Impact of COVID-19 on Mobile Networks – Project #2

The project aim was to analyse the LTE mobile radio network behaviour before and after the outbreak of the COVID-19 disease, in the metropolitan area of Milan, hence studying the traffic related to the months of January, February and March of 2020: considering the period before the 16th of February as a "Covid free" observation period, and full lockdown starting from the 9th of March.

The chosen approach has the following structure:

- Retrieving the geographical positions of the various antennas from the "Coordinates_MILANO.csv" file.
- Checking and organizing the original given data for further processing.
- Analysing the behaviour in the month of January, in terms of designated Key Parameters Indicators, to obtain median weekly signatures of the traffic (that would later be useful to discriminate the type of area) and computing the calculation of the median for the periods of February and March as well.
- Clustering through the KMeans Classifier in an unsupervised Machine Learning environment, using January as reference.
- Focusing on a limited number of eNodeBs, selected from the KMeans clusters, we compared the different trends found in the three months taken into consideration, highlighting what happened because of the outbreak of the disease.
- Verifying the correctness and coherence of our results with other general information.

In the upcoming paragraphs, we'll discuss each of these logical steps and analyse the obtained information to complete the request.

The whole project was carried out on Jupyter Notebook, because of its ease of use given the possibility to work step by step with singular cells rather than with a whole python program.

Please Note: the ".ipynb" files that we used for the project work fine with our own file path to the correct directory; in order to check the results' correctness or just to reproduce our work, one may be careful to put the proper files path. Regardless of this problem, the use of Jupyter Notebook allows the share and the check of the work with the results already stored in (cells are runned and the ipynb checkpoints are kept).

1. Retrieving eNodeBs' position

In the "cells_addresses.ipynb", we achieved reverse geocoding from "Coordinates_MILANO.csv". In order to make things clearer and visualize these data, we decided to find the location of every 4G antennas in the city of Milan listed in the file.

We exploited the "Nominatim" geopy library's command to extract the exact addresses, with Google Maps as user agent, then we managed to plot them on a map thanks to "plotly_express" and the use of a mapbox access token. Afterwards, for the purpose to make the visualization clearer, we simply assigned different colours to the eNodeBs to highlight the administrative subdivision group (meaning the "Municipio") they belonged to.

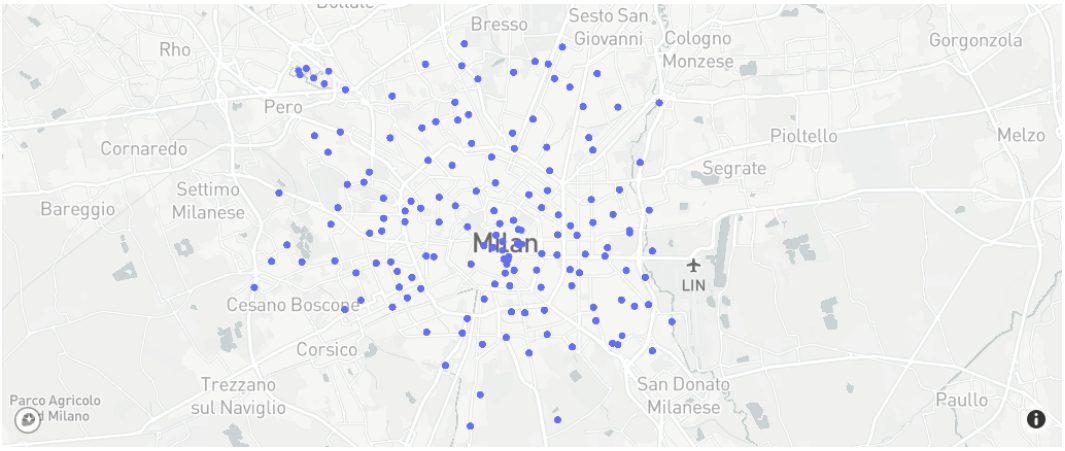


Fig.1: eNodeBs in Milan.

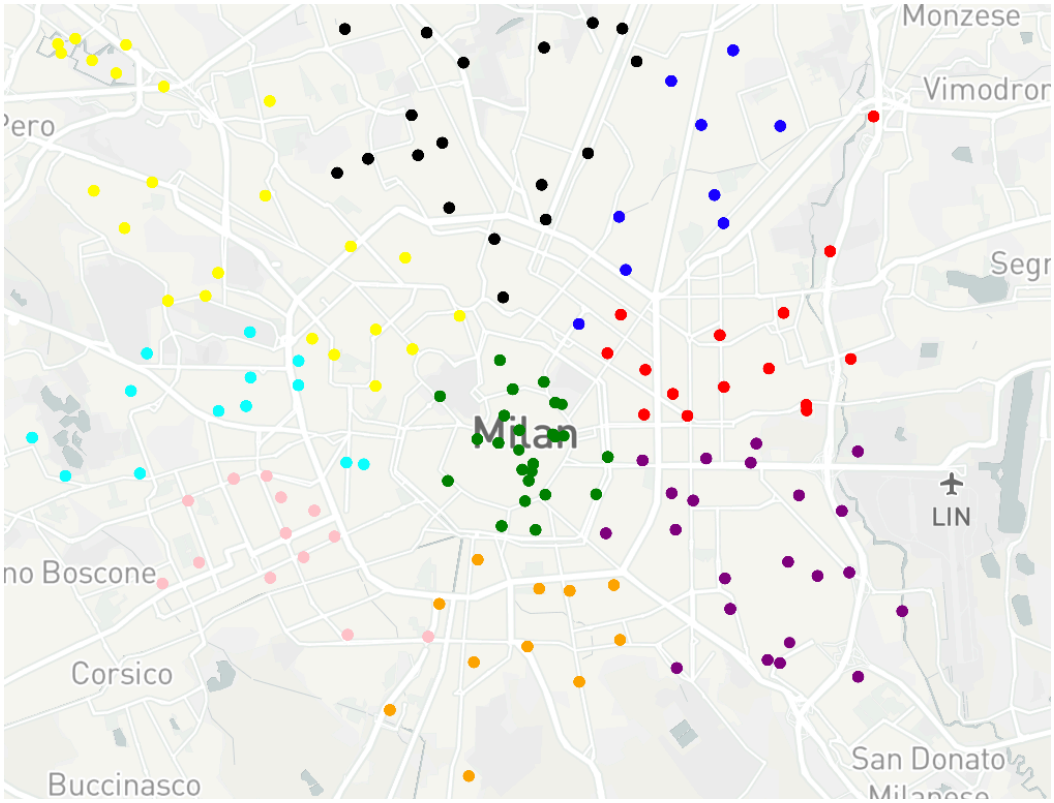


Fig.2: eNodeBs in Milan divided by “Municipio”.

The results of this first steps of the project have been saved in the “Cell_address.csv” file, available in the “data/location_info” path.

2. Data organization

Before proceeding with further steps into the analysis, we decided to check whether the available files contained data related to the same antennas, to understand if we could later use a systematic approach to study their traffic. This inspection was carried out in the “data_check.ipynb” notebook.

We noticed that, in the "Milano_800_January_MRN.csv" file, there were 398 unique ECELL_IDs, while in the February and March files, there were 395 consistent cells, with the following being absent:

- 5a6ea208eb7b2dcbf59dc457706b8b04a5a8e1b2
- b713a4f2dc59f35a0dfa1e598cf4cf01659bd2e8
- b7eed02edd74696bf65d055085787772478961de

We decided to drop those three reported above and focus on the remaining 395, which, we double checked to be the same in the rest of the months: the outcome file was saved as "Milano_800_January_New_MRN.csv".

The exact procedure was applied for the "Cells_address.csv", where we found 567 additional cells, meaning those weren't used to register traffic data over the three months, therefore they were disregarded. As a result of this, we managed to obtain a coherent set of locations as well, under the name "Cells_adress_updated.csv" (also in the "location_info" folder).

3. Median weekly signatures

For this step, our goal was to extract median weekly signatures of the traffic, related to a single chosen KPI, from the month of January. This was done to have a baseline for the dataset to later use as input for the KMeans Classifier: since one could safely say January is a Covid Free period of observation, we determined it would be the most appropriate to classify the cells in different areas, but this section of the project will be discussed in the next paragraph.

After many considerations, we decided to focus mainly on the KPIs below, respectively belonging to the categories of usage, mobility, and accessibility:

- "DL_VOL" and "UL_VOL" values, namely the downlink and uplink traffic volume. We hypothesized they might have been fit parameters to analyse in this case scenario, so to shape the changes in the throughput and number of active DL users (those with active data transmission in the downlink buffer), and active UL users (similarly).
- The "InterF_Hout_SR" attribute, related to the success rate of outgoing inter-frequency handovers.
- The call connection set up success rate, indicated with "CS_SR".

In the "median_weekly_signatures_jan.ipynb" code, we started from creating new csv files for each day of the week with the corresponding data gathered by every one of the 395 ECELL_IDs: those were individually saved in the "data/weekdays/january" directory.

We proceeded by creating seven new data frames with hourly medians of the selected KPIs, for the single corresponding day: referring to Fig. 3, we calculated the medians over all the Mondays of the month at 00:00, then at 01:00, at 02:00, at 03:00...etc, iterating over the 24h and repeating the process for all the involved cells. Subsequently, we replicated this course of action for the Tuesdays, Wednesdays, Thursdays, Fridays, Saturdays, and Sundays.

	ECELL_ID	Hours	Median DL_VOL	Median UL_VOL	Median CS_SR	Median InterF_Hout_SR
0	b7de6cd46e855aaec8e28b26d1767e7a738819db	Monday at 0	18709124.0	2260072.0	12320.0	95.287304
1	b7de6cd46e855aaec8e28b26d1767e7a738819db	Monday at 1	9233704.0	616256.0	10839.0	89.370714
2	b7de6cd46e855aaec8e28b26d1767e7a738819db	Monday at 2	2807488.0	326460.0	9835.0	91.911765
3	b7de6cd46e855aaec8e28b26d1767e7a738819db	Monday at 3	3334156.0	223312.0	9933.5	88.709677
4	b7de6cd46e855aaec8e28b26d1767e7a738819db	Monday at 4	1373080.0	185400.0	9474.5	85.641026
...	...	...	...	...	...	...
9475	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 19	7264136.0	1007712.0	6678.0	98.671096
9476	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 20	11664984.0	868896.0	4760.0	99.431818
9477	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 21	9863712.0	1346504.0	4532.0	99.047619
9478	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 22	11684384.0	684760.0	3661.0	98.701299
9479	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 23	10293712.0	1162848.0	3049.0	96.428571

9480 rows × 6 columns

Fig.3: eNodeBs in Milan divided by “Municipio”.

The next step was aggregating all these data frames in a collection organized in specific way: we achieved to have one data frame per cell, where all its particular statistics data over the course of the month was stored with one hour granularity from Monday through Sunday. Thanks to this and the “matplotlib” library, we were able to plot the median weekly signatures of every ECELL_ID: we reserved ourselves the possibility of graphing them all and only graphing a selected one, for later use in the study of the network trends and changes.

	ECELL_ID	Hours	Median DL_VOL	Median UL_VOL	Median CS_SR	Median InterF_Hout_SR
0	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 0	4667236.0	461656.0	1638.5	97.916667
1	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 1	4532380.0	298996.0	1470.5	100.000000
2	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 2	445604.0	23504.0	1373.0	100.000000
3	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 3	774052.0	21664.0	1170.5	100.000000
4	dccb348b4375dfd2ab29330d847eccb81784e650	Monday at 4	151776.0	17840.0	1079.5	100.000000
...	...	...	...	...	...	...
163	dccb348b4375dfd2ab29330d847eccb81784e650	Sunday at 19	8013108.0	642772.0	3903.0	97.894847
164	dccb348b4375dfd2ab29330d847eccb81784e650	Sunday at 20	4224040.0	582452.0	4034.0	99.009901
165	dccb348b4375dfd2ab29330d847eccb81784e650	Sunday at 21	8219692.0	2088104.0	4026.0	97.080039
166	dccb348b4375dfd2ab29330d847eccb81784e650	Sunday at 22	9207340.0	834000.0	3246.5	96.527778
167	dccb348b4375dfd2ab29330d847eccb81784e650	Sunday at 23	5857788.0	637596.0	2563.5	100.000000

168 rows × 6 columns

Fig.4: Random data frame chosen from the created collection of data frames, referred to the cell “dccb348b4375dfd2ab29330d847eccb81784e650”: only its related medians are shown.

The same execution will be found in the "median_weekly_signatures_feb.ipynb" and "median_weekly_signatures_mar.ipynb" notebooks. At this point, we had the foundations for the main objective of the project: the data analysis.

4. KMeans Classifier

With the aim of reducing complexity and group network sites according to spatial and temporal dynamics of the served traffic, we concluded the KMeans algorithm was needed in order to keep a data driven approach, with the rationale that the network activity of eNodeBs in the same cluster can be explained with the same model.

As reported in "dataset_for_kmeans.ipynb", we made the choice of only using "DL_VOL" medians as input for the classifier and that can be explained with two main reasons:

- Through trial and error, we found out that too many KPIs' statistic values were detrimental for the outcome of the clusters, meaning that we couldn't spot consistent patterns.
- We directed our attention on a usage prospective and given that nowadays the traffic is not evenly distributed between downlink and uplink, the former being more stressed than the latter, we finally decided on the DL_VOL.

Then, we exploited the command "pivot", to reshape the collection of data frames previously created and explained. The goal was to have one unique ECELL_ID per row, and its medians for every hour of every day of the week per column, as seen in Fig.5.

We obtained a dataset fit to replicate the weekly trend of each cell and to interpreted by the machine learning environment. It was saved under the name "weekdays.csv" in the "data/k-means/dataset" path.

Hours	Monday at 0	Monday at 1	Monday at 10	Monday at 11
ECELL_ID				
00199ffaae391a819e512627bdaff10347c4fd13	33069296.0	22457216.0	15670440.0	12483012.0
019cfa281650073eb1ed07c2202dbc289d0423c5	19330088.0	8285528.0	11551984.0	13562752.0
01e15165b5683c9f4922e7ffdc82471e479e7779	10074440.0	4853896.0	4769392.0	5219648.0
0424fc3a83b05aa9108aabdd8cbb087814a34bd1	15496852.0	18425572.0	6530156.0	10216108.0
046b22f684551e03511a06d4c2489c4ce47e9446	627112.0	311420.0	2018144.0	2569952.0
...	...	...	...	...
fbb375f9f0a4f004475be02539cd98ad87edefac	7522292.0	4018588.0	19569780.0	28475784.0
fc151d76149c8066a9d19e0279dcf9f5fbfbb2e0	3258596.0	1298648.0	13649216.0	14264768.0
fde3676655dbd432364e7468ba06a2daaf385e4c	38738648.0	36594952.0	3148840.0	5227320.0
fe64706273bed80a3fa4a13093f4d8d825ab2228	8337952.0	6019552.0	29810248.0	35337528.0
ffbb644800b874c89350a70d5acdc8d0fe051755	8004672.0	12455744.0	13941792.0	15784076.0

395 rows × 168 columns

Fig.5: Dataset for KMeans Classifier partially shown.

After importing the said file in the “k-means.ipynb” code, we proceeded with the necessary pre-processing requirements, which are performed by the “StandardScaler” command and the “scaler.fit_transform()”, in order to standardize the dataset features removing the mean and scaling to unit variance. Then, we used the algorithm, choosing k=3, since our goal was to categorize antennas in residential, non-residential and transports areas. The corresponding clusters’ centroids calculated (with “kmeans_model.cluster_centers_”) are located in the directory “data/k-means/result” under “centroids_kmeans.txt”. The final clusters can be found in “Clusters_weekdays.csv”, in the same folder.

At this point, we had no deterministic way to validate the outcome of the algorithm, nonetheless we tried plotting the median weekly signatures of the cells, cluster by cluster. This work was carried out in the “graph_for_cluster_check.ipynb”, where we learned the antennas had been grouped as the following:

- cells in transportation area: 202

- cells in residential area: 38

- cells in non-residential area: 155.

The acquired graphs showed promising results: by simply scrolling through them, we could recognize the three different curves we were looking for and they seemed coherent enough inside the single cluster.

Moreover, since we were working in an unsupervised scenario, we opted for cross checking this outcome with a better appropriate approach.
We realized that the true differences between the plots could always be found at the end, where the data related to the days Friday through Sunday were represented.

We thought this “deal breaker” could be exploited to clean up the clusters and that’s how we proceeded: in the “k_means_weekend.ipynb” code, we repeated the same steps as in “k_means.ipynb” but we filtered for only Friday, Saturday and Sunday, we saved the cluster’s centroids as “centroids_kmeans_weekend.txt” in the “result” folder as well, and the final cluster obtained were imported in the “graph_for_cluster_check.ipynb”, then, we compare them with the previous ones and only kept the matching clusters.

We learned the information below:

- cells in transportation area after crosscheck: 11

- cells in residential area after crosscheck: 30

- cells in non-residential area after crosscheck: 20.

We moved forward with graphing them again, and, being satisfied with the attained clusters, we selected one cell each to be subjected to the “pre and after lockdown” analysis:

- Residential: d055f65a544c72b0ac3aafc30e2721ff4348b273.
- Non-residential: 99e9658069bd9ab5c29e413305bf15a4855ef553.
- Transportation: 087aa4fdc05a3b32fc33f8af4fc17d1fbda02ecc.

5. Data analysis

The analysis was focused on the study of the weekly median signatures for the above-mentioned cells, over the months of January, February and March. In this section, we'll show and discuss the trends of the graphs, in order to understand how the Covid-19 disease has impacted data traffic in the metropolitan area of Milan.

- ***Residential***

Considering a cell which covers a residential area, we may expect a consistent daily trend in the downlink and uplink volume, since it's reasonable to hypothesize that there shouldn't be particular events causing an increase in the amount of traffic in one specific day rather than another. In addition, due to the fact that people generally come back to their house in the evening (after work or other daily activities), we may expect an increase in the internet usage during those hours.

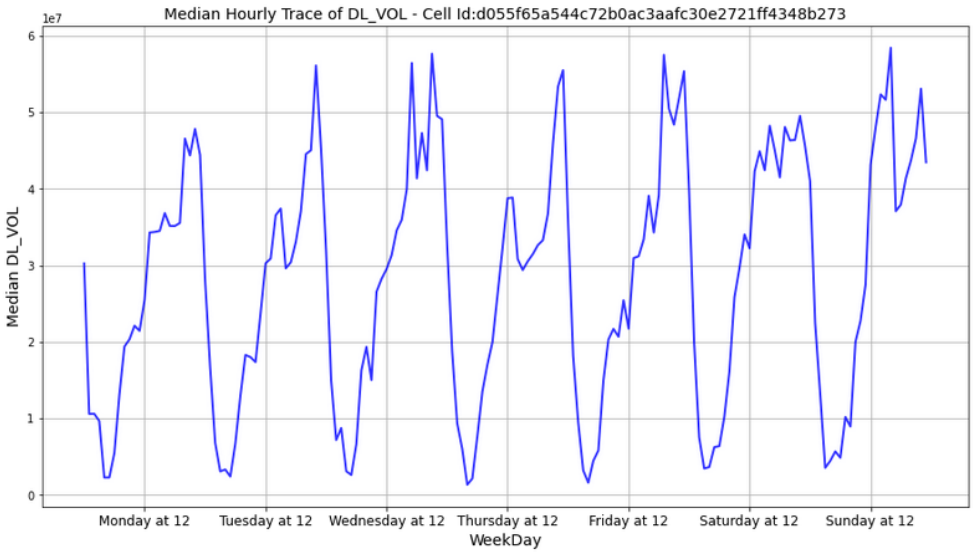
In Table 1, we reported the weekly median signatures of "DL_VOL" obtained for each month for the chosen cell. At first glance, the trends matched our expectations, since the daily curves are more or less identical and, towards the evening, the downlink volume values reach their peaks. Up to this point, given these results, our analysis may be considered successful.

For the purpose of understanding if and what changes may have happened because of Covid-19, and all its consequences, one may check the actual values represented on the y axis. Even if the graphs seem very similar (nearly identical) to each other, in the first two months of the year we have numerical values, related to the lowest and highest peaks, comprised between 3 and 5; in March, the situation changes, with the previous interval becoming 2 to 4, and the highest peaks becoming less narrow and wider, while some smaller peaks can always be spotted in the morning time.

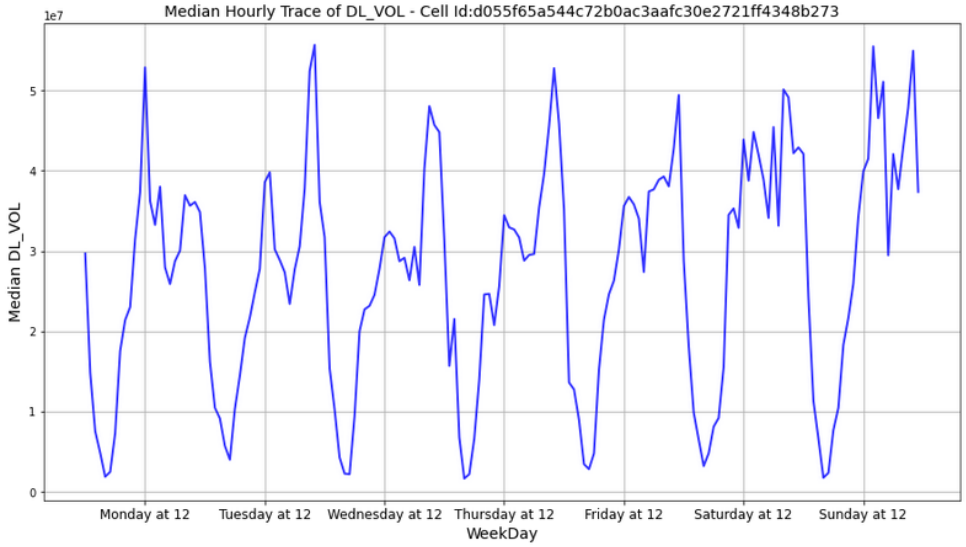
What we observed makes sense because of the reason that lockdown officially started only on the 9th of March, thus effecting the people usage of the download link. One may especially highlight that the students of every grade started online classes and the adults started working by remote during the morning hours: this might explain the presence of said smaller peaks before 12:00 PM, which, on average, weren't above 4 in the months before.

Still taking into account a common school schedule, the visible depression occurring right after that time could be elucidated by the students turning off their devices for a lunch break: this is an aspect present also in January and February, although, from their signatures, we can see that in their cases the apices only ever happened after the lunch hours, when the students would commonly return home from school.

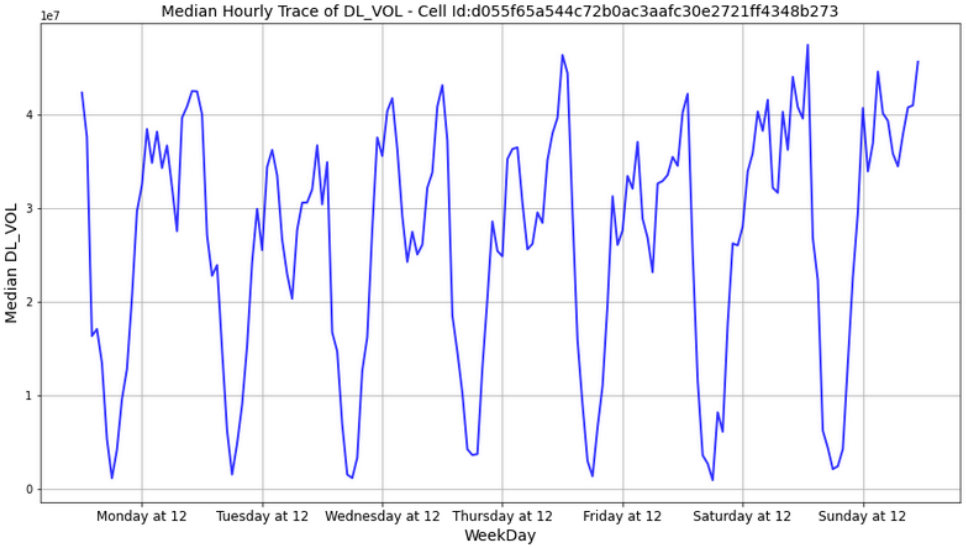
The period after the 16th of February has to be considered as "Covid affected" so we could have expected some differences compared to January. This has not actually been seen but, it has to be noted that in this period no restrictions or changes in habits and lifestyle were happening yet: so the graph might be considered coherent with reality, and even if there were some side effects, choosing the median as a statistical metric has led to a sort of "weighted" graph, canceling them out.



January



February



March

Table 1: DL_VOL

In Table 2, the “UL_VOL” weekly median signatures for the interested cell are available. As one may expect, the trends are very similar to the ones seen for the downlink parameter. In the first two graphs, we can still observe peaks, in a 3-5 range again, during the hours of the evening, generally the time window in which all family members (workers/students) should be at home: in the morning, on the other hand, the values have the4 tendency to stay under the 4 threshold. The biggest noticeable difference in March is found in this times: it may be noticed the apices increases to 4-5: nonetheless, also here a decline is spotted around lunch hours, probably due to the same reasoning explained above. Talking about numbers, as expected from the beginning, the upload link is significantly less stressed than the downlink, with values in the order of 10^6, instead of 10^7 as seen previously.

In Table 3, we explicated the “CS_SR” (RRC Call Connection Set Up Success Rate) weekly median signatures for the chosen residential cell. In this case, being indeed a success rate, we can do a qualitative description rather than a quantitative one, taking into consideration changes in the accessibility of the network.

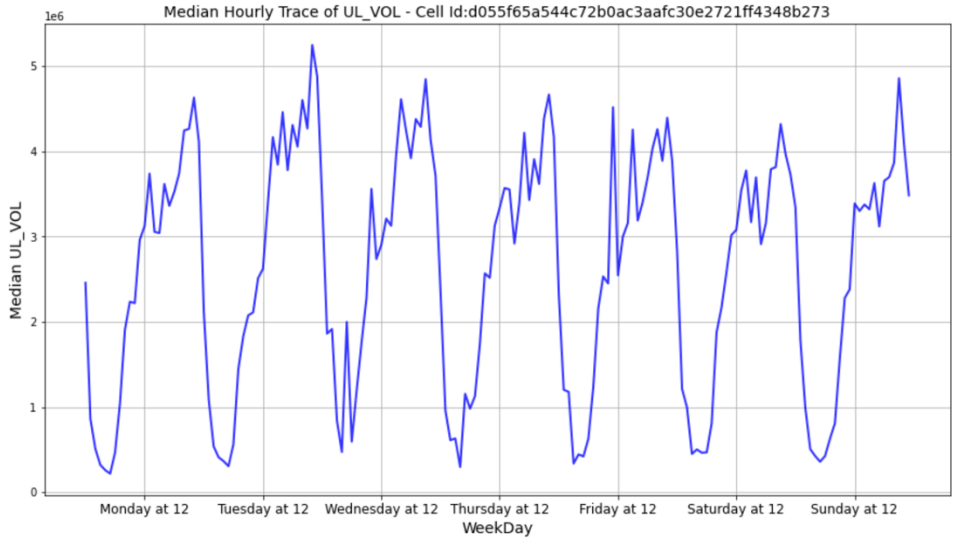
The trends from January and February seem consistent with wide peaks around 12:00PM and again in the late afternoons, likely due to the use of devices after school/job hours and the social life events at night: there’s no surprise that the highest crest corresponds to Saturday evenings. One may say these first two graphs keep the same evolution over the week, as opposed to what it is shown in March: here, the greatest value reached is around 25000, while before it was around 35000, so we register a 28,6% decrease. Moreover, the apices are not as “periodic” and regular, they are rather narrow with frequent and small ups and downs. These observations indicate that the Call Connection Set Up Success Rate experienced a decline, which translated into a poorer accessibility, probably due to the sudden rise in users trying to connect at the same time, during lockdown.

Lastly, the “InterF_Hout_SR” (Inter Frequency Outgoing Handover Success Rate) weekly median signatures are reported in Table 4: again, the analysis will be qualitative as for the previous parameter.

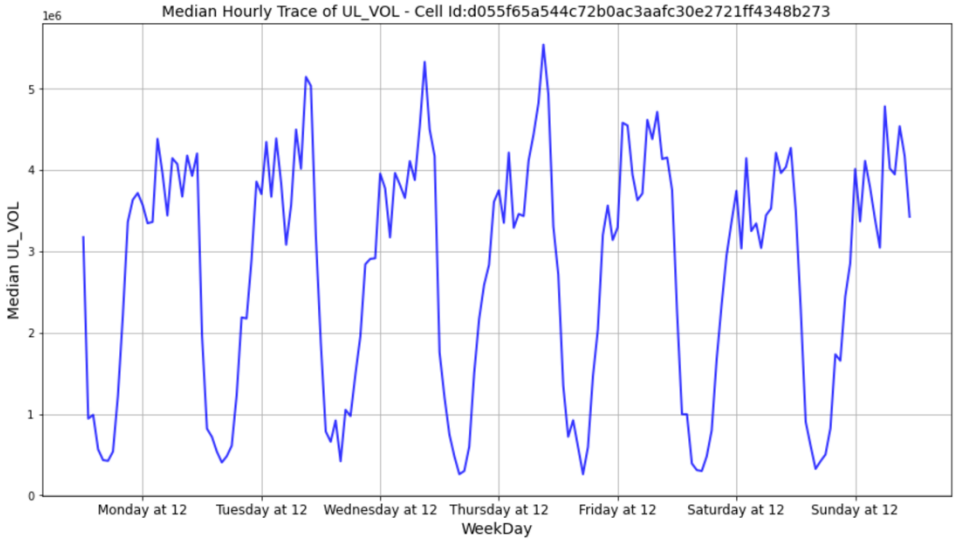
The main aspect that can be noticed is that, generally, the handover success rate is almost always near 100%, with higher values registered during the day while lower ones during the night. The situation plotted on the January graph worsens month by month, with the maximum threshold decreasing by 2% circa in February, and almost by 4% in March, indicating many outgoing handover procedures didn’t go through correctly.

In terms of mobility, the mandatory lockdown reduced the number of individuals going around the city, therefore needing to update their location and switch cells, but this parameter being a rate of success we can’t deduct any quantitively information from it to confirm this aspect.

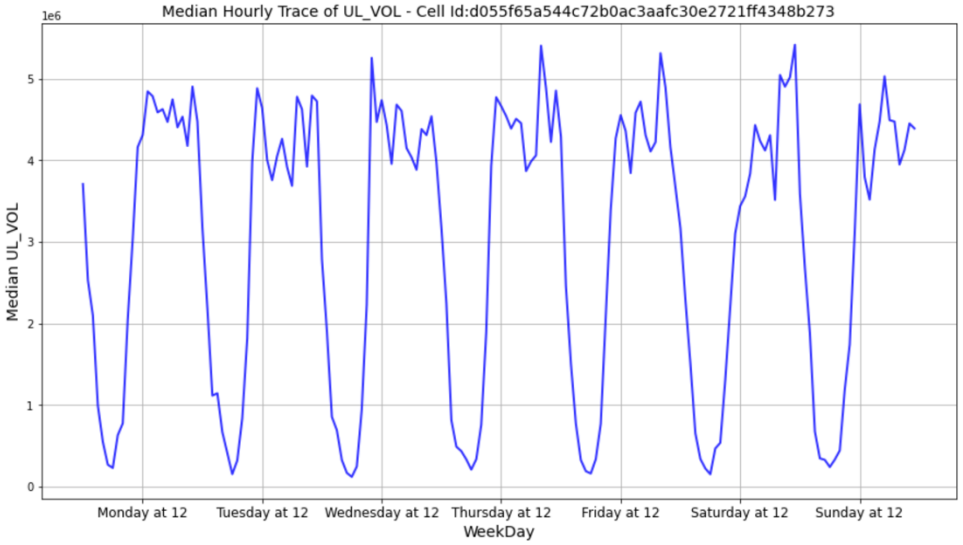
We can only state that a more stable behavior around the 100% threshold was expected in this period of observation, since less people were going out, and although these conjectures weren’t matched, the results are interesting to speculate on. The causes for a bad handover performance can vary, in our opinion, the most likely in our scenario might be: congestion and uplink interference. Such phenomena are especially plausible during the “Internet rush hour”, in the late afternoon when there is very high demand: hence, we do spot declines in those times in all the graphs, of course for said reasons they appear steeper in the March one.



January

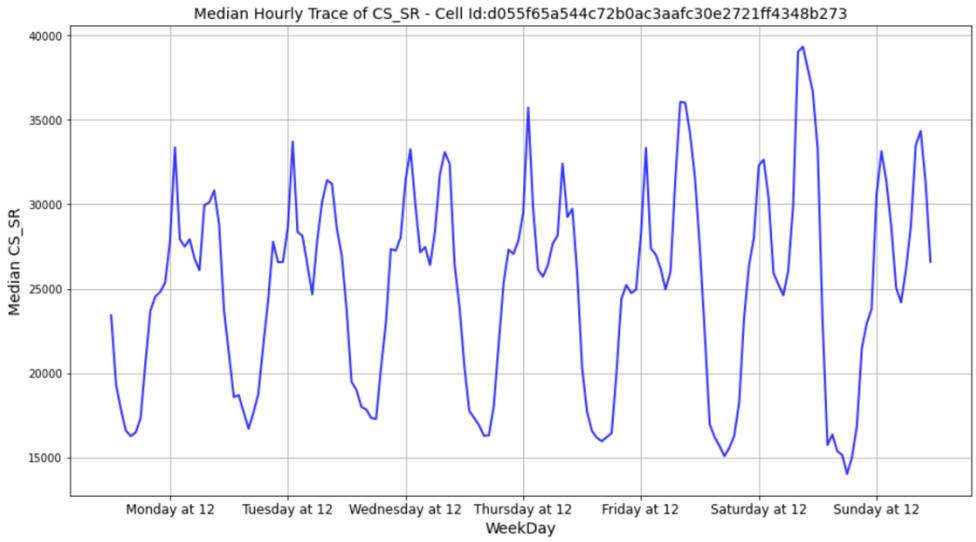


February

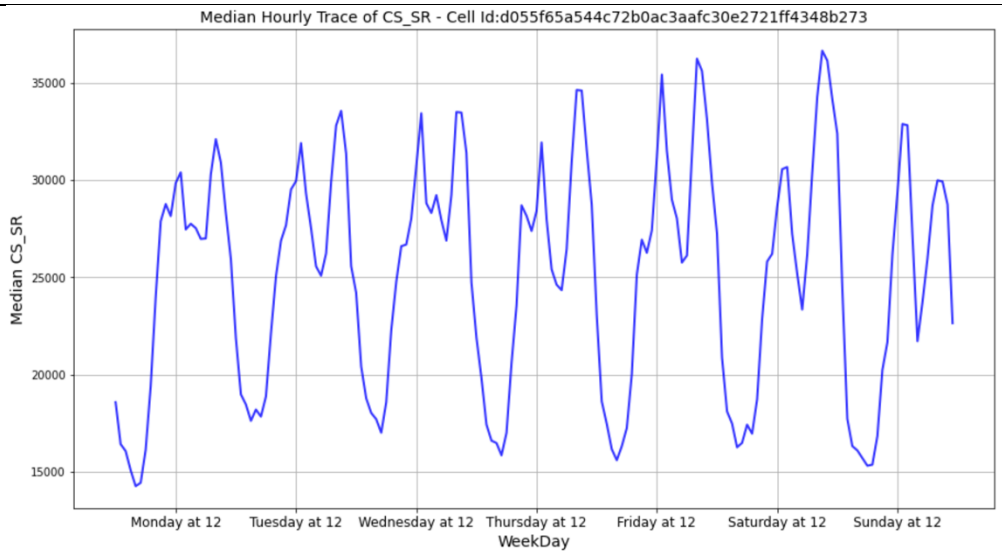


March

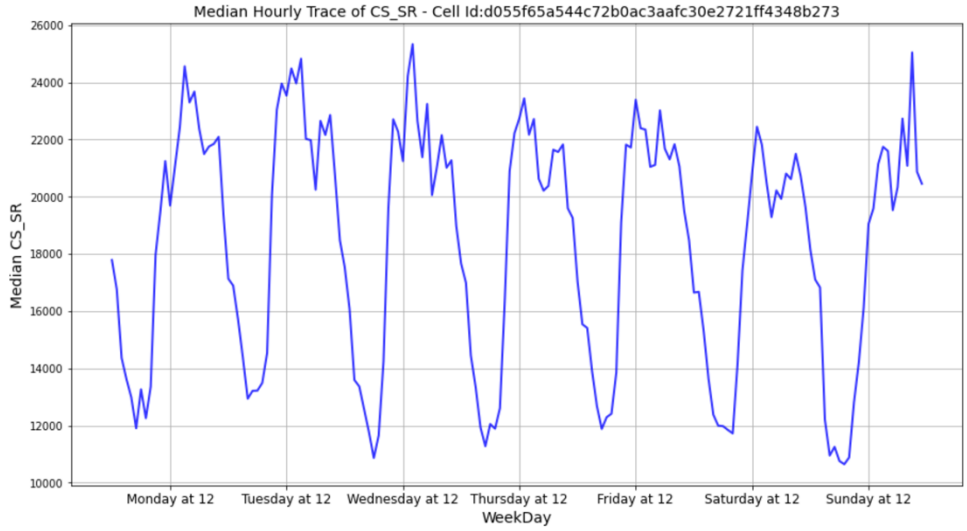
Table 2: UL VOL



January

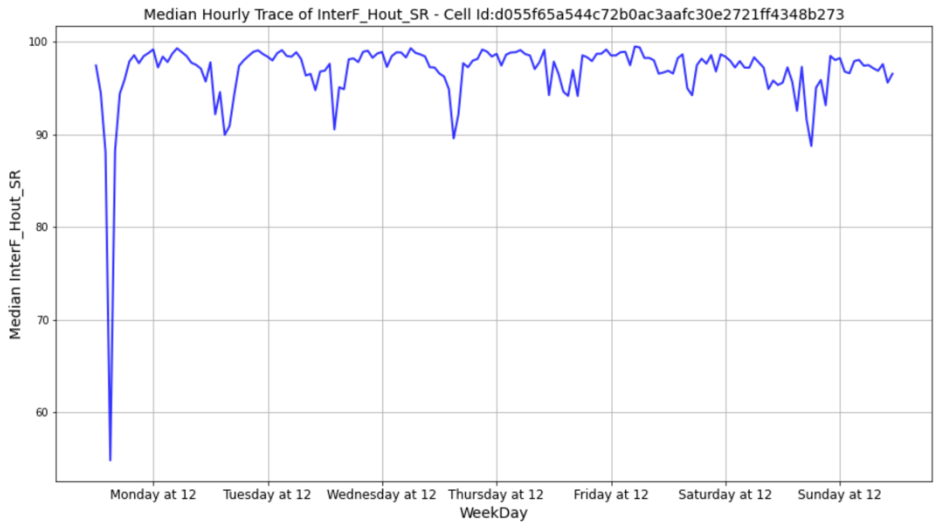


February

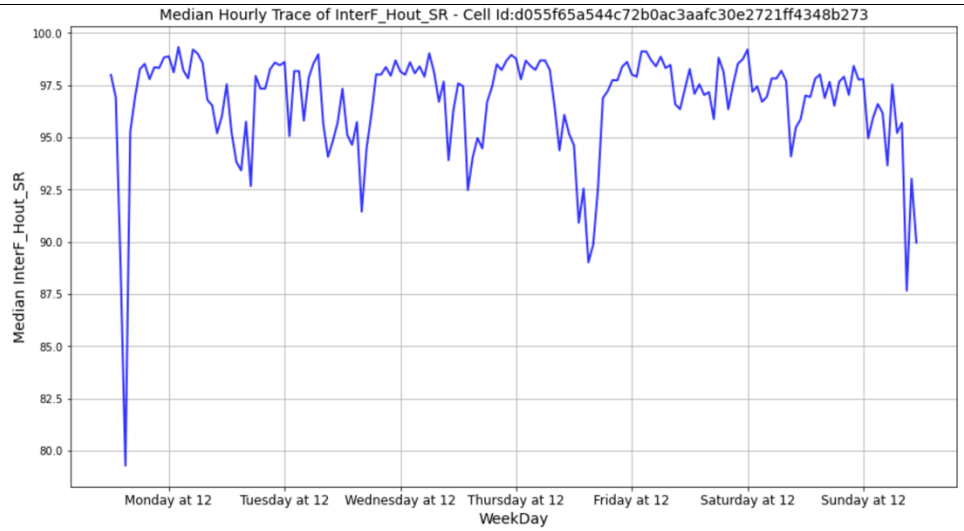


March

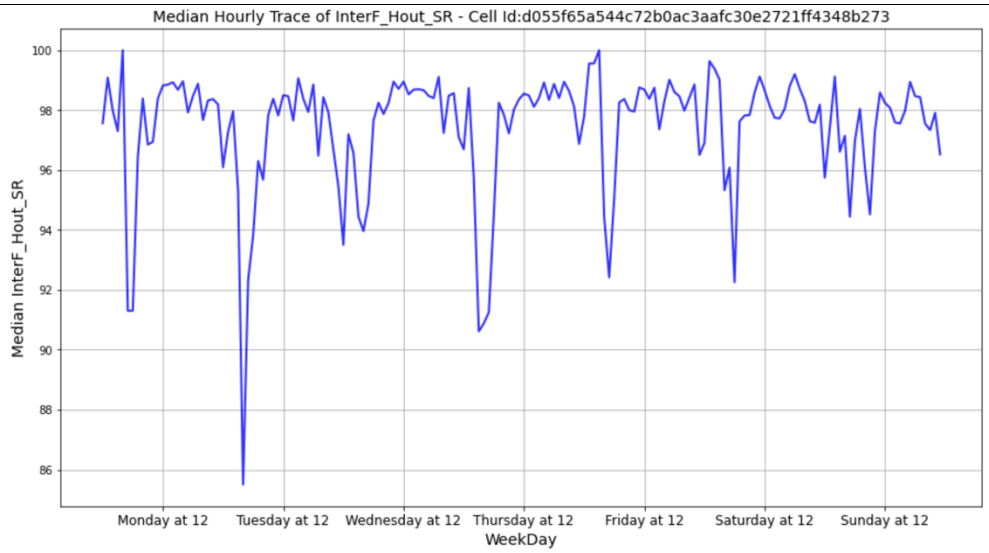
Table 3: CS_SR



January



February



March

Table 4: InterF_Hout_SR

- ***Non-residential/Business***

Considering a cell which covers a non-residential or business area (including both offices and Universities/schools) in normal conditions, we may expect higher downlink traffic volume values during the weekdays (Monday – Friday), while little to nothing activity during the weekend, since those structures should be closed.

In light of the outbreak of the disease and the "stay at home" order, it's reasonable to think that there won't be too much traffic activity in the considered area, resulting in a graph with lower values compared to the ones awaited for the first two months of the year.

In Table 5, we reported the weekly median signatures of "DL_VOL" obtained for each month for the chosen cell.

The trends of January and February (in which, the same consideration from the "Residential" section applies) strongly confirm our expectations, showing an activity from Monday to Friday (peaks comprised between 2 and 3) and less intensity during the weekend (around 0.5 and 1). Moreover, as it is shown in the March graph, our suppositions were found correct again: not only the traffic is nearly evenly distributed during the whole week, without any major difference between weekdays and weekend, but also the values are dramatically reduced, reaching 1.2 as the highest value.

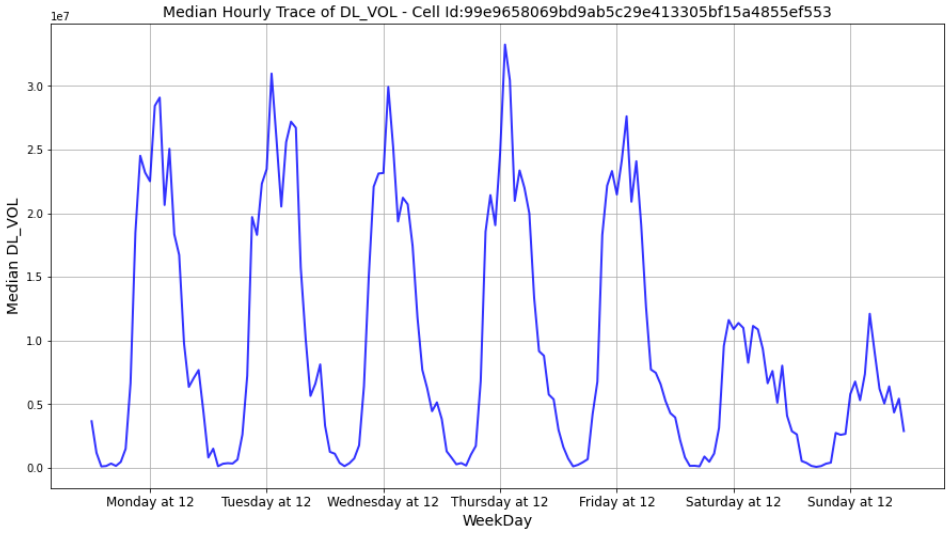
A common element one may observed is that the daily pinnacles are always around 12:00 PM, a trend that stays consistent also in the "UP_VOL" and "CS_SR". It is likely due to the office hours: we see a rapid increase in the morning times and a steady decline in the afternoons.

Moving on to Table 6 for the uplink plots, again, some of the already discussed aspects for the residential area cell are noticeable as well. The main trait to be pointed out is that March shows curves under a maximum threshold of 2.5, whilst before the apices were around 5: so that would be a decrease of exactly 50% of the peaks. Furthermore, the behavior of said month presents itself to be less regular and to have more spikes throughout the week, in comparison to the January and February ones.

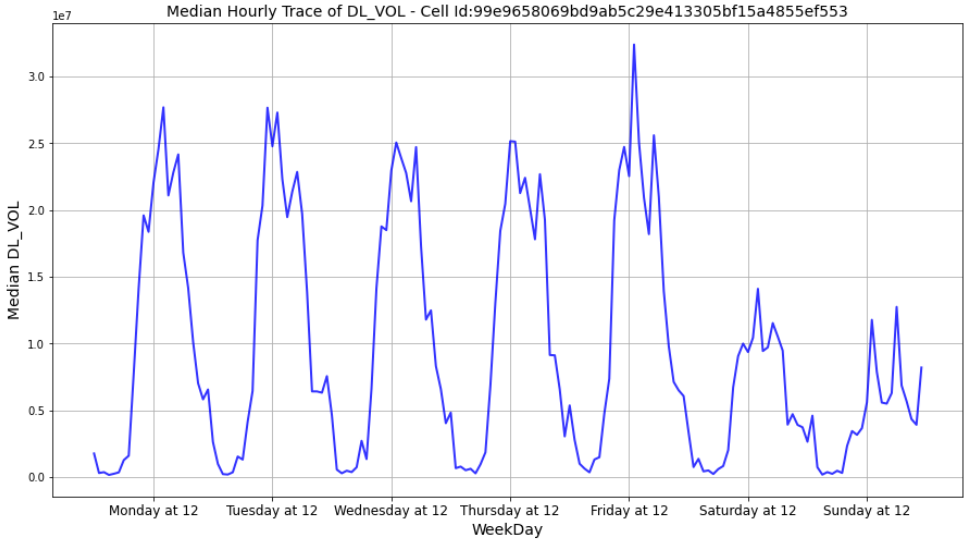
In Table 7, the "CS_SR" weekly median signatures for the interested cell are available. The trends look similar to the ones noticed for the parameters above, although the intensities have much smaller values, with the highest being around 50000 for January and February and 8000 for March: the latter would be 6,25 times lower between the "Covid Free" and lockdown observation periods.

In Table 8, the outgoing handover KPI signatures of the business cell are shown. In respect on what has been seen for the residential cell, we can state that the graphs have a tendency to stay at the 100% success rate for longer periods, rather than just rapidly going up and down around said mark. Additionally, no big changes are detectable throughout the months, but we could speculate on two details:

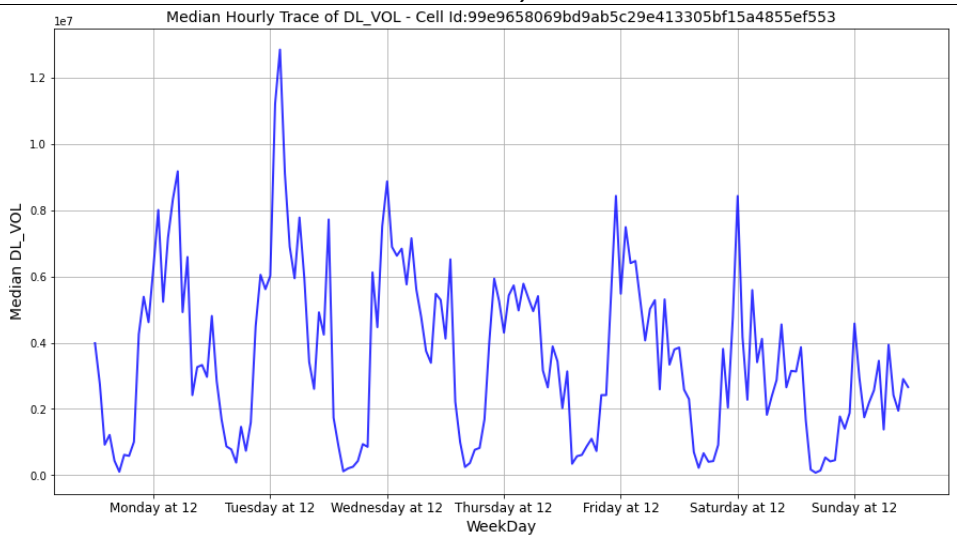
- One may say the January plot almost has a "periodic" behavior over the week, more or less repeating the same pattern every day. On the other hand, from the February and March graphs the trends start to seem irregular, with recurrent deep depressions, principally during night hours.
- In March, the graph seems to have wider and more "constant" apices, especially on Thursday nights and Sunday mornings, meaning that, once the curve goes up to the maximum threshold, it stays there for a prolonged time.



January

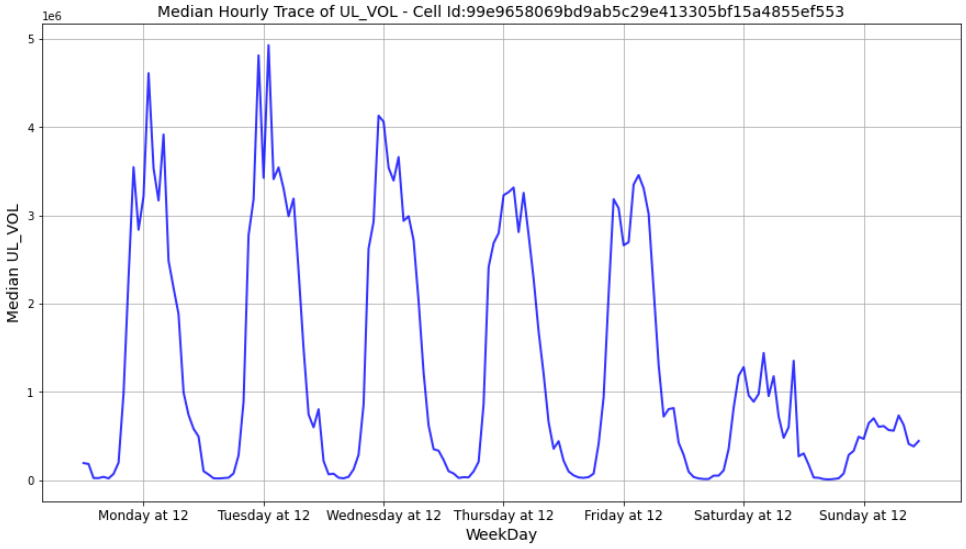


February

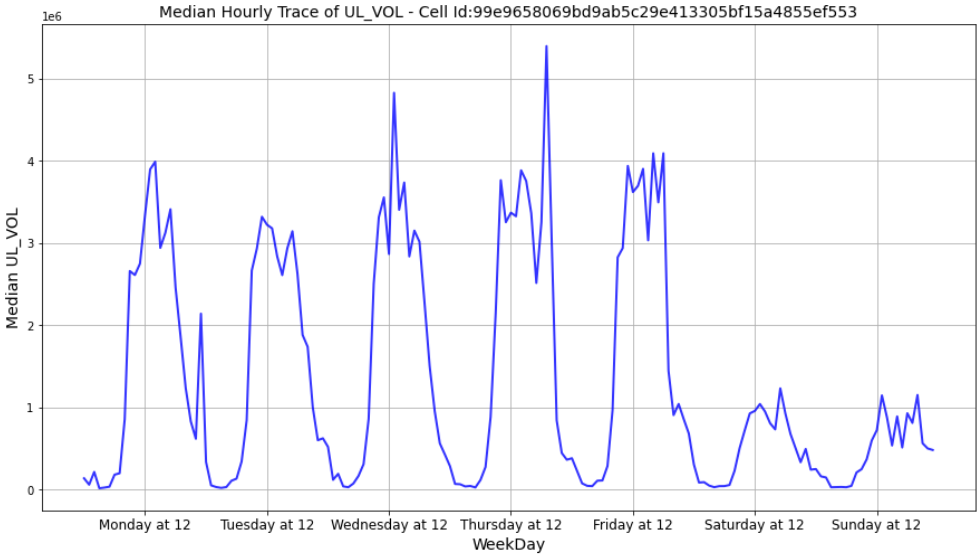


March

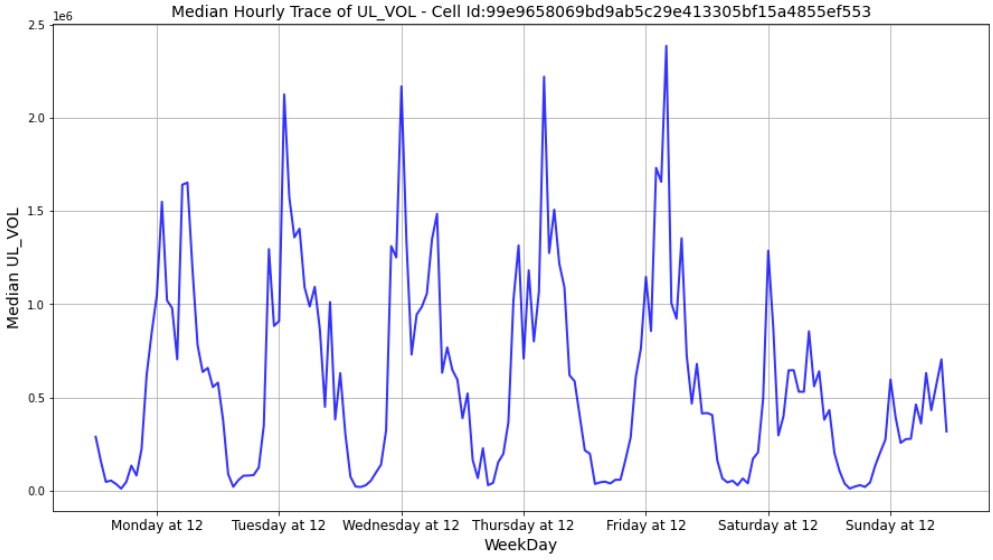
Table 5: DL_VOL



January

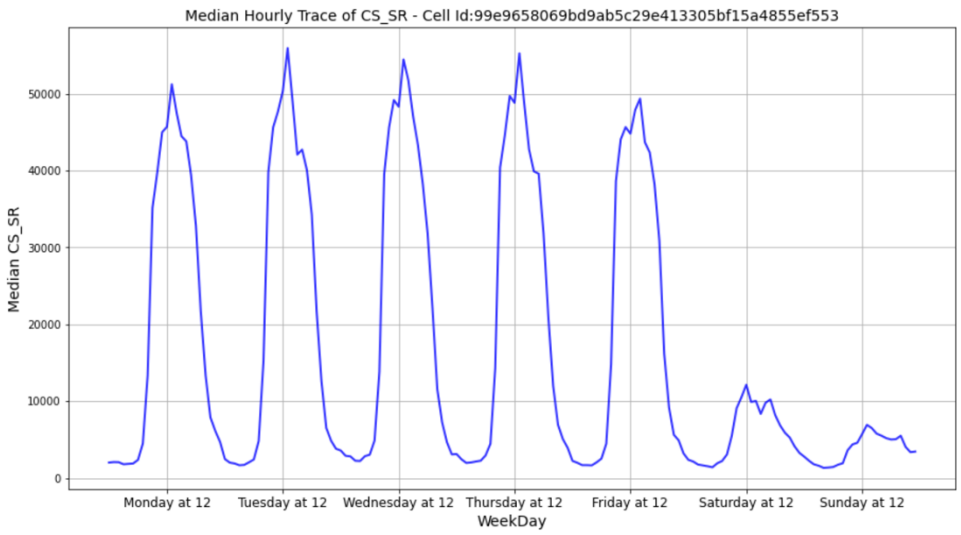


February

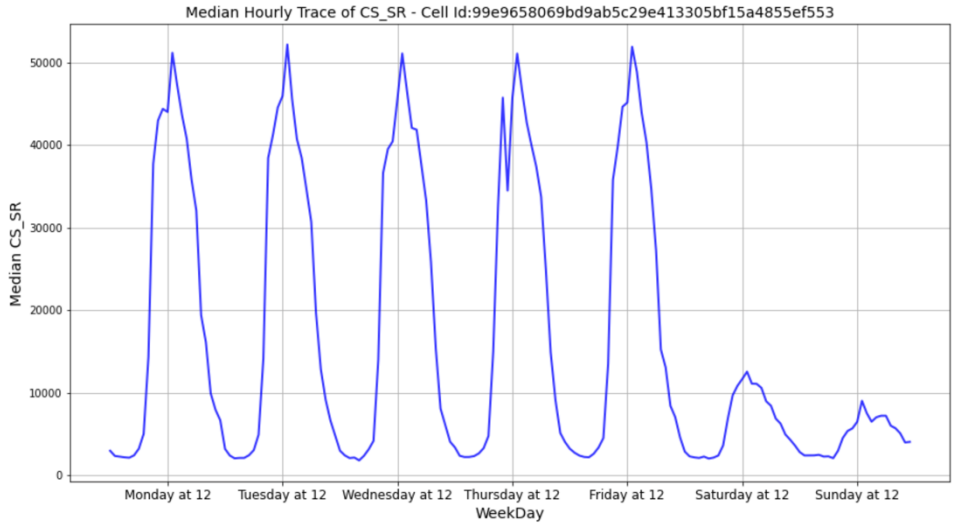


March

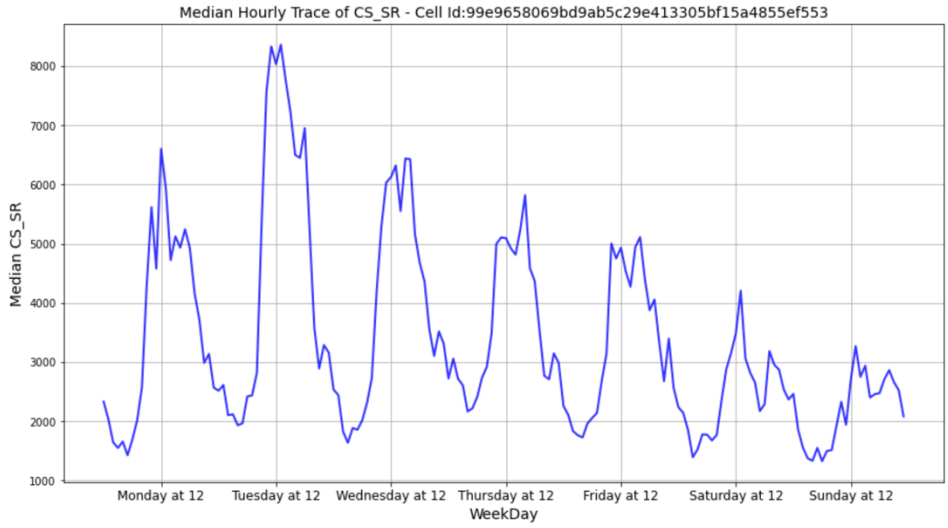
Table 6: UL_VOL



January

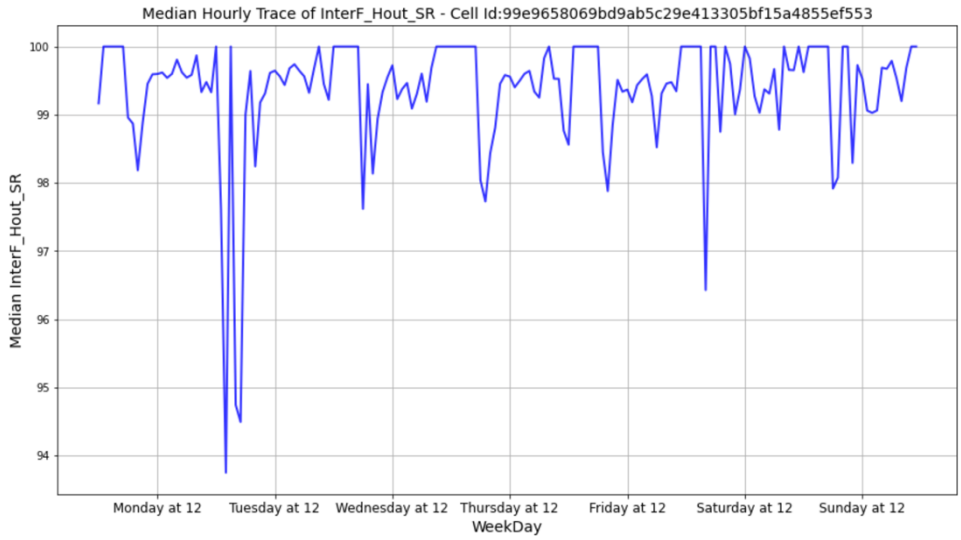


February

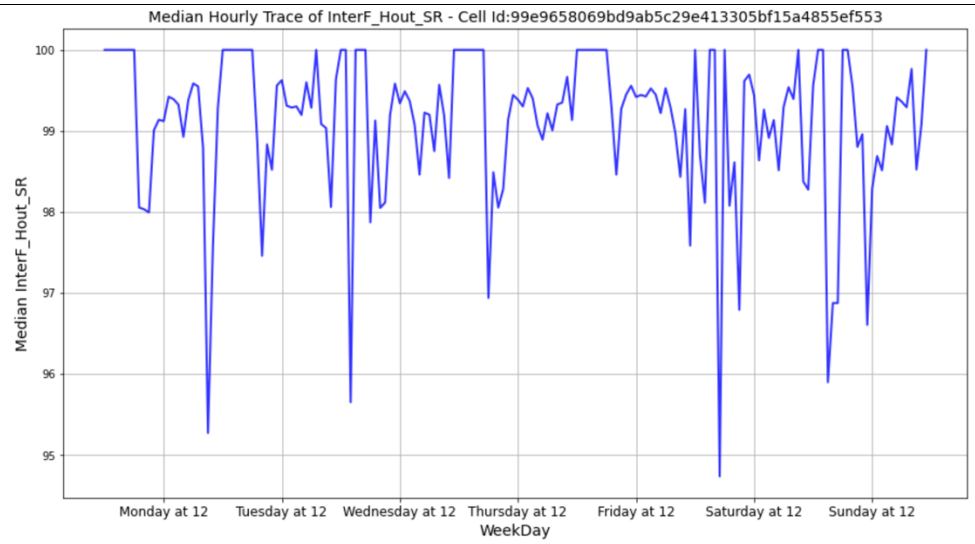


March

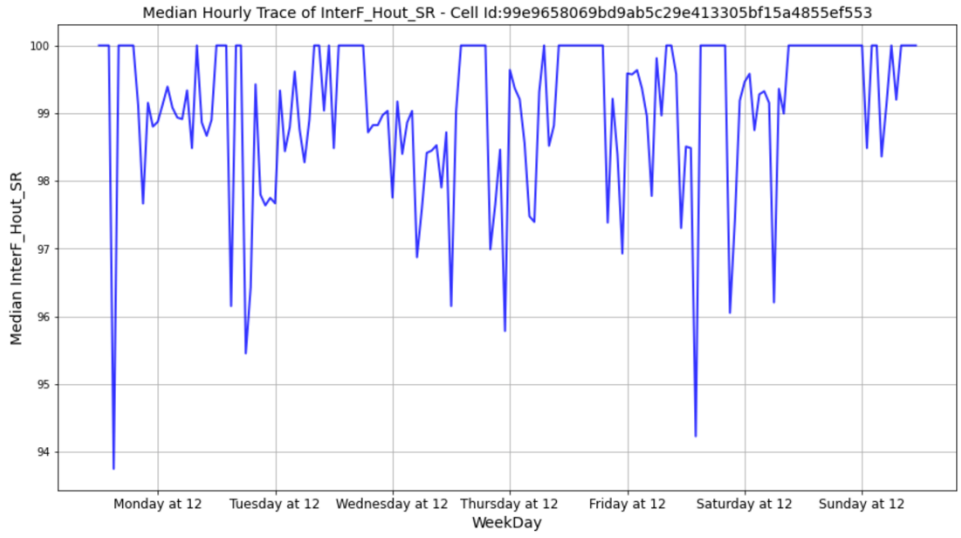
Table 7: CS_SR



January



February



March

Table 8: InterF_Hout_SR

- ***Transportation***

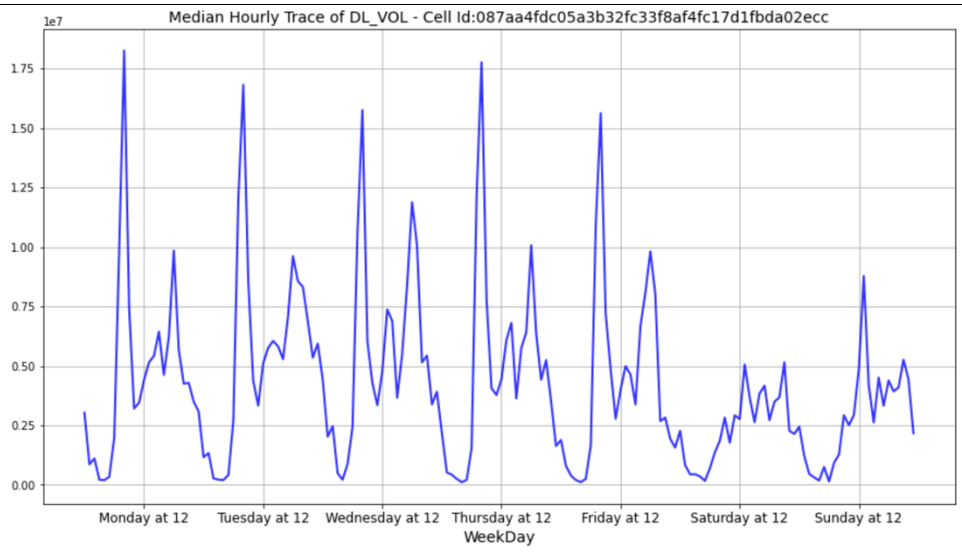
Considering a cell which covers a transportation area, we could expect both a trend similar to the one of: the non-residential cells, respectively a certain daily trend for the weekdays and a lower one for the weekend, and a steady behavior throughout: this depending on the business and the location of the transportation area. However, we proceeded our analysis with the former. The main difference between the two type of cells should be found in the hour in which peaks are present: while generally there would be one major peak hour for a work/study scenario, in the case of means of transport, we expect multiples ones, related to the times people go either to their job or to school.

This awaited situation is indeed depicted in the graph related to January and February (see Table 9), with the highest daily peak in the morning and some lower ones during the rest of the day, especially in the evening. During the weekend, as expected and understood from the previous cases, the plots' values are very low and the activity drops of 72,4%. The same can be said about the median weekly signature of "UL_VOL" in Table 10: where the considered parameter appears very similar to the one just described.

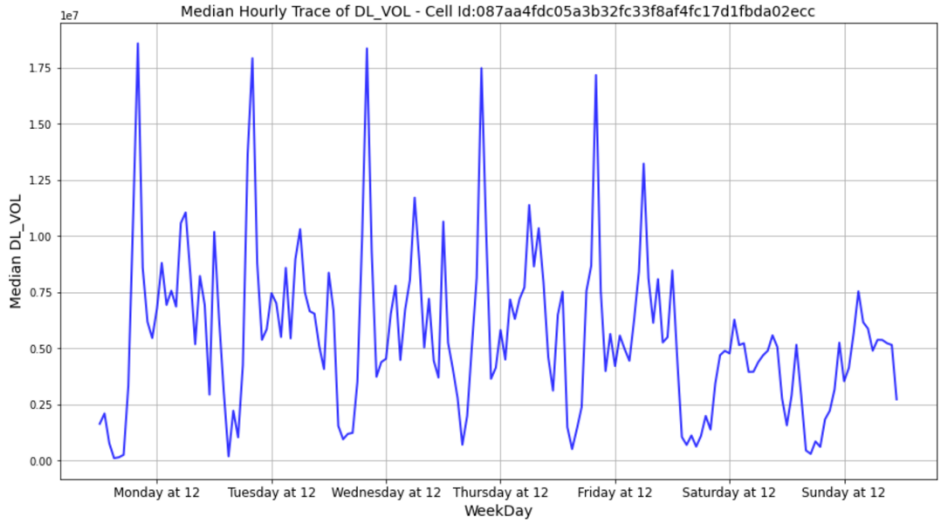
Regardless, in both tables, the March's graph presents a strange behavior: showing even higher peaks in the morning than the ones recorded in the previous two months (going up to almost twice the value), while always keeping the rest of the day around 0,5. According to local newspapers, it seems that in this period of time, even though the lockdown was happening, the public transportations, whose rides were limited, have stayed completely full, due to the individuals that weren't or maybe even couldn't work remotely and were physically going to their job place: that could actually justify the results we attained.

In Table 11, we have reported the "CS_SR" weekly median signatures for the selected transportation cell. Coherently with what we have seen so far with the other cells, the graphs' shape are similar to the ones of the "DL_VOL" and "UL_VOL" attributes. January and February are characterized by peaks comprised between 6000-8000 for the weekdays and around 4000-5000 for the weekend; in March, there isn't the same recognizable trend as the above one, and the weekdays' values have experienced a decrease of around 25% and are now comparable with the weekend ones, still around 4000-5000.

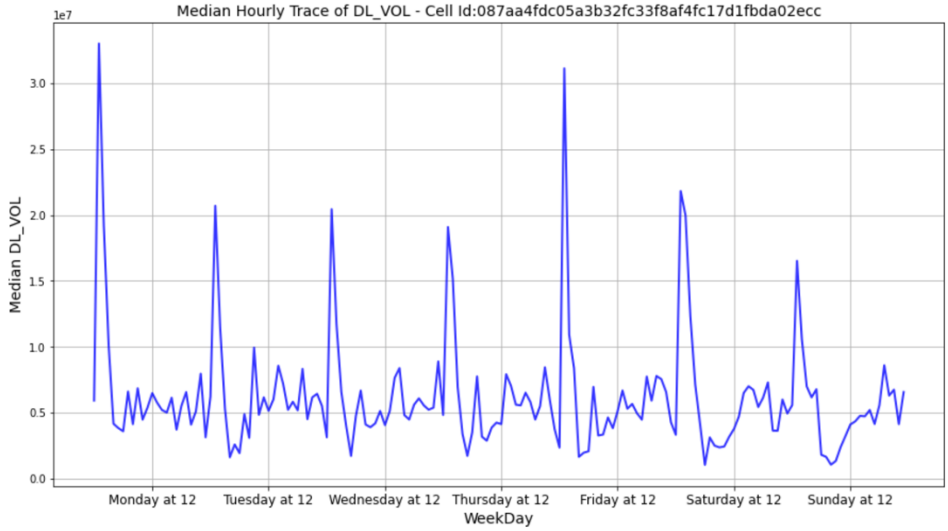
The last parameter that will be commented is the "InterF_Hout_SR" KPI, represented in Table 12. Compatibly with what we have seen so far, there's a general tendency to stay near 100%, especially in February, where there is only one noticeable negative peak and the rest of the curve is between the maximum threshold and 98%. For this type of cell, the month of January looks to be the worst one in terms of successful handovers, maybe due to the fact that the capacity and the rides available were still at their regular numbers, therefore the cell being more congested. Finally, in all three months, it's easy to detect a daily trend, hence noticing that the higher values are registered mostly at night during the weekdays and there seems to be small ups and downs in the weekends.



January

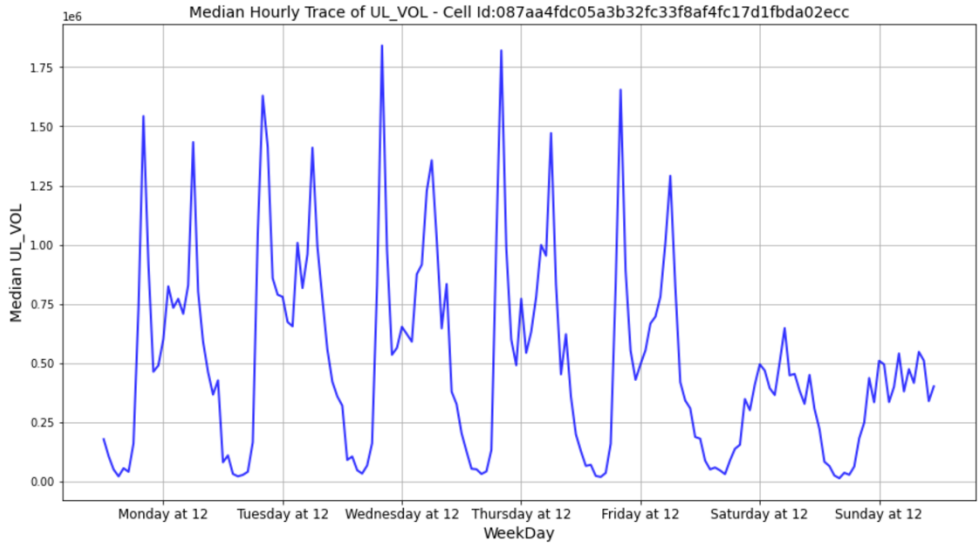


February

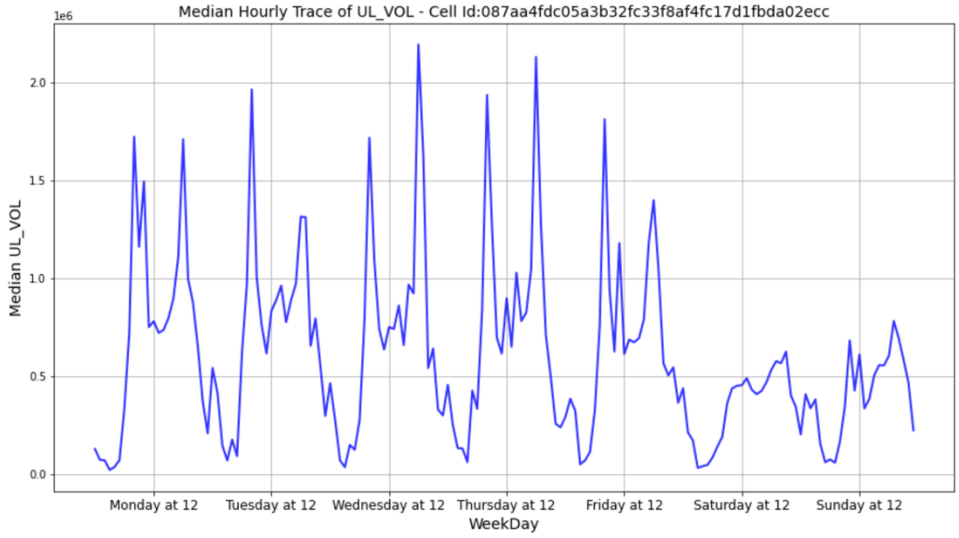


March

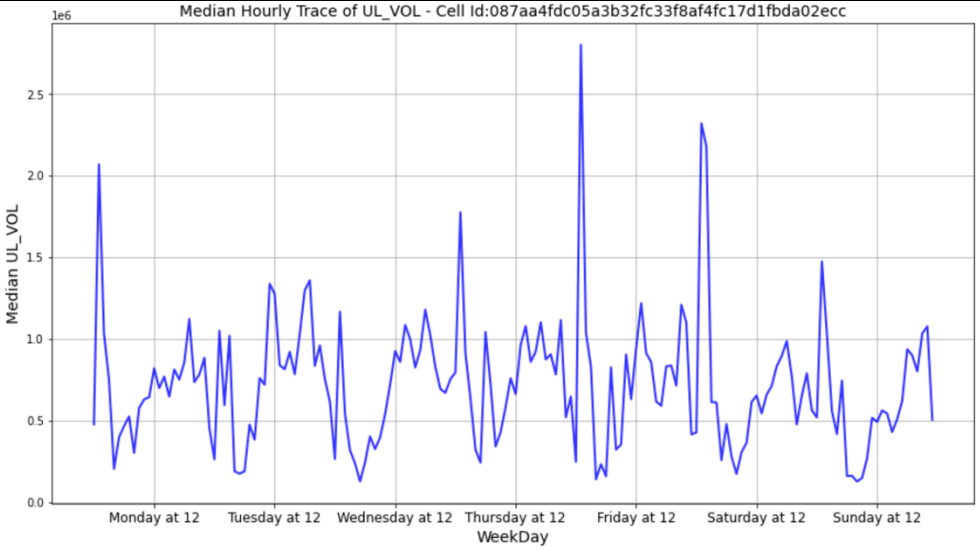
Table 9: DL_VOL



January

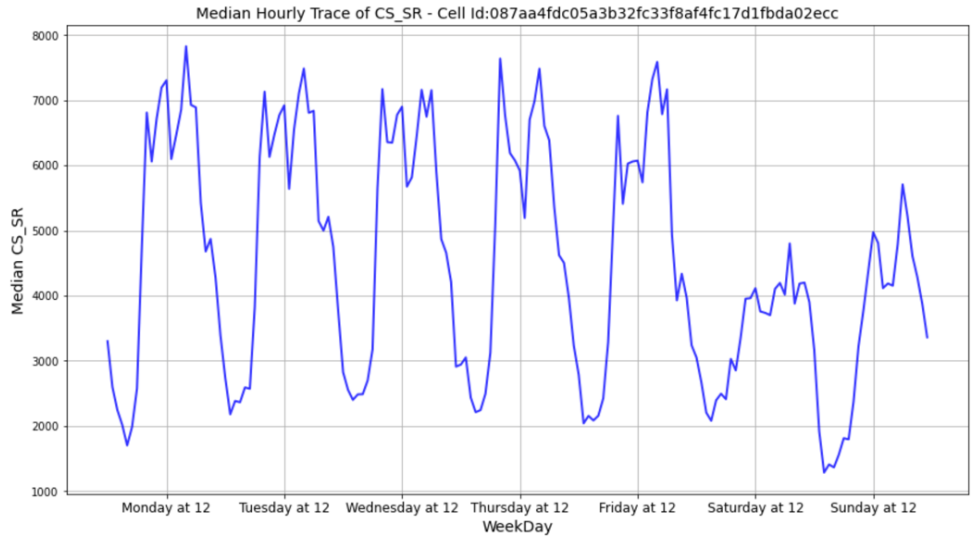


February

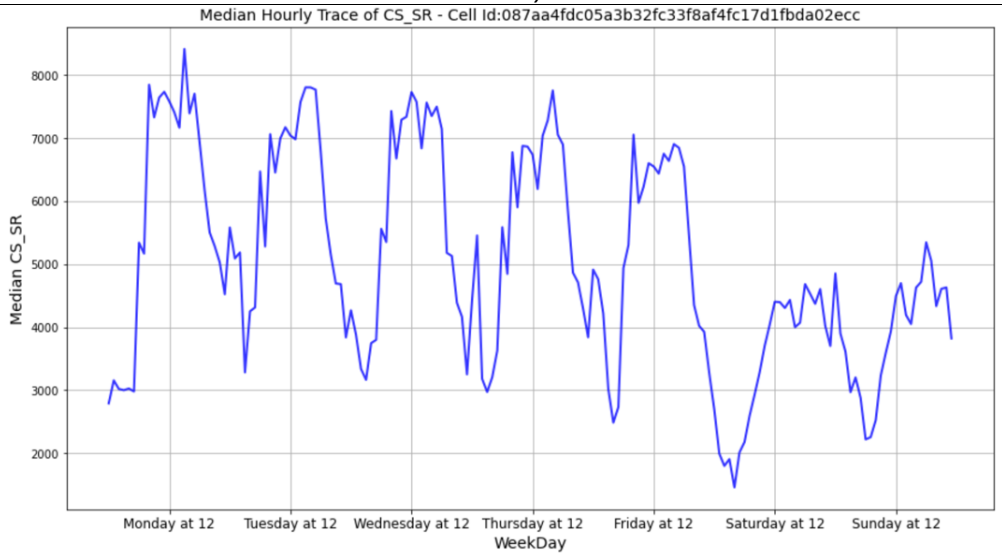


March

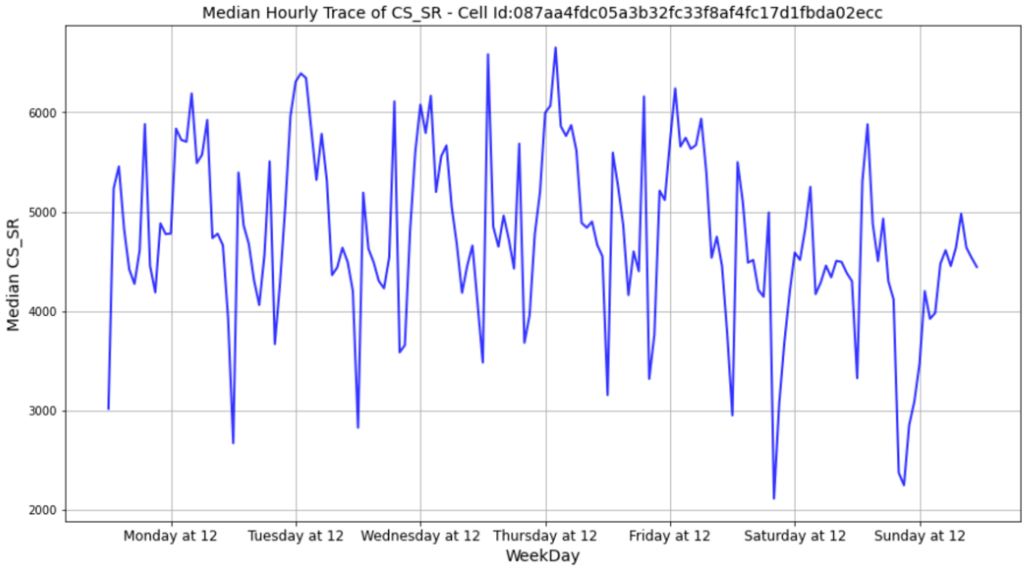
Table 10: UL_VOL



January

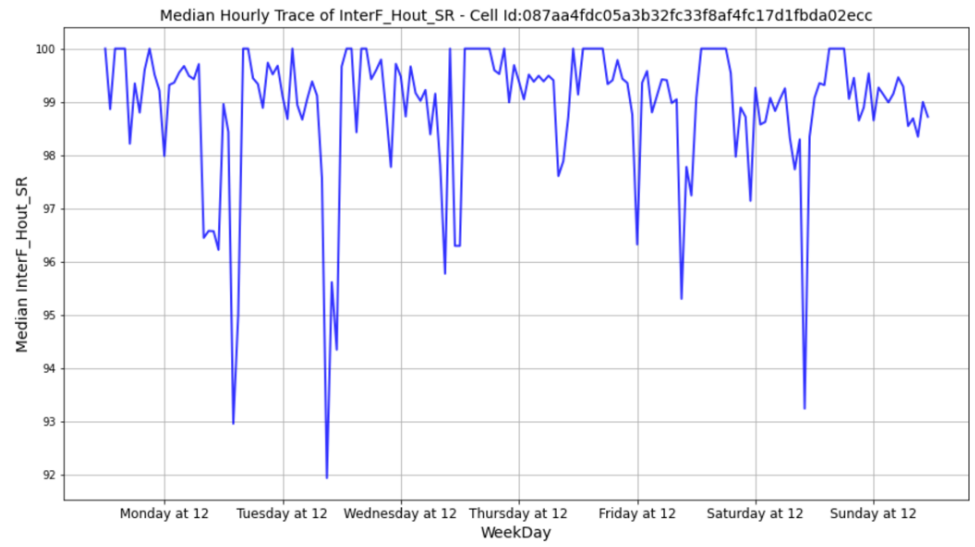


February

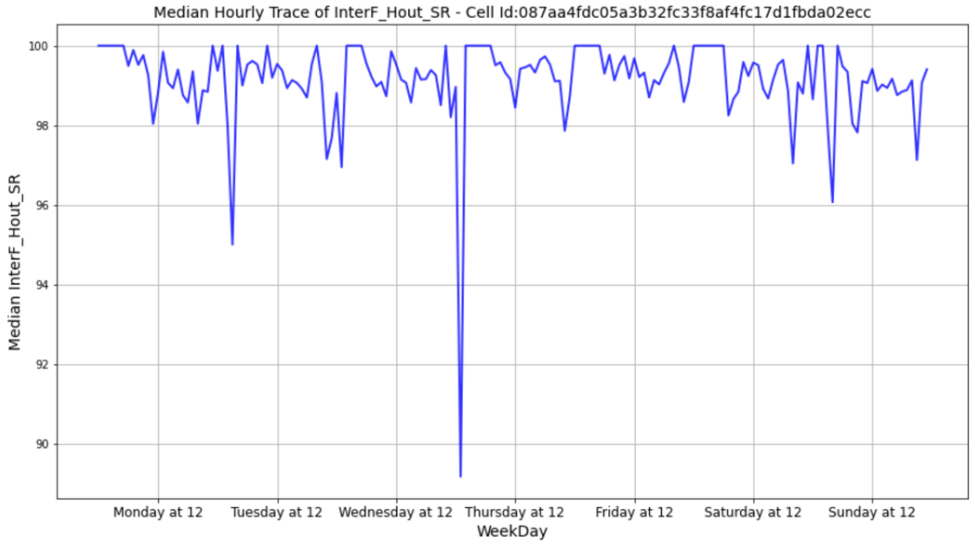


March

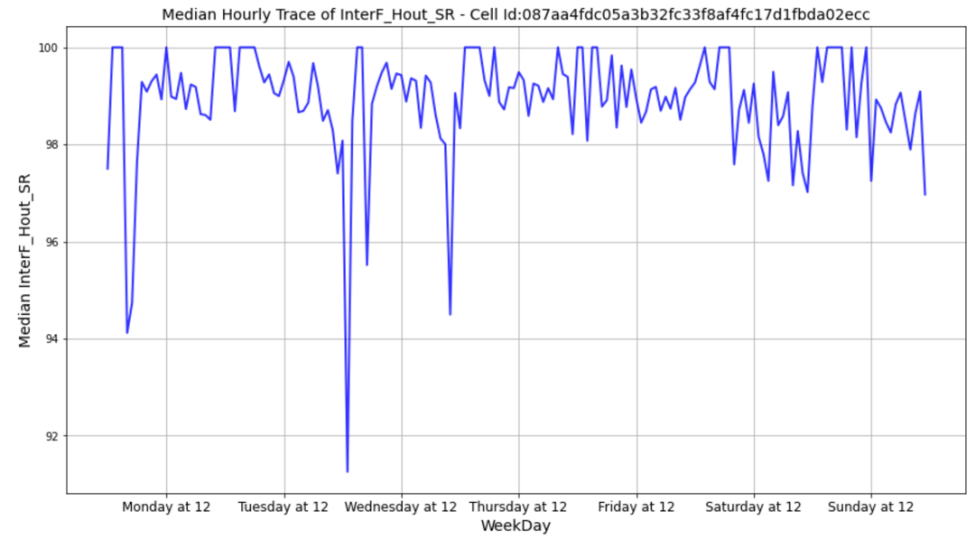
Table 11: CS_SR



January



February



March

Table 12: InterF_Hout_SR

6. *Comments and conclusion*

Aiming to confirm what we have stated in the previous paragraphs, we decided to look for locations of the cells we had selected, so to verify whether they were in the type of area the clustering process had assigned them to.

The complete addresses are reported below:

- The chosen residential cell is in Viale Sarca 61, Segnano, Municipio 9, Milano, Lombardia, 20125, which is the most densily populated administrative subdivision of the city.
- The business/non-residential one was found exactly in the center of Milan, precisely in Via San Giovanni sul Muro 17, Duomo, Municipio 1, Milano, Lombardia, 20121, an area where lots of stores are present.
- The trasportation one, was located at Cascina Gobba (M2) in Via Padova, Parco Lambro - Cimiano, Municipio 3, Milano, Lombardia, 20132, so in the upper north-east side, where the “green” underground line links the outer part of Milan with the centre.

Fig.6: The selected cells mapped.

In addition to our suppositions, this is a strong confirmation that the KMeans clustering process has worked fine and produced satisfactory results.

To conclude this project, we may say the analysis of data through various Key Parameter Indicators, has allowed us:

- To understand and observe how the mobile radio network in the metropolitan city of Milan behaves both in normal conditions and in particular ones, such as the current pandemic.
- To find strengths and weaknesses of a network and so to possibly prevent the consequences the changes in everyday lifestyle of the population cause on the usage, accessibility, and mobility of the traffic.

In the GitHub repository, you can find all the files and the codes we have used and discussed so far to complete our work.

https://github.com/chiaradraghini/COVID19_impact_MRN_project

Please Note: some of the files, due to their dimensions, are only available in their raw version, others were uploaded in this repository as ".zip" and can be visible if downloaded.