2021

Battle of the Neighbourhoods

A COMPARATIVE ANALYSIS OF COLOGNE, GERMANY

SIMON THALHEIM

Table of Contents

Introduction

Background Information

Gentrification is a common phenomenon happening in big and small cities. The question is: when/how is a neighbourhood starting its gentrification process? What attracts the newcomers? If we evaluate a neighbourhood by its rent price, density and its current local businesses, then compare it to the similar and trendier neighbourhoods of that same city, could we predict what type of upcoming business will potentially be the next one to establish itself, in order to kick-start the gentrification? In that way, could we help local councils choose which development project to choose, in order improve the life quality of their citizens?

Problem statement

In this report, we will pick a neighbourhood of Cologne, Germany called Nippes, which is rather residential. We will put it against the other similar neighbourhoods of that same city with similar densities, similar cluster of local business and higher rent prices, then analyse what type of businesses would be hypothetically the next one to open up.

Target audience for this report

The information gathered from the web and its analysis will be a good basis to make an informed decision, when it comes to forecast what kind of business would be the best fit to improve a specific neighbourhood. In that case, entrepreneurs looking for inspiration, businesses looking to expand, or even city council members could welcome this report as valuable study.

Data

Data Harvesting

We will base our research with the following data:

- **Average price per m² per postal code in Cologne**
  - This data is available through web scraping from this webpage
    - https://www.miet-check.de/mietpreise/plz/koeln/2243/
- **Name of neighbourhoods per postal code**
  - This data is available through web scraping from this webpage
    - https://www.postleitzahlen.de/plz/koeln
    - For the sake of simplifying the processus of harvesting the data, it will be saved and stored in a csv file on Github.
- **Population and density per postal code**
  - Population
    - This data is available from this report from the City of Cologne
      - https://www.stadt-koeln.de/mediaasset/content/pdf15/statistik-standardinformationen/k%C3%B6lner_postleitzahlgebiete_strukturdaten_2016_si_3_2017.pdf
  - Density
    - This data is available on each page of each postal code
      - https://www.suche-postleitzahl.org/koeln-plz-50667-51467.4c37
  - For the sake of simplifying the process of harvesting the data, both information will be handpicked and stored in a csv file on GitHub called PopulationCologne.csv
- **Venues in all different neighbourhoods of Cologne**
  - This information is gathered through Geocoder and Foursquare. Once we attach the proper GPS coordinates to each neighbourhood, we'll be able gather businesses in a radius from that GPS location.

Processing of data

We'll be gathering the data into panda data frames from different web sources mentioned above.

Clean dataset

If we are missing numeric information such as rent prices for a specific postal code, we'll be using the average of that same neighbourhood from other postal codes. Then we'll be removing any NaN, should there be any left, and making sure that only relevant data stays into the datasets.

Exploratory data analysis

We will be using histograms and Folium maps to look at the different rent prices and density.

Once we get to the clustering of the postal codes into similar groups, we will use one-hot encoding and K-Means, as well as Folium map. We will build up a Top 10 of businesses per postal code.

Once this is all completed, we will then triangulate between higher rent group, similar density group and similar businesses cluster. We will reach to a narrow list of similar postal codes to Nippes. Once

this is done, we'll be able to gather all types of businesses by frequency within these postal codes and remove the ones already existing in Nippes, to end up with a top 5 of potential future businesses.

Harvesting and gathering the initial data

After some Python manipulation we're able to group the information coming from the postal code, rent and population files.

	Postal_code	Neighbourhood	Rent	Population	km2	Density
0	50667	Altstadt-Nord,Altstadt-Süd	15.94	4615	1.27	3633.86
1	50668	Altstadt-Nord,Neustadt-Nord	17.29	12281	1.59	7723.90
2	50670	Altstadt-Nord,Neustadt-Nord	16.47	18122	1.89	9588.36
3	50672	Altstadt-Nord,Neustadt-Nord	18.42	9640	1.09	8844.04
4	50674	Neustadt-Süd,Neustadt-Nord,Altstadt-Süd	16.92	19385	1.43	13555.94
5	50676	Altstadt-Süd	16.02	15842	1.37	11563.50
6	50677	Neustadt-Süd,Altstadt-Süd	16.76	13789	1.06	13008.49
7	50678	Neustadt-Süd,Altstadt-Süd	17.50	19158	1.46	13121.92
8	50679	Mülheim,Deutz	14.78	15463	5.39	2868.83
9	50733	Nippes,Weidenpesch	15.55	36526	3.06	11936.60

Figure.1 Postal code, Rent & Population merged

By using the library Geocoder we'll be able to locate the GPS coordinates of each postal code.

	Postal_code	Neighbourhood	Rent	Population	km2	Density	Latitude	Longitude
0	50667	Altstadt-Nord,Altstadt-Süd	15.94	4615	1.27	3633.86	50.939020	6.953905
1	50668	Altstadt-Nord,Neustadt-Nord	17.29	12281	1.59	7723.90	50.950230	6.961745
2	50670	Altstadt-Nord,Neustadt-Nord	16.47	18122	1.89	9588.36	50.949745	6.944331
3	50672	Altstadt-Nord,Neustadt-Nord	18.42	9640	1.09	8844.04	50.942165	6.934012
4	50674	Neustadt-Süd,Neustadt-Nord,Altstadt-Süd	16.92	19385	1.43	13555.94	50.931635	6.933289
5	50676	Altstadt-Süd	16.02	15842	1.37	11563.50	50.930275	6.951211
6	50677	Neustadt-Süd,Altstadt-Süd	16.76	13789	1.06	13008.49	50.920870	6.951421
7	50678	Neustadt-Süd,Altstadt-Süd	17.50	19158	1.46	13121.92	50.923660	6.960874
8	50679	Mülheim,Deutz	14.78	15463	5.39	2868.83	50.935033	6.982215
9	50733	Nippes,Weidenpesch	15.55	36526	3.06	11936.60	50.964880	6.952868

Figure.2 Postal code, Rent & Population merged

Once we obtained the GPS coordinates of each postal codes, we'll be able to call Foursquare and gather information about the venues in proximity to those specific locations. We'll cap it at 75 venues and 1000-meter distance. This will help to regulate how much information is acquired and help all postal codes to have the chance to contribute equally to the research. With those restrictions, we gathered a total of 1482 businesses, spread out around the GPS locations of all postal codes.

	Postal_code	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	50667	50.93902	6.953905	KOLUMBA Kunstmuseum	50.938530	6.954084	Art Museum
1	50667	50.93902	6.953905	Sattgrün	50.938441	6.954965	Vegetarian / Vegan Restaurant
2	50667	50.93902	6.953905	Espresso Perfetto	50.938852	6.953973	Coffee Shop
3	50667	50.93902	6.953905	REWE Richrath	50.938520	6.950774	Supermarket
4	50667	50.93902	6.953905	MUJI	50.939165	6.955877	Miscellaneous Shop
5	50667	50.93902	6.953905	Fanshop 1. FC Köln	50.936579	6.953707	Sporting Goods Shop
6	50667	50.93902	6.953905	LEGO Store	50.937042	6.956564	Toy / Game Store
7	50667	50.93902	6.953905	Apple Schildergasse	50.936552	6.955877	Electronics Store
8	50667	50.93902	6.953905	Vinoteca da Rino	50.938629	6.953072	Italian Restaurant
9	50667	50.93902	6.953905	Fitness First	50.936639	6.951151	Gym / Fitness Center
10	50667	50.93902	6.953905	Brave New World	50.937645	6.953775	Toy / Game Store
11	50667	50.93902	6.953905	Craftbeer Corner	50.937222	6.958928	Beer Bar
12	50667	50.93902	6.953905	Aussichtsplattform des Kölner Doms	50.940893	6.957263	Scenic Lookout
13	50667	50.93902	6.953905	Theater am Dom	50.938586	6.951731	Comedy Club
14	50667	50.93902	6.953905	Excelsior Hotel Ernst	50.942222	6.956282	Hotel

Figure.3 Venues in Cologne

Methodology - Exploratory data analysis

We'll explore 3 aspects as previously mentioned, which are the rents, the density and the clustering of venues for each postal codes.

Rent

We'll first of all put all the different rents into a histogram and check where Nippes stands with its 15.55 EUR/m^2. It will be displayed below by a red line.

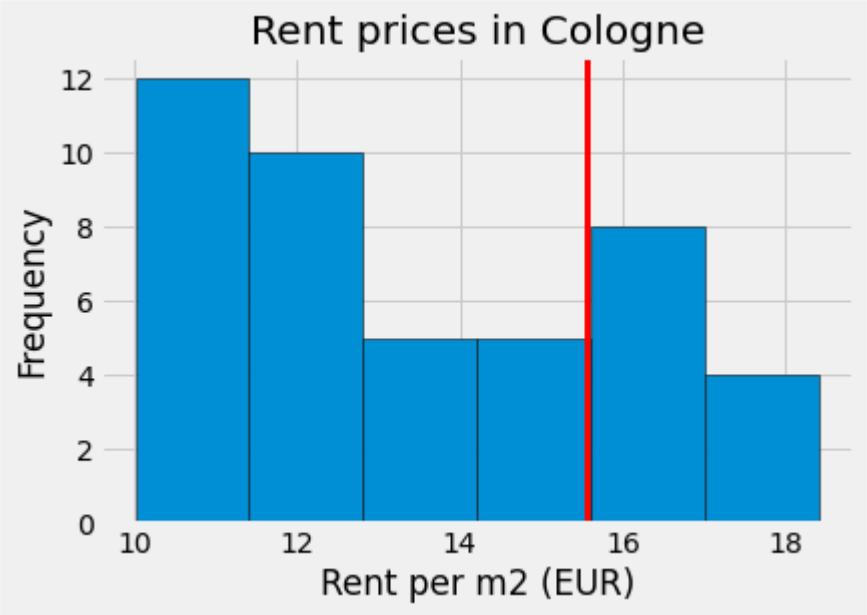

Figure.4 Rent prices in Cologne

As we can see, it's on the verge of joining the upper 2 brackets of the rent prices in Cologne. That means that we will have 2 more expensive groups (8 and 4 postal codes) to look up to, a bit later. What are their respective price ranges? We'll have colours following traffic lights, starting at green for the lowest rent prices, all the way up to red and firebrick for the most expensive rents.

	Bracket	Count	Colour
0	(10.002, 11.412]	12	green
1	(11.412, 12.813]	10	yellow
4	(12.813, 14.215]	5	gold
3	(14.215, 15.617]	5	darkorange
2	(15.617, 17.018]	8	red
5	(17.018, 18.42]	4	firebrick

Figure.5 Brackets for rent prices in Cologne

The upper brackets are respectively **15.62 - 17.02** EUR/m^2 and **17.02 - 18.42** EUR/m^2. Let's push them onto a map of Cologne by using the tool Folium Map.

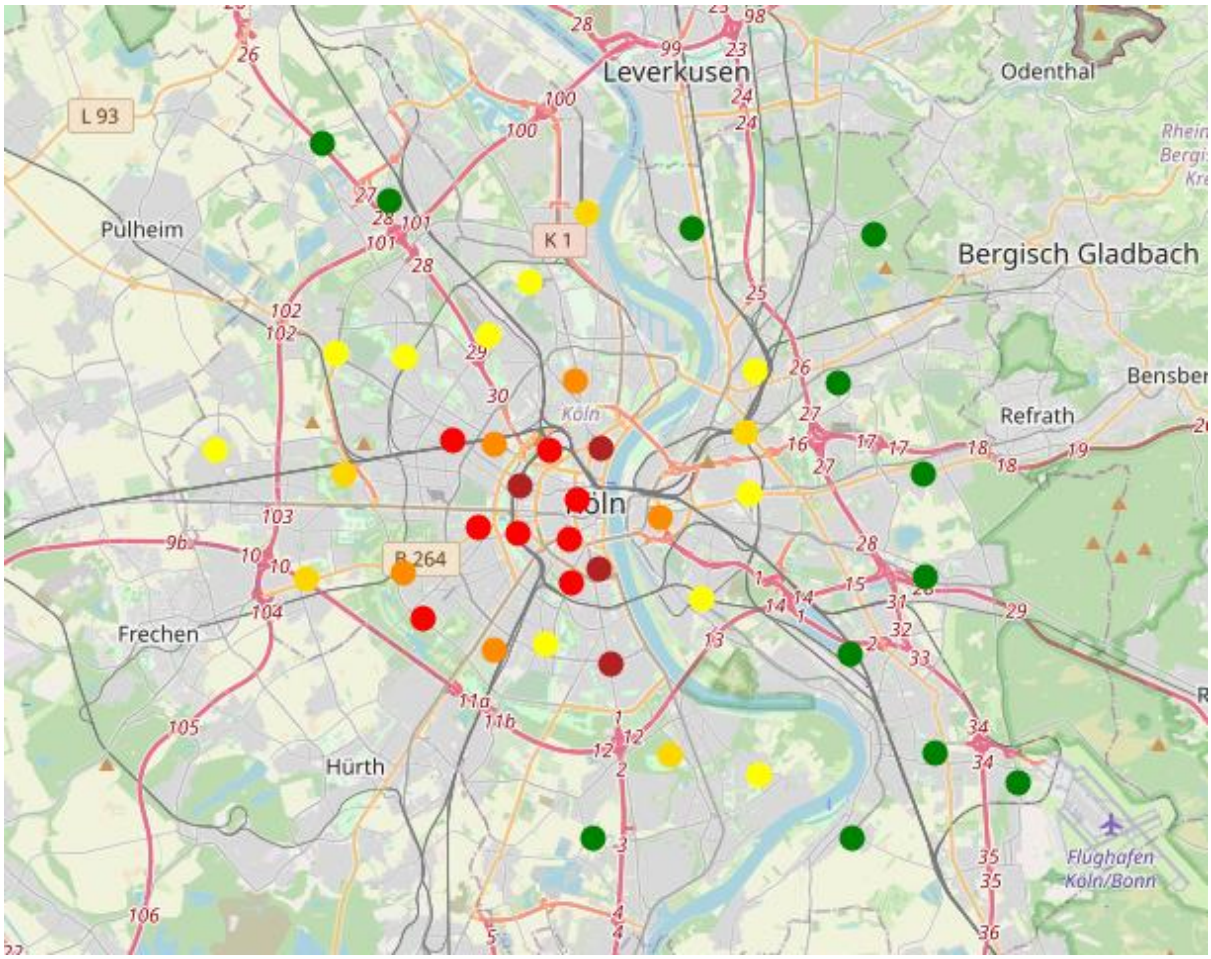

Figure.6 Folium Map of the rent prices in Cologne

If you wish to explore map and click on the different pins, I recommend to visit the Jupyter Notebook used for this analysis. The link to this specific map can be found here.

We can obviously witness a correlation between the proximity to downtown and the rent prices. Three interesting things:

1 - Living right *"in the core of the old town"* (**50667** (Altstadt-Nord) & **50676** (Altstadt-Süd) – the 2 red pins right in the center) is not as expensive as living "*inside Neustadt, but outside the old town*" (**50668** (Neustadt-Nord), **50672** (Neustadt-Nord) & **50678** (Neudstadt-Süd) – the 3 firebrick-coloured pins in peripheral).

2 – To the exception of **50679** (Mülheim, Deutz - orange dot on the right) because it is probably transitioning into being gentrified, there's an obvious and drastic difference of price, should one cross the Rhine. If you ask a local, they all have a theory about it. It even goes way back to Roman times 2000 years ago, when the barbaric tribes (Germanen) would live across the Rhine.

3 - **50968** (Marienburg - the firebrick pin at the bottom of the map) is an obivous outlier. There is an historical reason behind it. At the end of the 19th century, over a period of around 30 years, a great interest in this neighbourhood arose from many architects of the time, which resulted in the development of a large number of palatial buildings. The development was largely completed by 1925 and was called "The villa area".

After World War 2, Bonn became the seat of government of the Federal Republic of Germany in 1949 and Marienburg became the location of some residences of foreign diplomatic missions and embassies. Still today, the villa area of Marienburg is a residential area characterized by magnificent buildings from the turn of the century, with extensive gardens, avenues and parks.

Density

Let's start by visualizing the density distribution in Cologne via a histogram. Let's single out our target neighbourhood with a red line at 11,937 inhabitants/km^2.

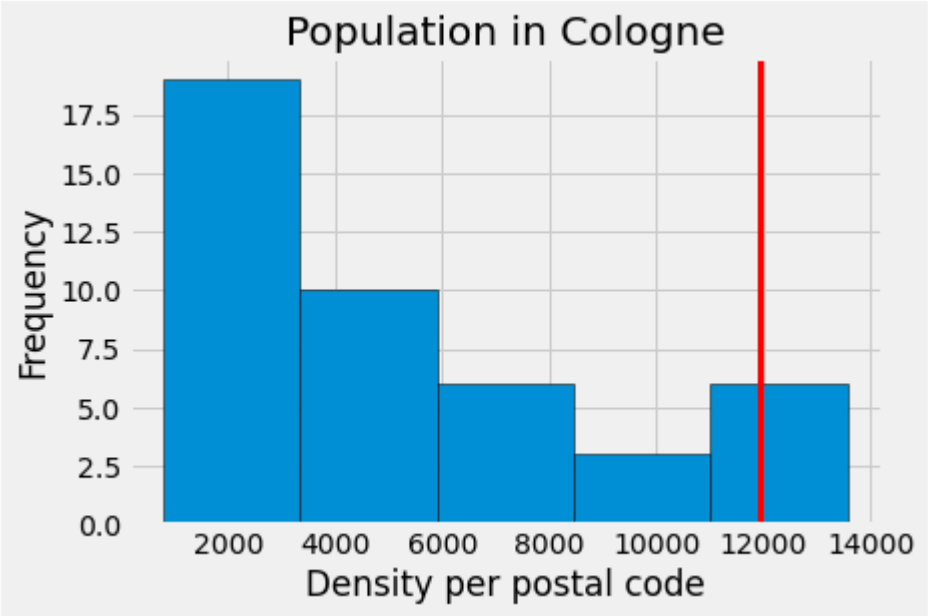

Figure.7 Density in Cologne

As we can see, Nippes is one of the densest neighbourhood in Cologne. That means that we will have 1 group (6 postal codes) to look up to, a bit later. What are their respective price ranges? We'll have colours following traffic lights, starting at green for the lowest densities, all the way up to red and firebrick for the densest postal codes.

	Bracket	Count	Colour
0	(770.808, 3338.052]	19	green
1	(3338.052, 5892.524]	10	yellow
3	(5892.524, 8446.996]	6	darkorange
4	(8446.996, 11001.468]	3	red
2	(11001.468, 13555.94]	6	firebrick

Figure.8 Brackets for density in Cologne

The upper bracket is **11,002 – 13,556** inhabitants/km^2. Let's push this onto a map of Cologne by using the tool Folium Map.

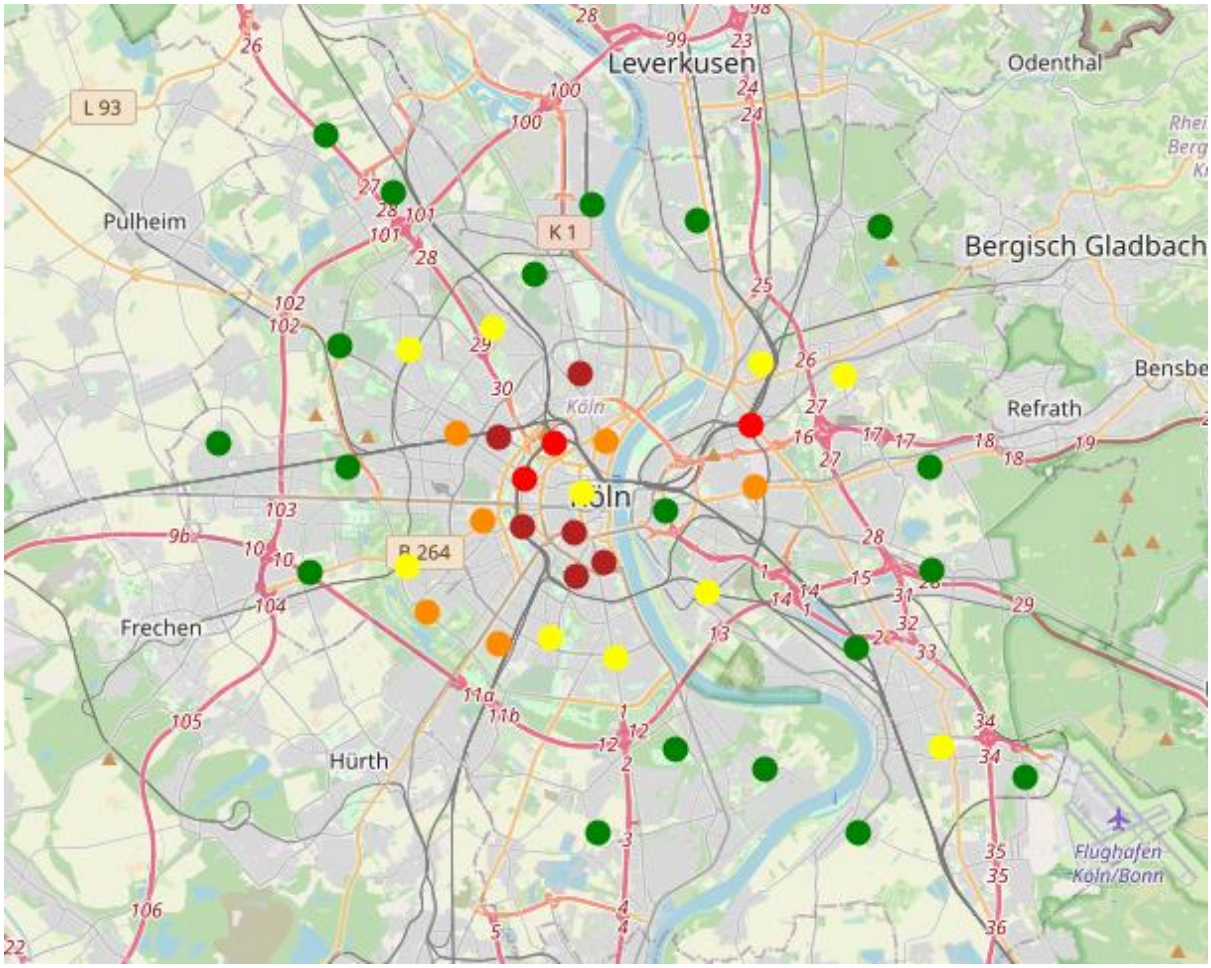

Figure.9 Folium Map of the rent prices in Cologne

If you wish to explore map and click on the different pins, I recommend to visit the Jupyter Notebook used for this analysis. The link to this specific map can be found here.

The correlation between the proximity to downtown and the density is quite obvious here, it decreases with radial distance to downtown. Some interesting facts:

1 - Two postal codes have a smaller density than expected. This can be explained by:

50667 ((Alstadt-Nord - yellow pin in the middle)): The presence of many hotels, squares, cathedrals, restaurants which take up the space instead of residential streets.

50679 ((Mülheim - green pin in the middle)): The presence of Deutz Messe (Cologne trade fair), arenas, hotels which take up the space instead of residential streets.

2 - All denser postal codes are within a close proximity to the Grüngürtel, a group of parks which surrounds the Neustadt. It can be translated to “Green belt”.

Venues

Now that we have a good visual of the spread of the rent prices and density in Cologne, now let's have a look at the venues we harvested from Foursquare. We'll be able to find correlation between all neighbourhoods through their type of venues and find which neighbourhoods have more in common with Nippes.

First, let's see how many venues were harvested per postal code.

Postal_code	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
50667	75	75	75	75	75	75
50668	75	75	75	75	75	75
50670	75	75	75	75	75	75
50672	75	75	75	75	75	75
50674	75	75	75	75	75	75
50676	75	75	75	75	75	75
50677	46	46	46	46	46	46
50678	75	75	75	75	75	75
50679	74	74	74	74	74	74
50733	75	75	75	75	75	75

Figure.10 Venues harvested per postal codes

We can see that we have some neighbourhoods which easily reached the maximum of 75 venues, where others could not reach it. That can be explained by the fact that those postal codes have their venues further out from the epicentre, meaning that they are more spread out, and unfortunately provided less venues. Of all 1482 different venues, we could identify 232 unique categories.

Now let's use the one-hot encoding technique to deal with the categorical data (the venues’ categories). It will help use Machine learning techniques to cluster the postal codes into groups.

	Postal_code	ATM	Airport	Airport Service	Airport Terminal	American Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Turkish Restaurant	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Vietnamese Restaurant	Water Park
0	50667	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0
1	50667	0	0	0	0	0	0	0	0	0	...	0	1	0	0	0
2	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
5	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
6	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
7	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
8	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
9	50667	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

Figure.11 One-hot encoding of the venues

By summing the one-hot encoding per categories and per postal codes, it will allow us to see which categories have a higher frequencies within specific postal codes, in relations to the whole city. Let's have a look at the top 10 categories per postal code.

	Postal_code	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	50667	Hotel	Plaza	Art Museum	Italian Restaurant	Pedestrian Plaza	Café	Gym / Fitness Center	Coffee Shop	Clothing Store	Sushi Restaurant
1	50668	Café	Italian Restaurant	Hotel	Turkish Restaurant	Coffee Shop	French Restaurant	German Restaurant	Movie Theater	Pizza Place	Bakery
2	50670	Italian Restaurant	Turkish Restaurant	Bar	Park	Café	Spanish Restaurant	Hotel	French Restaurant	Burger Joint	Record Shop
3	50672	Café	Cocktail Bar	Hotel	Restaurant	Bar	Sushi Restaurant	Bakery	Plaza	Breakfast Spot	Park
4	50674	Café	Italian Restaurant	Cocktail Bar	Bar	Nightclub	Greek Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Tapas Restaurant	Thai Restaurant
5	50676	Italian Restaurant	Hotel	Café	German Restaurant	Sushi Restaurant	Event Space	Restaurant	Pub	Hostel	Nightclub
6	50677	Italian Restaurant	Café	Bakery	French Restaurant	Plaza	Restaurant	Sushi Restaurant	Ice Cream Shop	Salad Place	Supermarket
7	50678	Café	Plaza	French Restaurant	Italian Restaurant	German Restaurant	Restaurant	Bar	Theater	Sushi Restaurant	Movie Theater
8	50679	Restaurant	Hotel	Bakery	Tram Station	BBQ Joint	Café	Platform	Pizza Place	German Restaurant	Supermarket
9	50733	Supermarket	Café	Park	Italian Restaurant	Bakery	Bar	Plaza	Doner Restaurant	Bookstore	Greek Restaurant

Figure.12 One-hot encoding of the venues

Now that we have one-hot coded all categories to all postal codes, we'll be able to use KMeans. It is a machine-learning technique that tries to partition the dataset into K-pre-defined distinct non-overlapping clusters where each data point belongs to only one cluster. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different as possible.

Colour	Name	Description
Firebrick	City Cluster	City + S-Bahn
Red	Suburbs	Access to KvB and/or S-Bahn
Orange & Yellow	Suburbs	End of lines of KvB
Other	Outliers	Blue, Purple, Green, Cyan, Magenta

Figure.13 Categories of clusters via KMeans method

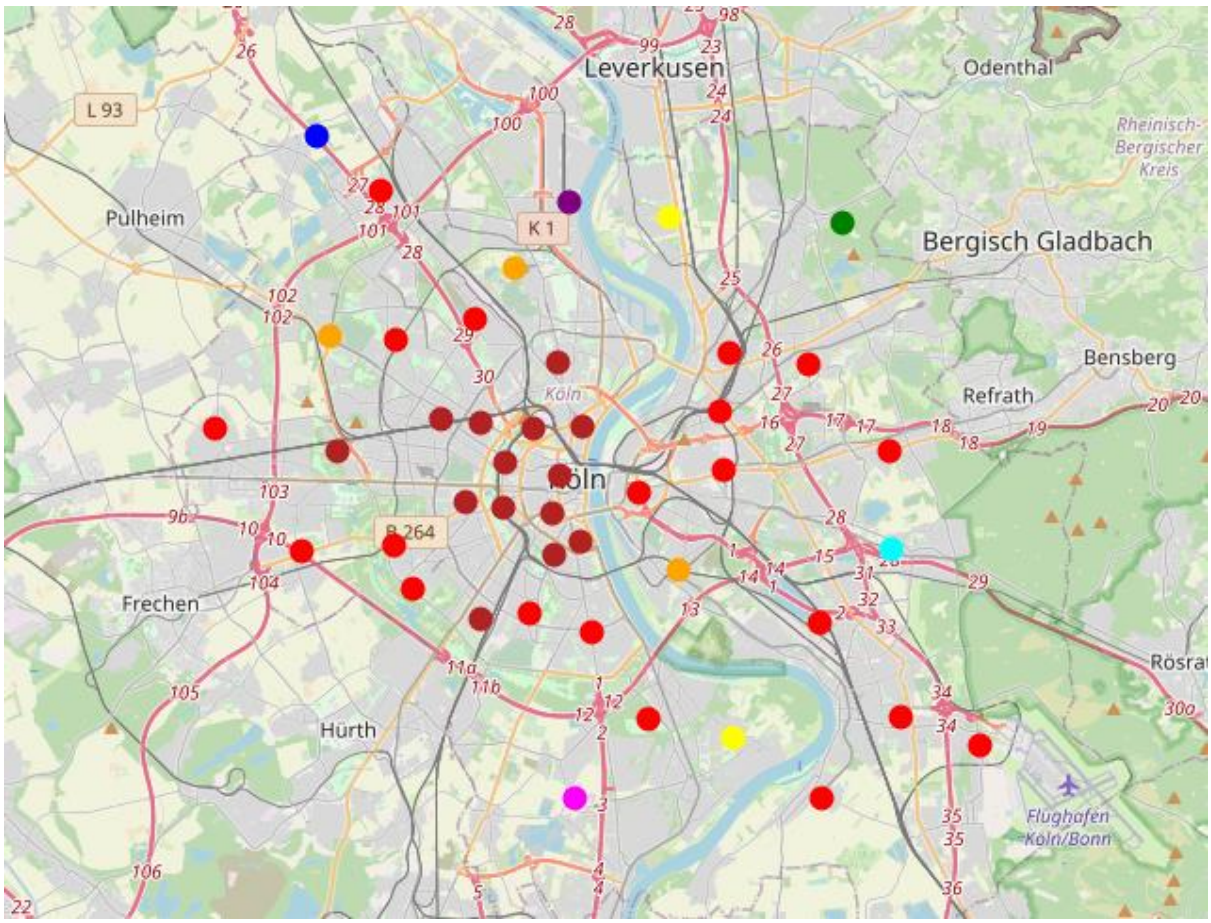

Figure.14 Map of the different clusters of Cologne based on venue types

If you wish to explore map and click on the different pins, I recommend to visit the Jupyter Notebook used for this analysis. The link to this specific map can be found here.

No surprises here. Postal codes closer to the city centre tend to belong to the same cluster, meaning there is a correlation between their types of venues. Let's call this firebrick-coloured cluster:

"The City cluster"

The Regio/S-Bahn train network (dark grey lines) seems to help suburbs and their venues develop like in the city:

- **50933** (Müngersdorf), **50939** (Klettenberg), which should have been technically red or yellow, are part of "*the city cluster"* and therefore has venues similar to the city.

The KvB train network (light grey lines) seems to support help correlate the red cluster and its venues:

- It's mostly visible with **50859** (Lövenich), **51063** (Mülheim) and **51109** (Brück).

The outliers (Blue, Purple, Green, Cyan, and Magenta) are the ones rather distant from downtown and/or not directly deserved by train tracks. They went their own way when it comes to their venues. They are also different from one another.

Results

Here are all scores for Nippes:

1. 3rd bracket in Rent prices (**dark orange**). We'll investigate the postal codes in the 2 upper brackets (**red** and **firebrick)**.
2. 1st bracket in density (**firebrick**). We'll investigate the postal codes in the current bracket.
3. *"The City Cluster"* in clusters. We'll investigate within the current cluster.

There are 4 postal codes which correspond to the criteria:

Postal Code	Neighbourhood	Colour Rent	Colour Density	Cluster
50674	Neustadt Nord/Süd	Red	Firebrick	City Cluster
50676	Altstadt-Süd	Red	Firebrick	City Cluster
50677	Neustadt-Süd	Red	Firebrick	City Cluster
50678	Neustadt-Süd	Firebrick	Firebrick	City Cluster

Now that we have the potential postal codes to be inspired from, let's sum up all of their top 10 categories and see what comes up the most frequently. From this, let's subtract the venues already existing in Nippes and let's make a top 5 of the other venues.

	Count
Sushi Restaurant	3
German Restaurant	2
French Restaurant	2
Nightclub	2
Movie Theater	1

Figure.15 Top 5 of missing venues in Nippes

The top 5 results are: Sushi restaurant, German restaurant, French restaurant, Nightclub and Movie theatre.

Discussion

Different analysis have been performed to find the ideal upcoming top 5 venues for Nippes, Cologne.

We considered rent prices, population density, and clusters of postal codes with similar venues. By narrowing it down based on those analysis, we ended up with a final group of 4 postal codes from which we could get inspired. Then we summed up their top 10s and sorted them by frequency. After subtracting the top 10 venues already existing in Nippes, we could find out that the following are the potential ideal venues for Nippes:

- Sushi restaurant
- German restaurant
- French restaurant
- Nightclub
- Movie Theatre

Those recommendations are based on the precision of the tools we used. At some point, we realized that the Foursquare API was not providing any results for the postal code 50769 Fühlingen-Roggendorf, so we had to take it out of the research. It was not at a great loss, since the majority of the postal code's area is a Ford factory and a golf course.

The KMeans analysis required some tuning when it came to decide the number of clusters to use. The variation was not that steep enough to find a significant winner, although we did pick the number of cluster with the highest slope (9 clusters).

There are also other factors which could have been taken into account, i.e. the average income per household, properties to sell within Nippes for business purposes, parking lots location, etc.

Conclusion

These recommended venues are a good starting point for further analysis for entrepreneurs looking for inspiration, businesses looking to expand, or even city council members looking to improve the lifestyle of their citizens.

As you may have realized with this quick analysis, a lot can be interpreted from data from many sources. Let's make sure we provide our decision makers with accurate data and relevant analysis for the benefit of our neighbours and fellow citizens.