Getting started on Elastic Cloud with a Sample Dataset “”

Objective:

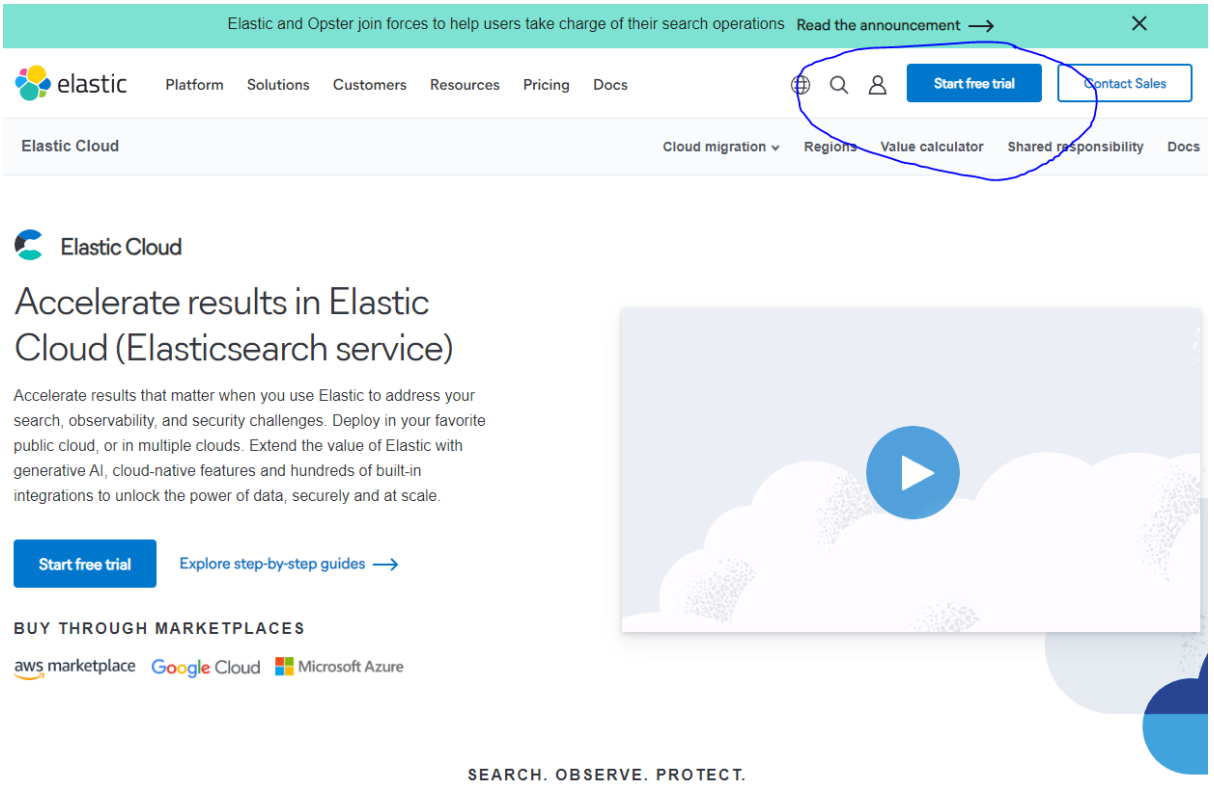
- Visit and create an account for Elastic Cloud
- Implement data sets
- Create Simple Graphs Visualizations

Platform Specifications:

- ElasticSearch & Kibana
- Set size to 1 GB RAM
- Set nodes per zone to 1

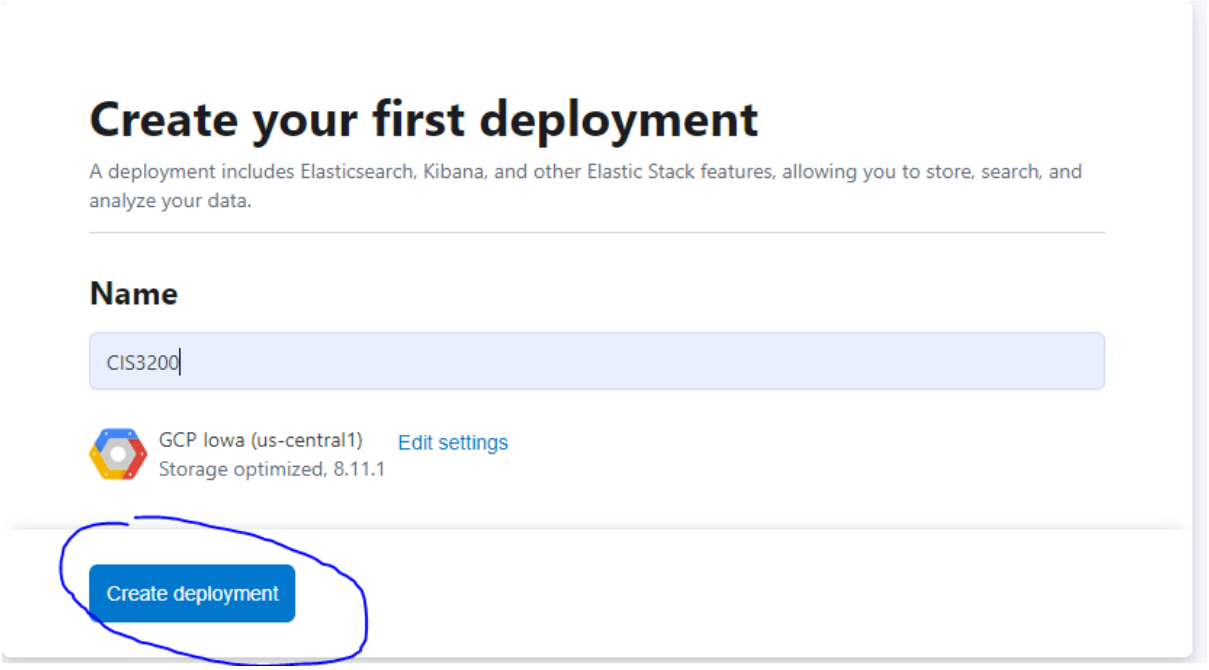
Step 1 Create an account for Elastic Cloud:

1. Go to https://www.elastic.co/cloud/elasticsearch-service/signup
2. Click the option said **Start free trial**

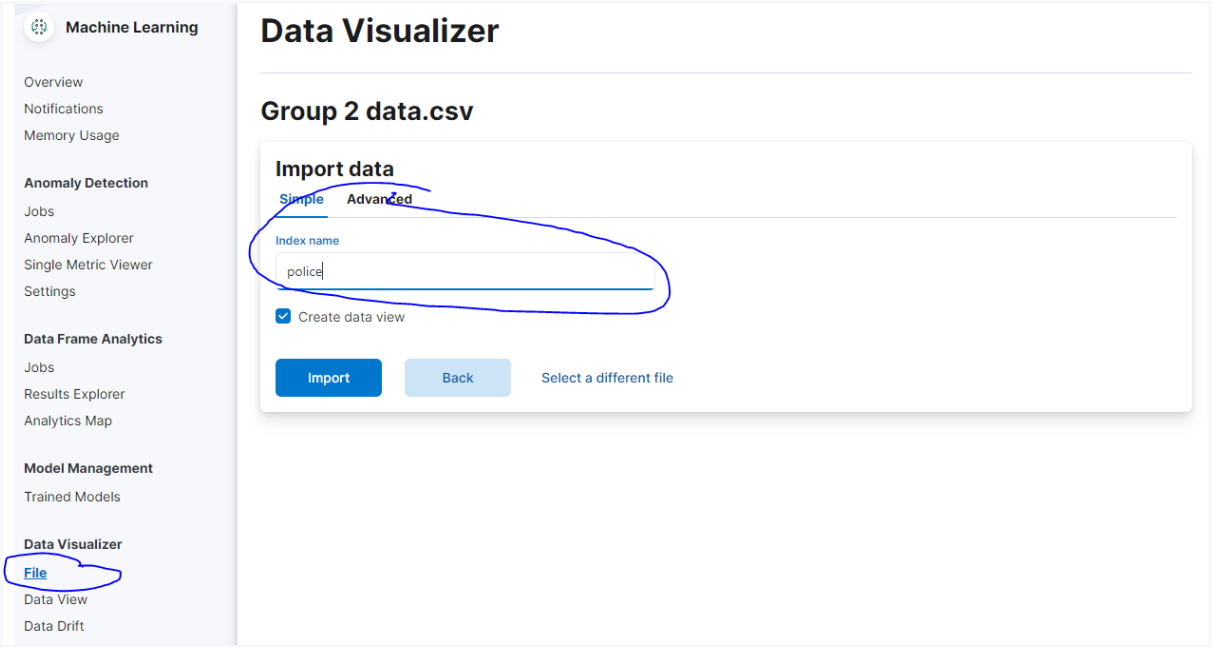


3. Next, you can register and Log into your Elastic Cloud account by using your school email.

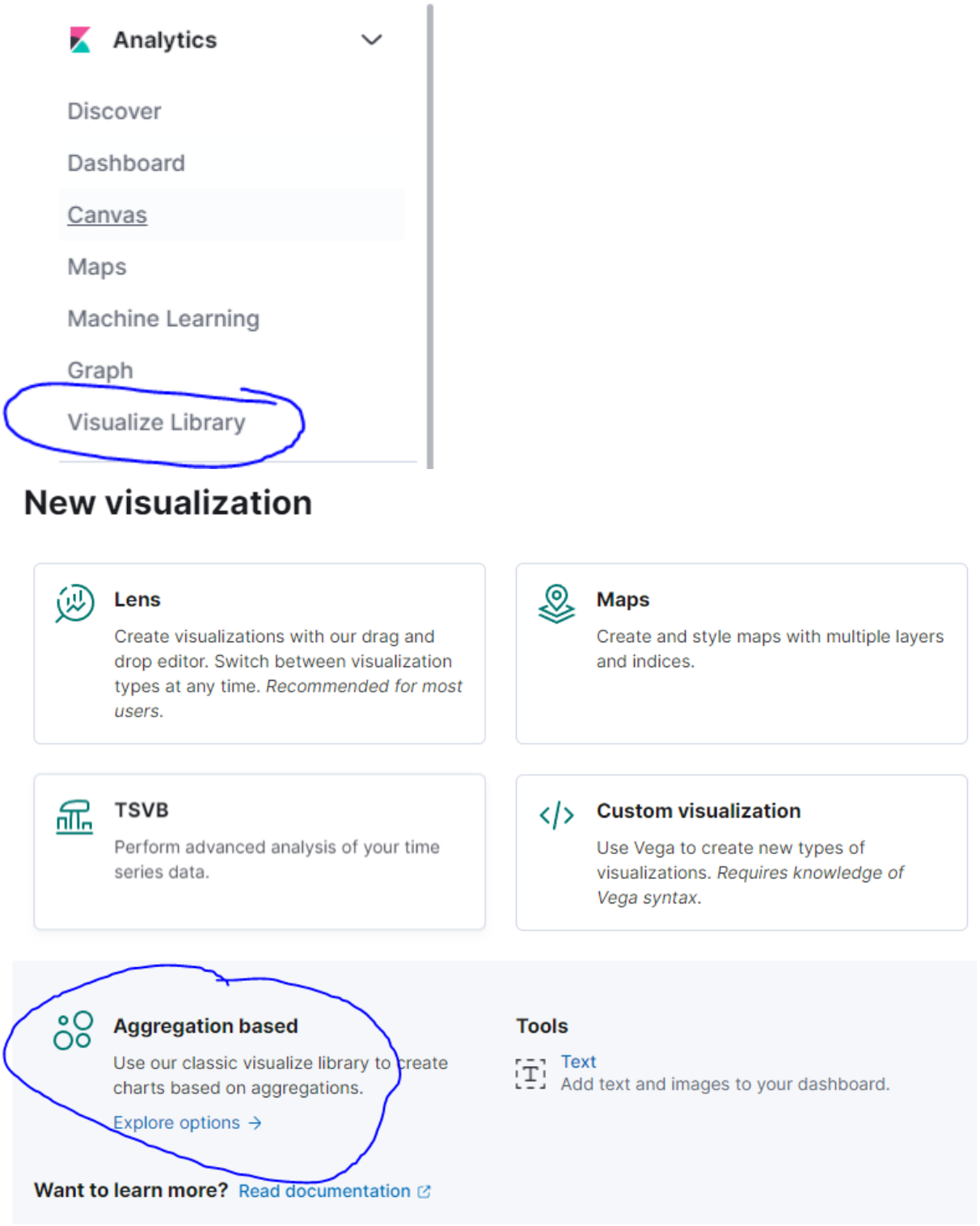
4. Once you have signed in you will see an page asking your name and company. Name your deployment and click on the "Create Deployment" button.



5. Next, you will need to download the data sets from https://github.com/ZehaoChen3200/CIS-3200-group-2/blob/main/Q2_2019_SFPD_Data.csv and https://github.com/ZehaoChen3200/CIS-3200-group-2/blob/main/SFpolice-2020Q2.csv
6. Go back to the home page of Elistic Coud and click the **Machine Learning and go to the Data visualizer->Files.** Set the Index name to police.

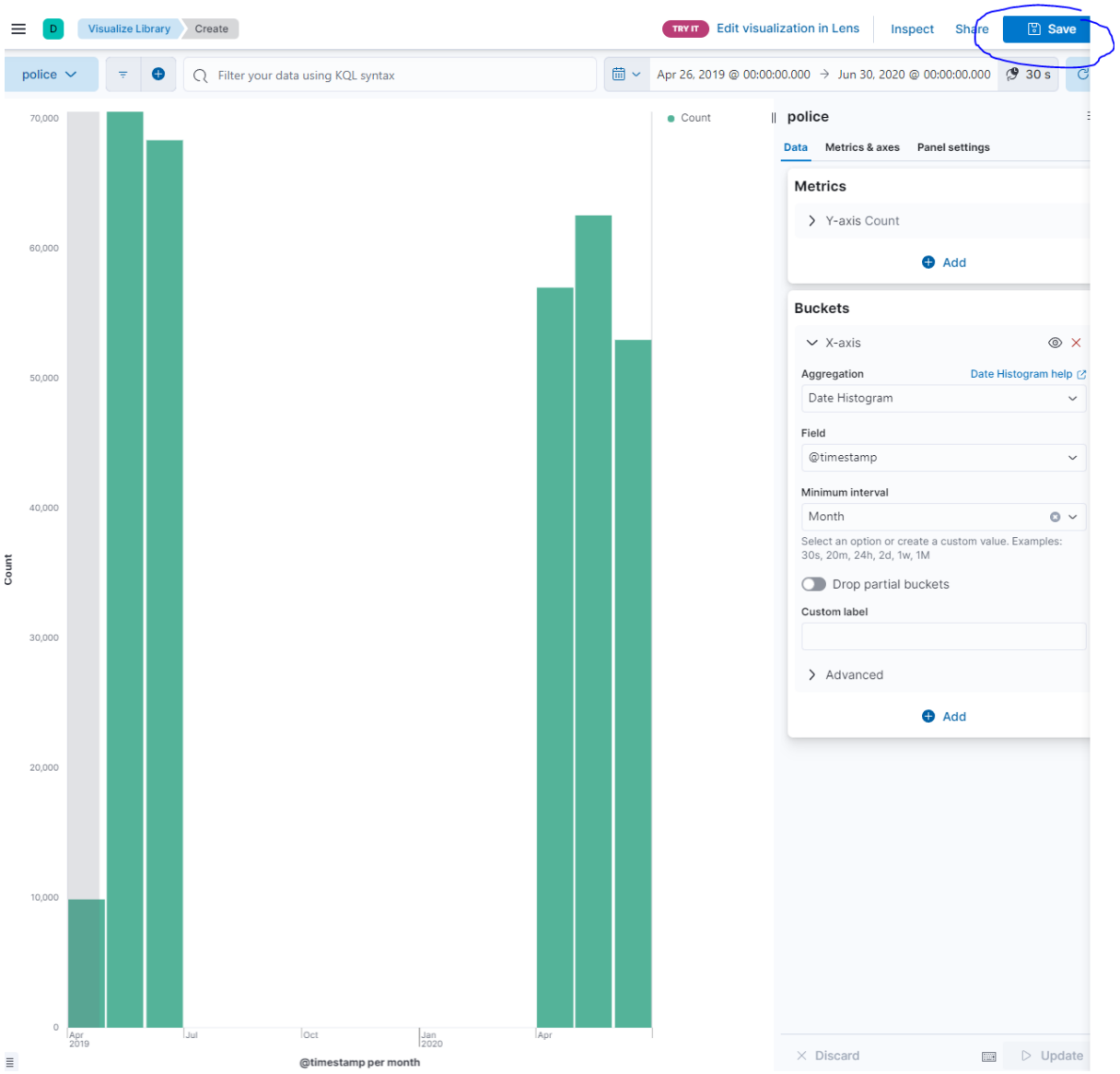


7. Click on “Visualize” tab to start creating the Visualization for the data,then click the Create visualization and choose Vertical Bar in the Aggregation based.



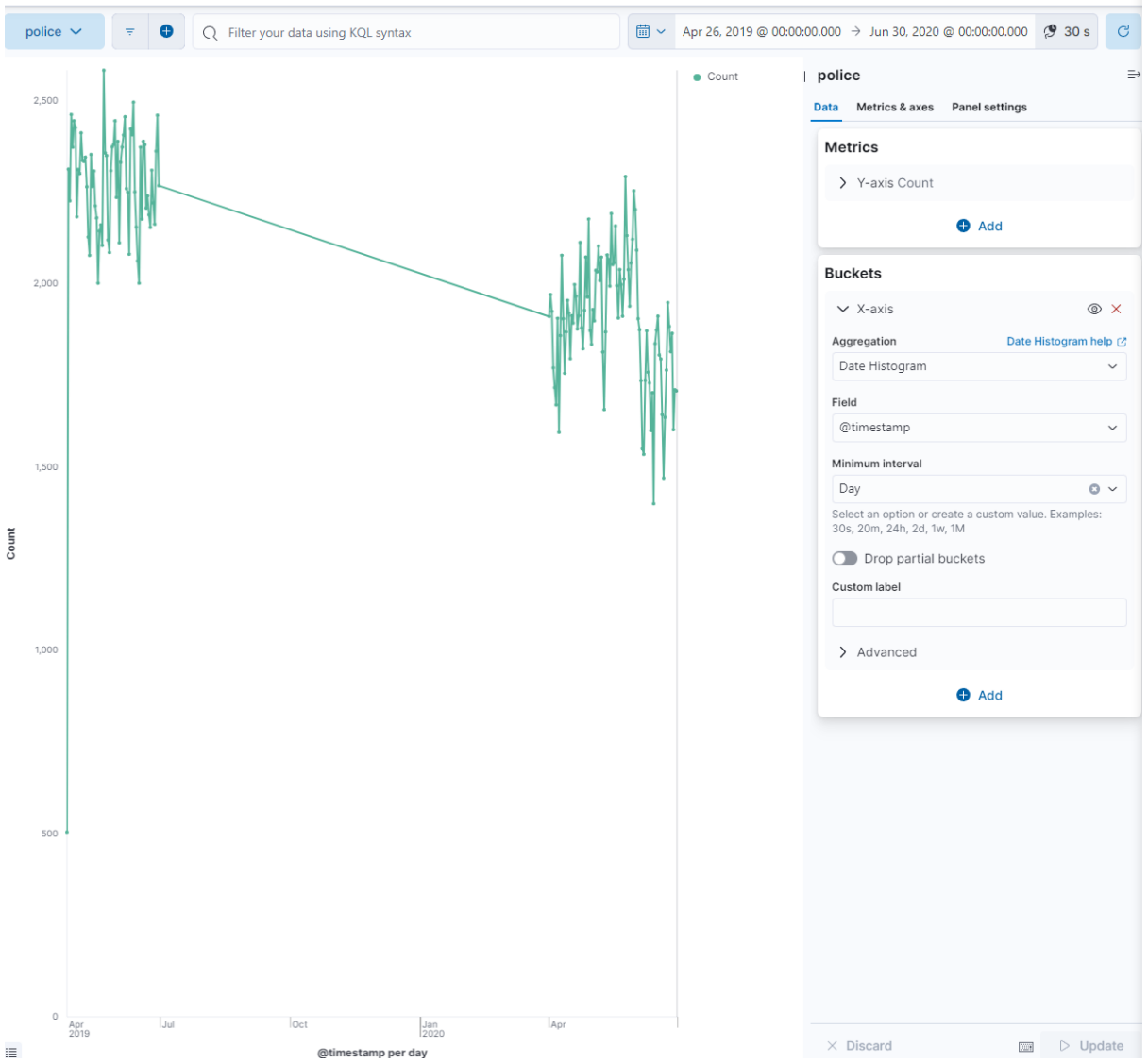
8. At the New Vertical Bar, click on the index pattern name “police”.
9. Next, set the time range from Apr.24 2019, 00:00:00, to Jun 30 2020, 00:00:00. and then add a X-axis into bucket and set Aggregation->Date Histogram, Field->timestamp, ->minimum

interval ->Month. You can see the difference between Q2 2019 and Q2 2020. Save it.



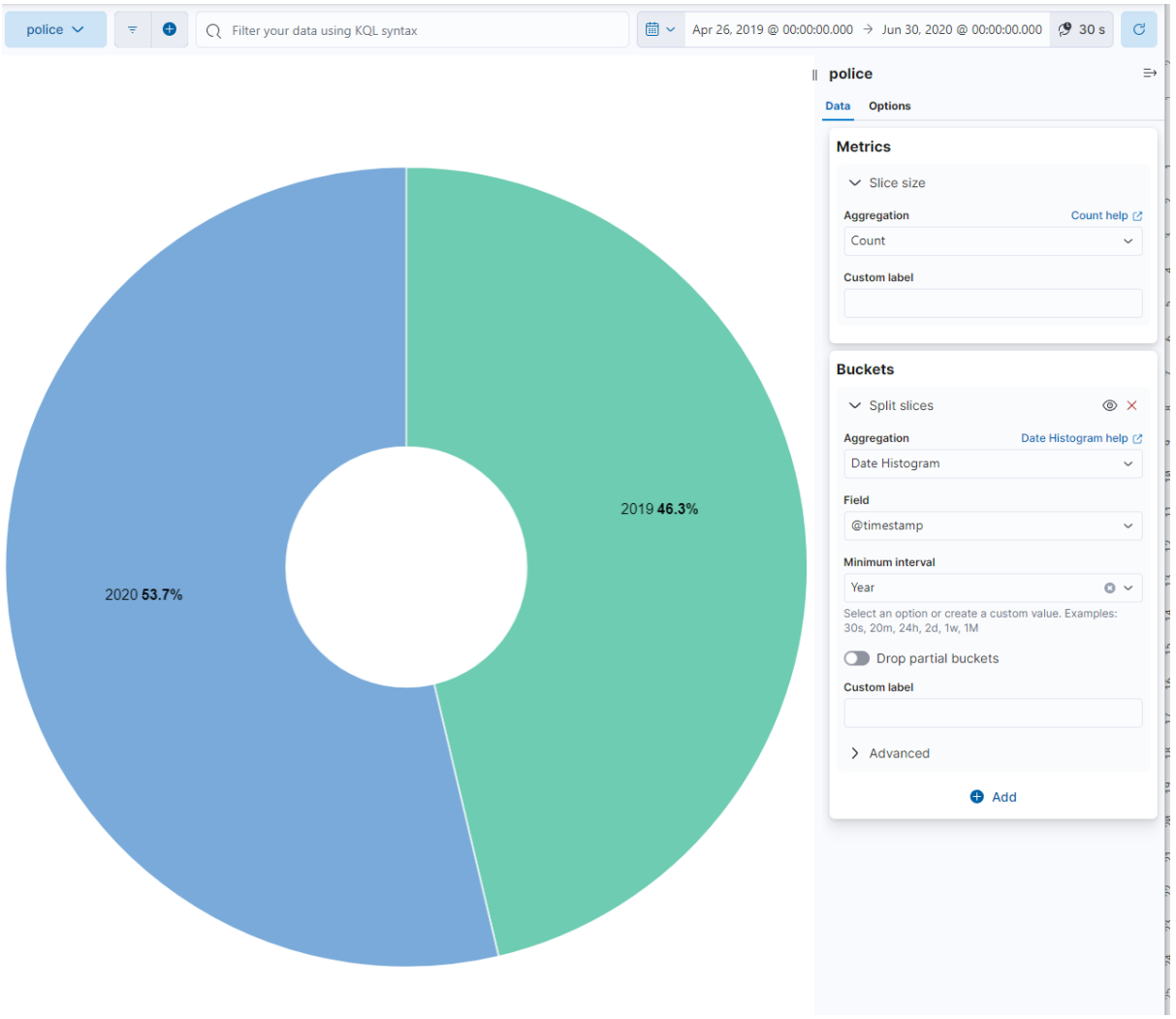
10. Next, go back to he visualize library and click the Create visualization again. Choose Line in the Aggregation based. Then click on the index pattern name “police”.
11. Set the time range from Apr.24 2019, 00:00:00, to Jun 30 2020, 00:00:00. and then add a X-axis into bucket and set Aggregation->Date Histogram, Field->timestamp, ->minimum

interval ->Day. Save it.

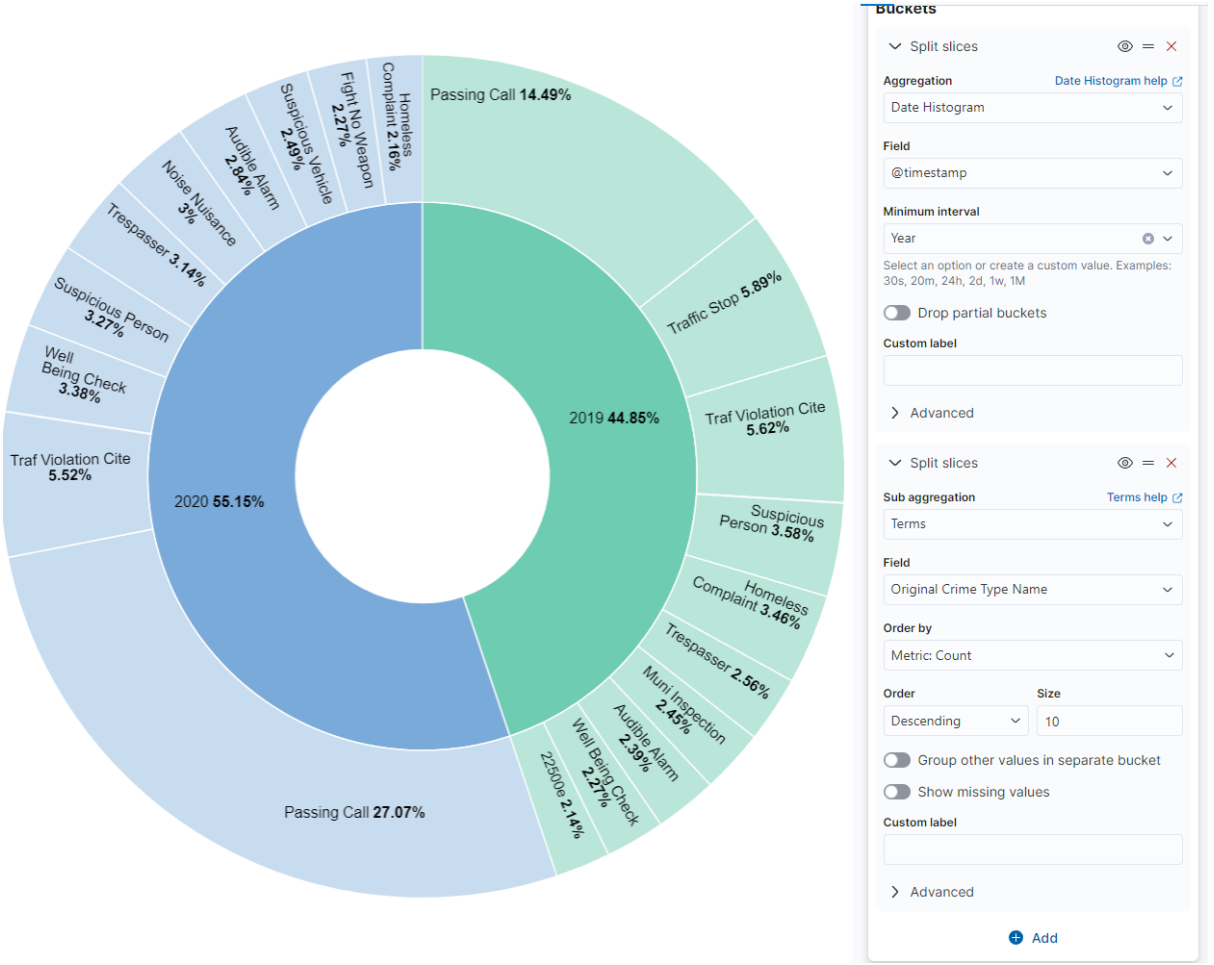


12. Next, go back to he visualize library and click the Create visualization again. Choose Pie in the Aggregation based. Then click on the index pattern name “police”.
13. Set the time range from Apr.24 2019, 00:00:00, to Jun 30 2020, 00:00:00. and then add a Bucket for Split slices. Set Aggregation->Date Histogram, Field->timestamp, ->minimum

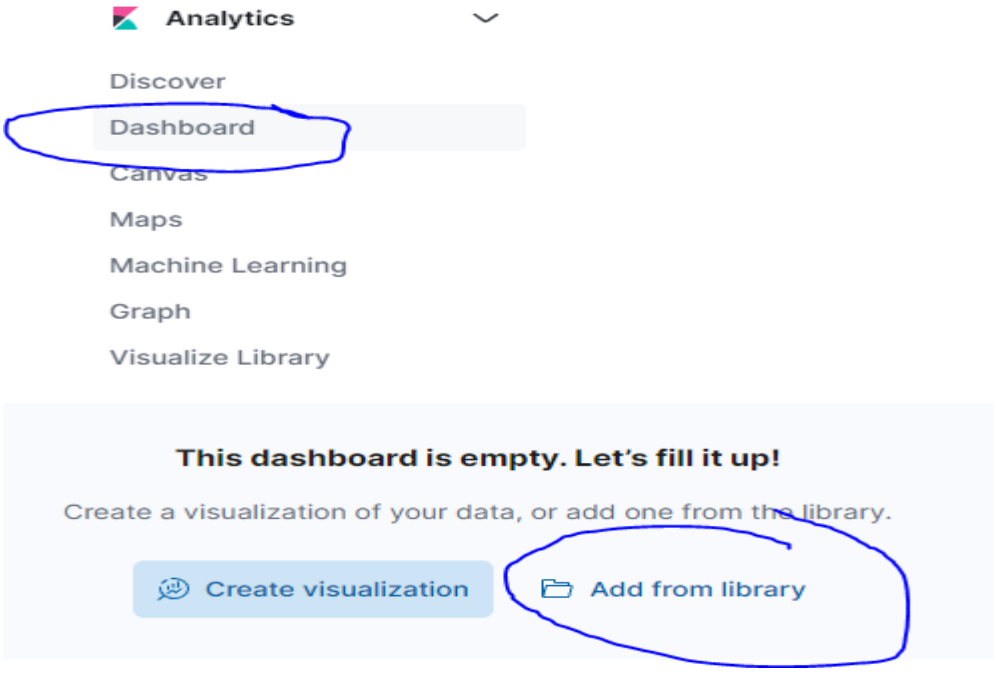
interval ->year.



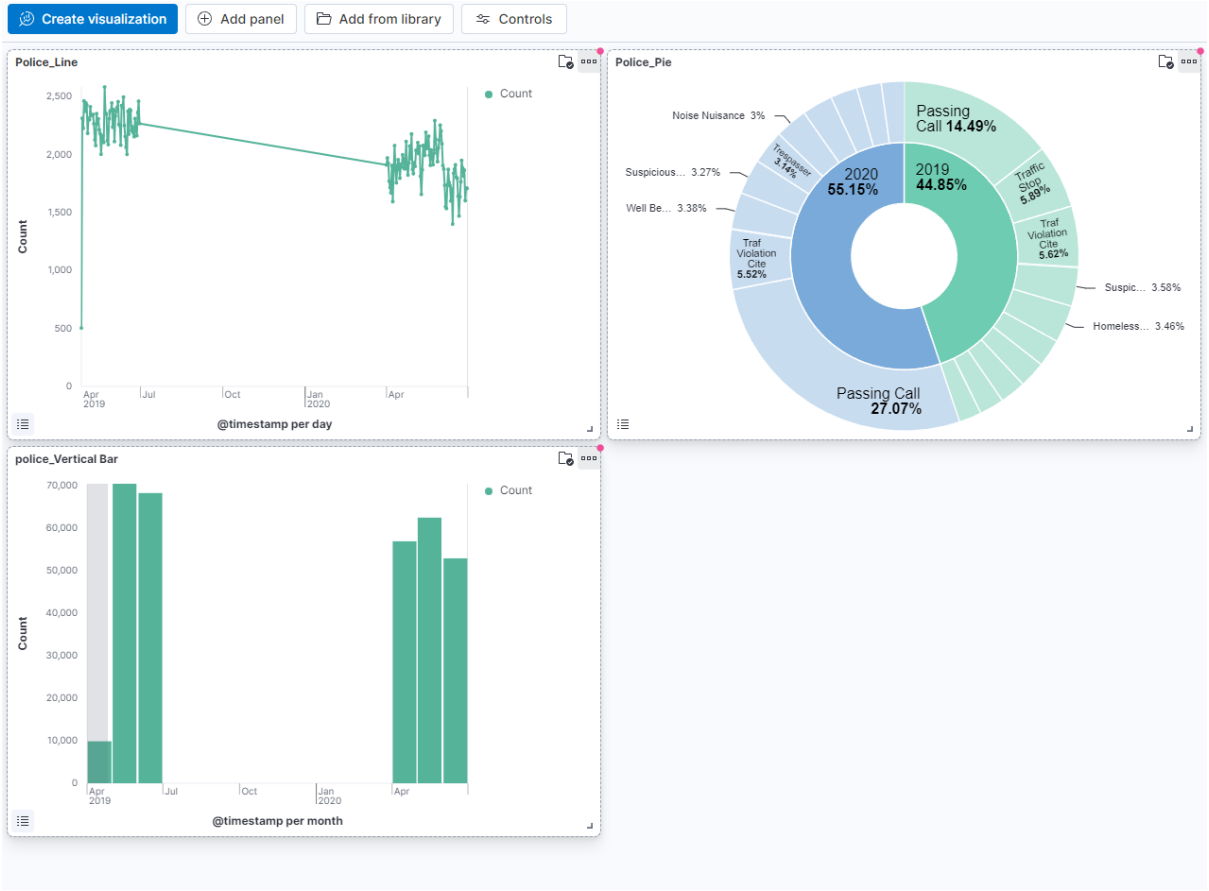
14. Add another split slices. Set sub aggregation->Terms, Field->Original Crime Type Name, Order by-> Metric:Count, Order->Descending, Size->10.Save it.



15. Click the dashboard and click the Create dashboard. Choose add from library and select all the chart you made in previous steps.

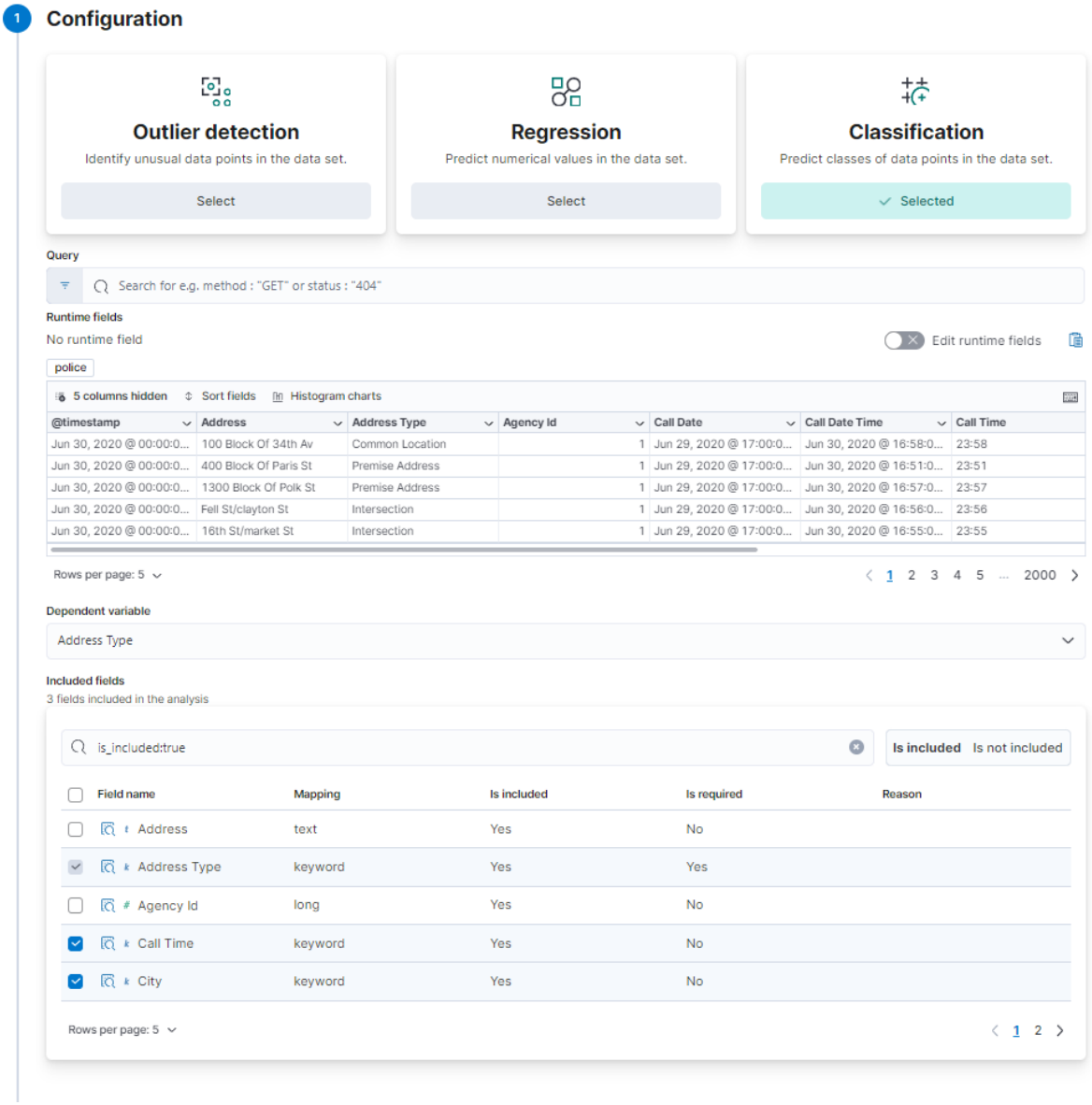


16. You will see the page similar to the following: Click Save

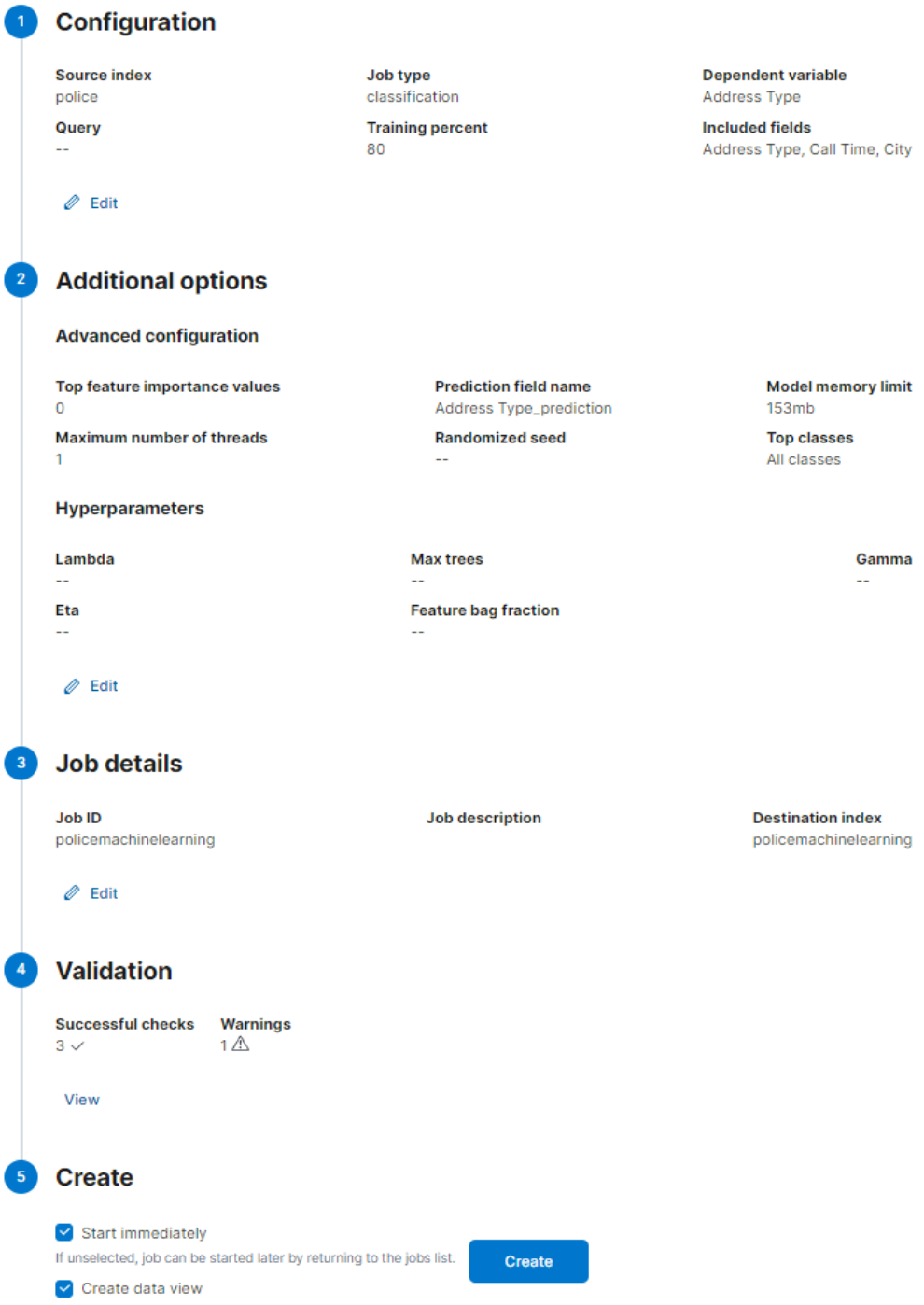


17. Go back to Machine Learning. Data Frame Analytics->job, click Create job, choose “police” which is the data we input in previous step.

18. Click Classification and set the Dependent variable to “Address type”. Included fields are address type, call time and city. Click continue



19. Set the model memory limit to 153mb. Set jobID to policemachinelearning and the index name should be same as your jobID. Click create.



20. You will got the result like the following: