Cream Shop.

**IDEAL AREA FOR A NEW ICE CREAM SHOP**



Changing the visualization of the other food venues from clusters to heat spots, a more comprehensive view of the possible places for the new Ice Cream Shop was generated. The map also shows the competitors on red circle markers, and for the ideal location, a green circle marker was added, covering an area that complies with all the established requirements.

## 5. DISCUSSION

As presented in the final map, most food-related venues within the VII. District of Budapest is concentrated in one half of the neighborhood. More specifically, it is possible to see that venues that directly serve the public are concentrated closer to the very center of Budapest, by the Danube River.

Another interesting insight is that little direct competitors were found throughout this analysis. A plausible explanation is that instead of having specialized venues that serve restrictively ice cream, it should be considered a more broaden category of venue, serving all sorts of desserts.

Finally, exploring the datasets and information in this notebook may present some good insights, however, finding a good location does not mean acquiring it. Real estate availability and budget planning are also crucial factors to achieve a more accurate result.

## 6. CONCLUSION

This study demonstrates the importance of in-depth analysis and an appropriate market study for the opening of a new business. The use of the tools shown here is extremely relevant, not only for academic and scientific circles but also for the expansion and refinement of business strategies.

It can be said that the results of this analysis were satisfactory in the spectrum of the demonstration of all the tools and applications learned in the course of the specialization in Data Science by IBM, through the Coursera learning platform. However, it is worth mentioning that the pre-established requirements to identify the ideal area for the new business could be expanded, generating a more in-depth analysis.

For future work, it is suggested to include other factors that can interfere directly with the incidence of potential clientele on the venue, such as the existence of tourist spots or metro stations nearby. Also, a more appropriate investigation of the product could be initiated, such as classifying flavor appreciation. Furthermore, an in-depth analysis of the competitors could be implemented, even using the same tools utilized in this work. For instance, a word cloud could be generated using the Foursquare API, showing the menu of the direct competitors.

It is also worth mentioning that the data acquired for this analysis were taken from an outsourced and non-governmental platform, and may present unreliable and/or outdated information. For the most reliable and professional application of such analyzes and tools, access to proprietary data portals to explored entities is recommended, such as open country, state or municipality data, or authorized institutions.