CIS3200 Term Project Tutorial CIS3200 Te Project Tutorial

Authors: Arne Muller, Erik Pacheco, Michael Do, Shadman Sayef, & Steven Vi

Instructor: Jongwook Woo

Date: 12/16/2023

Lab Tutorial

Storm Events Data Analysis using Elastic Cloud

Objectives

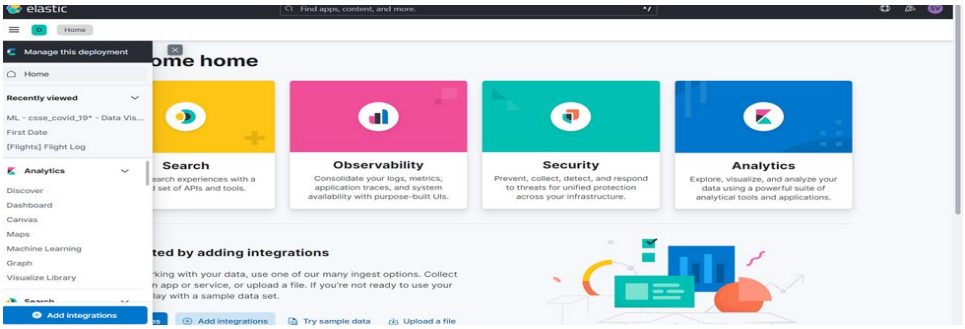
List what your objectives are. In this hands-on lab, you will learn how to:

- Get data manually
- Create prediction analysis, geo-map, & visual charts
- Using Elasticsearch cloud system
- Using Elasticsearch cloud to perform the analysis.
- Visualization

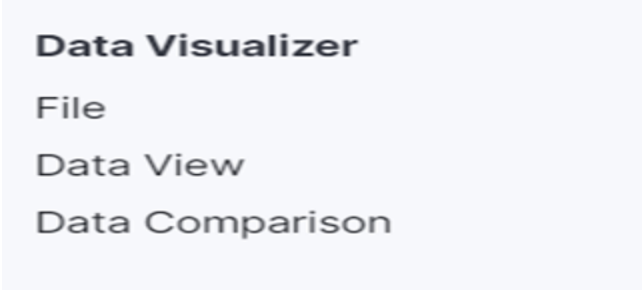
Platform Spec

- Windows 11
- CPU Speed: 1.35 GHz
- Number of CPU cores: 16 cores
- Number of nodes: 4
- Total Memory Size: 31 GB

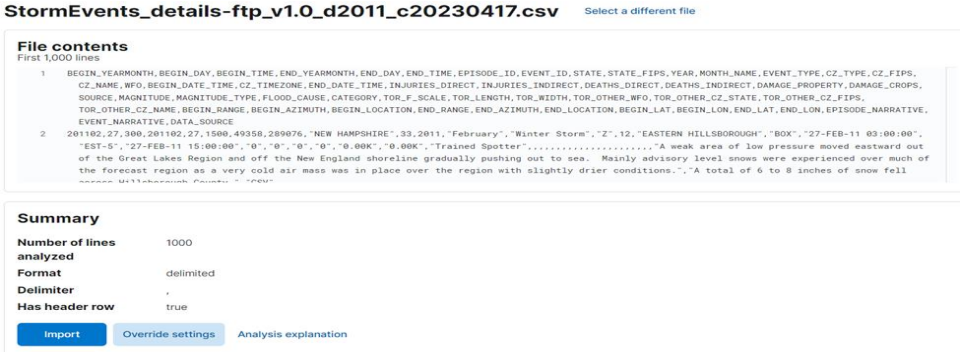
Machine Learning



1. After logging into the cloud server, go to Machine Learning.



2. Go to Data visualizer and go to file.
3. From there, upload the CSV files. In this case, the storm detail files from either 2010 and 2011.



4. Import the files as is.
5. Make the index name: *storm_events_2011*
6. Import the file.

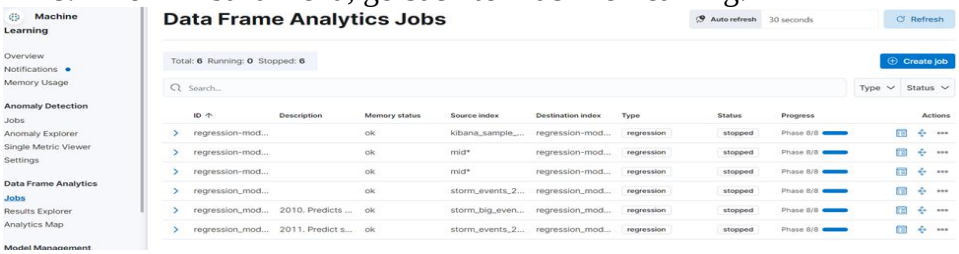
7. From Kibana menu, go to Dev Tools. Type in

```
PUT storm_events_2011/_mapping
{
  "properties": {
    "CZ_NAME": {
      "type": "text",
      "fielddata": true
```

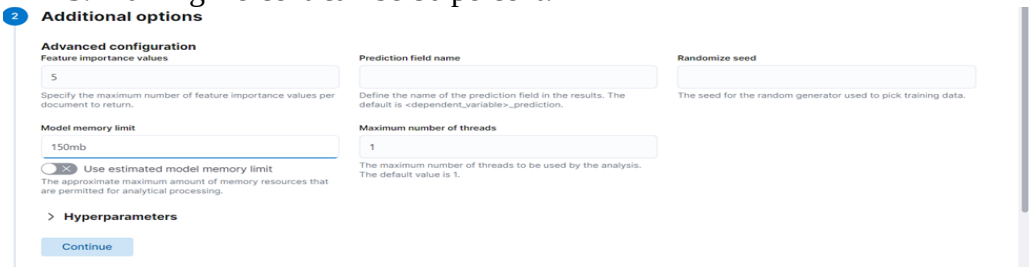
```
    }
  }
}
```

- **This is necessary for both storm_events_2010 and storm_events_2011.**

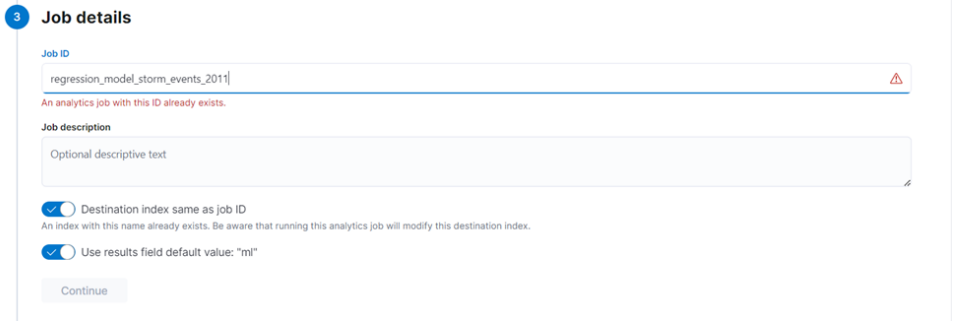
8. From Kibana menu, go back to Machine Learning.



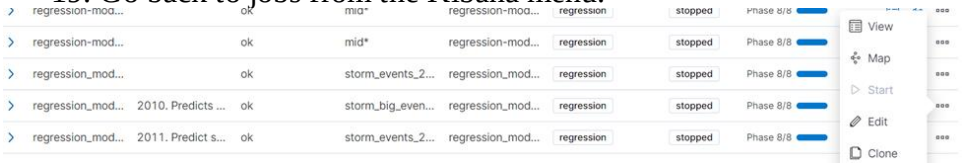
9. Go back to jobs and create job.
10. From there, pick the dataset storm_events_2010 or storm_events_2011.
11. Select Regression.
12. From dependent variable, select MAGNITUDE.
13. Training Percent can be 80 percent.



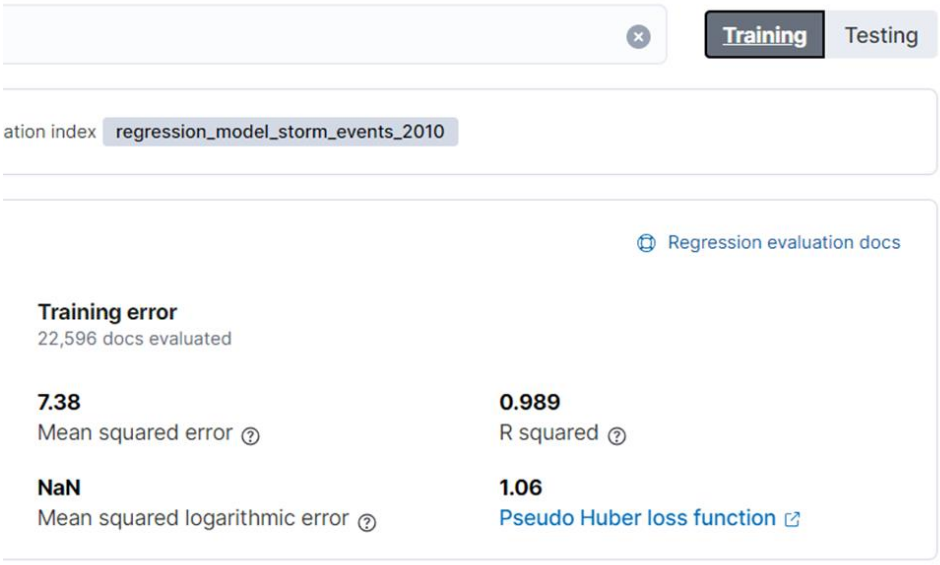
14. Put 5 in feature importance values and 150mb in model memory limit.



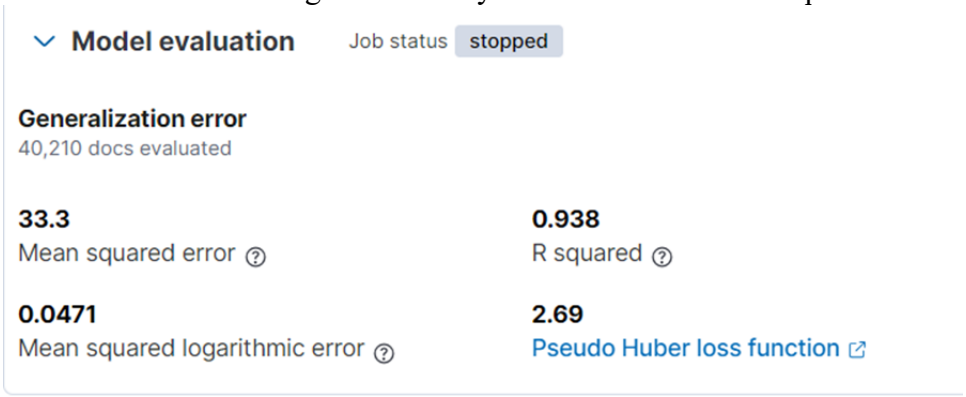
15. For job id, type in regression_model_storm_events_2010 or 2011.
16. After validation, press continue.
17. Finally, just wait for the data to be uploaded into the cloud.
18. You must do these same steps for both storm events 2010 and storm events 2011. From steps 2-17 to have them successfully uploaded to the cloud server.
19. Go back to jobs from the Kibana menu.



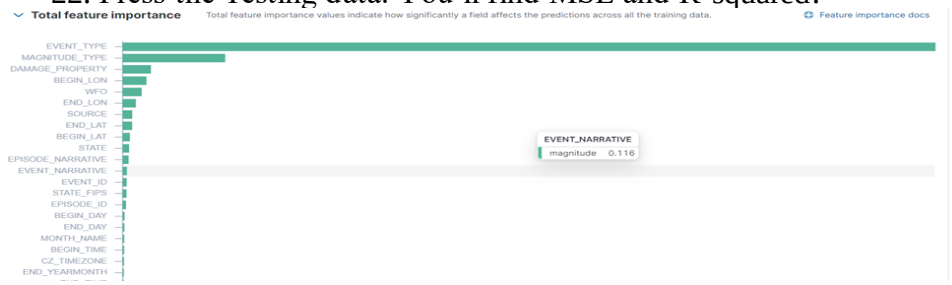
20. On the right, go to the models, click view.



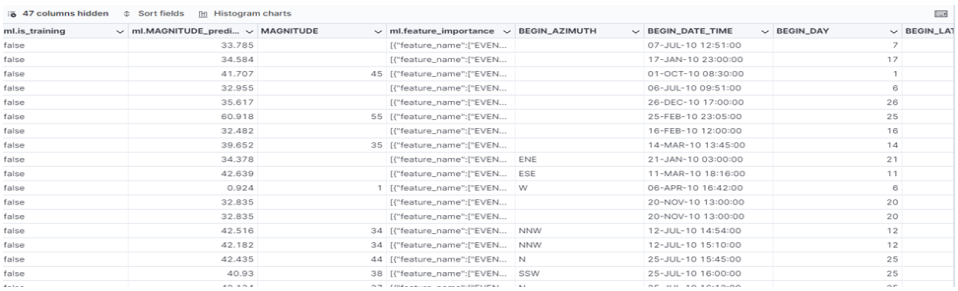
21. Press the training button and you'll find MSE and R-squared.



22. Press the Testing data. You'll find MSE and R-squared.



23. Scroll down and you'll find total feature importance. You'll find that EVENT_TYPE has the highest magnitude.

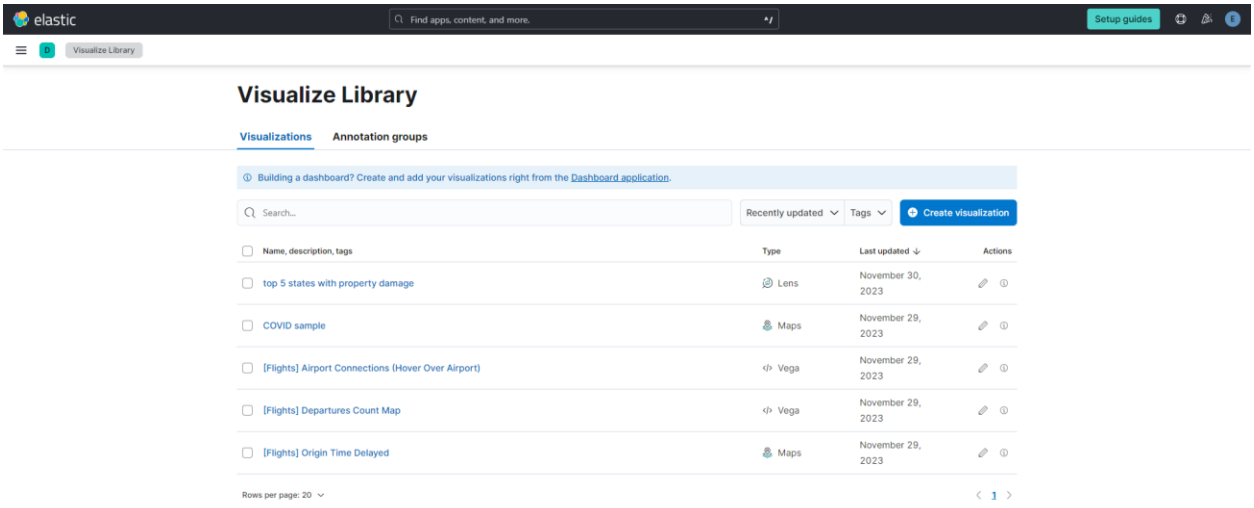


24. Scroll down further and you'll find ML_feature importance.

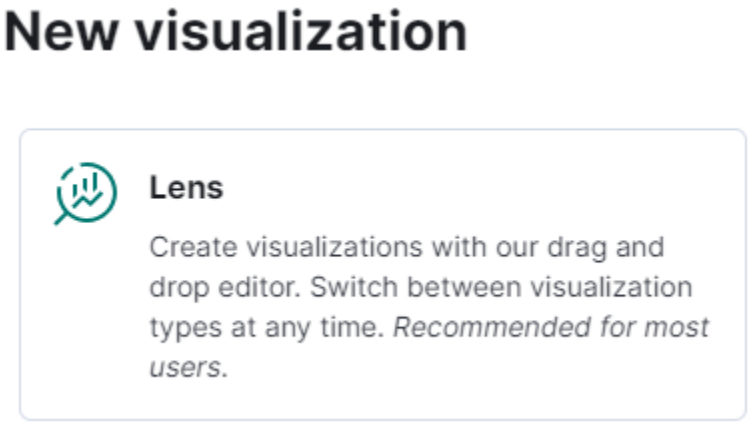
25. From here, click on one of the labels, and you'll see how much of an impact each feature has with the dataset. We learn that EVENT_TYPE has the highest magnitude of 9.95 from the 2010 dataset.
26. That's it for the tutorial on how to use machine learning for these CSV files.

Create Visualizations

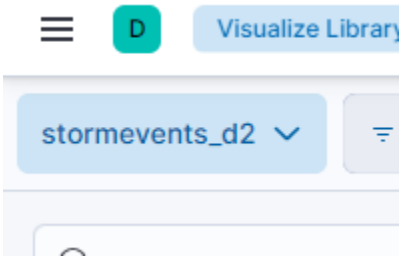
1. Open the Visalize Library in Elastic.
2. Click Create Visualization



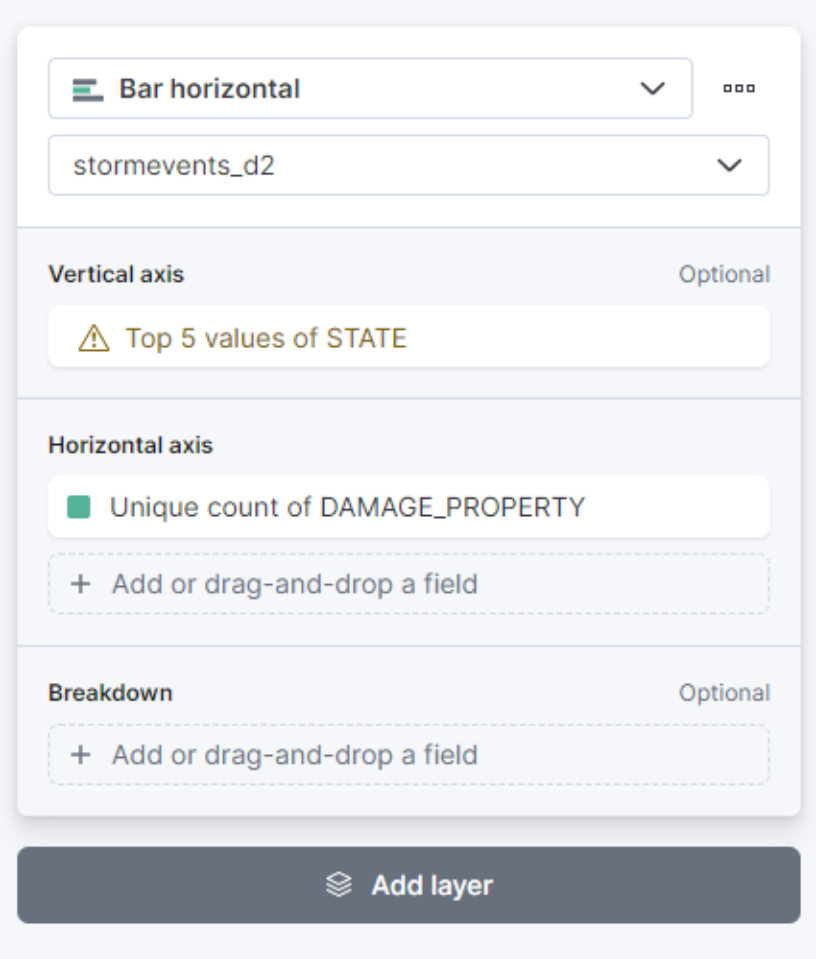
3. Select Lens.



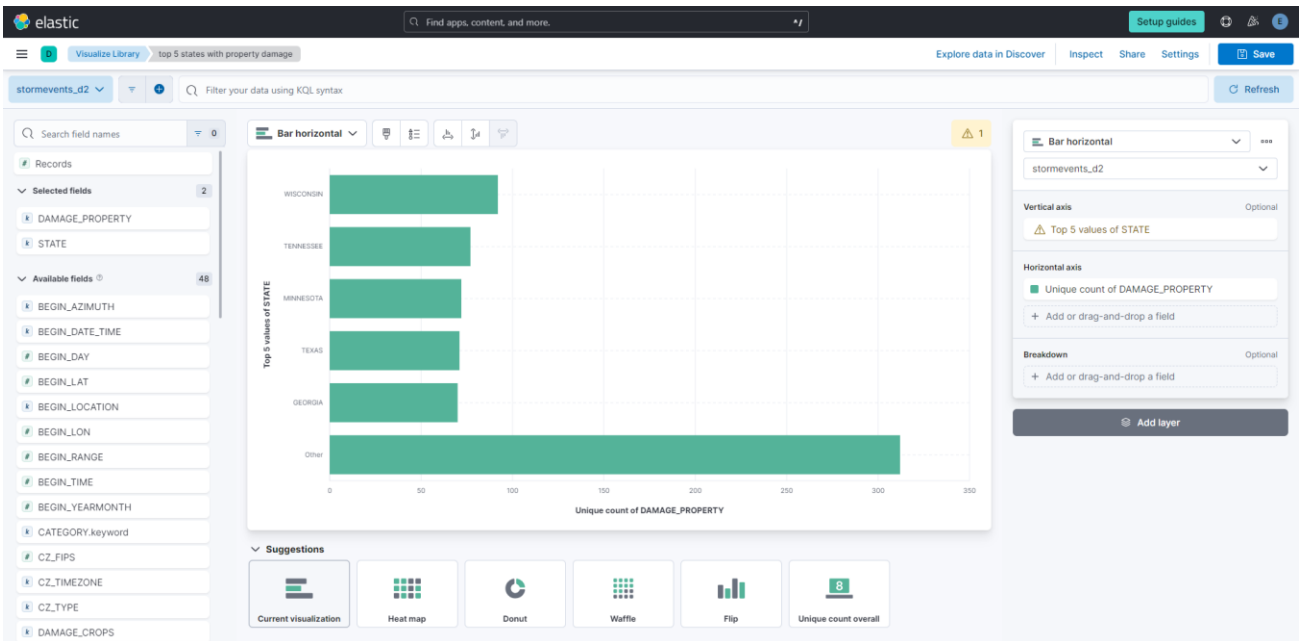
4. Make sure the stormevents selection is made:



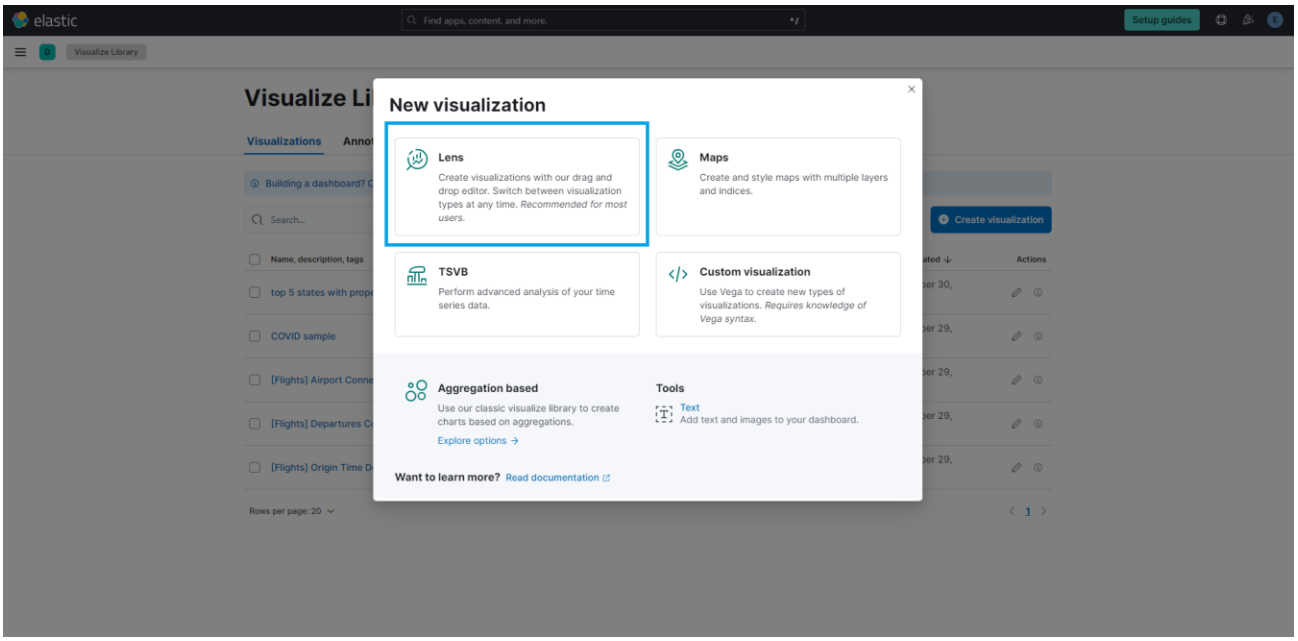
5. Drag and drop the TOP 5 VALUES OF STATE and UNIQUE COUNT OF PROPERTY DAMAGE



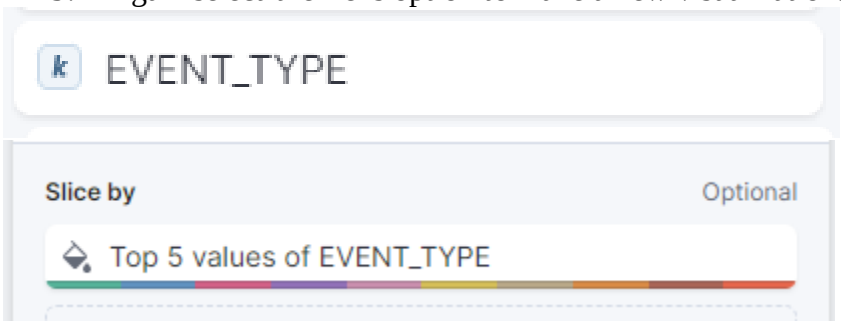
6. Select Bar Horizontal



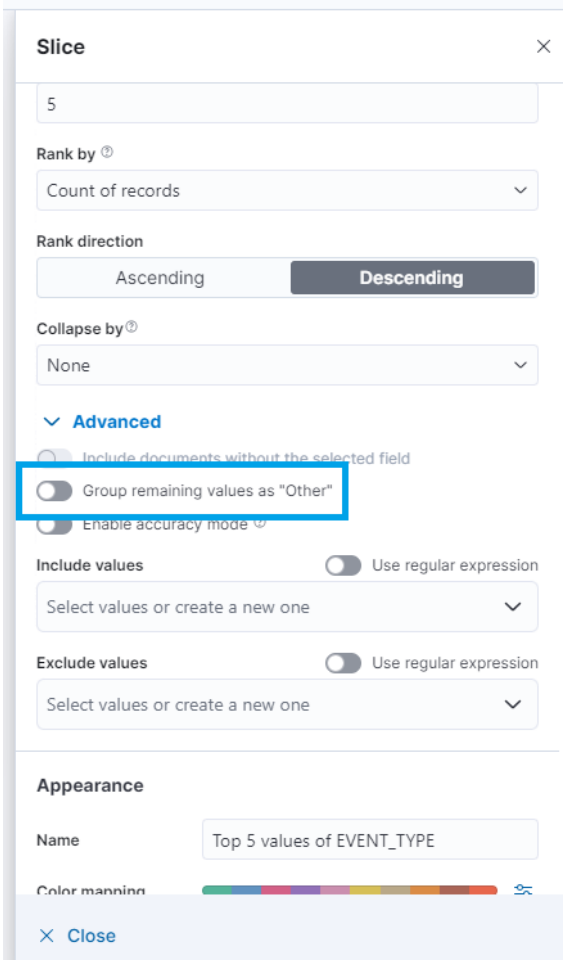
7. Save and name the Bar Chart Visualization.
8. Open the Visualize Library in Elastic to make another visualization.



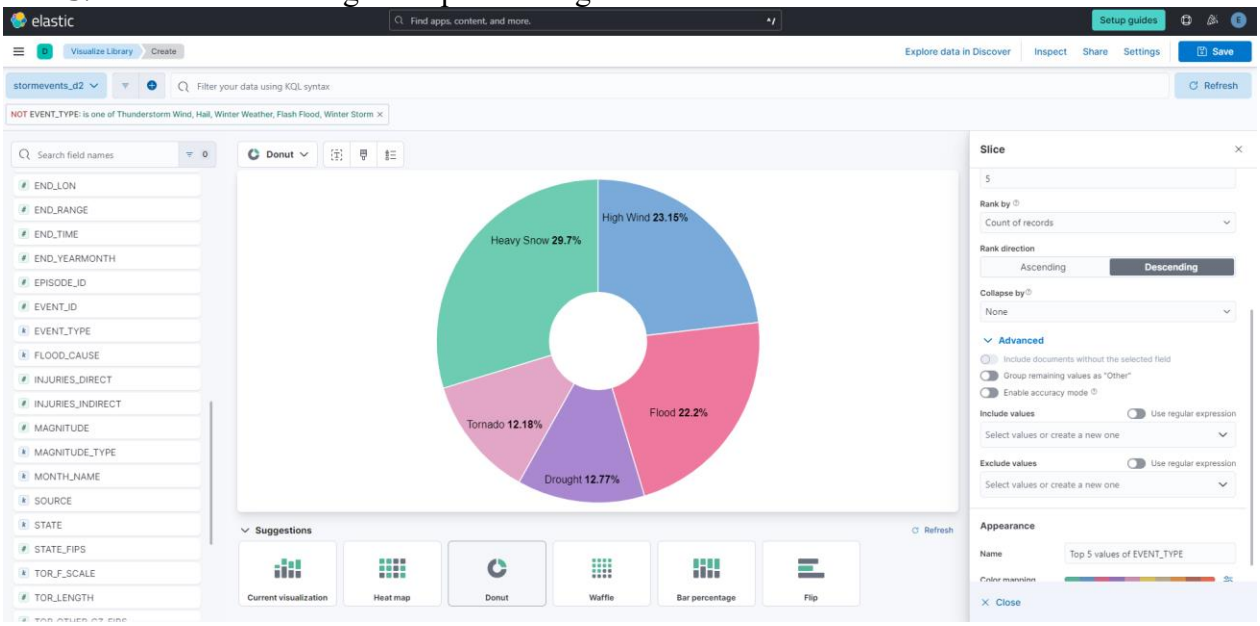
9. Again select the Lens option to make a new visualization.



10. Drag event type into the visualization space.
11. Select pie chart.
12. Click on Top 5 Values of EVENT TYPE, and click on Advanced settings.



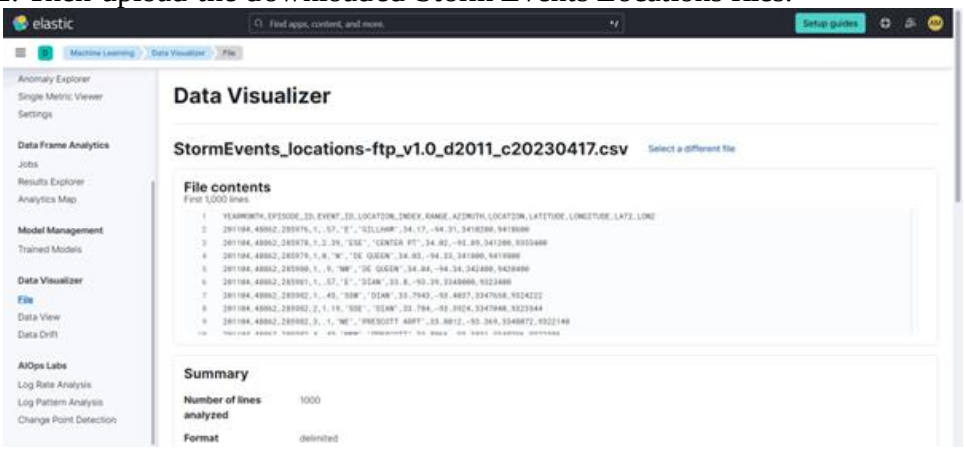
13. Turn off the setting: Group remaining values as “Other”.



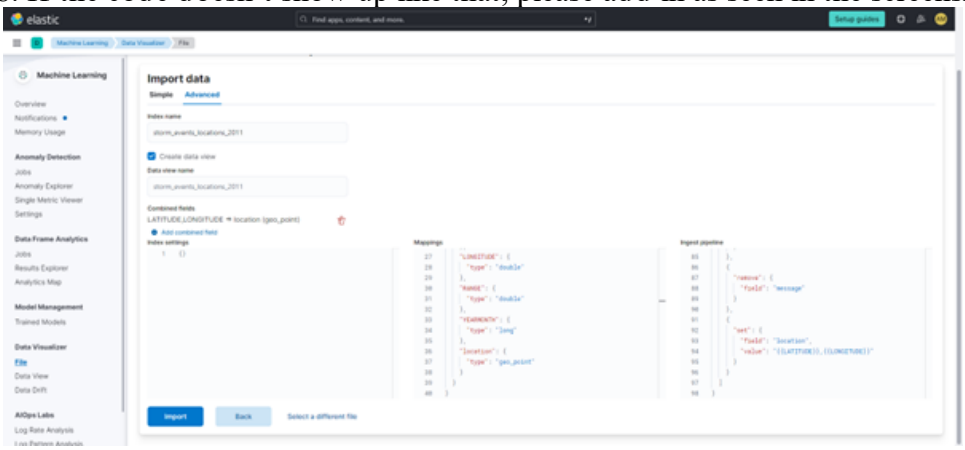
14. Save the Visualization as TOP 5 TYPES of STORM EVENTS. Save and exit.

Uploading the location data sets for 2010 and 2011:

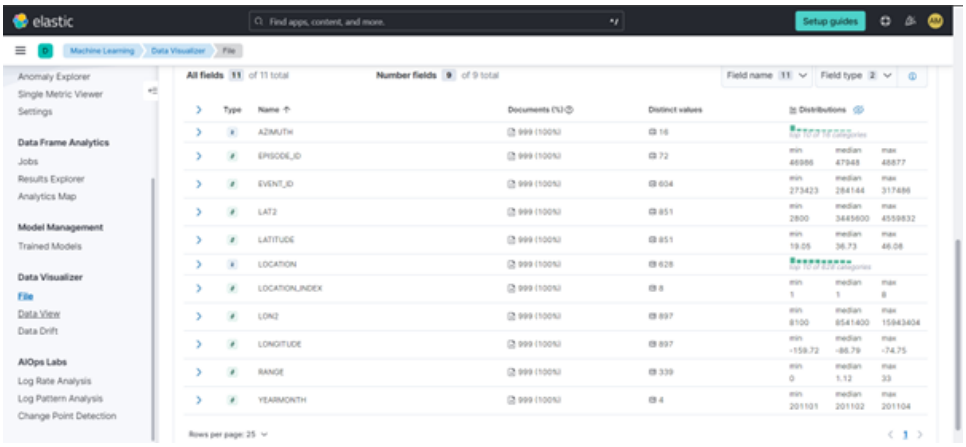
1. Go to: Top left menu -> Analytics -> Machine Learning -> Data Visualizer -> File
2. Then upload the downloaded Storm Events Locations files.



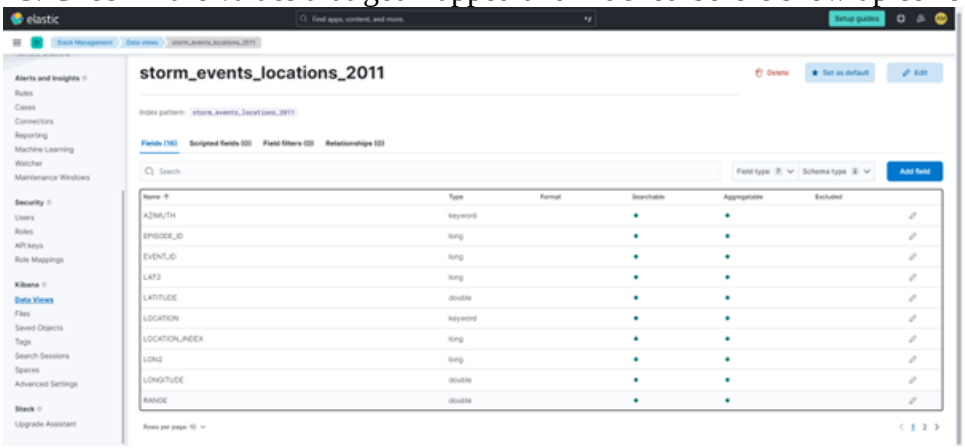
3. Check the indexing and mapping of the uploaded files.
4. Now check if latitude, longitude and location is indexed and mapped right in the advanced view of the imported data.
5. The lines need to show up for “latitude” and “longitude” as shown in the screenshot on the Mappings window on line 27. Also “location” needs to show up with the type of “geo_point” as in line 36 and 37 of the Mappings window.
6. Additionally, in the Ingest pipeline window, the code needs to be shown as in line 92 to 94.
7. Set needs to have “location” for the field type and “{{LATITUDE}}, {{LONGITUDE}}” for value type.
8. If the code doesn’t show up like that, please add in as seen in the screenshot.



9. Check if the map values show up as intended
10. After the upload is finished, scroll down and have a look at the data structure of the uploaded file, to check if latitude, longitude and location is indexed and mapped correctly as in the screenshot.

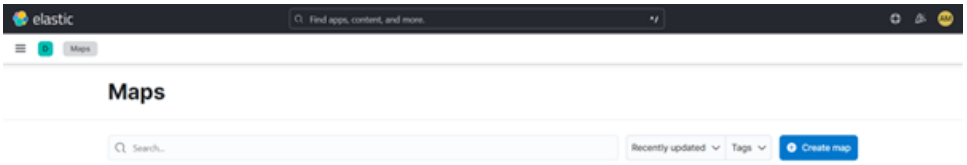


11. Check if the values show up in Kibana Data View
12. Go to: Top left menu -> Stack Management -> Kibana -> Data Views -> select your file
13. Check if the values that got mapped and indexed before show up correctly in the data view.

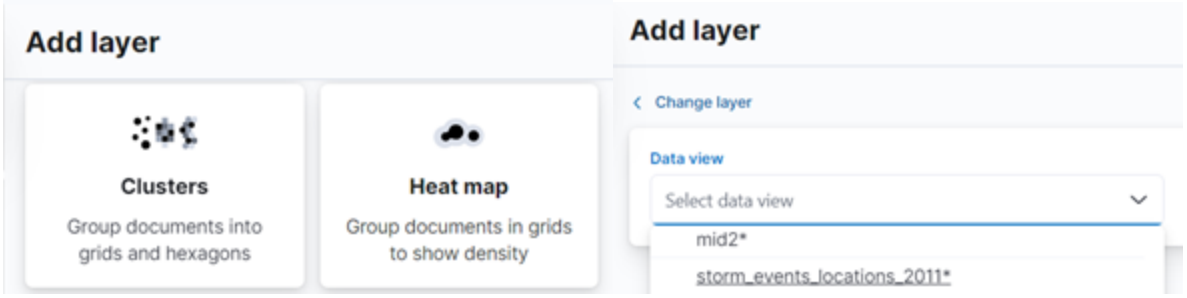


Creating the Storm Events Locations map:

1. Now go to Elastic main page -> Analytics -> Maps, and click on create map in the top right corner.



2. Then in the new view, click on “add layer” at the right side, scroll down and select clusters.
3. For the data view now select the uploaded mapped and indexed locations file.

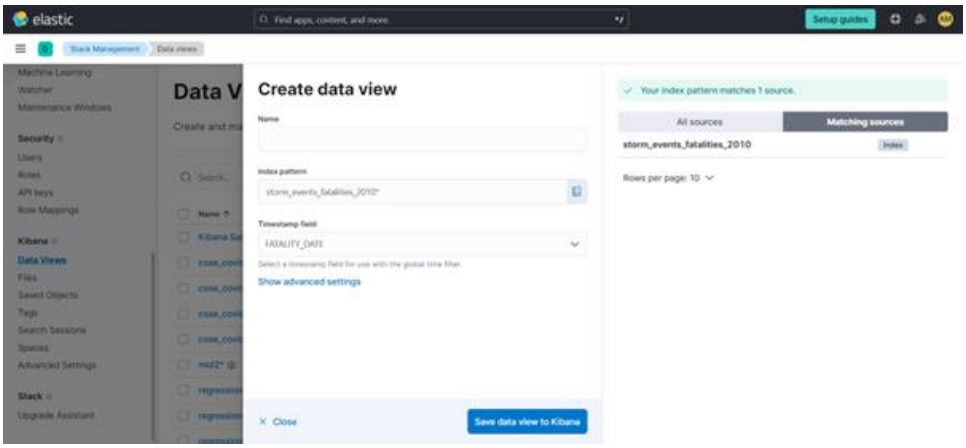


4. Now select “location” for cluster field and adjust the time span on the top right corner.
5. Select January 1st with a time of 00:00 am to December 31st with a time of 23:30 pm and either for the whole year of 2010 or 2011 depending on the file that you working with right now.
6. Update and refresh, the map should look like shown in the screenshot now. Save the view.

7. Save the map and exit.
8. Repeat the steps above to have the maps for both years.

Creating the storm events fatalities data view:

1. Upload the data sets for the fatalities as described in earlier sections.
2. Create a visualization as described in earlier sections.
3. Go to: Top left menu -> Stack Management -> Kibana -> Data Views -> Create Data View

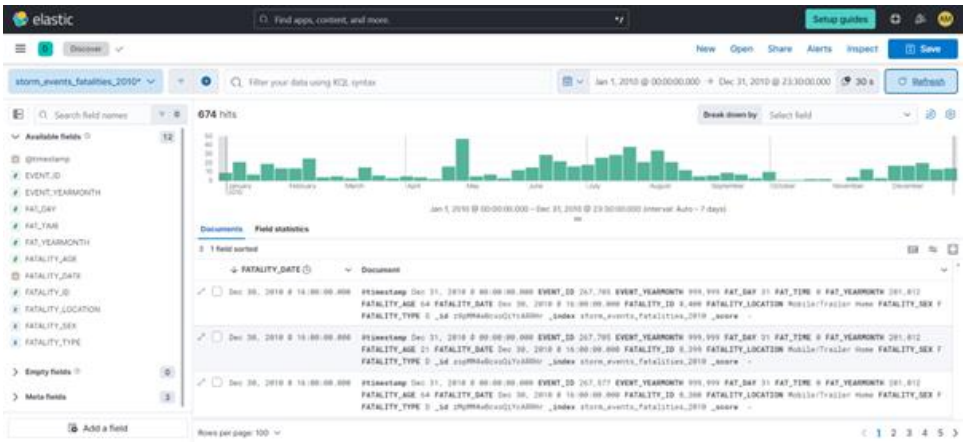


4. Create a bar chart in the same way as you already before in the visualizations section, but this time use the storm_events_fatalities_2010* data set that you uploaded.
5. Select Bar Vertical.
6. Select descending order.
7. Select show numbers in bar.
8. Adjust the date to the whole year with absolute dates in the top right corner.

9. Save the view and exit.
10. Repeat the steps for the 2011 data set.

Creating the storm events fatalities discovery view:

1. Go to: Elastic main page -> analytics -> discover
2. Select the storm_events_fatalities_2010* dataset that you uploaded earlier.
3. Adjust the date to the whole year with absolute dates in the top right corner.



4. Save and exit.
5. Repeat the steps for the 2011 data set.

References

1. URL of Data Source, https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/

-StormEvents_details-ftp_v1.0_d2010_c20220425.csv.gz

-StormEvents_details-ftp_v1.0_d2011_c20230417.csv.gz

-StormEvents_fatalities-ftp_v1.0_d2010_c20220425.csv.gz

-StormEvents_fatalities-ftp_v1.0_d2011_c20230417.csv.gz

-StormEvents_locations-ftp_v1.0_d2010_c20220425.csv.gz

-StormEvents_locations-ftp_v1.0_d2011_c20230417.csv.gz

2. URL of your Github: https://github.com/MichaelCIS/3200_Stormevent.git