Business Intelligence Final Project

Unveiling the Factors Impacting Network Quality of Mobile Phones: Based on the investigation of ARCEP

Yi-ting KUO

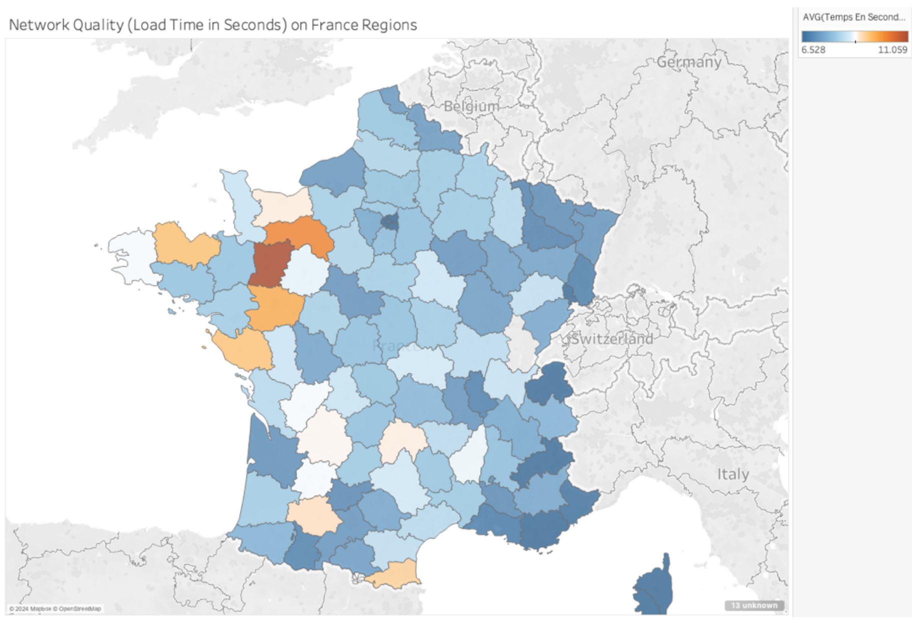

Figure 1. Map showing the level of network quality throughout the French departments (Orange: bad (slow load time); Blue: good (fast load time))

1. Introduction

1.1. Problem definition

The performance of mobile connectivity has always been one of the core concerns of the choice of telecom operators for customers. Regardless of marketing strategies and promotions campaigns, the bad quality of network would certainly churn customers away. However, the reasons behind the latency and weak signals of internet connection are somehow complicated and intertwined, which caused telecom engineers a lot of headaches, trying to capture the big picture of quality of service within different areas and discover the potential factors that might affect the efficiency of network.

Given this challenge, in order to respond to the CTE (Chief Telecommunications Engineer) of Network Operations Center in a mega French Telecom group – Bouygues Telecom, I as a

Business Intelligece Engineer, shall engage in the data and keen to find out the reasons on my side. Specifically, CTE has been aware of the increasing complains about the quality of network that it doesn't perform as expected in some certain area in the city, whether it is Toulouse or Paris, allegedly in the underground metro, some private facilities and residences, a few communities or even arrondissments. The loading time of webpage and data transmission (text, image, video etc) time show certain ineqality in different areas. Even with repairing mechaism, engineers on field are having hard time identifying the problems of poor quality.

Henceforth, This project addresses the issue of poor network quality in specific areas of France Hexagon and Grand Paris. The key problems are set as followed:

1. Identifying areas with suboptimal network performance.
2. Understanding the attributes and functions of these areas to determine the causes of poor network quality (e.g., usage of land, population density, building structures, ongoing urban construction).
3. Analyzing the time patterns to identify when the network quality issues occur.

The main data source of telecom quality investigation was conducted by Arcep, a French independent administrative authority who is responsible for regulating electronic communications. With partnership with French government (data.gouv.fr), they were able to assemble a set of measurements carried out in the field by agents mandated by Arcep, in order to assess the quality of networks of different mobile operators at a given time and location. These data focused on two branches: transport axes and residential areas. The former were conducted in mainland France and latter even includes overseas territories (Guyana, Martinique…).

In practice, there are a variety of methodology to define network performance and quality of service. Here I use "RSRP" and "elapsed time to load page" as two main indicators of network quality. RSRP (Reference Signals Received Power) is the key measure of signal level and quality for modern LTE networks. In cellular networks, when a mobile device moves from cell to cell and performs cell selection/reselection and handover, it has to measure the signal strength/quality of the neighbor cells. RSRP normally falls between -80dBm to -100dBm. Beyond -80dBm means excellent signal strength and below -100dBm is nearly without signal.

Other data sources like France geographical hierarchical encoding (Region, Department, Commune and Area) and datetime data are required to proceed the geographical and temporal analysis on network quality. Also, I include other dimensions like population distribution, transportation, building information, urbanization construction planning, in order to have comprehensive perspectives.

1.2. Justification for the Choice of Tools

Visualization: Tableau
Ideal for visualizing spatial and temporal data on different level. It can create interactive maps and time-series graphs, as well as Dashboards and Story telling to include multiple graphs and

statistics, which are crucial for deeper understanding of network performance across different regions and times.

Data preparation: Python
Being more flexible and developper-wise than Tableau Prep, Python is useful for data cleaning and preparation using the Pandas and Numpy library. It can handle large datasets and perform complex cleaning, type checking, NA checking and filling, condition filtering, transformations. Especially with powerful pd.apply() method that could apply any customized function for data transformation. Tables and columns are easily decomposed and recombined with others. These strengths would ensure the data is ready for analysis in Tableau.

2. Data modeling and preparation

2.1 Data preprocessing

1. Geographical Encoding
   Dropped unnecessary columns like district and neighborhood level information, changed postal code to string type and finally split into 3 hierarchical data frames (region, department and commune).
2. INSEE Population
   Fill None value with 0 and reduce the size on axis of time by only selecting the most recent year (2020) since the table originally contained the data since 1968.
3. Urban Planning
   Filtered the data with construct status equals to "Under construction" and create latitude and longitude columns out of coordinate column.
4. Building
   Dopped unnecessary columns like legal permit related fields. Cleaned geo coordinates data to usable latitude and longitude. Created "year" column from datetime. Transformed postal code since it was encoded differently than usual (75000001 instead of 75001).
5. QoS residential & QoS transport
   Each year has one csv file since 2016. Yet the size of one year data already reaches around 100 MB and we only care about the most present situation of networks, so I only chose 2023's dataset as year of interest. Then I transformed time to datetime format, processed numerical data from string type. Created commune code column out of address column.

All tables are exported as individual csv file.

2.2 Data Schema

I use Galaxy Schema as my data modelling. Fact tables are QoS residential and QoS transports, whose commune code connects to Commune table. Commune table then direct to Department table by Department code, which then goes to Region table by Region code, forming a hierarchy. Building, Urbanization and Population also possess commune code as foreign key so that they could connect to geographical table.

The reason why QoS residential and QoS transports should be treated separately instead of combined in one table is that they have different feature structure. For example, in QoS transports there is no commune code and address since the data is collected on metro or train. The place category in both tables also differs. For QoS residential it classifies zones (dense, rural, touristic, intermediary), while in QoS transport it is transport types (metro, RER, TGV, highway).

Also available on: https://dbdiagram.io/d/667818c75a764b3c722ea7e0

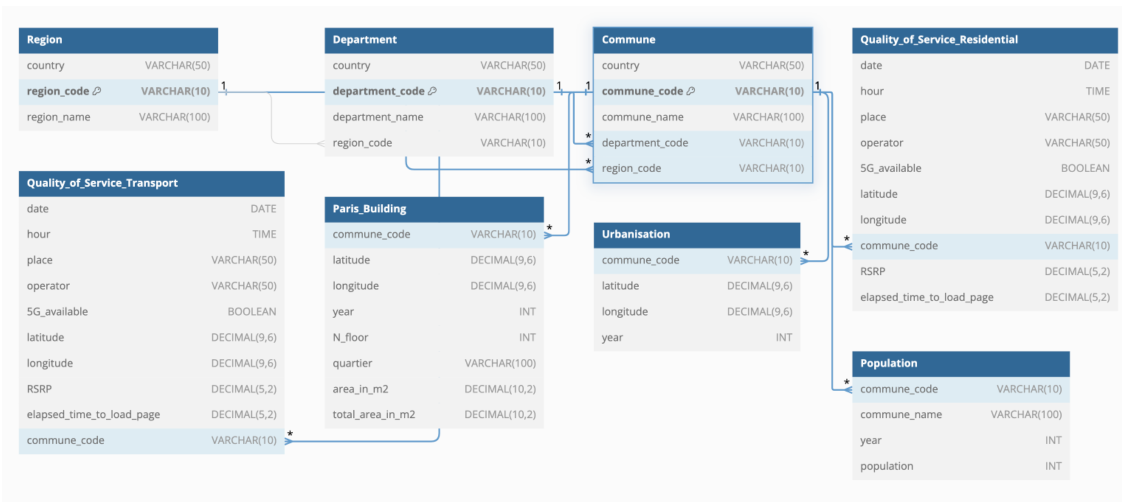

Figure 2. Galaxy Schema of this project

3.Application

3.1 Geographical and Demographical Analysis on Network Quality

From the map we could see network quality in region and department level. North-western regions like Pays de la loire and Normandie are suffered from latency and poor quality, and the department Mayenne and Orne. Another phenomena is that the rural area is not necessarily weak but the suburban area around the big city is often bad . Also, over here we could see that the mountain area perform in average better than most of the rural area.

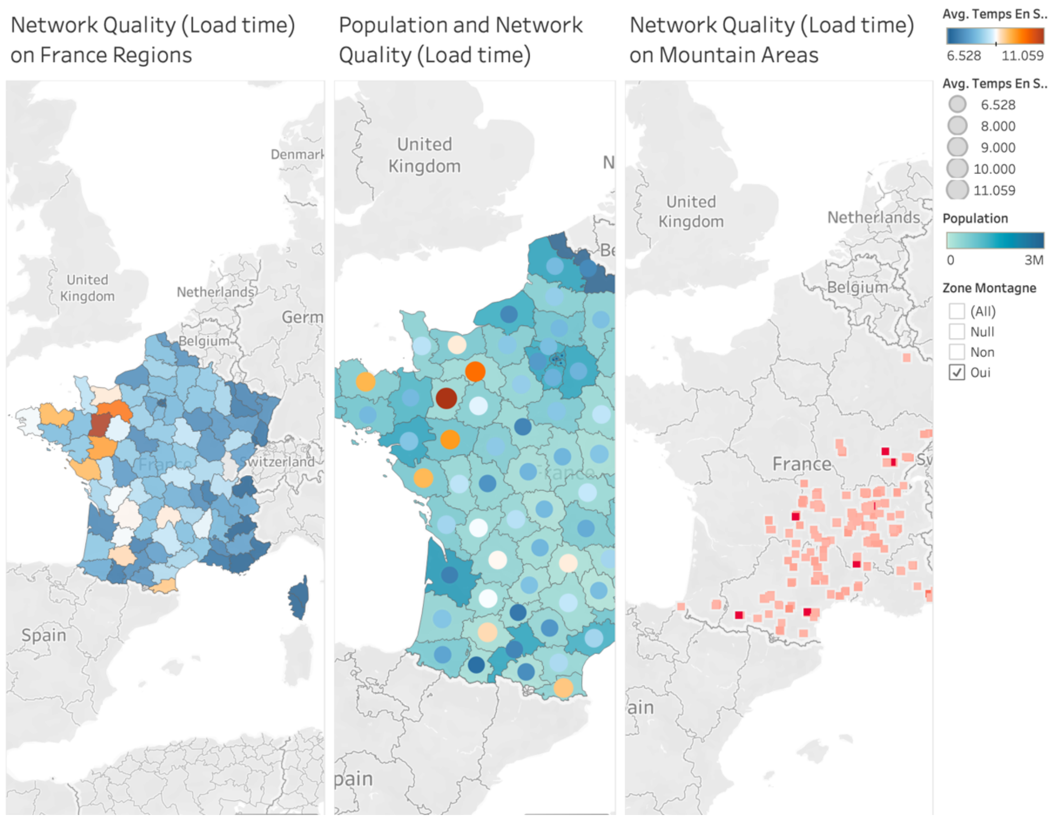

Figure 3. Geographic distribution of network performance

3.2 Competitors Analysis: Bouygues, Orange, SFR and Free

From coverage map of 4 operators, we could easily spot some dark blue areas which indicates the bad quality (long elapsed time), especially in north-western departments. Overall, the quality of Orange is more equally-distributed in territory than the others, while SFR has the poorest quality in north-western areas and the most unequal. The main rival of Bouygues, Orange overall performs better. However, Bouygues's network quality outperforms Orange in some area like highways (Routes) and Transilien trains. However, Bouygues's networks are inferior than the others in TGV systems.

The recommendations for Bouygues would be maintaining current performance in highways and trains, while attempt to fix the weak signal issues in TGV. On the other hand, the performance in metros and Intercités are sufficiently stable thus despite being behind Orange, there is no need to emphasize on these systems.

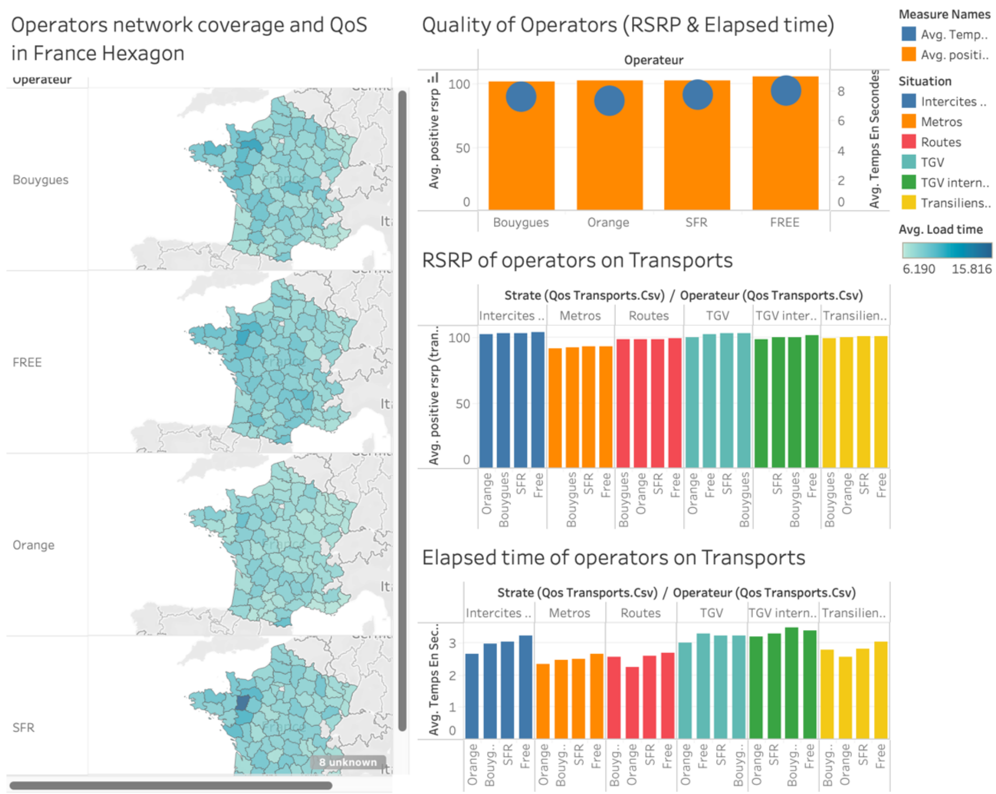

Figure 4. Competitors comparison

3.3 Temporal Analysis

Another approach for quality assessment is to observe the change within a day and a year, denoted by both RSRP and elapsed time. For hourly data, elapsed time goes down after 6pm in every zone but RSRP is rather stable throughout the day. In TGV international, it suddenly takes more time to load a webpage during 1pm to 3pm. For quality throughout the year, we could see that the quality becomes better in populated urban area (blue line) during mid-July, while rural areas (red line) become worse during mid-May.

The recommendation to Bouygues CTE would be implementing “dynamic supply” of network service according to demand shift. Since the quality in urban zones becomes stronger during summer, implying that most of the people left the town potentially for the vacation. Under this situation, Bouygues could then reallocate networks resources to other places that is in need for more network service. Additionally, future research on this topic could collaborate with SNCF real time data of passenger flow, to see if this periodic change in quality is correlated with trains organization and real-time volume in stations.

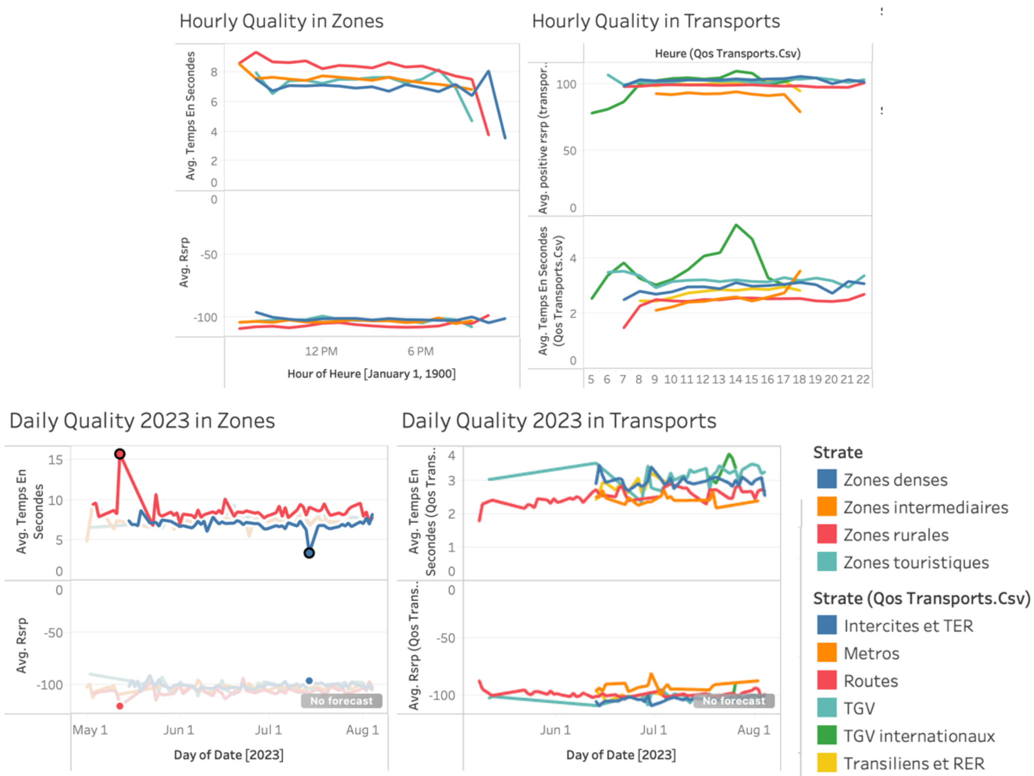

Figure 5. Temporal pattern of network

3.4 Architectural Influence on Network Quality in Paris

In order to identify the relationship between internet connection and building, I map two variables together, where map color (yellow-green) indicates the number of floors of building and purple circles indicate RSRP (the darker the worse). Same technique for floor area.

It is observable that the average height of building (denoted by number of floors) is higher in 17th and 18th arrondissements, with RSRP being averagely good. The height in 5th and 9th, however, with a merely average height in Paris, have the worst quality among all districts. This shows that the height of building is not a perfect explanatory of network quality. Instead, since the building age in 17th and 18th is generally younger or renovated, it might provide better infrastructure and environment for internet signal, and vice versa for 5th and 9th.

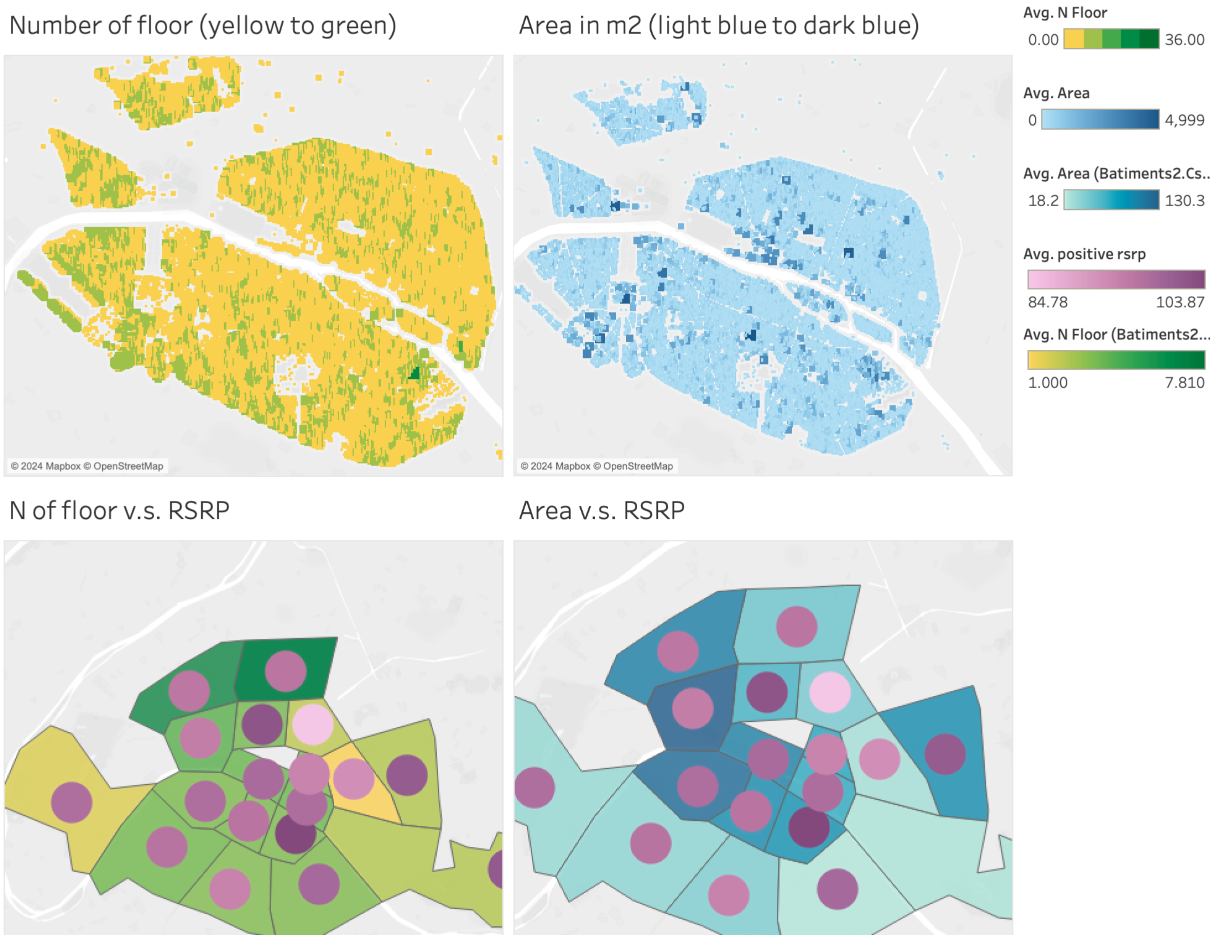

Figure 6. Construction features and network performance in Parisien Area

3.4 Network Quality on Public Transports of Paris

RSRP below -100 indicates nearly no signal, thus classified as “Slow Conncetion” (marked in red), with other value being normally operational (in blue). We could observe that there is way more spots with weak connection in metro and TGV. For metro, the slow internet occurs mostly in the mega junctions of metro lines, especially in Gare Saint-Lazare and Gare de Lyon. For TGV, they have the most serious network delay when it approaches the terminal station (like Gare de l’Est, Gare du Nord and Gare Montparnasse).

It would be then suggested to Increase the base stations of networks within those stations concerned. For TGV, potentially due to the “Faraday cage” effect, which is the shielding effect played by the TGV passenger compartment against electromagnetic waves, explained by Director of Network Technique in Orange.

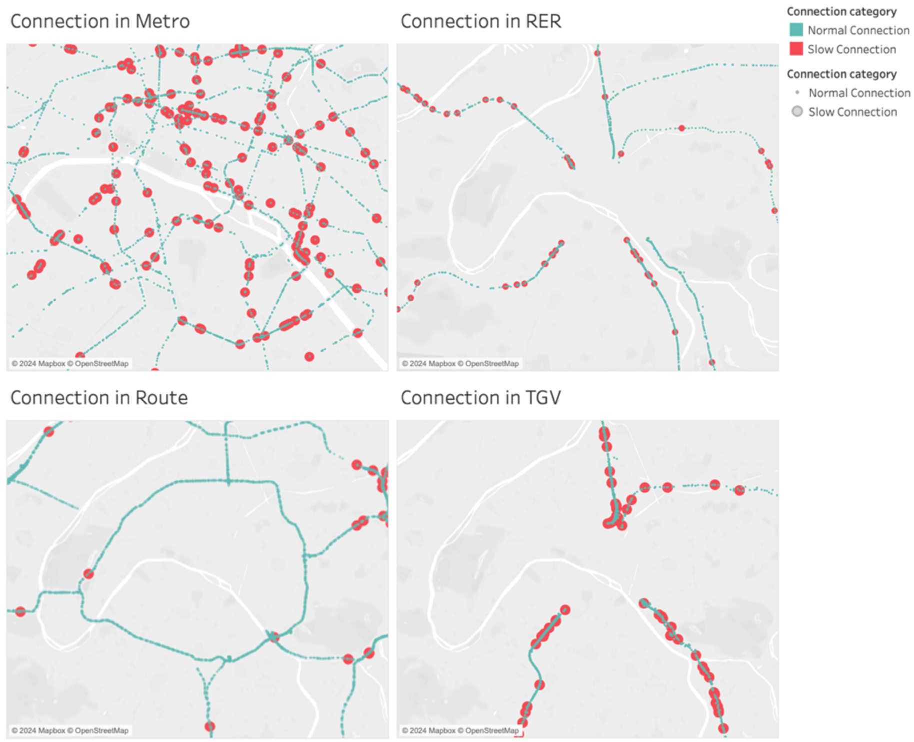

Figure 7. Public transports and network performances in Parisien Area

4.Conclusion

Python's robust data manipulation capabilities allowed us to clean and prepare large datasets efficiently. This step was critical as it ensured that the data was accurate and ready for further analysis. Once prepared, the data was imported into Tableau, where its powerful visualization features came into play. Tableau enabled us to create detailed interactive maps and time-series graphs, which were essential in pinpointing specific areas with network performance issues and understanding the temporal patterns of these issues.

Reflecting on the tools and data used, it is clear that my choices were well-founded. Python's flexibility and Tableau's intuitive visualizations complemented each other perfectly, allowing us to derive meaningful insights from complex datasets. Moreover, the data we used, sourced from ARCEP, along with additional geographical and demographic information, was appropriate for our analysis.

From this project, I learned several key insights about network performance. We observed significant regional variations, with some areas, such as Pays de la Loire and Normandie, experiencing more network issues. Additionally, we identified temporal patterns in network performance, indicating that issues fluctuated throughout the day and year. This insight suggests that dynamic resource allocation could improve service during peak times and seasons. We also discovered that building infrastructure, including factors like height and age, impacted network quality, with newer buildings generally offering better conditions for network signals. Furthermore, our analysis highlighted considerable network issues within public transport systems, particularly in metro and TGV networks, with major junctions and terminal stations being notable problem areas.

If I were to undertake this project again, there are several improvements. Integrating real-time data from public transport systems and other dynamic sources could provide more granular insights into network performance variations. Additionally, employing advanced analytical methods, such as machine learning algorithms, could help us predict network issues and optimize resource allocation more proactively. Incorporating real-time user feedback could also enable quicker identification and resolution of problem areas. Expanding the scope of data to include variables like weather conditions, traffic patterns, and other environmental factors would provide a more holistic understanding of the factors influencing network quality.

In conclusion, the business intelligence tools and data used in this project were not only appropriate but highly effective in addressing the network quality issues for Bouygues Telecom. The insights gained from this analysis are invaluable and will guide future efforts to enhance network performance and customer satisfaction.

5. Sources and appendices

5.1 Dataset

Fact table: ARCEP Dataset (2023)

- Name : 2023_QoS_Metropoe_data_habitations.csv / 2023_QoS_Metropoe_data_transports.csv
- Source : ARCEP
- https://data.arcep.fr/mobile/mesures_qualite_arcep/index.html

Population

- Name: insee_rp_hist_1968.xlsx
- Source: INSEE (2024)
- https://www.insee.fr/fr/information/2837787

Geographical Encoding Reference

- Name: georef-france-commune
- Source: Opendatasoft (2024)
- https://public.opendatasoft.com/explore/dataset/georef-france-commune/information/

Urbanization information

- Name: dossiers-recents-durbanisme
- Source: Opendata Paris (2024)
- https://opendata.paris.fr/explore/dataset/dossiers-recents-durbanisme/information/

Building information

- Name: volumesbatisparis
- Source: Opendata Paris (2024)

- https://opendata.paris.fr/explore/dataset/volumesbatisparis/information/

5.2 Reference:

1. RSRP and RSRQ
https://wiki.teltonika-networks.com/view/RSRP_and_RSRQ

2. Pourquoi le Wi-Fi dans le train est-il aussi mauvais?
https://www.bfmtv.com/tech/actualites/telecoms/pourquoi-le-wi-fi-dans-le-train-est-il-aussi-mauvais_AV-202307250292.html#:~:text=R%C3%A9seau%20mobile%20instable,Lugagne%20Delpon%20%C3%A0%20Ouest%2DFrance.