# Capstone Project 1 – Data Story

## Top 5 & Bottom 5 Grids

totalGridActivity dataframe has the total volumes of SMS-In, SMS-Out, Call-In, Call-Out, Internet Activity for each Grid over the 2 months. Tables below shows the top 5 and least 5 Grid Ids with respect to their volumes

Fig 1: Top 5 Grids

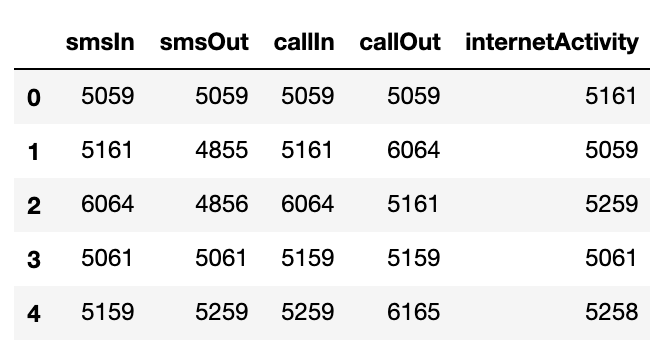
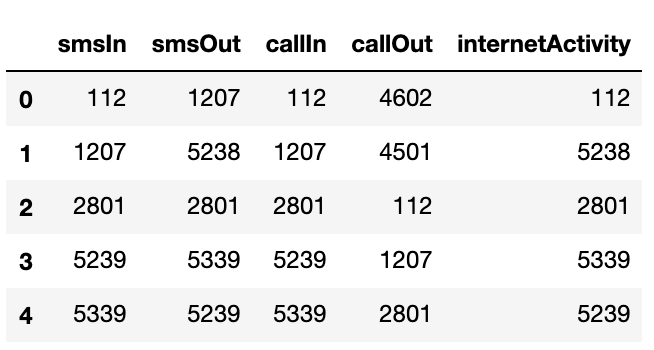
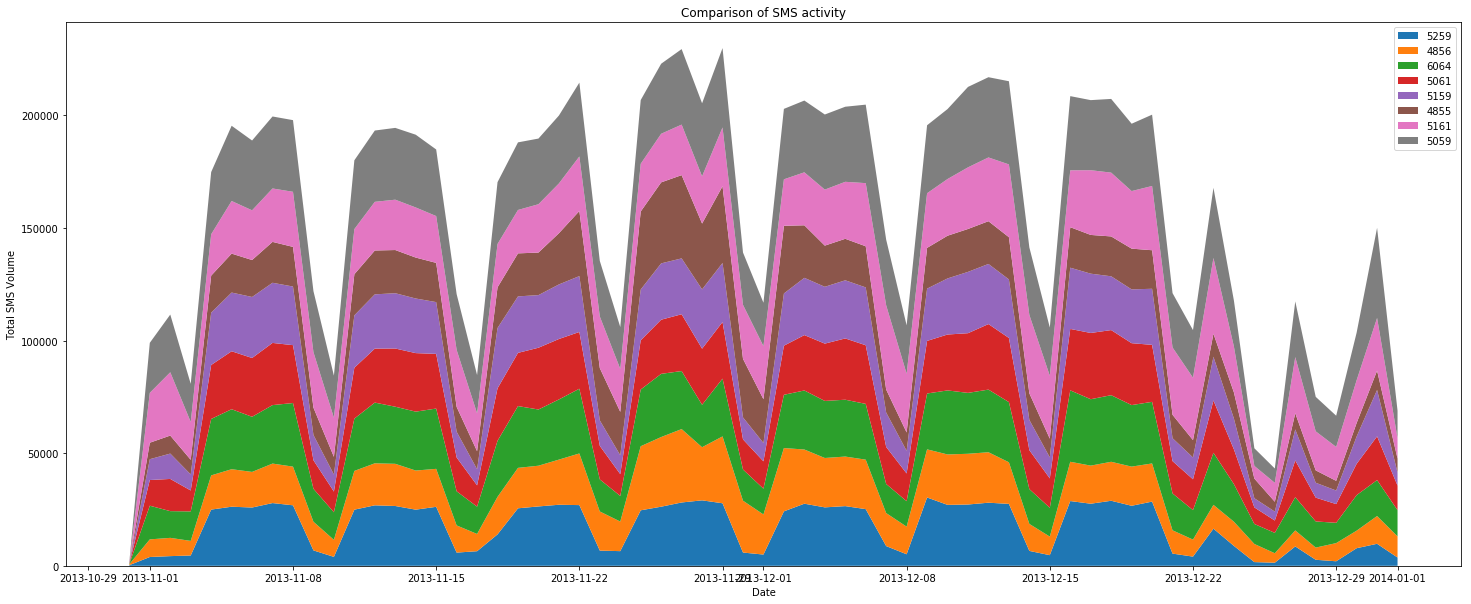


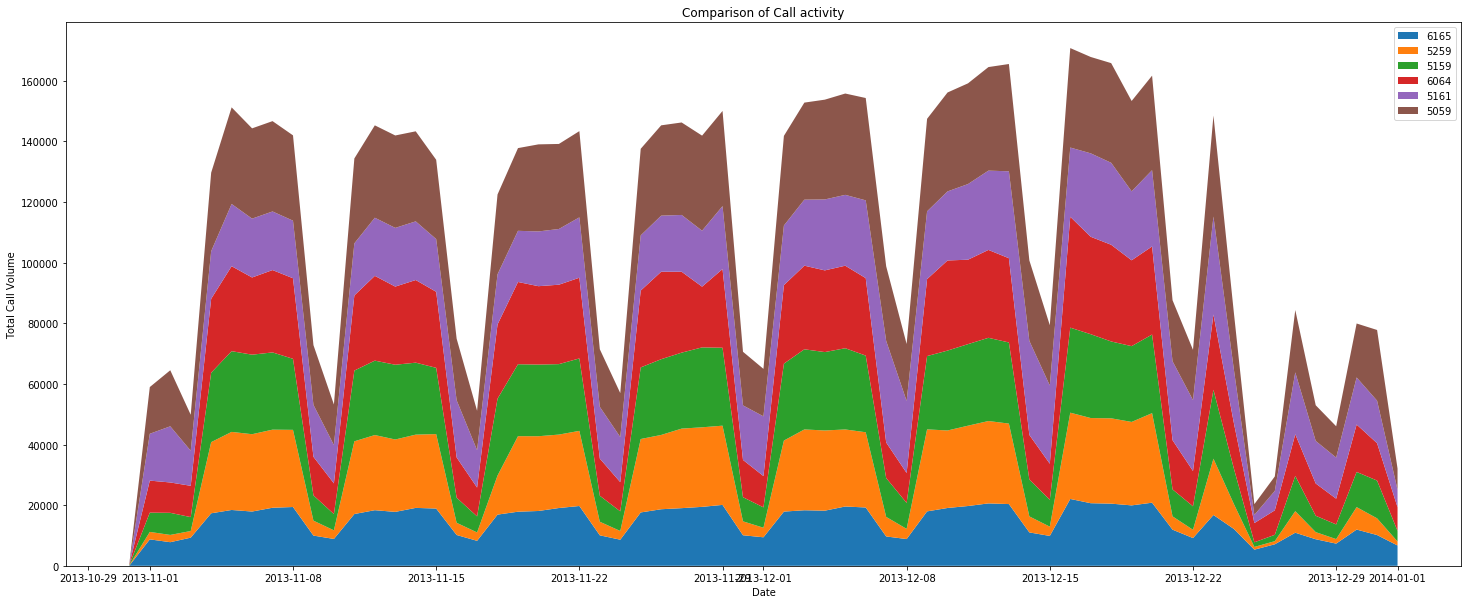
Fig 2: Bottom 5 Grids

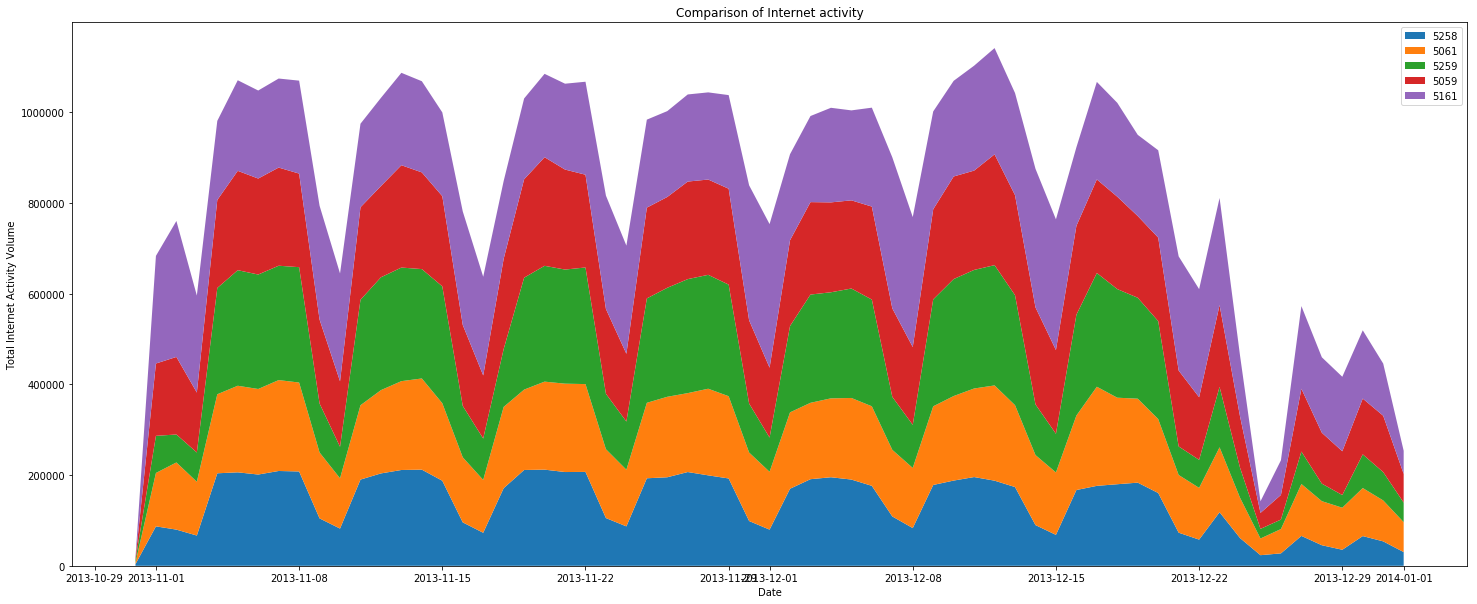


## Net volumes of SMS, Calls and Internet Activity

Stacked area plot below shows the volumes of telecommunication activities that occurs in top Grids. There aren’t any significant changes in the volumes among the top grids. They also have similar seasonality. However, Grid 5059 handles the highest amount of SMS & Calls and Grid 5161 handles highest internet activity. In the next section, we will see that these two grids are in the center of the city with very high population density.



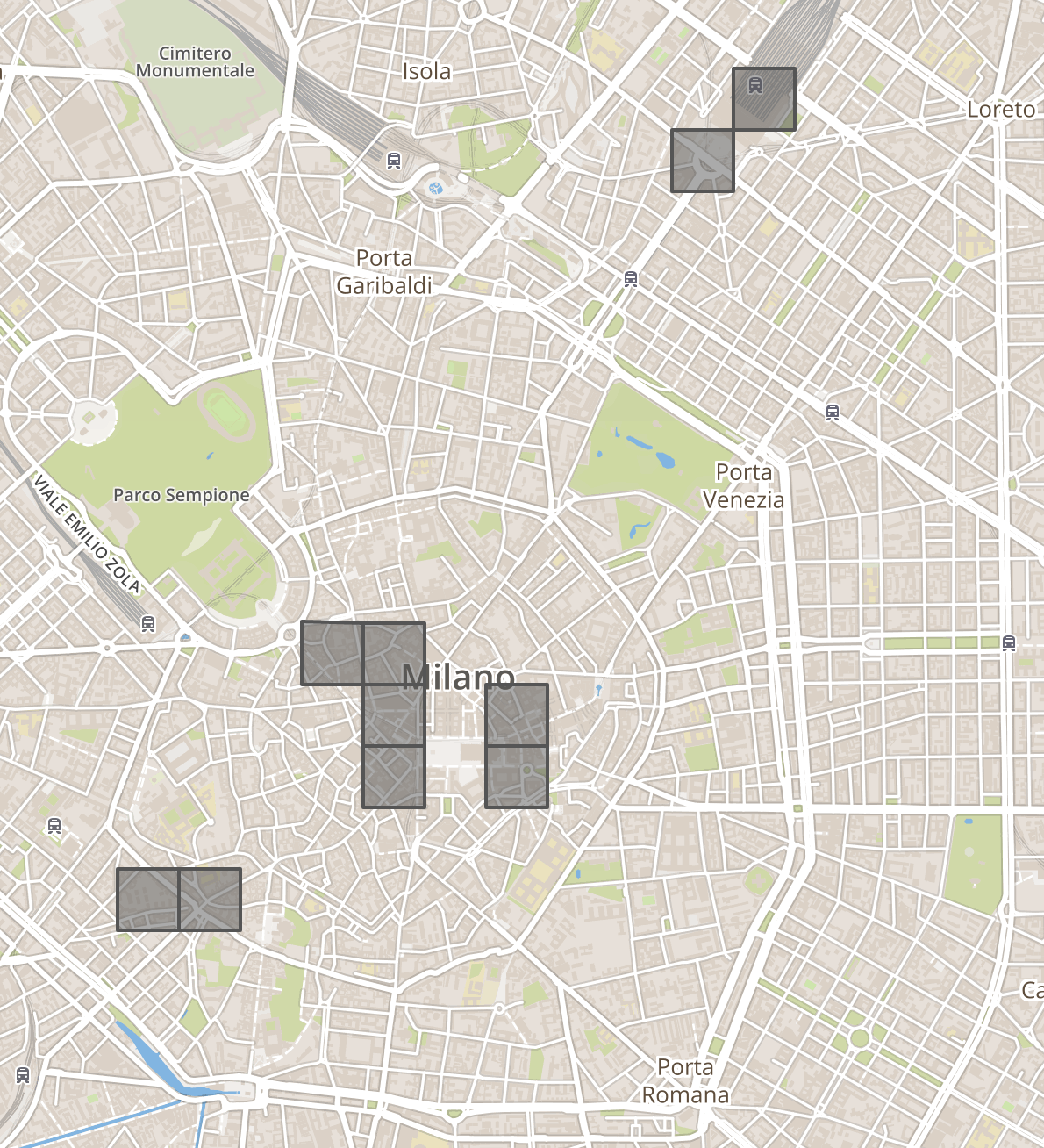




## Mapping Grids with real geographic locations

By mapping the top and bottom Grid Ids with the provided geojson file and Milan’s population density map, we note that the top 5 Grids are located in a highly populated area whereas least 5 Grids are located in sparsely populated area.

Fig 3: Grids with highest telecommunication activities are in high population density area



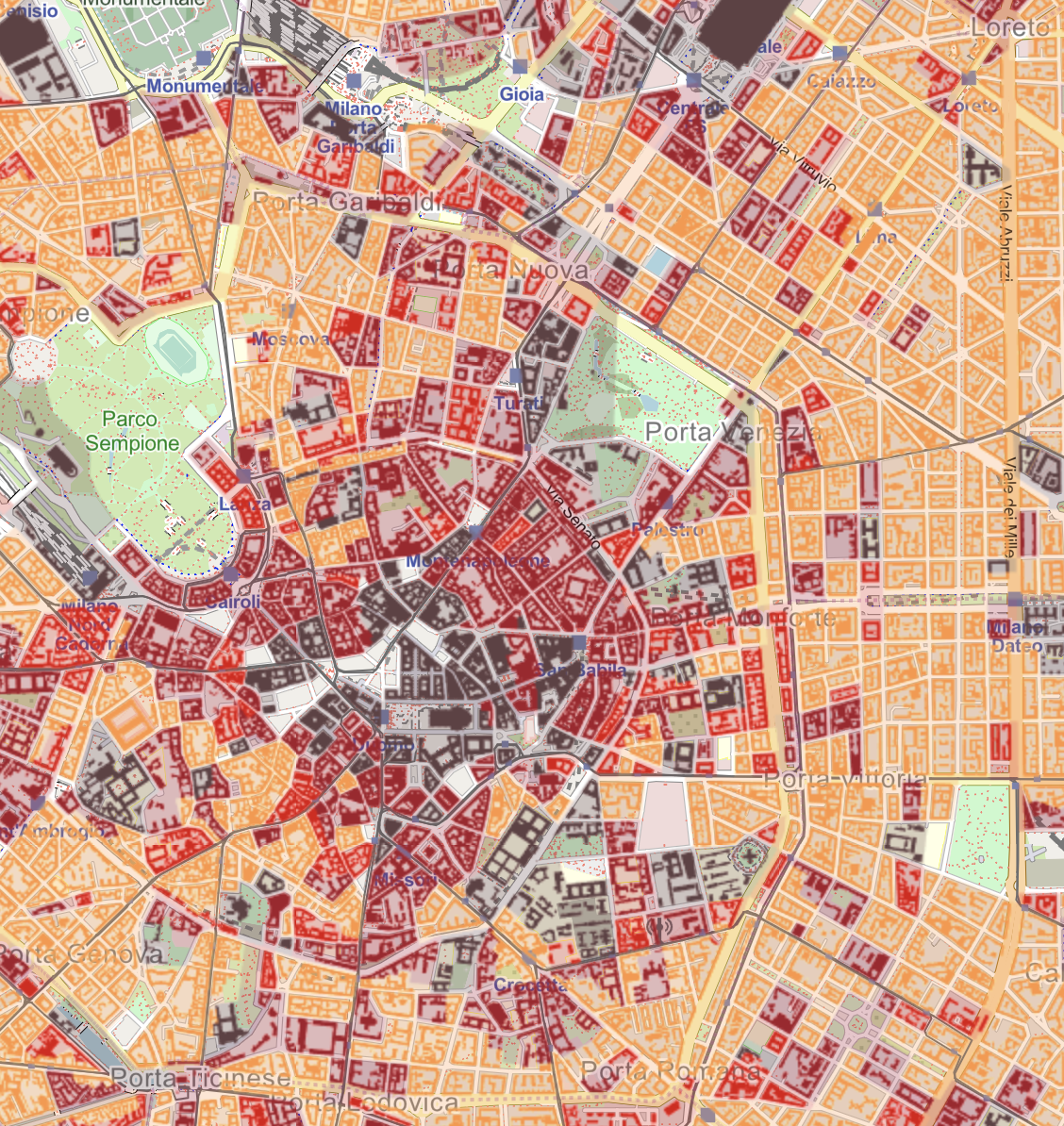
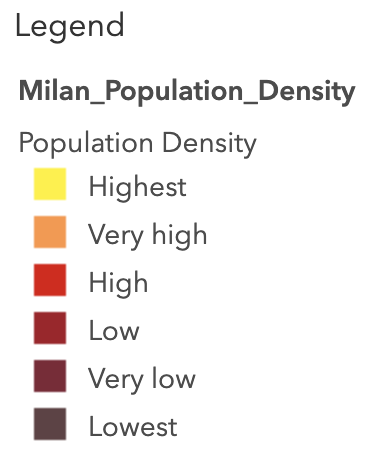
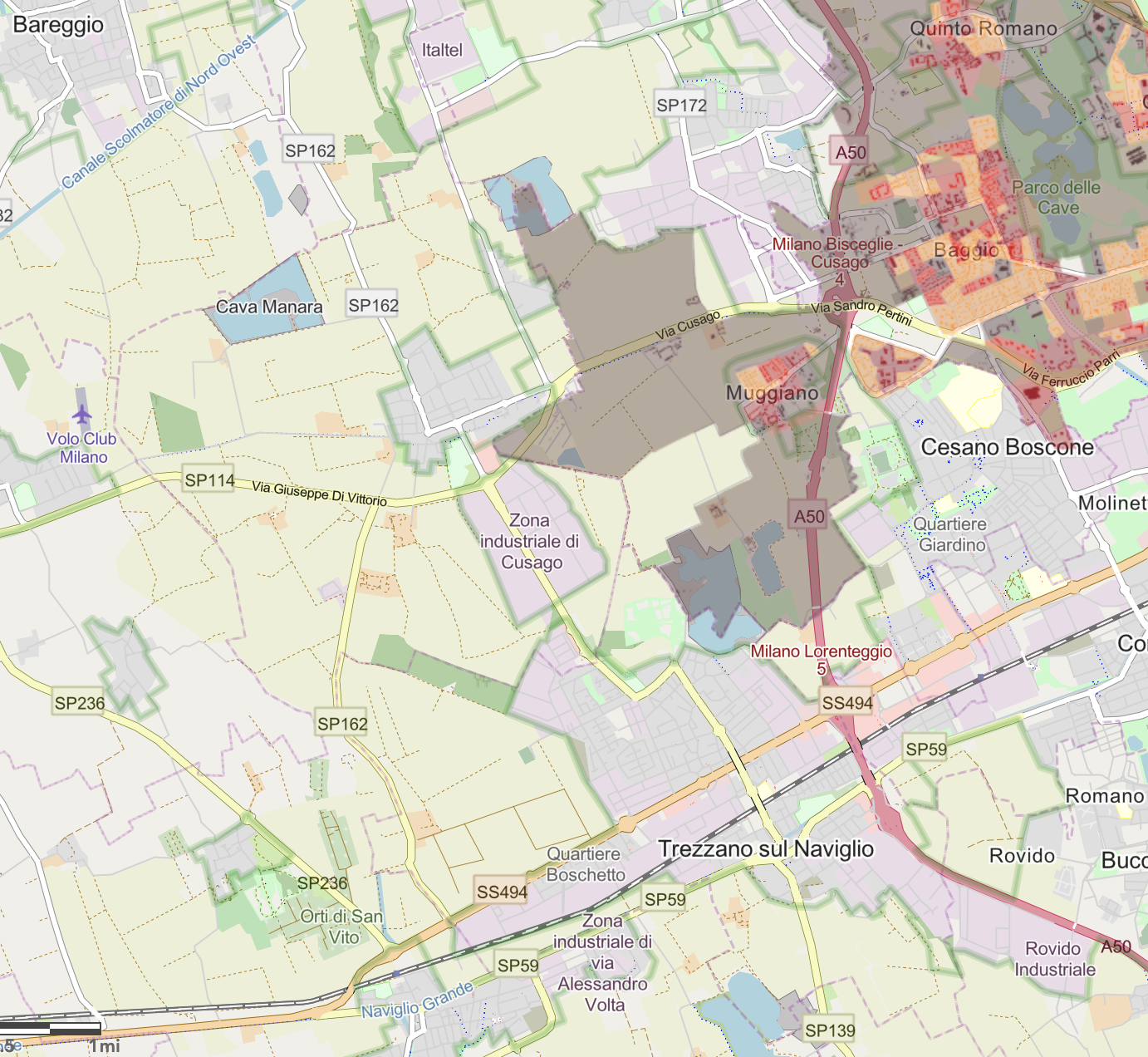
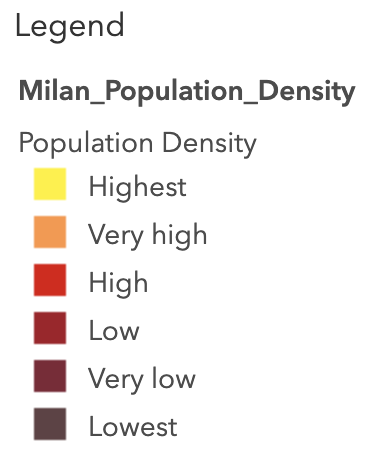
 

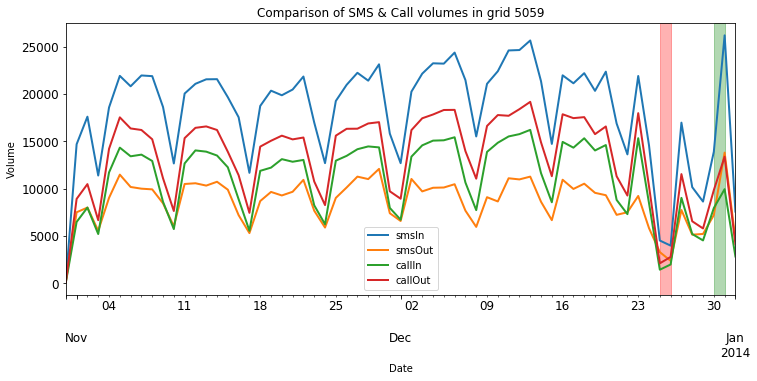
Fig 4: Grids with least telecommunication activities are in low population density area

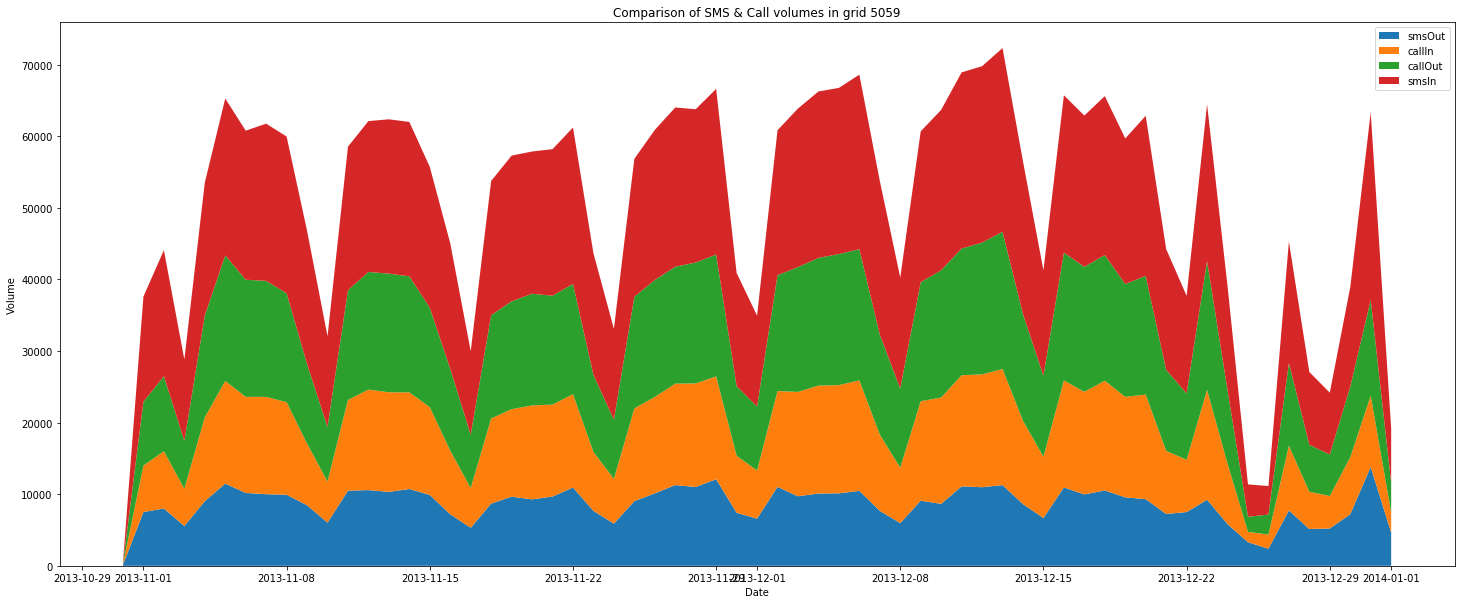


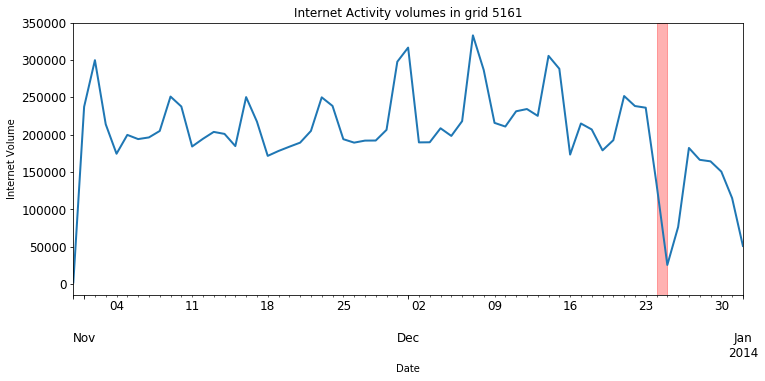
 

## Comparison of telecommunication activities within the top Grid

We will now compare the SMS In/Out, Call In/Out volumes in grid 5059 and Internet activities in grid 5161.







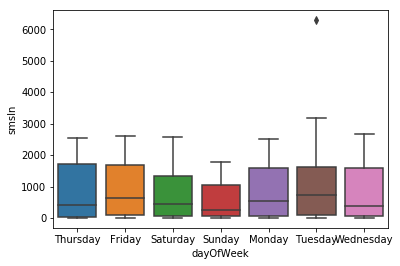
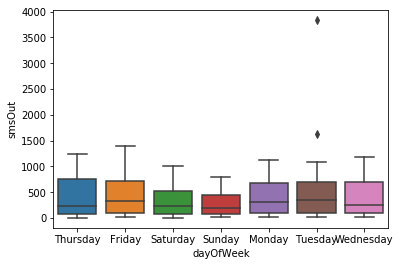
From both time series plot and stacked area plot we can conclude that grid 5059 receives more volumes SMS than calls. We also observe seasonality in the time series plot. There is a sudden drop in all activities around Christmas holiday, Dec 25th 2013, as seen in the red shaded region and a steep increase in calls and SMS volume around New Year Eve.

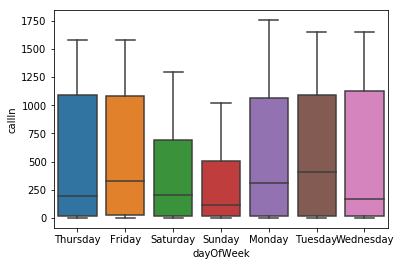
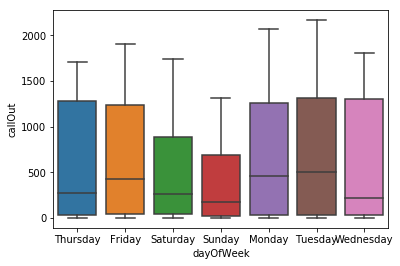
Internet activity volumes in grid 5161 also shows seasonality, and a sudden drop in volumes around Christmas holiday.

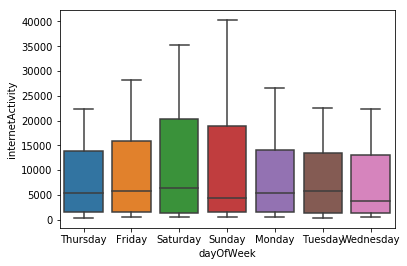
## Understanding the seasonality

### Box plot:

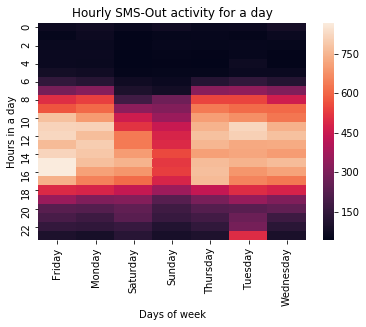
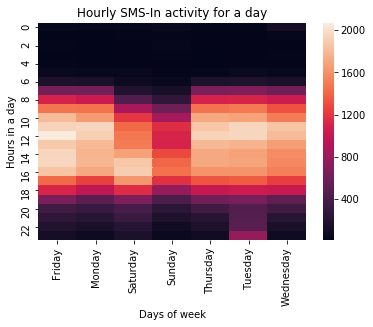
Box plot helps us understand the seasonality in the time series plots we observed earlier. We see that SMS and call volumes decreases during weekends while Internet activity volumes increases during weekends.

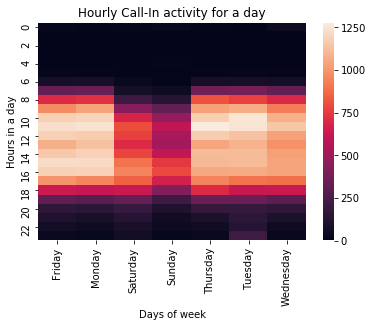
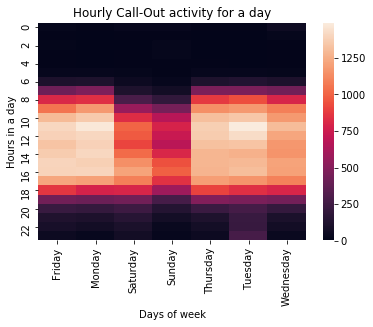
 

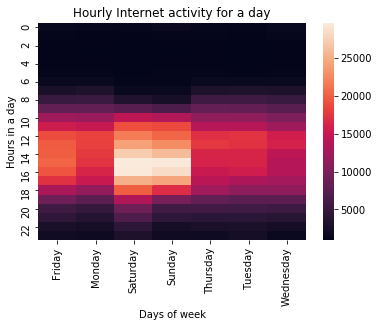
 



### Heat map:





On weekdays, SMS and call activities typically starts around 6am, and internet activity starts around 9am. During weekends, all activities starts around 8am. Internet activity on Saturdays extends late into the evening, until 10pm. As observed in box plots, heat maps show during weekends there is decrease in SMS & Call activities Internet activity increases.