1. To find the most congested time periods, I aggregated all the data values for each time period. Using these aggregated values for the data fields it is possible to plot them on a graph. To rank them I used a custom point system giving the top 5 ranking congestions points from 5 to 1. Maybe it would have been smarter to just use the total mass of the values but this is how I ended up doing it. For calculating the total value, I ended up normalizing the values by dividing them by their means, since on first glance it seemed that internet usage was 10 times larger than the mobile usage. I don't really know if this was a good idea, but when doing it without this normalization the graphs seemed to look quite similar. It at least made it quite difficult to compare the total mass so maybe it should not have been done.
2. To find the most called provinces, I just aggregated the provinces by their counts and plotted the values on a histogram plot.
3. To get the languages, I aggregated by user and created a field which contained an array of the distinct languages they had used. This allowed for me to easily count the number of instances the language was used by a user and to get both the top 5 and the number of Finnish tweeters.
4. To get the three day span distributions plotted, I combined them and plotted them on one graph. In terms of total volume of data communications we can notice that during christmas it is considerably lower. The number of mobile activity seems to spike on christmas day during midday, and on 26th it is considerably lower than normal, I guess because people are returning home or some such.
5. For the weather I tried to look at multiple weather stations and ended up plotting some difference between rainy days and not rainy days. There weren't that many rainy days in general to choose from and one of the good rainy days was December 26th which didn't seem like a good thing to try to compare. So in the end I picked two kind of rainy days and two non rainy days which were the same weekdays. One of the two rainy days seems to contain much more activity but for the other the activity is pretty much the same. I believe that is the day where most of the rain happened during the night so it is not such a surprise to see such results.
6. For the heatmaps, I ended up splitting the data into day and night time intervals. With these intervals I summed up all the data that there was. After all the data was aggregated, I transformed it into a matrix that resembles the actual location coordinates that were specified in the documentation. With this it was possible to draw a heatmap by interpreting the matrix as an image. The heatmap for Milan clearly shows that during the day most of the data communication happens in the center of the city as it would make sense. This can be seen when you look at the real map of Milan and compare the heatmaps. During the nighttime the hotspots are more spread out as people are home and not at work as it would make sense. For Trentino, we can see Trento and Rovento the two larger cities in this area and the region between them quite lit up with communications. In between the two larger clusters there is a mountain range where there is no real communication happening as it would make sense. In the day and night comparison we can see that many people commute to Trento from Rovento and at night both cities and their surrounding areas are more evenly lit up pretty much as expected.

The datasets did not contain data for all points, so there had to be some manual imputation for points that contained no data. Especially for the Trentino area, there were thousands of squares without any data.

7. I joined two datasets, for the temperature and CO in the air and it showed that when temperature is lower, there is more CO in the air. This would make sense as when it gets cold, people need to heat their houses and likely will prefer cars over walking and producing CO. Similarly, as wind speed rises, the CO in the air is reduced. After a point however, there are diminishing returns.