Where to live in Kyiv? Apartment rent prices, population density and venues analysis of Kyiv districts.

Oleksandr Tsapin

December 14, 2020

Introduction/Business Problem

Kyiv, the capital of Ukraine, is one of the biggest cities in Eastern Europe and the center of business in Ukraine. The city has a population of about 3 mln people, out of which about 2/3 are local work migrants. People from all over the country are trying to find work here, which resulted in a near double population increase in the metropolitan area from 2.5 to 5 mln during the last 10 years.

The first step for the local migrants is to find a place to live, which at least meets their minimum criteria. Often this is a big problem for people who are new in the city. Thus, the aim of this project is to help local work migrants to soften the problem of relocation to Kyiv. We implement this by giving the chance to look at the Kyiv map based on possible apartment search criteria, namely - wallet size (or apartment rent price), population density in different city districts and available infrastructure around specific sub-districts.

Data

For this project, we need the following set of data: a list of Kyiv districts and sub-districts, their locations, rent prices, population density, list of venues in different sub-districts. List of districts and sub-districts we downloaded from Wikipedia. Geospatial location data we received from OpenStreetMap service using geopy Python client. We use the SV-Development website to get the average rent prices. From the city administration website, we downloaded the information on population density. Finally, we use Foursquare location data to download a list of venues in a radius of 500 meters around each sub-district.

Kyiv consists of 10 districts (borough), and 200 sub-districts. In the report we explore only 166 sub-districts as 34 of the observations were lost during the data cleaning process.

First, using the data mentioned above, our goal is to explore the diversity of Kyiv 10 districts based on the average apartment rent price and different population densities. Second, we want to compare all sub-districts and determine how similar or dissimilar they are in terms of available infrastructure. Third, we show all this information on the Kyiv map to simplify the decision process for our imaginary work migrant.

Methodology & Data Analysis

In this project, we are focusing our efforts on the difference of districts in Kyiv in terms of rent price and population density, and similarities or dissimilarities of Kyiv sub-districts in terms of venues. Unfortunately, the availability of data limits us to use rent price and population density only for the districts but not for more narrowed sub-districts.

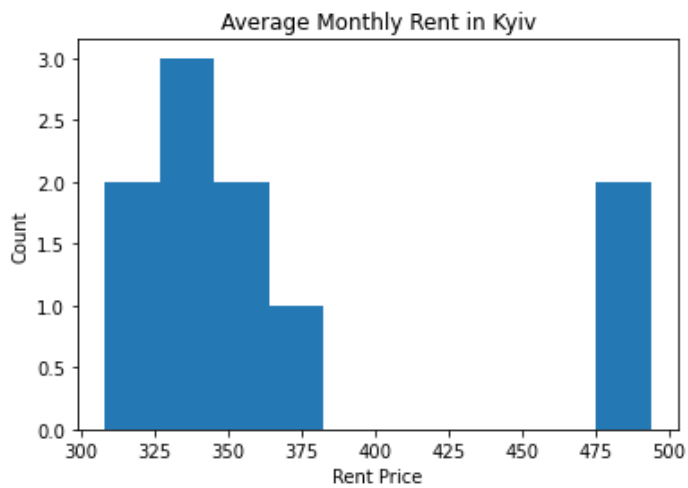
As a first step, we collected the required data - list of all districts in Kyiv with their locations, average monthly rent price for apartments and average population density. The location data we downloaded from OpenStreetMap service using geopy Python client.

Table 1. Kyiv Districts: Rent Price and Population Density

	DISTRICT	AVERAGE_MONTHLY_RENT	POPULATION_DENSITY	PRICE_GROUP
0	Golosiivskyi district	329	1516	Low
1	Solomianskyi district	352	8653	Low+
2	Sviatoshynskyi district	333	3075	Low
3	Darnytskyi district	308	2420	Low
4	Shevchenkivskyi district	493	8482	High+
5	Pecherskyi district	476	7439	High+
6	Dniprovskyi district	324	5075	Low
7	Desnianskyi district	339	2456	Low+
8	Obolonskyi district	377	2886	Medium
9	Podilskyi district	360	5715	Low+

There are 10 districts in Kyiv, for 70% of which average rent price is USD 300-350 per month. We see the significant gap between prices in two districts versus all others - very expensive Pecherskyi and Shevchenkivskyi districts - where rent cost starts at USD 470+.

Chart 1. Average Monthly Rent Price in Kyiv

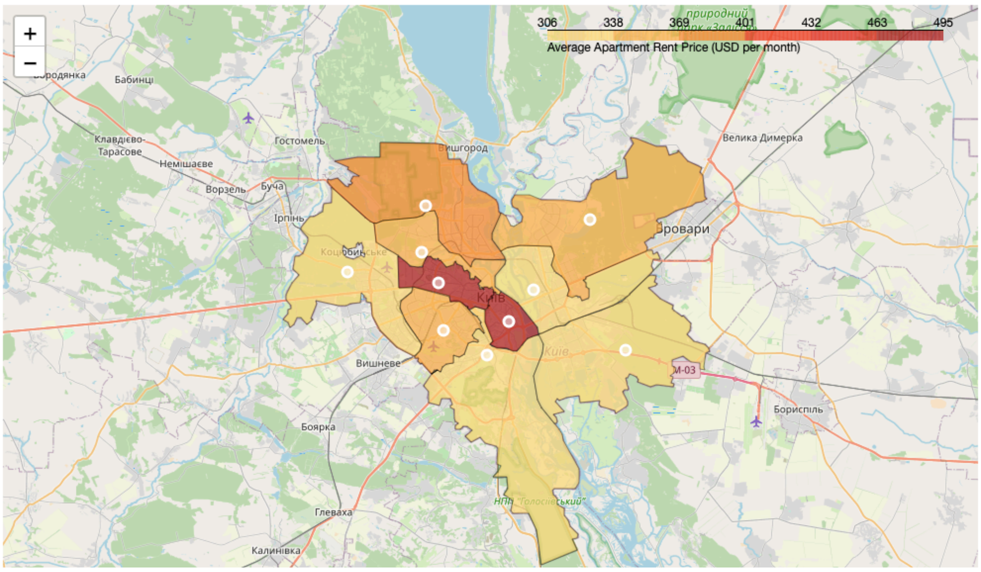


We grouped all the districts into 6 groups based on rent prices: Low, Low+, Medium, Medium+, High, High+. As we see, 7 out of 10 districts, lined in Low/Low+ price group. They are located on both sides of the Dnipro river in the west and south-east parts of Kyiv.

Chart 2. Rent Price Groups

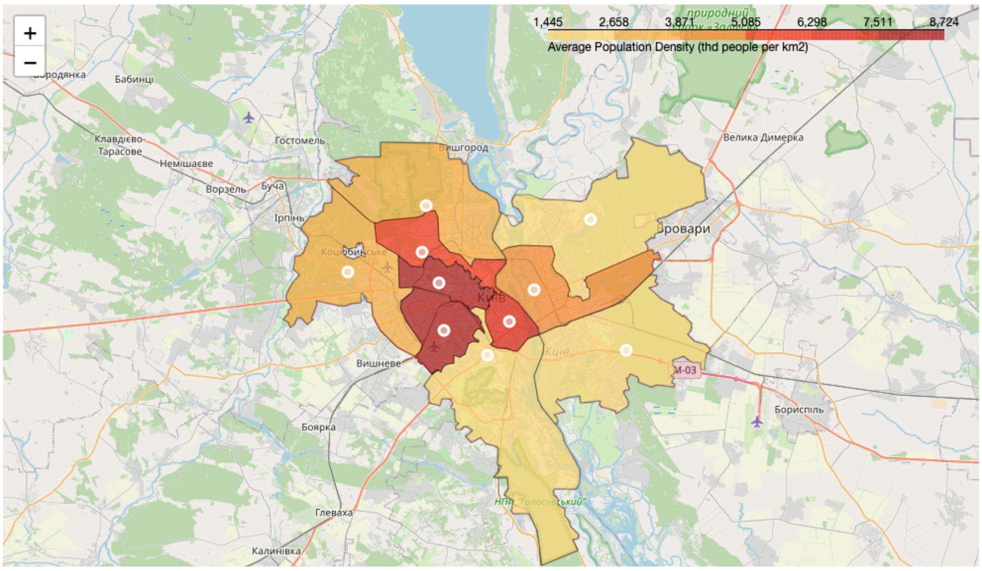
Let's look at the choropleth map of Kyiv to visualize the different rent prices with a colour. For this visualization, we use the folium Python library.

Chart 3. Average Apartment Rent Price in Kyiv Districts (USD per month)



In near the half of Kyiv districts, people live with a low density, in another half, the population density is from medium to high. The closer you move to the city center, the more densely populated areas you have. The highest density is in central historical areas – Solomianskyi, Shevchenkivskyi and Pecherskyi districts. Let's show it on a map differentiating population density in different districts by different colours.

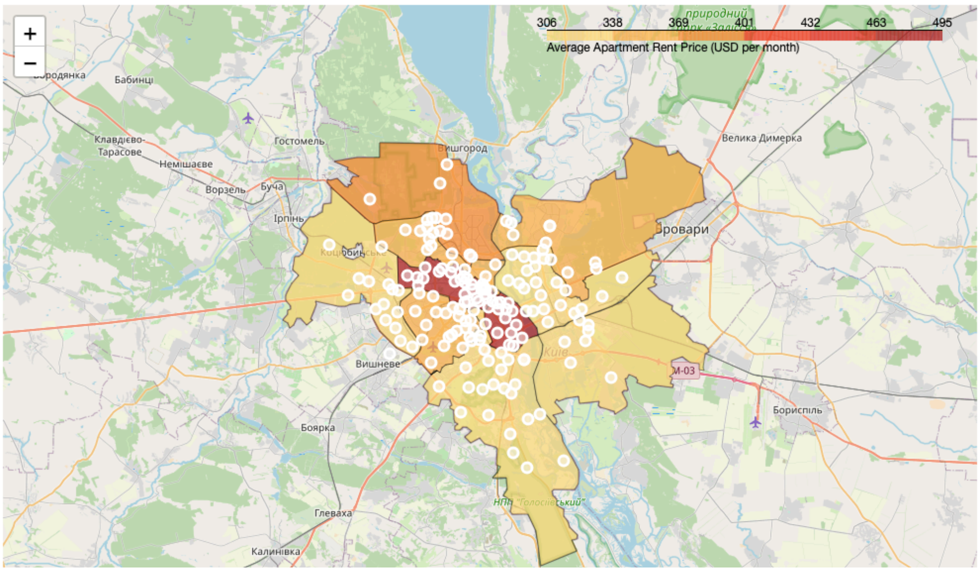
Chart 4. Average Population Density in Kyiv Districts (thd people per km2)



The second step in the project is the collection of the list of all sub-districts in Kyiv and visualization of their locations on the city map.

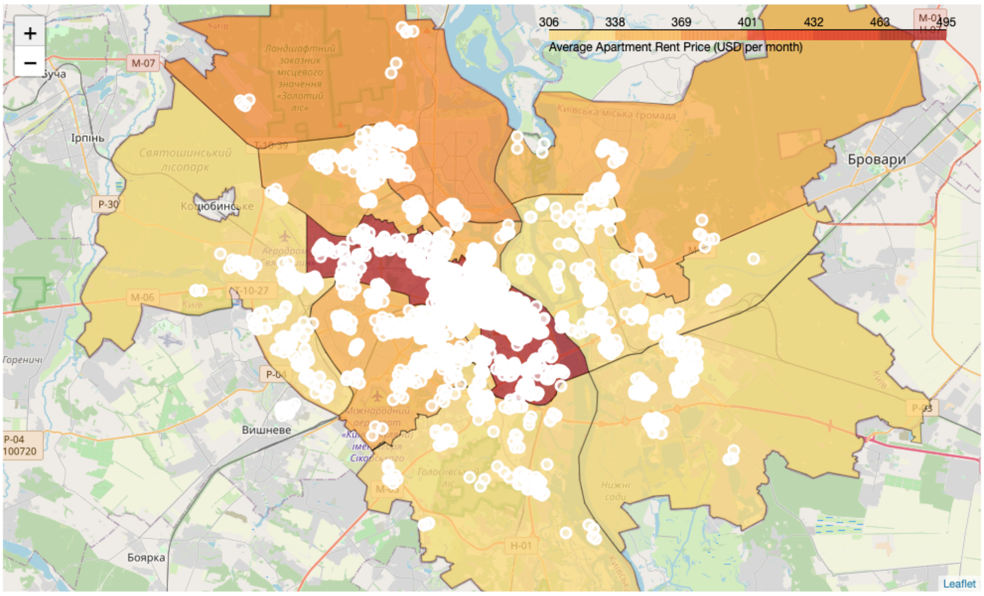
Since we weren't able to find a list of sub-districts in Kyiv from any official municipal resource, we downloaded the list of historical locations (sub-districts) from the Wikipedia webpage. This resulted in 200 names, out of which we lost 34 values during the data cleaning process and stopped with the exploration of the rest 166 sub-districts. With the help of geopy Python client from OpenStreetMap client, we download location data and add it to our dataset. Let's look at the allocation of sub-districts over the map of Kyiv districts coloured based on rent price.

Chart 5. Location of Kyiv Sub-Districts



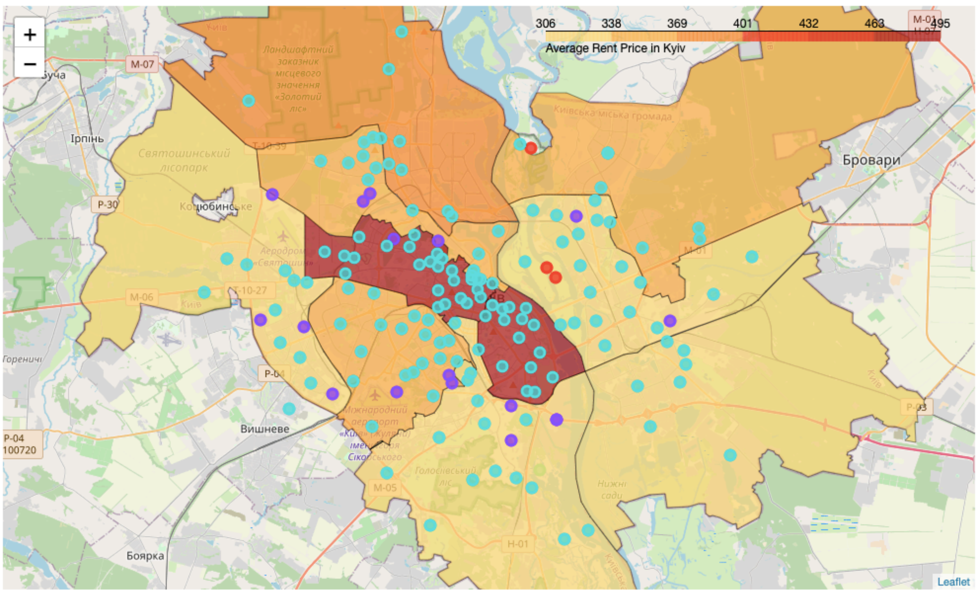
Third and the final step, is to collect data on venues and explore sub-districts based on similarities of existing venues. Using Foursquare API, we obtained all venues near our sub-districts in a radius of 500 meters. We found 3691 venues within 350 unique categories. All venues are shown on the map below.

Chart 6. Location of Venues



Before running the clusterization algorithm, we employ the one-hot encoding technique to venues data, which converts the categorical values into dummies so they can be used for machine learning. Next, we grouped sub-districts by taking the mean of the frequency of occurrence of each venue category. This data we are using for the following clusterization process. Finally, we apply the K-means approach, an unsupervised machine learning algorithm, to cluster sub-districts based on similarities of their venues. To remind, we explore only venues in a radius of 500 meters from a sub-district. To identify the optimal number of clusters we use the Elbow method. As a result, we divided the sub-districts into 4 clusters. All clusters are shown on the map below.

Chart 7. Sub-District Clusters



We see that Kyiv is a quite homogeneous city with sub-districts mostly belong to Cluster 2, consist of 87% or 144 sub-districts (colour of the sea wave on Chart 7). But, the cluster is incredibly diverse. You can find here almost everything - clothing malls and supermarkets, café/pizza places/coffee shops/and restaurants, parks/lakes/gyms/fitness and dance centers.

Table 2. Sub-District Clusters

Cluster	Number of Sub-Districts
2	144
1	17
0	3
3	2

In contrast, the second biggest cluster, Cluster 1 (17 sub-districts), is monotonous, "boring" and quiet. Venues mostly consist of auto workshops and garages, fishing spots and fishing stores, fish/farmers/flea markets, and outdoor gyms.

Two remaining clusters (Cluster 0 and 3) include only 2-3 sub-districts, small and similar to previous Cluster 1. We can find here mainly lakes, beaches, fishing spots/stores and different markets.

Results & Discussion

For a better understanding of results let's show two maps side by side - a map of sub-district clusters with district rent prices and a map of population density. To remind, the background colour on the map identifies the level of rent price or level of population density. The more saturated the colour, the higher the value of rent price or population density.

Chart 8. Sub-District Clusters & Rent Price

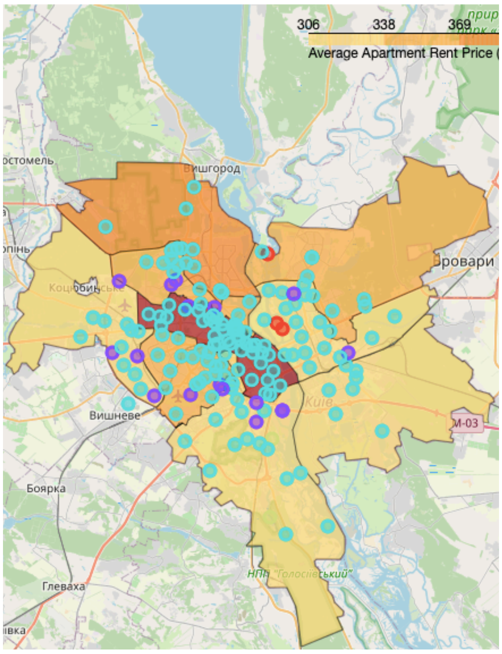
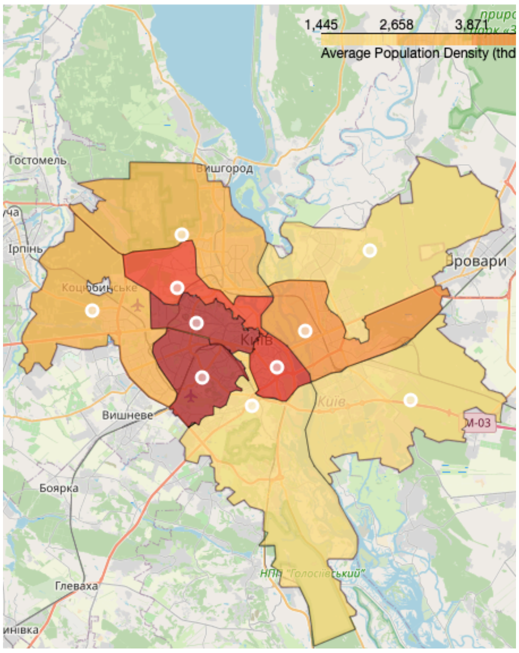


Chart 9. Avg Population Density



Now we can clearly see the rent price and population density in a particular district of Kyiv, and how sub-districts differ based on available venues around.

The result on rent prices is pretty much expected - the highest prices are in the center of the city in two districts, Pecherskyi and Shevchenkivskyi. The rent price starts here at USD 470+ and the gap between the closest price is circa USD 100, which is notable. This means that to move from a Low-price district to a Low+/Medium-price district you should pay only USD 5-17 extra on average. But to move to the city center districts the extra cost increases 5 times to USD 100.

When you move to the city you should take into account another important factor, the population density in different city parts. Some locations are highly overcrowded, other - more calm and "rural". This is a personal choice, so we just describe the situation and show it on the map. As expected, the historical center of the city is more crowded compared to the surroundings. Thus, the highest density of population in Solomianskyi, Shevchenkivskyi and Pecherskyi districts. The lowest density in relatively recently inhabited locations highlighted with light orange on the map - Desnianskyi, Holosiivskyi and Darnytskyi districts.

When we look at more narrowed areas of the city, sub-districts, we realize that Kyiv is highly homogeneous. The most of sub-districts, 144 out of 166 examined (87% of total), belong to same cluster. This means that all these sub-districts are very similar to each other when we compare them by the surrounding venues. In terms of venues, this cluster includes almost everything you need near your home - gyms and dancing studios, coffee shops, bars and restaurants, different shopping places, green areas like parks or lakes. The second biggest cluster represents only 17 sub-districts or 10% of total. Locations in this cluster are more quiet, and what is interesting, the 1-st most popular venues in this cluster is the auto workshops and garages. Other popular venue categories include different types of markets, fishing spots and fishing shops, outdoor gyms.

Next, we want to discuss some limitations and suggestions for future research. First, in this project, we explore rent prices and the population density of 10 Kyiv districts. The area of each district is very big, that is why rent prices and population density may vary significantly within each particular district. This may lead to some bias in the result since the situation in smaller areas (sub-districts) may differ considerably from a situation in bigger areas (districts). Unfortunately, we weren't able to find more narrow-location data to analyze rent price and population density based on smaller sub-districts. Second, clustering the sub-districts we explored only venues in a radius of 500 meters, comfortable walking distance. This also may include some bias as it limits our research mostly to the people without a car. So, it would be interesting to see how the result would be different if we increase the radius, for example, to 1000 meters. We currently leave all these questions open, which keeps room for improvement for future research.

Conclusion

This project aims to help local work migrants, and other people, in their search for the most suitable place to live in Kyiv. We tried to answer the following questions: "From which part of Kyiv I should start searching for an apartment? Whether the selected location be comfy for me?" With this goal in mind, we explored the city from three sides: rent prices in different districts, the density of population in these districts, and going to smaller sub-districts how similar the city locations are based on their venue infrastructure.

As a source of our data, we used Wikipedia, SV-Development, OpenStreetMap and Foursquare services. This project was performed based on Python 3 and Jupyter Notebook. The research was implemented using pandas, matplotlib, numpy, requests, folium and scikit-learn Python libraries. We used geopy Python client to download geospatial data for Kyiv districts and sub-districts. With folium library, we were able to build choropleth maps where city districts coloured by average rent price or population density. Employing Foursquare API requests, we downloaded a list of venues around each Kyiv sub-district in a radius of 500 meters. Utilizing this information, we applied the K-means approach, an unsupervised machine learning technique, to cluster the sub-districts into 4 clusters based on similarities of their venues.

Concluding the results, first, we confirmed the obvious :) - the closer you to the city center, the higher the rent price. Also, we found that rent prices in different parts of Kyiv are not uniformly distributed. Most of the districts, 7 out of 10, are closer to the lower price level with a significant price gap to the most expensive two districts located in the center of Kyiv. Second, we explored that old historical areas in the center have higher population density, while relatively recently inhabited locations on south and east parts of the city tend to have lower density. Finally, exploring the similarity of sub-districts we discovered that Kyiv is a highly homogeneous city. About 87% of all sub-districts belong to the same cluster, which means that all these locations are very similar to each other when we compare them by the surrounding venues. In terms of venues, this cluster is saturated with extremely diverse venue categories, it includes almost everything you need just near your home.

References

1. Wikipedia: list of Kyiv districts, population density
2. SV-Development: average rent prices in Kyiv districts
3. Wikipedia: list of Kyiv sub-districts
4. Foursquare API: information on sub-districts venues
5. Google Maps

PS. The code for the project can be found on my github repository.

PPS. This is my first data science project, so there might be some mistakes. Please feel free to if you find some so that I can correct them and make this research better.