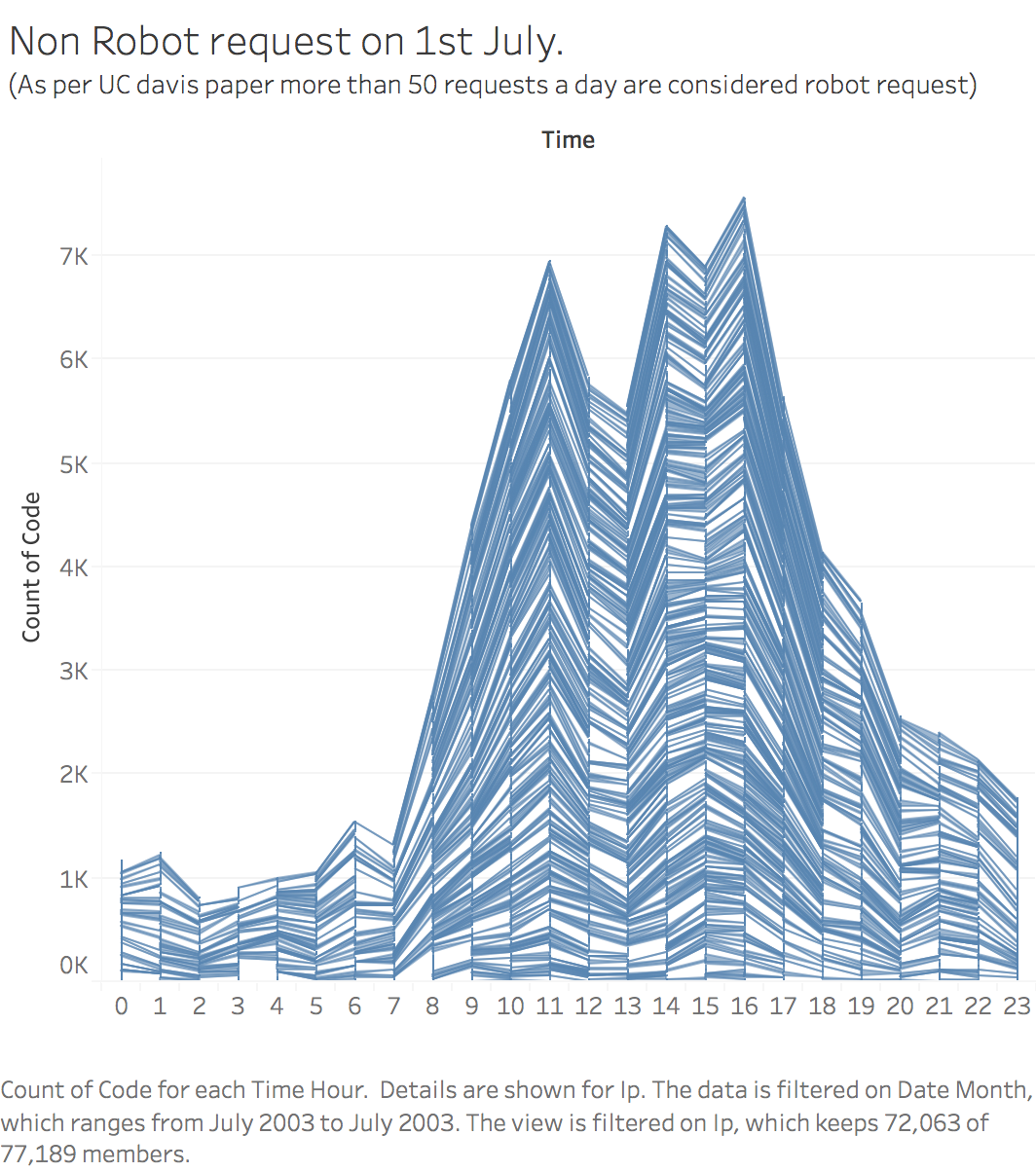


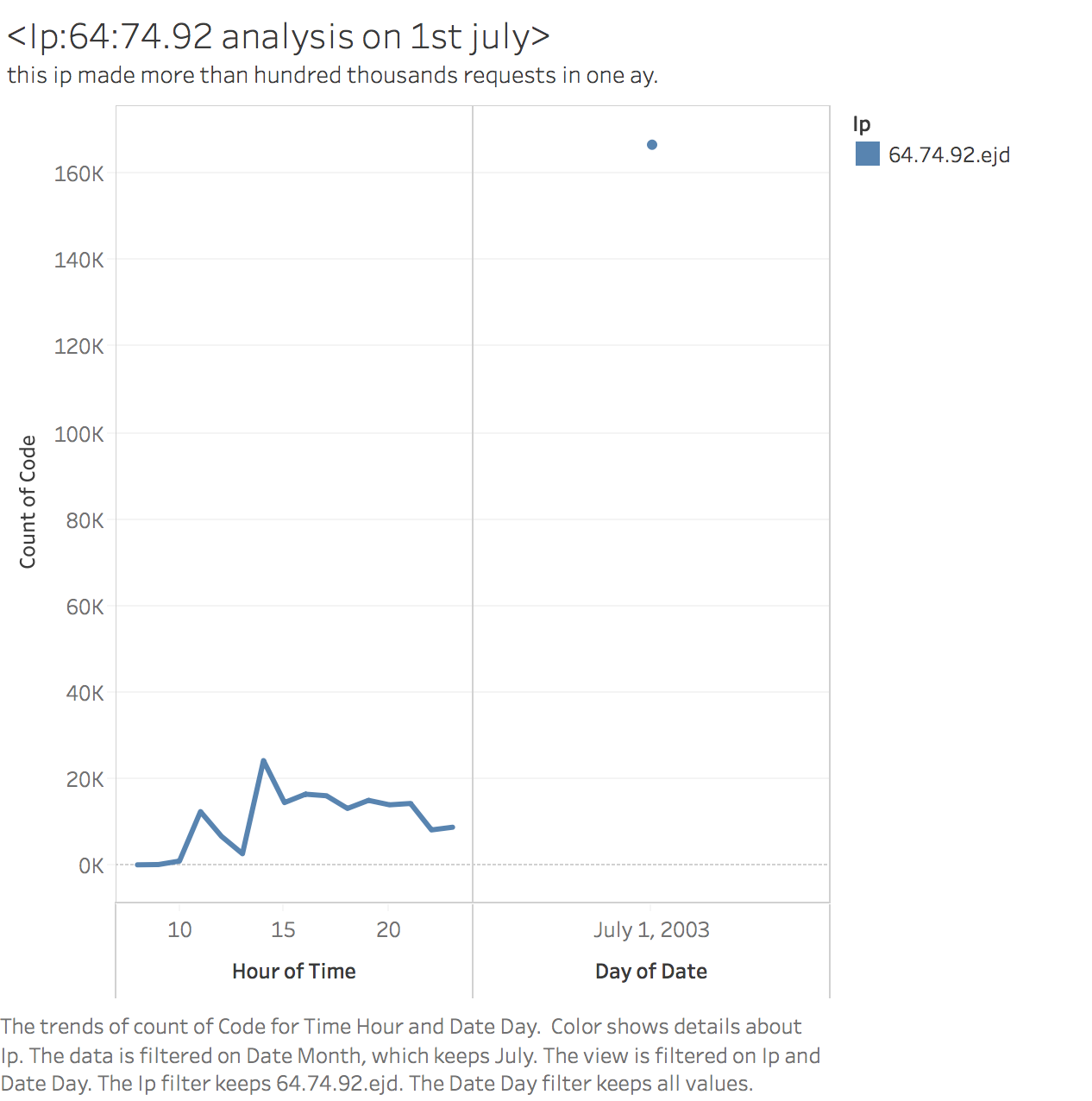
**Analysis**

**1 st july has highest number of hits recorded. Hence, analysis of traffic on 1st july. As per UC Davis paper on edgar datasets log files, more than 50 request per day from a particular IP are considered Robotic. The above graph shows IP’s with more than 50 requests over 24 hours of the day. The highest number of request were made during 2 pm-3 pm. The distribution is also more/dense during trading hours i.e 10am-5pm.**

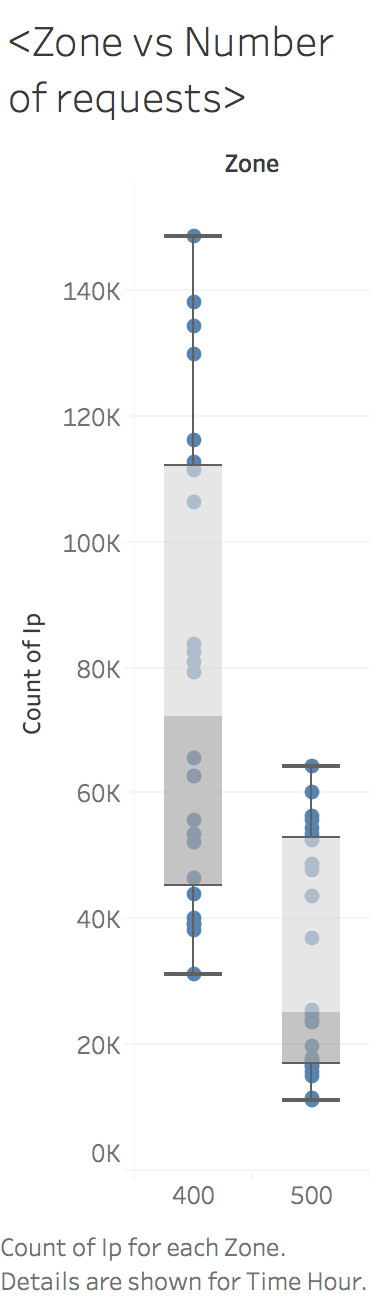


**Analysis**

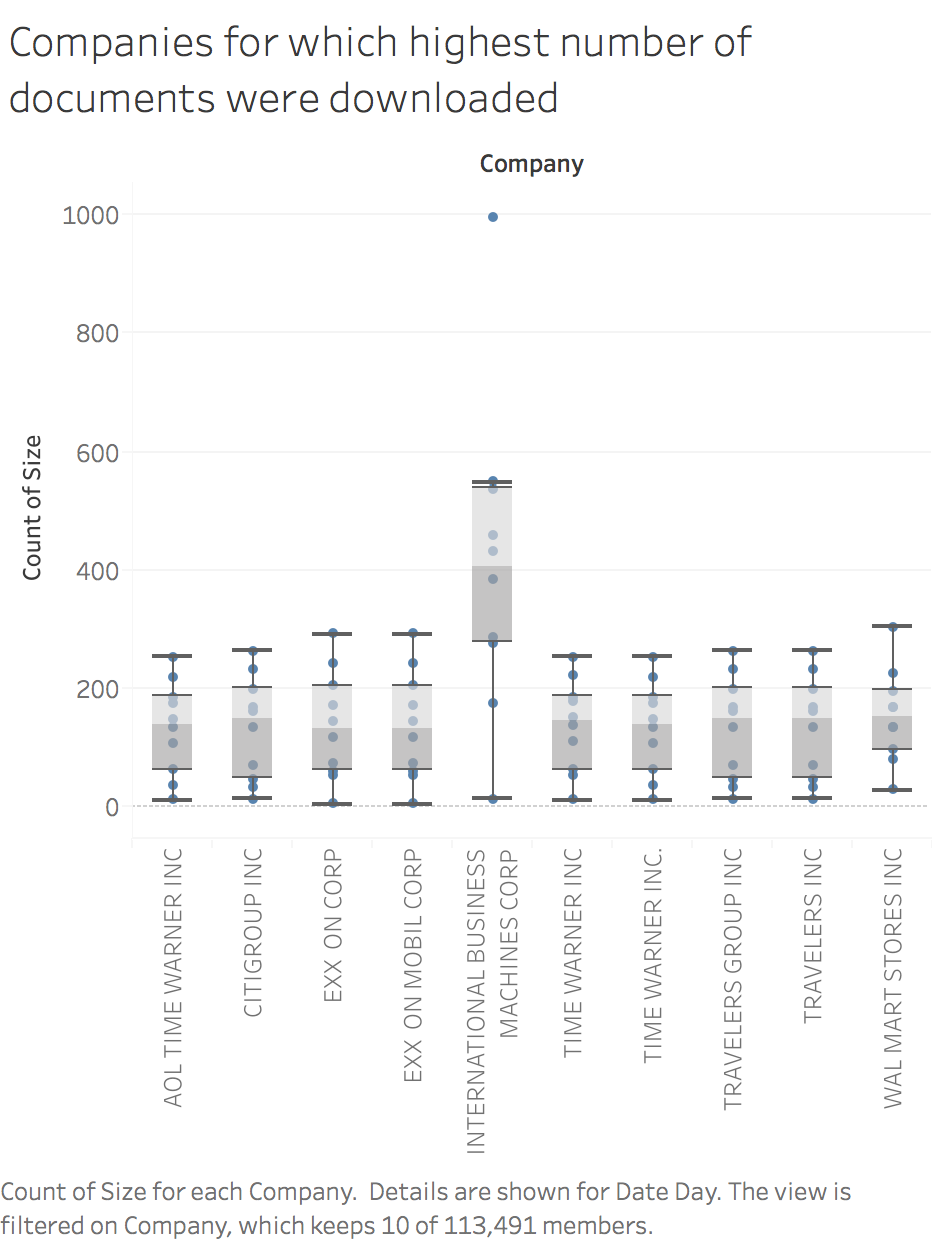
**1 st july has highest number of hits recorded. Hence, analysis of traffic on 1st july. As per UC Davis paper on edgar datasets log files, less than 50 request per day from a particular IP are considered Non Robotic. The above graph shows IP’s with less than 50 requests over 24 hours of the day. The distribution is evenly distributed across the trading hours whereas the graph for robotic request peaks at 2pm.**



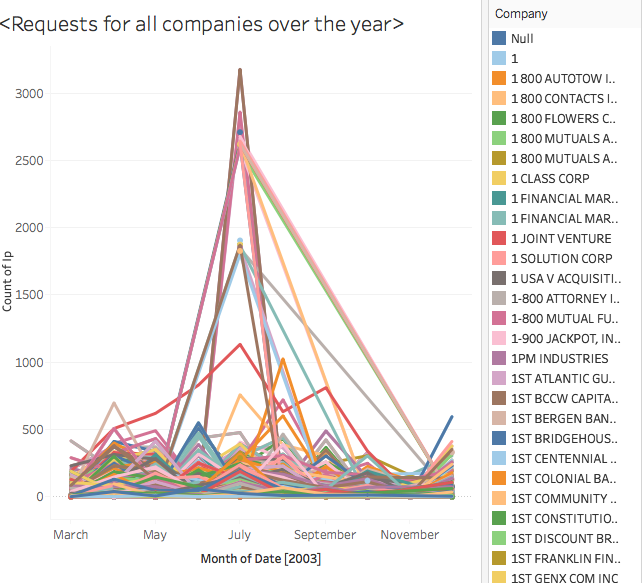
**IP 64.74.92.ejd made more than hundred thousands requests on 1st july. This was the highest hit count from any IP over the year. The above graph shows how the request were made by this IP over the whole day. The highest hits of around 23,000 were made at 2pm.**



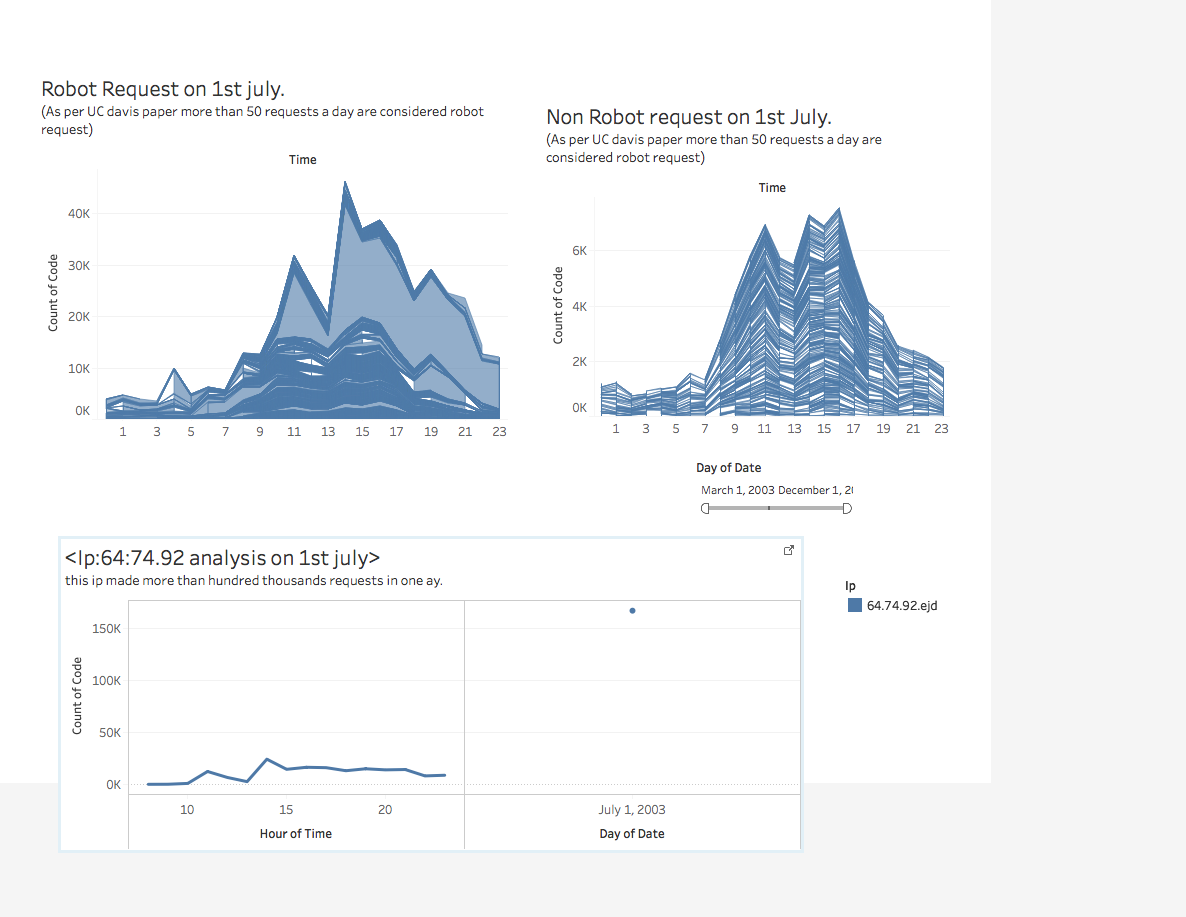
**The log file data analysis shows two zones i.e 400 and 500. The 400 zone received more hits than 500 zone over the 200 year. The above box plots shows the same analysis.**



**The above graph shows the list of companies for which highest number of documents( .i.e 10q or 10k or company performance documents ) were downloaded over 2003 . Highest download was for IBM.**

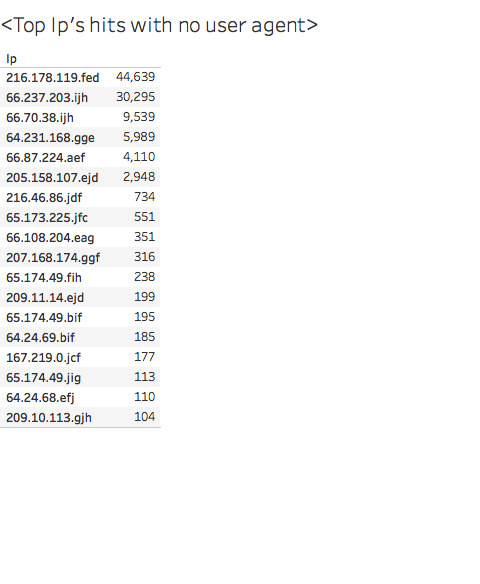


**We created a file with scrapping which contains CIK-CompanyName mapping. The following file was joined on CIK with 2003 log data set file in tableau. Following graph plots all companies and number of hits for them over the year. The outlier are Merill Lynch (highest hits on 1 s july) As per wikipedia In 2003, a employee was charged with civil securities fraud by the** [**U.S. Securities and Exchange Commission**](https://en.wikipedia.org/wiki/U.S._Securities_and_Exchange_Commission)**. The number of hits can be attributed to the news of fraud. The line in red shows IBM for which the hits were for the entire year. It can be attributed as most popular company for investments.**



Analysis for few companies as per number of hit over the year

**Aggregated dashboards**



**The no user agent can be attributed to request made by non human or robots. The list above shows all the IP’s over year with no user agent.**