Comparative Analysis of Urban vs. Suburban Usage

1st Khaled Waleed Farouk
dept. of Computer Engineering
Nile University
Elsheikh Zayed, Egypt
K.waleed@nu.edu.eg

2nd Yusif Khaled Rashed
dept. of Computer Engineering
Nile University
Elsheikh Zayed, Egypt
Y.Rashed@nu.edu.eg

3rd Eyad Yasser Badawi
dept. of Computer Engineering
Nile University
Elsheikh Zayed, Egypt
E.yasser@nu.edu.eg

4th Mostafa Assad
dept. of Computer Engineering
Nile University
Elsheikh Zayed, Egypt
M.assad@nu.edu.eg

***Abstract*—This paper explores mobile application usage patterns in urban and suburban areas of six French cities—Paris, Lyon, Marseille, Le Mans, Nancy, and Orléans—using data from the NetMob 2023 Challenge dataset. We aim to determine when and where individuals engage with popular entertainment and mobility apps (Instagram, Google Play Store, and YouTube) to identify optimal times and locations for marketing outreach. Data preprocessing involves replacing missing values with means and removing outlier days using the Interquartile Range (IQR) method. Heat maps are used to visualize usage by day of the week and to reveal average traffic trends. The elbow method is applied to determine the optimal number of clusters, and the K-Means algorithm is used to group geographic areas based on daily app activity. Our results provide useful insights for app developers and advertisers seeking to optimize pricing strategies and improve audience targeting.**

***Keywords*—Outlier Detection, Clustering, Classification, Data Traffic, NetMob, App Usage**

I. Introduction

Nowadays, people spend a significant amount of time on their phones using various mobile applications. This has shifted advertising strategies from traditional media to digital platforms, where knowing when and where users engage most is essential for effective ad placement. Additionally, app owners need insights to set appropriate pricing for hosting advertisements based on user activity. This paper analyzes usage patterns of Instagram, Google Play Store, and YouTube across urban and suburban areas in multiple French cities. By identifying peak and low usage periods and spatial differences, we aim to provide valuable information for advertisers to optimize their campaigns and for app owners to maximize monetization. The data will be cleaned by detecting and removing outliers, and clustering techniques will be applied to simplify and interpret the usage patterns across geographic regions.

II. Literature Review

The widespread adoption of smartphones and mobile applications has profoundly reshaped the digital ecosystem, influencing both user behavior and advertising strategies. Numerous studies have investigated mobile app usage patterns, emphasizing the importance of user segmentation, advertising impact, and clustering techniques. These insights are particularly relevant in understanding geographic usage variations between urban and suburban areas.

Previous research has highlighted essential preprocessing methods such as imputation for handling missing values and the Interquartile Range (IQR) method for outlier detection, both of which enhance data reliability. Clustering algorithms, particularly K-Means, are widely used for segmenting user behavior patterns, with the elbow method aiding in determining the optimal number of clusters to balance accuracy and efficiency.

The literature underscores the importance of identifying spatial and temporal usage trends to support targeted advertising and effective app monetization strategies. In this context, our study applies these methodologies to analyze Instagram, Google Play Store, and YouTube usage across six French cities, including both urban (Paris, Lyon, Marseille) and suburban (Le Mans, Nancy, Orléans) areas. We utilize mobile traffic data from the NetMob 2023 Challenge, sourced from Orange. According to Radics et al. (2023), this dataset is useful for estimating population dynamics at the grid-cell level and provides valuable insights into urban mobility and behavioral patterns.

III. Methodology

A. Dataset Description

This research aims to draw analysis and conclusions in the advertising field. The target is to manipulate data from 6 cities in France (Paris, Lyon, Marseille, Le Mans, Nancy, and Orléans) to reach a statistical and visual analysis that assists people investing in the advertising field. The data used is from NetMob's 2023 Challenge, which has been collected using open-source geospatial data and measurements from Orange, a major company in the mobile network field. Originally, it was collected utilizing coverage information from a commercial radio-frequency signal propagation tool. Its computation involves probabilities and applying Bayes' theorem. The study focuses on usage data from three popular applications—Instagram, Google Play Store, and YouTube—due to their widespread use and relevance to both urban and suburban users.

B. Application Scope and Temporal Resolution

For each selected city, this study analyzes mobile usage patterns for the applicationsInstagram, Google Play Store, and YouTube. The dataset provides network traffic statistics for each app in 15-minute intervals, spanning a continuous period of 77 days—from March 16, 2019, to May 31, 2019. This high temporal resolution enables the detection of detailed usage patterns and temporal trends across both urban and suburban regions.

C. Analysis Strategy

The core variable in this analysis is the network traffic over time. Each application—Uber, Netflix, and YouTube—is examined individually to observe traffic fluctuations within each city. Once the peak usage hours for each app in every city are identified, a comparative analysis between the cities is conducted to reveal spatial usage trends.

D. Data Cleaning and Missing Value Handling

Prior to analysis, the dataset underwent a preprocessing phase to identify and correct anomalies. Each day was scanned for missing or non-numeric traffic values that could disrupt computations. These anomalies were handled by replacing them with the mean traffic value for the corresponding day. Notably, such anomalies were only observed on March 31, 2019, during a four-hour period starting at 8:00 PM.

E. Outlier Detection via Interquartile Range (IQR)

To further refine the dataset, statistical outliers were detected and removed. For each hourly time slot, the daily average traffic was calculated. Outliers—data points significantly deviating from typical traffic behavior—were excluded using the Interquartile Range (IQR) method. The IQR is defined as the range between the first quartile (Q1, the 25th percentile) and the third quartile (Q3, the 75th percentile). This technique isolates the middle 50% of the data, ensuring that the analysis remains robust and representative.

$$IQR = Q3 - Q1$$

$$LowerBound: Q1 - 1.5 \times IQR$$

$$UpperBound: Q3 + 1.5 \times IQR$$

F. Data Structuring and Aggregation

The data was manipulated using data frames and the NumPy library in Python methods. After removing all values that do not belong in the Interquartile Range and fixing the miswritten data, a new data frame was created. All the remaining days were grouped together, each according to their respective day of the week (e.g., all Mondays together, all Tuesdays together, etc.). Consequently, all days were reduced into a list of seven groups representing the average network traffic of their respective day of the week.

G. Heat Map Visualization

After the data cleaning and reduction part, a heat map was created to visualize the average traffic for each day of the week. This assisted in identifying patterns and variations across different days, providing a clear and intuitive understanding of traffic distribution. Furthermore, the average traffic of all days was calculated to represent a single average traffic of each city.

H. Clustering with K-Means

The K-Means clustering technique was used to group geographic areas based on their daily average traffic patterns. This involves calculating the daily average traffic for each area and then applying the K-Means algorithm to partition the areas into clusters with similar traffic patterns.

$$\text{K-Means Clustering:} \quad d(x_i, c_j) = \sqrt{\textstyle\sum_{m=1}^{M} (x_{im} - c_{jm})^2}$$

$$c_j = \tfrac{1}{|C_j|} \textstyle\sum x_i \in C_j \left\| x_i - c_j \right\|^2$$

I. Determining Optimal Clusters with Elbow Methods

To determine the optimal number of clusters (K) for the K-Means algorithm, the elbow method was used. It plots the sum of squared distances (inertia) from each point to its assigned cluster center for different values of K. The optimal K is the point where the plot shows a sharp bend (elbow), indicating a little decrease in clustering performance with increasing K.

$$Inertia = \sum_{j=1}^{K} \sum_{x_i \in C_j} \left\| x_i - c_j \right\|^2$$

Each of the previous methods was done for all six cities applications and the results and findings will be compared and discussed in the analysis section.

IV. Analysis

Our methodology was applied to a set of urban and suburban cities in France to ensure broad representation and consistent findings rather than isolated patterns. We selected three urban cities—Paris, Lyon, and Marseille—and three suburban cities—Orléans, Le Mans, and Nancy. The same analytical process was applied to each city to maintain consistency. For each application (YouTube, Instagram, and Google Play Store), we conducted spatial geographic analysis, heatmap visualizations, and statistical aggregation to capture temporal and spatial usage patterns. By comparing these results across urban and suburban environments, we aim to uncover differences in app usage behavior that reflect contextual, infrastructural, or demographic factors.

A. *Google Play Store*

TABLE I
MEAN VALUES ACROSS TIME INTERVALS FOR URBAN CITIES — GOOGLE_PLAY_STORE

Day	Paris	Marseille	Lyon
Friday_0	136410.421	24826.485	22881.080
Monday_0	89987.708	21926.804	21731.013
Saturday_0	100160.077	23059.756	21484.723
Sunday_0	72684.655	18288.533	17484.725
Thursday_0	135938.231	25081.902	30753.259
Tuesday_0	140330.098	31171.485	32478.313
Wednesday_0	139663.214	28742.419	29926.523

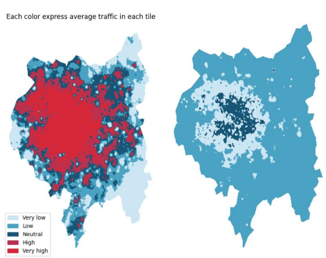

Fig. 1. for Paris

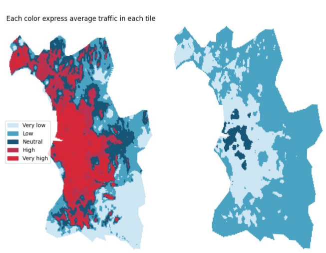

Fig. 2. for Marseille

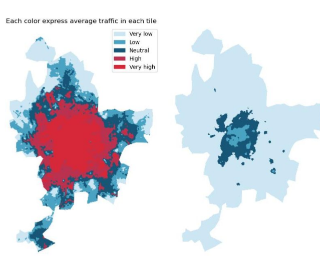

Fig. 3. for Lyon

TABLE II
MEAN VALUES ACROSS TIME INTERVALS FOR SUBURBAN CITIES — GOOGLE_PLAY_STORE

Day	Orléans	Mans	Nancy
Friday_0	9852.680	9451.388	19598.622
Monday_0	8230.936	7221.135	16631.814
Saturday_0	8156.980	6143.428	15289.268
Sunday_0	6315.537	5652.067	12792.451
Thursday_0	8967.142	8277.569	18466.744
Tuesday_0	9018.298	10843.942	21261.531
Wednesday_0	8606.489	5582.939	21102.296

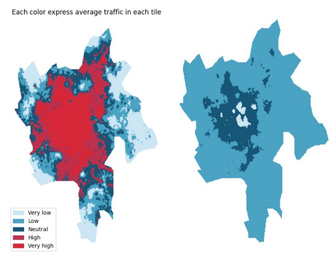

Fig. 4. for Orleans

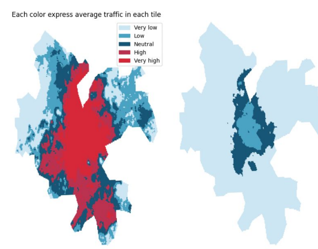

Fig. 5. for Mans

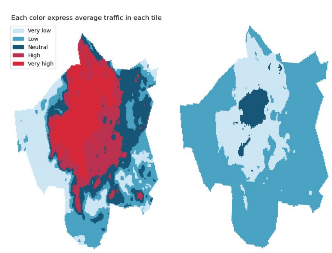

Fig. 6. for Nancy

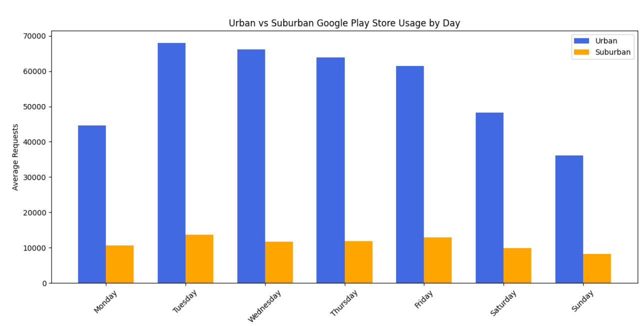

Fig. 7. Comparison between Urban and Suburban of Google play store

B. Youtube

TABLE III
MEAN VALUES ACROSS TIME INTERVALS FOR URBAN CITIES — YOUTUBE

Day	Paris	Marseille	Lyon
Friday_0	283884.439	50322.234	48776.401
Monday_0	252381.051	56906.054	51562.415
Saturday_0	232814.855	53107.261	48927.743
Sunday_0	228180.085	54429.477	50970.140
Thursday_0	258845.127	56779.550	49953.626
Tuesday_0	259790.826	52319.972	52033.734
Wednesday_0	261225.879	57413.272	52726.195

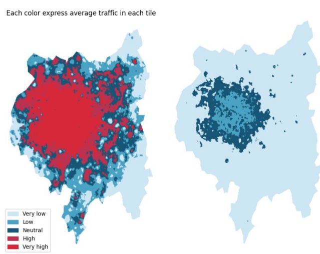

Fig. 8. for Paris

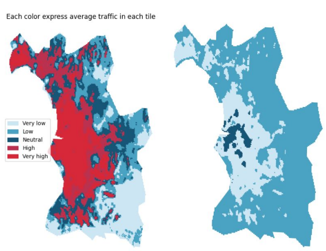

Fig. 9. for Marseille

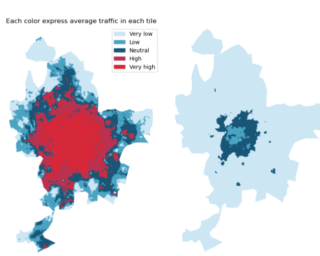

Fig. 10. for Lyon

TABLE IV
MEAN VALUES ACROSS TIME INTERVALS FOR SUBURBAN CITIES — YOUTUBE

Day	Orléans	Mans	Nancy
Friday_0	11078.350	9439.391	29216.190
Monday_0	12145.717	9551.351	31626.330
Saturday_0	11020.842	10122.704	25524.529
Sunday_0	11328.750	9360.241	25065.458
Thursday_0	12825.986	10139.683	29157.650
Tuesday_0	12129.916	10387.288	30949.880
Wednesday_0	12420.770	9580.710	33055.771

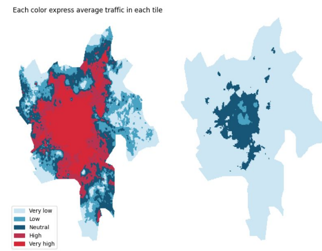

Fig. 11. for Orleans

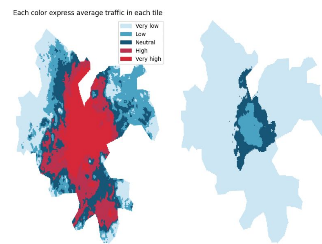

Fig. 12. for Mans

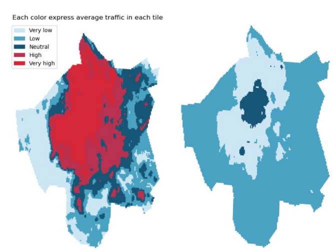

Fig. 13. for Nancy

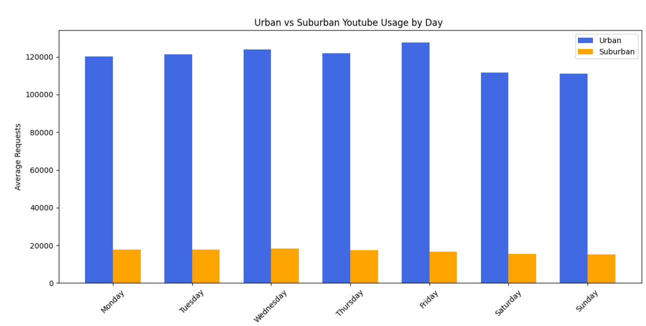

Fig. 14. Comparison between Urban and Suburban of Youtube

C. Instagram

TABLE V
MEAN VALUES ACROSS TIME INTERVALS FOR URBAN CITIES — INSTAGRAM

Day	Paris	Marseille	Lyon
Friday_0	739537.501	161015.969	152309.082
Monday_0	704837.978	164304.184	150657.942
Saturday_0	686063.823	157641.344	136982.085
Sunday_0	677062.521	157438.872	137709.852
Thursday_0	727812.443	159814.105	153111.716
Tuesday_0	710533.725	160916.352	150358.724
Wednesday_0	708628.987	160939.064	148928.574

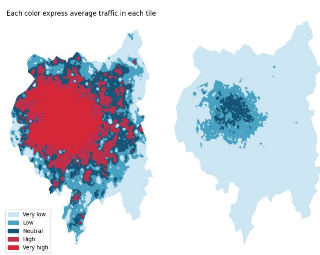

Fig. 15. for Paris

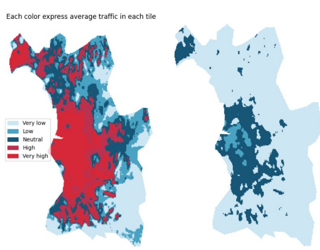

Fig. 16. for Marseille

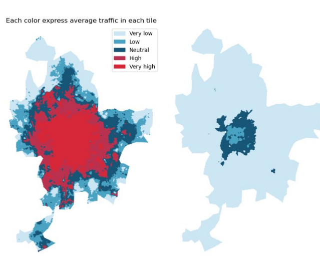

Fig. 17. for Lyon

TABLE VI
MEAN VALUES ACROSS TIME INTERVALS FOR SUBURBAN CITIES — INSTAGRAM

Day	Orléans	Mans	Nancy
Friday_0	152384.179	136913.843	84339.691
Monday_0	152304.863	135921.441	96790.319
Saturday_0	143280.531	124178.883	82238.230
Sunday_0	141931.177	121370.919	75804.345
Thursday_0	152322.408	135110.580	97096.258
Tuesday_0	152108.337	135718.867	95284.921
Wednesday_0	150123.342	135894.724	99162.106

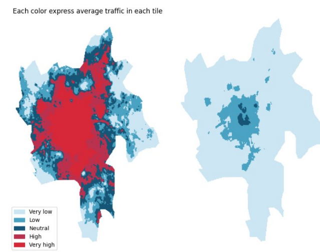

Fig. 18. for Orleans

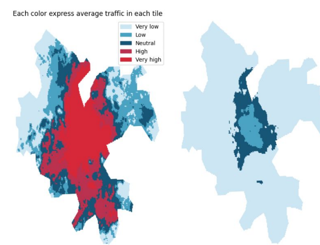

Fig. 19. for Mans

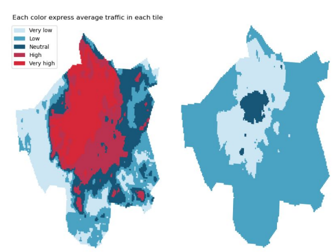

Fig. 20. for Nancy

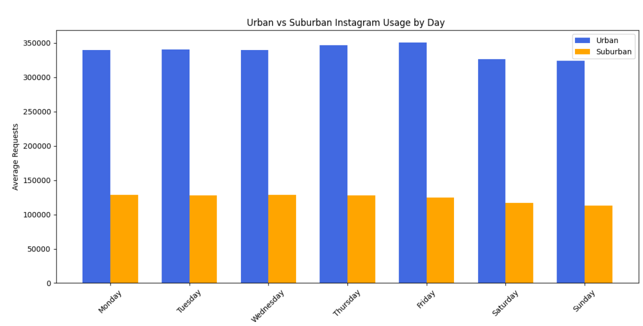

Fig. 21. Comparison between Urban and Suburban of Instagram

V. Findings

A. Outliers

Before cleaning, the dataset showed variation in quality and completeness across different cities. Notably, Lyon had a portion of missing entries, which could be attributed to collection gaps or system limitations. This missingness positions Lyon as potentially less reliable for consistent analysis. Paris, on the other hand, exhibited clean and stable data, while Orléans (suburban) also showed minimal anomalies—suggesting more consistent usage or data collection methods.

B. Classification And Clustering

From the clustering graphs, it can be observed that urban cities like Paris and Lyon showed a wider spatial spread in usage, reflecting the high density and diverse population distribution. In contrast, suburban cities like Le Mans and Orléans had tighter clusters of activity. This could be an advantage for advertisers—targeting in suburban areas requires covering a smaller region to reach a similar percentage of the population, leading to potentially lower campaign costs.

Moreover, the clustering patterns appear to mirror the population density of each city. Paris, with its vast and dense urban fabric, had clusters spread across many zones, while Le Mans showed more concentrated and localized usage, consistent with its suburban layout. This insight reinforces the importance of tailoring marketing strategies not just by city size, but by user concentration zones within each area.

C. Heat Maps

The heat maps show noticeable differences in user behavior between urban and suburban areas. In urban cities such as Paris and Lyon, activity peaks around 1 PM, likely corresponding to lunch breaks. At the same time, there is a clear drop in usage over the weekend, suggesting that people in urban areas may prefer to spend their weekends outdoors or away from their devices.

In contrast, in suburban cities like Le Mans and Orléans, the activity is more evenly distributed throughout the day, without strong spikes. This implies that time-targeted advertising may be more effective in urban areas, while in suburban areas, it may be better to maintain consistent exposure across the day rather than focusing on specific time slots.

Furthermore, analyzing usage patterns across different apps in the same city provides deeper insights. For instance, YouTube traffic tends to spike later in the day, which makes sense given the longer-form content on the platform. On the other hand, Twitter sees activity increase around lunchtime, likely due to the ease of consuming short posts in a limited break period. Instagram, meanwhile, shows multiple spikes throughout the day, reflecting more frequent but shorter interactions.

These differences highlight the importance of tailoring advertising strategies not only to the geographic context but also to the specific usage behavior of each platform.

VI. Summary And Conclusions

This analysis explored differences in app usage between urban and suburban areas in France, focusing on user behavior across time and platforms. Urban cities such as Paris and Lyon demonstrated clear activity peaks during midday hours, while suburban cities like Le Mans and Orléans showed flatter usage patterns. These findings suggest that time-targeted advertising would be more effective in urban areas, whereas consistent all-day strategies may perform better in suburban regions.

The clustering analysis further revealed that urban areas tend to have more concentrated traffic zones, which can enable more efficient advertising with less geographical spread. Suburban areas, in contrast, displayed a wider distribution of activity, indicating the need for broader campaign reach to achieve similar exposure.

Heat map comparisons and platform-specific traffic patterns emphasized the importance of aligning marketing strategies with both the temporal habits of users and the nature of each application. For example, Twitter is best utilized during short daily breaks, while YouTube is more suitable for evening engagement. These distinctions reinforce the need for a tailored, data-driven approach when planning digital marketing strategies in diverse geographic settings.

Ultimately, understanding the spatial and temporal dynamics of user activity can significantly improve the precision and effectiveness of mobile service deployment and digital outreach efforts.

Recommendations

We recommend scaling this analysis to a broader dataset covering more cities and a longer time span. The current study was limited by time and computational resources, which constrained the depth of insights. Expanding the scope will allow for more robust pattern recognition, uncovering regional and temporal nuances in app usage. This deeper understanding can directly inform the development of more precise, efficient, and profitable advertising strategies tailored to both urban and suburban audiences.

References

[1] P. Christidis, M. Vega-Gonzalo, and M. Radics, "Modelling daily mobility using mobile data traffic at fine spatiotemporal scale," *Transportation Research Part C: Emerging Technologies*, vol. 145, 2023.