# Project Title

**Find the location spot for Café in Kyiv, Ukraine**

# Introduction/Business Problem section

*Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem*

Ukrainian people love to drink coffee. Especially those ones who live in big cities like Kyiv and used to work in the offices.

Kyiv is the capital of Ukraine, the multimillion city and the competition in the market of HoReCa is very tense. In addition to “traditional” restaurants and cafés that serve all kinds of coffee, there are so-called “moving coffee points” (automobiles redesigned to operate as cafés and equipped with professional coffee machines).

My friend is intended to open a small new brick&wall café. And I aim this research to find the best place for him to settle his new business. The ideal location should have a high people traffic (next to office centres, or even better if consider work-from-home trend which is to in 2021 – next to supermarkets in so-called “sleeping areas”. Of cause, the ideal spot should not be crowded with competitors (other brick&wall cafes and “moving coffee points”), so we probably should avoid big shopping malls and metro stations where they usually flourish.

# Data section

*Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.*

Unfortunately, I have not found ready-to-use tabular information about Kyiv City boroughs, neighborhoods, or even postal coded. All the information I have found was in Cyrillic, but to match the list of postal codes / addresses with corresponding geospatial information (i.e. latitude, longitude) data should be re-translated into Latin alphabet. The problem is that each geo map system has its own version of transliteration from Ukrainian Cyrillic into Latin alphabet. Thus, for several borough names I literally had to iterate over all options of spelling to find the right one that has the match to proper address in geo map and allows to retrieve latitude and longitude.

The only list of Kyiv boroughs (historic regions) I have found in latin, was from this source:

<https://en.wikipedia.org/wiki/Subdivisions_of_Kyiv>

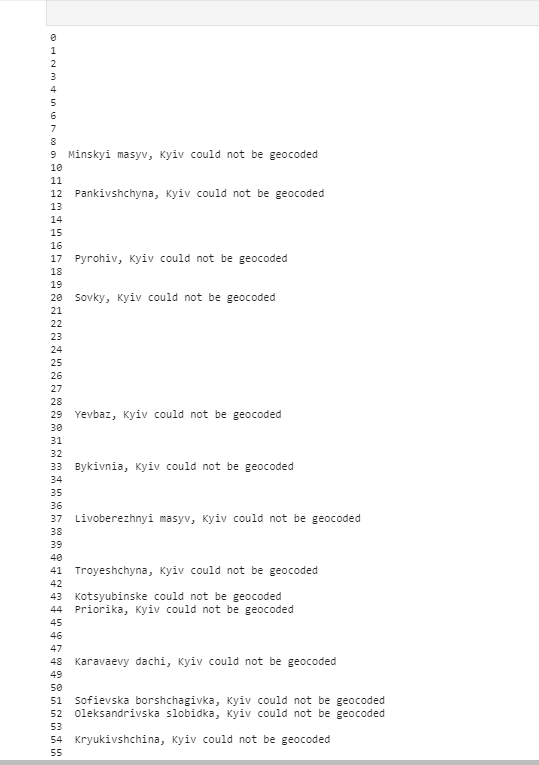
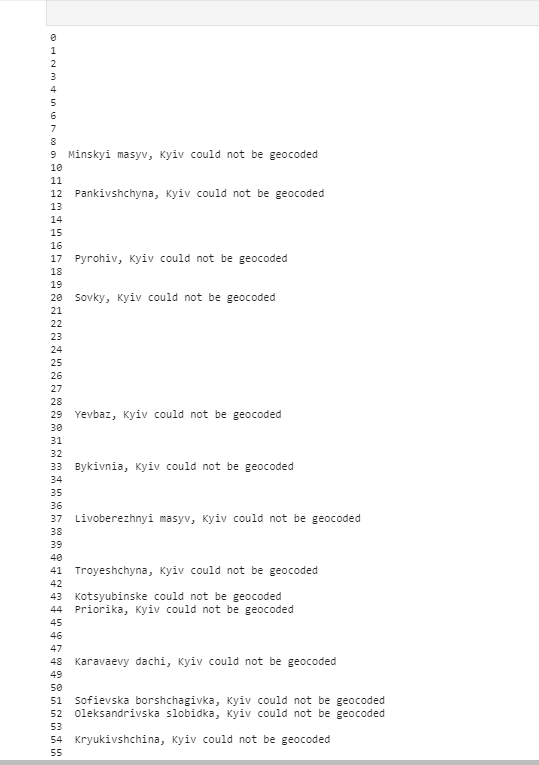
So, I used the list of boroughs from Wikipedia as the base and topped up manually with missing borough (fortunately, I live in Kyiv and know the city well 😊).

To match the list of Kyiv City boroughs with latitude and longitude coordinates I used GeoPy.

While retrieving geo information, I’ve made the check if there's a data associated with the borough name, and for those ***\_Borough\_*** that don't have geo information in geopy, filled in the cells in dataframe as ***\_NaN\_:***

******

Results of the code run:

******

As I mentioned earlier, for those \_Borough\_ that didn’t find the match in geopy, I iterated manually to find out the right Latin spelling of the Borough name.

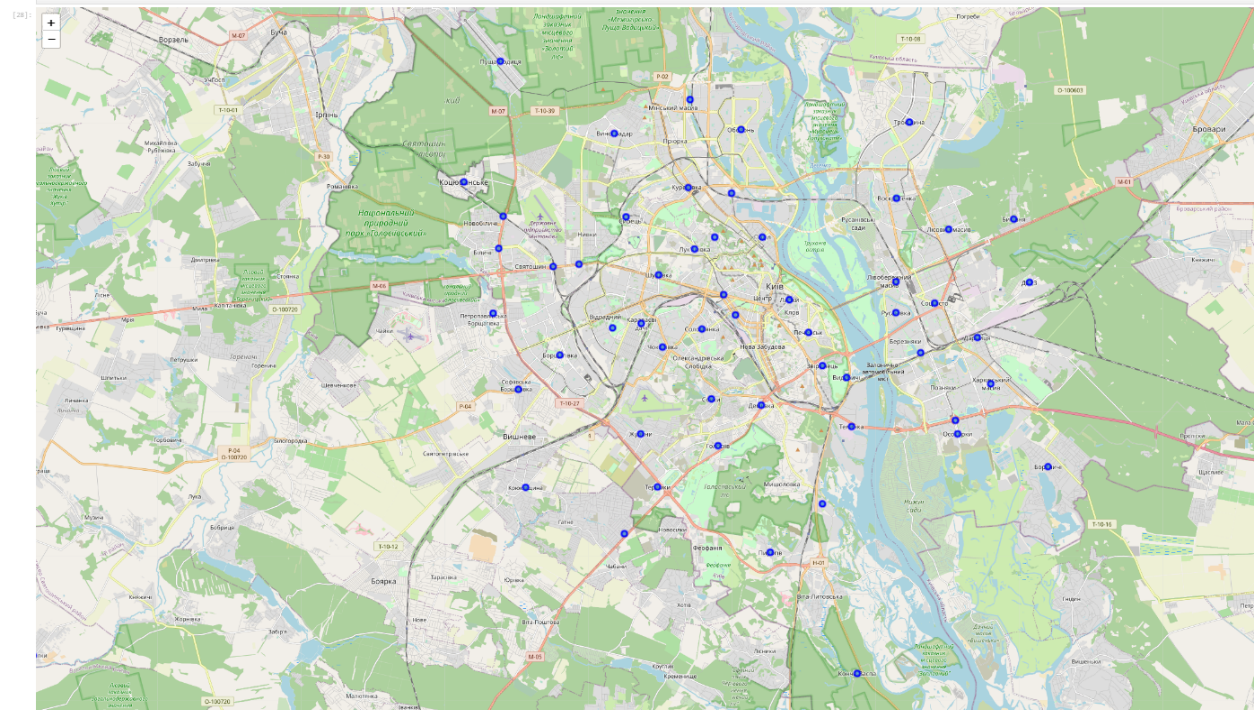
And, finally, I created the following dataframe consisting of 56 Kyiv city boroughs:

# Methodology section

*Discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why*

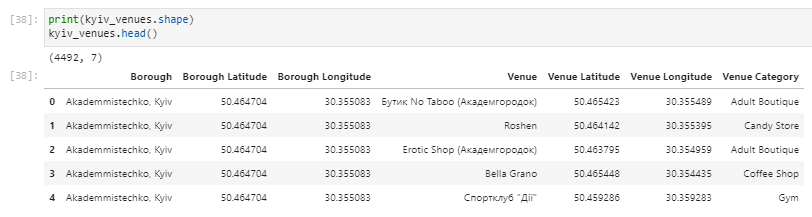
Firstly, I used python folium library to visualize Kyiv City boroughs dataframe and plotted them on the map. I used latitude and longitude values to center the boroughs’ on the map:



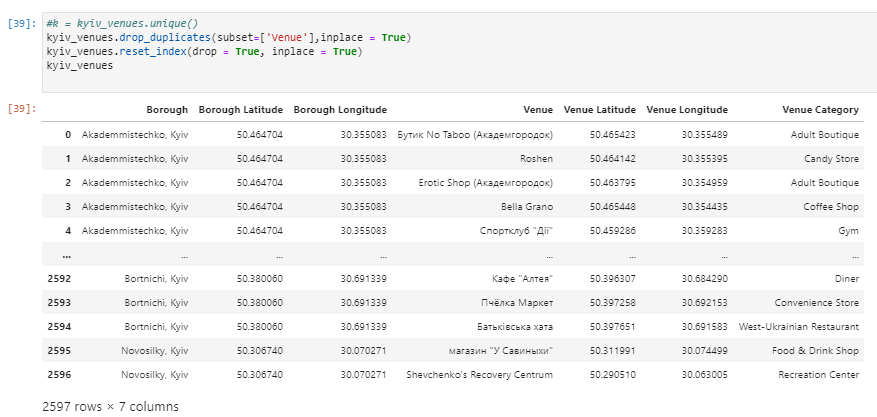
Secondly, I used Forsquare API to get the most common venues of the 1st borough in the dataframe (Akademmistechko). In fact, it is the borough where I live, so this step allowed me to examine how full is the Foursquare data for Kyiv.

In the radius of 500 meters around the Akademmistechko center, Forsquare API has found only 24 locations. Which is very poor coverage of real-life list of venues. But according to the task description, we are obliged to use Forsquare API, so I decided for further analysis of all Kyiv boroughs to widen the radius of venue search up to 2000 meters.

Finally, I have created the dataframe of all velues available for all Kyiv boroughs in my dataframe. List of venues included 4492 spots.



I checked for duplicated venues and dropped it from the dataframe. Thus, the final dataframe accounted for 2597 venues which fall into 341 unique categories.:



Next, I have displayed the top 10 venues for each borough. Not surprisingly, for 27 out of 56 boroughs cafés and coffee shops were the 1st most common venue in the borough. 3 more boroughs had restaurants\shopping mall as the 1st most common venue.

That still left me 26 boroughs for further consideration…

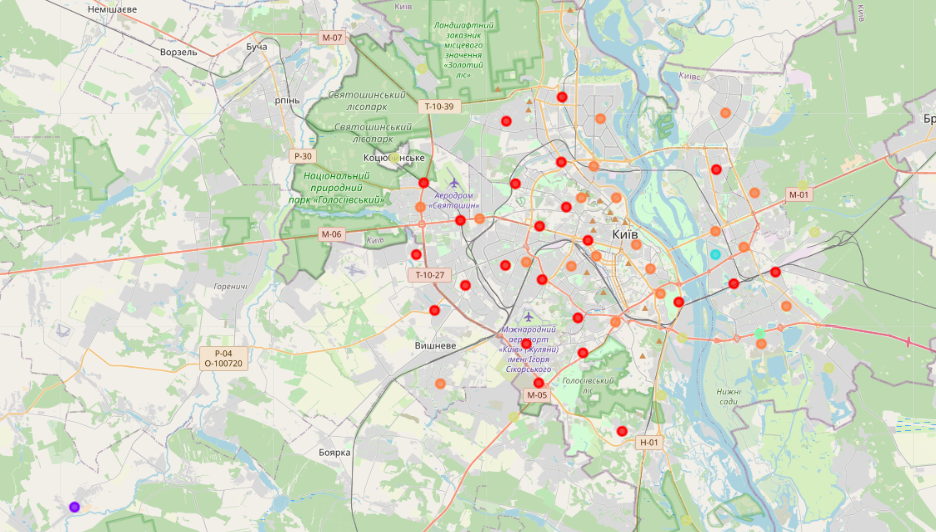
Finally, I decided to cluster boroughs to see the hidden patterns that can help me to eliminate unproper boroughs from the list of 26 boughs left for consideration.

To cluster Kyiv boroughs I used K-Means algorithm. I tested the Elbow point, and find out the optimum k value = 7.

# Results section

*Discuss the results*

K-Mean clustering with optimal k value showed following clusters on the map:

**

Clustering algorithm divided all 56 Kyiv boroughs into 7 cluster:

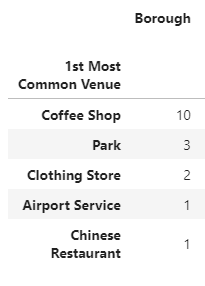
* 23 boroughs in the 1st cluster
* 1 borough per cluster 2 – 5
* 9 boroughs in the 6th cluster
* 20 boroughs in the 7th cluster

Let’s discuss the peculiarities of each cluster in the next section

# Discussion section

*Discuss any observations you noted and any recommendations you can make based on the results*

**1st Cluster. Consists of 23 boroughs, 17 of which are crowded with coffee shops:**

**

Boroughs in the 1st cluster are mostly “old”, inhabited and has high competition for a new coffee shop. In addition, according to my knowledge of Kyiv, these boroughs are the preferred place of living for blue-collar workers who spend all day in the working place in other borough, even in the face of COVID-19 pandemic and lockdown.

That’s why I don’t consider the boroughs from this cluster as the promising locations for opening a new small Café

**2nd Cluster. Contains 1 distant borough - Novosilky, Fastiv.**

It is Kyiv urban area, where blue-collar workers mainly live. They don't work-from-home thus not spending day-time in the resident area



That’s why I don’t recommend the boroughs from this cluster to be considered as the promising locations for opening a new small Café

**3rd Cluster. Contains 1 borough – Chaiky**

It is also Kyiv sub-urban area, but not so distant as the 2nd cluster.

Besides, it is very fast developing area of Kyiv, has very fast and convenient transport connection to Kyiv.

In addition, Chaiky is very green (lots of parks and fresh air) and comfortable for young families with small kids.



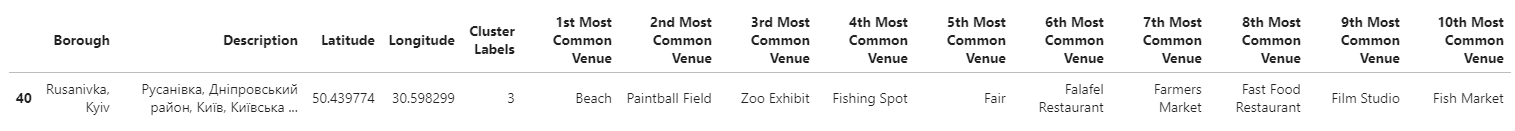
To summarize, this borough can be very promising location for opening a new Café

**4th Cluster. Contains 1 borough – Rusanivka - Very promising borough to settle a new Cafe:**

1. It is situated almost in the Kyiv city center, on the bench of Dniper River, and has astonishing views on the Kyiv downtown and sunsets

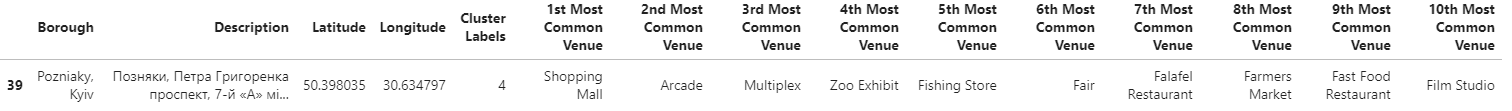
2. This borough has dense building, and still in active development. The cost of apartments for rent and purchase is above average. So Rusanivka is populated with well-to-do people, who work hard and need to recreate and enjoy the life on the river's bank (of cause in the company of their beloved coffee drink!)

3. Fortunately, Foursquare don't show many Cafes and coffeeshops in the borough, which can be a great opportunity for new business spot opening!



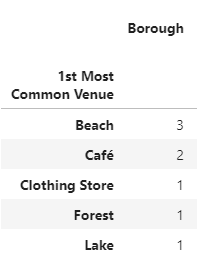
**5th Cluster. Includes a single yet the biggest borough of Kyiv - Poznyaky**

Poznyaky has very dense building and crowded with people living there. But the density of Cafes, restaurants and shopping malls is also extremely high. In addition, people from this borough usually recreate by visiting shopping malls, not walking. And the rent cost of space in shopping mall is unaffordable for small Café



**6th Cluster. Includes 9 boroughs located in Kyiv outskirts**

Beach & Cafe are the most popular venues for this cluster of boroughs:

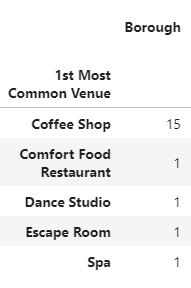


Considering their distant location and not very dense population, I assume that the competition of coffee restaurants per 1000 people for this cluster is relatively high

So, I decided not no recommend boroughs from this cluster to be considered for possible locations of future small Café

**7th Cluster. Consists of 20 boroughs**

All boroughs from this cluster has well-developed infrastructure. And the competition for the new Cafe is the sky-high: Coffee shop & restaurants is the 1st most popular venue for 16 boroughs out of 20. Thus, I would call this cluster of boroughs as the "red ocean" for coffee shops:



# Conclusion section

*Concludes the report*

Thanks to K-Means clustering algorithm results, the consideration list of boroughs to open a small new Café has narrowed from 26 to only 2 boroughs: **Chaiky** and **Rusanivka. Both of them looks very promising** … but remembering that completeness of Foursquare data for Kyiv City leaves much to be desired, both boroughs should be thoroughly explored on the place.