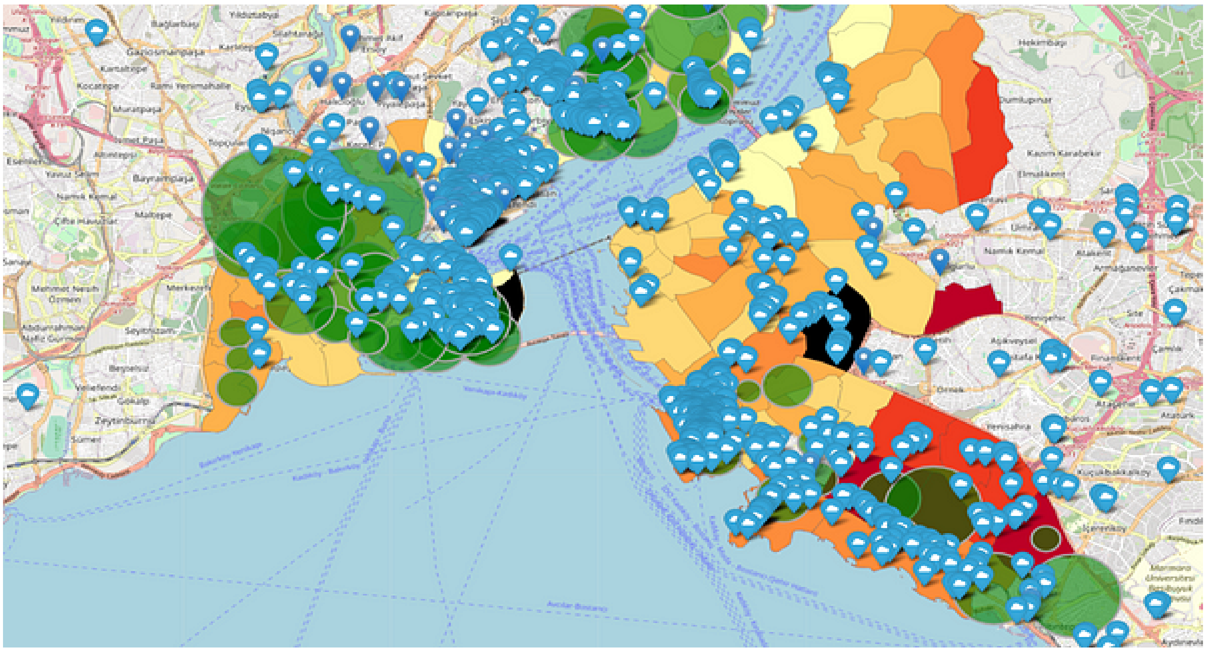
Unraveling the Coffee Code: Data-Driven Analysis into Istanbul Café & Airbnb Locations



In the following script, I will be excavating into the cityscape of Istanbul. The threefold objective is to examine the city's population distribution, the geographical spread of coffee shops, the average Airbnb rental prices across various neighborhoods and examining the potential correlation between them.

Data

The data used in this project were obtained from various sources:

Population & Demographics Data

The population data obtained from Turkish Statistical Institute. The dataset contained every neighborhood population in Turkey. For the purpose of this project, it has been distilled to focus solely on Istanbul's districts.

Geographical Data

To locate the neighborhoods, I leveraged on Nominatim Open Street Maps project. From the API, neighborhood boundaries as polygon coordinates were converted to geojson files for easier processing.

Cafe Location Data

The list of coffee shops was obtained by querying YELP through the API. As I use a free-tier account, the results of my query -coffee- has their limits. Through the various libraries I have managed to fetch more result.

Airbnb Data

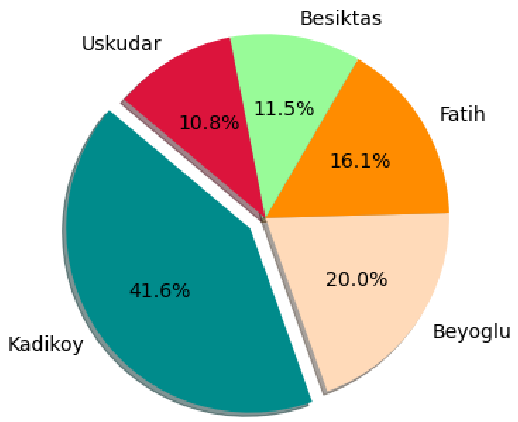
The dataset detailing Airbnb rental prices was sourced from Kaggle

Listing all the cafes in the given coordinates

The first part involves listing all the cafes in given coordinates. Leveraging the Yelp API, I have created a query for coffee shops in the neighborhoods of Istanbul.

The script employs requests and JSON libraries to retrieve and process data from the Yelp API. Yelp limits the results to 50 per query. However, in a city like Istanbul, where cafes are numerous, this limit needed to be expanded. This requires a combination of offset and while loop as well as time library to fetch more results and not to get banned.

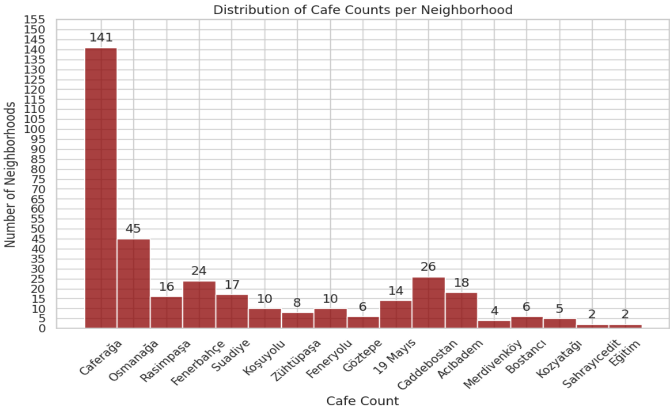
By using the term **'coffee'** and providing the latitude and longitude coordinates, I retrieve a list of cafes in the area. The while loop continuously queries the API for cafes until there are no more left (990 cafes listed). For each cafe, I have extracted the name, coordinates, category, rating, and review count. This data is then appended to our DataFrame, which provides us with a structured dataset of cafes(dfcafes).



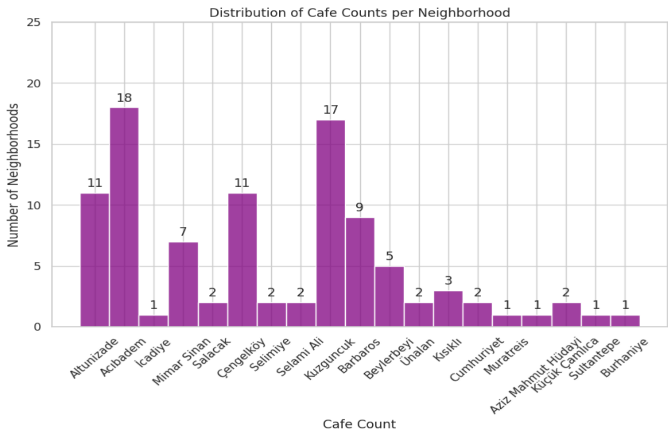
In the boundry box file of the cafes in terms of the location, I had only coordinate values. In order to look more closely, I designed the function to enhance a dataset of Istanbul cafes with neighborhood and district information. It achieves this by using OpenStreetMap's Nominatim API, which allows users to extract place information based on geographical coordinates.

The code defines two main functions: get_neighborhood() and add_neighborhood(). The first function fetches and cleans neighborhood data for a given set of coordinates, while the second one iterates over a dataframe, calling the first function for each row's latitude and longitude.

As a precaution against overloading the API, the second function pauses for 10 seconds every 50 rows. When executed, the code adds a new 'Neighborhood' column to the dataframe. At the end, the function goes through 990 cafes and append the dataframe with the neighborhood and district values for each row/cafe which allows me to perform more detailed analysis and visualization of the distribution of cafes across different Istanbul neighborhoods.

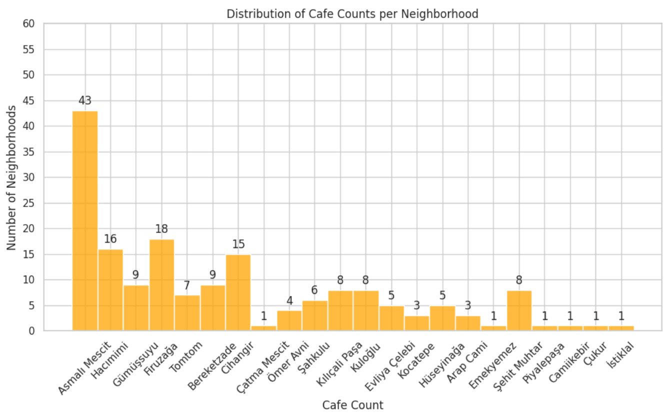


Kadıköy

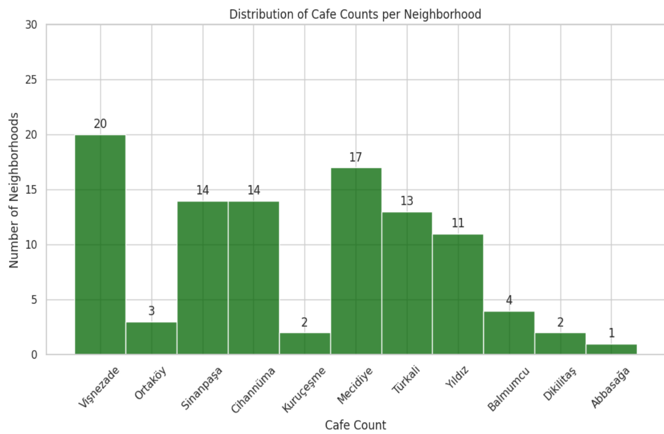


Üsküdar

Caferağa stands out as a significant outlier with 141 cafes, a figure that vastly surpasses the count in any other listed neighborhood. The second highest, much smaller, is Moda. This stark difference underscores Caferağa's unique position as a central hub for cafes in Istanbul, hinting at a vibrant coffee culture within this locale. The remaining neighborhoods, Osmanağa, Caddebostan, and Suadiye, also has a remarkable count compared to other neighborhoods in Istanbul.

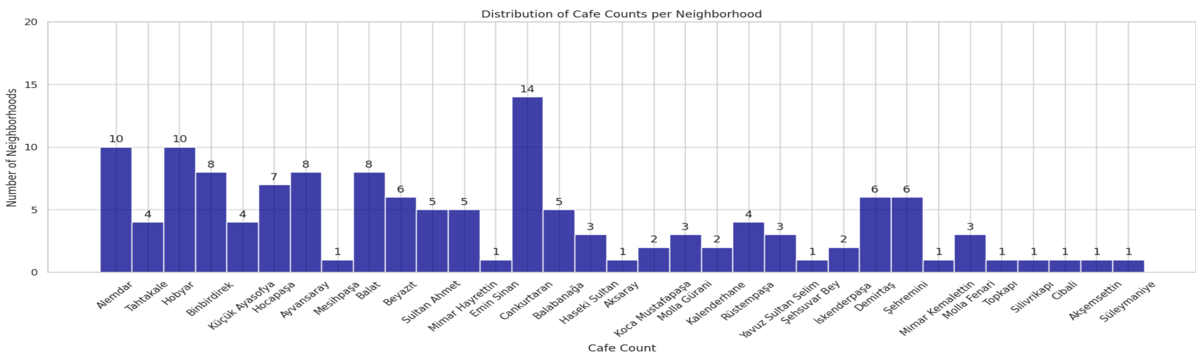


Beyoğlu



Beşiktaş

Surprisingly, the cafe count is lower than expected in vibrant Istanbul areas such as Beyoğlu's Asmalı Mescit, Gümüşsuyu, and Cihangir, which are often known for their vibrant day-to-night rhythm.
In the Beşiktaş district, there is a considerable number of cafes as well. Vişnezade takes the lead in this district, closely followed by Mecidiye, Cihannüma, and Sinanpaşa.



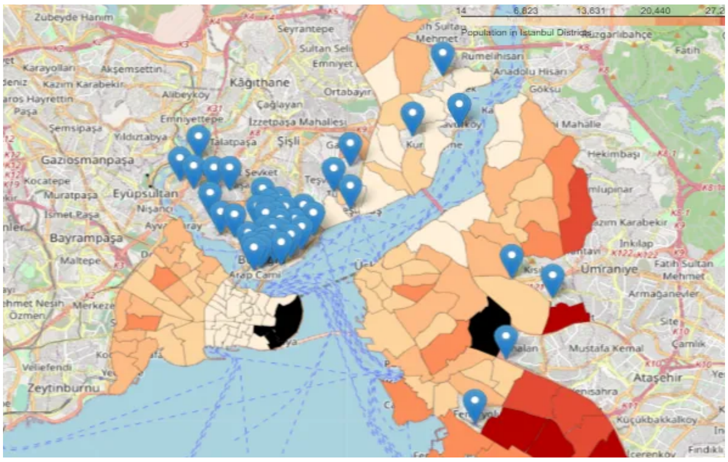
Fatih

Shaping the borders of neighborhoods of Istanbul

The next challenge is to identify the boundaries of Istanbul's neighborhoods. Unfortunately, no API or data file provides this information. Therefore, we will use the names of the neighborhoods from the Turkish Statistical Institution(TUIK) population data and retrieve the geometry data that encloses each neighborhood.
Every neighborhood in the given neighborhood list + Istanbul, searched through the Open Street Map API. The result constructs a feature dictionary that represents the neighborhood.
The feature dictionary stores the name and the geographical data of the neighborhood, as in below.

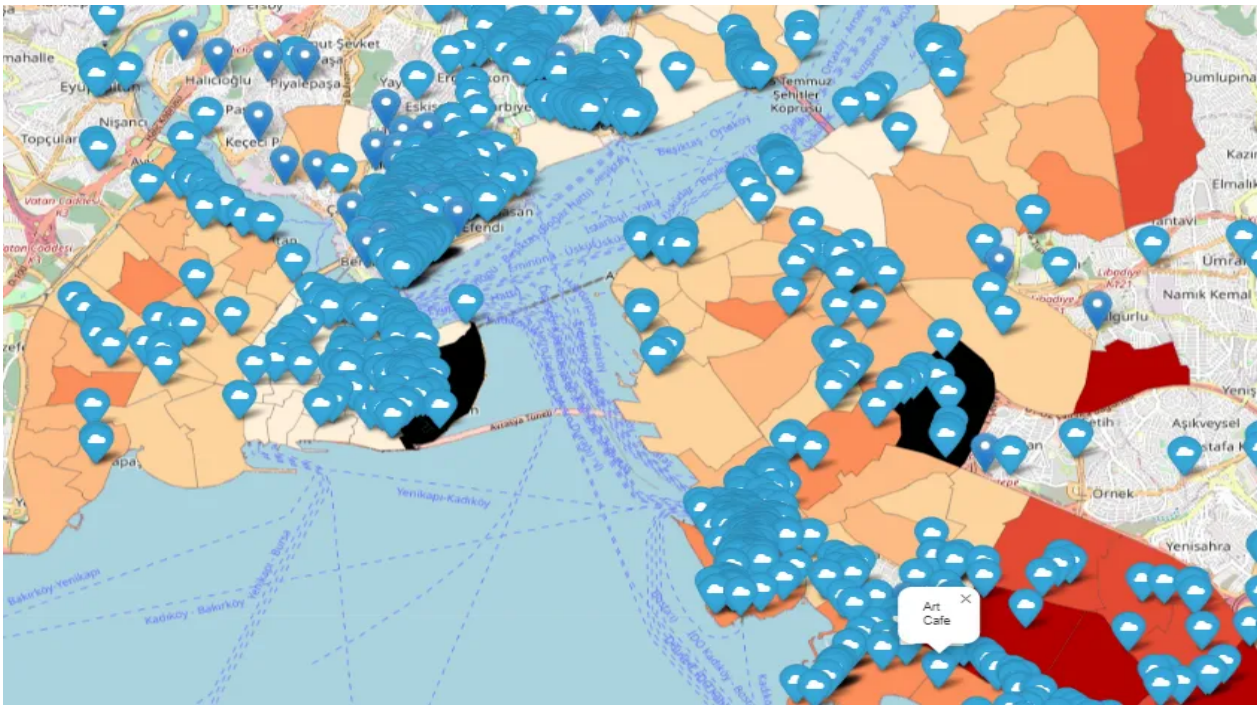
```
name \                                                  geometry
0 Abbasağa Mahallesi, Beşiktaş, İstanbul, Marm  0 POLYGON ((29.00319 41.04803, 29.00321…
1 Akat Mahallesi, Beşiktaş, İstanbul, Marmara…  1 POLYGON ((29.01997 41.09076, 29.02036…
2 Arnavutköy, Bebek Arnavutköy Caddesi, Arnav…  2 POINT (29.04327 41.06718)
3 Balmumcu Mahallesi, Beşiktaş, İstanbul, Mar…  3 POLYGON ((29.00983 41.05668, 29.01034
```

The blue markers on the map below, represent the "POINT" geometry in the output. The geometry of the certain neighborhood cannot be found. It will be seen with the blue markers.



Then visualize these districts on a map using Folium, a Python library that provides a high-level interface for drawing attractive and informative statistical graphics.

Merging the population data with the GeoDataFrame data, allows us to have both polygon coordinates and the population of the neighborhood at the same time

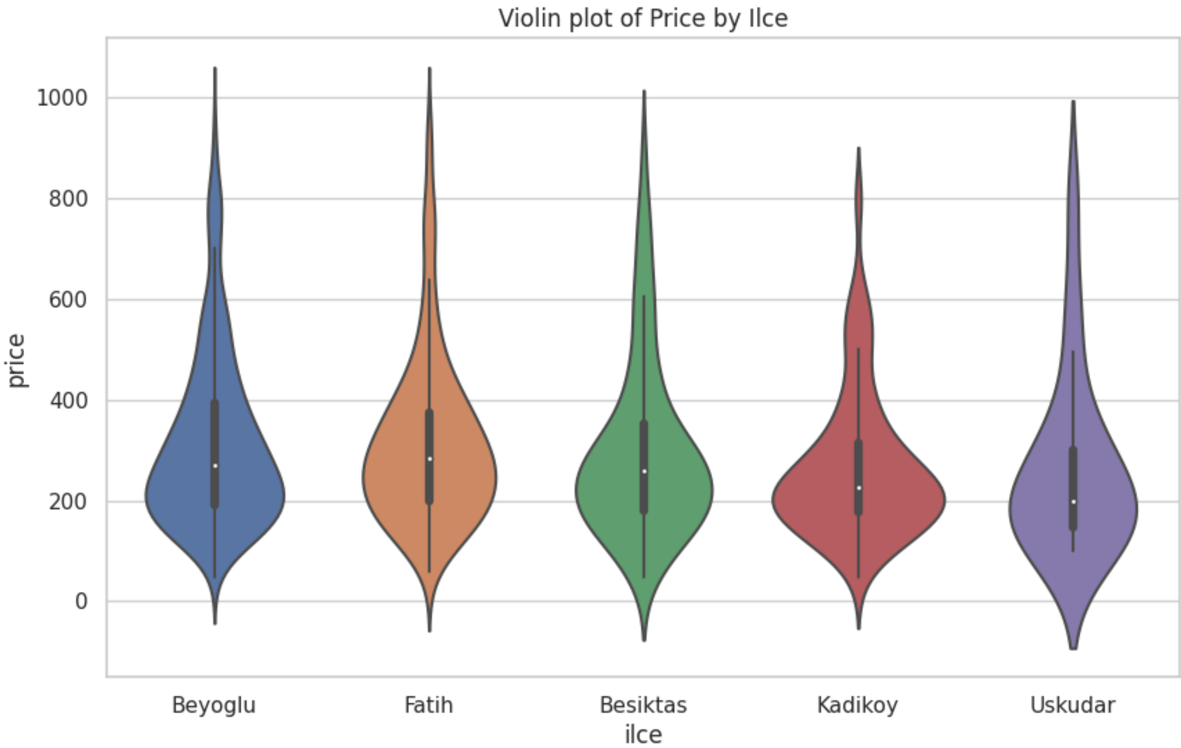


Airbnb Exploration

In this section, I will add another layer to the analysis — the average Airbnb rental prices for each neighborhood. I have an Airbnb Istanbul dataset up to 2019. However, due to inflation, I will adjust the prices to reflect the current rates.

We will perform some cleaning and filtering operations on the data. We will limit the analysis to only five neighborhoods: Kadikoy, Fatih, Üsküdar, Beyoğlu, and Besiktas. We will also filter the data to include only listings that are either a private room or an entire house.

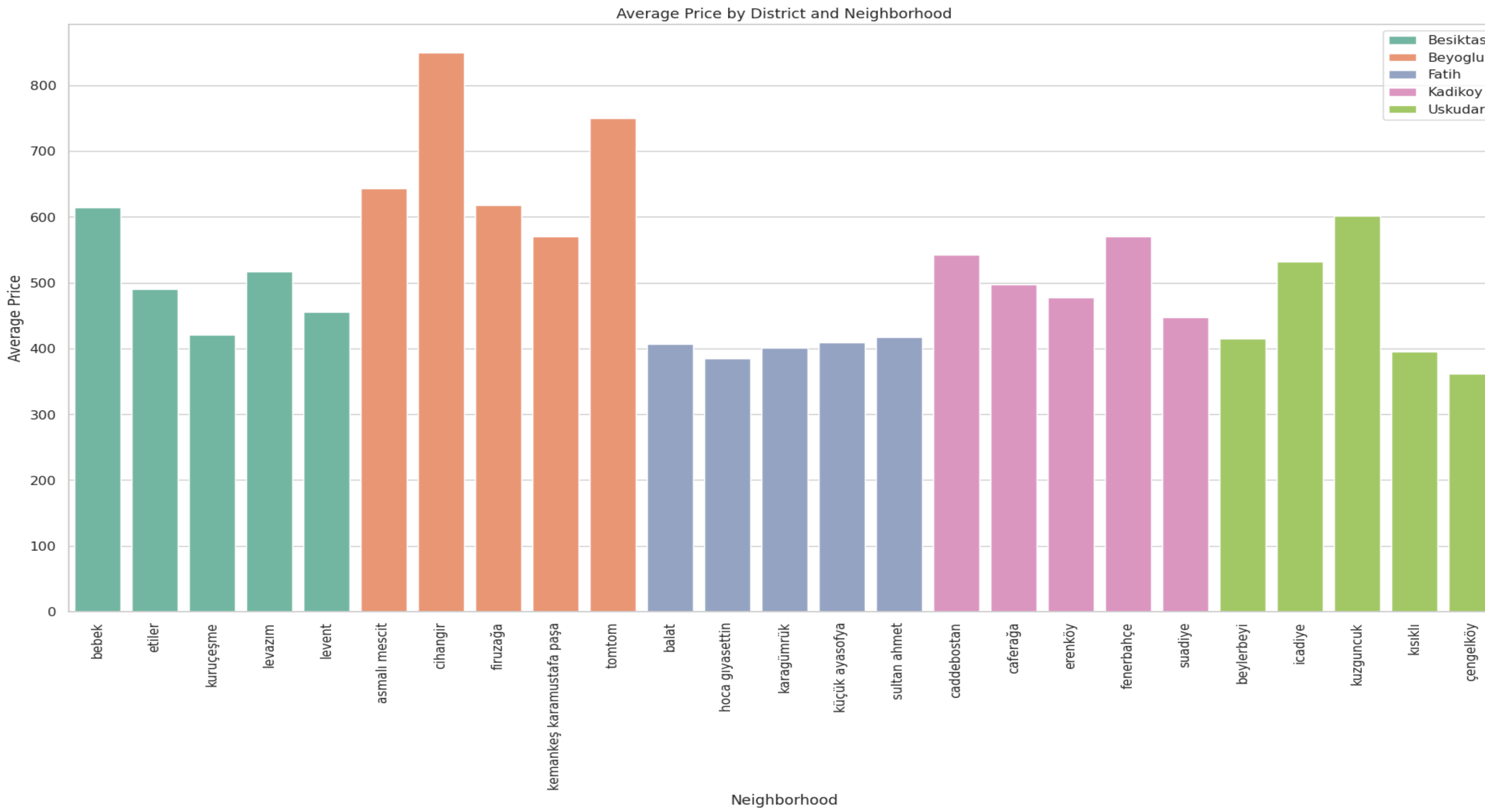
Optionally, we can filter further by only including listings that have received reviews, and ones that receive more than 0.5 reviews per month. These filters could help ensure that we're only including popular, actively rented listings in our analysis.



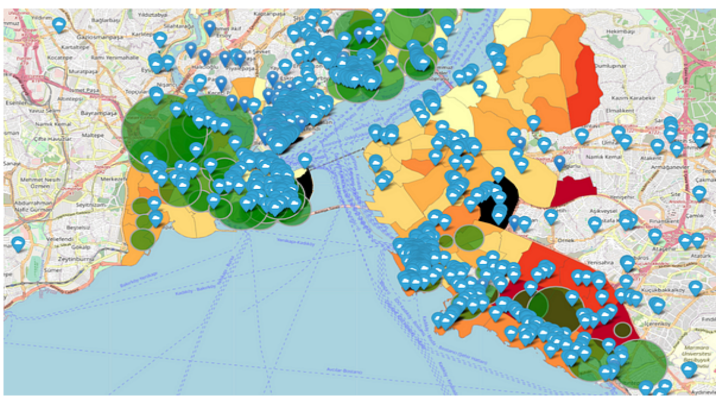
The distribution of prices in Beyoğlu showed a single peak, suggesting a unimodal distribution, whereas Beşiktaş and Kadıköy demonstrated a bimodal distribution with a significant portion of prices at both lower and higher ends. Fatih and Üsküdar had their prices leaning towards the lower end with a long tail extending towards the higher end, indicating a right-skewed distribution.

In terms of location data, airbnb dataset have district (tr. ilçe), latitude and longitude values. As I have constructed our base on neighborhood, I will use the the Nominatim service again, but this time only find the neighborhood. As I had previously went through the provided functions get_neighborhood() and add_neighborhood() designated for extracting the neighborhood with the given coordinate values. This time we do not have a JSON file to clean so I run airbnb values only in the add_neighborhood() function on each row's latitude and longitude. The function goes through the 3800 airbnb register and appended the data frame with the neighborhood values.

For Beşiktaş, the highest prices are seen in the Bebek neighborhood, followed by Levazım and Etiler. In Beyoğlu, the Cihangir neighborhood commands the highest prices, followed by Tomtom and Asmalı Mescit. For Fatih, the most expensive neighborhood for Airbnb rentals is Sultan Ahmet, while in Kadıköy, the highest prices are found in Fenerbahçe. Lastly, in Üsküdar, the Kuzguncuk neighborhood has the highest average rental prices.



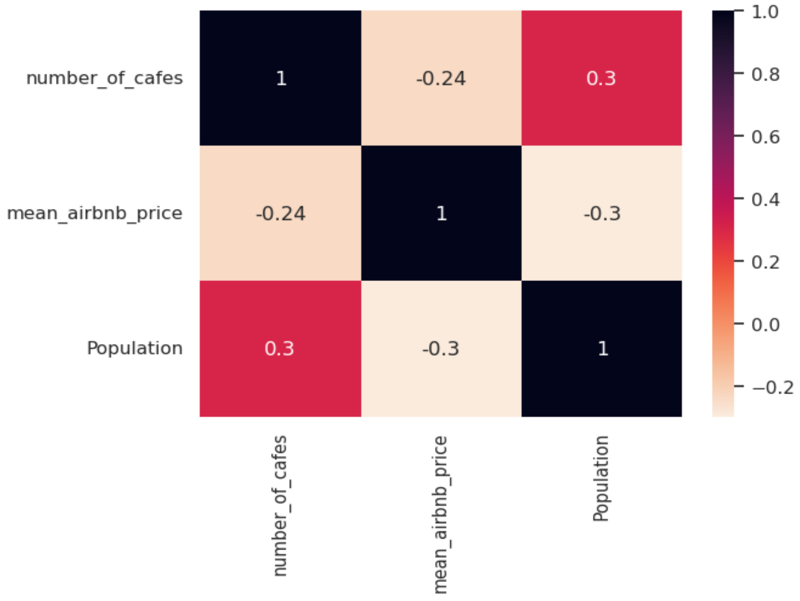
Finally, we visualize the data. We create a map with the neighborhoods, cafes, and Airbnb locations. ***(see the first image)***



Blue cloud markers are scattered across the map, symbolizing the widespread presence of cafes in Istanbul. While providing a visual representation of each cafe, it also illustrate the geographic concentration and distributio.

Green circles reflect the mean Airbnb rents. Each of these circles corresponds to a specific neighborhood, with the circle's radius embodies of the average daily rent for an entire home within that particular district.

Finally, the correlation matrix among any potential variable is shown in below.



The number of cafes and population size show a mild positive correlation of 0.304, suggesting that districts with larger populations tend to have more cafes. This could be due to:
Demand and Supply: One straightforward explanation for this positive correlation might be the basic economic principle of supply and demand. In areas with larger populations, there's likely to be greater demand for coffee shops. This demand can sustain a larger number of cafes, allowing more to exist in these areas compared to less populated neighborhoods.
Social Gathering Places: Cafes are often viewed as social gathering places, and a larger population might necessitate more such spaces. As population density increases, cafes become not just venues for coffee but also community hubs for socializing, working, and leisure activities.
Diversity of Preferences: With a larger population, there's likely to be a wider variety of preferences and tastes. Different preferences can support a larger number of cafes as well.
Higher Foot Traffic: Neighborhoods with larger populations often have higher foot traffic, which can benefit cafes. More people passing by can mean more potential customers, whether it's people heading to work in the morning, meeting friends during the day, or those seeking a late-night coffee fix.

However, both the number of cafes and population size show a negative correlation with the mean Airbnb price, **-0.237** and **-0.296** respectively. This suggests that areas with more cafes and larger populations tend not to have high Airbnb rental prices. This could be due to:
Competition: More cafes could mean more competition, which may keep prices, including Airbnb rentals in lower degree.
Type of Neighborhood: Areas with high cafe density might be more commercial such people prefer to visit rather than reside.
Demographics: Neighborhoods with a higher population might have more varied demographics, including students or younger populations who might not have high spending power.
Preference for Experiences: Airbnb guests sometimes look for 'authentic' experiences. Areas with a higher concentration of local cafes might be perceived as more 'authentic' compared to areas with more tourist-centric attractions.