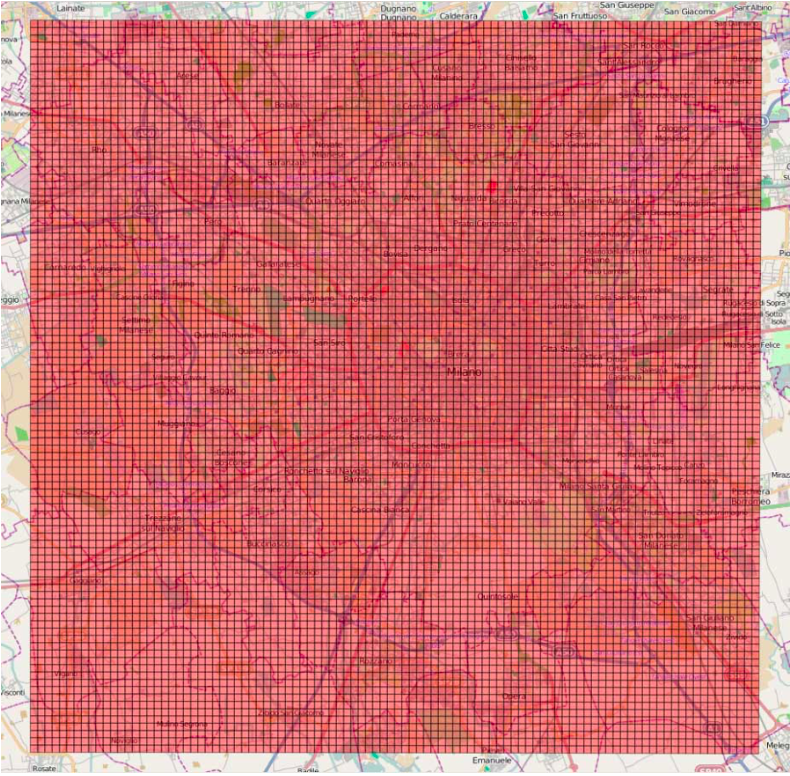
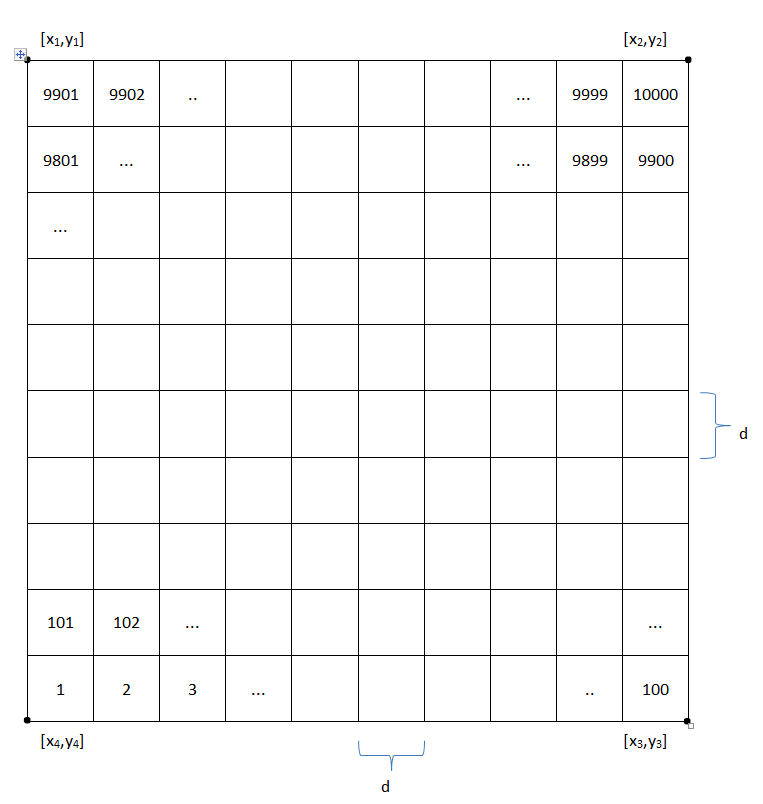
Milano Grid



* In this above picture, The Grid is overlaid to the city of Milano.
* **Grid** – Some of the datasets are spatially aggregated using a regular grid overlayed on the territory. The Grid dataset provides the geographical reference of each square which composes the grid in the reference system: WGS 84-EPSG:4326.
* Square id: Identification string of a given square of the Milan or Trentino Grid;
* Time Interval: The cell geometry expressed as geoJSON and projected in WGS84(EPSG:4326).



***WGS84 (EPSG:4326)***

**[x1,y1] = [**9.011533669936474, 45.56821407553667 **]**

**[x2,y2] = [** 9.312688264185276, 45.56778671132765 **]**

**[x3,y3] = [**9.311521155996243, 45.356261753717845 **]**

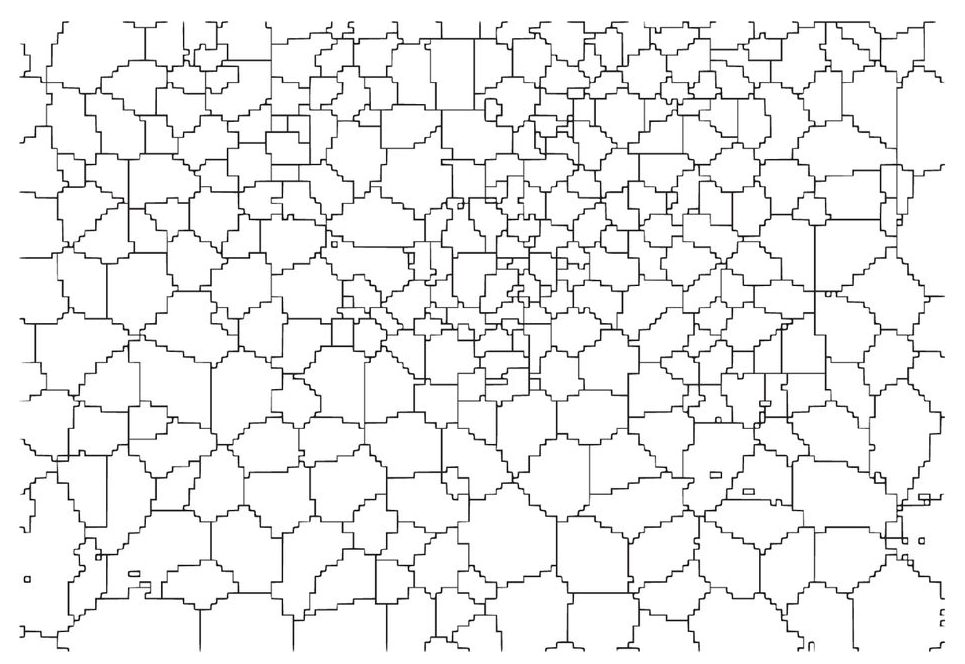
**[x4,y4] = [**9.011490619692509, 45.356685994655464 **]**

***d: 235 m***

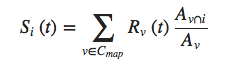
* In above picture, squares are numbered with ids. The square id numbering starts from the bottom left corner of the grid and grows till its right top corner.
* Milano Grid has total of 10,000 cells. Each cell has a cell Id.
* Each cell Id has 4 sides and each side’s length is 235 m. So Total area of one cell Id would be 55225 m2 .

Telecom Italia data-set

* The data-set refers to the traffic of voice/sms/data for each squared cell of the Milan area, measured during November and December 2013.
* It provides information about telecommunication activities in the city.
* Squared cells are numbered through ID, instead cell geometry.
* Cell geometry is expressed as GEO-JSON format and projected in WGS84 (EPSG:4326).
* **Telecommunications** - The Telecommunication datasets provide data about the telecommunication activity in the city of Milan and in the Province of Trentino. Specifically, we are releasing three different datasets, one for telecommunication activities and two for telecommunication interactions.
* Telecommunications activity – This dataset [<https://dandelion.eu/datagems/SpazioDati/telecom-sms-call-internet-mi/resource/>] serves as measure of the level of interaction between the users and the mobile phone network.
* Square id: Identification string of a given square of Milan/Trentino GRID;
* Time Interval: Start interval time expressed in milliseconds. The end interval time can be obtained by adding 600,000 milliseconds (10 min) to this value;
* SMS-in activity: Activity proportional to the amount of received SMSs inside a given square id and during a given Time interval. The SMSs are send from the nation identified by the Country code;
* SMS-out Activity: Activity proportional to the amount of send SMSs inside a given Square id during a given Time interval. The SMSs are received in the nation identified by the Country code;
* Call-in activity: Activity proportional to the amount of received calls inside the Square id during a given Time interval. The calls are issued from the nation identified by the Country code;
* Call-out activity: Activity proportional to the amount of issued calls inside a given Square id during a given Time interval. The calls are received in the nation identified by the Country code;
* Internet traffic activity: Number of CDRs generated inside a given Square id during a given Time Interval. The Internet traffic is initiated from the nation identified by the County code;
* Country code: The phone country code of the nation.
* The Call Detail Records (CDR) are provided by the Semantics and Knowledge Innovation Lab (SKIL) of Telecom Italia. Every time a user engages a telecommunication interaction, a Radio Base Station (RBS) is assigned by the operator and delivers the communication through the network. Then, a new CDR is created recording the time of interaction and the RBS which handled it. From the RBS it is possible to obtain an indication of the user’s geographical location. The coverage maps Cmap which associates each RBS to the portion of territory which it serves (AKA coverage area).
* An example of coverage map of Milan.



* In order to spatially aggregate the CDRs inside the grid, each interaction is associated with the coverage area v of the RBS which handled it. Hence, the number of records si(t) in a grid square I at time t is computed as follows:



* Where Rv,j(t) is the number of records in the coverage area v at time t, Av is the surface of the coverage area v and Av∩I is the surface of the spatial intersection between v and the square i.
* There are many types of CDRs and Telecom Italia has recorded the following activities:
* **Received SMS** a CDR is generated each time a user receives and SMS
* **Send SMS** a CDR is generated each time a user sends an SMS
* **Incoming Call** a CDR is generated each time a user receives a call
* **Outgoing Call** a CDR is generated each time a user issues a call
* **Internet** a CDR is generated each time a user starts an Internet connection or ends an Internet connection. During the same connection a CDR is generated if the connection lasts for more than 15 min or the user transferred more than 5 MB.
* The shared datasets were created combining all this anonymous information, which a temporal aggregation of time slots of ten minutes. The number of records in the datasets S’I (t) follows the rule:

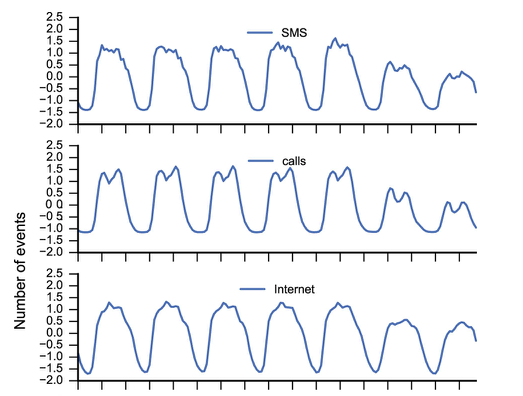
S’i (t) = Si (t) k

Where k is a constant defined by Telecom Italia, which hides the true number of calls, SMS and connections.

* CDRs log the user activity for billing purposes and network management every ten minutes, creating 144 records for each day.
* The data has been collected over two months, from November 1st, 2013 to January 1st, 2014 and the information is geo-referenced to the city of Milan and to the Province of Trentino. Milan is the main industrial, commercial, and financial centre of Italy. The city has a population of about 1.3 million. Trentino is an autonomous province of Italy, located in the northern part of the country. It covers an area of more than 6,000 km2, with a total population of about 0.5 million. Link for telecommunication data is shown below.

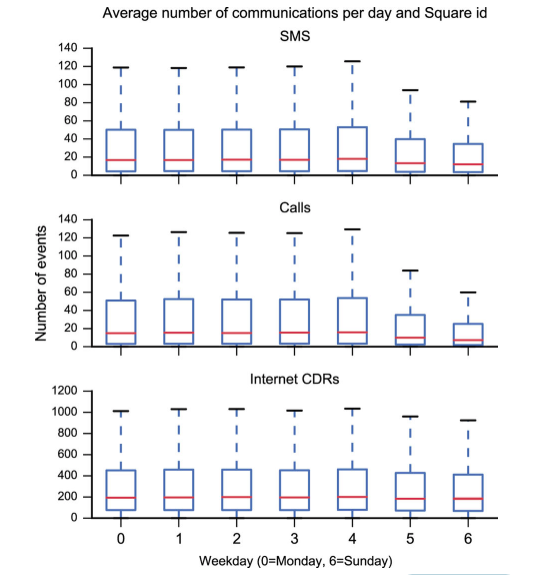
<https://dandelion.eu/datagems/SpazioDati/telecom-sms-call-internet-mi/resource/>

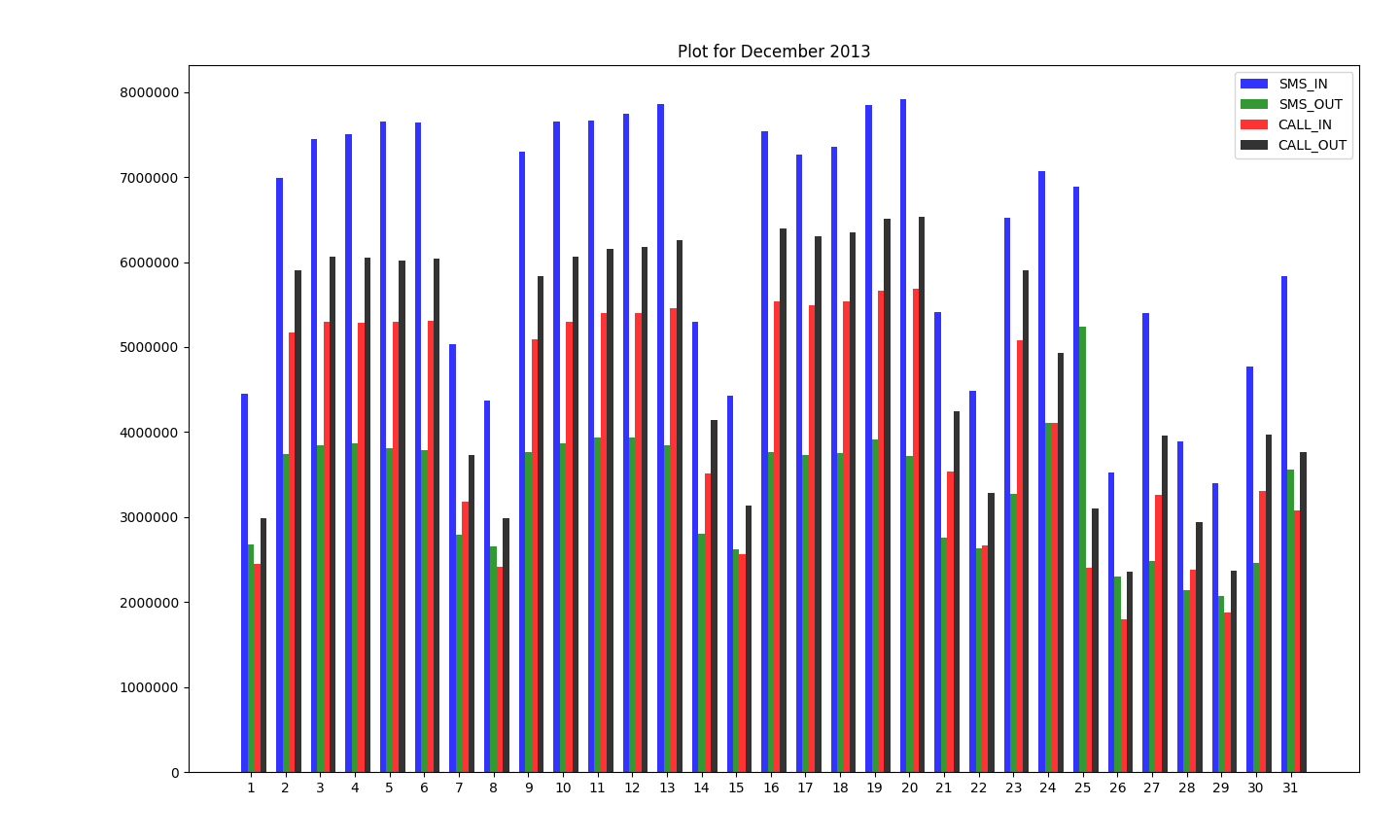
* Technical Validation – The technical quality validation of the datasets is limited due to the absence of similar datasets to compare our results with. Hence, in this section we propose a statistical and visual characterization with the aim of supporting the naïve correctness of the information provided.
* Temporal aspects – Generally, people perform different activities during the day, Many of them are repeated on a daily basis (e.g., eating at noon, jogging in the evening etc.), others on a weekly basis (e.g., watching the favourite football team at the stadium). From Figs 1 and 2 it is possible to observe a strong daily seasonality which usually starts at 7:00, when people turn on their phones and probably commute to work and then slowly decreases in the evening when people return home and sleep. Moreover, there is also a weekly seasonality due to the work cycles behavior of people (e.g., working days versus weekends).
* Figure 1: Weekly Z-scaled behavior of SMS, Calls, Tweets and Internet CDRs in Milan.



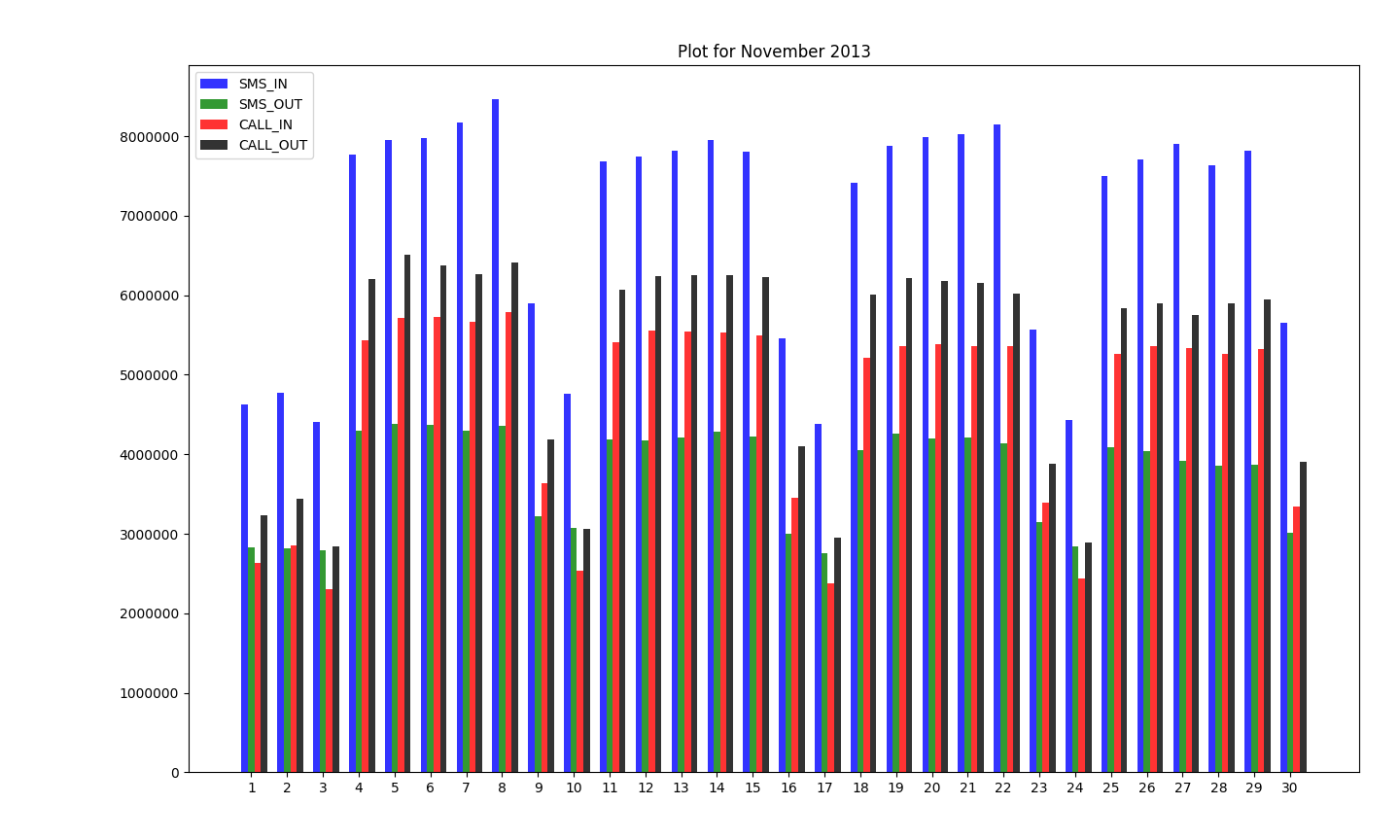
Hour (in a week)

* Figure 2: Box-plots showing the calls, SMS, and Internet CDRs distributions per weekday and per cell in Milan.

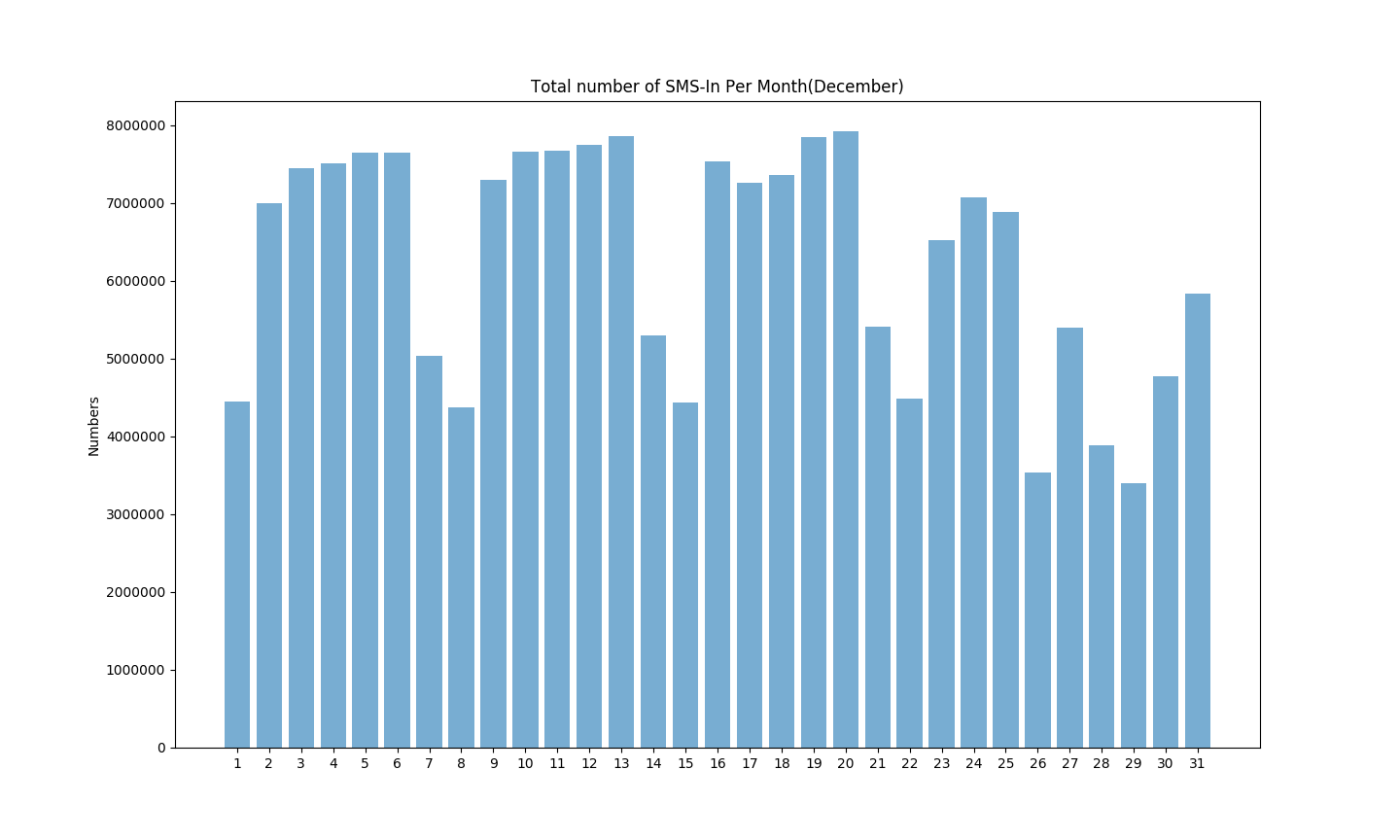




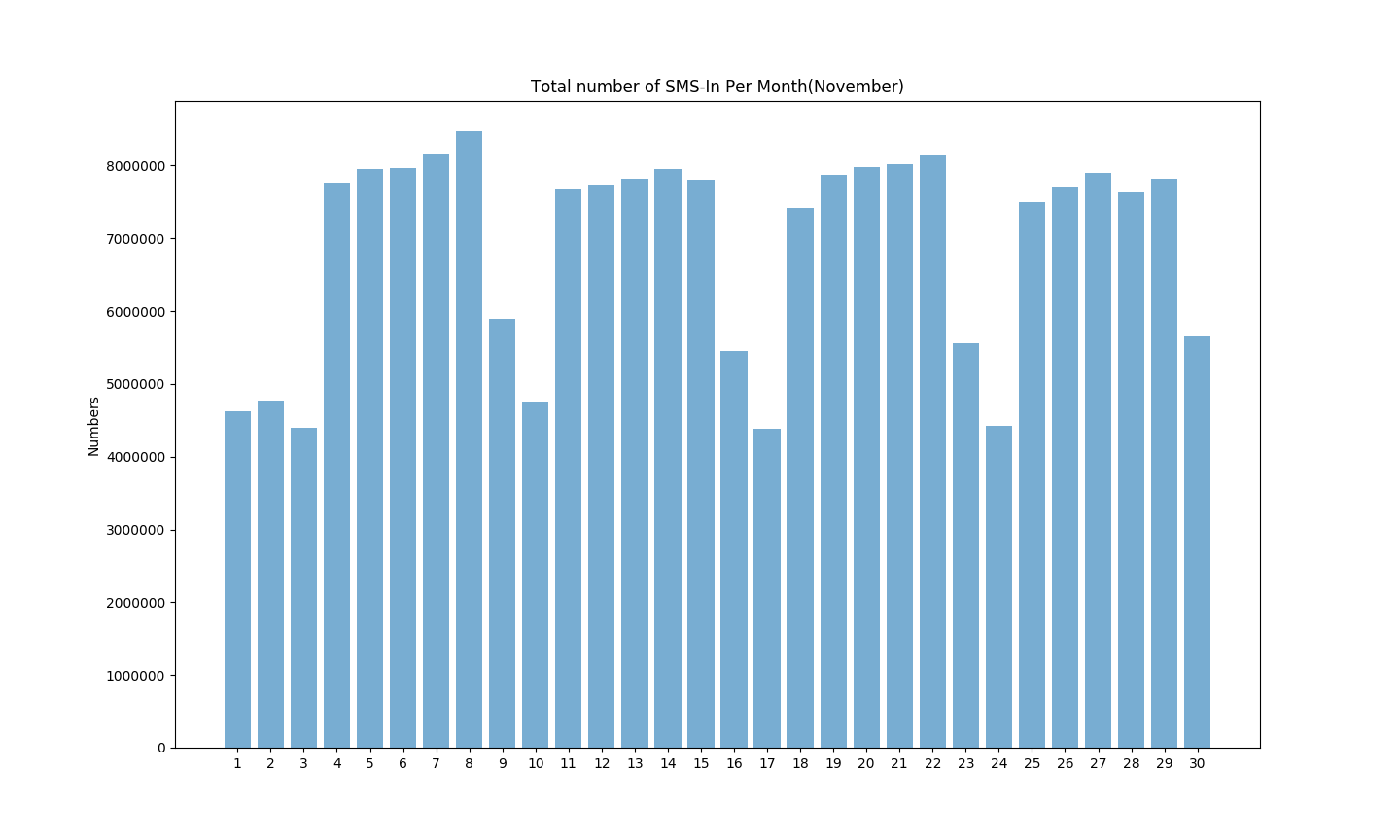
* This above bar plot shows number of activities like SMS-In, SMS-Out, Call-In and Call-Out.
* As we can see here that number of SMS-In for almost every month is higher than all other three activities.
* Number of SMS-Out for most of months is lower than all of the other three activities.
* The lowest record we found is 1796979.26 for Call-In on 26th of December month and highest record we found is 7917101.16 for SMS-In on 20th of December month.



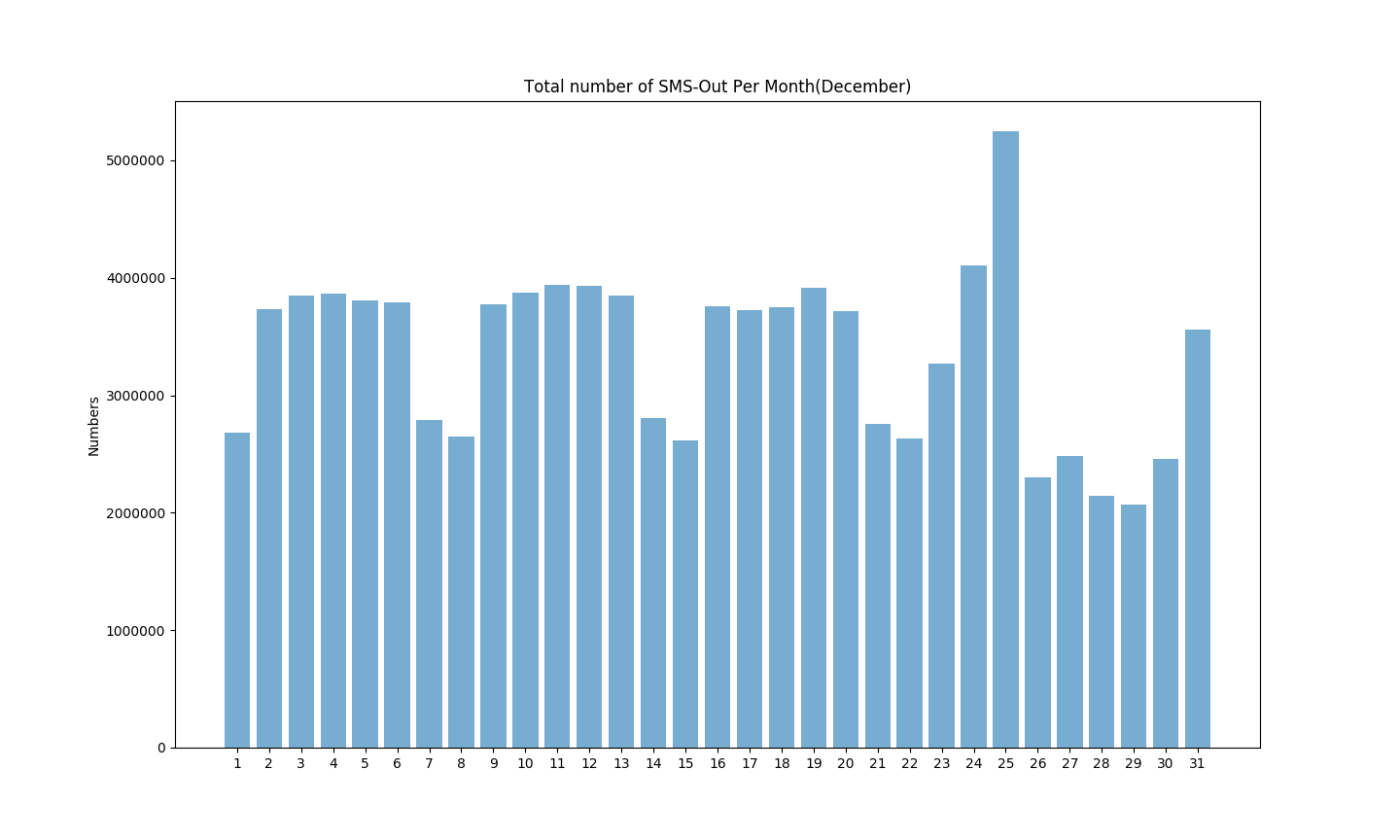
* This bar graph shows the same activities like above.
* Here also we have number of SMS-In activities is higher than other three activities for all the days of November month.
* SMS-Out activity for most of months is lower than other three activities.
* The highest activity we found is 8469033.67 for SMS-In Activity on 8th of November and lowest activity we found is 2299718.11 for Call-In Activity on 3rd November.
* Bar plot is shown below for SMS-In of December Month in 2013.



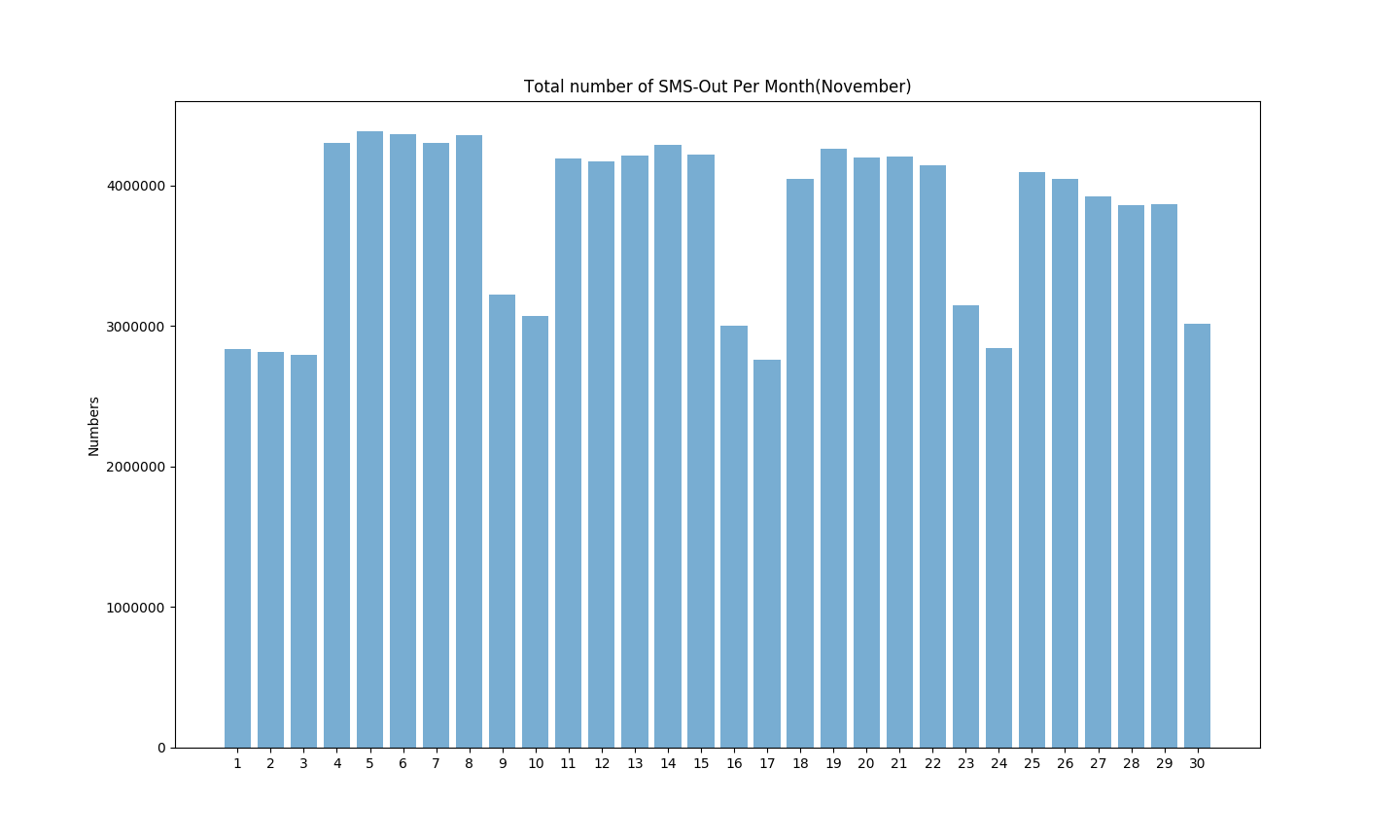
* Bar plot is shown below for SMS-In of November Month in 2013.



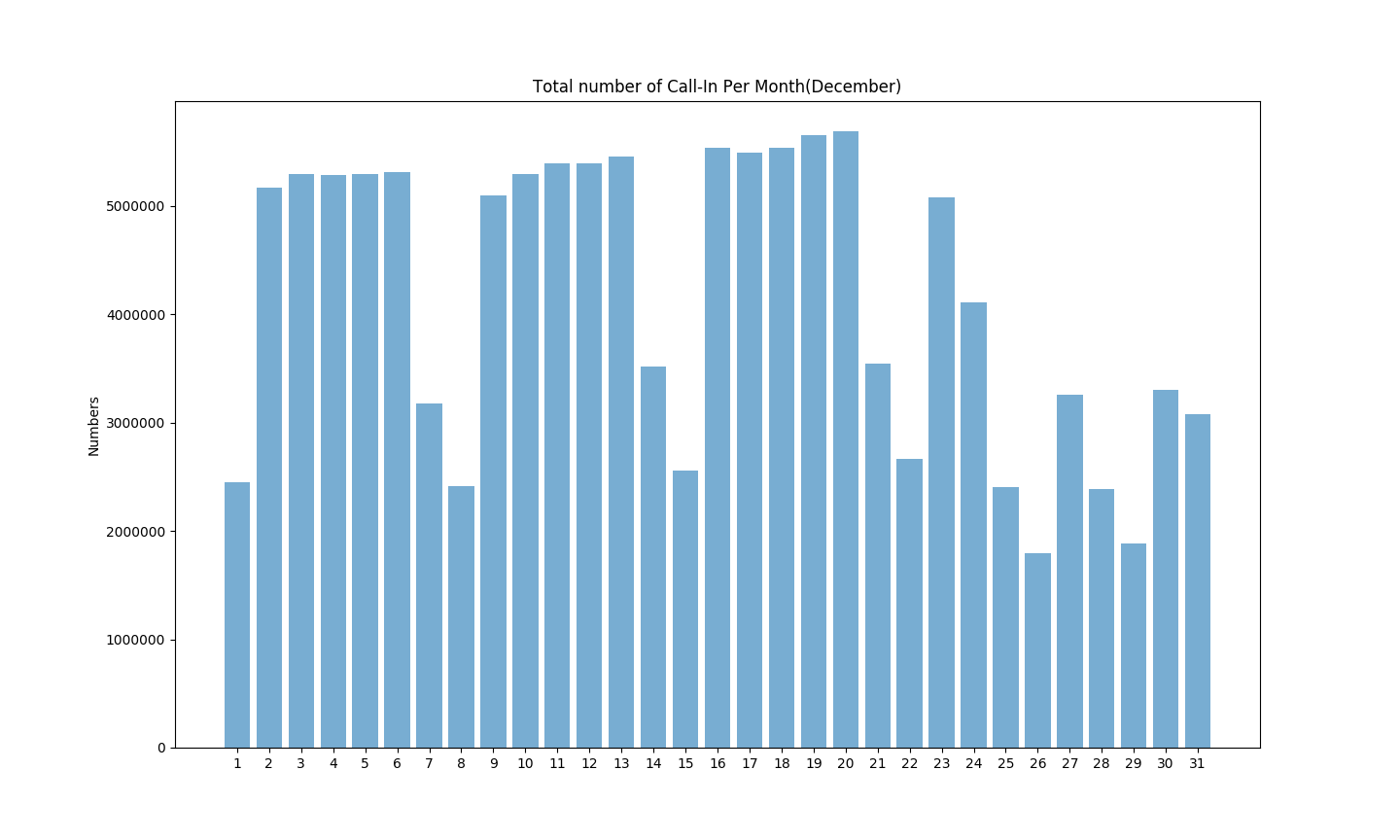
* Bar plot is shown below for SMS-Out of December Month in 2013.



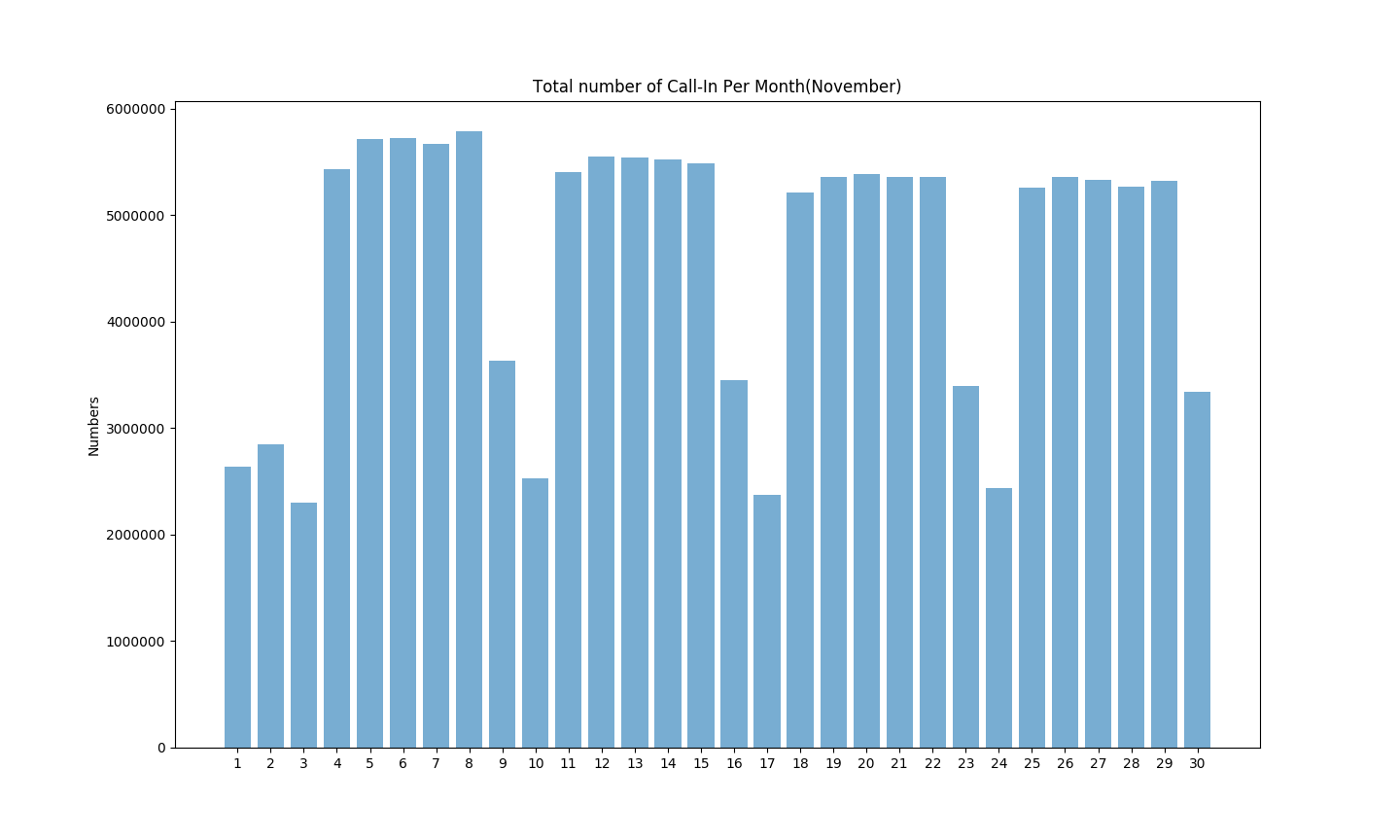
* Bar plot is shown below for SMS-Out of November Month in 2013.



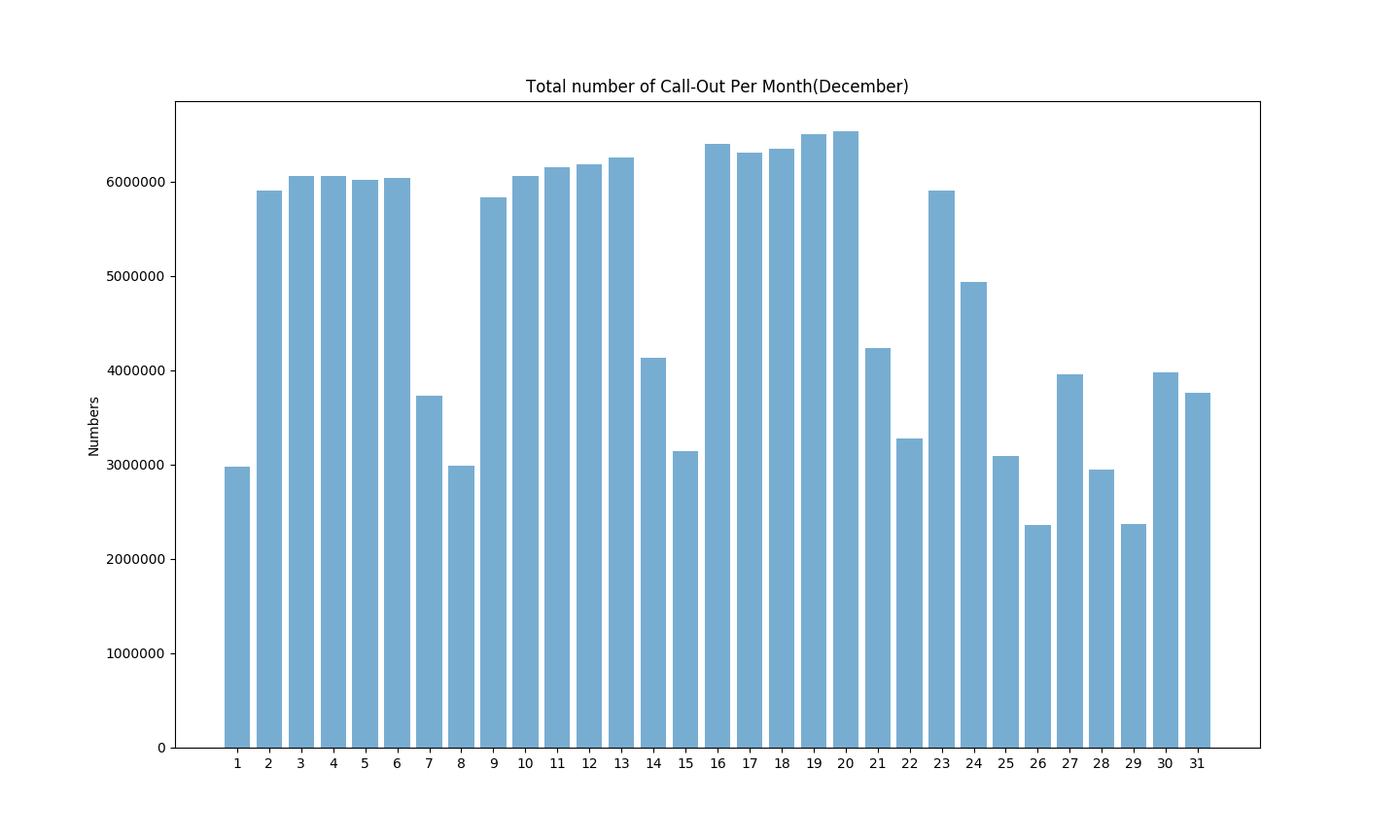
* Bar plot is shown below for Call-In of December Month in 2013.



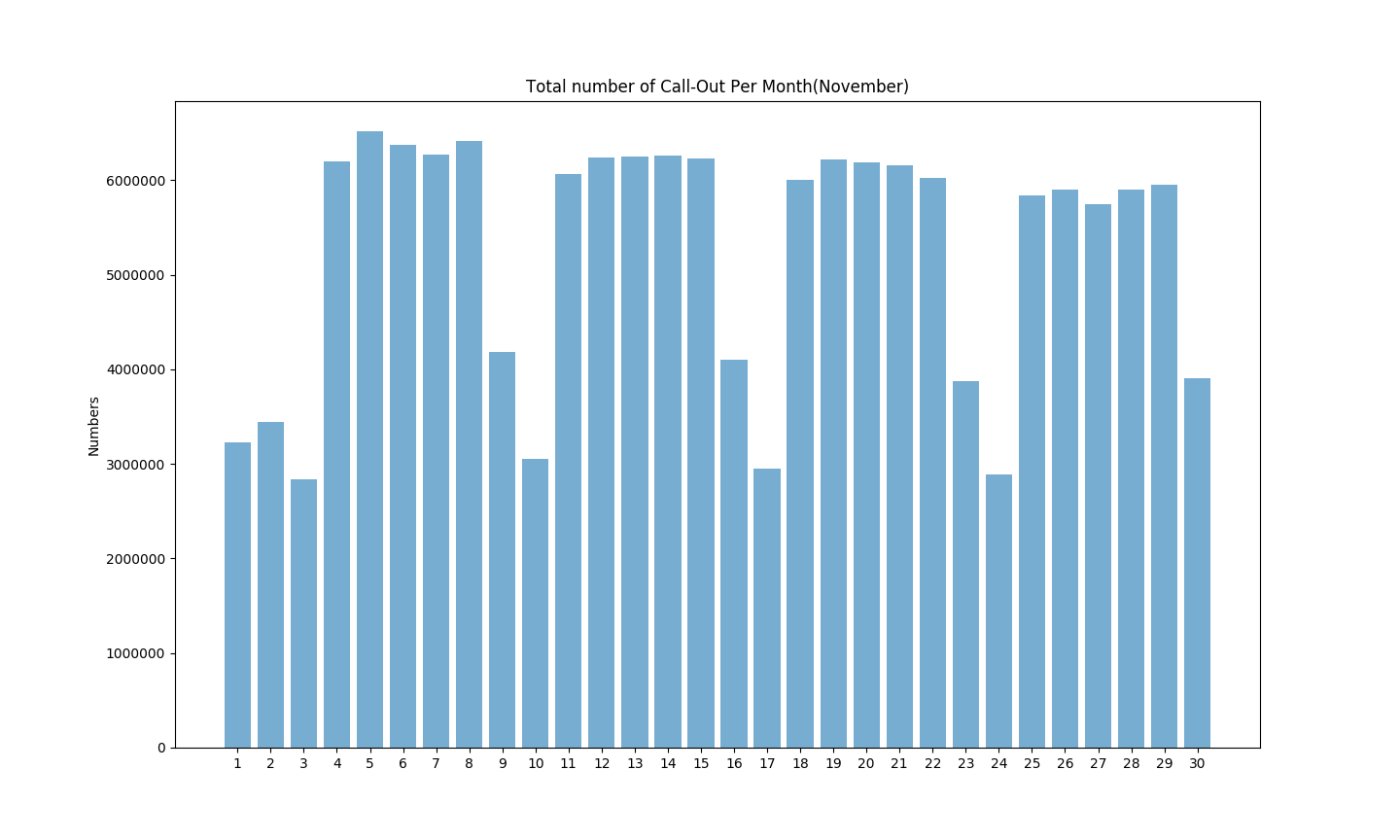
* Bar plot is shown below for Call-In of November Month in 2013.



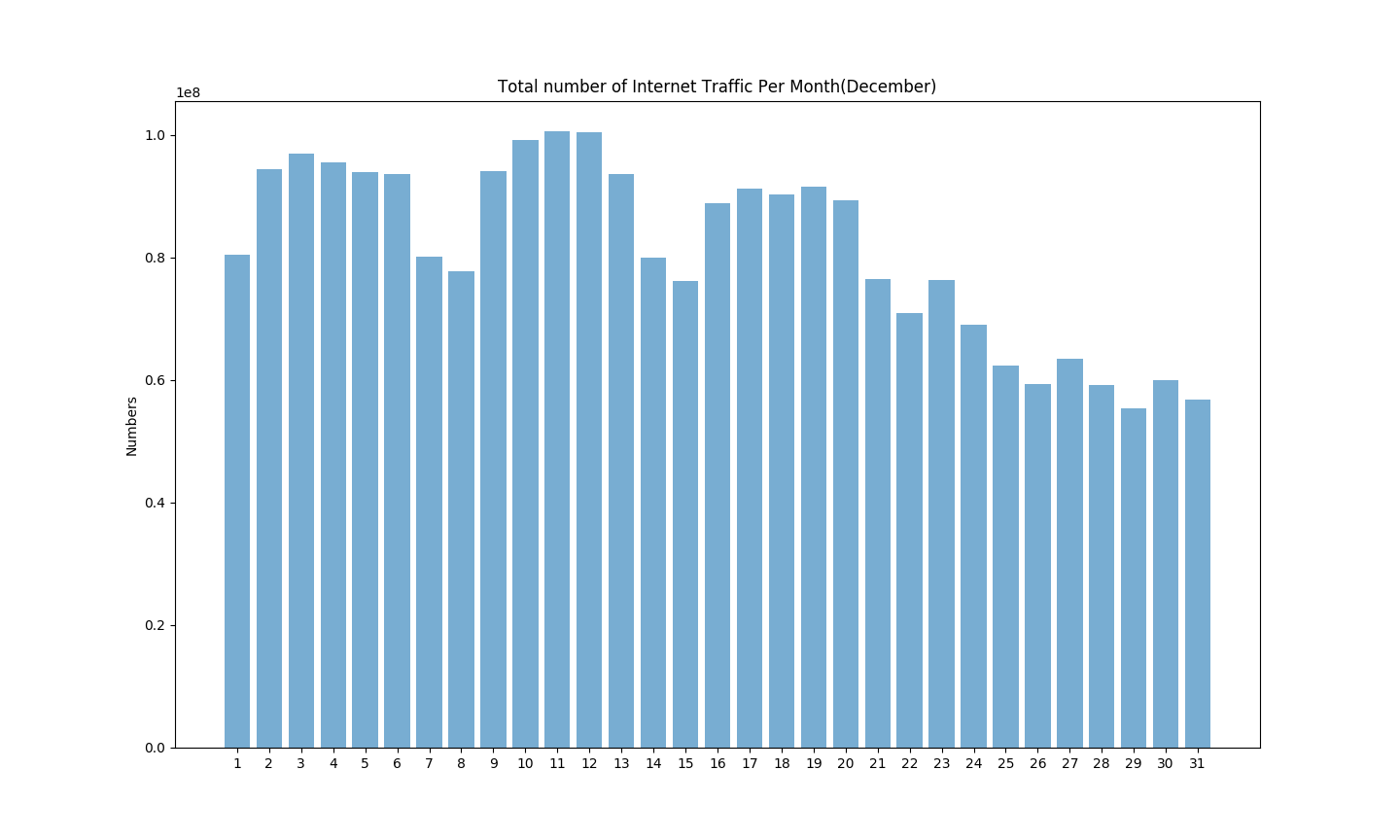
* Bar plot is shown below for Call-Out of December Month in 2013.



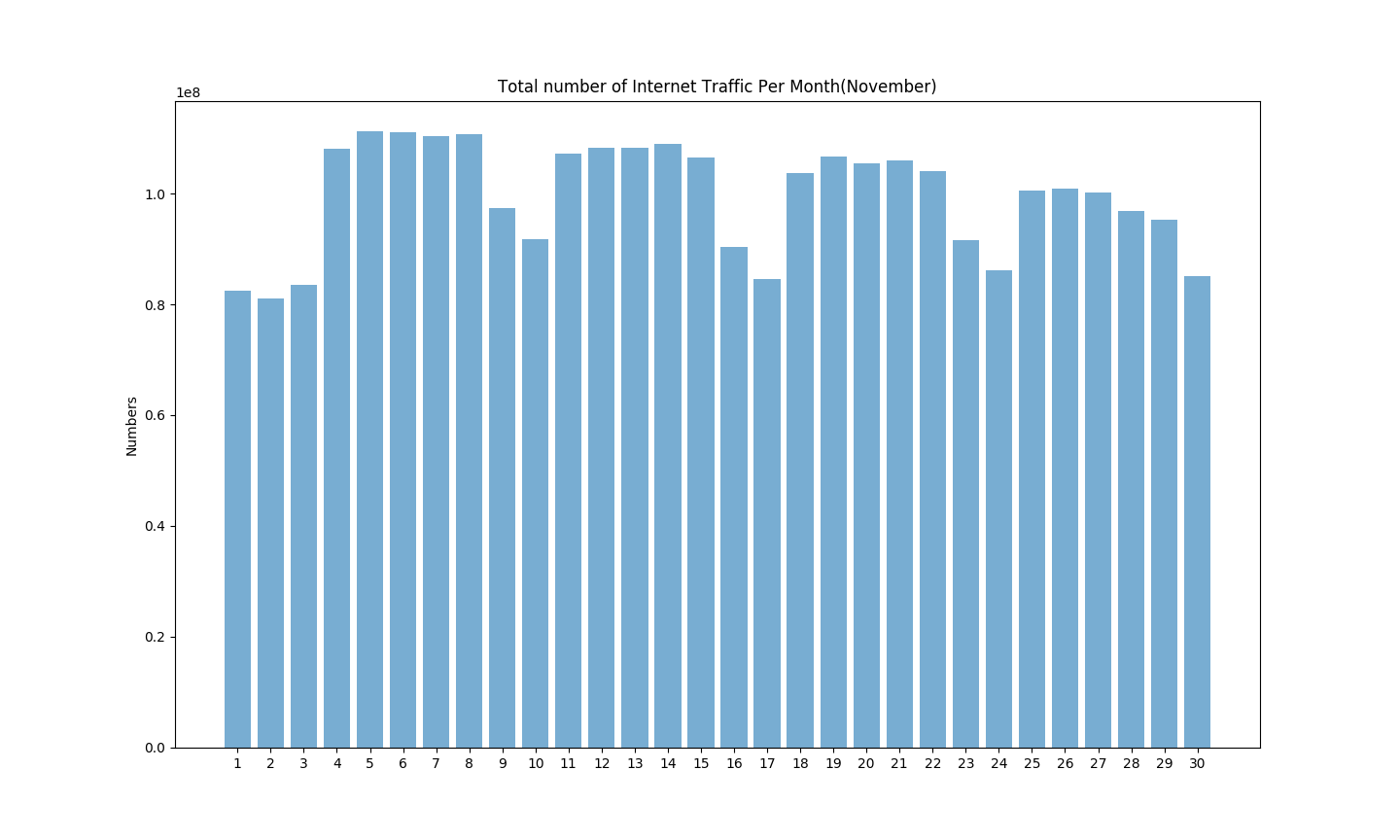
* Bar plot is shown below for Call-Out of November Month in 2013.



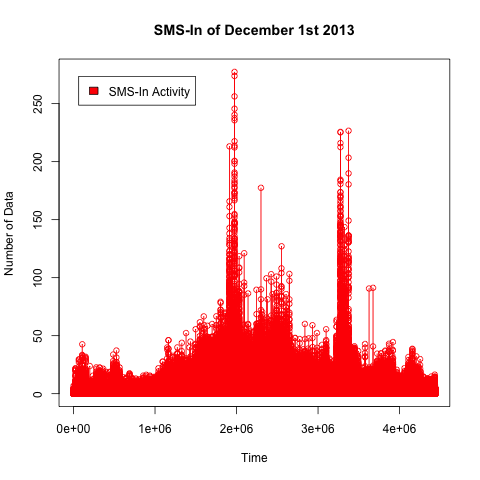
* Bar plot is shown below for Internet Traffic of December Month in 2013.



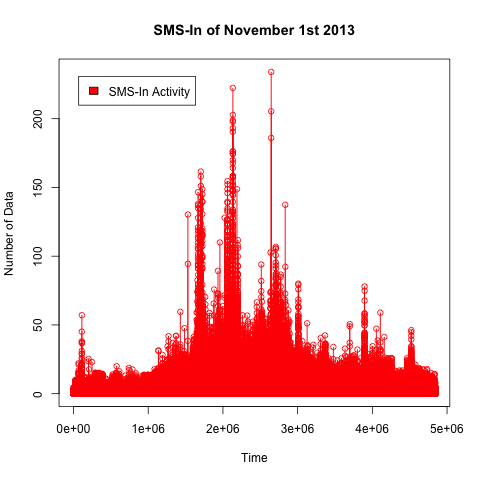
* Bar plot is shown below for Internet Traffic of November Month in 2013.



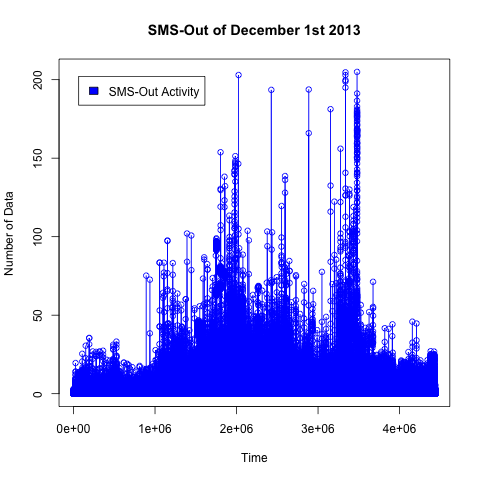
* Now This line plot is for 1st December of 2013 for SMS-In Activity. This is data of 4438331 lines of Time Interval. There is maximum activity we found is 277.35 and minimum activity is 0.



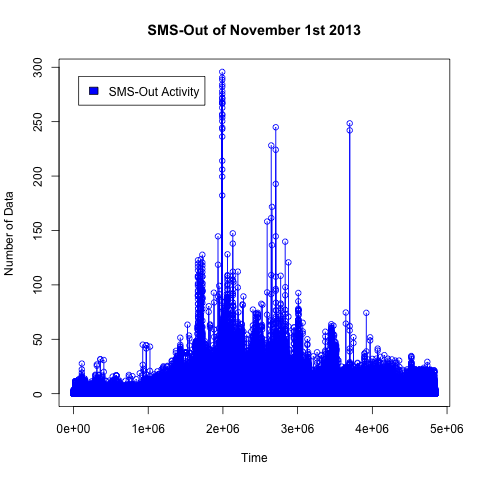
* Now this line plot is for 1st November of 2013 for SMS-In Activity. This is data of 4842625 lines of Time Interval. There is maximum activity we found is 234.02 and minimum activity is 0.



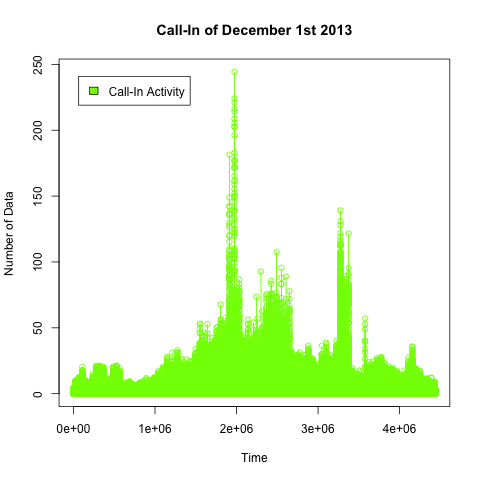
* Now this line plot is for 1st December of 2013 for SMS-Out Activity. This is data of 4438331 lines of Time Interval. There is maximum activity we found is 204.94 and minimum activity is 0.



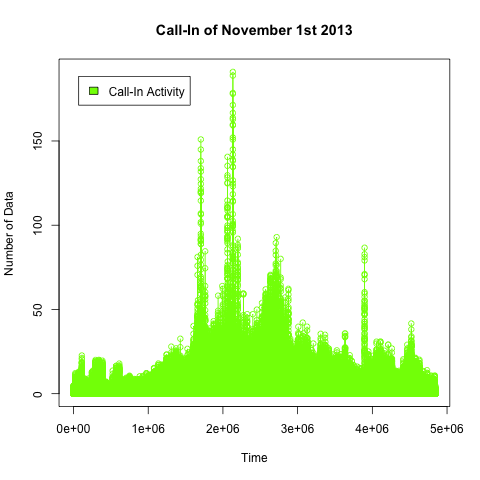
* Now this line plot is for 1st November of 2013 for SMS-In Activity. This is data of 4842625 lines of Time Interval. There is maximum activity we found is 295.73 and minimum activity is 0.



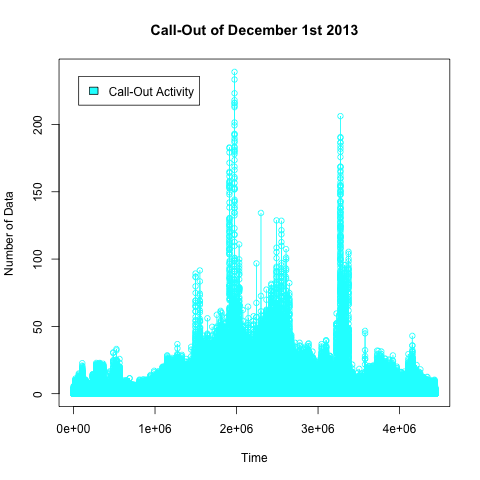
* Now This line plot is for 1st December of 2013 for Call-In Activity. This is data of 4438331 lines of Time Interval. There is maximum activity we found is 244.26 and minimum activity is 0.



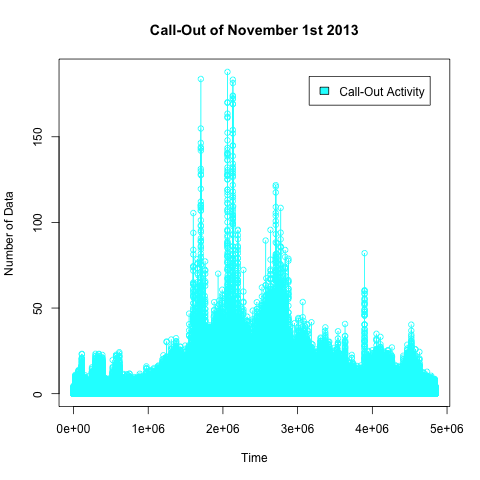
* Now this line plot is for 1st November of 2013 for Call-In Activity. This is data of 4842625 lines of Time Interval. There is maximum activity we found is 190.99 and minimum activity is 0.



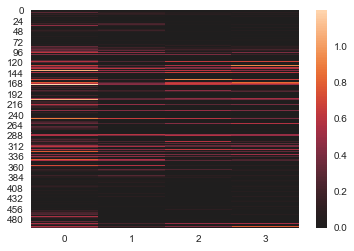
* Now This line plot is for 1st December of 2013 for Call-Out Activity. This is data of 4438331 lines of Time Interval. There is maximum activity we found is 239.04 and minimum activity is 0.



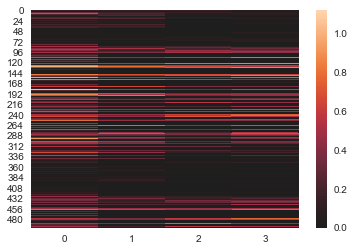
* Now this line plot is for 1st November of 2013 for Call-Out Activity. This is data of 4842625 lines of Time Interval. There is maximum activity we found is 187.87 and minimum activity is 0.



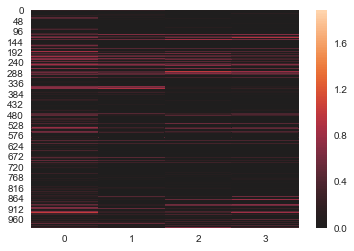
* Below figure will show heat map of first 500 number of data of December 1st 2013. Heat map covers all the four activities including SMS-In, SMS-Out, Call-In and Call-Out. Where data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



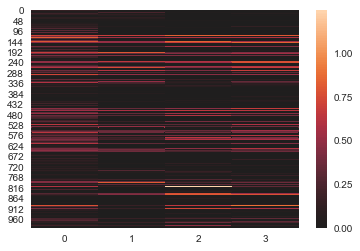
* Below figure will show heat map of first 500 number of data of November 1st 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



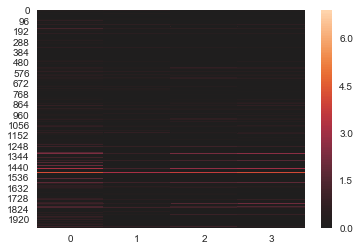
* Below heat map covers first 1000 number of data of December 1st 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



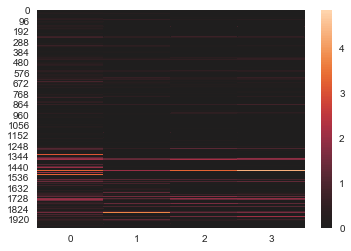
* Below heat map covers first 1000 number of data of November 1st 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



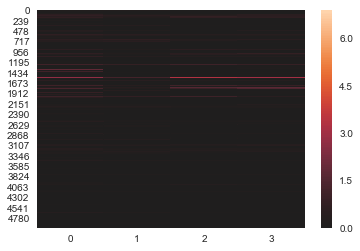
* Below heat map covers first 2000 number of data of December 1st 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



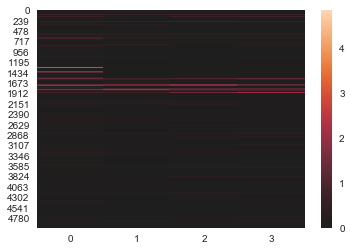
* Below heat map covers first 2000 number of data of November 1st 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



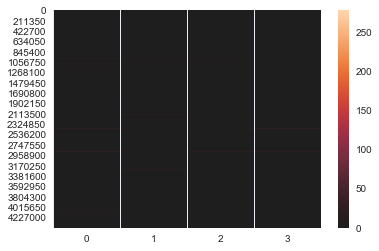
* Below heat map covers first 5000 number of data of December 1st 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



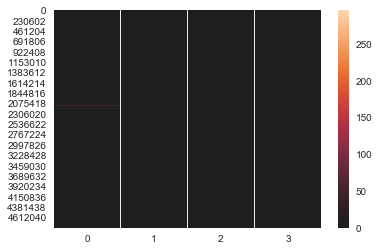
* Below heat map covers first 5000 number of data of November 1st 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



* This is the final and big heat map. It covers number of 4438331 lines of data and that is one whole day and it’s 1st December 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



* This heat map covers number of 4842625 lines of data and that is one whole day and it’s 1st November 2013. Here data on field 0 is SMS-In. Data on field 1 is SMS-Out, data on field 2 is Call-In and data on field 3 is Call-Out.



Reference: <https://www.politesi.polimi.it/bitstream/10589/126134/3/thesis.pdf>

<https://www.nature.com/articles/sdata201555#f2>