Study the impact of the lock-down due to COVID-19 pandemic on the 800 MHz frequency layer of an operator's LTE network in the city of Milan

Davide Andreotti, Giuseppe Gattulli

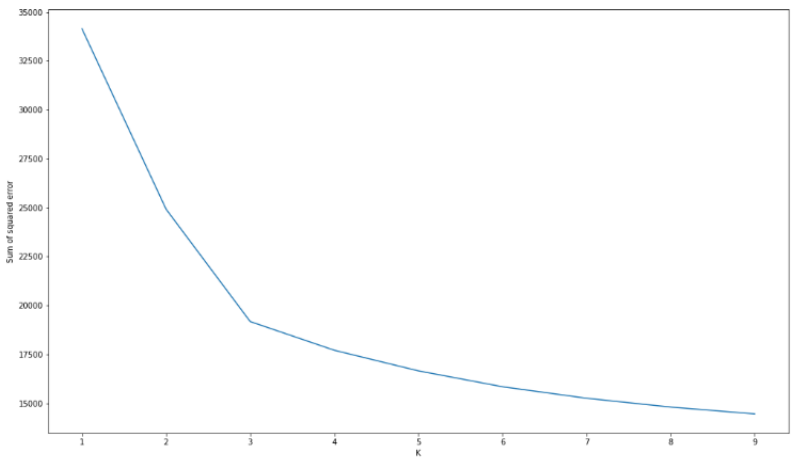
Contents

Clustering

The first thing we did was trying to understand how to do clustering: since it's known that human activities have weekly periodicity, the goal was to divide cells in several groups that had similar traffic behavior during the hours of the week.

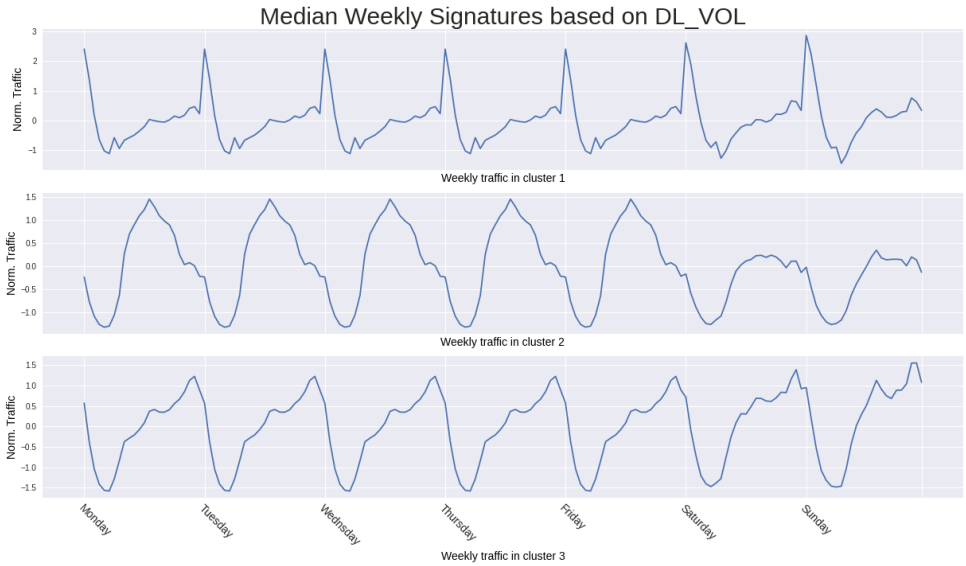
Starting from the Median Weekly Signature, we took the dataset with January data and focused on the DL_VOL KPI, which seemed to us the most useful for our objective; as explained in class and in the article [1], we divided the month in weekdays (analyzed all together) and weekends (separating Saturday and Sunday) and calculated the median of the KPI in each hour of each group. The result was a value for each of the 24 hours for the weekdays and the same for Saturday and Sunday.
Duplicating the weekdays medians over the whole five days and concatenating Saturday and Sunday medians gives us the Median Weekly Signature of a cell.

After doing this for each cell in the dataset we used a Standard Scaler (as done in article [2]) to normalize data over each cell, since some cells may have the same traffic behavior but different scales of values (in the city center the traffic volume could be higher than in the suburban areas).



Then, by using the k-means algorithm, after setting it up with an elbow plot (in the figure) on giving us three clusters, we had all the cells clustered.

The algorithm returns the coordinates of the three centroids, that in our case represent the Median Weekly Signatures of the three groups representatives:



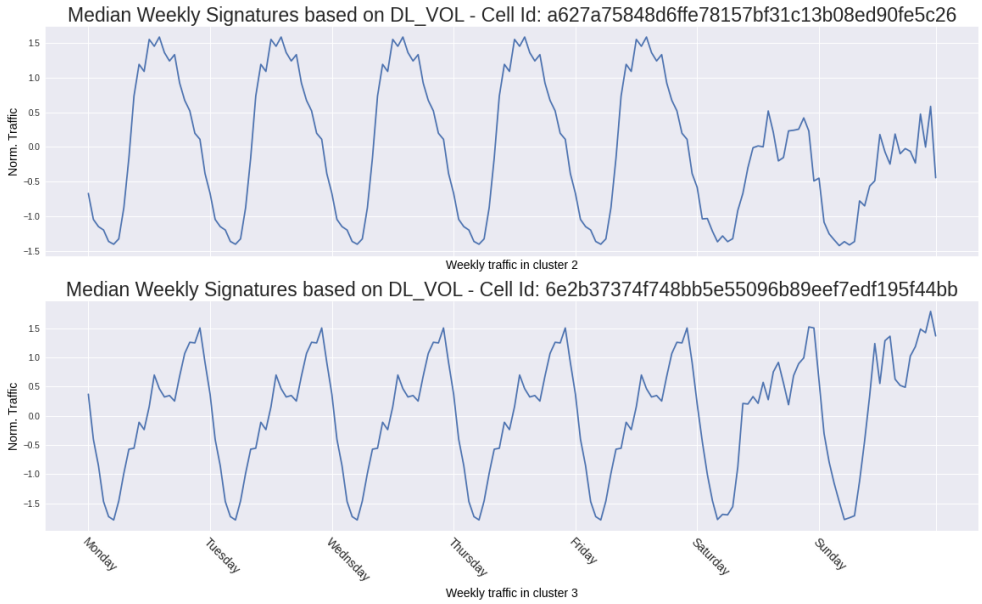
We can see that cluster 2 has a typical Business area behavior: grows during the working hours, decreases during the evening and is low on weekends, while cluster 3 has a Residential area behavior, staying low during the day and growing in the evening.

We are not certain about cluster 1: we expected to have a cluster with Transport behaviors, but this seems more similar to a residential area, in particular the unexpected result was that peak during midnight, that probably unbalanced all the data. We decided in the end to just neglect this cluster.

After the clustering procedure, we found that the cells were distributed like in the table:

	Number of cells	**Percentage**
Cluster 1 (neglected)	95	23.8%
Cluster 2 (Business)	117	29,4%
Cluster 3 (Residential)	186	46,7%
Total	398	100%

Next, we tried to find, for each of the two clusters, the cell with nearest coordinates (Median Weekly Signature) to the centroid, the result was this:

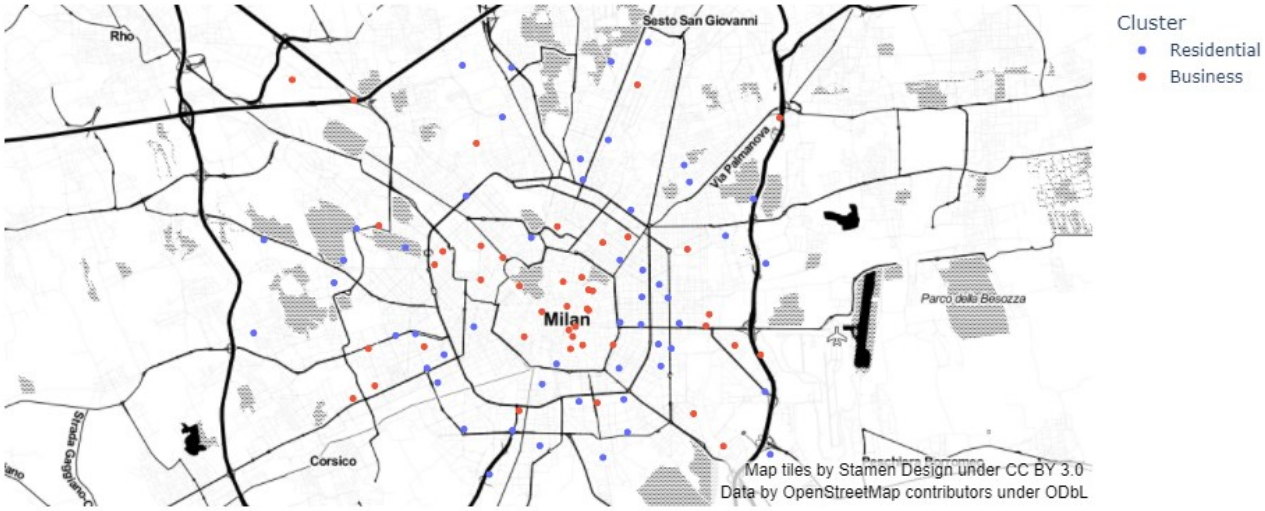


Which confirms the previously obtained trend.

Map

After this first analysis we wanted to localize each cluster on the map.
Since each location where an eNodeB is typically has about 3 cells, that would lead to a bad visualization, so we merged the chosen KPI values of each of the cells in a single eNodeB (since we are looking at a cumulative traffic volume, we decided to sum the values of the three cells).

Doing this and replicating the same clustering procedure previously explained the result is the following:



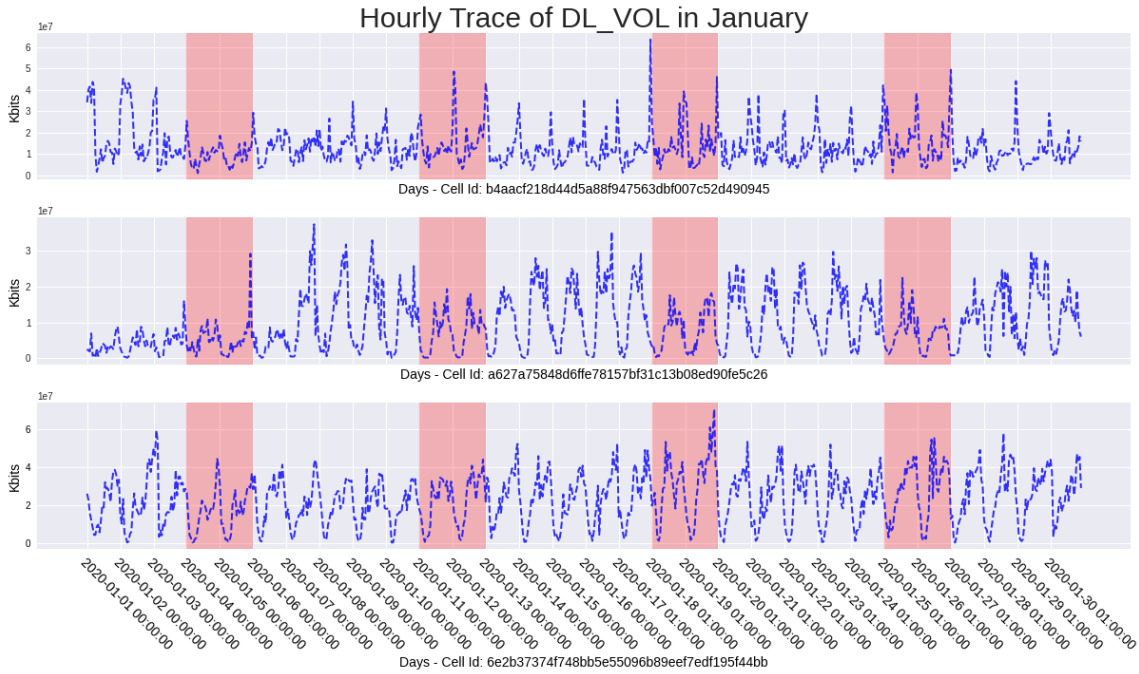
As expected, as we can clearly see, business clusters are concentrated mostly in the center of the the city, while residential areas are distributed around the center.

After having done the clustering procedure on eNodeB's instead of single cells we tried to look at the Median Weekly Signatures of centroids and cluster representatives, to see if the issues were

solving or the plots showed better results, but the trends and the clusters were more or less equal. For this reason we stayed on the first decision, thinking that we could find more precise results.

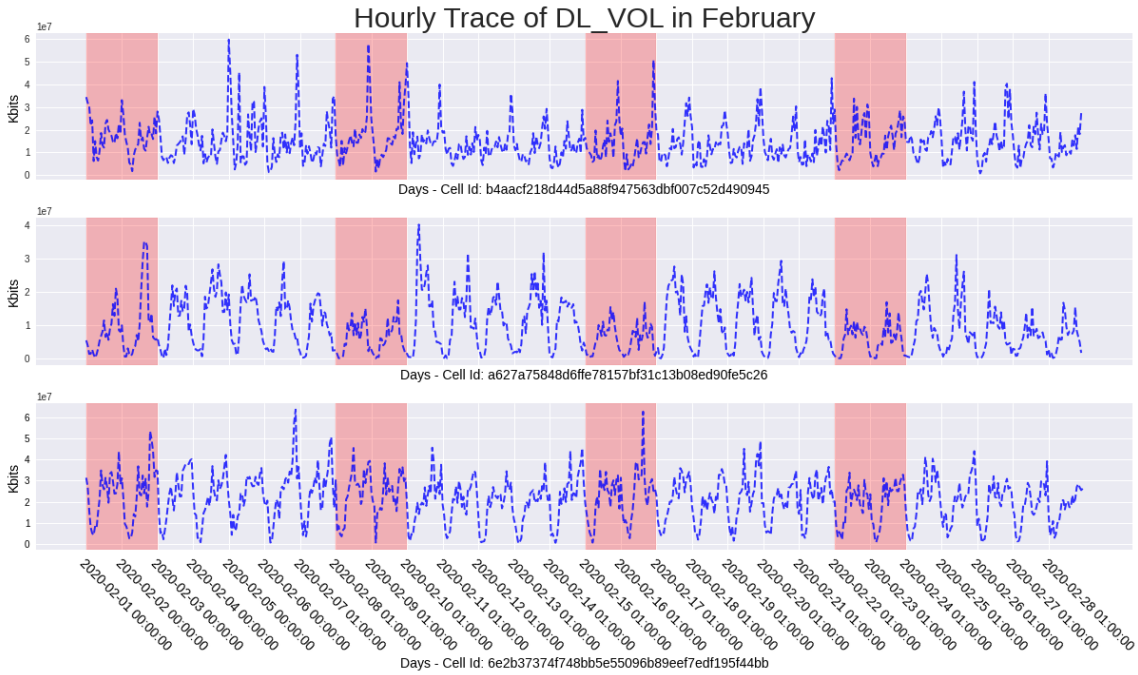
Monthly Trends

Here we show the comparison between the three clusters in the three months: we wanted to see how the monthly trend went on and in particular if the Median Weekly Signature behavior was respected. The three plots of each month represent respectively each of the three clusters in the correct order.



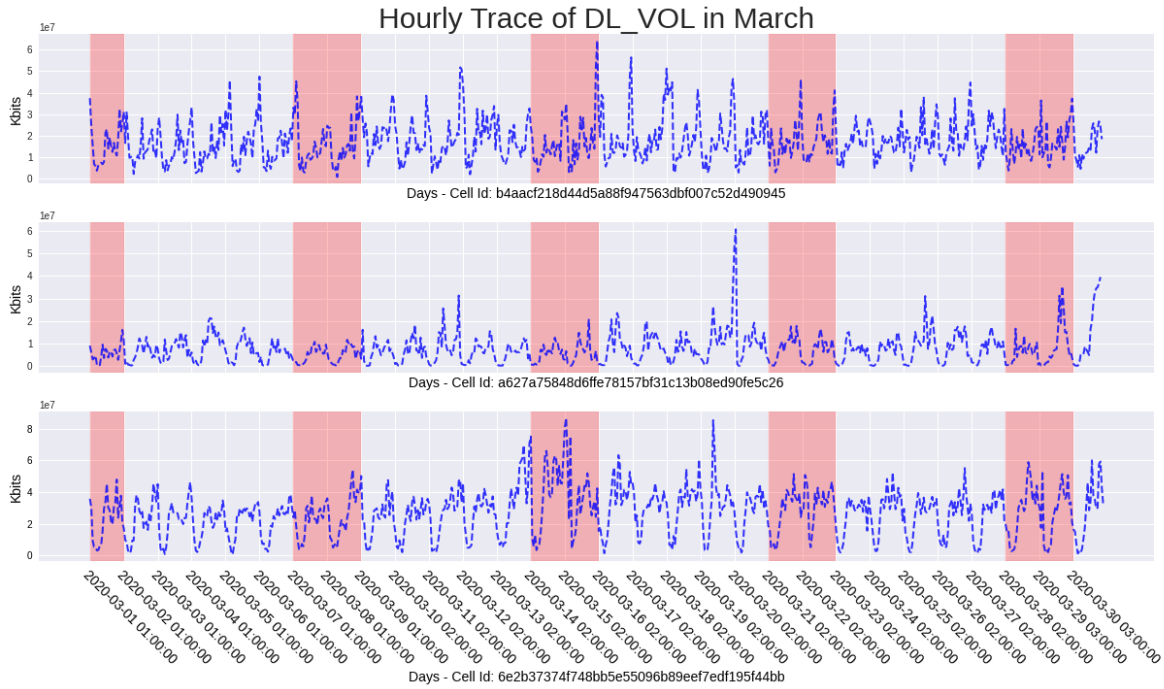
The red area indicates the weekends

In January we can see that the Business cluster has low traffic on the first week, because of the christmas holidays, and then goes back to normal, having high traffic on weekdays and low traffic on weekends.
Residential cluster is as expected, that is growing during evenings and the trend is the same over all the week.
The first cluster keeps having that peak as explained before.



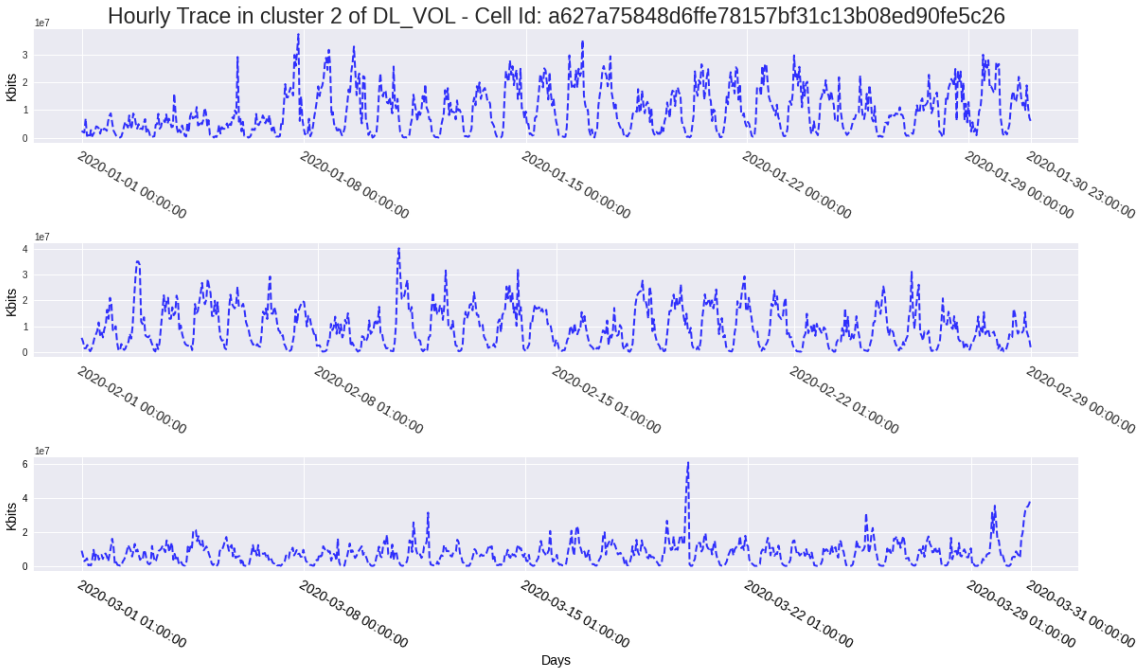
The red area indicates the weekends

In February the same trend goes on, until, in the end of the month, we can start seeing a slight decrease of traffic in the Business cells, probably because of the first restrictions being imposed in the Lombardy region.



The red area indicates the weekends

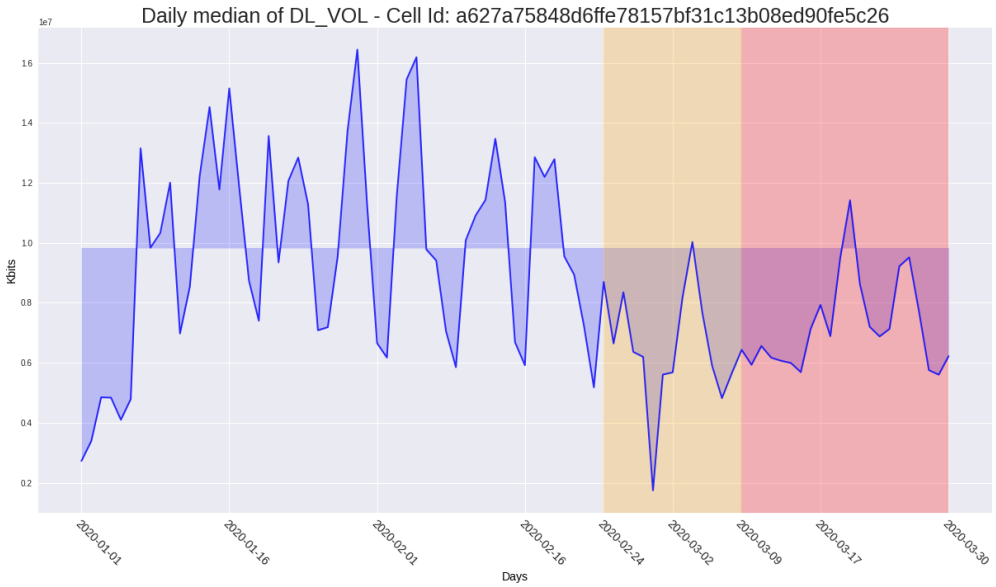
In March restrictions show their effect: Business cells have slightly lower traffic than the usual trend and Residential cells have a different behavior in most of the days: there is no more low traffic in the morning and a growth in the evening, but the traffic is almost constant during the daytime, sign that people stayed at home instead of going out for work.



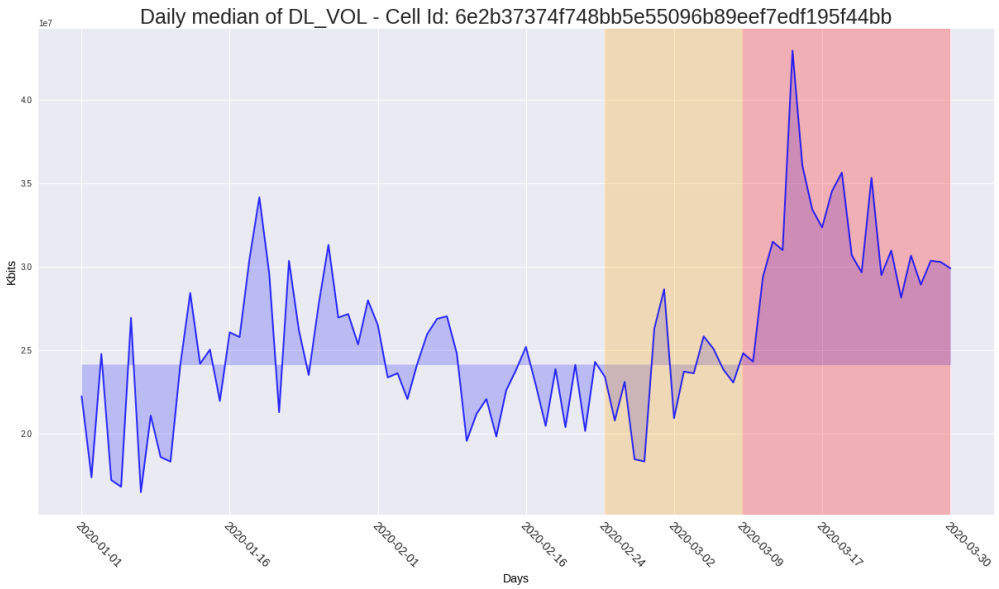
Considerations done previously can be seen more clearly on this last plot, which compares the three months of the Business cell analyzed.

Covid-19 effects

Last we wanted to show better the increase or decrease of traffic in the two clusters during the lockdown: the yellow area represents the start of the Covid-19 restrictions in Milan while the red area represents the strict restrictions in Italy. The blue line connects the median of DL_VOL of each day of the three months, while the blue area tells if the DL_VOL value of a day is above or below the median of DL_VOL value calculated between January 1st and February 23rd.

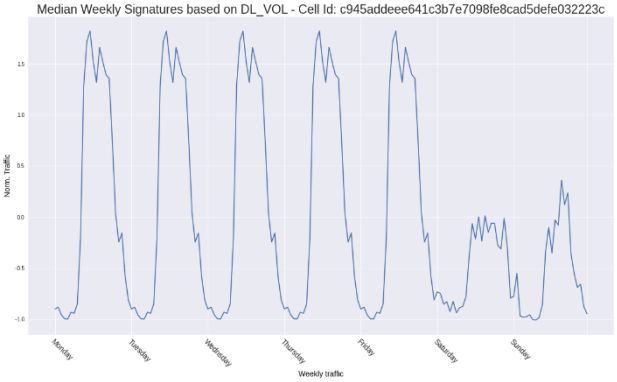


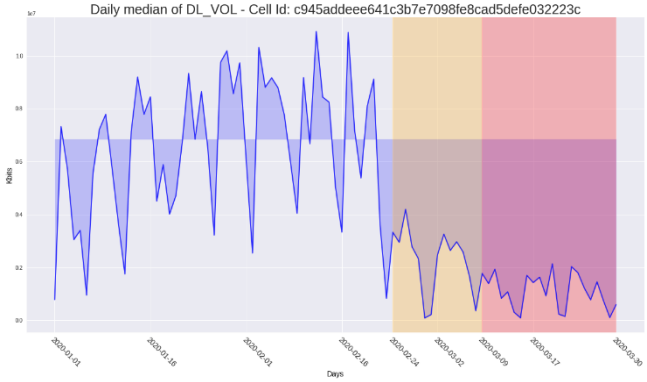
The plot above refers to the Business cell considered, it can be seen that from February 23rd on the daily traffic has a decreasing trend from the median. Also the first days of January show a negative trend, because of Christmas holidays.



The plot above represents instead the Residential cluster cell we are considering. We can see that during the national lockdown period an increase of traffic is present.

As a double check we also analyzed the cell presented in class from the professor as an example of Business cell in the city center:





On the left there is its Median Weekly Signature, that is coherent with all the results already found and on the right we show the traffic change during the three months period. Here it's even more visible that from the start of the restrictions in the end of February the traffic dramatically decreases in the cell.

Conclusion

In the end we found more or less the results we expected, based on what we have seen in class (about clusters shape) and intuition (about the traffic increase/decrease in March).

We worked for a long time to fix the issue related to the cluster with strange behavior but we couldn't manage to find a correspondence with real life as well as understanding what in our analysis was shaping data like that, but in the end we remained satisfied with the other two clusters, that seem to show the expected results.

Contributions

Davide Andreotti	Code review and development, plot review, report
Giuseppe Gattulli	Initial code draft, code review, plots, report

References

[1]: A. Pimpinella, F. Di Giusto, A. E. C. Redondi, L. Venturiniy, A. Pavon, "*Forecasting Busy-Hour Downlink Traffic in Cellular Networks"*

[2]: A. Furno, M. Fiore, R. Stanica, C. Ziemlicki, Z. Smoreda, "*A Tale of Ten Cities: Characterizing Signatures of Mobile Traffic in Urban Areas*"