Mobile Radio Networks
Project Report: Model Mobile Traffic Dynamics
Bahram Hedayati
Mahsa Delaram

1. Introduction

The outbreak of the Covid-19 pandemic in early 2020 had a profound impact on various aspects of daily life, including the dynamics of mobile radio networks in Italy. As the virus spread rapidly across the country, stringent measures were implemented to control its transmission, leading to widespread lockdowns, travel restrictions, and remote work arrangements. These unprecedented changes in social behavior and mobility patterns significantly affected the usage and traffic dynamics of mobile radio networks in Italy. In this report, we will explore how the Covid-19 pandemic influenced the traffic patterns of a mobile radio network operator in Milan. By analyzing data from before and during the pandemic, we aim to understand the shifts in user behavior, network congestion, and the overall impact on the performance and resilience of Milan's mobile infrastructure. This investigation will shed light on the challenges faced by the telecommunications industry during this crisis and provide valuable insights into adapting network strategies to meet the changing demands of a pandemic-altered society.

2. Data Analysis

We are given a dataset containing 4 files in which there are some features and information related to the network sites and the covered sectors located in Milan in a period of 3 months. In particular, the dataset includes the information of January, February, and March of 2020 in which the February 16th is considered as the date when the full lock-down was applied by the government. It is worth noting that all the modelling implementation is done by the Python programming language in the Jupyter platform. In the following sections, we explain the details briefly.

2.1. Map Visualization

First, we demonstrated the exact location of each eNodeB on the map using Folium library to obtain a general insight of geographical positions of the network sites.

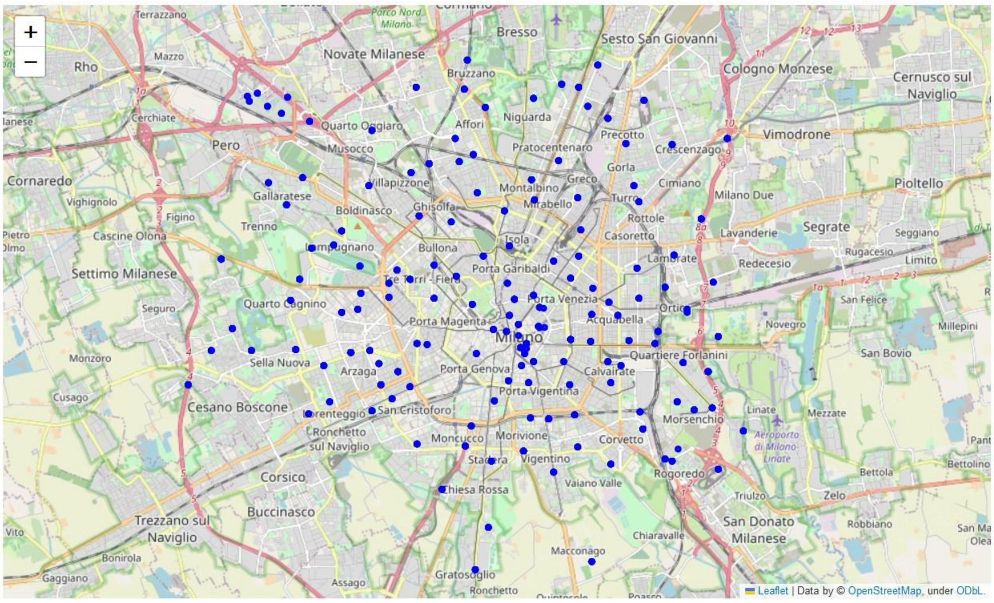

Figure 1: The location of network sites in Milan

2.2. Data Selection and Transformation

Since there were too many information values in the dataset, we decided to discard five features that had no value in the knowledge discovery process as they had the same or meaningless value. Furthermore, we combine the four files related to January, February, March network sites' information and the related locations into one single DataFrame and then split them to two separated DataFrames with respect to the full lock-down start date (February 16th). In addition, some required columns such as *weekday*, *time*, *total traffic*, etc. were added to facilitate the analysis process. At the end of this step, we had two DataFrames to analysis: *covid_free_data* and *lock_down_data*.

2.3 Median Weekly Signature

To reduce the complexity of the dataset, we exploit the concept of MWS which we learned in the Laboratory sessions of MRN course. MWS is the representation of the median behavior of an object with respect to a selected KPI. Precisely, for a fixed network site, a fixed KPI and an observation period T, the sample value of the site's MWS at hour i of day d equals the median of all the samples values observed at same hour of each day d contained in T. One of the selected KPIs in our project is the cumulative Downlink and Uplink traffic that we call it Total Traffic Volume[1]. It is the main KPI used in "A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas" paper which is written by Furno et al.

To create MWS of the KPI correctly, we extracted the amount of total traffic volume for each eNodeB with respect to the day. Then we computed the median value and saved it into MWS column of the dataset through grouping the filtered data (in terms of eNodeB and day) by TIME.

2.4 Location-based Graph Visualization

We tried to obtain a general view of KPIs' changes between the covid-free and lock-down periods of time in some special locations in Milan via creation of various plots in terms of MWS. Some of the plots represented significant changes in eNodeBs that are in the closest distance to the Politecnico di Milano (Leonardo campus), Centrale Station and Tre Torri are illustrated below as instance. The figures 2, 3 and 4 display Call Setup Success Rate, UL/DL Volume and RRC Success Rate for the mentioned locations, respectively. Other figures are attached to the report for providing more details.

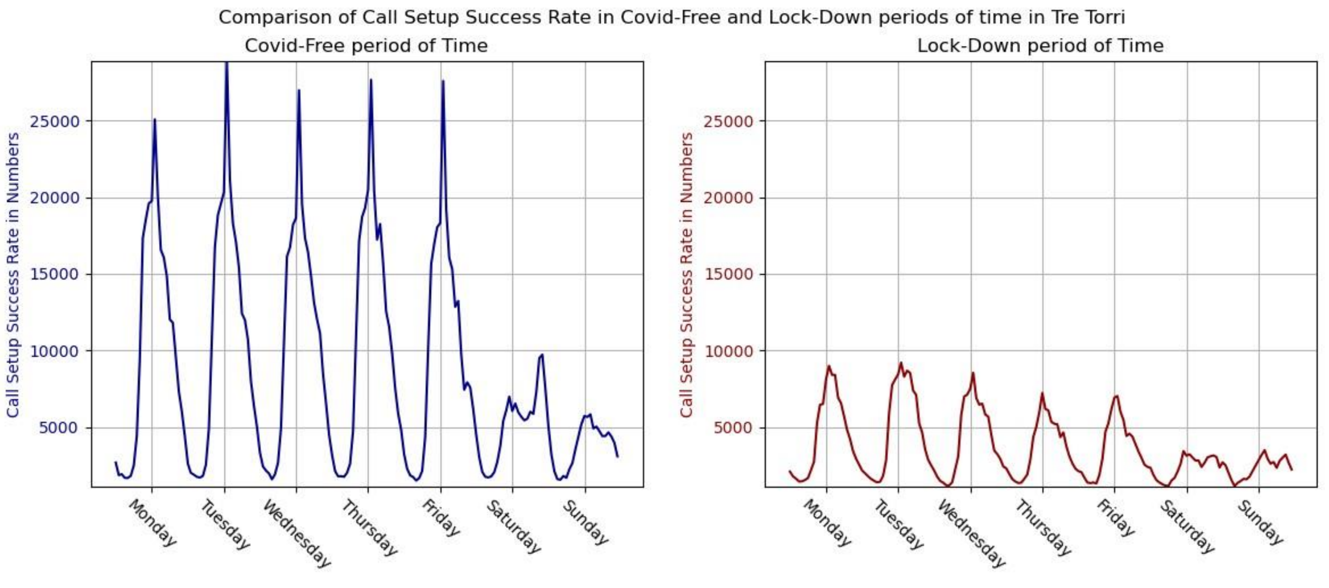

Figure 2: Extreme Reduction of Call Setup Success Rate in Tre Torri

---

[1] In the source code you can find it by the name UL_DL_VOL

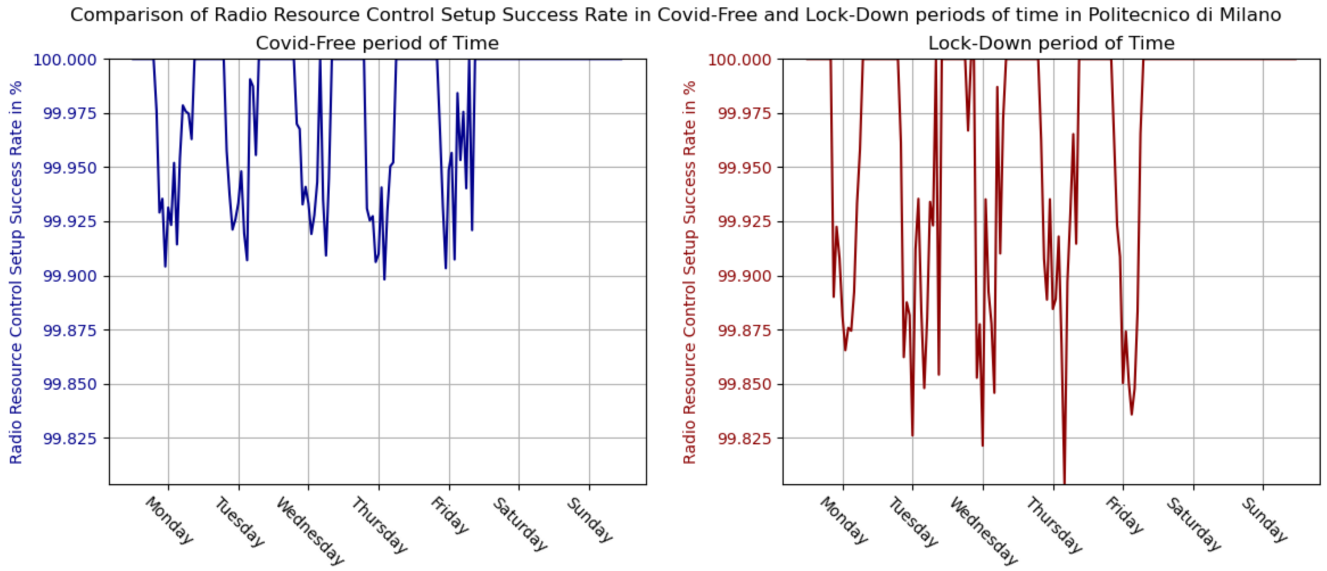

Figure 3: Degration of RRC Success Rate in Politecnico di Milano (Leonardo Campus)

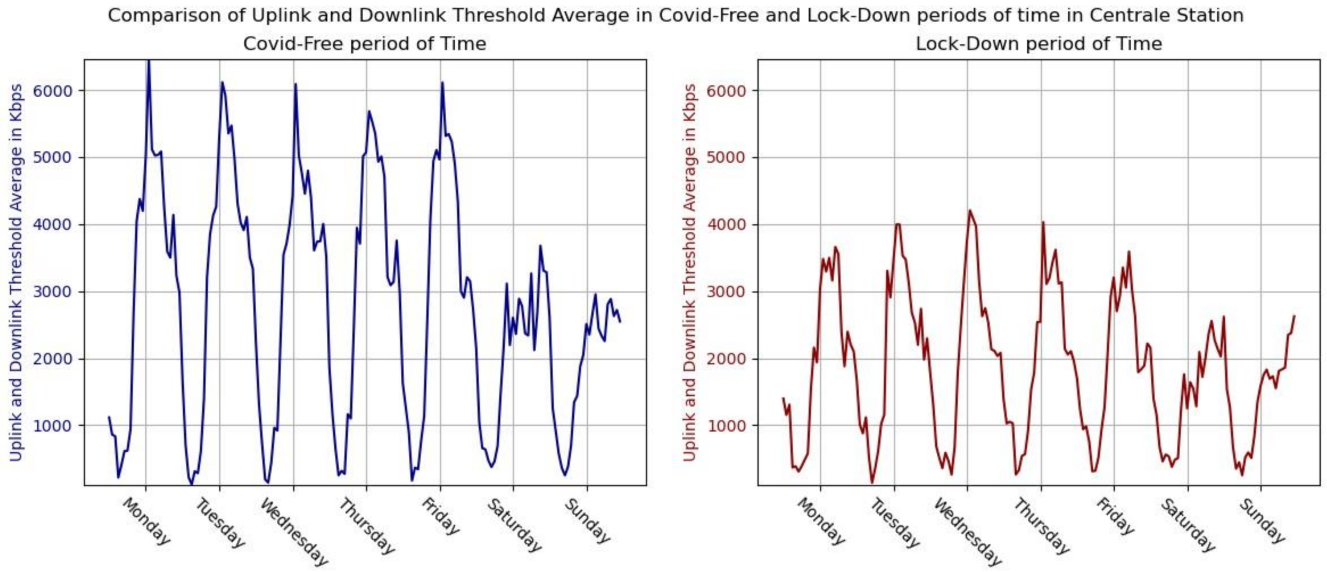

Figure 4: Reduction of UL/DL Threshold Average in Centrale FS

2.4 Clustering

To differentiate the residential and non-residential, we exploit the K-Means partitioning clustering algorithm. K-means is an iterative clustering algorithm used to partition data into K clusters, where K is a user-defined number. It starts by randomly placing K centroids, then assigns each data point to the nearest centroid. Next, it recalculates the centroids as the mean of the points in each cluster. The process repeats until convergence. K-means aims to minimize the sum of squared distances between data points and their assigned centroids, producing compact and well-separated clusters.

Because of the large scale of values in total traffic volume, we provided the normalized MWS with respect to the related eNodeB as the input to K-Means. The normalization method which we used in the project is standardization or z-score normalization. It is a data preprocessing technique used to scale numerical features in a dataset. It transforms each data point in a feature to a value representing the number of standard deviations it deviates from the mean of that feature. After some attempts, we find K=3 is the best number of clusters to partition pre-covid dataset. These clusters correspond to business, residential and transportation areas. To reduce the complexity of each cluster visualization, we consider the centroids as the representatives of the clusters. Therefore, we found the eNodeB which has the closest distance to the centroid of cluster and after that plot the MWS of the representative eNodeB for each cluster. Finally, we found the same eNodeB on the after-covid dataset to make comparison. The diagrams below show the clusters with respect to the UL/DL Volume KPI.

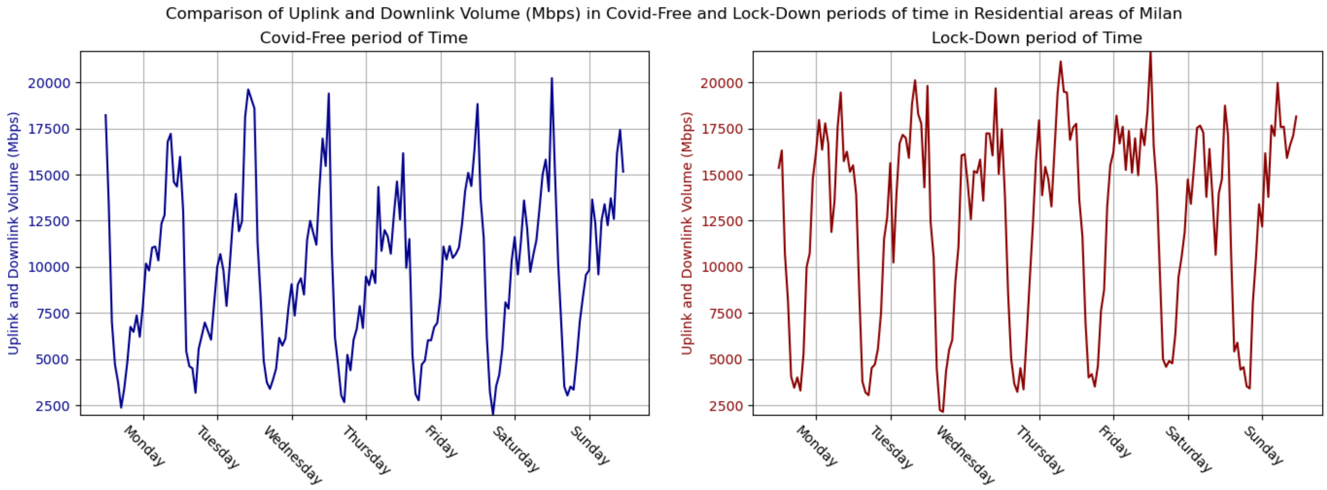

Figure 5: Growth of UL/DL Volume in Residential Areas

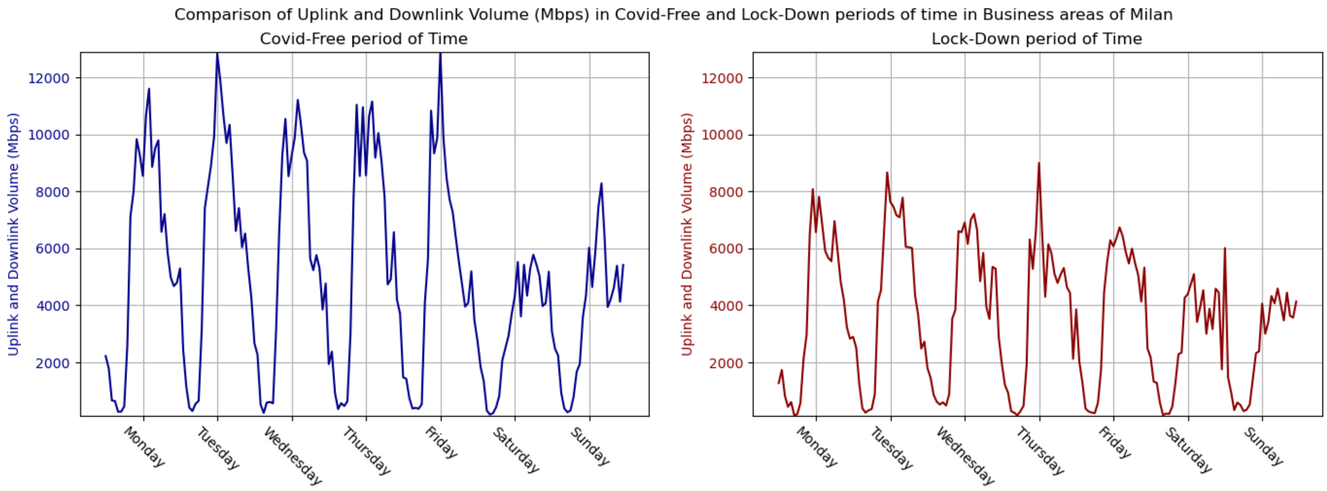

Figure 6: Dramatic Decrease in UL/DL Volume in Business Areas

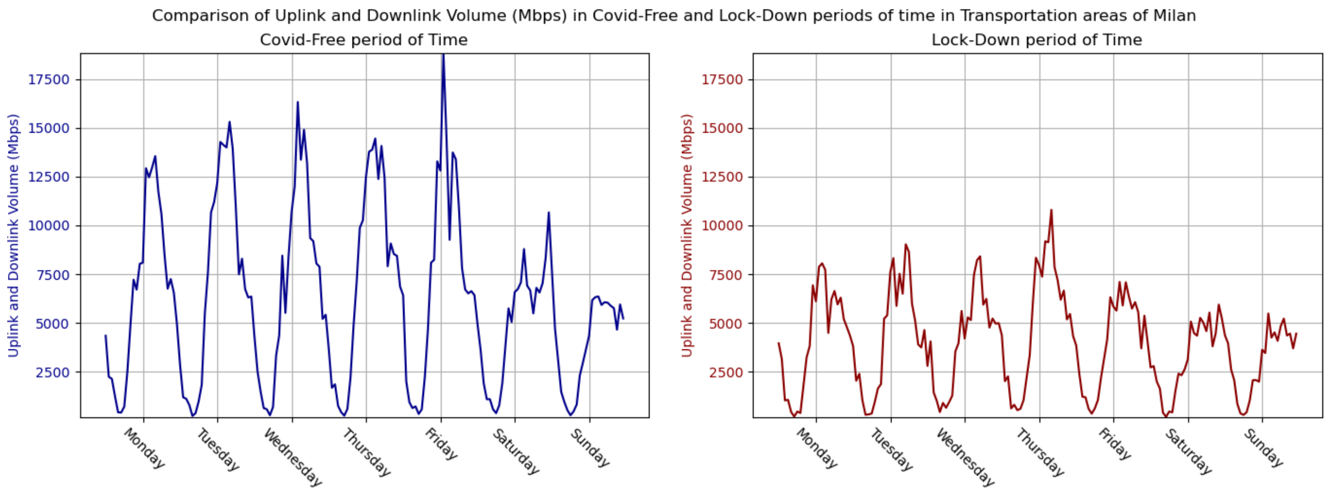

Figure 7: Radical Decline in UL/DL Volume in Transportation Areas

We did the similar analysis considering the RRC_RE_SR (Radio Resource Control Re-establishment Success Rate) in pre-covid and after-covid periods of time. The figures 8, 9 and 10 illustrate the changes of this KPI during the 3 months.

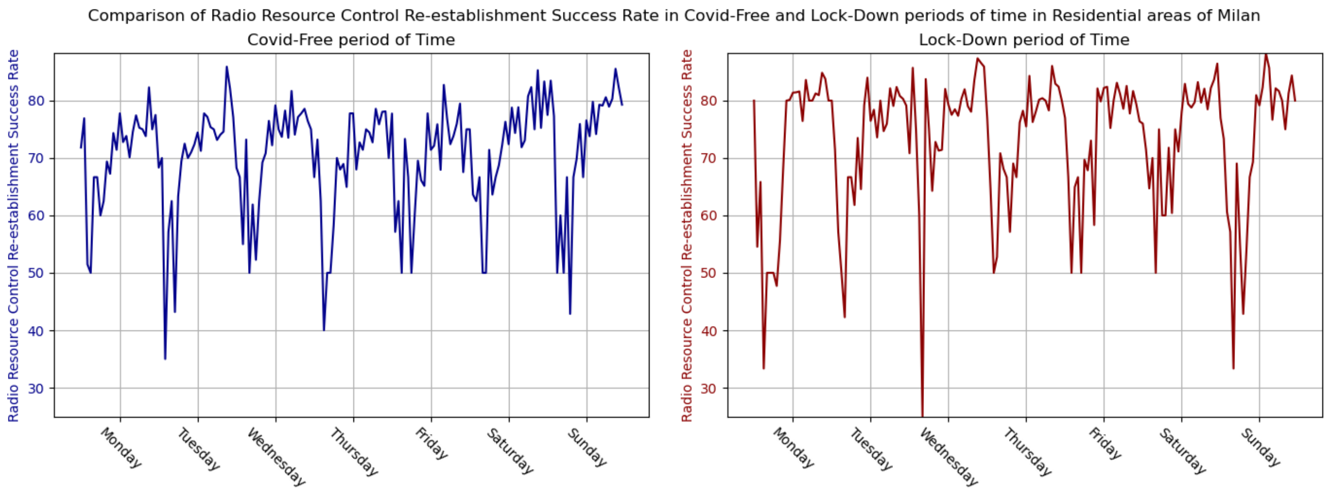

Figure 8: Reduction of RRC_RE_SR Rate in Residential Areas

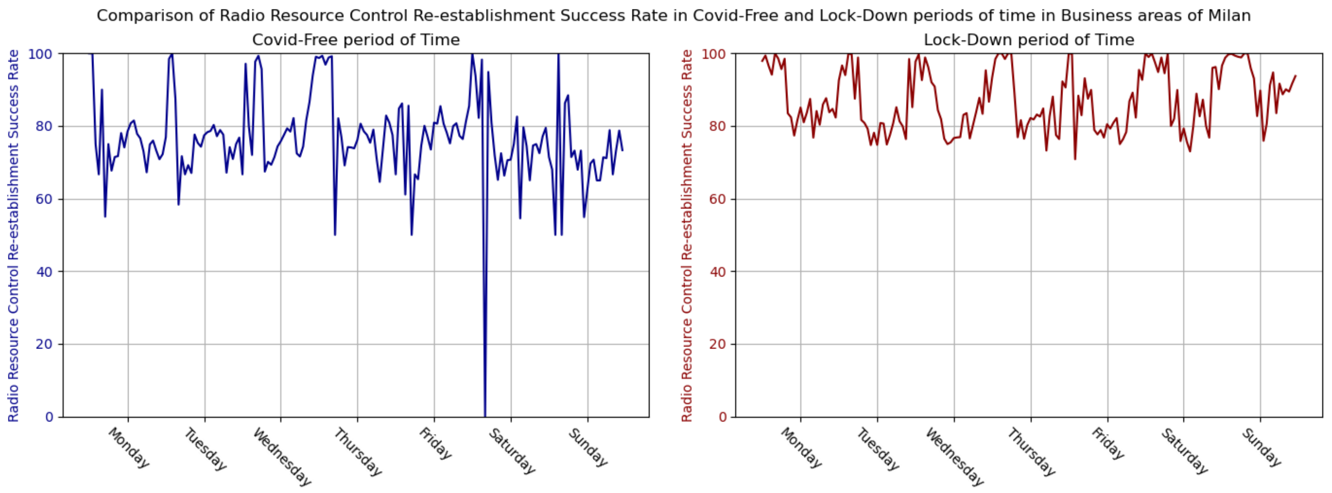

Figure 9: High Improvement of RRC_RE_SR Rate in Business Areas

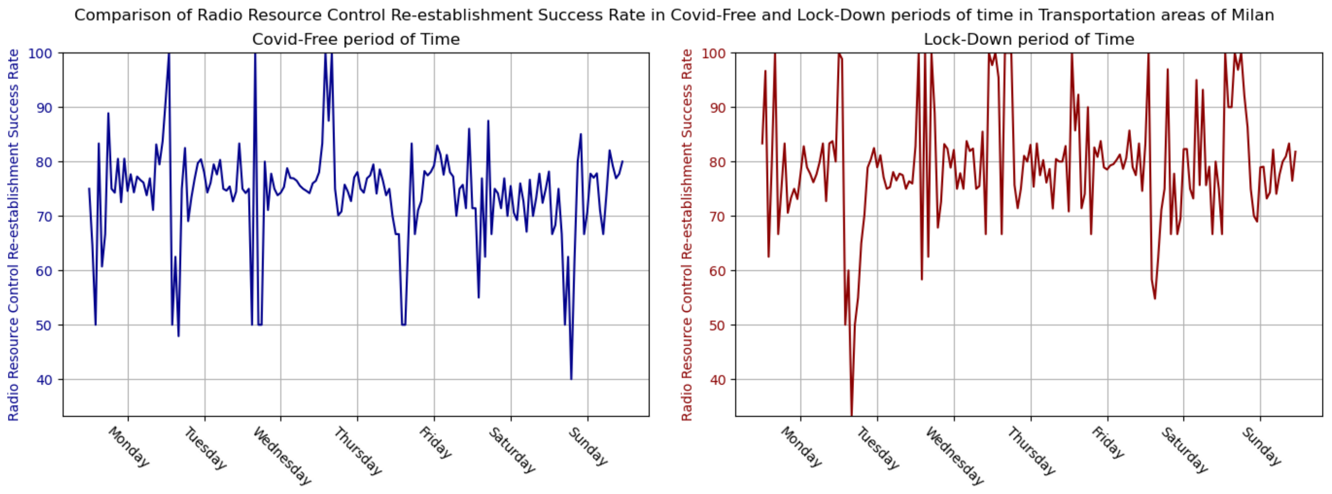

Figure 10: Improvement of RRC_RE_SR Rate in Transportation Areas

3. Results

Considering the total amount of Uplink and Downlink Volume as the reference KPI, the clusters described interesting patterns. In this section, we describe some of the results obtained from data analysis.

First cluster represents the eNodeBs that are located in residential areas as it shows a little amount of traffic in early mornings while the UL/DL volume grows up during the day with some spikes in nights. It can be seen that after announcement of the government on the February 16th in 2020 about the full lock-down, the maximum volume of UP/DL traffic increased by around 11% in residential areas, since many people stayed at home. Also, this led to more congestion in residential areas' network which can be seen in degration of RRC_RE_SE KPI in residential areas.

By looking at the second cluster, we can recognize the signature of business areas because of increasing trend of DL/UL traffic volume in the morning and existence of spikes in the evenings. In addition, the number of consumed DL/UL on Saturdays and Sundays is significantly smaller than weekdays. The 6-week period of lock-down due to the pandemic reveals approximately 17% reduction in the mentioned KPI as many companies' authorities decided to assign the tasks to the employees remotely or shut down the activities temporarily. In contrast to residential areas, we can see that the average network congestion disappeared during the lock-down period as the RRC_RE_SE indicator showed some improvements in successful re-establishment of radio resource control.

The most amount of decline is related to the third cluster which corresponds to transportation areas. Since a wide majority of people decided to stay at home or use their own vehicles, we can witness around 43% of changes through considering the maximum values of UL/DL volume during the representative week signature. It is worth noting that there were improvements in RRC_RE_SE KPI during the majority period of second half of observed time duration compared to the pre-covid time slice as the network congestion was strictly decreased.

4. Conclusion

Thanks to clustering technique, we can reduce the complexity of MRN datasets that lead to meaningful representation and more clarification. Furthermore, a wide variety of patterns can be revealed when we are facing special events. Network operators can exploit the strength of machine learning algorithms to describe their past data, current status and even predict the probable phenomenon that may happen in the future. In this way, they can be ready for sudden changes and modify their policies to stabilize the network, increase the satisfaction rate of subscribers and reduce the amount of customer churn.

5. Responsibilities

This project is done in a teamwork manner. Mahsa Delaram was responsible for studying the materials especially the papers provided by professor Pimpinella to gain sufficient knowledge about the project and various KPIs. After some meetings and discussions, we started to implement the knowledge discovery process using python and its powerful libraries. In particular, Bahram Hedayati wrote and debugged the codes. Moreover, he is the person who suggested the visualization styles for better presentation. All in all, both of us cooperated effectively in theory and practical parts of the project and we proud ourselves that we can did it finally, despite the challenges we faced.