On the Preferred Location of a Multifaceted Bar Business in Paris

I - Introduction and business problem description

Paris has been one of the most touristic cities in the world for a while now. It actually tops tourists' destination since at least the beginning of the 2010s, alternating this first place with London (see AFP, 2014; AFP, 2019a). [1]

Around 16 to 17 million foreign visitors visit Paris every year, with around 450 million nights spent by tourists for accommodation (as of 2018), including hotels, campsites, and youth hotels— yet excluding home-sharing platforms like Airbnb, (see AFP; 2014)[2]. The main 60 touristic sites or so total around 80 million visits yearly, with —according to the *Office du Tourisme et des Congrès de Paris*[3]—roughly 14 million visitors for the *Cathédrale Notre-Dame de Paris*, 11 million for the *Basilique du Sacré-Cœur de Montmartre,* or 9 million for the *Musée du Louvre*.

Meanwhile, as of 2015, Paris is one the most expensive—and as a corollary, on average, wealthy—city in the world (see AFP, 2019b quoting a 2019 Economist Intelligence Unit report),[4] trailed by Zurich, Geneva and Osaka.[5] Even more, Paris was the only eurozone capital in the top 10 most expensive cities in the world, with a woman's haircut and a two-piece business suite respectively quoted around 150 and 1700 euros in 2018.

Against this background, I was approached by a foreign investor who envisages to open a *bar/pub* in Paris. The investor seeks advice on the location of such a business, a location that would tap both the tourists' population—acknowledging the fact that numerous, if not a fair majority of, tourists are as interested in the city's nightlife, as they are in its cultural sites—as well as the local, resident clientele which will support smoothing the business's revenues in view of the seasonality of tourism influxes. The business would also function as a coffee shop during daytime. Tapping both foreign and resident clienteles, as well as mainly night (and secondarily day) consumers needs is intended to ensure a solid, long run, financial viability of the project.

---

[1] See respectively https://www.thelocal.fr/20140311/paris-still-worlds-top-tourist-destination, and https://www.thelocal.fr/20190410/france-retains-crown-as-most-visited-country-on-earth.

[2] See https://www.thelocal.fr/20190410/peak-year-for-tourists-in-france-despite-yellow-vests-and-rail-strikes.

[3] See https://pro.parisinfo.com/etudes-et-chiffres/enquetes-et-dossiers.

[4] https://www.thelocal.fr/20190319/paris-tops-ranking-of-worlds-most-expensive-cities-to-live-for-the-first-time.

[5] https://www.france24.com/en/20190319-paris-france-world-most-expensive-cities-singapore-hong-kong; see also https://www.novinite.com/articles/195890/Paris+is+Now+the+Most+Expensive+City+in+the+World.

The preceding paragraphs have laid down some of the main components of the business problem which can be simplistically stated as follows: **what is the best location(s) for starting the aforementioned, multifaceted business in Paris?**

The area of the city exceeds 100 km2 indeed. It counts 2.2 million inhabitants with wide spatial variation in income levels and other sociodemographic characteristics, and is organized in 20 *arrondissements* (let us view this as a borough), and 80 administrative neighborhoods. In this preliminary report, I identify such a location at the *arrondissement* level.[6]

Figure 1[7]

Paris's map
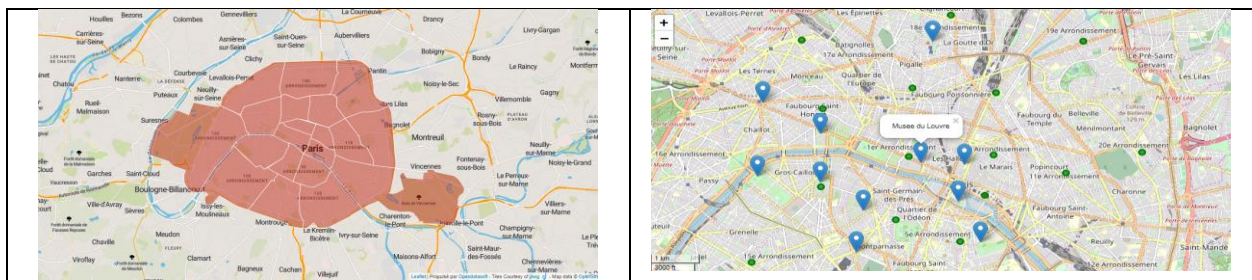
Note: Please click here to interact with the map and identify the main touristic sites in Paris along with the *arrondissement* they belong to. See also the notebook for the report's related Python code.

II - Data description

Giving the investor solid scientific insights into the most favorable locations for the business under consideration requires a diverse set of data—a total of 16 variables that required preprocessing and cleaning to a significant degree—to capture various characteristics of the prospective catchment area(s) that will be advised. Here, I describe the data and its sources, along with some visualization for descriptive—rather than analytical—purposes.

II.1 Foursquare sourced data

I make use of Foursquare to obtain a first group of variables that relate to the business's direct and indirect competitors, namely the bars, nightclubs, and restaurants. A high concentration of such businesses in a particular area is assumed to reduce the attractiveness of that location for our investment.

Among the 90 different categories of venues rendered by Foursquare, it appeared that the category *Bars* are referred to by the following labels: Bar, Pub, Brasserie, Bistro, Brewery, Nightclub, Beer Garden, Gastropub, Lounge. Similarly, the *Restaurants* category is referred to

---

[6] I will be using the French word *arrondissement* throughout this report.
[7] Please note that this report uses APA guidelines for figures and tables titles formatting.

through the labels Restaurant, Diner, Creperie, Fish & Chips Shop, Snack Place or Tea Room; while the labels Theater, Comedy, Club and Movie Theater refer to our category *Theaters*.

I also use Foursquare to extract data related to two more categories that I label *Hotels* and *Museums* (the latter covered by the labels Museum, Art Museum, History Museum, Monument / Landmark, and Science Museum). These are expected to have a favorable impact on my recommendation for a given location, in view of the fact that a high concentration of these increases our business's patronage, by night and day respectively. Figure 2 gives a snapshot of the five categories' counts.

Figure 2

The counts of Bars, Restaurants, Theaters, Museums, and Hotels per arrondissement, as per Foursquare data
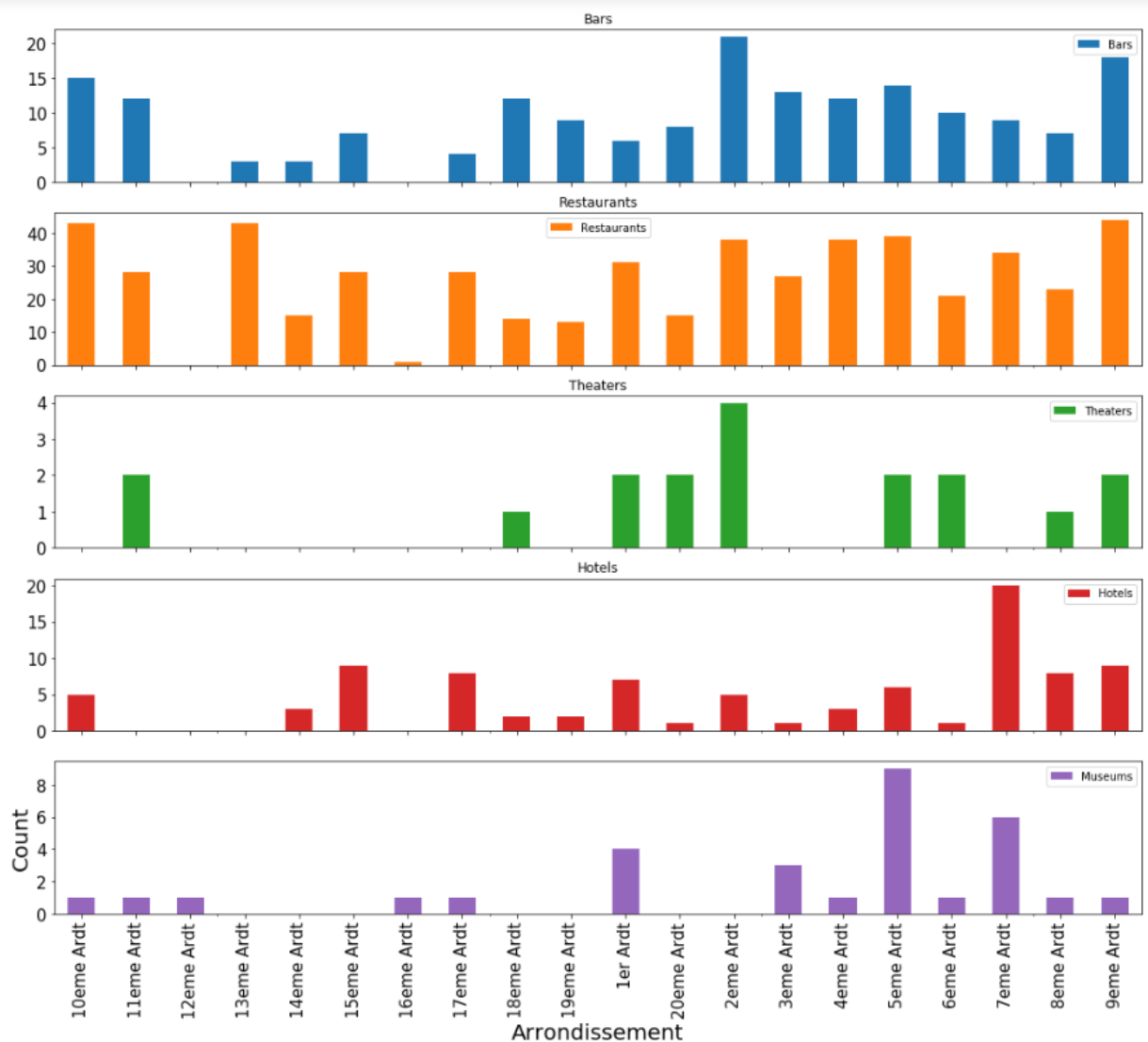
It is important to note that this data does not represent the full count of businesses in Paris's 20 *arrondissements*, but only those businesses that appear in Foursquare, which in the case of a non-anglophone city like Paris, reflect tourists' testimonies. This warrants strong skepticism over utilizing such data for decision-making purposes in a real-life problem solving.

Below are a couple of more figures that compare *Bars* and *Restaurants* counts (Figure 3). This shows, for example, that *Restaurants* testimonies are more frequent than Bars testimonies, even though there are differences in the *arrondissements* in this respect.

Figure 3

*Visualization of the counts of the Bars and Restaurants categories*
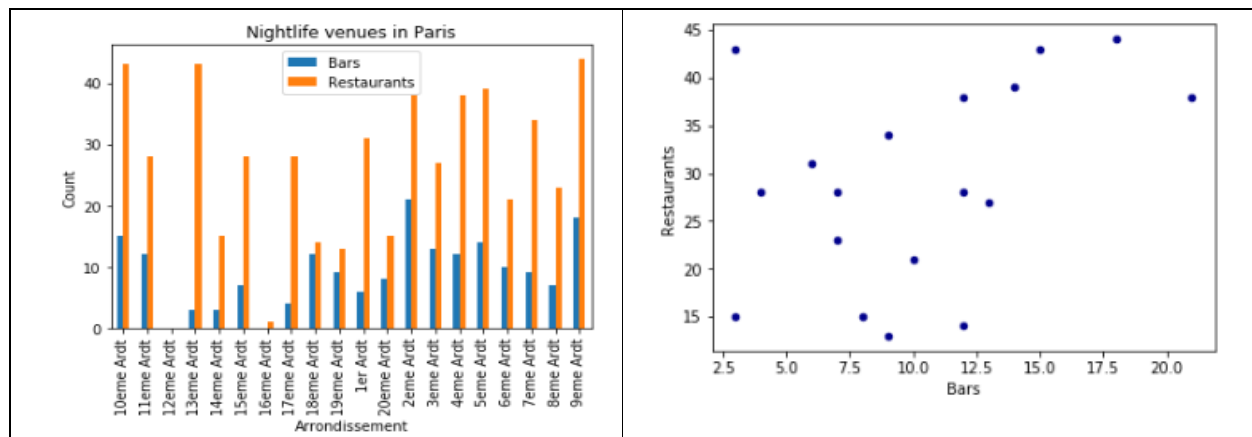
Secondly, and I will get back to this later, while this data gives a sense of the raw numbers of businesses listed in Foursquare, it needs—at least part of it—to be scaled to take into account the differences of the area and population of the *arrondissements* if it is to be used meaningfully in a quantitative analysis. Hence the *Bars* data for the 4th and 7th *arrondissements* for example, are not directly comparable as their areas are 7 and 16 square kilometers, and their populations of 182 000 and 165 000 respectively.

Consequently, I scale the *Bars*, *Restaurants*, and *Theaters* counts by dividing these by the *arrondissements'* population densities (number of inhabitants per square kilometers). Further making use of a variable that classifies the *arrondissements'* three categories which I label *Least*, *Moderately*, and *High touristic* according to their attractiveness to tourists (see paragraph II.2.4), I am able to produce a matrix scatter plot that gives interesting insights into this extracted Foursquare data (see Figure 4). For instance, one can see that high touristic sites have received more assessments for their bars and restaurants which lends support to the fact that those businesses that are established in high touristic locations rely proportionately more on tourists' consumption for their revenues and profits—assuming the behavior of tourists, as it relates their decision to assess bars and restaurants in Foursquare, is not skewed across touristic attractiveness or patronage.

Figure 4

Matrix scatter plot and densities of the Hotels and Museums counts, and the normalized Bars, Restaurants, and Theaters indexes
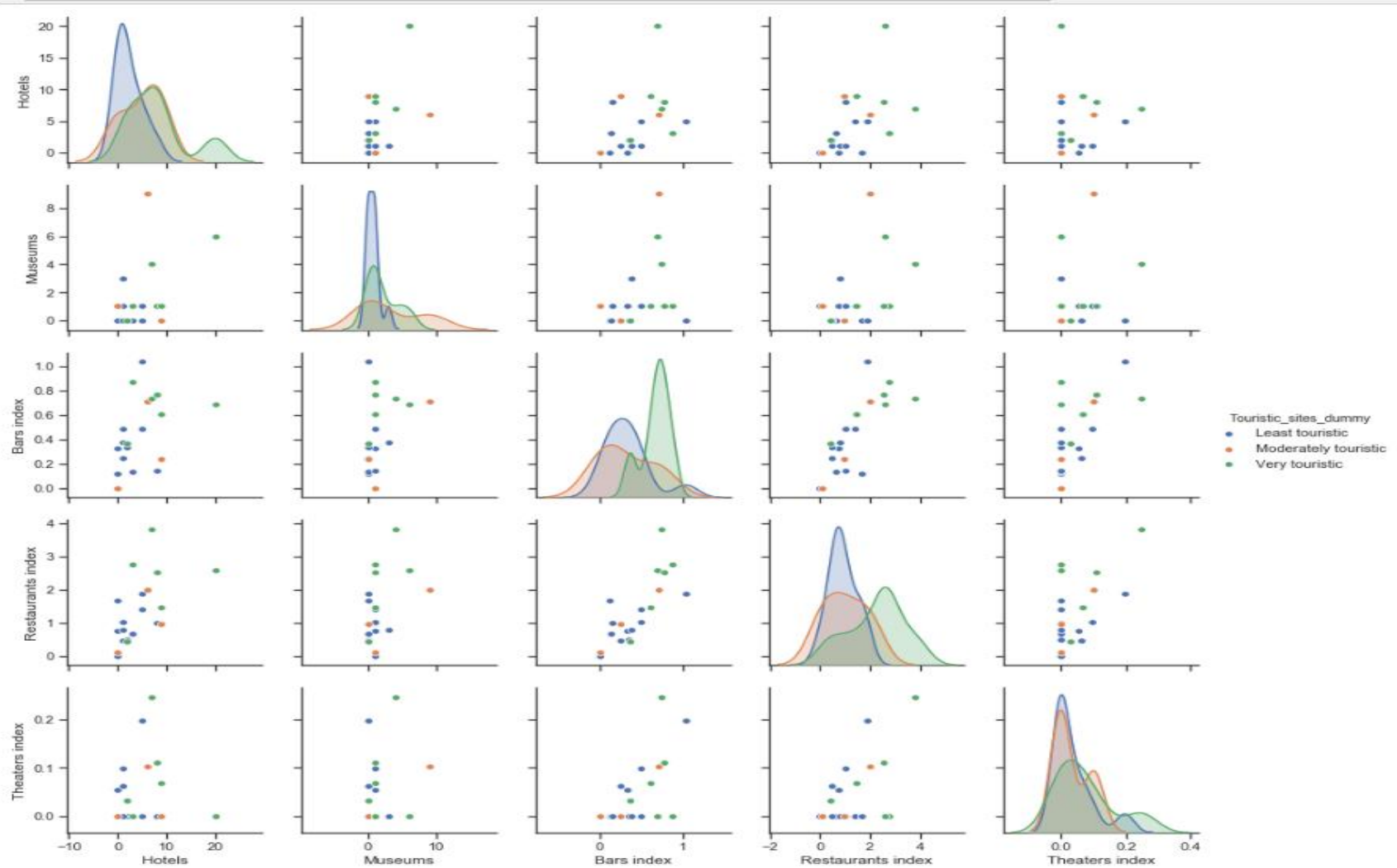
II.2 Socioeconomic, demographic, and touristic data

A second set of data are socioeconomic, demographic, and touristic data, per arrondissement, which I source in official statistics. These related to wealth, population, real-estate, and touristic sites statistics.

II.2.1 Wealth data

Wealthier locations are more attractive for any business. I use three socioeconomic indicators to capture that wealth effect, namely the poverty rate, the median annual income, and the percentage of taxable income households (sourced in French National Institute of Statistics and Economic Studies, INSEE). [8] Based on these, I construct a composite indicator, after normalizing these variables using the min-max method. This allows the ranking of the *arrondissements* according to their wealth level.

II.2.2 Population data

Similarly, I create a composite indicator based on three variables related to the *arrondissements'* population, namely the total population, the population density, and the percentage variation of population over 2013-2016. A higher value on that indicator increases the attractiveness of a given location.

II.2.3 Real-estate price data

Locations with high real-estate prices negatively impact my recommendation for a given location, even though they indirectly correlate to the wealth indicator. I retain two variables, the 2018 square meter price in euro, and the percentage change of that price over 2018–2019 that captures the short run trend in real-estate prices, which I use to create the real-estate composite along the same lines described in the preceding to paragraphs.

II.2.4 Other demographic and tourism indicators

The final set of indicators first includes one demographic variable, the 15–64 age category population density—our target local clientele— which is assumed to positively impact the assessment of an *arrondissement* location. The last two indicators are related tourism. The first of these adds up the tourism data of the touristic sites—among the 15 most visited sites in 2018— belonging to the various *arrondissements*, while the last variable classifies the arrondissements into three categories according to the degree to which they were preferred destinations for tourists.

---

[8] All socioeconomic and demographic data is official statistics provided by the French National Institute of Statistics and Economic Studies. See https://www.insee.fr/en/accueil.

Table 1

The set of supplementary demographic and tourism indicators

Arrondissement Number	15-64 population density	Cumulative tourism data*	Tourism patronage
10eme Arrondissement	23328		Least touristic
11eme Arrondissement	29496		Least touristic
12eme Arrondissement	5988		Least touristic
13eme Arrondissement	17385		Least touristic
14eme Arrondissement	16703		Least touristic
15eme Arrondissement	18903	1.17	Moderately touristic
16eme Arrondissement	6369	1.4	Moderately touristic
17eme Arrondissement	20720		Least touristic
18eme Arrondissement	23558	11	Very touristic
19eme Arrondissement	18724		Least touristic
1er Arrondissement	6266	9.13	Very touristic
20eme Arrondissement	22745		Least touristic
2eme Arrondissement	15868		Least touristic
3eme Arrondissement	22173		Least touristic
4eme Arrondissement	12231	17.45	Very touristic
5eme Arrondissement	16076	3.62	Moderately touristic
6eme Arrondissement	12633		Least touristic
7eme Arrondissement	8463	16.5	Very touristic
8eme Arrondissement	6458	7	Very touristic
9eme Arrondissement	19877		Very touristic

* Based on the 15 most touristic sites in Paris in 2018.

III. Methods

For the sake of ensuring the robustness of the recommendations, I rely on three ranking methodologies. Two of these are classification methods: *kmeans clustering*, with four clusters, and *agglomerative hierarchical clustering* which are robust techniques that rely on different algorithms for suggesting grouping of the *arrondissements* according to their proximity based on the 16 indicators described in the preceding paragraph. Hence, the *arrondissements* that show similar magnitudes in their statistical figures will be grouped together.

The third approach relies on the construction of a multilevel nested composite index of the 16 indicators as per best practices (see OECD, 2008).[9] Transformation, aggregation, and weighting choices are kept rather simple in the context of this report. The methodology can be illustrated schematically as shown in the following chart. At each level, the indicators are normalized, their order reversed where appropriate, then summed—a way of taking their simple average— and

---

[9] See the JRC/OECD guidelines on https://www.oecd.org/sdd/42495745.pdf

finally the resulting index is normalized once again. Paragraphs II.2.1 to II.2.3 describe the indicators involved at the fist nesting level.

The indicators and indexes of the second level are then normalized, reversed if needed, averaged, and finally normalized again to obtain the final, overall score per *arrondissement*.

Figure 5

*The multilevel nested composite index structure*
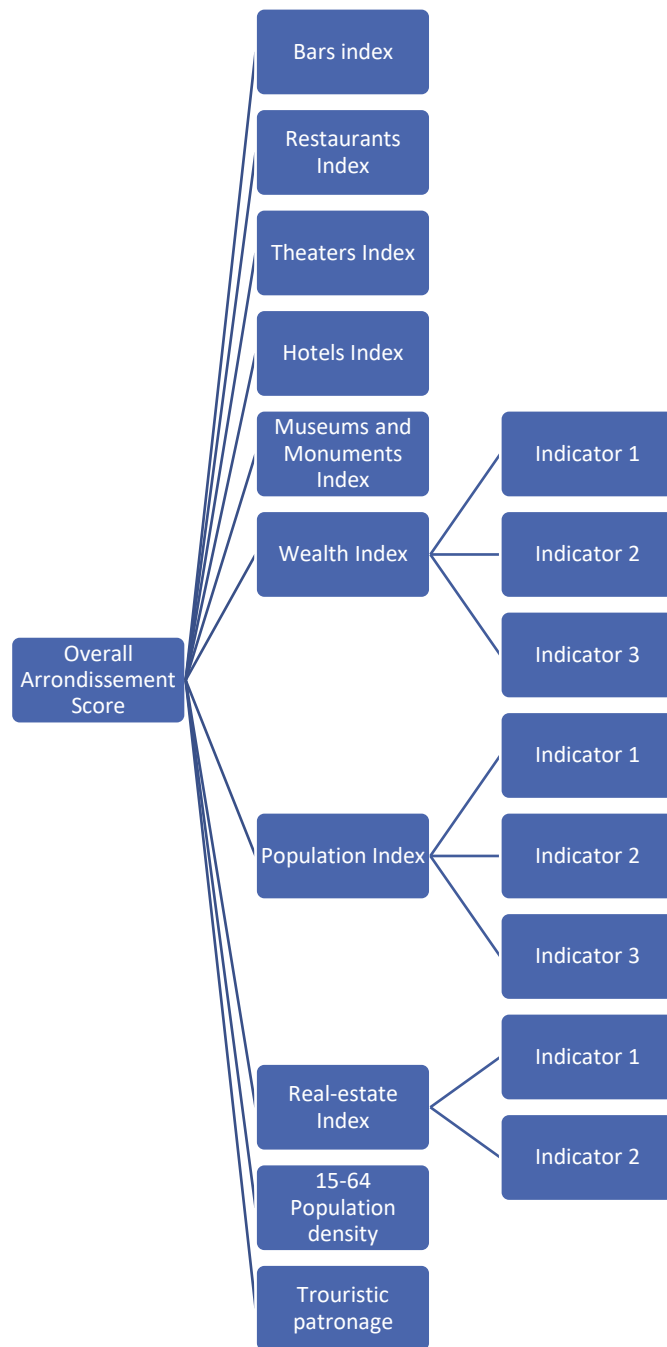
In sum, this multipronged quantitative approach yields more assurance in my final recommendations based on the comparison of three widely used alternative ranking and clustering methodologies.

IV- Results and discussion

Figure 6 summarizes the clustering results and the overall attractiveness scores for the 20 *arrondissements*. It is important to note the clusters numbers order, and the correspondence between these across the two clustering methods has no qualitative meaning. Simply put, with clusters numbers spanning from 0 to 3 for both methods,[10] the fact that an *arrondissment* is classified in cluster 3 does not mean it is more attractive than another that has been classified in cluster 1. Hence, the 1st and 2nd *arrondissements* were classified in cluster number 3 according to the kmeans classification, while they show the lowest attractiveness across all *arrondissements* according to the 16 indicators selected for this study.

Figure 6

Cluster numbers and overall attractiveness score of the 20 arrondissements

---
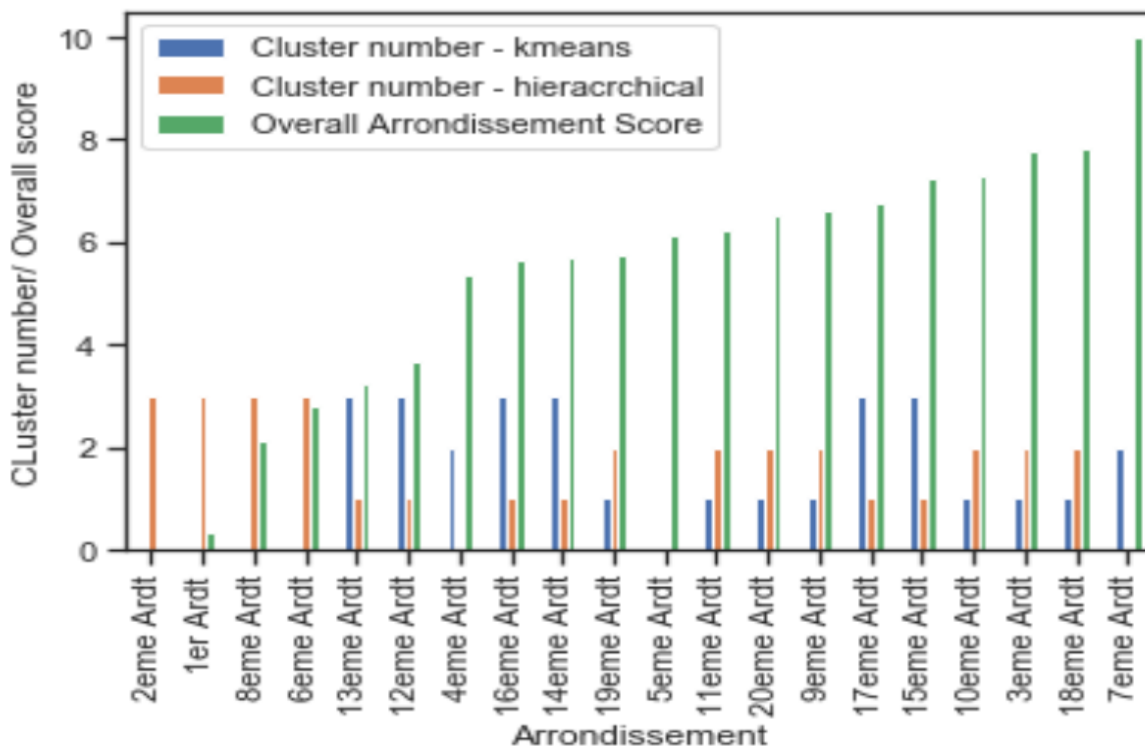
[10] I set the cutting number to 4 in the hierarchical clustering method for the purpose of comparing the arrondissements' classifications across both clustering methods.

Figure 6 warrants several comments. Firstly, the most attractive *arrondissements* appear to be the 7th, with an important lead, followed by the 18th and the 3rd. However, several other *arrondissements* are also worth considering, notably the 10th , 15th, 17th and 9th. The contributions of the various indicators to the high attractiveness scores of these "second best" arrondissements—which one can assume to differ to some extent—remain to be investigated though.

Secondly, the most attractive clusters are classified in cluster number 2 (which only includes the 7th *arrondissement*) and 1 by the kmeans algorithm, and in cluster number 3 by the hierarchical clustering algorithm. A perfect correspondence is also apparent between cluster 3 and 1, and clusters 1 and 2, in kmeans clustering and hierarchical clustering respectively. Further, theres is a correspondence, at the low end of the attractiveness distribution, between cluster 0 in kmeans clustering, and cluster 3 in hierarchical clustering.

In order to identify qualitatively the characteristics of the *arrondissements* pertaining to the four clusters, I investigate 1) the average values of the various indicators for both classification methods (Table 2); and 2) the correlation structure of the indicators, cluster numbers, and overall attractiveness score (Figure 7). Table 2 shows the two clustering methods actually yield very similar averages, reinforcing confidence in the conclusions one can draw from these methods.

Kmeans clusters 1 and 2 show the highest overall attractiveness scores yet the characteristics of the *arrondissements* they group are very different (Table 2). In effect, while cluster 1's arrondissements show low bars, restaurants and theaters competition, this competition is much fiercer for cluster 2 as regards bars and restaurants. Similarly, cluster 1 groups arrondissements with much more favorable local population dynamics, while cluster 2's arrondissements are much more favorable in terms of the income levels their residents, as well as from a tourism revenues perspective, through a high patronage for their touristic sites, and more hotels penetration. Meanwhile both clusters are characterized by favorable real-estate price dynamics. In sum, if the objective is to tap touristic revenues, both through day and night time, cluster 2's arrondissements are our best shot, yet at the price of a much fiercer direct and indirect competition. Cluster's 2 arrondissements would be the choice of a business more directed towards a local resident, more populous, less wealthy clientele.

As for the remaining two clusters, cluster 0 resembles cluster 1 to a great extent, yet with much less favorable real-estate prices and fiercer business competition; while cluster 0 shares many of cluster 1 's features with unfavorable population dynamics.

The analysis of the averages of the hierarchical clustering classes leads to similar conclusion in view of the clusters' correspondences previously mentioned.

Table 2

Average clusters values for the socioeconomic, demographic, and touristic indexes

Cluster number - kmeans	Hotels	Museums	Bars index	Restaurants index	Theaters index	wealth zscore	realestate zscore	population zscore	15-64 population density	Cumulative_tourism_data	Overall Arrondissement Score	Cluster number - hieracrchical
0	0.270000	0.333333	0.277899	0.411769	0.388472	0.769790	0.668508	0.178050	0.232780	0.226361	0.228500	2.4
1	0.142857	0.095238	0.621549	0.782960	0.876017	0.347663	0.749715	0.786962	0.716990	0.090053	0.685442	2.0
2	0.575000	0.388889	0.248114	0.298172	1.000000	0.822139	0.715841	0.269312	0.185426	0.972779	0.768986	0.0
3	0.166667	0.055556	0.898732	0.809020	1.000000	0.577075	0.163386	0.689791	0.355482	0.024546	0.537718	1.0

Cluster number - hieracrchical	Hotels	Museums	Bars index	Restaurants index	Theaters index	wealth zscore	realestate zscore	population zscore	15-64 population density	Cumulative_tourism_data	Cluster number - kmeans	Overall Arrondissement Score
0	0.483333	0.592593	0.270234	0.359160	0.862523	0.798947	0.631350	0.330625	0.266661	0.717670	1.333333	0.71786
1	0.166667	0.055556	0.898732	0.809020	1.000000	0.577075	0.163386	0.689791	0.355482	0.024546	3.000000	0.53771
2	0.142857	0.095238	0.621549	0.782960	0.876017	0.347663	0.749715	0.786962	0.716990	0.090053	1.000000	0.68544
3	0.262500	0.166667	0.268755	0.394428	0.338697	0.774097	0.720043	0.109250	0.183693	0.231089	0.000000	0.13172

Figure 7

Correlation matrix of the various indexes and the overall attractiveness score
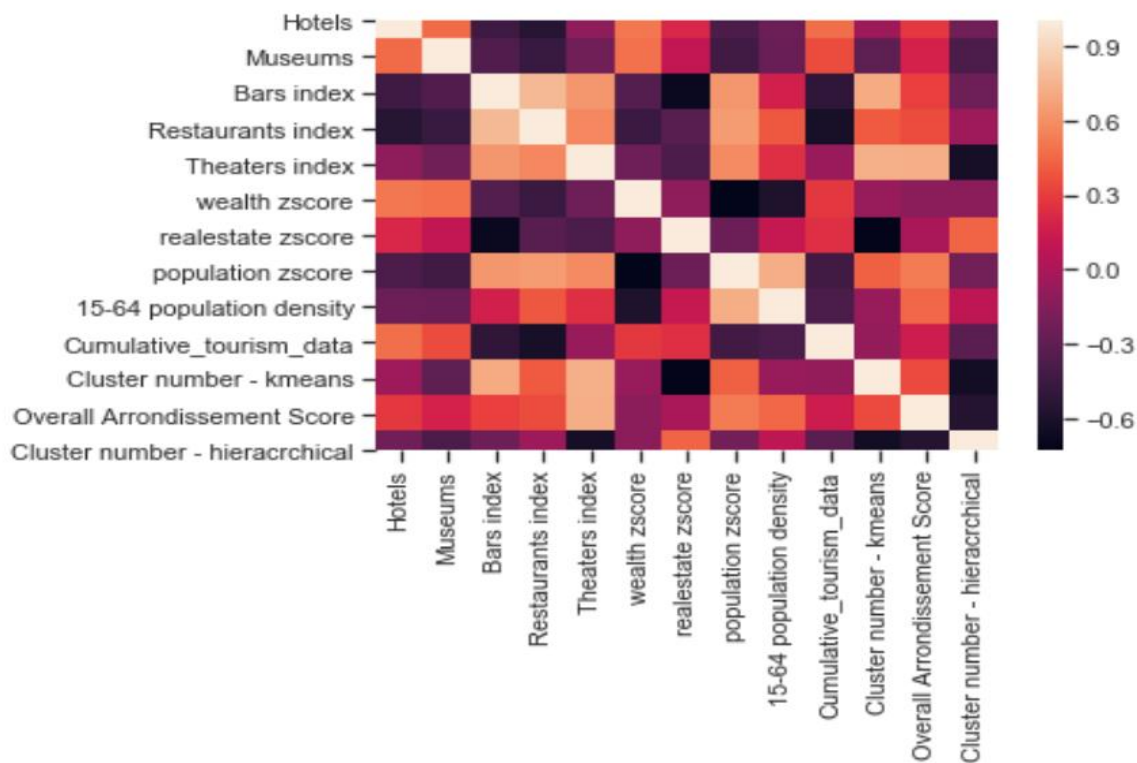
The previous analysis has made it clear that the *arrondissements'* high attractiveness scores build on very different constituents across the clusters. However, Figure 7—which notably pictures the correlation structure of that composite index with the various other indicators through a heat map—yields insightful association between attractiveness and the other variables. Specifically, an arrondissement with a higher attractiveness tends to have less competition, more favorable local population dynamics, yet less favorable real-estate prices and local residents' income levels.

Next, the recourse to a four-dimensional representation of selected indexes brings very interesting messages for our business problem. Figure 8 among others exemplifies, once more, one of the most important arbitrages we have to consider, clearly showing that more favorable businesses competition is highly associated with favorable population dynamics, yet higher real-estate prices.

Figure 8

Bars index, Population index and Real-estate index (bubble size) for the 20 arrondissements in Paris
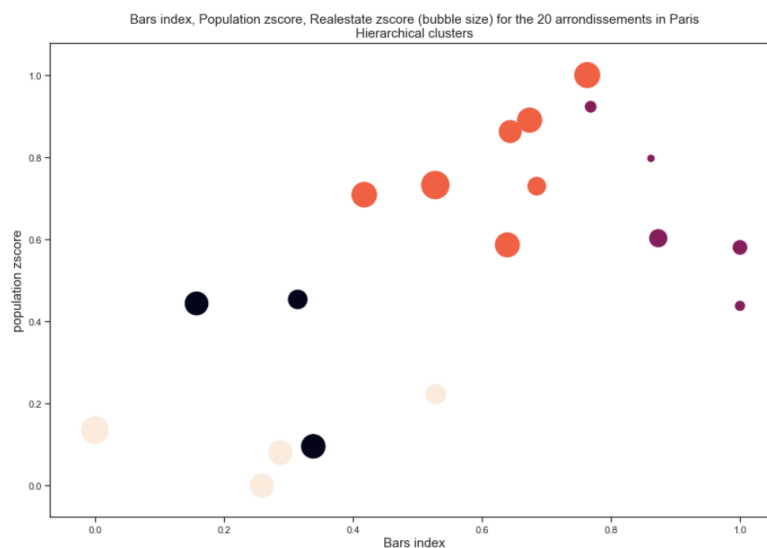
Aa a final perspective, let's order our data according to touristic patronage, bars competition, real-estate prices, and wealth levels. Figure 9 shows an excerpt of the resulting data allowing us to decide our issue. Our attention here is on the most touristic sites in Paris. Clearly, the *arrondissement* to recommend to our investor will have to strike a balance between the various factors within those arrondissements where these touristic sites are located.

Take the 4th *arrondissement*: while it hosts the most touristic sites in Paris, has a strong business competition, a low hotel penetration, and a very unfavorable youth population profile. As for the 1st *arrondissement*, while it has a fair tourism patronage score, it scores very low across the board with the exception of a couple of indicators.

Finally, our choice reduces to the 7th and 18th *arrondissement* which are both very touristic with favorable real-estate prices, yet differ drastically on all other indicators. Indeed, establishing our business in the 7th *arrondissement* as compared to the 18th, would lead to a relatively higher share of revenues coming from tourists during daytime, and taping a wealthier local albeit less important resident population, while facing a stronger business competition. This competition would be much less pronounced in the 18th *arrondissement* with our business relying much more on a younger, more important, yet less wealthy resident population.

Figure 9

Most touristic arrondissements characteristics

	Hotels	Museums	Bars index	Restaurants index	Theaters index	wealth zscore	realestate zscore	population zscore	15-64 population density	Cumulative_tourism_data	Cluster number - kmeans	Overall Arrondissement Score
4eme Ardt	0.15	0.111111	0.157628	0.275228	1.000000	0.644277	0.693803	0.443448	0.265569	1.000000	2	0.537972
7eme Ardt	1.00	0.666667	0.338600	0.321116	1.000000	1.000000	0.737878	0.095176	0.105283	0.945559	2	1.000000
18eme Ardt	0.10	0.000000	0.643890	0.887118	0.875023	0.122461	0.644563	0.862359	0.747405	0.630372	1	0.780785
1er Ardt	0.35	0.444444	0.287648	0.000000	0.000000	0.707303	0.731103	0.079737	0.011826	0.523209	0	0.032995
8eme Ardt	0.40	0.111111	0.258954	0.338439	0.554166	0.915989	0.707504	0.000000	0.019993	0.401146	0	0.212883

V- Conclusion

The most suitable location for our business, which is envisaged to tap both touristic and local populations, during night and daytime alike, involves various arbitrages that would impact the relative contributions of these parameters to the revenues and profits of our business. My recommendation boils down to the 7th and 18th *arrondissements* which appear to be promising locations for establishing our business.

From here, the analysis will have to consider that location at a lower spatial level, starting with the administrative neighborhood level (the so-called *quartiers administratifs*). Further, we need to take into account the official records regarding the numbers businesses actually established in those *arrondissements*. Finally, a further enrichment of the study would make use of the weighting of the Foursqaure data used so far with the clients' evaluations of the businesses.