IBM Capstone Project - Analysising Berlin boroughs

Introduction/Business Problem

House prices in Berlin have gone up dramatically in the last few years. This presents a critical challenge for policy makers and the general population. One of the main culprit is often said to be short lets.

The purpose of this notebook is give an overview of all areas in Berlin, and finding out what factors are most relevant when determining a dramatic change in prices. This will help figuring out which area in Berlin may be most convenient to move in for a young professional and for other cohorts, and which area has the highest potential for return on investment for real estate purchase.

Data

I will be leveraging the Foursquare API to provide geolocation information about the different boroughs of Berlin.

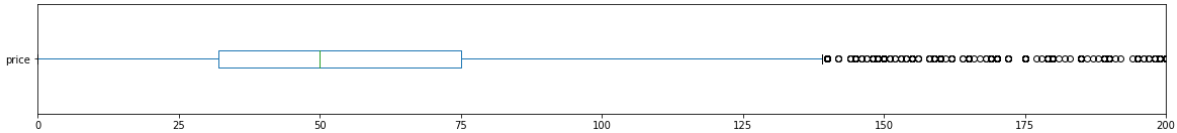
For this analysys, we'll look at the difference in pricing for short lets in each area, and then correlate that with the venues and facilities available. AirBnb prices tend to affect house prices in a particular area, so this will be one of the data sources I will be using. InsideAirbnb provides a dataset of listings for Berlin, including prices, area, and type of listing.

The amount of venues in an area, especially restaurants and cafes, also tends to be an indicator of a popular area, as well as offices, tech companies. For this, I will be using the Foursquare datasets.

Proximity to public transport, availability of services and facilities and schools are also factors to be taken into account. This data will be retrieved using the Foursquare API.

Let's visualise Airbnb prices in Berlin:

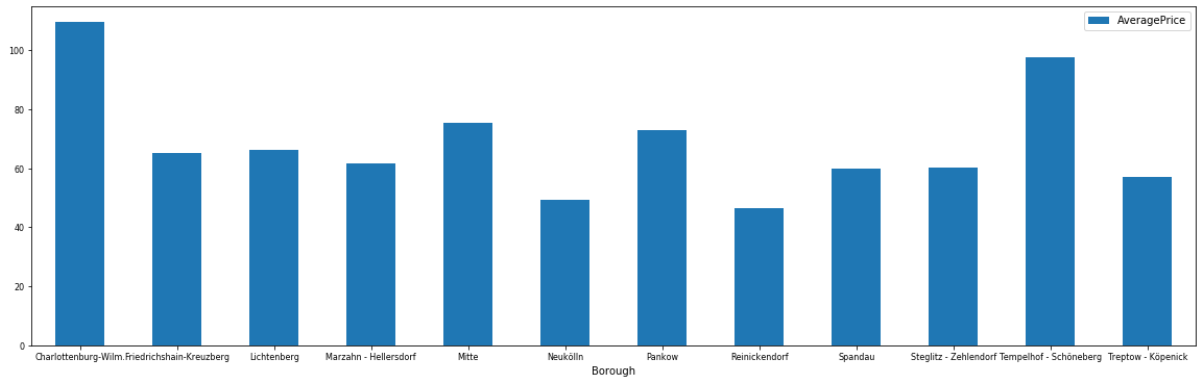
```
count    24395.00
mean        70.85
std        214.40
min          0.00
25%         32.00
50%         50.00
75%         75.00
max       9000.00
```

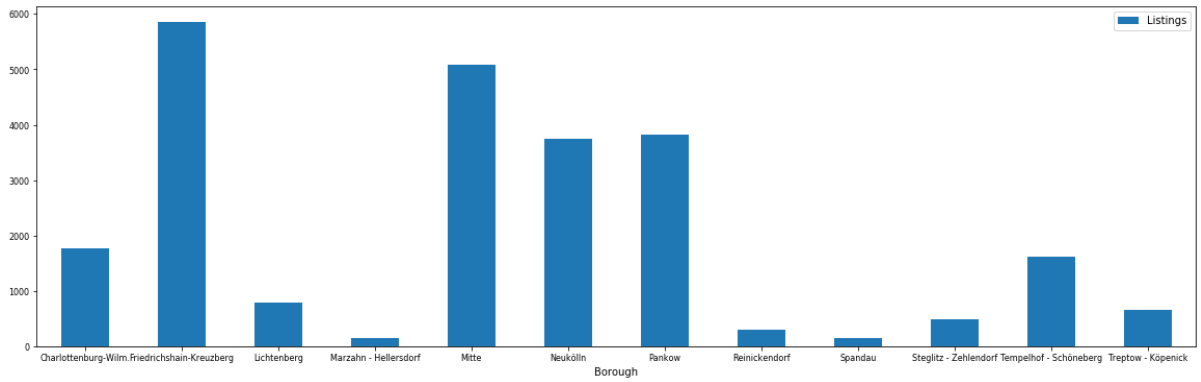


Let's now work on this data to get meaningful information. Let's add how many listings there are for each borough.

	Borough	Neighborhood	Listings	AveragePrice
0	Charlottenburg-Wilm.	Barstraße,Charlottenburg Nord,Düsseldorfer Str...	1766	109.46
1	Friedrichshain-Kreuzberg	Frankfurter Allee Nord,Frankfurter Allee Süd F...	5854	65.17
2	Lichtenberg	Alt-Hohenschönhausen Nord,Alt-Hohenschönhausen...	792	66.10
3	Marzahn - Hellersdorf	Biesdorf,Hellersdorf-Nord,Hellersdorf-Ost,Hell...	142	61.42
4	Mitte	Alexanderplatz,Brunnenstr. Nord,Brunnenstr. Sü...	5082	75.46
5	Neukölln	Britz,Buckow,Buckow Nord,Gropiusstadt,Köllnisc...	3753	49.19
6	Pankow	Blankenburg/Heinersdorf/Märchenland,Blankenfel...	3818	72.89
7	Reinickendorf	MV 1,MV 2,Nord 1,Nord 2,Ost 1,Ost 2,West 1,Wes...	304	46.61
8	Spandau	Brunsbütteler Damm,Falkenhagener Feld,Gatow / ...	138	59.83
9	Steglitz - Zehlendorf	Albrechtstr.,Drakestr.,Lankwitz,Ostpreußendamm...	489	60.36
10	Tempelhof - Schöneberg	Friedenau,Lichtenrade,Mariendorf,Marienfelde,S...	1610	97.66
11	Treptow - Köpenick	Adlershof,Allende-Viertel,Altglienicke,Altstad...	647	57.02

At this point, we can start visualising the data we retrieve to show some insights on the Berlin short-lets market.





Methodology

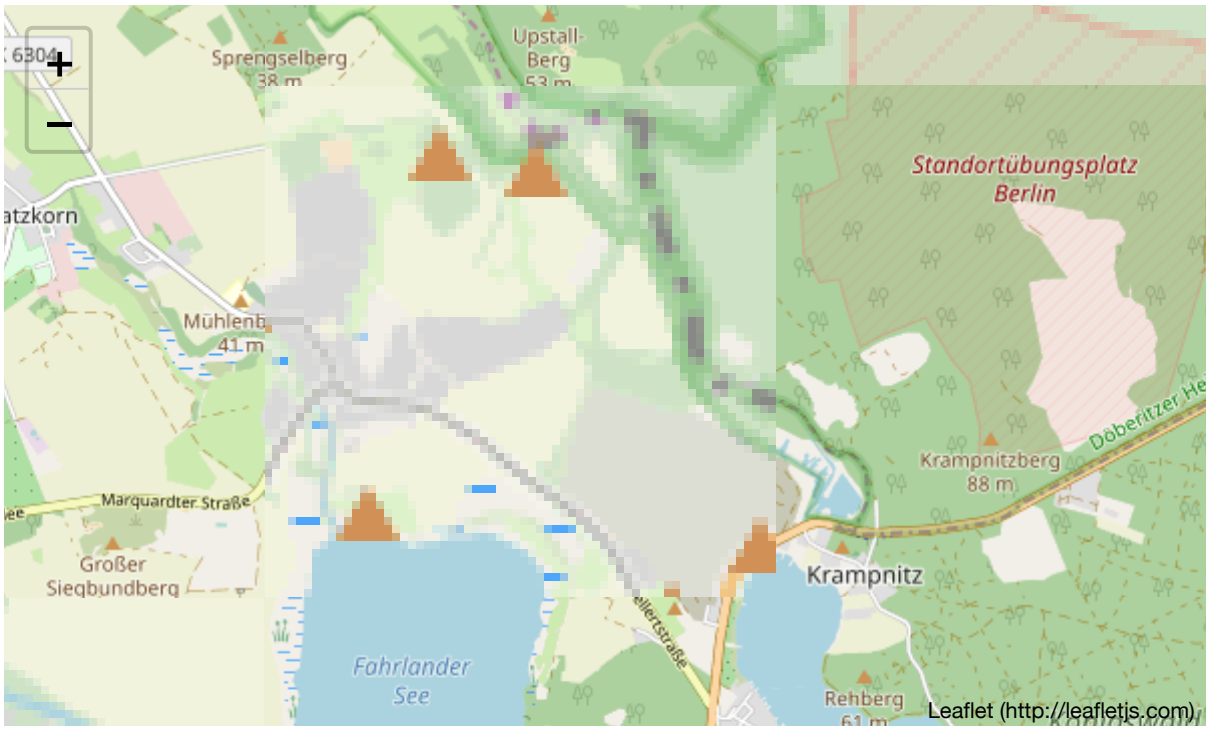
After retrieving geodata from the mentioned APIs, I will analyse each neighborhoods using various available data sources.

First of all, I'll retrieve available Berlin areas from a dataset available at InsideAirbnb. Then, from the same website, I'll retrieve current listing data for Airbnb rentals to check how many listing are active in each neighborhood, and the average price. This should give a good indication on how attractive the area is, and the average price people can expect for a short rental.

Another important factor will be public transport. From data publicly available at BVG and DB, we will check the average distance from the closest station, and the time it would take to reach the city center. As a city center I've selected the Stadtmitte U-Bahn station.

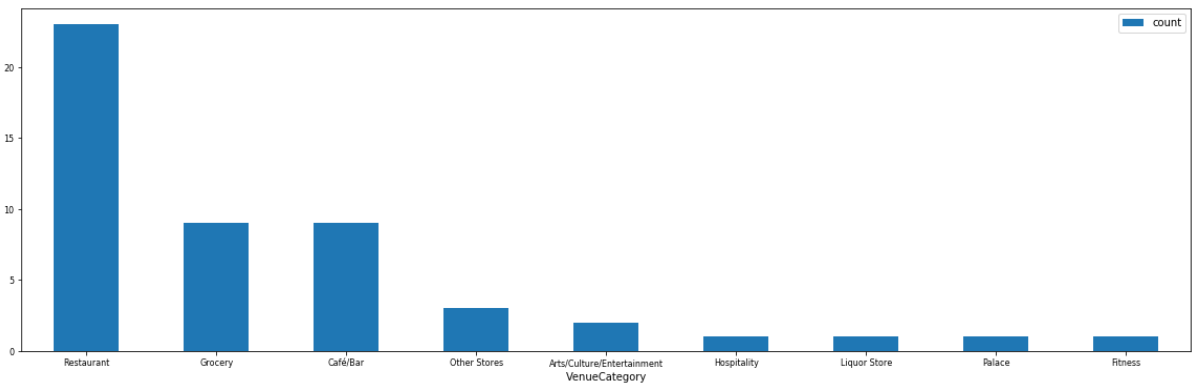
Finally, we will check each area for venues availability, analysed for each cohort. Young professionals will be more interested in proximity to city center, cafes and restaurants, while families will value proximity to schools, healthcare providers and public parks.

I have divided Berlin neighborhoods in clusters, to better visualise the current short letting situation, using K-means Clustering. Here's the result:

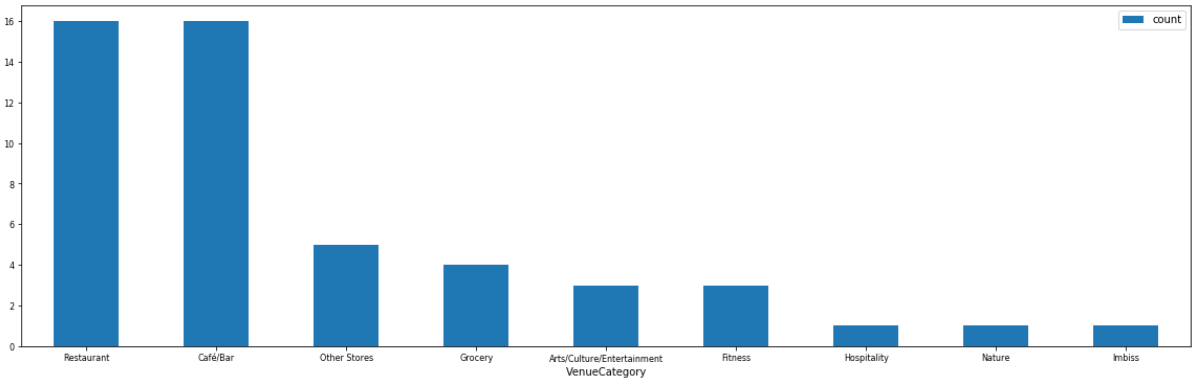


In addition, using the Foursquare API, we can see the most common venues on each borough.

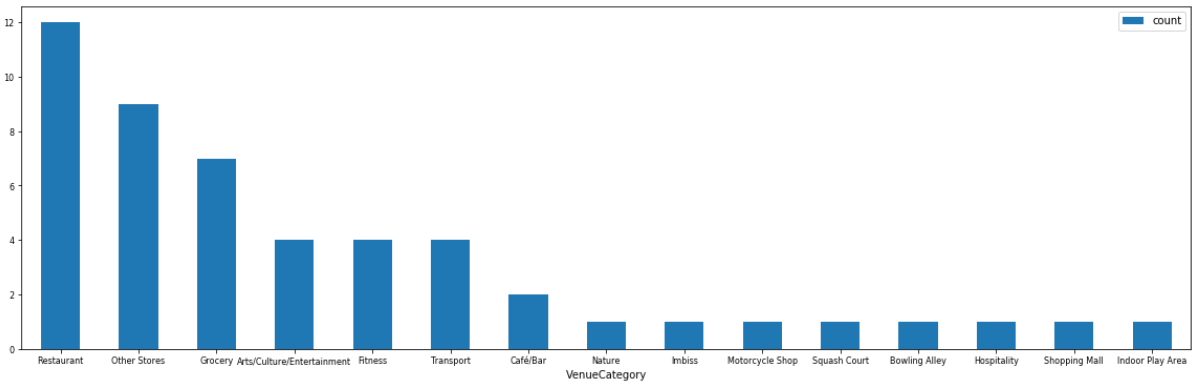
Charlottenburg-Wilm.



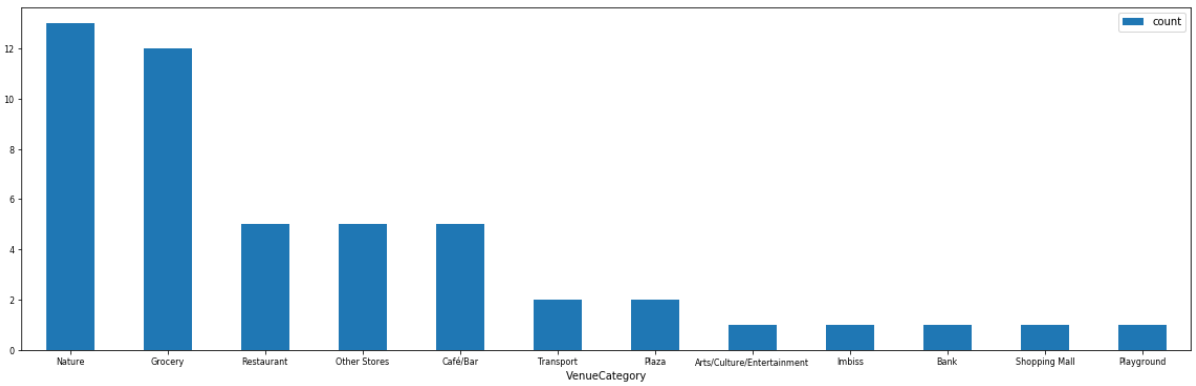
Friedrichshain-Kreuzberg



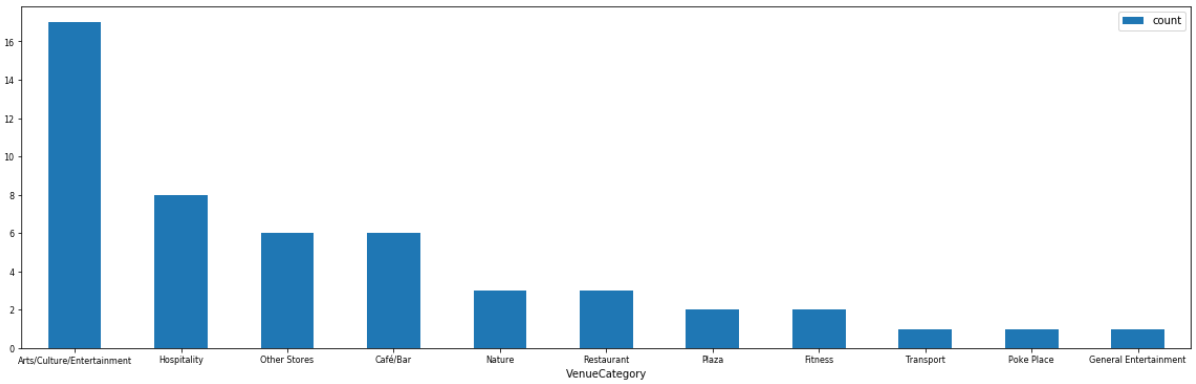
Lichtenberg



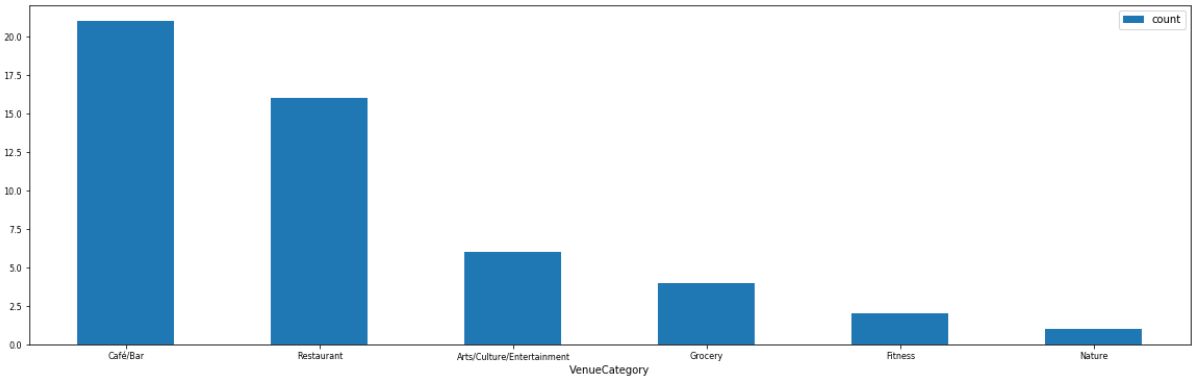
Marzahn - Hellersdorf



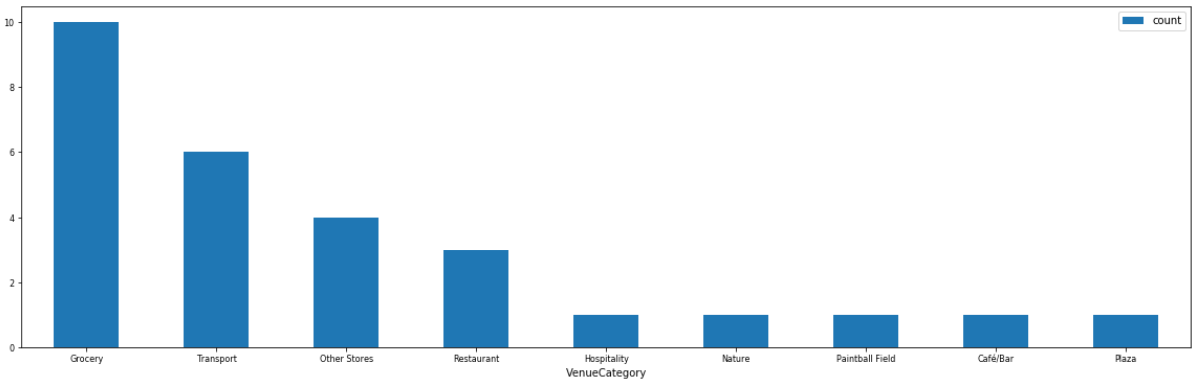
Mitte



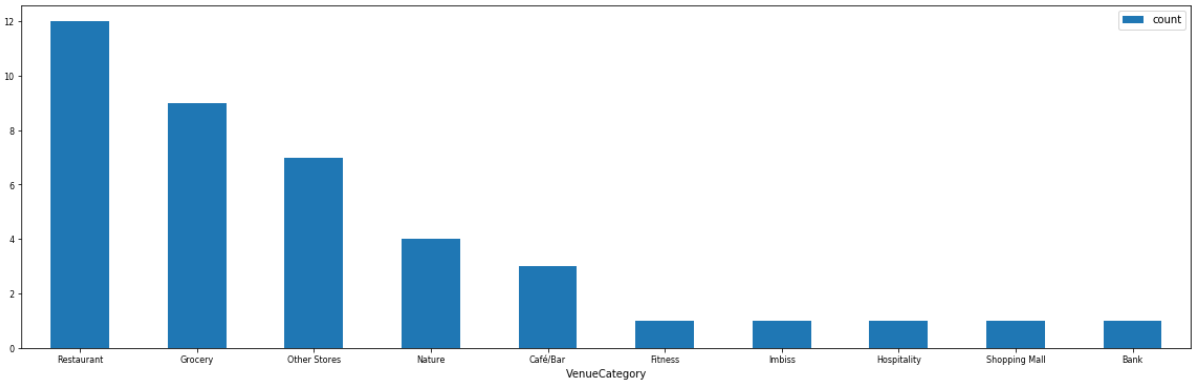
Neukölln



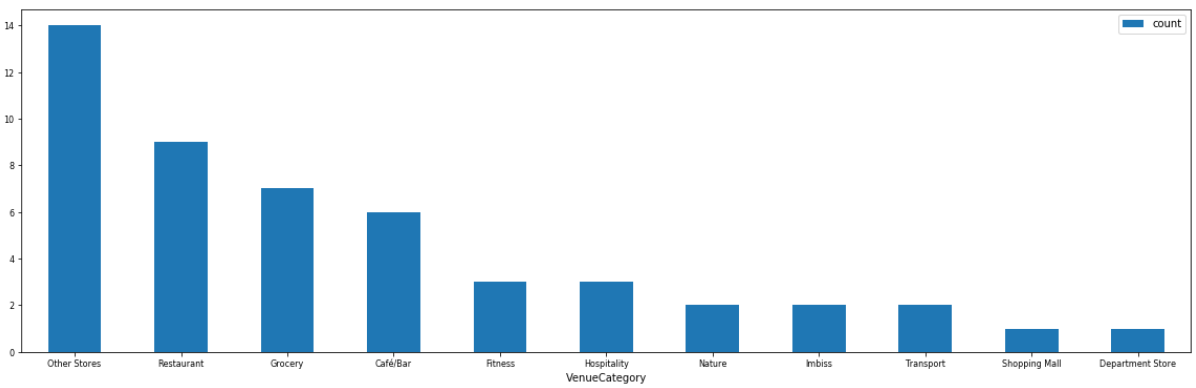
Pankow



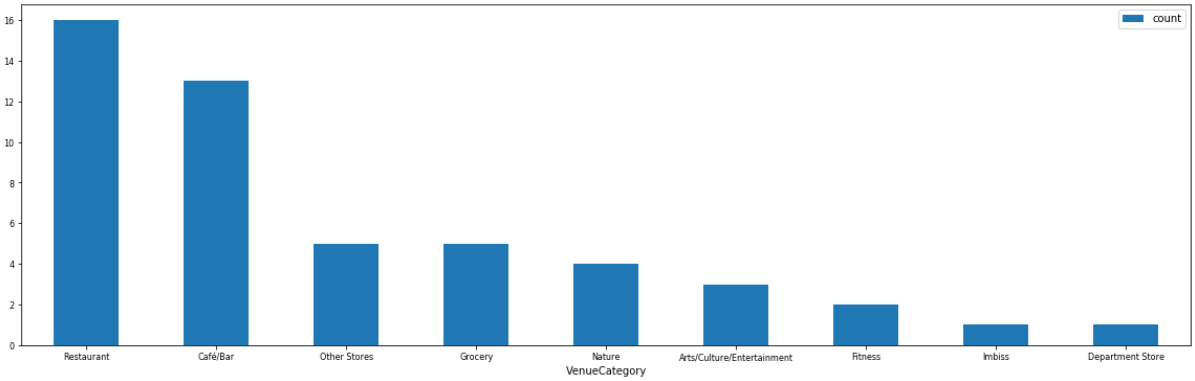
Reinickendorf



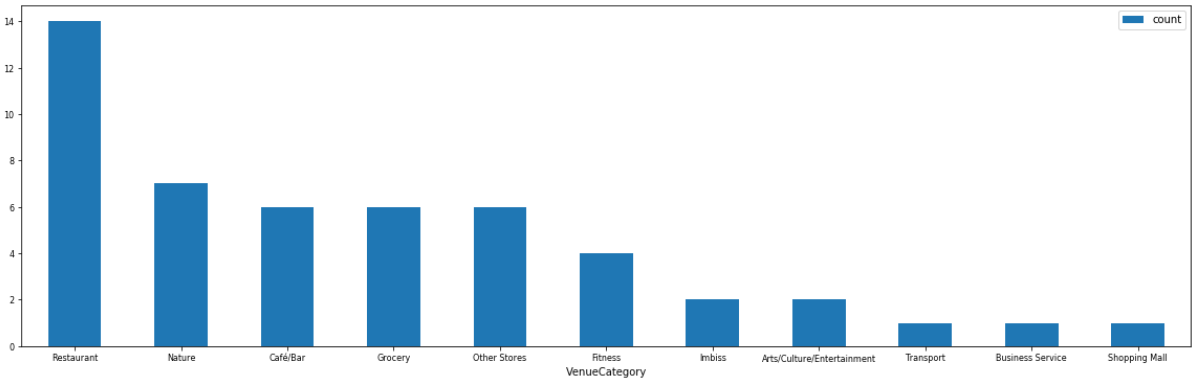
Spandau



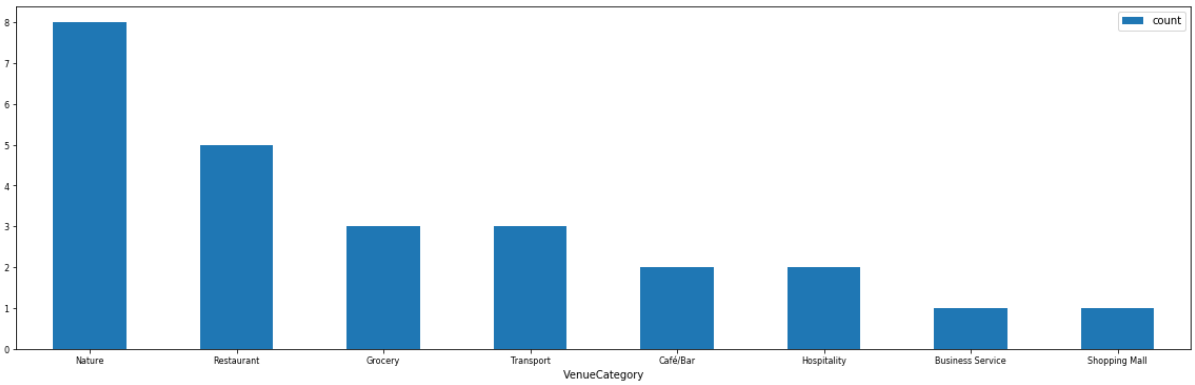
Steglitz - Zehlendorf



Tempelhof - Schöneberg



Treptow - Köpenick



Results and Discussion

After having analysed all boroughs in Berlin, we are now able to look at a clear picture of the short-letting prices, in relation to the venues and proximity to the city centre.

Being directly in the centre, Mitte has the highest amount of short-lets, along with a prevalence of hotels, museums and restaurants. For this reason, it's not the best option for rentals.

Boroughs such as Lichtenberg, Tempelhof - Schöneberg, Spandau, and Reinickendorf have a higher prevalence of Supermarkets, Restaurants, and other facilities that are more relevant for a residential area. In the same way, there are a low number of short-lettings in these areas.

This is also shown by the Clustering results I have shown above. The boroughs just mentioned are all in Cluster 0, while Mitte is together with Friedrichshain-Kreuzberg in Cluster 2, being touristic areas with higher availability of short-lets.

Conclusion

Using the data available in this report, you will be able to have a better idea of the different areas of Berlin, in relation to the facilities and venues available. A stakeholder will then be able to pick an area depending on the personal interest and business, such as a residential, quiet area for a young family, or a more up-and-coming area for young professional with many cafes and bars.