

**Real Estate Investments:
Commercial Buildings,
Cologne, Germany**

for REI Commercial

by BJS Analytics

10. October 2019

1	INTRODUCTION	3
2	DATA	3
3	METHODOLOGY	4
3.1	Foursquare queries	4
3.2	Clustering	5
4	RESULTS	5
4.1	Population density	5
4.2	Venues	7
4.3	50733 - Nippes	8
5	CONCLUSION	9
6	REFERENCES	10

1 Introduction

A large real estate investment company named "REI Commercial" - would like to invest on commercial buildings in Cologne. Cologne is the 4th biggest city in Germany and it's crossed by the Rhein river.

To maximize their profit, REI Commercial wishes to invest in an upcoming neighbourhood where land and existing buildings are on the raise but not yet reached extremely high prices.

REI Commercial needs to know which kind of building it should invest on i.e. restaurants, supermarkets, shopping centres. The city area to invest on should have already an existing good offer of housing and services as this assures that shops are visited during working hours as well as after working hours raising the profit of the shops and therefore the value of the rents.

2 Data

The work here presented will be based on the number of inhabitants for each postal code area of the city of Cologne as well as the existing venues information stored by Foursquare.

The number of inhabitants will be acquired from 'suche-postleitzahl.org'. The names of the neighbourhoods in the respective postal codes is here collected from 'offenedaten-koeln.de' and the geometric coordinates of the different neighbourhoods from 'opendatasoft.com'. Foursquare data will be retrieved through its API.

Foursquare data will be used to understand the already existing consumer trends in the different areas of the city as well as the available and missing services in each respective area.

Ideally, house pricing and socio-economic status associated with the different postal code areas should be included but due to their high costs this data will not be included in this first report. We will instead assume that house pricing and socio-economics correlate with population density i.e. highly dense regions will have higher house pricing due to a high search and demand. Although neighbourhoods with a low population density can also associate with high house costs and incomes per

household such neighbourhoods are not attractive for investment in commercial buildings and can therefore be ignored for this analysis.

3 Methodology

3.1 Foursquare queries

Given the limited number of queries that can be performed and the number of results per query that can be retrieved from the Foursquare API we 479 queries over the city of Cologne (Figure 1). More precisely, queries were spread through 479 approximately equally distant locations. Each query covered a radius of 293 m.

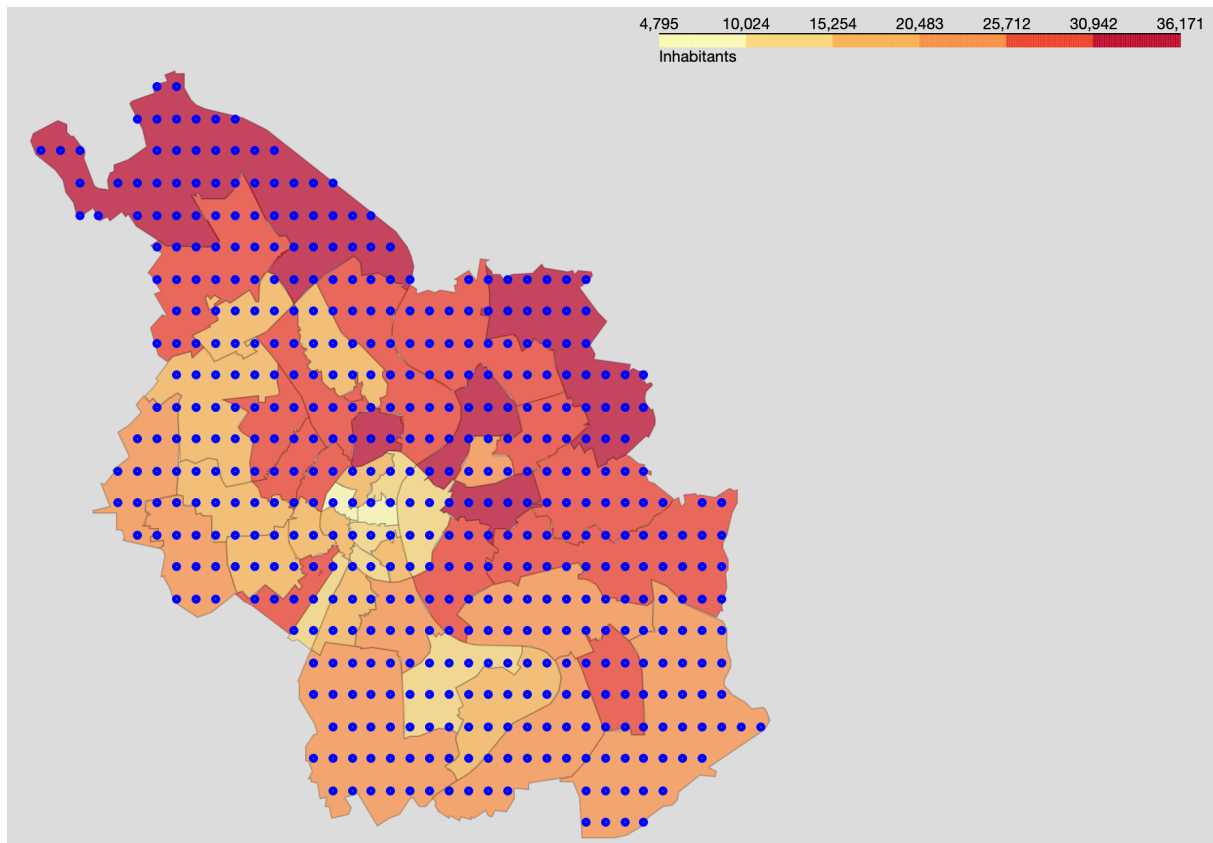


Figure 1. Foursquare queries coverage of the city of Cologne.

Due to the high variation between postal code areas when it comes to their size, population numbers, and number of venues data is often shown in a normalized fashion e.g. instead of “inhabitants / area size” the “ $\log_2(\text{inhabitants} / \text{area size})$ ” is shown. The sole purpose of this normalization is the avoidance of a colour mapping that is being fully driven by an outlier value with either too high or too low values that then might lead to the loss of contrast between the remaining city areas.

3.2 Clustering

To better understand the overall trends in the different parts of the city, postal code areas were clustered based on their venues categories using K-means clustering. k-Means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. k-Means minimizes within-cluster variances (squared Euclidean distances) (Ref. 1). To decide on the number of clusters (k), clusters in the range of 1 to 43 were tested and the respective within-cluster sum-of-squares (a reflection of the distance of each element to the centre of its cluster) was plotted in function of k (Figure 2).

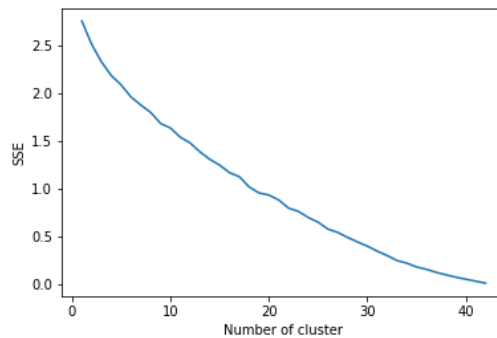


Figure 2. Elbow plot, K-means clustering.

As the optimal number of clusters could not be readily identified we visualized the different clusters on the map of Cologne given a variety of cluster numbers. A k of 3 was chosen as it nicely overlapped with population and venues density while being sense full for city with the size of Cologne which is crossed on the east by a river.

4 Results

4.1 Population density

Population density analysis identified the centre of the city as the most densely populated area (Figure 3).

Highly populated areas which have not yet reached the high density of the city centre could also be identified – 50733, 50823, 50674, 50676, 50677, and 50678 (Figure 4).

Given the assumption that middle-to-high populated areas correspond to up-coming neighbourhoods, this 6 areas represent possible target investment zones.

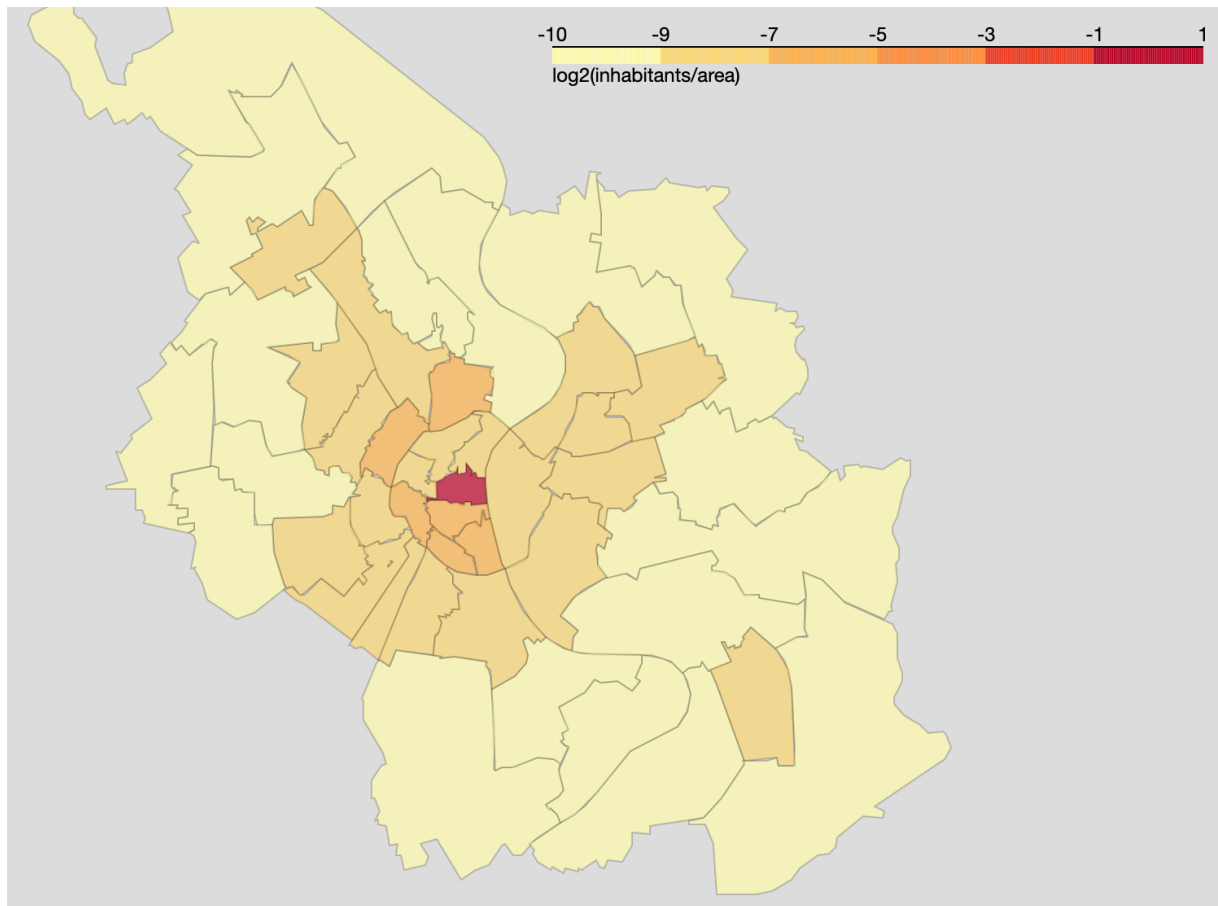


Figure 3. Population density.

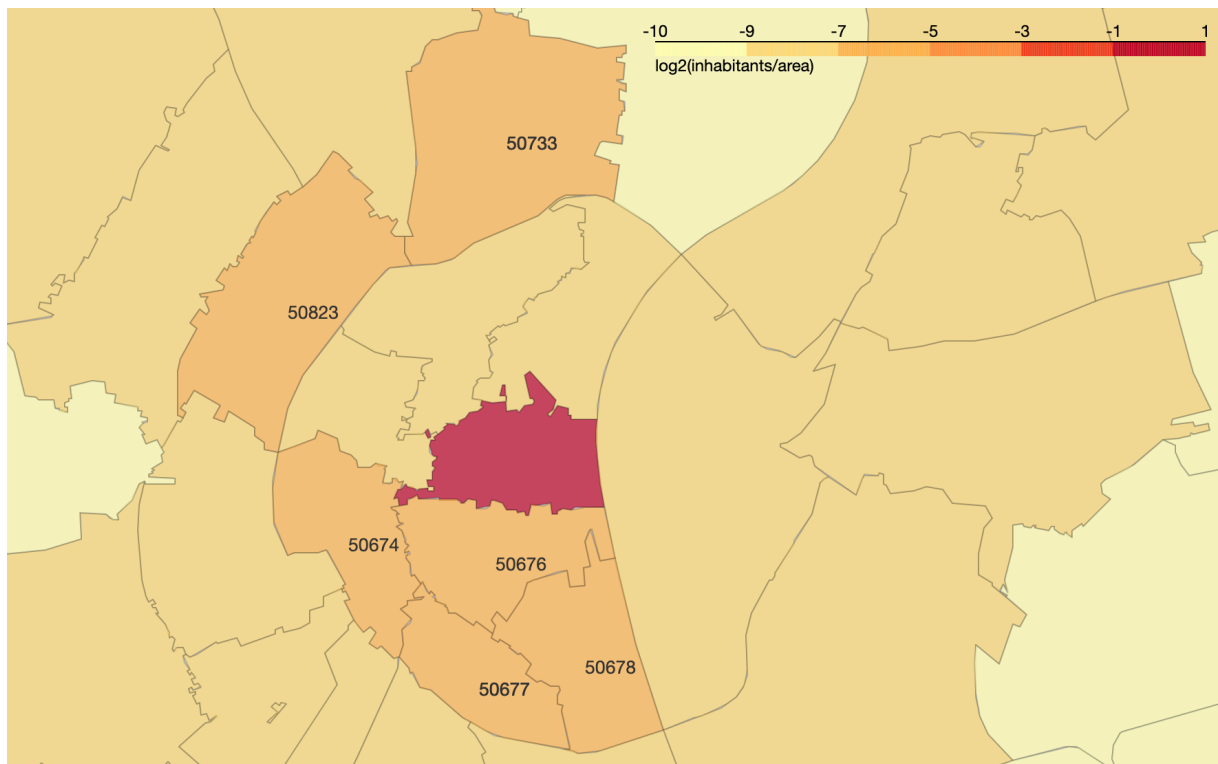


Figure 4. Population based identification of emerging neighbourhoods.

4.2 Venues

Foursquare data retrieval led to the identification of 3 areas with a high venues/area ratio (Figure 4, darker colour).

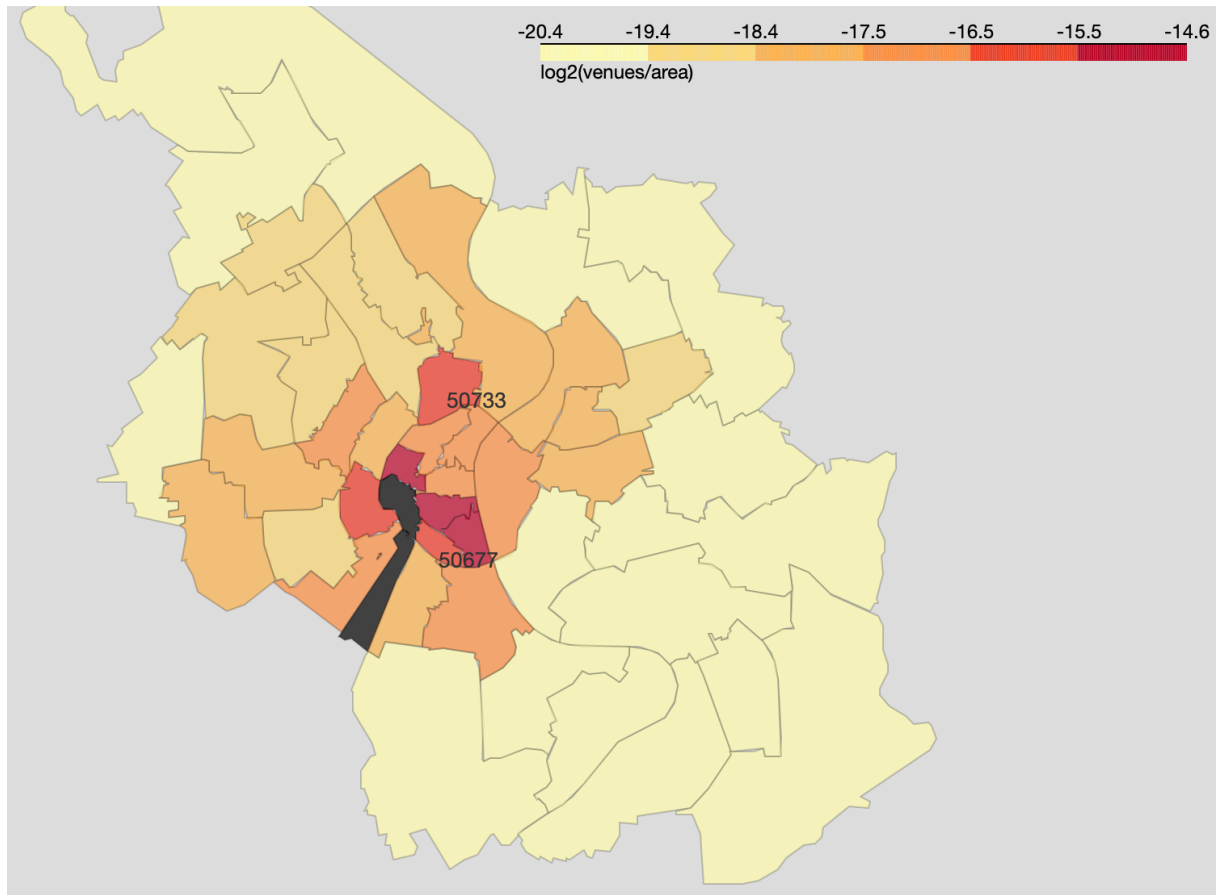


Figure 5. Venues density. The "number of venues / area size" was used to generate a color map.

Given again the assumption that these areas will suffer from high search and demand and consequently high investment costs these areas are out of our interest list. Two areas emerged with a middle-to-high venues/area ratio which had already been identified as potential candidate areas during our population analysis – 50733 and 50677.

Geographically, 50733 shows a bigger separation from regions with a high density of venues. Thus, habitants of this region will have in average longer distances to venues not yet existing in their neighbourhood. This makes 50733 an ideal place for investment.

4.3 50733 - Nippes

To analyse the characteristics of the current venues in area 50733 in relation to the rest to the city, k-Means clustering was used to cluster the different city areas in agreement with the characteristics of its respective venues (Figure 5).

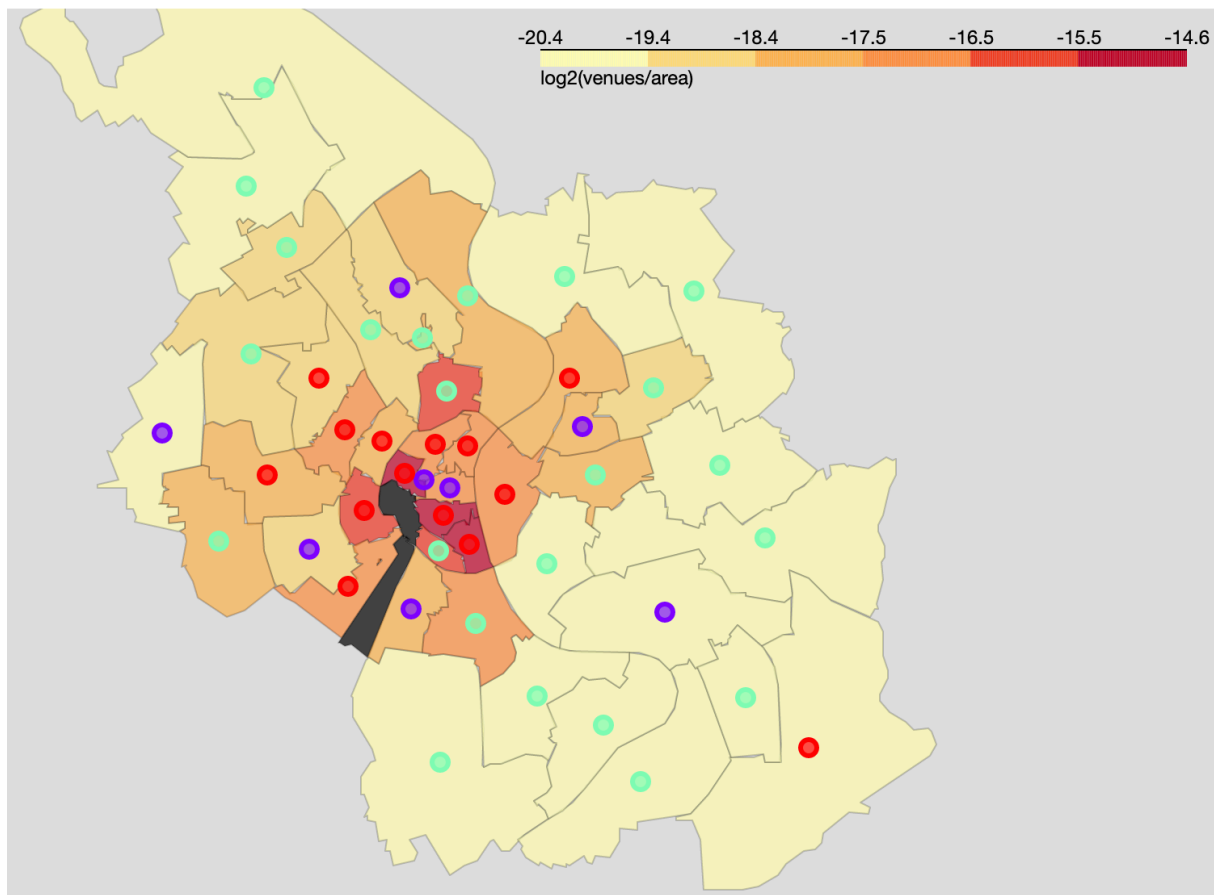


Figure 6. K-means clustering of different city areas based on the characteristics of their venues.

This allows us to separate existing trendy areas of the city from outskirts based on the characteristics of their venues. Interestingly both 50733 and 50677 areas do not group within the trendy areas cluster (red) although based on our population/area, venues/area, and geographical analysis they should both represent upcoming neighbourhoods. A cluster containing the city centre (purple) and some peripheric regions is also a markable result. This is probably driven by specific types of venues like shopping centres which lead to the agglomeration of certain characteristics which are normally not found in standard venues. In agreement with this, the purple marker on the south-east part of the city overlaps with the city's airport.

Having segregated city areas based on their venues we compared Nippes to the cluster that contained most of the trendy areas of the city (marked in red).

Table 1. Difference between 50733 (Nippes) venues and Cologne's trendy city areas. A higher value represents a needed venue for 50733 while a more negative value shows overrepresented venues. Top 20 needed and existing venues are shown.

	needed venues		existing venues
Hotel	0.072	Supermarket	-0.105
Bar	0.028	Bakery	-0.057
Italian Restaurant	0.026	Drugstore	-0.046
Nightclub	0.022	Doner Restaurant	-0.039
Restaurant	0.022	Martial Arts Dojo	-0.024
German Restaurant	0.016	Thai Restaurant	-0.024
Pizza Place	0.016	Kebab Restaurant	-0.024
Pub	0.013	Department Store	-0.024
Burger Joint	0.013	Bank	-0.024
Spanish Restaurant	0.012	Farmers Market	-0.024
Snack Place	0.012	Cantonese Restaurant	-0.024
Trattoria/Osteria	0.012	Plaza	-0.023
Asian Restaurant	0.012	Organic Grocery	-0.022
Austrian Restaurant	0.011	Music Venue	-0.022
French Restaurant	0.010	Bookstore	-0.022
Tapas Restaurant	0.009	Flea Market	-0.022
Automotive Shop	0.009	Modern European Restaurant	-0.021
Theater	0.009	Cocktail Bar	-0.019
Fast Food Restaurant	0.009	Café	-0.019
Water Park	0.008	Breakfast Spot	-0.018

Nippes – 50733 – could clearly profit from hotels, bars, and restaurants. On the other hand, the neighbourhood already shows a strong offer of supermarkets, bakeries, drug stores, department stores and bookstores.

5 Conclusion

Given the available data on population and venues we could identify the postal code area 50733, locally known as Nippes, as an ideal area for an investment in commercial buildings. Both population and venues numbers indicate that this neighbourhood is trending but not yet as saturated as other city areas. Nippes is geographically isolated from other trendy areas of the city which makes it's venues a potential hotspot for its residents but also for the surrounding neighbourhoods.

Nippes shows an overrepresentation of supermarkets, bakeries, drugstores, department stores, banks, book stores, and cafes. Such kind of venues are typical for residential neighbourhoods and are part of consumers everyday life. This clearly indicates that there is a high flux of people through the commercial venues of this neighbourhood. Additionally, although certain types of restaurants are already over-represented (e.g. Thai, Kebab) others are still under-represented (e.g. Italian, French).

Given the venues already existing in **Nippes** and the venues existing in the remaining trendy areas of the city we recommend investing in either a large space suitable for an **hotel** or smaller spaces suitable for **restaurants**. Both hotels and restaurants do not suffer from competition with online shopping and will therefore keep on being looked for in the years to come.

6 References

[1] wikipedia.com