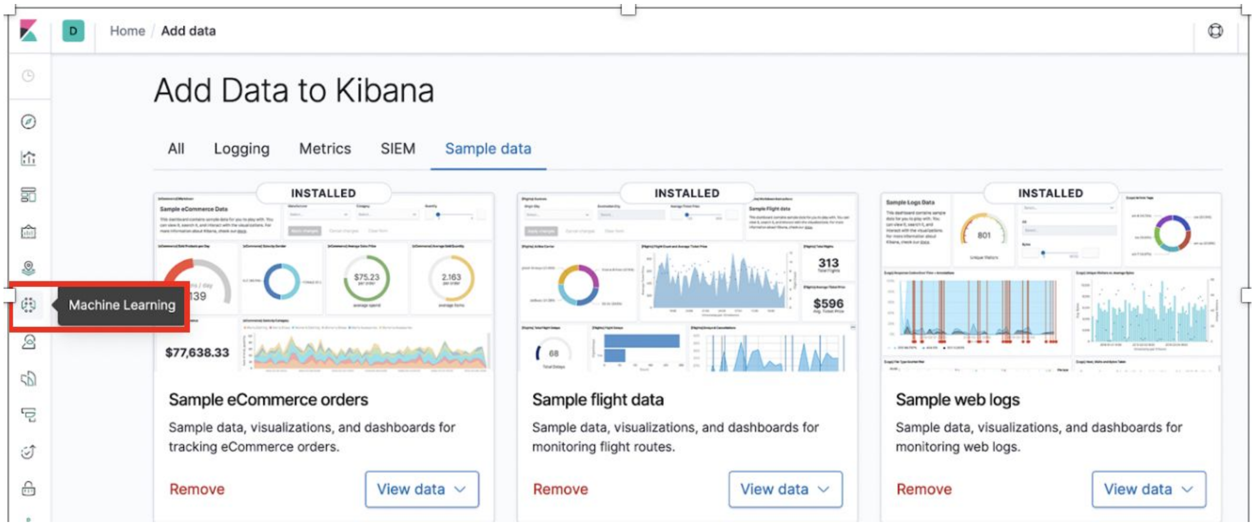
Lab 1 - Single and Multi-Metric Jobs, Forecasting

In this lab, we will be performing the following:

a. Set up a single metric job
b. Perform forecasting
c. Set up a multi-metric job
d. Add a custom URL to the multi-metric job

A - Set Up a Single Metric Job

1. Click on the “Machine Learning” link on the left side of Kibana.



Add Data to Kibana

Sample eCommerce orders

Sample data, visualizations, and dashboards for tracking eCommerce orders.

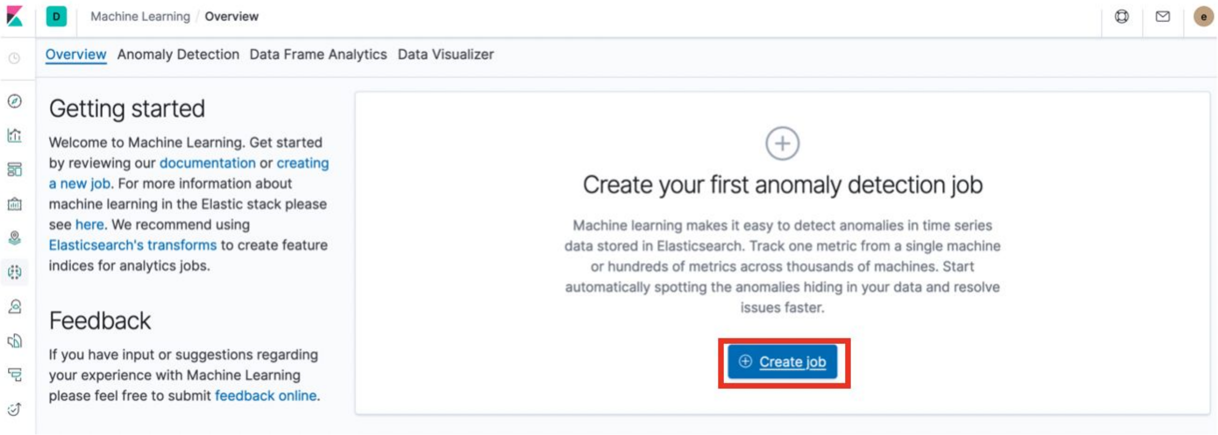
Sample flight data

Sample data, visualizations, and dashboards for monitoring flight routes.

Sample web logs

Sample data, visualizations, and dashboards for monitoring web logs.

2. Click on the “Create Job” link to set up your first Machine Learning job!



Getting started

Welcome to Machine Learning. Get started by reviewing our documentation or creating a new job. For more information about machine learning in the Elastic stack please see here. We recommend using Elasticsearch's transforms to create feature indices for analytics jobs.

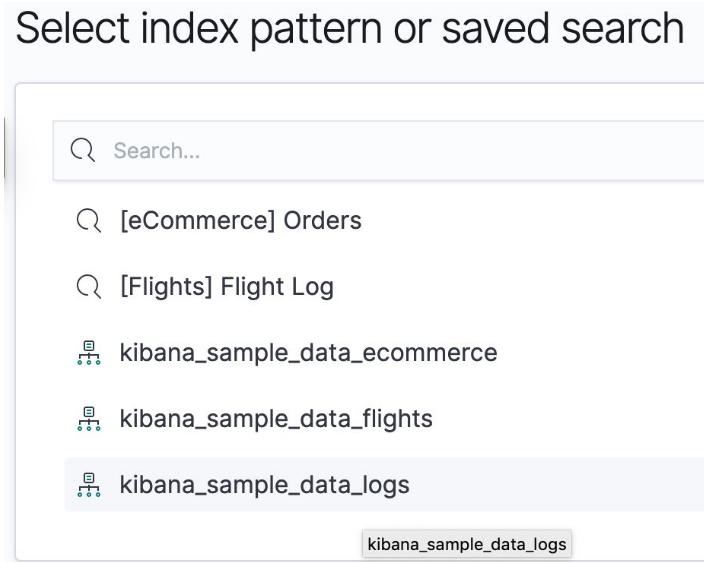
Feedback

If you have input or suggestions regarding your experience with Machine Learning please feel free to submit feedback online.

Create your first anomaly detection job

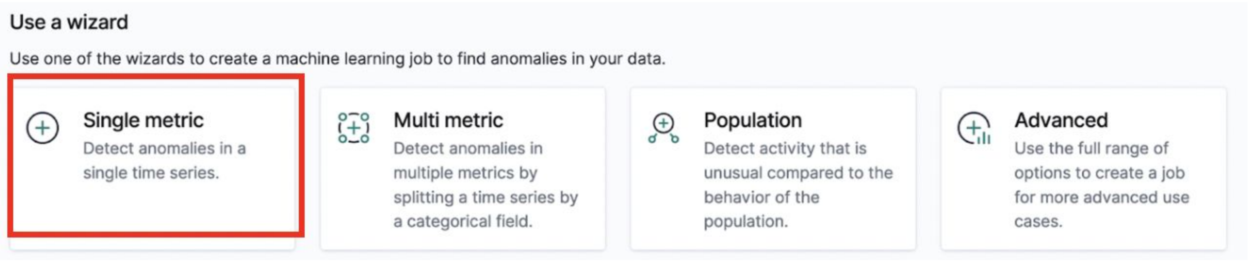
Machine learning makes it easy to detect anomalies in time series data stored in Elasticsearch. Track one metric from a single machine or hundreds of metrics across thousands of machines. Start automatically spotting the anomalies hiding in your data and resolve issues faster.

3. Select “kibana_sample_data_logs” index.



Select index pattern or saved search

4. Select the link to create a “Single metric” job



Use a wizard

Use one of the wizards to create a machine learning job to find anomalies in your data.

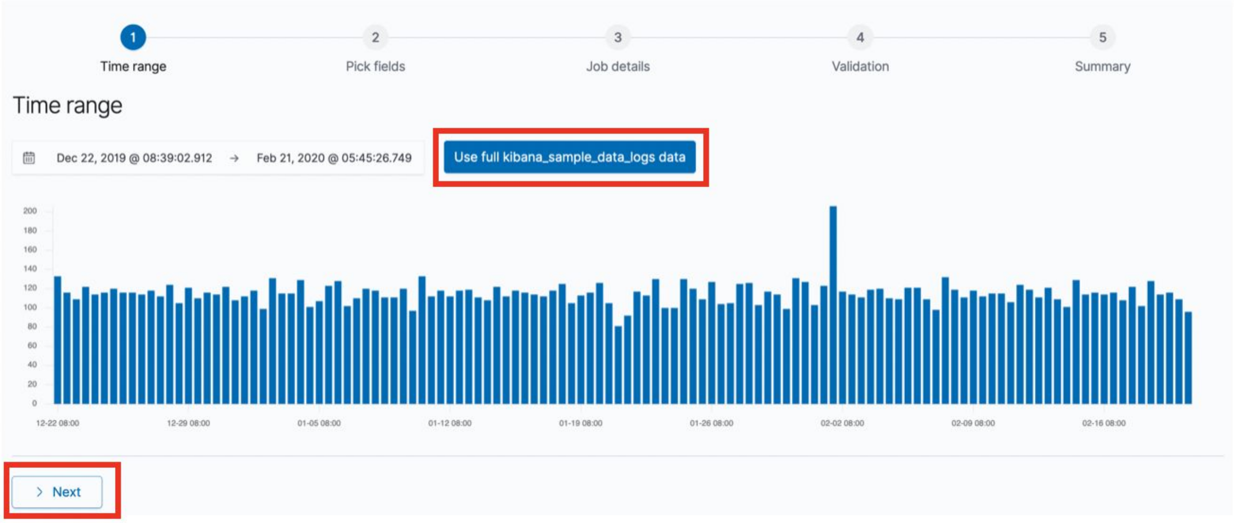
Single metric Detect anomalies in a single time series.

Multi metric Detect anomalies in multiple metrics by splitting a time series by a categorical field.

Population Detect activity that is unusual compared to the behavior of the population.

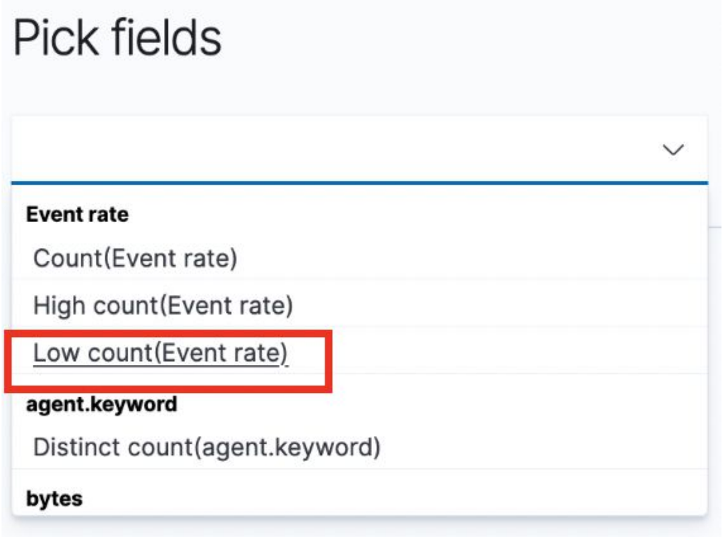
Advanced Use the full range of options to create a job for more advanced use cases.

5. Click on the “Use full kibana_sample_data_logs data” button, then click on “Next”

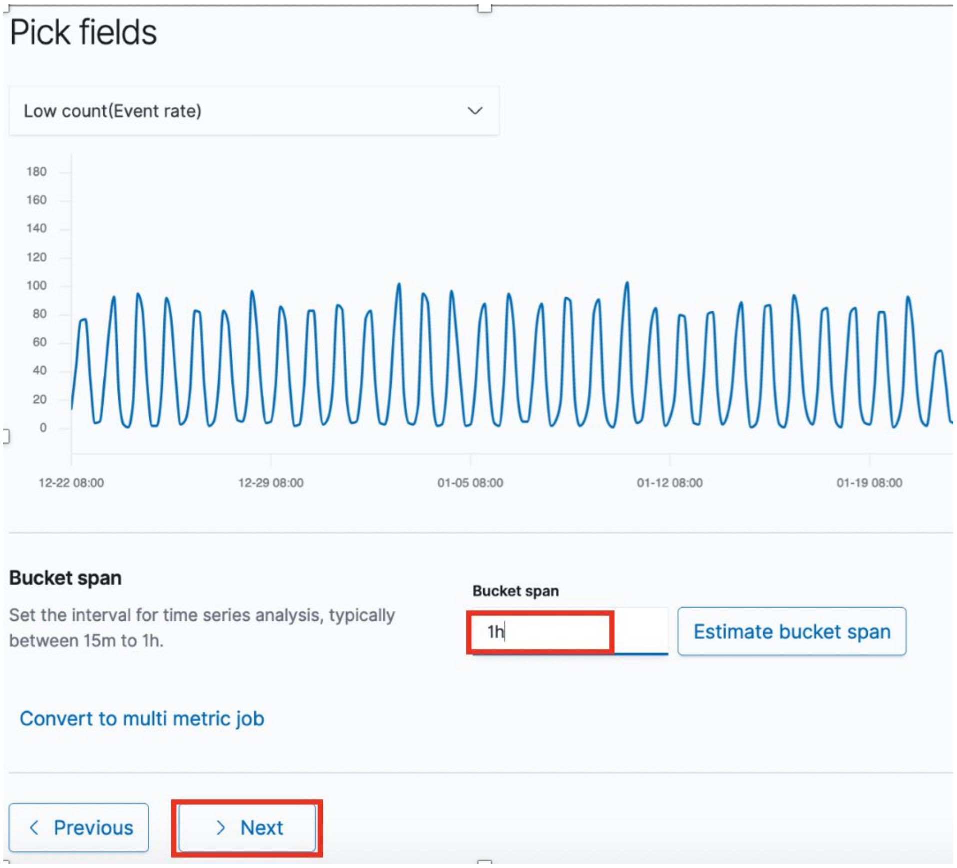


Dec 22, 2019 @ 08:39:02.912 → Feb 21, 2020 @ 05:45:26.749

6. Select “low_count” function



7. Enter 1h for the bucket span and click on “Next”


Pick fields

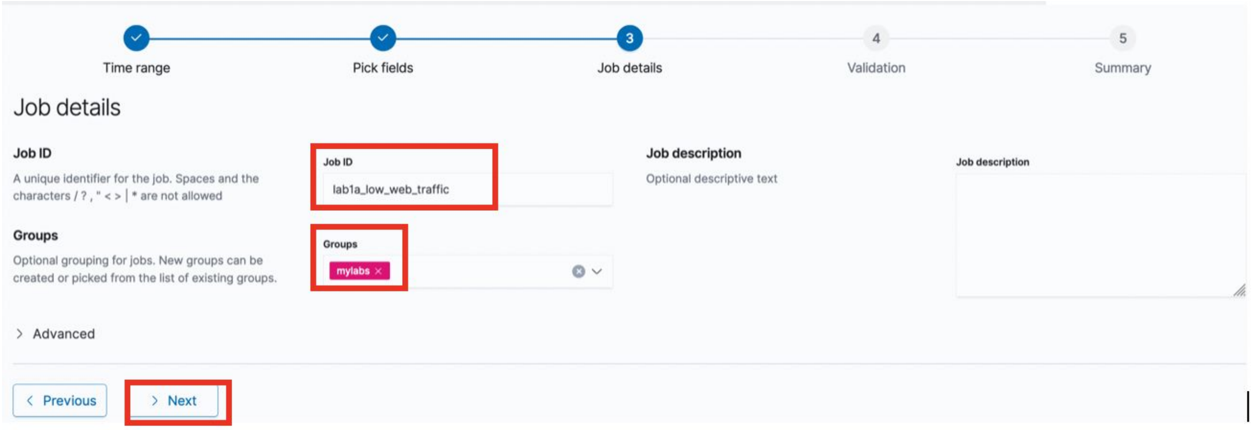
Low count(Event rate)

Bucket span

Set the interval for time series analysis, typically between 15m to 1h.

Bucket span: 1h

8. Enter “lab1a_low_web_traffic” as the Job ID, and “mylabs” as the Group name, then click on the “Next” button


Job details

Job ID

A unique identifier for the job. Spaces and the characters / ? , " < > | * are not allowed

Job ID: lab1a_low_web_traffic

Groups

Optional grouping for jobs. New groups can be created or picked from the list of existing groups.

Groups: mylabs

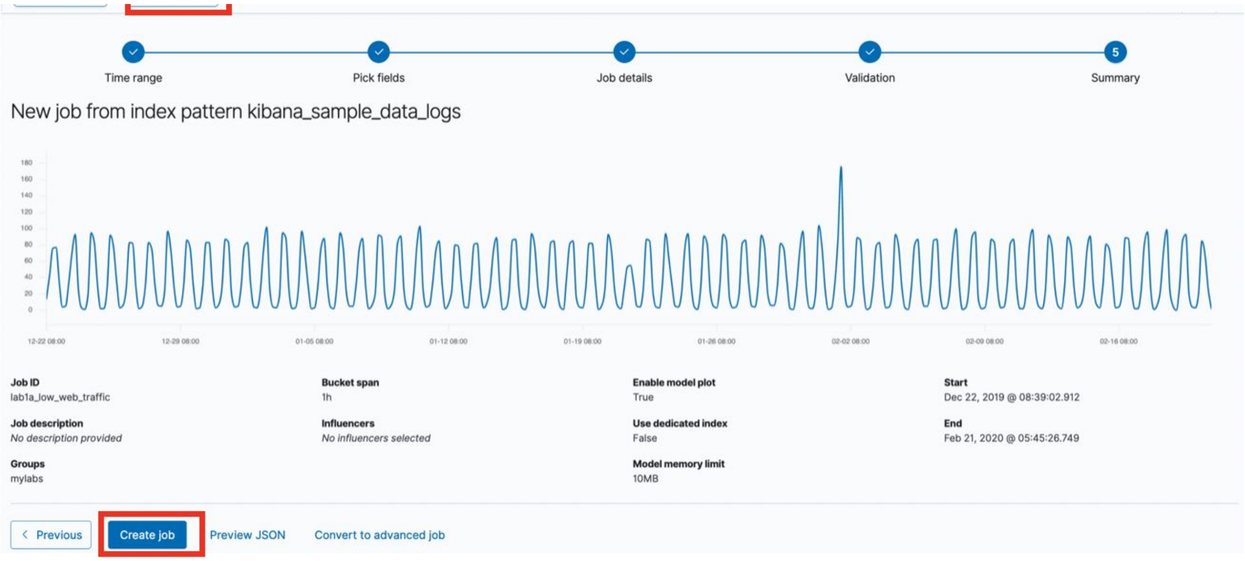
Job description

Optional descriptive text

Job description:

9. The data validation step should pass without problems. Click on the “Next” step to proceed.

10. Review the job configuration, and click on the “Create Job” button to start the ML job.



New job from index pattern kibana_sample_data_logs

Job ID
lab1a_low_web_traffic

Job description
No description provided

Groups
mylabs

Bucket span
1h

Influencers
No influencers selected

Enable model plot
True

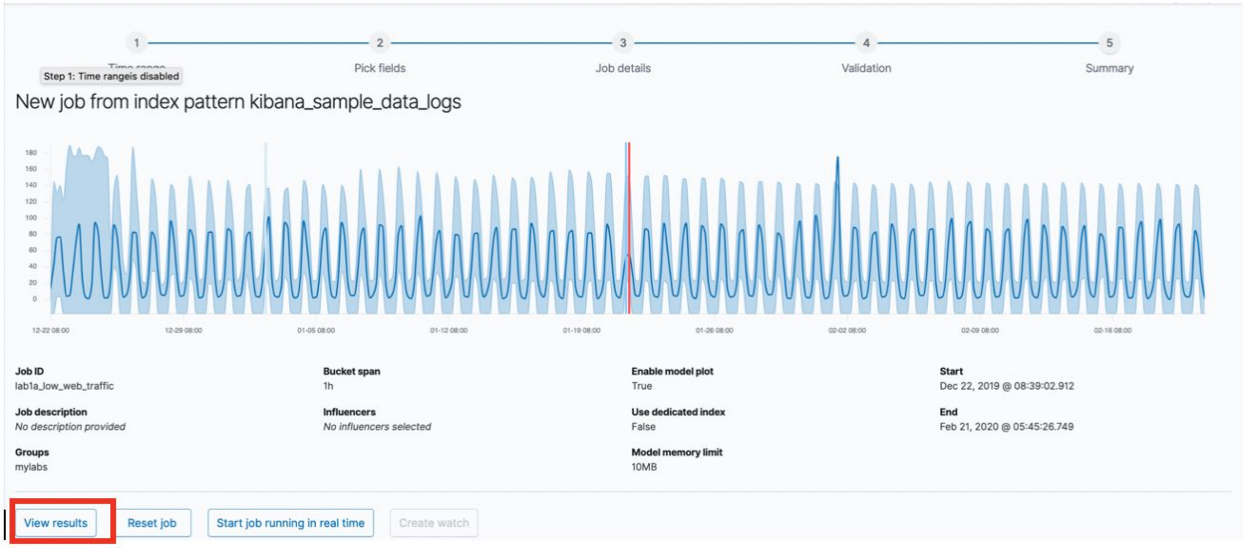
Use dedicated index
False

Model memory limit
10MB

Start
Dec 22, 2019 @ 08:39:02.912

End
Feb 21, 2020 @ 05:45:26.749

11. The job should take seconds to complete. Once done, please click on the “View Job” button to view the results:



New job from index pattern kibana_sample_data_logs

Job ID
lab1a_low_web_traffic

Job description
No description provided

Groups
mylabs

Bucket span
1h

Influencers
No influencers selected

Enable model plot
True

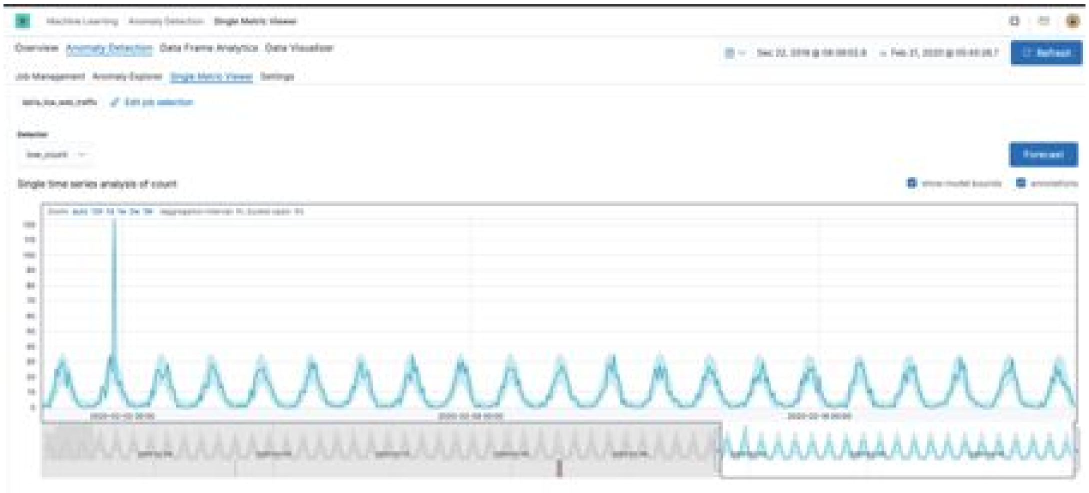
Use dedicated index
False

Model memory limit
10MB

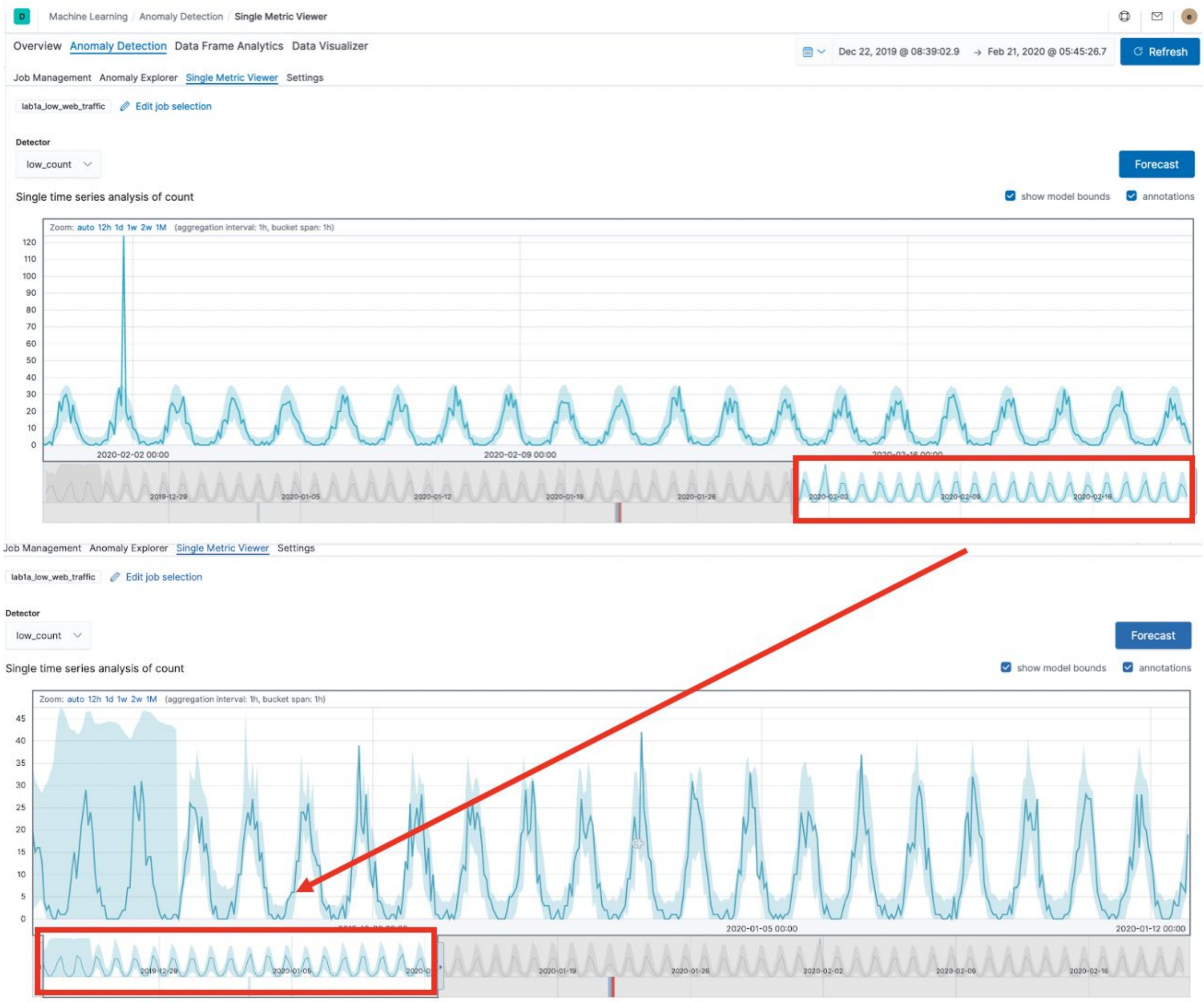
Start
Dec 22, 2019 @ 08:39:02.912

End
Feb 21, 2020 @ 05:45:26.749

12. This is how the results would look like:



13. You can drag the “timeline” bar at the bottom to the beginning of the time period, to see how ML “built” the “model” (after about 3 cycles at the beginning of the timeframe)



Dec 22, 2019 @ 08:39:02.9 → Feb 21, 2020 @ 05:45:26.7

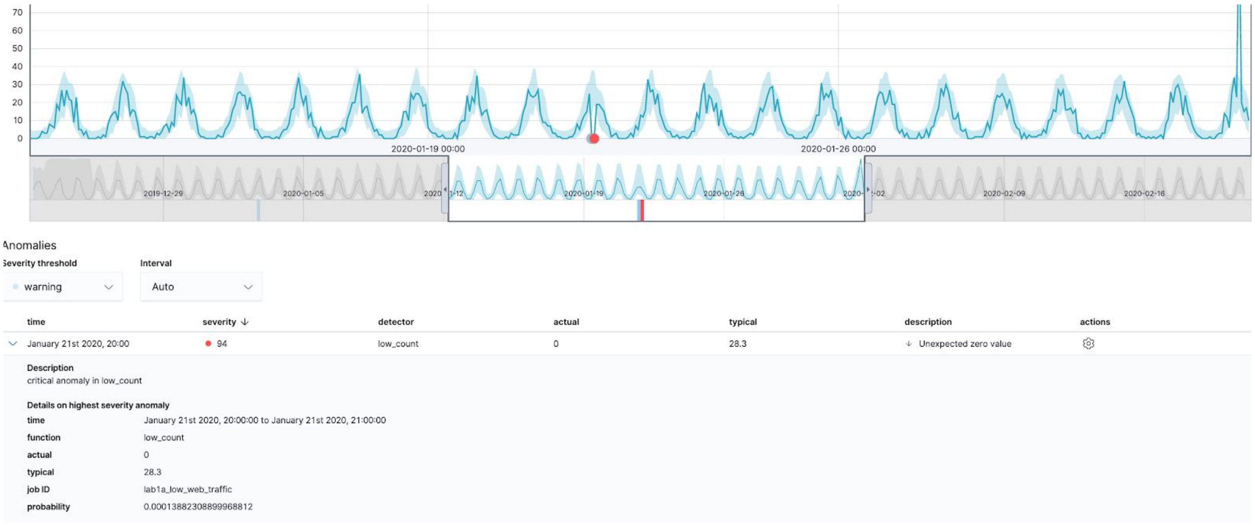
lab1a_low_web_traffic

Detector: low_count

Single time series analysis of count

Zoom: auto 12h 1d 1w 2w 1M (aggregation interval: 1h, bucket span: 1h)

14. Next, drag the timeline to the area with a “red” line to check out the anomaly found:



Anomalies

Severity threshold: warning

Interval: Auto

time	severity ↓	detector	actual	typical	description	actions
January 21st 2020, 20:00	● 94	low_count	0	28.3	↓ Unexpected zero value	⚙

Description
critical anomaly in low_count

Details on highest severity anomaly

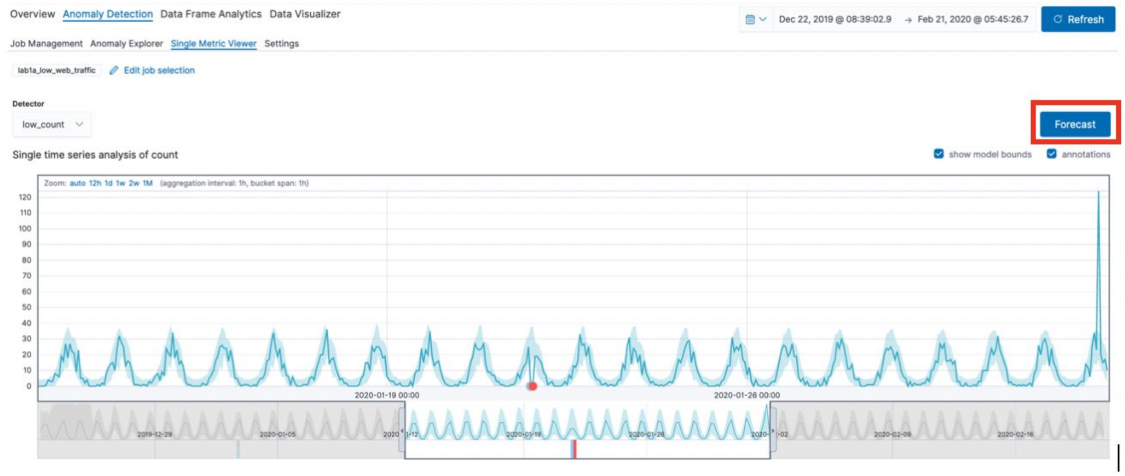
time	January 21st 2020, 20:00:00 to January 21st 2020, 21:00:00
function	low_count
actual	0
typical	28.3
job ID	lab1a_low_web_traffic
probability	0.00013882308899968812

Note that detailed information of the anomaly can be found in the anomalies at the bottom of the page. The anomaly was given a severity of 94. The expected count according to the model (for that time period) was 28.3, however the actual count was 0, hence the high severity.

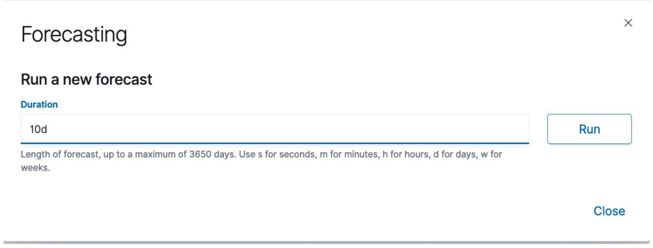
15. Note also that:
    - The drop in traffic was given critical severity score
    - Spikes in traffic on the other hand were not anomalous, given that we were looking only for anomalies on the low side (“low count" function)

B - Run a Forecast

1. Continuing on the ML Single Metric Job Results Page, note that there is a “Forecast” button at the top right hand side. Click on the button.



2. Enter the duration which you would like the forecasting to be calculated for and click on the “Run” button:



Forecasting

Run a new forecast

Duration

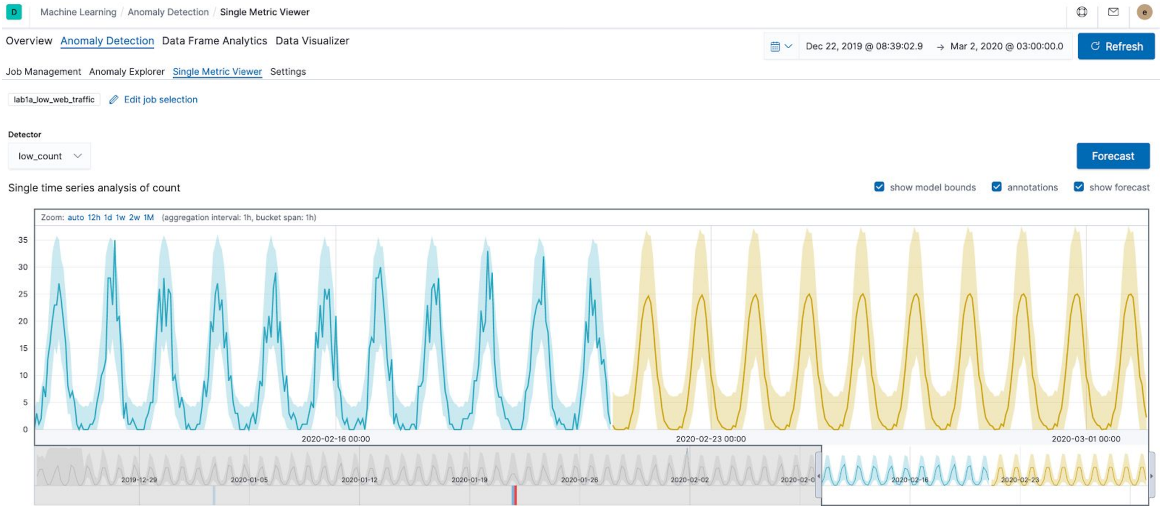
10d

Run

Length of forecast, up to a maximum of 3650 days. Use s for seconds, m for minutes, h for hours, d for days, w for weeks.

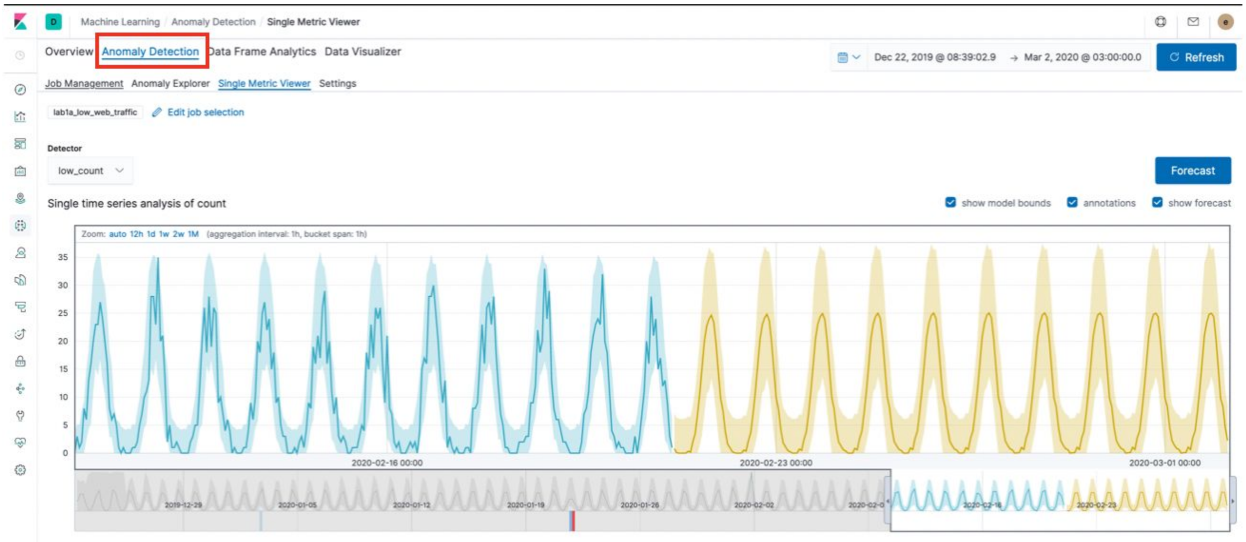
Close

3. The forecasted results are in Yellow lines:

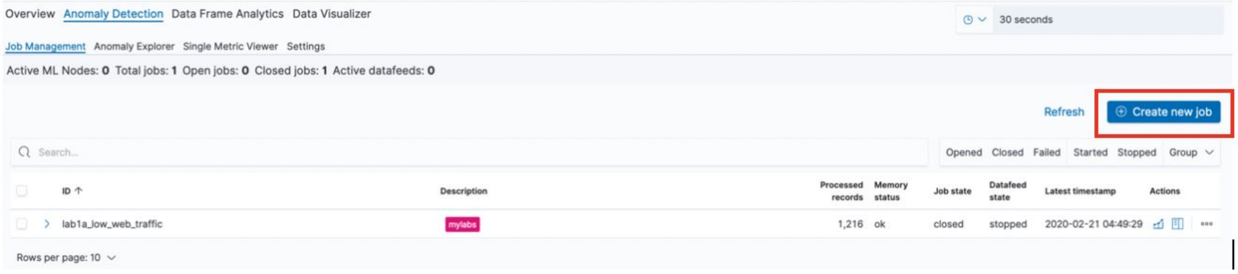


C - Run a Multi-Metric Job

1. Click on the “Anomaly Detection” link at the top of the page in Kibana.



2. Click on the link to “Create new job”

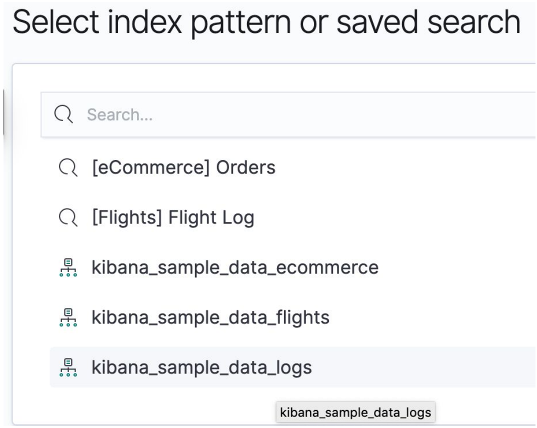


Active ML Nodes: **0** Total jobs: **1** Open jobs: **0** Closed jobs: **1** Active datafeeds: **0**

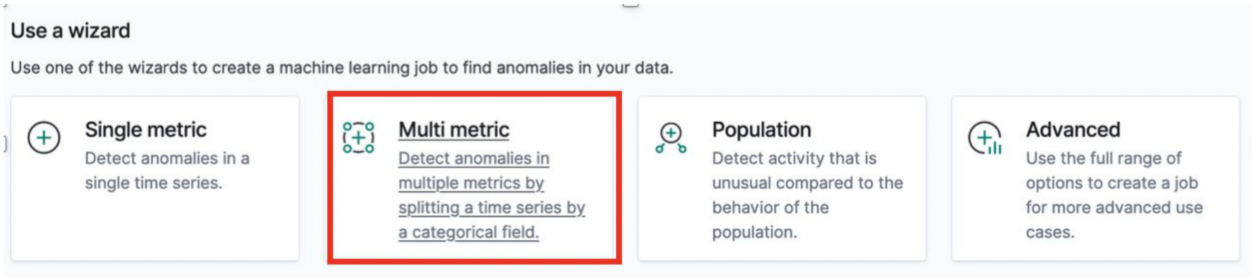
ID ↑	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
lab1a_low_web_traffic	mylabs	1,216	ok	closed	stopped	2020-02-21 04:49:29	

Rows per page: 10

3. As before, select the “kibana_sample_data_logs” index.



4. This time, select the link to create a “Multi metric” job



Use a wizard

Use one of the wizards to create a machine learning job to find anomalies in your data.

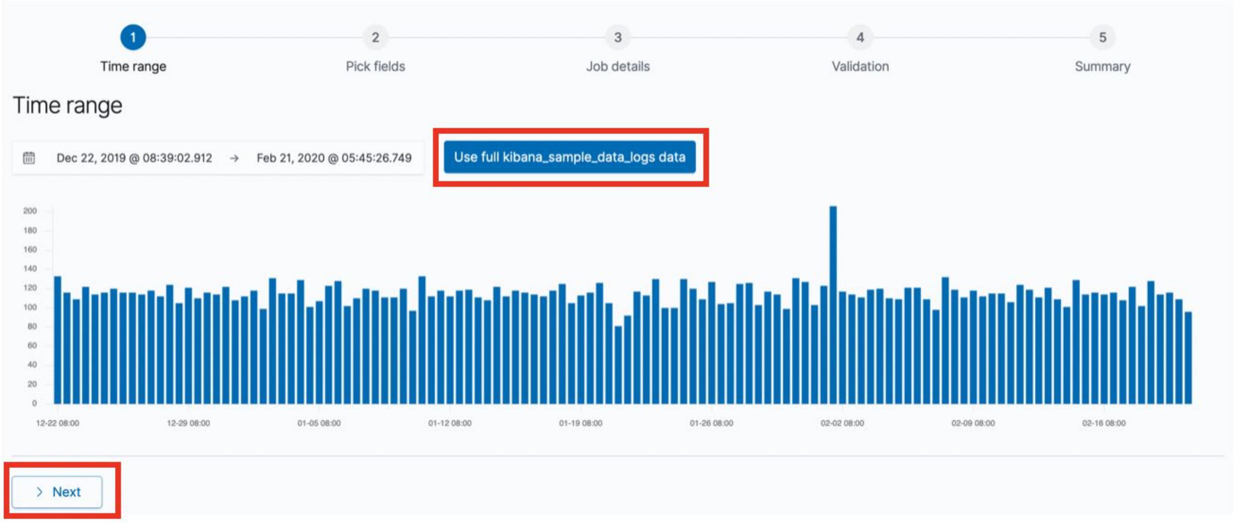
Single metric
Detect anomalies in a single time series.

Multi metric
Detect anomalies in multiple metrics by splitting a time series by a categorical field.

Population
Detect activity that is unusual compared to the behavior of the population.

Advanced
Use the full range of options to create a job for more advanced use cases.

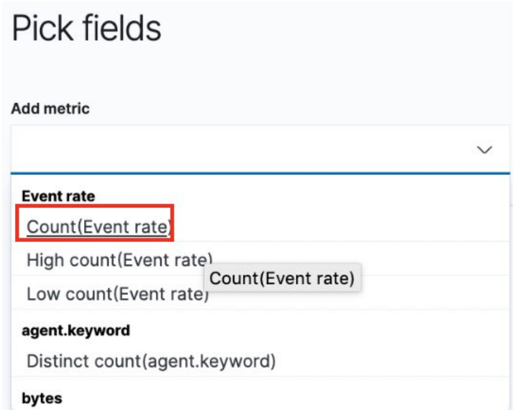
5. Click on the “Use full kibana_sample_data_logs data” button, then click on “Next”



Time range

Dec 22, 2019 @ 08:39:02.912 → Feb 21, 2020 @ 05:45:26.749

6. Select "Event rate" (count) function



Pick fields

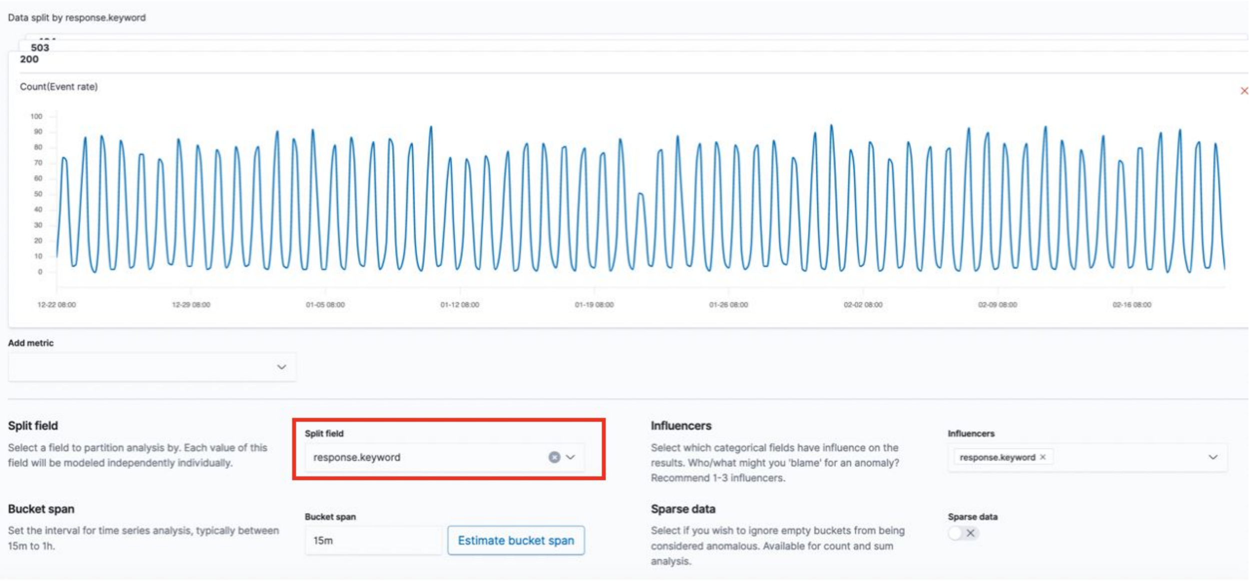
Add metric

Event rate
Count(Event rate)
High count(Event rate)
Low count(Event rate)

agent.keyword
Distinct count(agent.keyword)

bytes

7. Under “Split Field”, split the data on **response.keyword** (the HTTP status code)



Split field

Select a field to partition analysis by. Each value of this field will be modeled independently individually.

Split field: response.keyword

Influencers

Select which categorical fields have influence on the results. Who/what might you 'blame' for an anomaly? Recommend 1-3 influencers.

Influencers: response.keyword ×

Bucket span

Set the interval for time series analysis, typically between 15m to 1h.

Bucket span: 15m

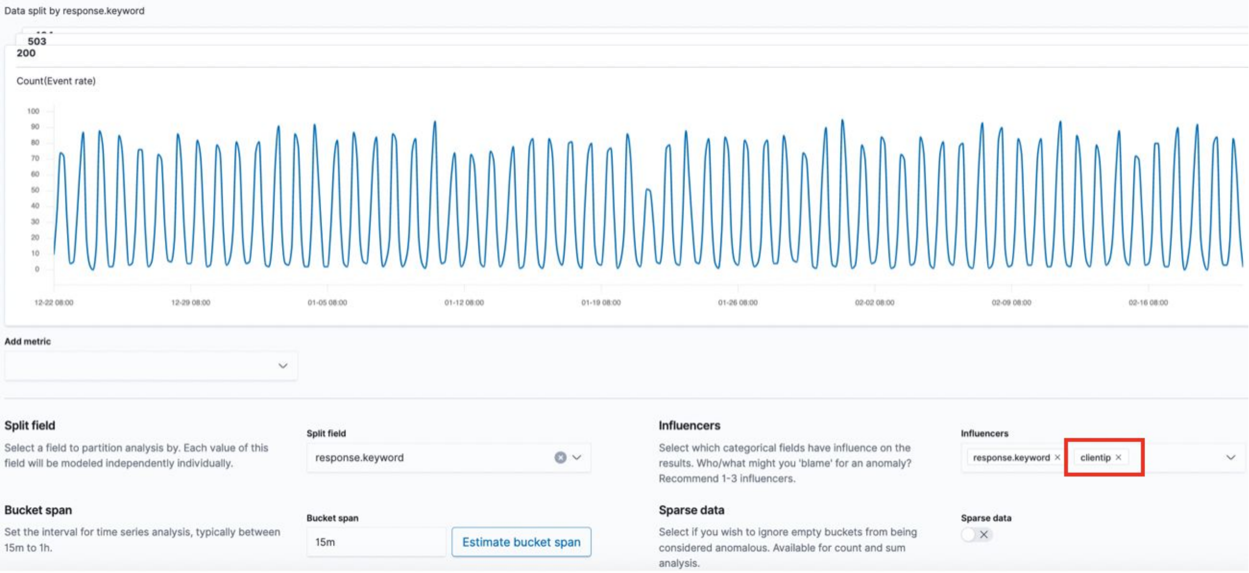
Estimate bucket span

Sparse data

Select if you wish to ignore empty buckets from being considered anomalous. Available for count and sum analysis.

Sparse data

8. Under Influencers, add "clientip" as an additional Key Field



Split field

Select a field to partition analysis by. Each value of this field will be modeled independently individually.

Split field: response.keyword

Influencers

Select which categorical fields have influence on the results. Who/what might you 'blame' for an anomaly? Recommend 1-3 influencers.

Influencers: response.keyword × clientip ×

Bucket span

Set the interval for time series analysis, typically between 15m to 1h.

Bucket span: 15m

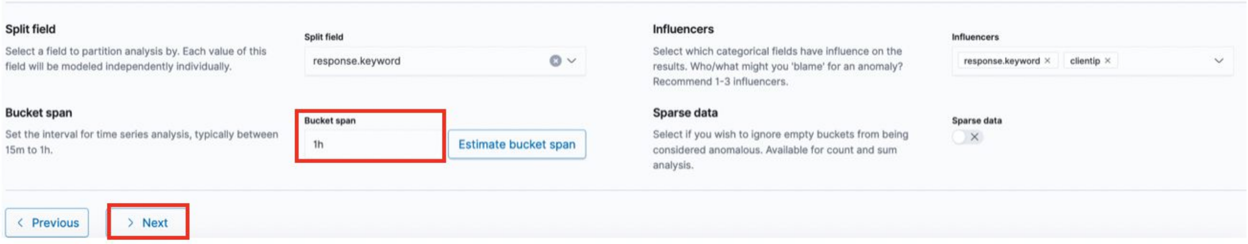
Estimate bucket span

Sparse data

Select if you wish to ignore empty buckets from being considered anomalous. Available for count and sum analysis.

Sparse data

9. Enter 1h for the bucket span and click on “Next”



Split field

Select a field to partition analysis by. Each value of this field will be modeled independently individually.

Split field: response.keyword

Influencers

Select which categorical fields have influence on the results. Who/what might you 'blame' for an anomaly? Recommend 1-3 influencers.

Influencers: response.keyword × clientip ×

Bucket span

Set the interval for time series analysis, typically between 15m to 1h.

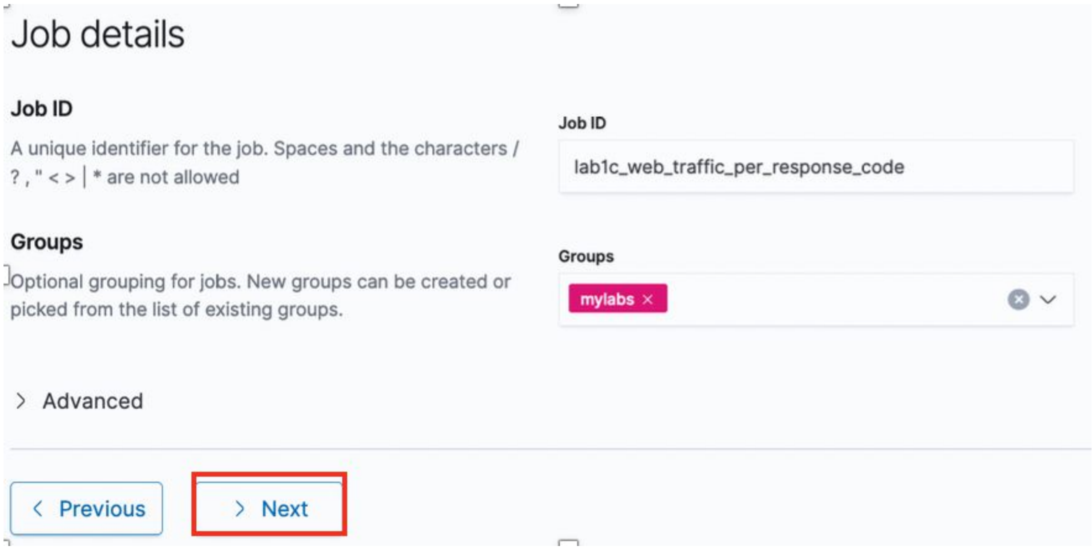
Bucket span: 1h

Sparse data

Select if you wish to ignore empty buckets from being considered anomalous. Available for count and sum analysis.

Sparse data: ✕

10. Name your job as "lab1c_web_traffic_per_response_code" and place it under “mylabs” group. Then click “Next”



Job details

Job ID

A unique identifier for the job. Spaces and the characters / ? , " < > | * are not allowed

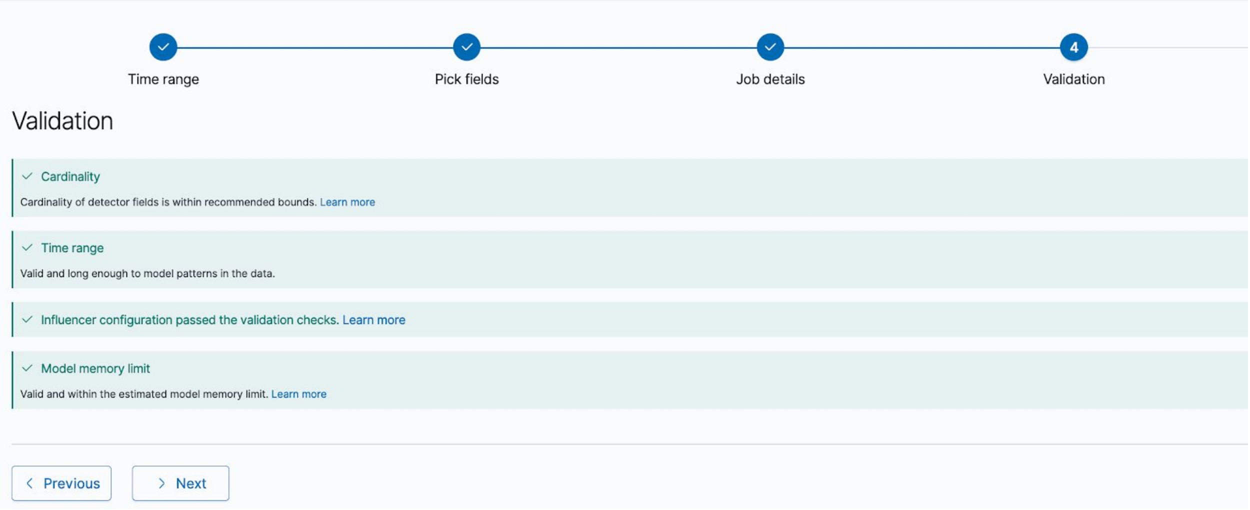
Job ID: lab1c_web_traffic_per_response_code

Groups

Optional grouping for jobs. New groups can be created or picked from the list of existing groups.

Groups: mylabs ×

11. The job should pass through the validation. Click Next.



Validation

✓ Cardinality

Cardinality of detector fields is within recommended bounds. Learn more

✓ Time range

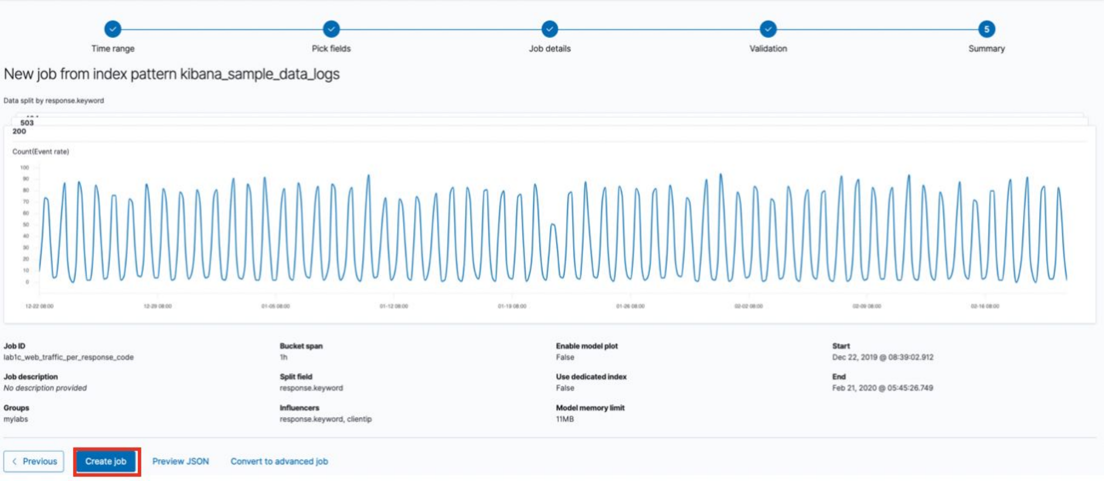
Valid and long enough to model patterns in the data.

✓ Influencer configuration passed the validation checks. Learn more

✓ Model memory limit

Valid and within the estimated model memory limit. Learn more

12. Review the job settings and click on “Create Job” to start the ML job



New job from index pattern kibana_sample_data_logs

Data split by response.keyword

Job ID
lab1c_web_traffic_per_response_code

Job description
No description provided

Groups
mylabs

Bucket span
1h

Split field
response.keyword

Influencers
response.keyword, clientip

Enable model plot
False

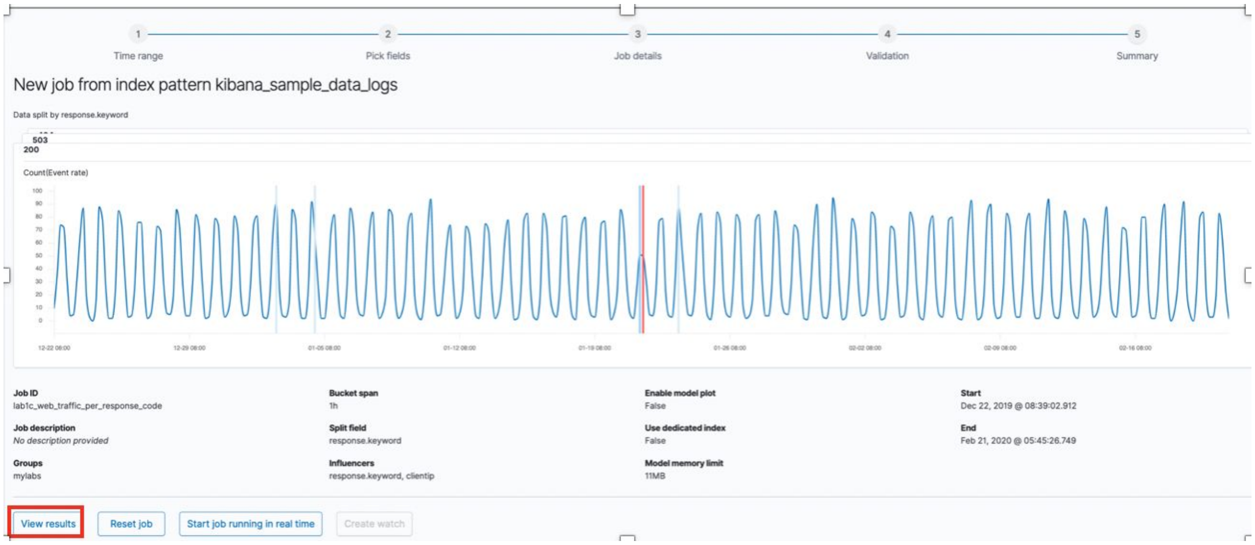
Use dedicated index
False

Model memory limit
11MB

Start
Dec 22, 2019 @ 08:39:02.912

End
Feb 21, 2020 @ 05:45:26.749

13. After the job has completed, click on the “View Results” link to drill down to the results



New job from index pattern kibana_sample_data_logs

Data split by response.keyword

Job ID
lab1c_web_traffic_per_response_code

Job description
No description provided

Groups
mylabs

Bucket span
1h

Split field
response.keyword

Influencers
response.keyword, clientip

Enable model plot
False

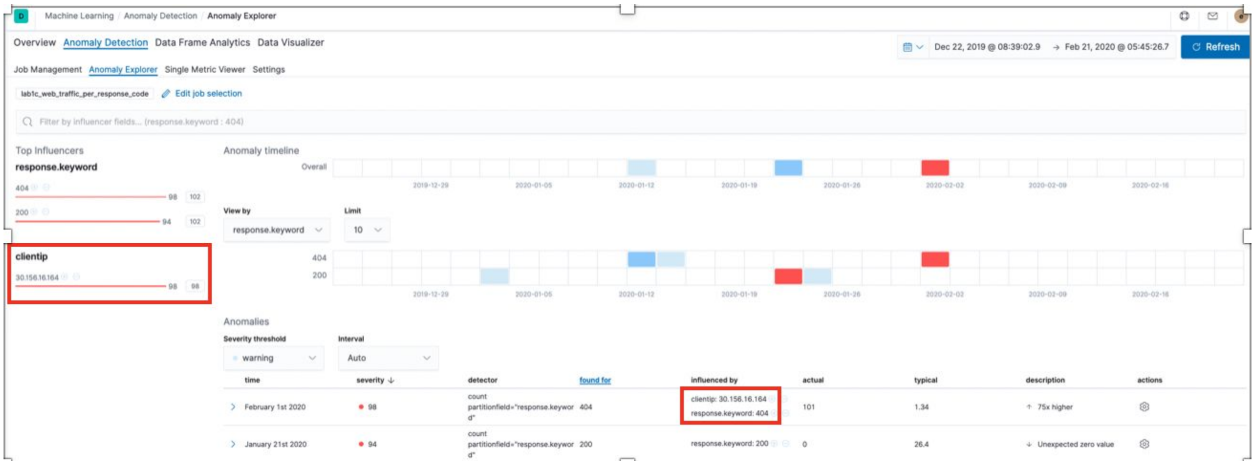
Use dedicated index
False

Model memory limit
11MB

Start
Dec 22, 2019 @ 08:39:02.912

End
Feb 21, 2020 @ 05:45:26.749

14. At a glance, we can tell that there were anomalies associated with response codes **404** and **200** over the period, and that the anomaly related to 404 seems to be caused by (influenced by) ip address: **30.156.16.164**

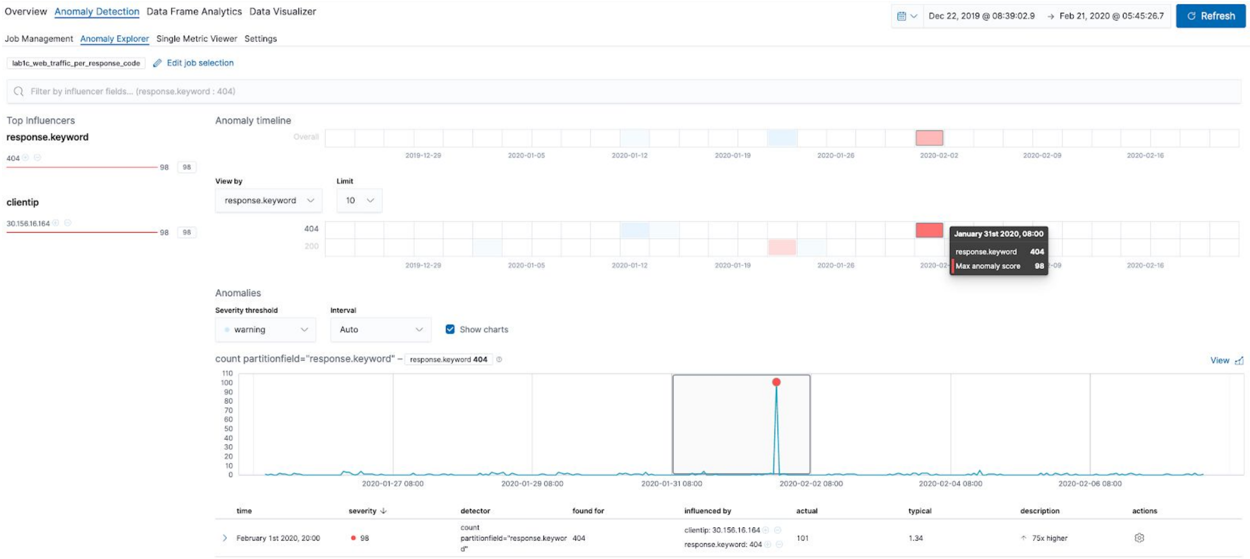


View by: response.keyword
Limit: 10

Severity threshold: warning
Interval: Auto

	time	severity ↓	detector	found for	influenced by	actual	typical	description	actions
>	February 1st 2020	● 98	count partitionfield="response.keyword"	404	clientip: 30.156.16.164 response.keyword: 404	101	1.34	↑ 75x higher	⚙
>	January 21st 2020	● 94	count partitionfield="response.keyword"	200	response.keyword: 200	0	26.4	↓ Unexpected zero value	⚙

15. Click on the red square for response code 404 to drill down further:



View by: response.keyword
Limit: 10

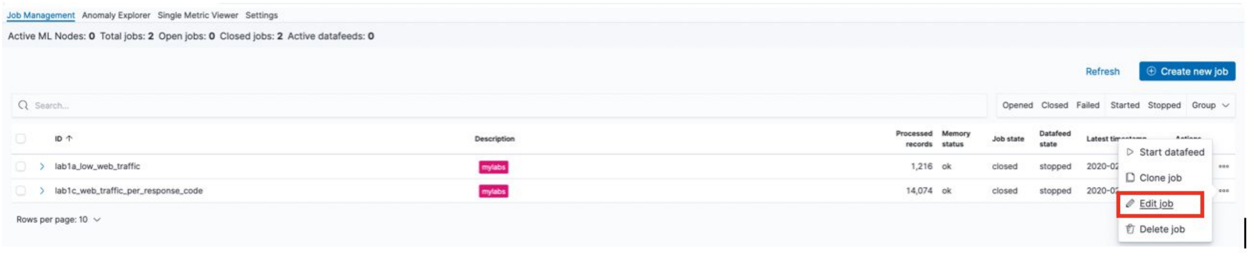
January 31st 2020, 08:00
response.keyword 404
Max anomaly score 98

Severity threshold: warning
Interval: Auto
☑ Show charts

	time	severity ↓	detector	found for	influenced by	actual	typical	description	actions
>	February 1st 2020, 20:00	● 98	count partitionfield="response.keyword"	404	clientip: 30.156.16.164 response.keyword: 404	101	1.34	↑ 75x higher	⚙

D - Add a custom URL to the Multi-Metric Job

1. Click on “**Job Management**” > “**Edit Job**” for “lab1c_web_traffic_per_response_code”



Job Management Anomaly Explorer Single Metric Viewer Settings

Active ML Nodes: **0** Total jobs: **2** Open jobs: **0** Closed jobs: **2** Active datafeeds: **0**

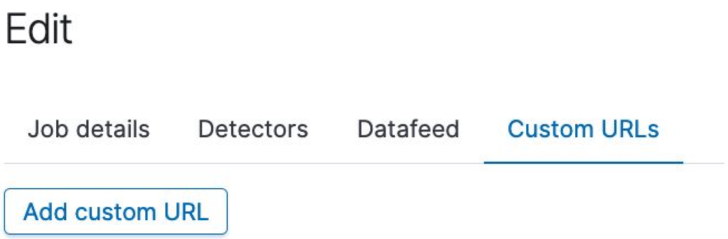
Refresh | Create new job

Search... | Opened Closed Failed Started Stopped Group

□	ID ↑	Description	Processed records	Memory status	Job state	Datafeed state	Latest tim
□	lab1a_low_web_traffic	mylabs	1,216	ok	closed	stopped	2020-02
□	lab1c_web_traffic_per_response_code	mylabs	14,074	ok	closed	stopped	2020-0

Rows per page: 10

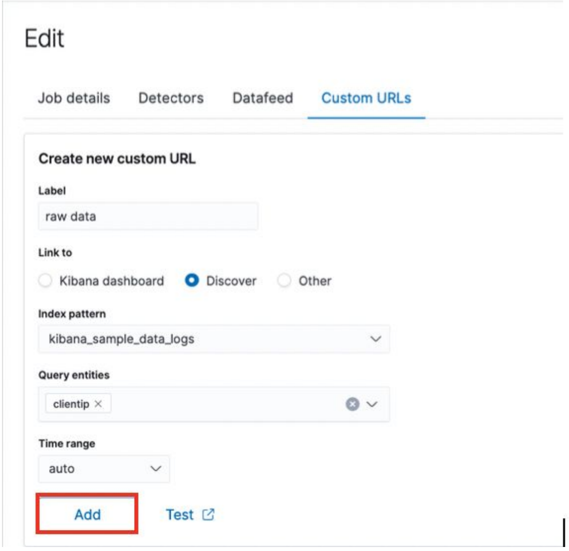
2. Click on the “Custom URL” tab and “Add custom URL”



Edit

3. Enter the following details:
    a. Label : “raw data”
    b. Link to : “Discover”
    c. Index Pattern : “kibana_sample_data_logs”
    d. Query entities : “clientip”

   Click on the “**Add**” button



Edit

Create new custom URL

Label: raw data

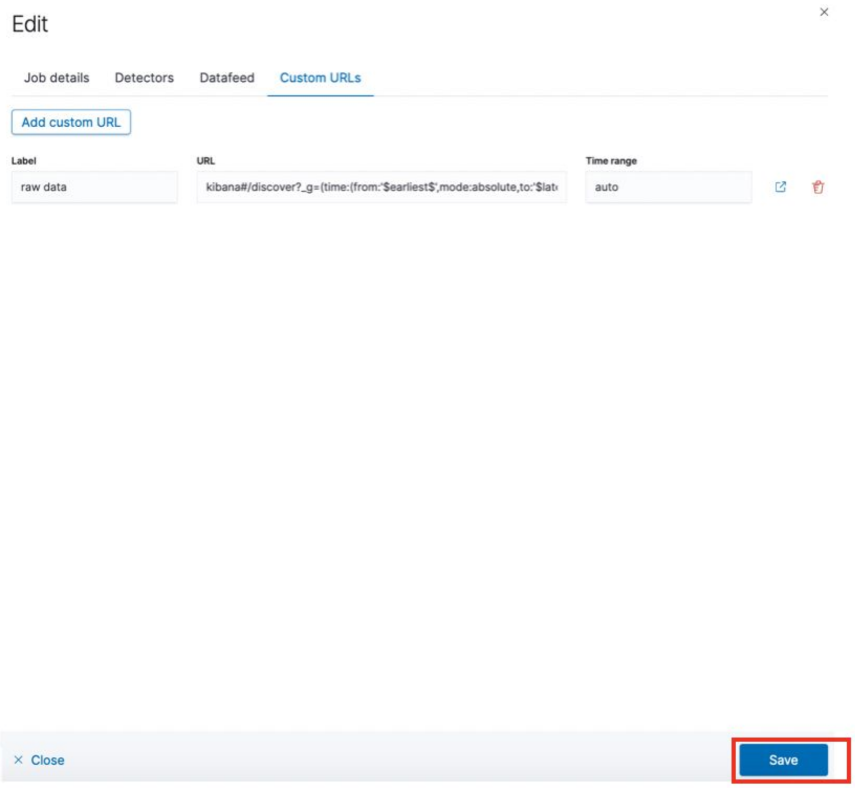
Link to: ○ Kibana dashboard ● Discover ○ Other

Index pattern: kibana_sample_data_logs

Query entities: clientip

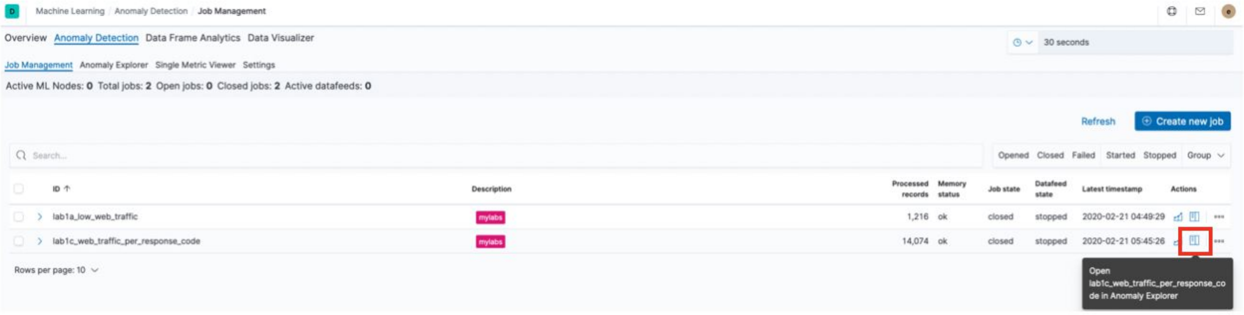
Time range: auto

4. Remember to click on the “Save” button



Label	URL	Time range
raw data	kibana#/discover?_g=(time:(from:'$earliest$',mode:absolute,to:'$lat	auto

5. Now, let’s go back to the previous results: Click on the Anomaly Explorer link for the lab1c job:

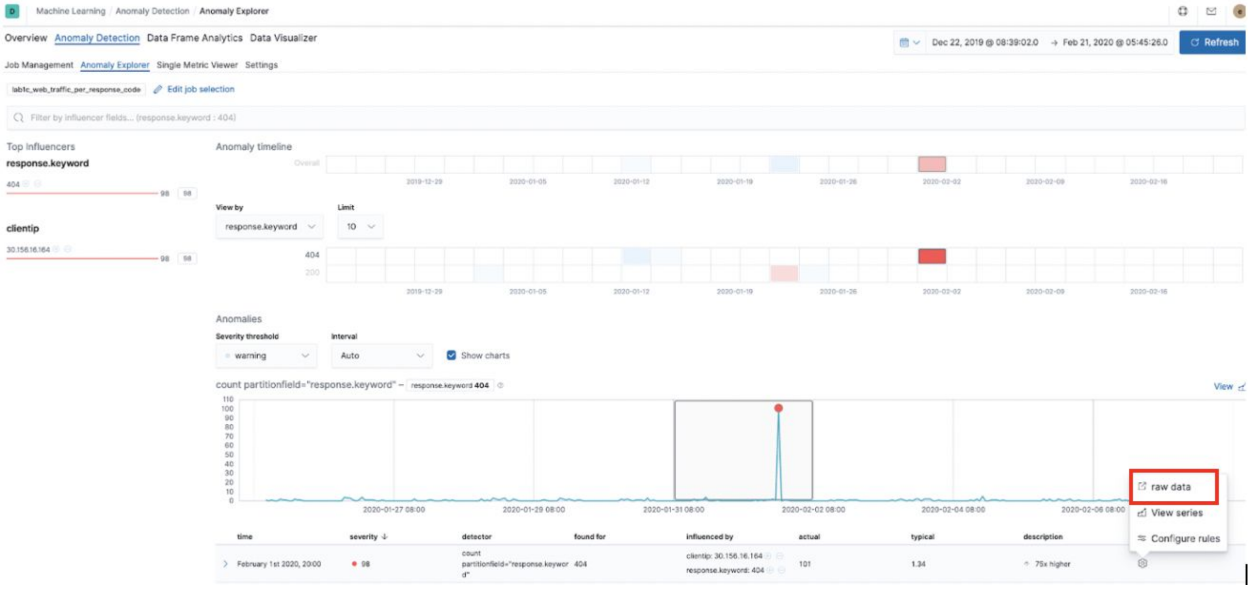


Active ML Nodes: **0** Total jobs: **2** Open jobs: **0** Closed jobs: **2** Active datafeeds: **0**

ID ↑	Description	Processed records	Memory status	Job state	Datafeed state	Latest timestamp	Actions
lab1a_low_web_traffic	mylabs	1,216	ok	closed	stopped	2020-02-21 04:49:29	
lab1c_web_traffic_per_response_code	mylabs	14,074	ok	closed	stopped	2020-02-21 05:45:26	

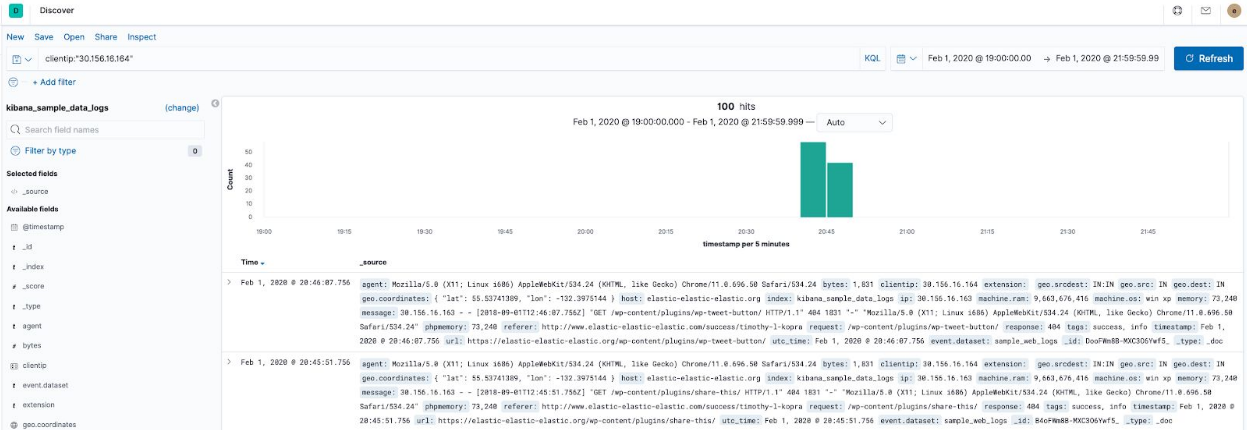
Open lab1c_web_traffic_per_response_code in Anomaly Explorer

6. Click on the red box again for 404. Click on the “Actions” icon at the top right-hand corner and click on the “raw data” link:



time	severity	detector	found for	influenced by	actual	typical	description
February 1st 2020, 20:00	● 98	count partitionfield="response.keyword"	404	clientip: 30.156.16.164 response.keyword: 404	101	1.34	75x higher

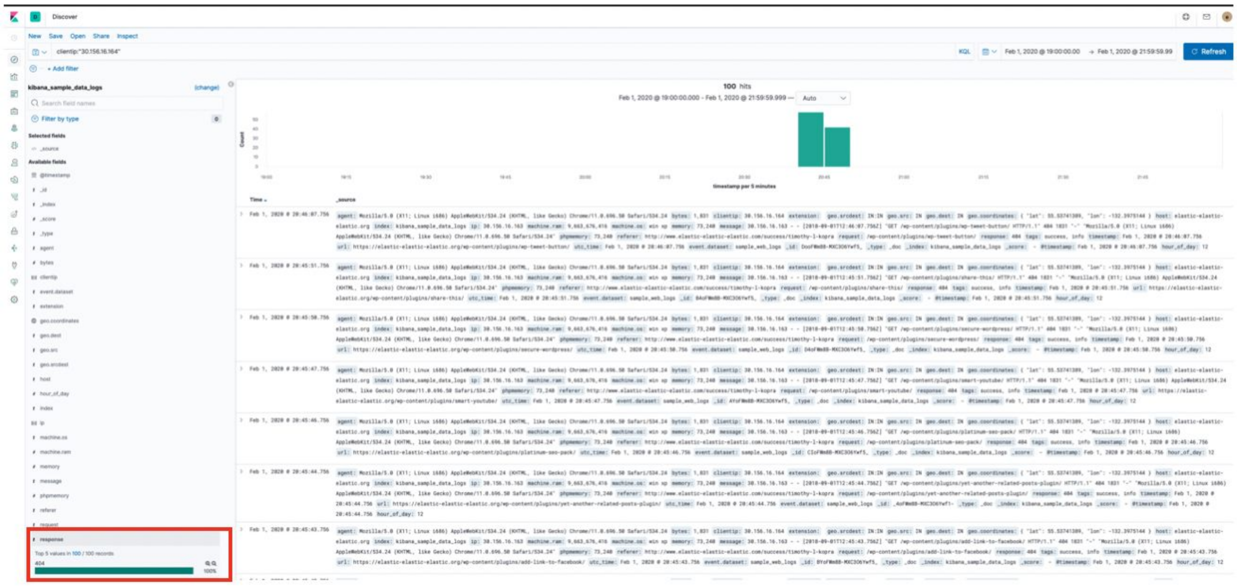
7. This brings us to the Discover page showing us all the relevant documents filtered by “clientip:"30.156.16.164"”



clientip:"30.156.16.164"

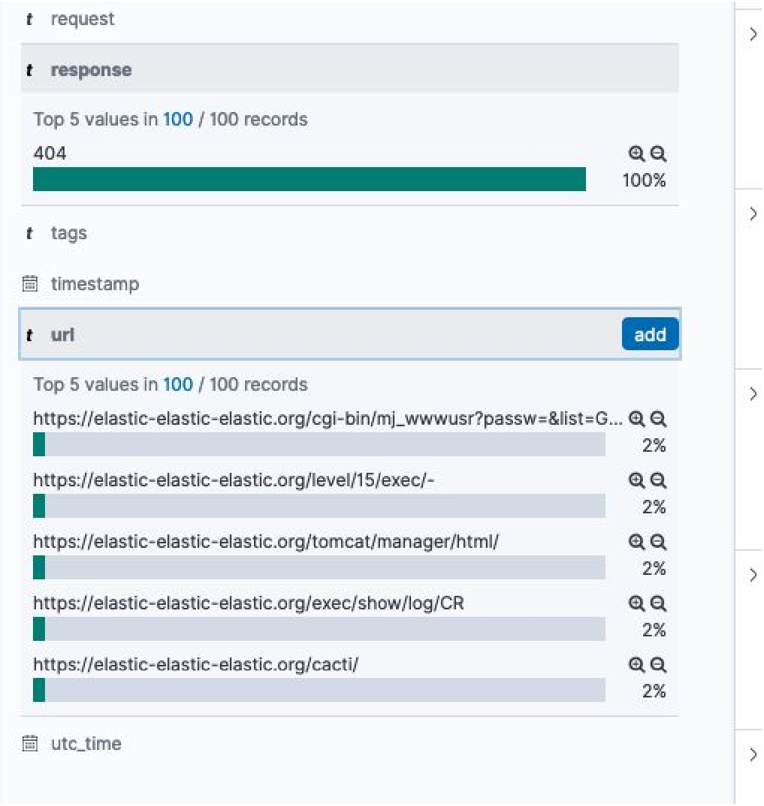
```
Time        _source
Feb 1, 2020 @ 20:46:07.756  agent: Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes: 1,831 clientip: 30.156.16.164 extension:  geo.srcdest: IN:IN geo.src: IN geo.dest: IN geo.coordinates: { "lat": 55.53741389, "lon": -132.3975144 } host: elastic-elastic-elastic.org index: kibana_sample_data_logs ip: 30.156.16.163 machine.ram: 9,663,676,416 machine.os: win xp memory: 73,240 message: 30.156.16.163 - - [2018-09-01T12:46:07.756Z] "GET /wp-content/plugins/wp-tweet-button/ HTTP/1.1" 404 1831 "-" "Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24" phpmemory: 73,240 referer: http://www.elastic-elastic-elastic.com/success/timothy-l-kopra request: /wp-content/plugins/wp-tweet-button/ response: 404 tags: success, info timestamp: Feb 1, 2020 @ 20:46:07.756 url: https://elastic-elastic-elastic.org/wp-content/plugins/wp-tweet-button/ utc_time: Feb 1, 2020 @ 20:46:07.756 event.dataset: sample_web_logs _id: DooFWm8B-MXC3O6Ywf5_ _type: _doc
Feb 1, 2020 @ 20:45:51.756  agent: Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes: 1,831 clientip: 30.156.16.164 extension:  geo.srcdest: IN:IN geo.src: IN geo.dest: IN geo.coordinates: { "lat": 55.53741389, "lon": -132.3975144 } host: elastic-elastic-elastic.org index: kibana_sample_data_logs ip: 30.156.16.163 machine.ram: 9,663,676,416 machine.os: win xp memory: 73,240 message: 30.156.16.163 - - [2018-09-01T12:45:51.756Z] "GET /wp-content/plugins/share-this/ HTTP/1.1" 404 1831 "-" "Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24" phpmemory: 73,240 referer: http://www.elastic-elastic-elastic.com/success/timothy-l-kopra request: /wp-content/plugins/share-this/ response: 404 tags: success, info timestamp: Feb 1, 2020 @ 20:45:51.756 url: https://elastic-elastic-elastic.org/wp-content/plugins/share-this/ utc_time: Feb 1, 2020 @ 20:45:51.756 event.dataset: sample_web_logs _id: B4oFWm8B-MXC3O6Ywf5_ _type: _doc
```

8. A quick click on the “response” field on the left would show us that all the 100 requests sent by this clientip encountered the 404 response code:



```
> Feb 1, 2020 @ 20:46:07.756  agent: Mozilla/5.0 (X11; Linux i686) AppleWebKit/534.24 (KHTML, like Gecko) Chrome/11.0.696.50 Safari/534.24 bytes: 1,831 clientip: 30.156.16.164 ... response: 404 tags: success, info ... url: https://elastic-elastic-elastic.org/wp-content/plugins/wp-tweet-button/ ...
> Feb 1, 2020 @ 20:45:51.756  ... request: /wp-content/plugins/share-this/ response: 404 ...
> Feb 1, 2020 @ 20:45:50.756  ... request: /wp-content/plugins/secure-wordpress/ response: 404 ...
> Feb 1, 2020 @ 20:45:47.756  ... request: /wp-content/plugins/smart-youtube/ response: 404 ...
> Feb 1, 2020 @ 20:45:46.756  ... request: /wp-content/plugins/platinum-seo-pack/ response: 404 ...
> Feb 1, 2020 @ 20:45:44.756  ... request: /wp-content/plugins/yet-another-related-posts-plugin/ response: 404 ...
> Feb 1, 2020 @ 20:45:43.756  ... request: /wp-content/plugins/add-link-to-facebook/ response: 404 ...
```

9. You can also click on the “URL” field to take a quick look at the URLs in question:



t request

t **response**

Top 5 values in 100 / 100 records

Value	Share
404	100%

t tags

timestamp

t **url** add

Top 5 values in 100 / 100 records

Value	Share
https://elastic-elastic-elastic.org/cgi-bin/mj_wwwusr?passw=&list=G...	2%
https://elastic-elastic-elastic.org/level/15/exec/-	2%
https://elastic-elastic-elastic.org/tomcat/manager/html/	2%
https://elastic-elastic-elastic.org/exec/show/log/CR	2%
https://elastic-elastic-elastic.org/cacti/	2%

utc_time