

# Socio-Economic and Demographic Patterns in Video Streaming: Unveiling City Dynamics in France

Aslıgül Aksan\*, İrem Betül Koçak\*, and Oğuz Yücel\*

aksan20@itu.edu.tr, kocaki@itu.edu.tr, yucelog@itu.edu.tr

\* Management Engineering Department, Istanbul Technical University, 34469 Maçka, Istanbul, Turkey.

**Abstract**—As part of the NetMob Data Challenge, this study aims to identify demographic and socio-economic similarities and differences between the city of Paris, Montpellier, and the various arrondissements within Paris. We achieve this by analyzing the usage data of different streaming platforms among residents of these cities. In our analysis, we employ two distinct techniques to uncover urban segregation trends: exploratory data analysis and a form of statistical analysis known as ANOVA. Our findings indicate that not only do the demographic and socio-economic characteristics of cities influence streaming service usage, but also global trends driven by popular TV shows and movies can lead to substantial spikes in platform usage. Interestingly, these trends may not necessarily correlate with significant variations in demographics and socio-economic levels.

**Index Terms**—mobile data, demographic pattern, urban dynamics

## I. INTRODUCTION

This study is part of the challenge of The NetMob23<sup>1</sup>, which focuses on cities in France and the internet usage data of different platforms and services.

We aim to understand whether the choice of streaming platforms in two cities in France reflects the socio-economic situation of the people or the demographics of specific communes. While looking at this, we will examine whether there is segregation in the urbanized areas of cities in France.

Schelling's segregation model sheds light on the field of urban segregation studies, revealing that people's housing preferences are influenced by their social, ethnic, and economic backgrounds [1]. This finding aligns with Tobler's first law of geography, which states that 'everything is related to everything else, but proximity has a stronger influence than distance' [2]. Thus, our primary focus in this study will be on examining maps, urbanization patterns, and various communes within the cities. This approach allows us to capture how the physical proximity and social connections shape the streaming platform preferences and, intriguingly, how similar neighborhoods can yield distinct outcomes.

In France the urban layout often takes on a spiral shape, spreading outwards instead of having a single city center. This results in the population being dispersed across larger regions of the urbanized centers and suburbs of France. The spread is influenced by strict building height limits and historical living habits that were shaped by rural immigration movements, particularly during World War II. This dispersal is even more pronounced in Southern France, where there are no specific city centers in these cities [3].

We have chosen Paris due to its status as the most populous and diverse city in France, and for comparison, we've selected Montpellier, a city in southern France [4]. However, our initial focus is on examining the population and development levels of these two cities. It is essential to clarify that we won't be working with the entire Paris arrondissements, which exceeded 10 million in population in

2019. Instead, we will specifically focus on the City of Paris, with a population of around two million residents distributed across 20 districts. Some of these regions houses approximately 500 thousand people, similar to the population of the City of Montpellier in the year of 2019. This similarity in population size allows us to make meaningful comparisons between the residents of the French Riviera and those of Paris.

## II. METHODOLOGY

The objective of this study is to analyze the usage of different video content platforms, classified as paid and unpaid, in selected cities and the communes of those cities. Since the dataset [5] belongs to 2019, we have neglected YouTube Premium as it was not popular among people at that time. Instead, we have selected YouTube and Dailymotion as representative choices for unpaid video content platforms, including watching videos, TV series, and movies. These platforms will be mentioned as Unpaid Platforms and will be denoted as  $U$  if necessary. To represent paid streaming services, we have included Netflix, Apple Video, and Orange TV, similar to former services, we will be mentioning them as Paid Platforms and will be denoted as  $P$ . However, there are some differences between these three platforms; Netflix and Orange TV are subscription-based, while Apple Video offers rental and purchase options for movies and TV series. In order to simplify the analysis, we have excluded browser streaming or TOR, as these platforms provide both paid and unpaid pirate services.

In this study, downlink data for the mentioned platforms was collected at 15-minute intervals for each day between April 1, 2019, and April 30, 2019. The total daily usage values of these platforms were used to aggregate the points in the cities where each platform had the highest data usage. This was done for two different cities.  $C_{Pxj}(t)$  and  $C_{Mxj}(t)$  represent the amount of internet data used on day  $j$  at time  $t$  for platform  $x$  in Paris and Montpellier, respectively. The total internet usage in the cities over a span of 30 days is denoted as  $C_{Px}$  and  $C_{Mx}$ , as shown Equation 1.

$$C_{Px} = \sum_{j=1}^{30} \sum_{t=1}^{96} C_{Pxj}(t) \quad C_{Mx} = \sum_{j=1}^{30} \sum_{t=1}^{96} C_{Mxj}(t) \quad (1)$$

The quantities of platform usage for YouTube, DailyMotion, Apple Video, Orange TV, and Netflix in city  $i$  is denoted as  $y_i$ ,  $d_i$ ,  $a_i$ ,  $o_i$ , and  $n_i$ , respectively. We represent the total internet usage on these platforms across all cities in our study as  $Y$ ,  $D$ ,  $A$ ,  $O$ , and  $N$ . We categorize them into Paid Platforms and Unpaid Platforms, as defined in Equation 2 and Equation 3. We will use the category total usages and individual platform usage for different tests and analysis.

$$u_i = y_i + d_i; p_i = a_i + o_i + n_i \quad (2)$$

$$U = Y + D; P = A + O + N \quad (3)$$

<sup>1</sup>The NetMob 2023 Data Challenge is a competition aiming at deriving and exploiting insights from massive mobile service consumption information. Details can be found at <https://netmob2023challenge.networks.imdea.org>

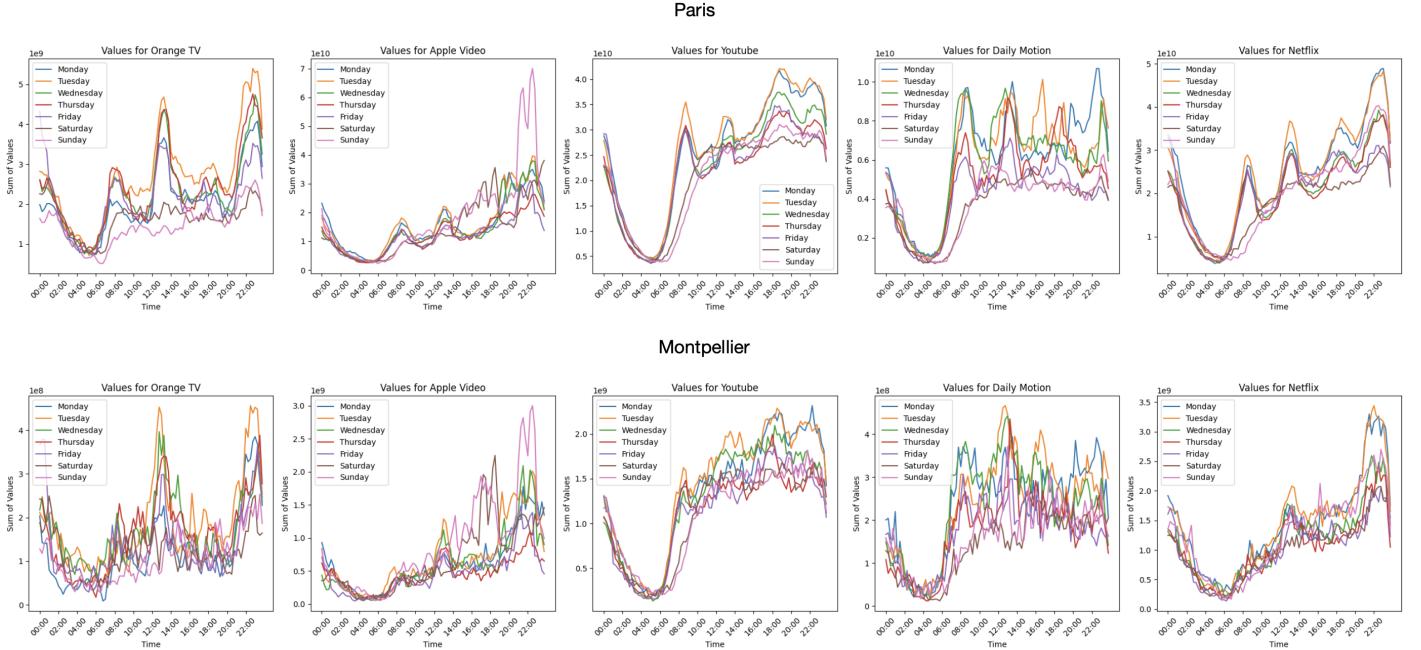


Fig. 1. Hour-Level Changes of Platform Usage for Paris and Montpellier.

We will begin by examining maps of these cities to visualize the dominant platform usage patterns and how these patterns change over the month of April, the time span we choose to analyze, for our exploratory data analysis. Following that, we will employ the statistical tool ANOVA [6] to gain insights into the differences between communes and cities, and subsequently, we will discuss the results of both analyses.

### III. EXPLORATORY DATA ANALYSIS

First and foremost, we are examining the times when people use these platforms throughout April. Aggregate daily usage of people living in cities of Paris and Montpellier is shown in Figure 1. In the figure, it is evident that people primarily use these platforms more during non-work hours such as lunch breaks and after work hours. This indicates that these platforms are popular leisure-time activities for people. On the other hand, it is observed that internet usage on these platforms is generally less on weekends than on weekdays. These findings align with the Acumen Daily Report on Youth Video Diet, which surveyed a sample of individuals aged 13-24 years. The results of the survey reveal that these individuals typically watch videos when bored or to fill their free time, with a preference for doing so after work or school and during lunchtime. Therefore, our findings underscore that this activity is indeed a leisure-time pursuit, with people choosing to engage in it primarily on weekdays after school or work hours.

It is important to note that our analysis represents only a small aspect of these individuals' lives. Through our analysis, we aim to gain insights into the demographics of areas within cities and the socio-economic conditions of neighborhoods, all based on a limited snapshot of people's leisure-time activities.

To investigate the similarity between the cities regarding platform usage, we employ the dissimilarity index in our study. The dissimilarity index was originally introduced by Duncan and Duncan in 1955 to assess ethnic and population disparities [7]. This index's values range between zero and one, where zero signifies complete integration, and

one indicates complete segregation between paid and unpaid platform usage.

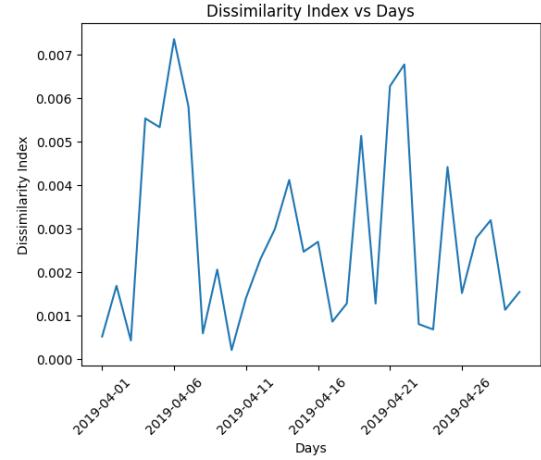


Fig. 2. Day-Level Change in Dissimilarity Index

The dissimilarity index is calculated to determine how much the two cities differ from each other in terms of daily platform usage, as shown in Equation 4. We conduct a thorough analysis by individually scrutinizing the fluctuations in the dissimilarity index values for each day. This examination allows us to pinpoint any potential outlier days within this dataset. Throughout the 30-day period presented in Figure 2, these values range between 0.001 and 0.007. These small values indicate that the cities are very similar in terms of their platform usage.

$$D = 1/2 * \sum_{i=1}^2 \left( \left| \frac{u_i}{U} - \frac{p_i}{P} \right| \right) \quad (4)$$

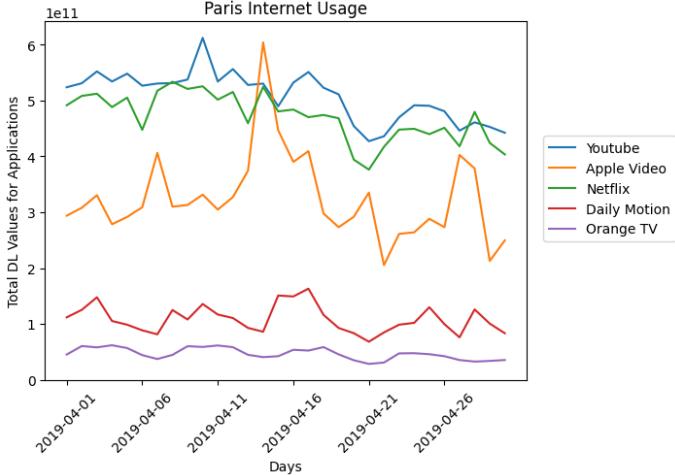


Fig. 3. Day-Level Change in Paris Internet Usage

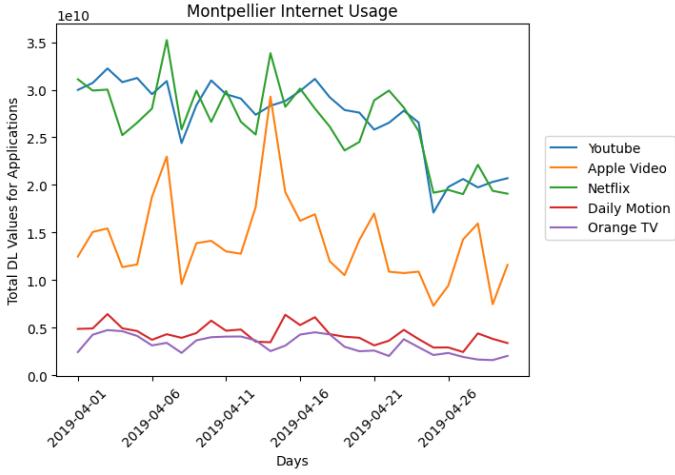


Fig. 4. Day-Level Change in Montpellier Internet Usage

To understand the proportion of how these platforms are used in relation to each other, we examine the usage of all platforms in the cities of Paris and Montpellier in Figures 3 and 4. Our goal is to discern whether certain days differ from others, providing insights into people's reactions to daily life and their preferences for streaming platforms. For example, we observe a significant increase in the usage of Apple Video on April 14, 2019. When we investigate the cause of this spike, we find that it coincided with the final season premiere of HBO's TV series Game of Thrones in Apple Video [8]. This finding demonstrates how the hype surrounding a TV series can lead to an extraordinary surge in platform usage and highlights the susceptibility of people's online video-watching habits to trends and hype. Following the premiere episode, we observe an increase in usage every Monday, but none of these subsequent increases match the magnitude of the first episode's spike. If our data extended to the series finale, we could also comment on people's behavior when it comes to watching both the first and last episodes. It's worth noting that we did not encounter another similarly dramatic instance like this on any of the other platforms.

As Figures 3 and 4 show, both cities primarily use YouTube and Netflix as their most popular platforms. Due to the dominance of these

platforms, it is unlikely to observe a clear distinction between paid and unpaid platforms, as previously mentioned in the dissimilarity index analysis. The same can be said for Apple Video, DailyMotion and Orange TV. Apple Video ranks in the middle in terms of usage in both cities; DailyMotion and Orange TV are the least used platforms. While we cannot discern a distinct difference in the usage trends of these platforms, the y-axis of Figures 4 and 3 reveals a dramatic contrast in the scales of platform usage. This discrepancy aligns with the population and size differences between the two cities.

In Figure 5, we present the usage distributions of various platforms for both Grand Paris and Montpellier. We initially examine Grand Paris to identify any areas within the city with higher internet usage and reconsider to exclude them from our analysis. Upon analysis, we find that the density of video-watching is notably high in the city of Paris. In contrast, for Montpellier, we consider the city as a whole rather than focusing solely on the city center, as our aim is to compare Paris and Montpellier, not just the central commune of Montpellier.

However, regardless of the city, we observe that the most dominant usage of all platforms tends to be concentrated in the city centers. In Paris, despite the greater prevalence of Apple Video and Netflix in the city center, Orange TV emerges as the most widespread platform across the entire Grand Paris. This aligns with our earlier findings explaining the lower usage of Orange TV in the city of Paris. In Montpellier, we don't observe a widespread distribution of any single platform across the city; instead, Apple Video and Orange TV are more prevalent in various areas compared to other platforms.

From a socio-economic perspective, we notice that in economically developed arrondissements of Paris, such as Champs-Élysées, all platforms experience more extensive usage. Conversely, in areas further from the city center, such as La Courneuve, where socio-economic challenges are more pronounced, platform usage is significantly lower. Figures 6 and 7 visually depict the spread of dominant applications across arrondissements in the City of Paris and Montpellier.

When we examine each region of City of Paris in Figure 6, we observe that only three platforms are the most used. YouTube and Netflix stand out as dominant platforms, while Apple Video is scattered specific parts of the central and western of the city. YouTube dominates the in the central and western parts of the City of Paris, while Netflix is the most used platform. The regions where Apple Video is most used are the northwest and central areas. DailyMotion or Orange TV, on the other hand, are not the most used platforms in any region of the city.

Similar to Paris, the number of most used platforms is also three in Montpellier, with YouTube and Netflix exhibiting a dominant spread in Figure 7. However, domination of Apple Video is observed in a much smaller number of regions compared to Paris. While YouTube is more heavily used in the center and south, Netflix is used much more in the northern part of Montpellier. Also, it is observed that the regions where Netflix is dominant are more in Montpellier compared to the City of Paris. DailyMotion and Orange TV, like in Paris, are not the most used platforms in any region of Montpellier.

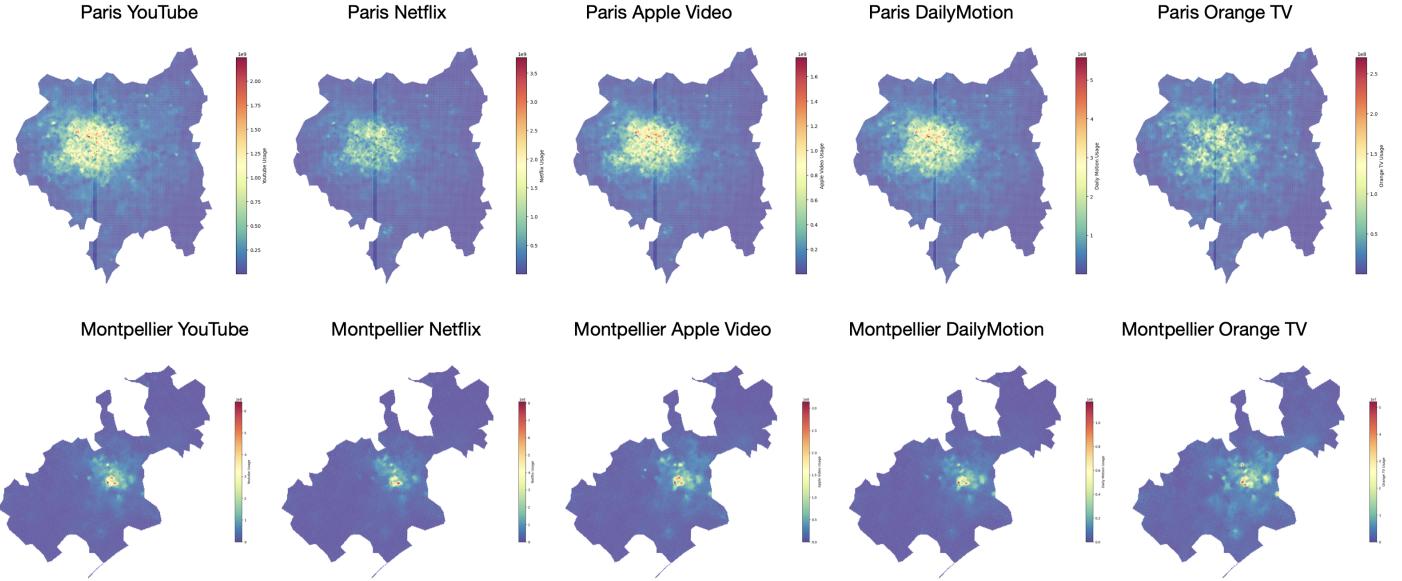


Fig. 5. Usage Spread of Platforms in Grand Paris and Montpellier

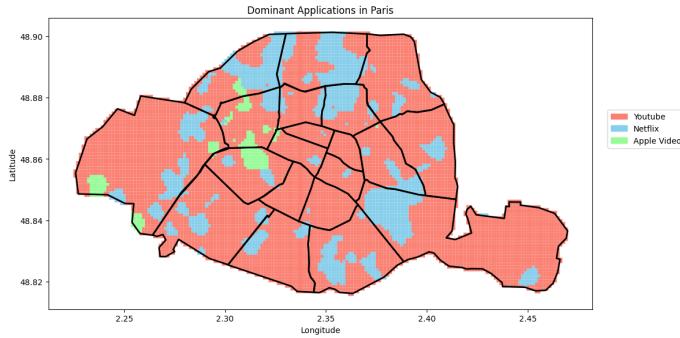


Fig. 6. Dominantly Used Platforms in Paris

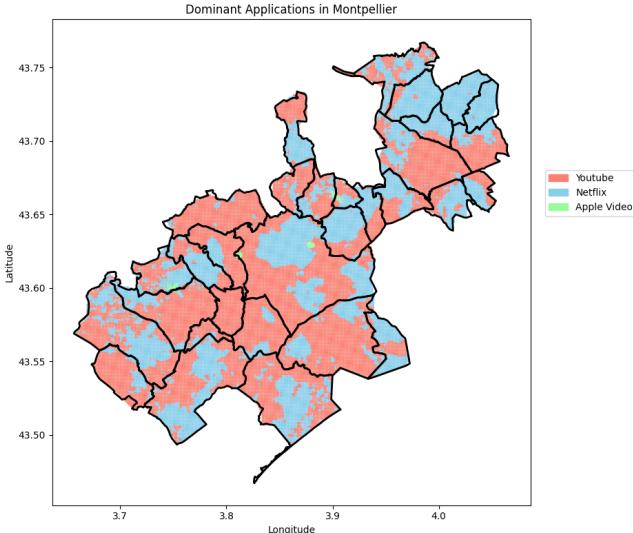


Fig. 7. Dominantly Used Platforms in Montpellier

#### IV. ANOVA ANALYSIS

As evident from the map in the preceding section, there is a significant disparity in population between the cities. Therefore, we'll be working with average data for each unit, which is normalized by its respective population. We plan to use ANOVA to detect any differences in the average usage data between the Paris commune with 20 arrondissements and Montpellier. This approach will help us gain insights into the demographics and socio-economic conditions of both similar and distinct regions within the cities.

In the context of an ANOVA (Analysis of Variance) analysis, equations 5, 6, and 7 pertain to examining whether the means of variables across more than two groups are statistically significantly different. These steps are formulated through equations. Within-group and between-group variances form the foundation of the analysis. In this context,  $SS_{within}$  represents the sum of squares within groups, while  $SS_{between}$  signifies the sum of squares between groups.  $n_i$  denotes the number of observations in the  $i$ th group,  $X_{ij}$  represents the value of the  $j$ th observation in the  $i$ th group,  $\bar{X}_i$  stands for the mean of group  $i$ , and  $\bar{X}$  represents the overall mean.

$$SS_{within} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \quad SS_{between} = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2 \quad (5)$$

$df_{between}$  symbolizes the degrees of freedom between groups, while  $df_{within}$  represents the degrees of freedom within groups.  $MS_{between}$  and  $MS_{within}$  represent the mean squares between groups and within groups, respectively.  $N$  represents the total number of observations in the study, and  $k$  represents the total number of groups.

$$MS_{between} = \frac{SS_{between}}{df_{between}} \quad MS_{within} = \frac{SS_{within}}{df_{within}} \quad (6)$$

$$df_{within} = N - k \quad df_{between} = k - 1 \quad F = \frac{MS_{between}}{MS_{within}} \quad (7)$$

In this study, an ANOVA analysis was conducted on 20 arrondissements of the city of Paris and the city of Montpellier, focusing on streaming platforms' usage categorized as paid or unpaid. A total of 21 different groups were analyzed, initially based on the regions' usage of paid applications. The analysis revealed a statistically significant difference among the groups. The F-value for the analysis was 357887.347, and the p-value was 0.000, indicating strong statistical significance.

As a statistically significant difference was observed among the groups, it became necessary to determine which specific groups exhibited this difference and whether it held statistical importance. To investigate this, a post-hoc test, namely Tukey's Honestly Significant Difference (HSD) test [9], was employed. This test allowed for the examination of differences between all pairs of groups. It is noteworthy that nearly all differences between groups were found to be statistically significant, as evidenced by p-values below 0.001. The outcome of the Tukey's HSD test illustrated the variations between groups and is depicted in the visual representation in Appendix.

According to Tukey's HSD test, there is a statistical difference in the use of paid streaming applications between Arrondissement 7, one of the regions where Netflix and Apple Video, both paid applications, are commonly used, and Arrondissement 12, where Netflix has a relatively larger share of usage in terms of location. Although Arrondissement 12 has a population approximately three times that of Arrondissement 7, their age distributions exhibit relatively similar percentage proportions. For instance, the population aged 15-29 in both regions is slightly less than 25%. In contrast, the percentage proportion of the population aged 30-44 is around 1/6 for Arrondissement 7 and slightly less than 25% for Arrondissement 12. Additionally, when comparing the level of education, Arrondissement 7 has a significantly higher proportion of individuals with university and postgraduate education compared to its population size when compared to Arrondissement 12.

Statistically, there is nearly no difference in the use of paid platforms between two neighboring regions, Arrondissement 19 and Arrondissement 20. Upon examining their population distributions, there are significant similarities both in terms of total population and age distribution. In these two regions, despite Arrondissement 19 having slightly more young population, Arrondissement 20 has a higher proportion of middle-aged individuals. The percentage of elderly population is approximately the same in both regions. Moreover, when considering higher education levels, Arrondissement 19 has a significantly higher percentage of individuals with undergraduate and postgraduate degrees, despite having only slightly more population [4]. Economically, Arrondissement 20 has a slightly higher number of employed individuals aged 15 and above, but proportionally, the regions are nearly identical. Taking age distribution, education, and economic conditions into account, it can be inferred that these two similar regions exhibit similar viewership tendencies regarding paid streaming platforms.

In order to assess potential disparities in the utilization of unpaid streaming platforms across various regions, a subsequent ANOVA analysis was carried out. The outcomes of this analysis indicated the presence of statistically significant differences among the regions concerning their usage of unpaid platforms ( $p\text{-value} = 0.000$ ,  $F\text{-value} = 265025.66$ ). To pinpoint which specific regions exhibit divergent usage patterns and ascertain the statistical significance of these distinctions, HSD test was subsequently employed. The inter-regional p-values and difference values are comprehensively documented in the Appendix for reference and further scrutiny.

When examining two regions, Arrondissement 14 and Arrondissement

17, which do not exhibit statistical differences in the usage of unpaid platforms, it is noteworthy that they display significantly different demographic and viewing patterns. Despite having a smaller population, Arrondissement 14 hosts a considerably higher number of undergraduate and postgraduate students. This is attributed to Arrondissement 14 being home to some of France's top institutions such as the Paris School of Economics and the University of Paris. In contrast, Arrondissement 17 has approximately 15,000 more employees in total. Although the usage of applications like YouTube and Daily Motion is quite similar in these two regions, there are radical differences in specific areas mentioned earlier.

One of the regions that statistically resembles Montpellier the most, with significant differences in the usage of both unpaid and paid applications, is Arrondissement 12. Both YouTube and Netflix are widely prevalent in both regions, and when examining these areas, it becomes evident that both regions have a young population with a percentage exceeding fifty percent. In Montpellier, the number of undergraduate and postgraduate students is approximately 1/7 of the population, whereas in Arrondissement 12, this proportion is approximately around 10%. When analyzing the number of individuals aged 15 and above who are employed, Montpellier shows a workforce participation rate of approximately 1/3 of the population, while in Arrondissement 12, this rate is approximately 45% [4].

## V. CONCLUSION

In this study, we exclusively concentrate on data related to people's usage of specific streaming platforms to uncover the livelier, younger, and wealthier areas within the cities of Paris and Montpellier. This research illuminates mobile data usage patterns in connection to people's leisure activities, with the goal of identifying unique areas of mobile data usage.

It becomes apparent that the choice of streaming platforms not only reflects the socio-economic conditions and demographics of neighborhoods but is also influenced by global trends and hype. Thus, the mid-level timespan of our study offers only limited insights into the broader aspects of these neighborhoods. To gain a deeper understanding, we could analyze year-level data and weekly reports from these streaming services, particularly focusing on the TV series they offer. Such an approach would provide us with insights into how these cities align with global trends. By examining their alignment with these trends, we can infer the age and education levels of these cities' populations.

Additionally, when we investigate these services individually, we discover that Orange TV streams football matches. Considering the presence of French football teams like PSG, Lille, and Lyon in the Champions League in 2019, and recognizing the importance of watching football as a leisure activity in less advantaged neighborhoods, we can draw connections between the availability of Orange TV and the socio-economic situations of the people in those areas. However, this would require smaller-scale, more focused studies to provide meaningful insights.

## REFERENCES

- [1] T. C. Schelling, "Models of segregation," *The American Economic Review*, vol. 59, no. 2, pp. 488–493, 1969.
- [2] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Economic Geography*, vol. 46, no. sup1, pp. 234–240, 1970.
- [3] T. Caplow, "Urban structure in france," *American Sociological Review*, vol. 17, no. 5, pp. 544–549, 1952.

- [4] T. N. I. of Statistics and E. Studies. Population en 2019. [Online]. Available: <https://www.insee.fr/fr/statistiques/6543200/>
- [5] O. E. Martinez-Durive, S. Mishra, C. Ziemlicki, S. Rubrichi, Z. Smoreda, and M. Fiore, “The netmob23 dataset: A high-resolution multi-region service-level mobile data traffic cartography,” *arXiv preprint arXiv:2305.06933*, 2023.
- [6] L. St, S. Wold *et al.*, “Analysis of variance (anova),” *Chemometrics and intelligent laboratory systems*, vol. 6, no. 4, pp. 259–272, 1989.
- [7] O. D. Duncan and B. Duncan, “A methodological analysis of segregation indexes,” *American Sociological Review*, vol. 20, no. 2, pp. 210–217, 1955.
- [8] H. Frishberg and L. Sarner, “Game of thrones season 8 premiere: How to stream hbo and what time,” *NYPost*. [Online]. Available: <https://nypost.com/2019/04/08/game-of-thrones-season-8-premiere-how-to-stream-hbo-and-what-time/>
- [9] J. W. Tukey, “Comparing individual means in the analysis of variance,” *Biometrics*, pp. 99–114, 1949.

# Appendices

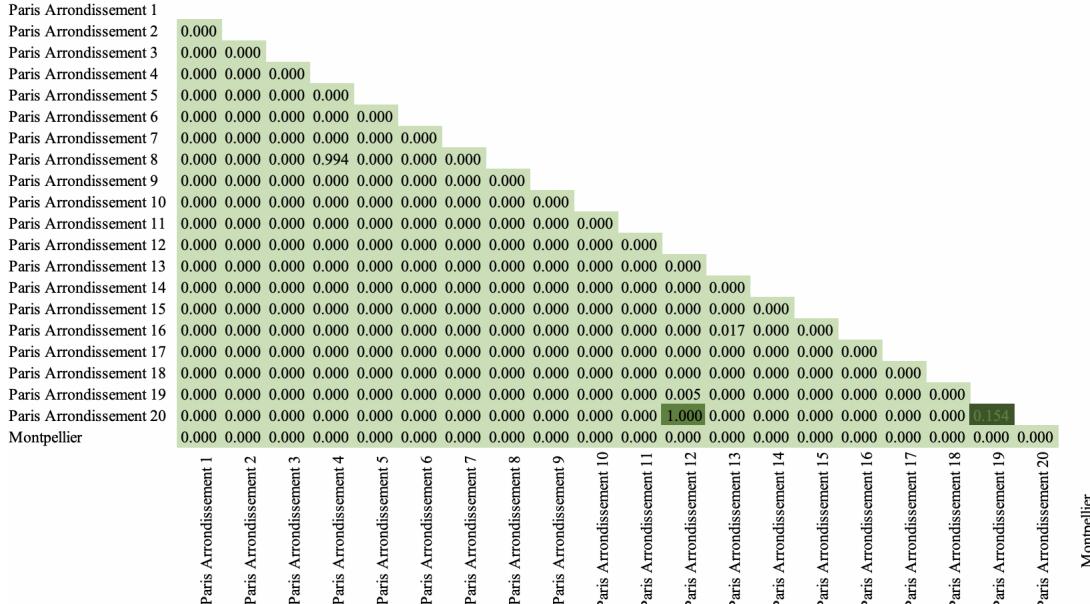


Fig. 8. Tukey's HSD Significance Differences for Paid Streaming Platforms

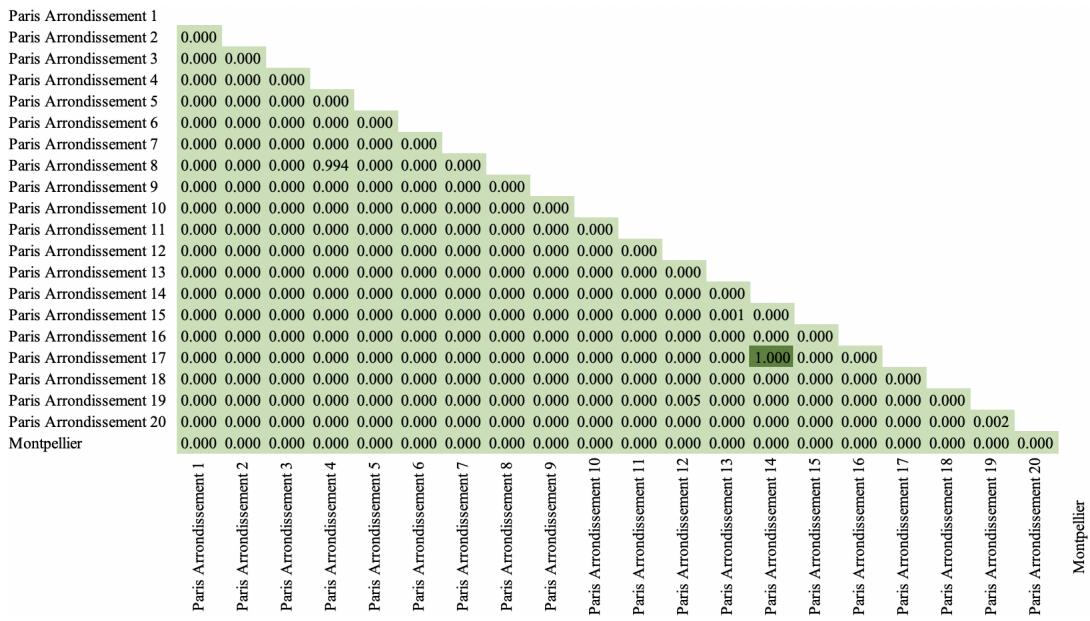


Fig. 9. Tukey's HSD Significance Differences for Unpaid Streaming Platforms

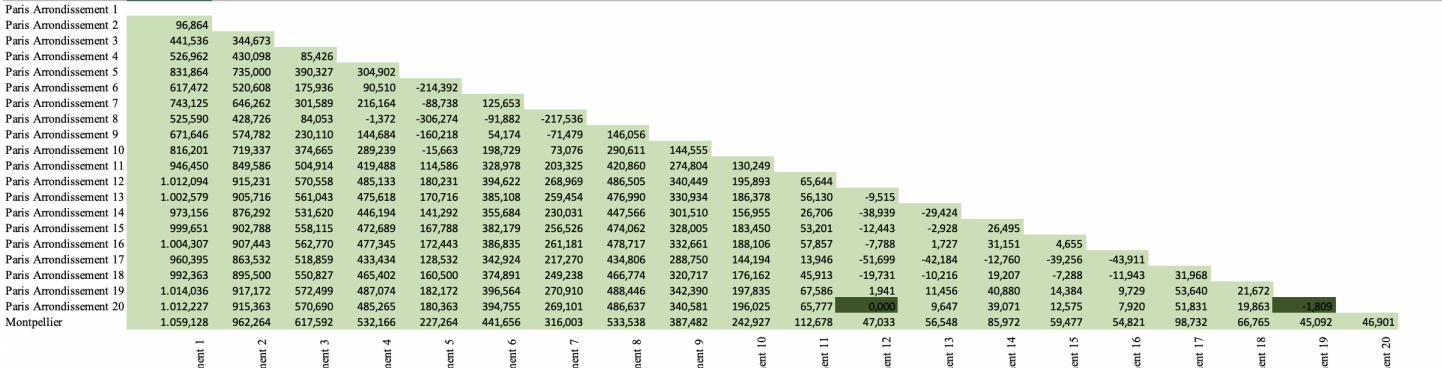


Fig. 10. Tukey's HSD Differences for Paid Streaming Platforms

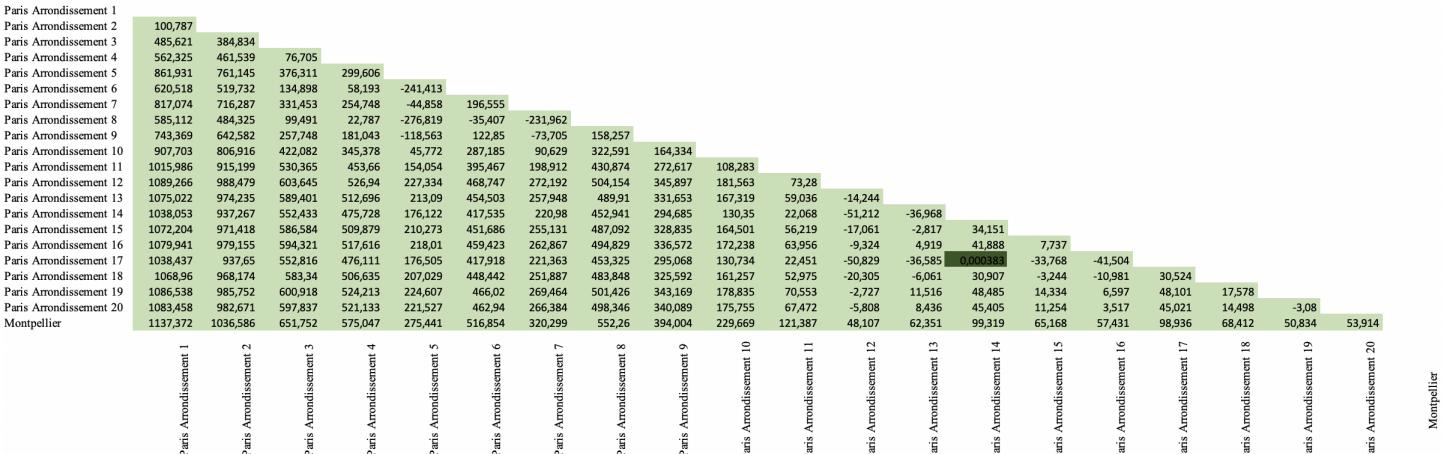


Fig. 11. Tukey's HSD Differences for Unpaid Streaming Platforms

Montpellier