**Introduction**

In this work, we are analyzing the Web syslog NetworkRouting.net. We are trying to track the daily pattern for the top three most popular visited webpages, number of 404 happens on a daily basis trend and finally, daily number of downloads for the first chapter of the book available in the website NetworkRouting.net. We also consider the unique visitors number during each day.

For the first data we consider the number of visit for all the URLs during days; then we consider the mean value that each URL has been visited during whole available days. Among those mean values we select top three of them and plot them with date series in X axis and number of visits in Y axis.

In Section two, forecasting methods for all datasets are implemented. These forecasting techniques include Moving Average with window size of 7; such as we forecasted the next day’s visit of that specific URL. Another Method we apply for this data set is Exponential Smoothing with alpha=0.8. Regression Method is the last forecasting Method we used for this data set.

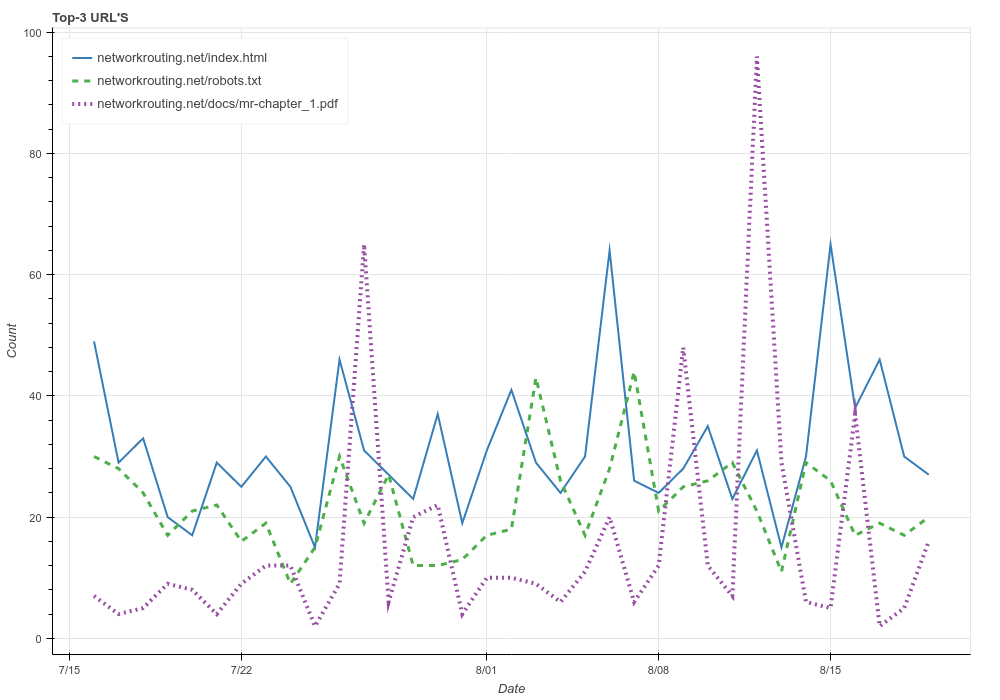
TOP Three Visited:

In this section, the number of visits for all available URLs for each day is estimated and the mean value of the visit of each URL is considered. Among all of the mean values for each URL the top three is plotted in fig. 1. As it is shown in the figure the top three webpages are Index.html, robots.txt, and first chapter of Network Routing book.

Robots.txt is a text file placed in the web server and tells webcrawlers such as [Googlebot](https://varvy.com/googlebot.html) if they can access a file. The robots.txt is among the top three visited pages and has almost a steady trend compared to two other graph which means the search engine constantly utilizing the robots.txt file to update the search engine server.

The Index.html is the most visited page and it shows more jumps at the second half of the graph which means as the school days will get close more students will check the website most probably.

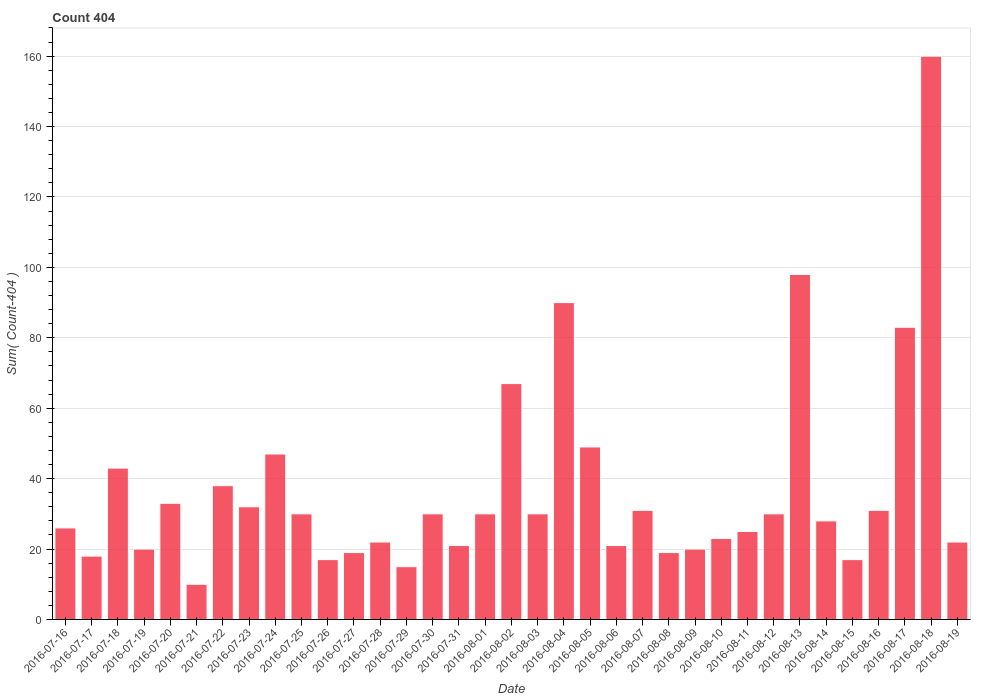
The second most visited webpage is chapter-1 with more jumps at the second half of the graph which that also may be the school dates get close and students want to decide to take a look at the book to take this course or even study.



404:

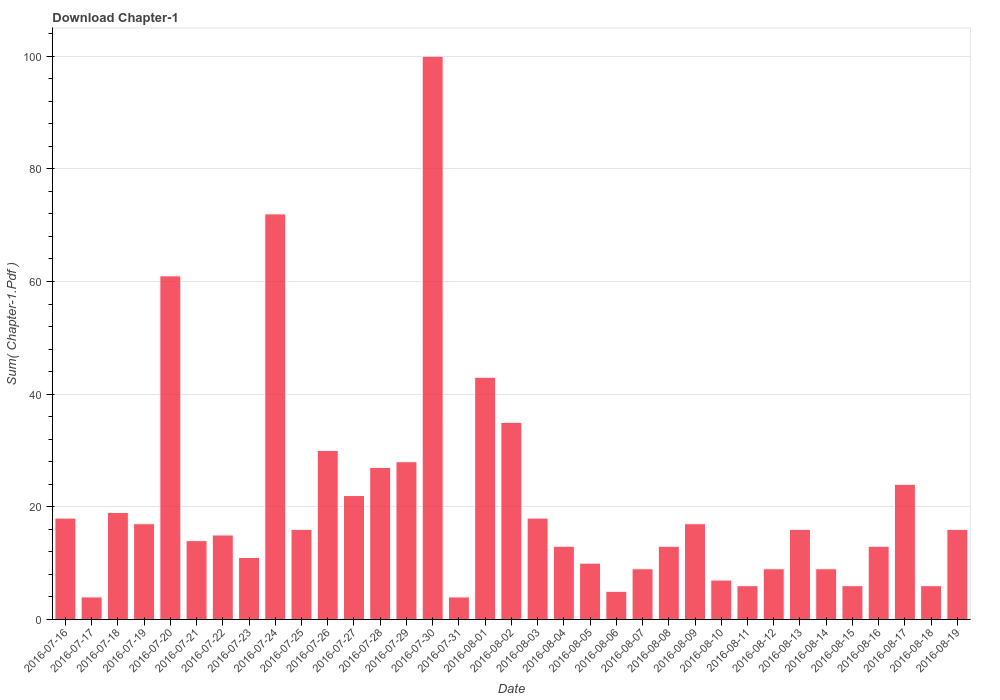
Figure 2 shows the number of 404 per day is all dates. The **404** is a HTTP response code indicating the client can communicate with server, but the server cannot find what is requested.

At the second half we have more number of 404 that that is because the more requests for URLs are ade at those days. 08/18/2016 is the pick of 404 numbers.



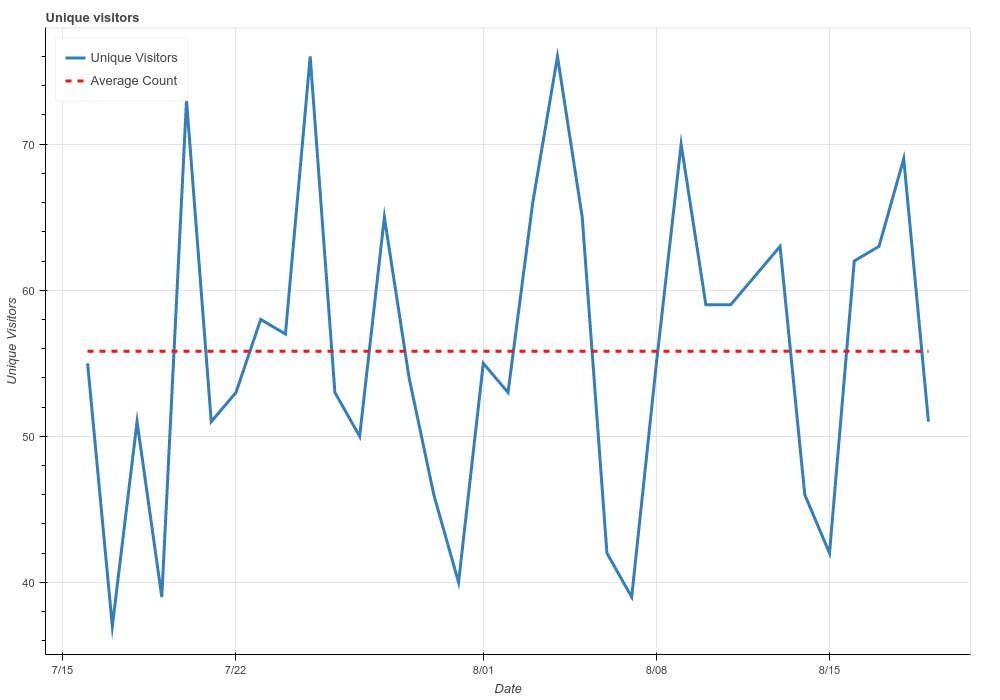
First chapter:

Pick at 2016/07/30 but why first half? In the first graph it seems second half was more. No clue according daily basic



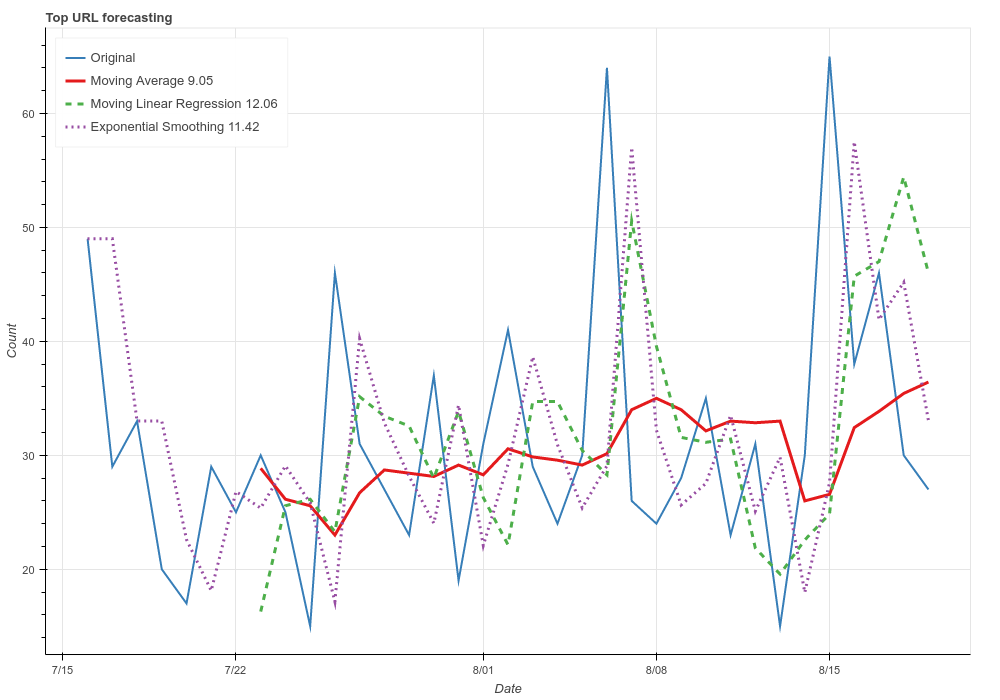
Unique visitors

Unique IP numbers that requested to visit the websites ate shown in figure 4. The data is inconsistence that no specific assumption can be made on that.



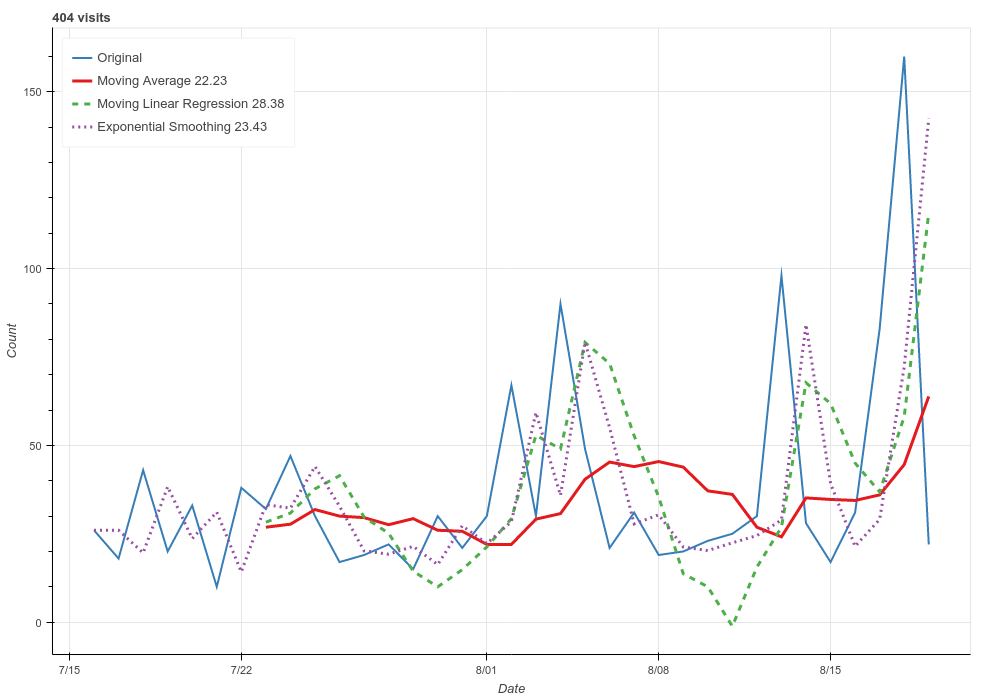
Forecasting:

For the top most visited website, three methods have been used. Moving Average which we presume a window of size 7 to get the forecast value of next day of the window. Exponential Smoothing with alpha value .08 and Regression. No specific method is ideal, but mostly on the picks the Moving Avarage is closer to the actual value.



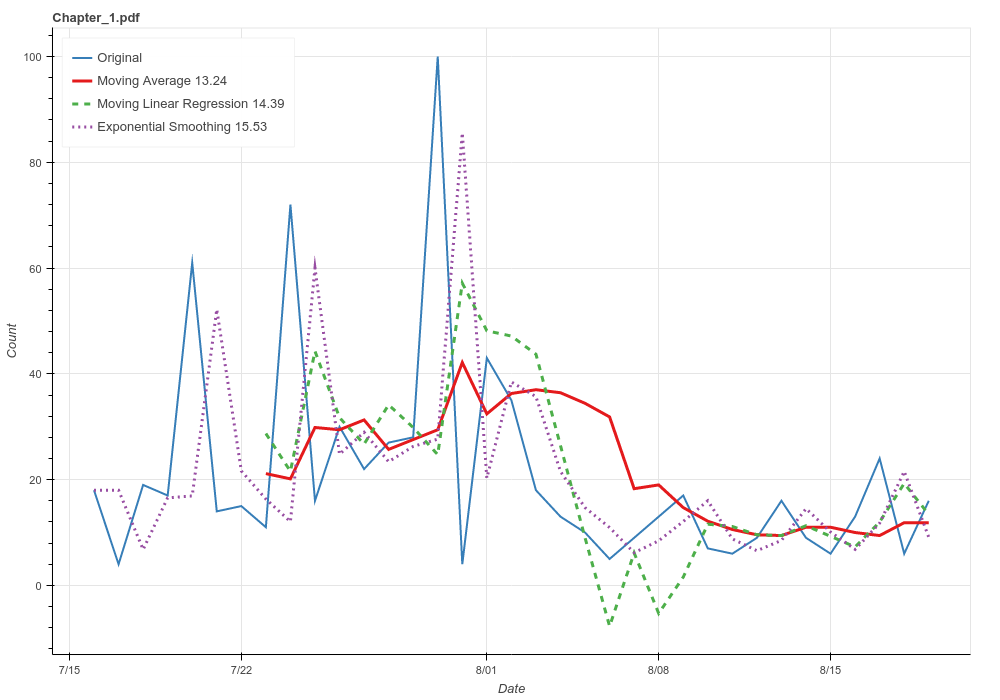
404

Forecasting for tis dataset shows that most of the time Exponential smoothing can be the most accurate forecasting method as the forecasted value is closer to actual value.



1st chapter

No specific clue☹



Unique

Sometimes regression and sometimes moving average

