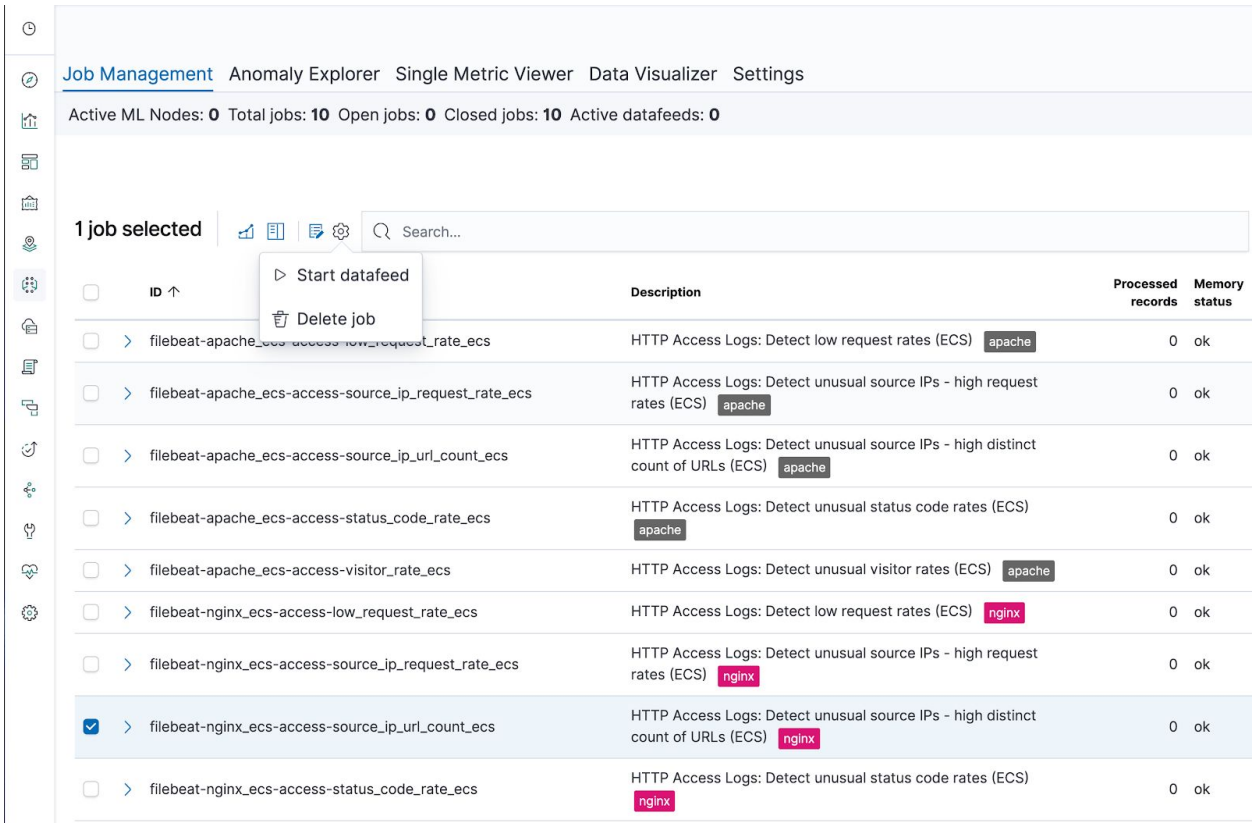
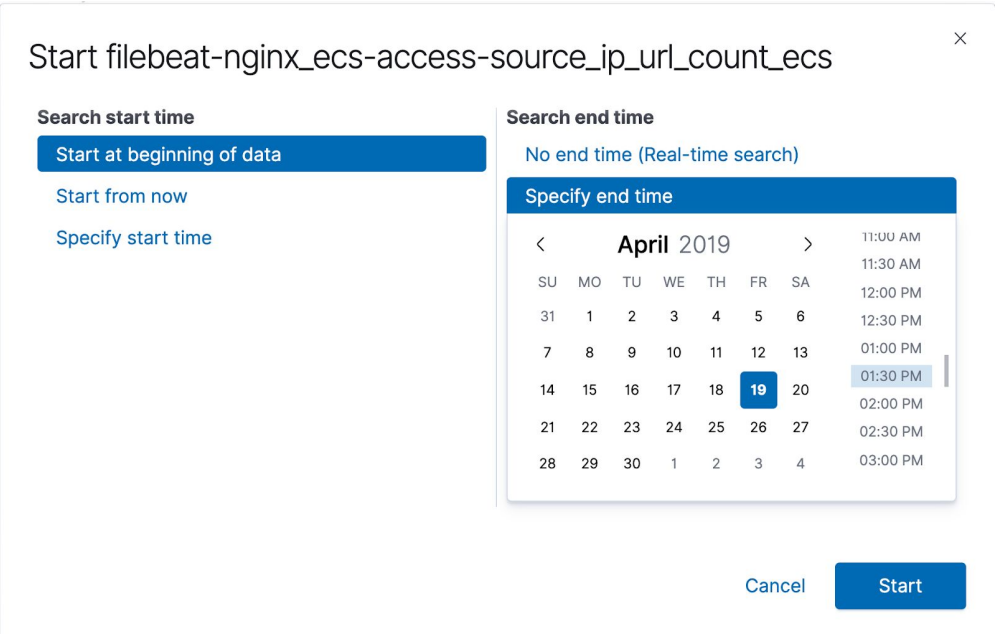
Elastic Observability Workshop

Lab 4 - Machine Learning

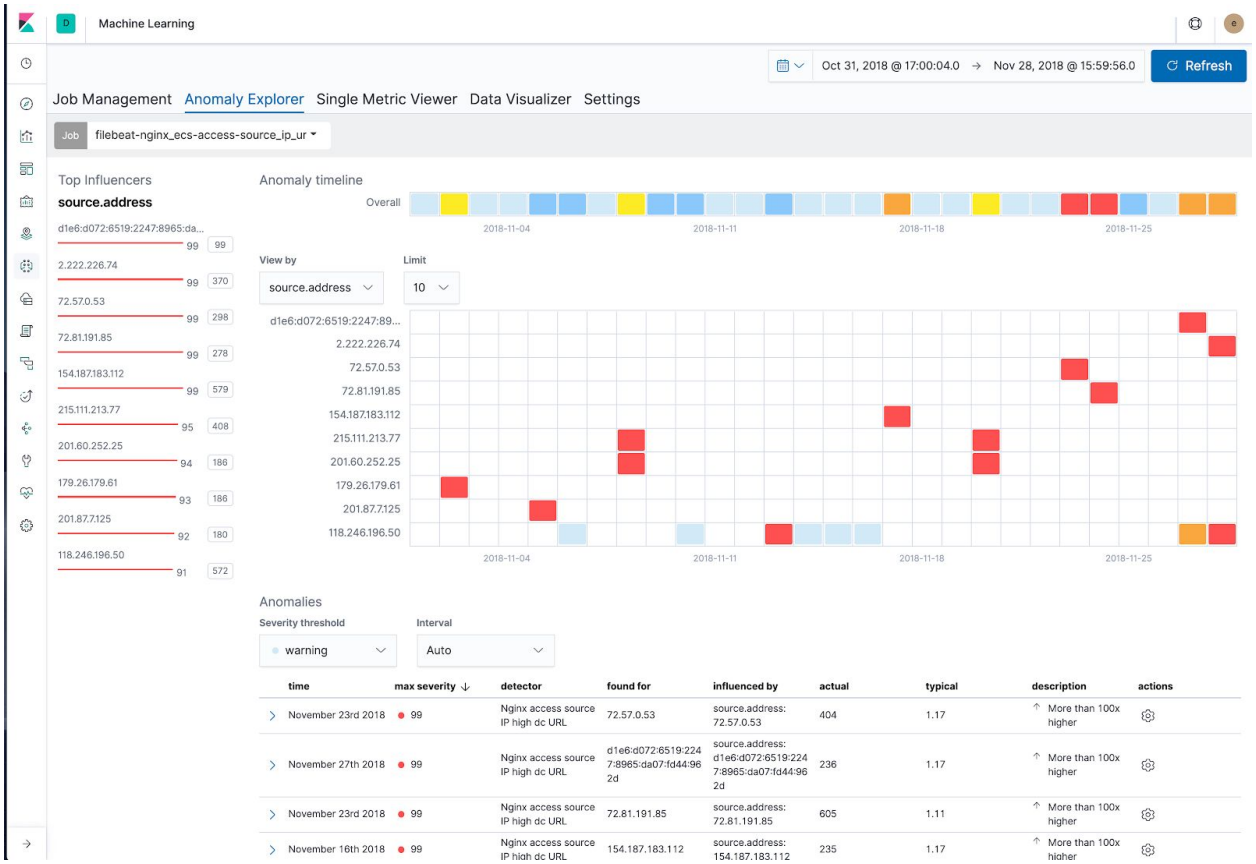
1) Click on **`Machine Learning`** app from the menu on left-hand side in Kibana. Do you recall the **`--setup`** command in the filebeat startup in Lab 2? All these jobs that we see here are a result of that - out of the box!!
Select **`filebeat-nginx_ecs-access-source_ip_url_count_ecs`** job from the list and start the data feed. Select Start at beginning of data and click on Start.





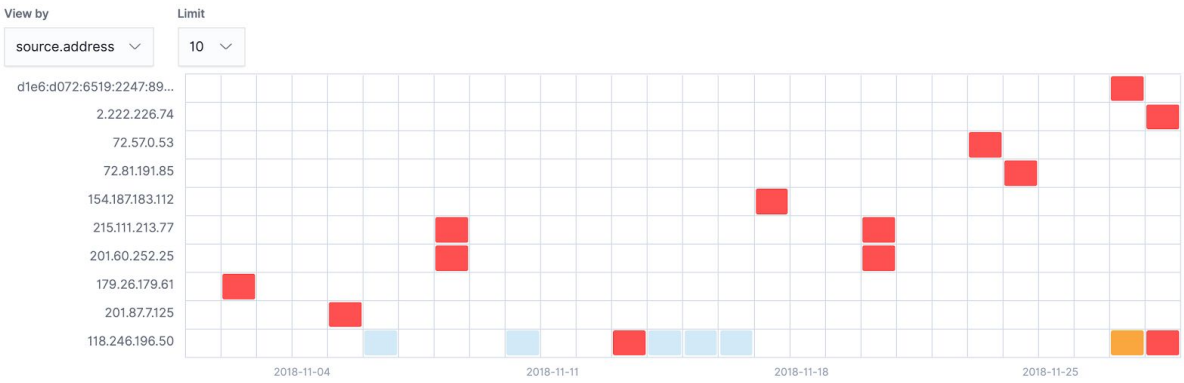
2) Wait until the datafeed state changes to **`stopped`** (about 984,887 processed documents) and click on the results icon to view the output of the job in Anomaly Explorer. (Deselect the job to enable the in-row action items)

3) You will see in the **`Anomaly Explorer`** view where the results are broken down by particular IPs. The influencers that you see are statistically significant elements in our data set contributing to the anomaly. You have an option to select a particular field in your data as influencer to understand its impact on the anomalies when you're creating a Machine Learning job.



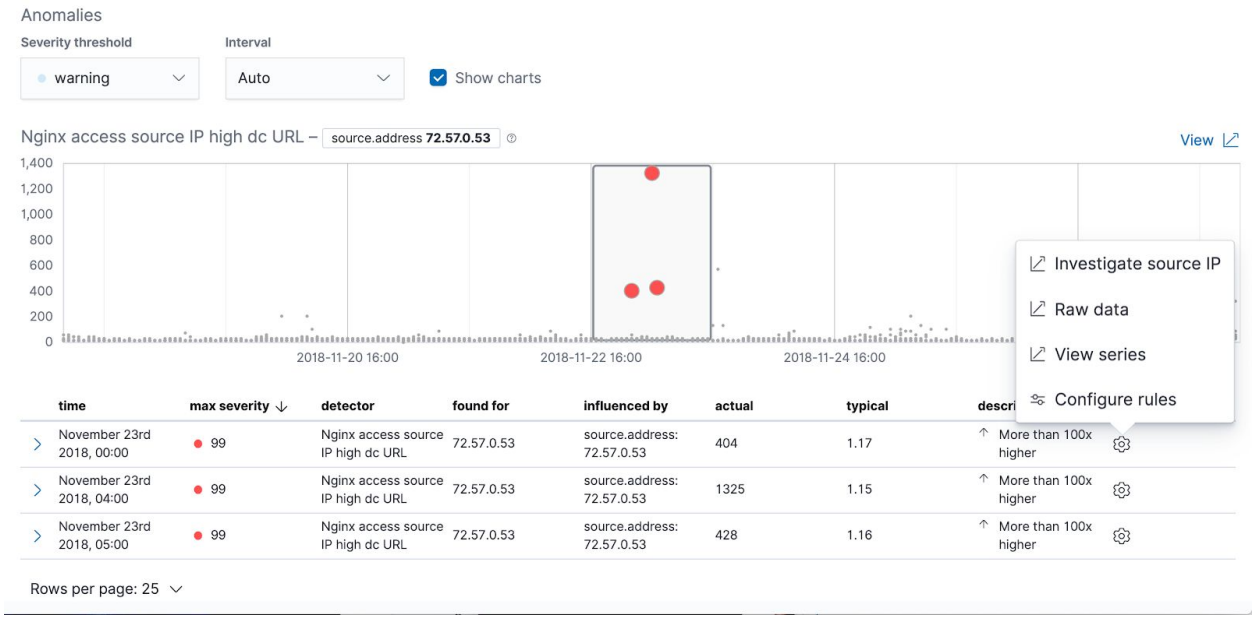
On the top in **`Anomaly Timeline`** you see overall result for the job and what kind of anomalies were detected.

and in the matrix below the **`Anomaly Timeline`** you'd see the result of the job for a particular IP.

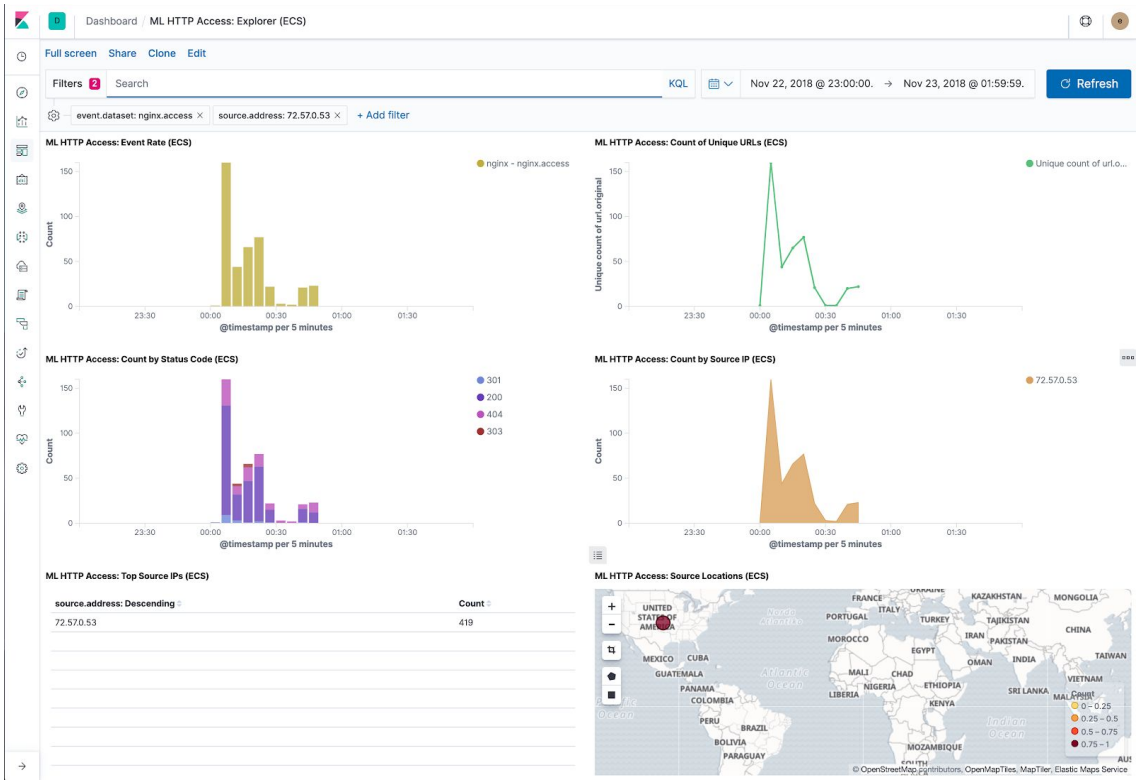


Hover over the overall results to see the score across the whole population and over particular IPs contributing to that same score. When the score for IP `118.246.196.50` is `90` and overall anomaly score in the `Anomaly Timeline` view is only `19` as opposed to for IP `72.57.0.53` with a score of `99` and overall anomaly score of `85`. Can you see why that might be?

3) We have the ability to link the anomaly results of ML job with particular views when we setup the job. When opening the views the metadata from the job can be passed to it. Click on “actions” next to top anomaly and click on `URL Explorer`.



Review the `Investigate Source IP` dashboard, note how its view was customized based on the IP for the anomaly (i.e. on the metadata from the ML job on the previous screen)



Summary: We took a brief look at how Machine Learning in the Elastic stack can be leveraged to identify anomalies in your data. In the above case we looked at nginx access logs and an ML job configured to find out anomalies in count of requests from IP was quickly able to identify outliers in the access logs. Imagine how hard would it be to come up with all the alerts for all the different IP addresses? Filebeat --setup command created a lot of other jobs as well out of the box for nginx module, feel free to look at them.