CIS3200 Term Project Tutorial (Group 2)

Authors:

Instructor: **Jongwook Woo**

Date 12/18/2020

Lab Tutorial

Christopher Valdepena (cvaldep3@calstatela.edu)

Jiancong Zhang (jzhang100@calstatela.edu)

My Truong (mtruon42@calstatela.edu)

Ryan Applegate(rappleg@calstatela.edu)

Getting started on Elastic Cloud with a Sample Dataset “Superstore”

Objectives

In this hands-on lab, you will learn how to:

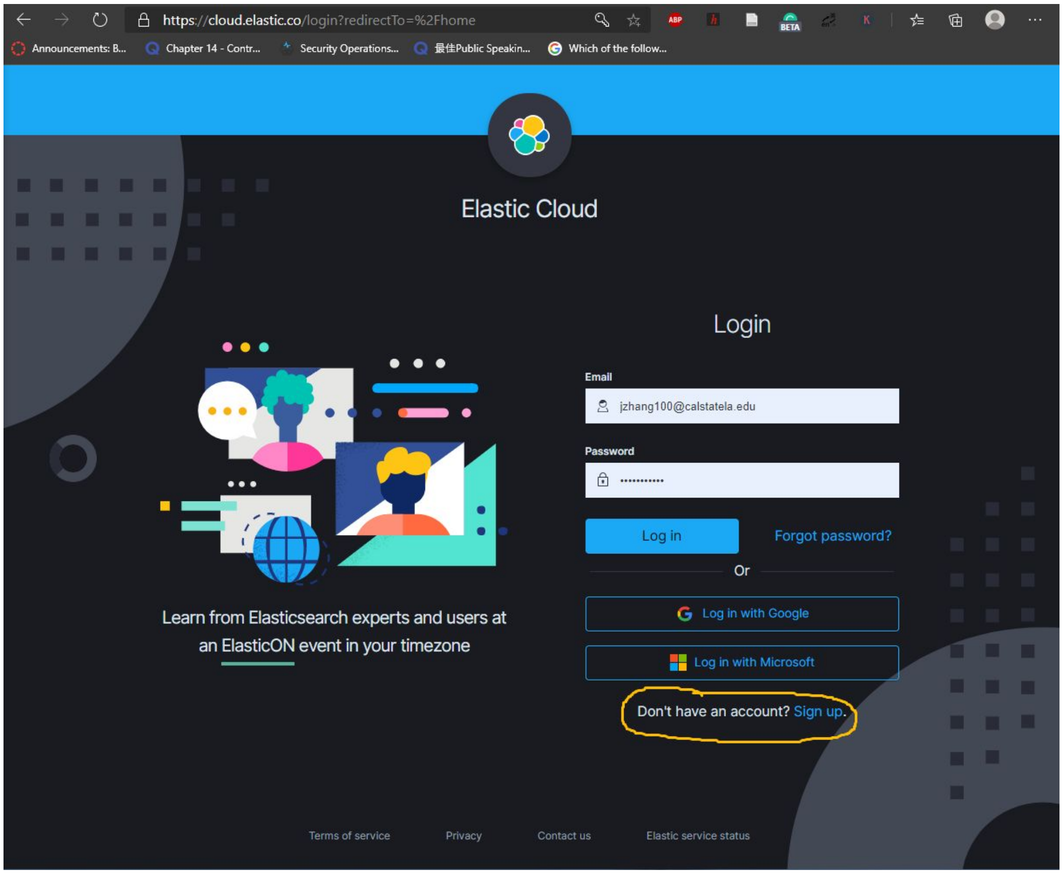
- Find clear data sets
- Visit and create an account for Elastic Cloud
- Implement data sets
- Create Simple Graphs Visualizations
- Create GEO_Map_Visualizations

Platform Specifications:

- ElasticSearch & Kibana
- CPU Speed: 3.0GHz
- # of CPU cores: 4
- # of nodes: 1 node
- Total Memory Size: 500MB

Step 1: Visit and create an account for Elastic Cloud

1. Go to https://www.elastic.co/cloud/as-a-service
2. Register then Log into your ES (Elastic Cloud) account by using your email account



https://cloud.elastic.co/login?redirectTo=%2Fhome

Elastic Cloud

Login

Email: jzhang100@calstatela.edu

Password: •••••••••••

Don't have an account? Sign up.

3. Click on the verification link in the email that was sent to your ".edu" address

4. After email verification, you will be prompted to create a password for your Elastic Cloud.

5. Log into your Elastic Cloud account

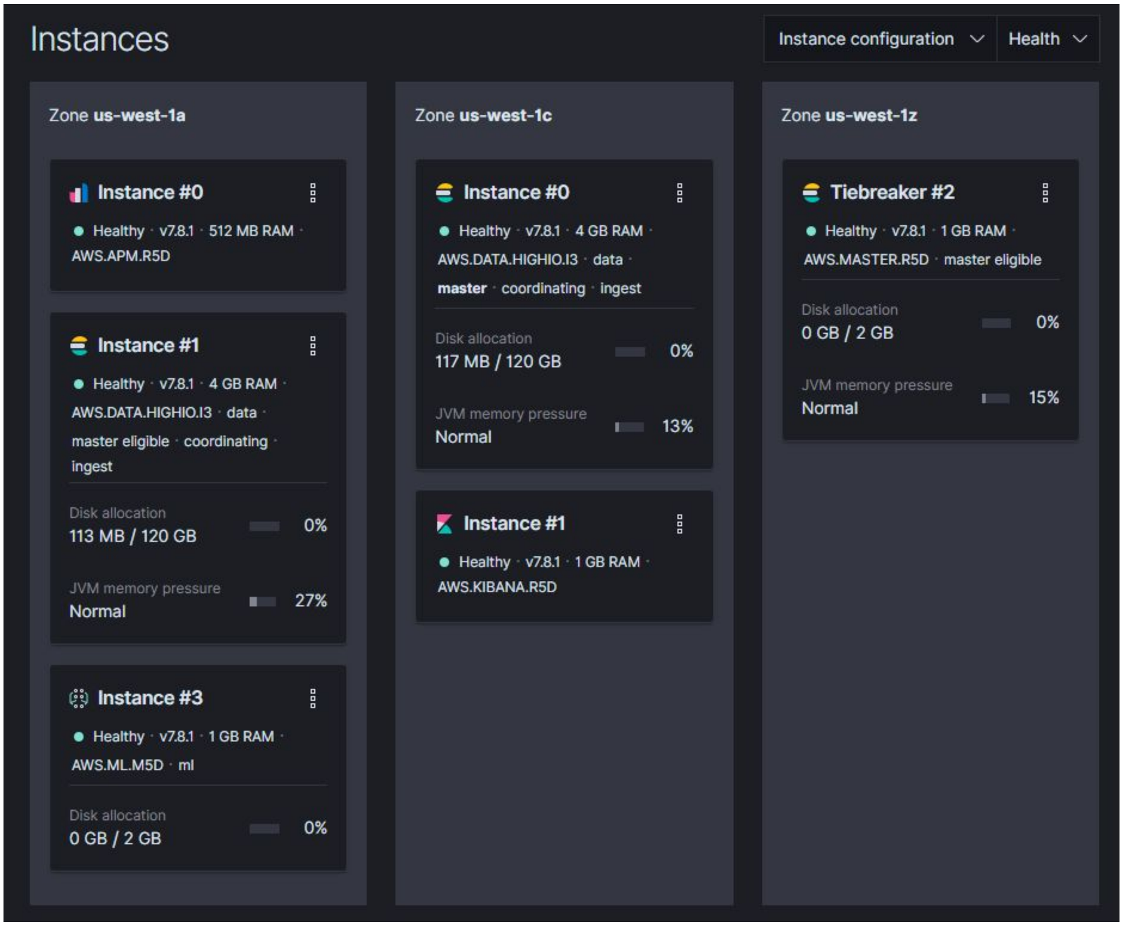
6. NOTE: If the verification email expires, go to https://cloud.elastic.co/forgot and enter the student email address to trigger a new verification email. Create your first hosted Elasticsearch cluster

Set up the Instances

7. Once you have signed in you will see an overview of your deployments. Click on the "Create Deployment" button

8. Name your deployment, choose the provider and region that you prefer. For trials the deployment size please set up the instances as the next graph showed

9. One thing please pay attention, Since the max memory option may change over time. please just select the maximum option that they provide you.



Instances

Zone us-west-1a

Instance #0
● Healthy · v7.8.1 · 512 MB RAM · AWS.APM.R5D

Instance #1
● Healthy · v7.8.1 · 4 GB RAM · AWS.DATA.HIGHIO.I3 · data · master eligible · coordinating · ingest
Disk allocation 113 MB / 120 GB 0%
JVM memory pressure Normal 27%

Instance #3
● Healthy · v7.8.1 · 1 GB RAM · AWS.ML.M5D · ml
Disk allocation 0 GB / 2 GB 0%

Zone us-west-1c

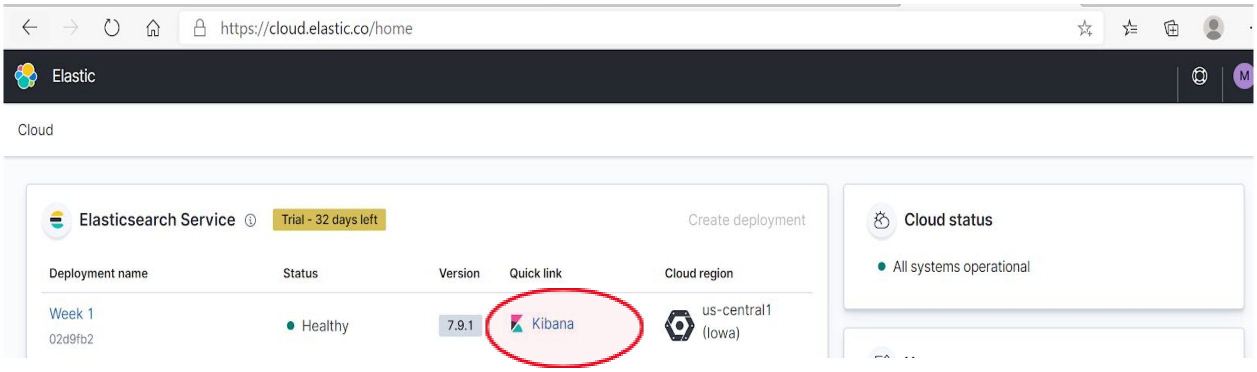
Instance #0
● Healthy · v7.8.1 · 4 GB RAM · AWS.DATA.HIGHIO.I3 · data · **master** · coordinating · ingest
Disk allocation 117 MB / 120 GB 0%
JVM memory pressure Normal 13%

Instance #1
● Healthy · v7.8.1 · 1 GB RAM · AWS.KIBANA.R5D

Zone us-west-1z

Tiebreaker #2
● Healthy · v7.8.1 · 1 GB RAM · AWS.MASTER.R5D · master eligible
Disk allocation 0 GB / 2 GB 0%
JVM memory pressure Normal 15%

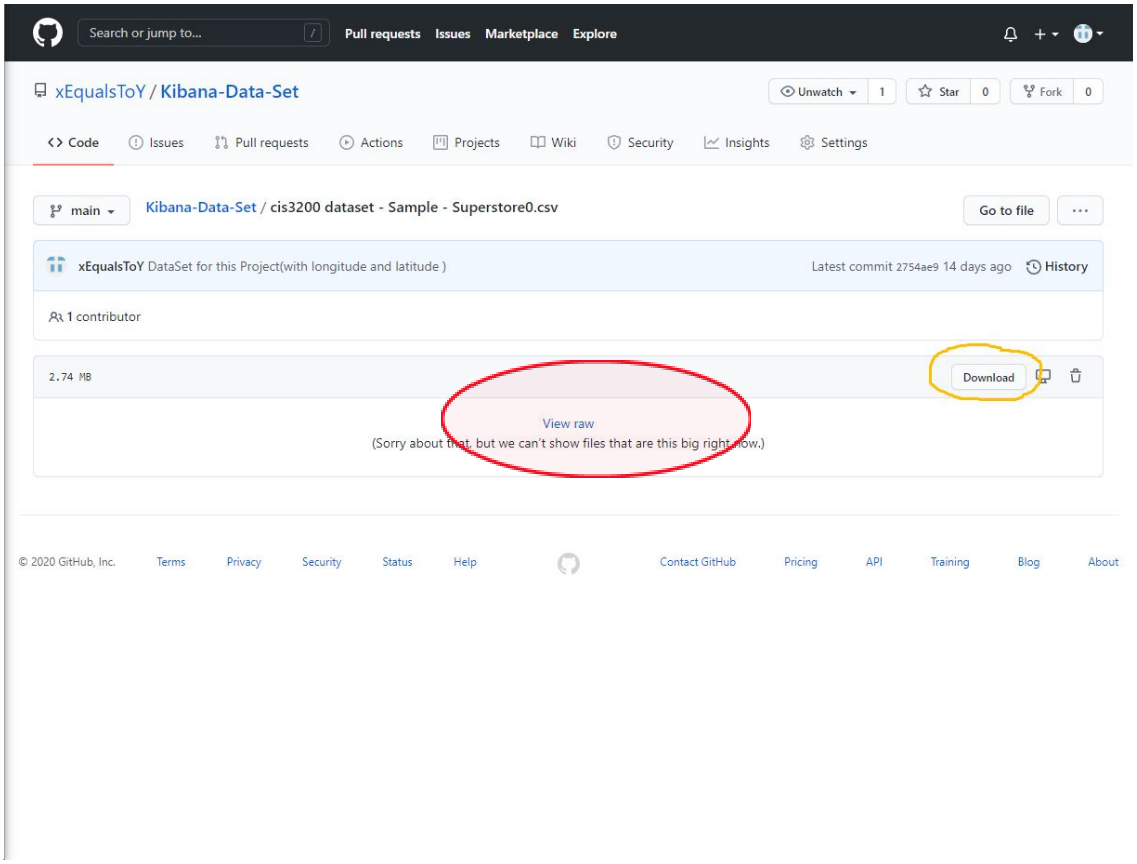
10. After logging into the ES account, you will see the following page as the image below. Then, click on your Kibana:



Elasticsearch Service ⓘ Trial - 32 days left

Deployment name	Status	Version	Quick link	Cloud region
Week 1 02d9fb2	● Healthy	7.9.1	Kibana	us-central1 (Iowa)

Cloud status
● All systems operational



xEqualsToY / Kibana-Data-Set

Kibana-Data-Set / cis3200 dataset - Sample - Superstore0.csv

xEqualsToY DataSet for this Project(with longitude and latitude)

Latest commit 2754ae9 14 days ago

1 contributor

2.74 MB

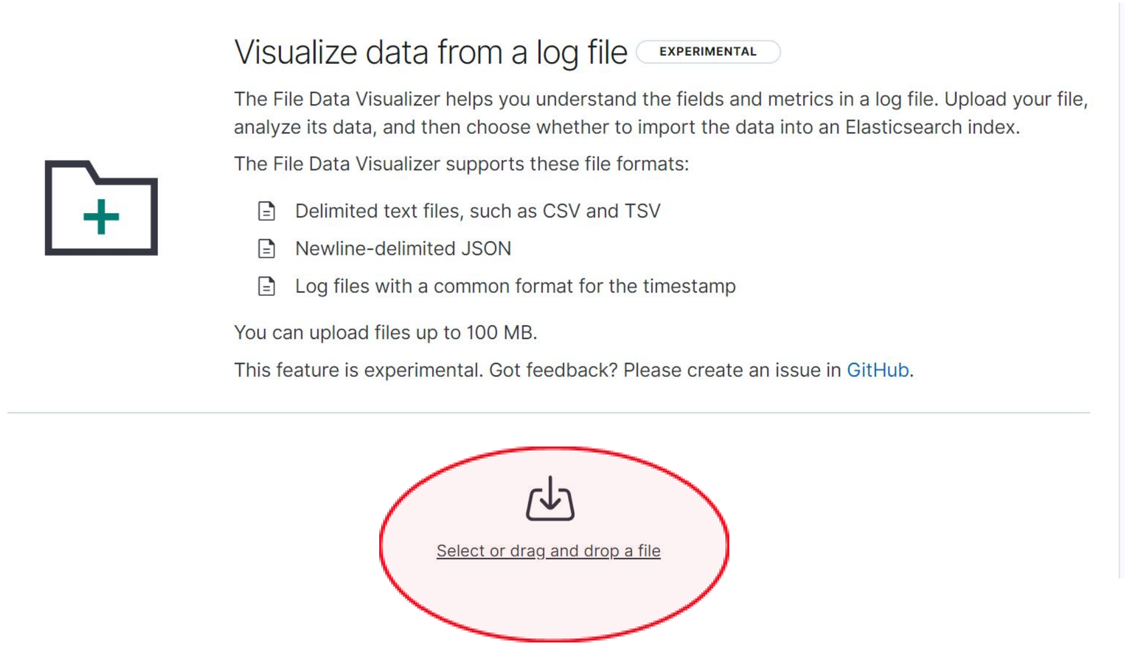
View raw

(Sorry about that, but we can't show files that are this big right now.)

© 2020 GitHub, Inc.

11. Next, you will need to download the data sets from [Kibana-Data-Set/cis3200 dataset - Sample - Superstore0.csv at main · xEqualsToY/Kibana-Data-Set (github.com)](github.com) (see next graph)

12. Come back to the Kibana pages, drag the data set into the red area.



Visualize data from a log file

The File Data Visualizer helps you understand the fields and metrics in a log file. Upload your file, analyze its data, and then choose whether to import the data into an Elasticsearch index.

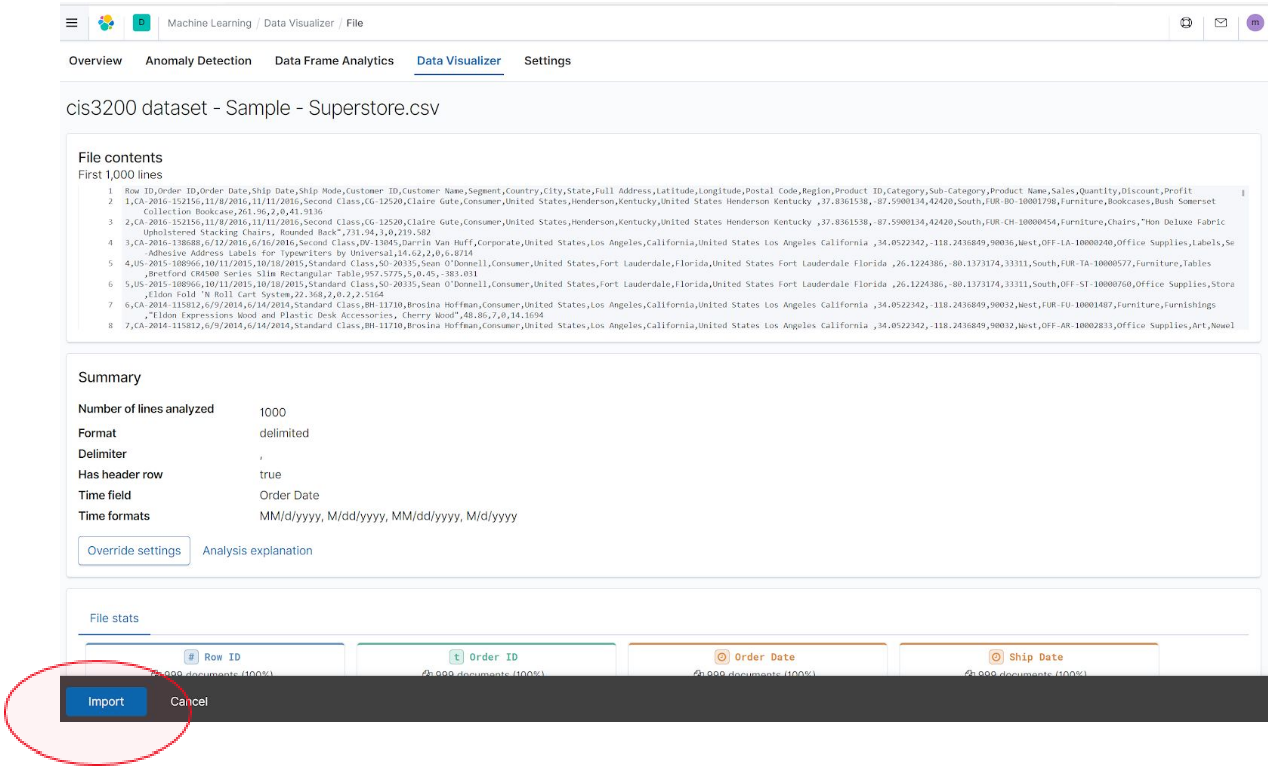
The File Data Visualizer supports these file formats:

- Delimited text files, such as CSV and TSV
- Newline-delimited JSON
- Log files with a common format for the timestamp

You can upload files up to 100 MB.

This feature is experimental. Got feedback? Please create an issue in GitHub.

13. Then, you will see this page on your screen.



cis3200 dataset - Sample - Superstore.csv

File contents

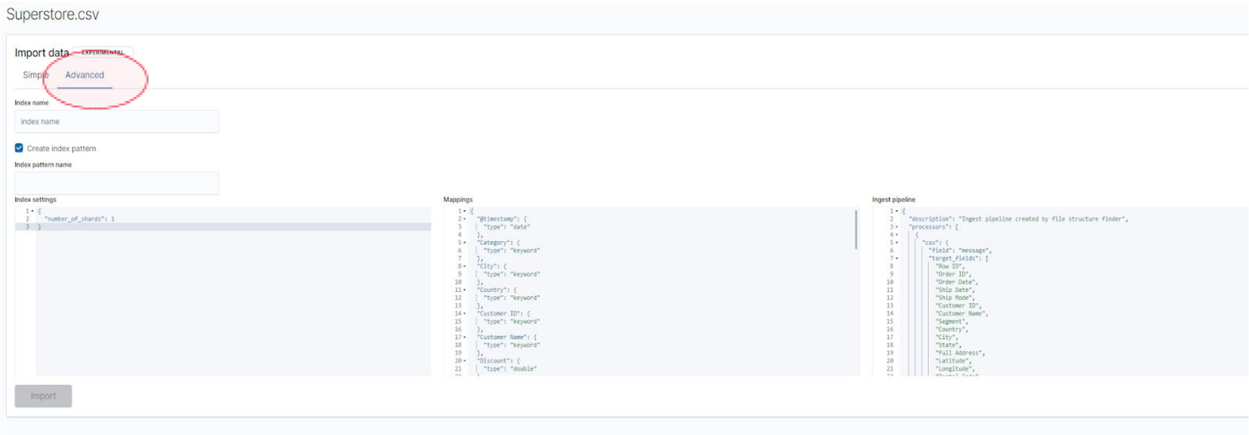
First 1,000 lines

```
Row ID,Order ID,Order Date,Ship Date,Ship Mode,Customer ID,Customer Name,Segment,Country,City,State,Full Address,Latitude,Longitude,Postal Code,Region,Product ID,Category,Sub-Category,Product Name,Sales,Quantity,Discount,Profit
1,CA-2016-152156,11/8/2016,11/11/2016,Second Class,CG-12520,Claire Gute,Consumer,United States,Henderson,Kentucky,United States Henderson Kentucky ,37.8361538,-87.5900134,42420,South,FUR-BO-10001798,Furniture,Bookcases,Bush Somerset
    Collection Bookcase,261.96,2,0,41.9136
2,CA-2016-152156,11/8/2016,11/11/2016,Second Class,CG-12520,Claire Gute,Consumer,United States,Henderson,Kentucky,United States Henderson Kentucky ,37.8361538,-87.5900134,42420,South,FUR-CH-10000454,Furniture,Chairs,"Hon Deluxe Fabric
    Upholstered Stacking Chairs, Rounded Back",731.94,3,0,219.582
3,CA-2016-138688,6/12/2016,6/16/2016,Second Class,DV-13045,Darrin Van Huff,Corporate,United States,Los Angeles,California,United States Los Angeles California ,34.0522342,-118.2436849,90036,West,OFF-LA-10000240,Office Supplies,Labels,Se
    -Adhesive Address Labels for Typewriters by Universal,14.62,2,0,6.8714
4,US-2015-108966,10/11/2015,10/18/2015,Standard Class,SO-20335,Sean O'Donnell,Consumer,United States,Fort Lauderdale,Florida,United States Fort Lauderdale Florida ,26.1224386,-80.1373174,33311,South,FUR-TA-10000577,Furniture,Tables
    ,Bretford CR4500 Series Slim Rectangular Table,957.5775,5,0.45,-383.031
5,US-2015-108966,10/11/2015,10/18/2015,Standard Class,SO-20335,Sean O'Donnell,Consumer,United States,Fort Lauderdale,Florida,United States Fort Lauderdale Florida ,26.1224386,-80.1373174,33311,South,OFF-ST-10000760,Office Supplies,Stora
    ,Eldon Fold 'N Roll Cart System,22.368,2,0.2,2.5164
6,CA-2014-115812,6/9/2014,6/14/2014,Standard Class,BH-11710,Brosina Hoffman,Consumer,United States,Los Angeles,California,United States Los Angeles California ,34.0522342,-118.2436849,90032,West,FUR-FU-10001487,Furniture,Furnishings
    ,"Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood",48.86,7,0,14.1694
7,CA-2014-115812,6/9/2014,6/14/2014,Standard Class,BH-11710,Brosina Hoffman,Consumer,United States,Los Angeles,California,United States Los Angeles California ,34.0522342,-118.2436849,90032,West,OFF-AR-10002833,Office Supplies,Art,Newel
```

Summary

Number of lines analyzed	1000
Format	delimited
Delimiter	,
Has header row	true
Time field	Order Date
Time formats	MM/d/yyyy, M/dd/yyyy, MM/dd/yyyy, M/d/yyyy

14. Once selecting the *Import* button, we want to navigate to the *Advanced* tab for set the mapping for geo spatial data.



Index name

index name

☑ Create index pattern

Index pattern name

Index settings

```
{
  "number_of_shards": 1
}
```

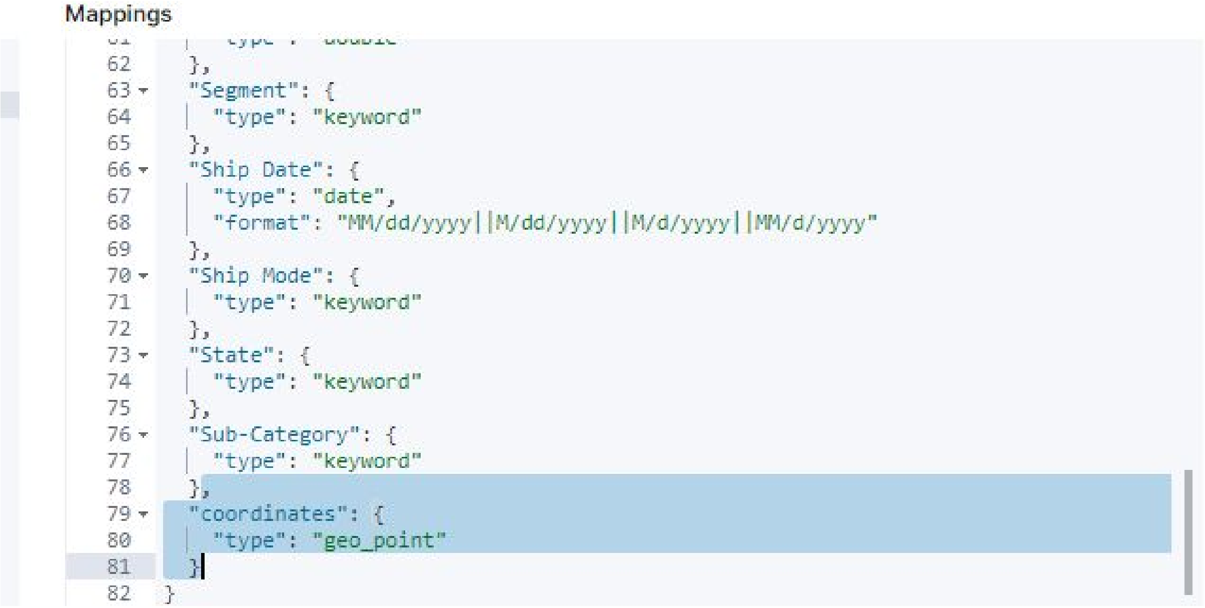
Mappings

```
{
  "@timestamp": {
    "type": "date"
  },
  "Category": {
    "type": "keyword"
  },
  "City": {
    "type": "keyword"
  },
  "Country": {
    "type": "keyword"
  },
  "Customer ID": {
    "type": "keyword"
  },
  "Customer Name": {
    "type": "keyword"
  },
  "Discount": {
    "type": "double"
```

Ingest pipeline

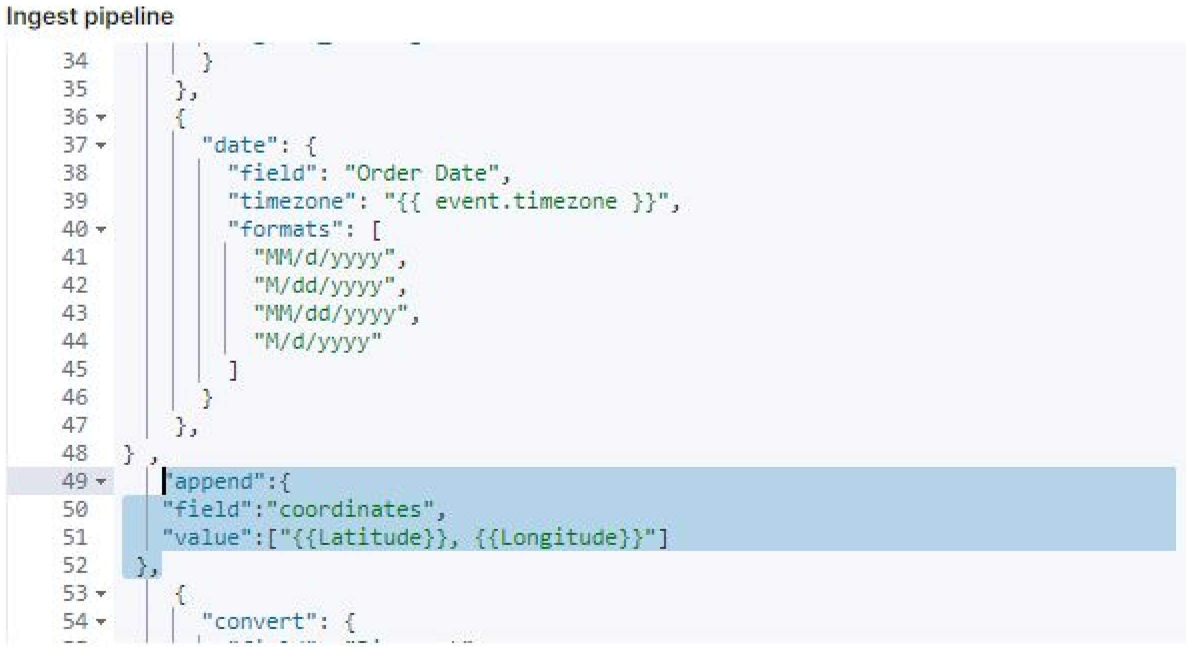
```
{
  "description": "Ingest pipeline created by file structure finder",
  "processors": [
    {
      "csv": {
        "field": "message",
        "target_fields": [
          "Row ID",
          "Order ID",
          "Order Date",
          "Ship Date",
          "Ship Mode",
          "Customer ID",
          "Customer Name",
          "Segment",
          "Country",
          "City",
          "State",
          "Full Address",
          "Latitude",
          "Longitude",
```

15. Now we add the “coordinates”: { “type”: “geo_point” } to declare the variable coordinates with the data type geo_point for our geo spatial visuals. We do not have to add any mapping for the *Longitude and Latitude* dimensions since they are already present in the downloaded dataset from our github.


Mappings

```
    "type": "double"
  },
  "Segment": {
    "type": "keyword"
  },
  "Ship Date": {
    "type": "date",
    "format": "MM/dd/yyyy||M/dd/yyyy||M/d/yyyy||MM/d/yyyy"
  },
  "Ship Mode": {
    "type": "keyword"
  },
  "State": {
    "type": "keyword"
  },
  "Sub-Category": {
    "type": "keyword"
  },
  "coordinates": {
    "type": "geo_point"
  }
}
```

16. Next, we head over to the *Ingest Pipeline* to declare the values for our *coordinates* dimension.

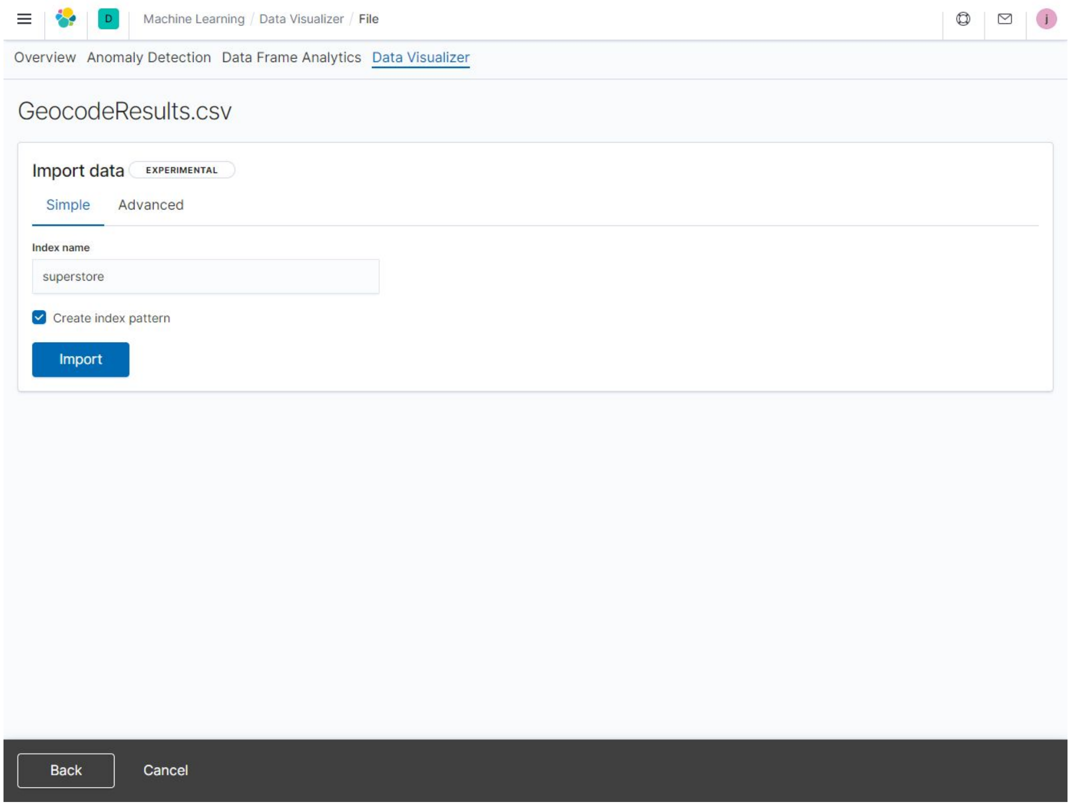


Ingest pipeline

```
      }
    },
    {
      "date": {
        "field": "Order Date",
        "timezone": "{{ event.timezone }}",
        "formats": [
          "MM/d/yyyy",
          "M/dd/yyyy",
          "MM/dd/yyyy",
          "M/d/yyyy"
        ]
      }
    },
} ,
   "append":{
   "field":"coordinates",
   "value":["{{Latitude}}, {{Longitude}}"]
 },
    {
      "convert": {
```

17. Then, create a name for your dataset at the “Index name” tab as you can see in the red circle. **One thing to pay attention to here, make sure to check the “create index pattern” box First!**

After that, please click import in the red circle. Then, you will see the following pipeline as the picture below which shows the imported process with geo_point. If there is no error, the process is completed.

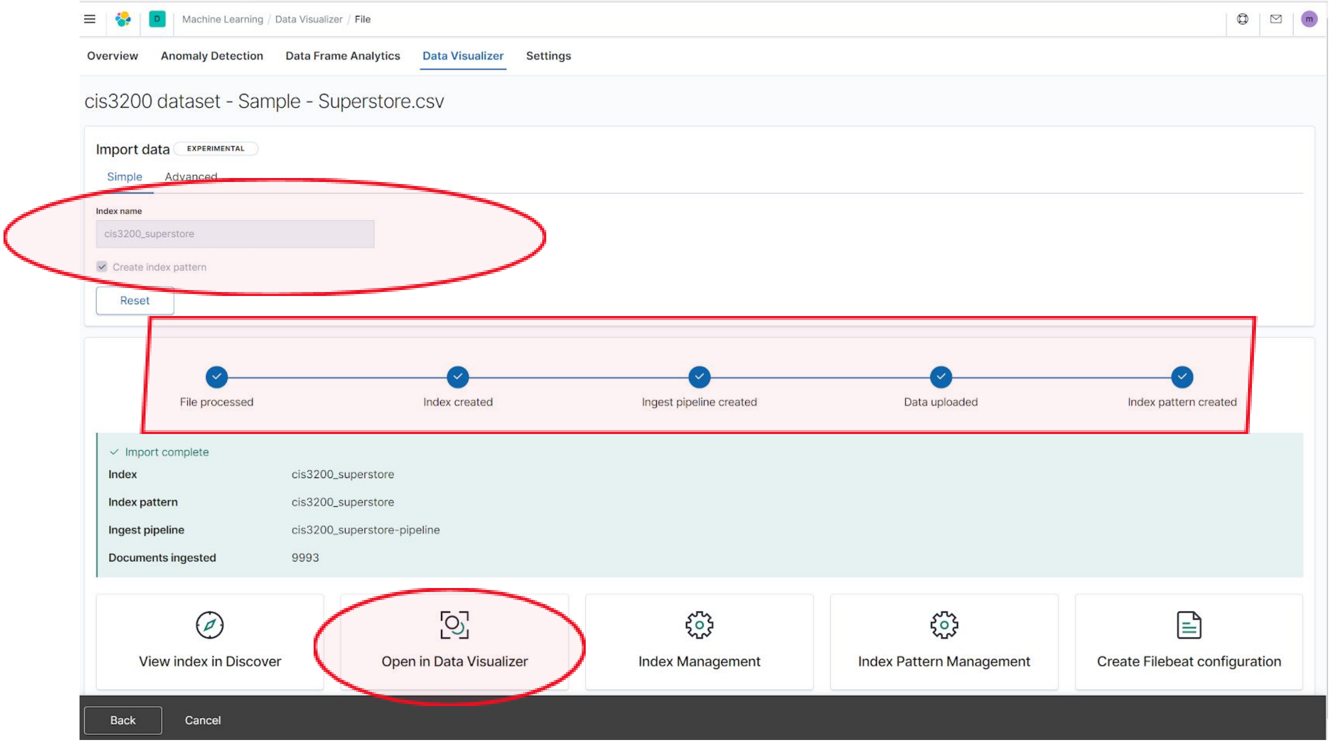


GeocodeResults.csv

Index name: superstore

☑ Create index pattern

18. Next, click on the “Open in Data Visualizer.”



cis3200 dataset - Sample - Superstore.csv

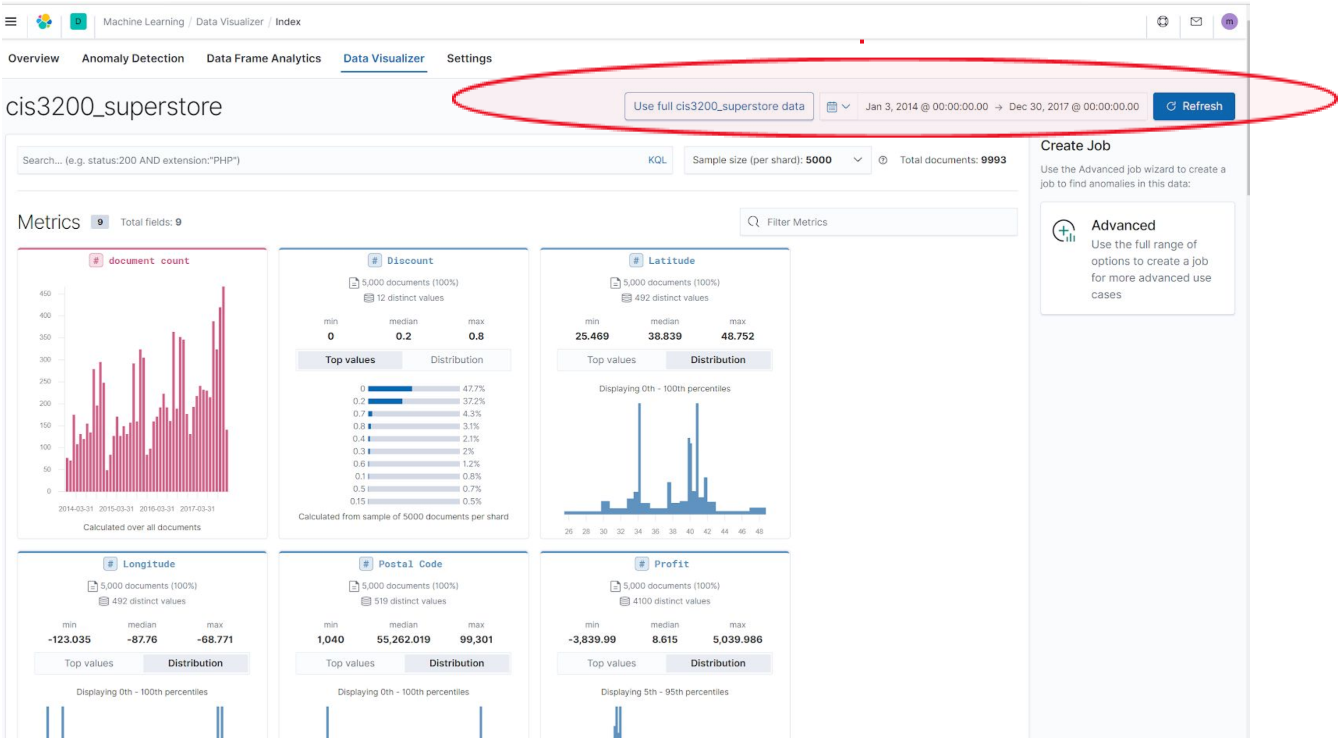
Index name: cis3200_superstore

☑ Create index pattern

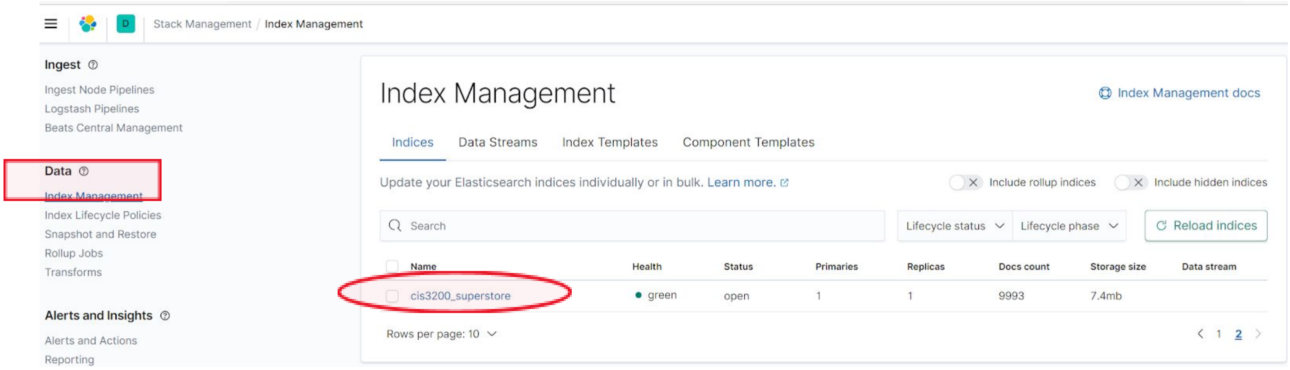
✓ Import complete

Index	cis3200_superstore
Index pattern	cis3200_superstore
Ingest pipeline	cis3200_superstore-pipeline
Documents ingested	9993

19. Then, it will take you to the window where you can check the time range, and also see metrics and total fields of your dataset.

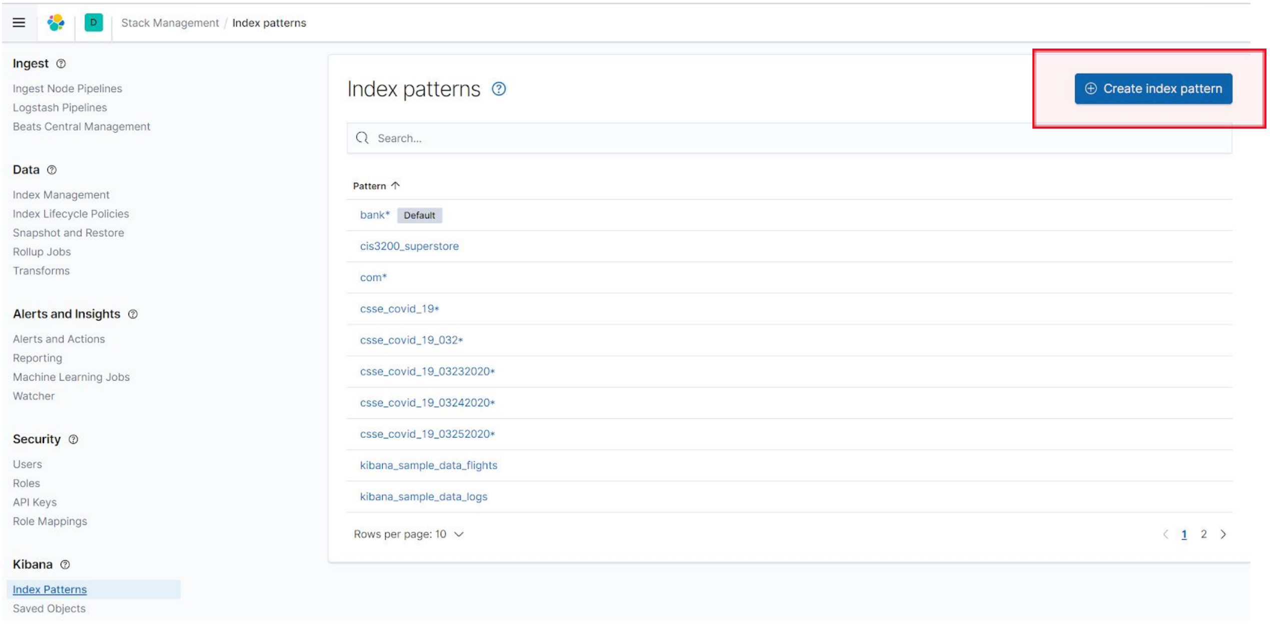


17. Next, click on the "Index Management" under the "Data" tab. It will take you to the step where you can create Index Management.

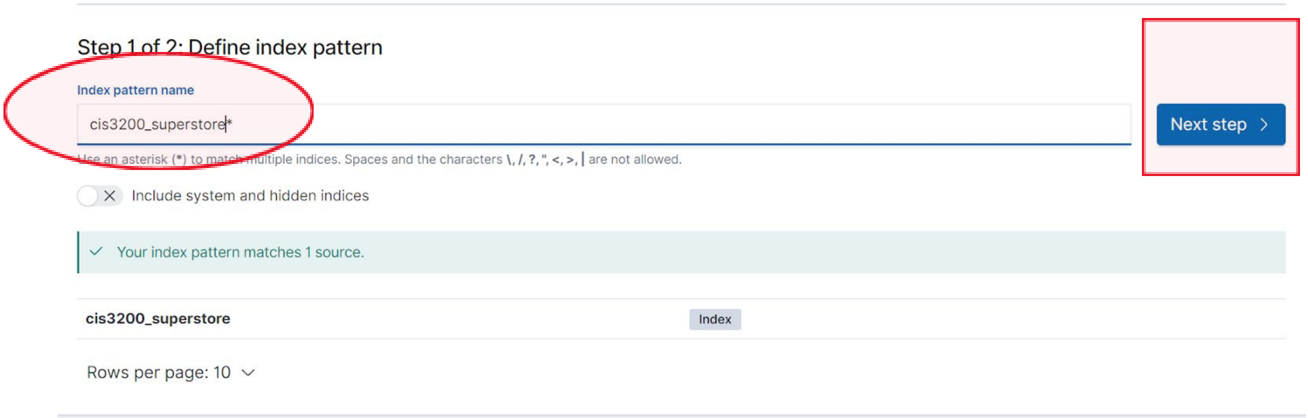


Name	Health	Status	Primaries	Replicas	Docs count	Storage size	Data stream
cis3200_superstore	● green	open	1	1	9993	7.4mb	

18. Then, click on "Create index pattern" to create the new one and follow with creating a new name for the index as you can see in the image below with the red circles are:



Pattern ↑
bank* Default
cis3200_superstore
com*
csse_covid_19*
csse_covid_19_032*
csse_covid_19_03232020*
csse_covid_19_03242020*
csse_covid_19_03252020*
kibana_sample_data_flights
kibana_sample_data_logs


Step 1 of 2: Define index pattern

Index pattern name: cis3200_superstore*

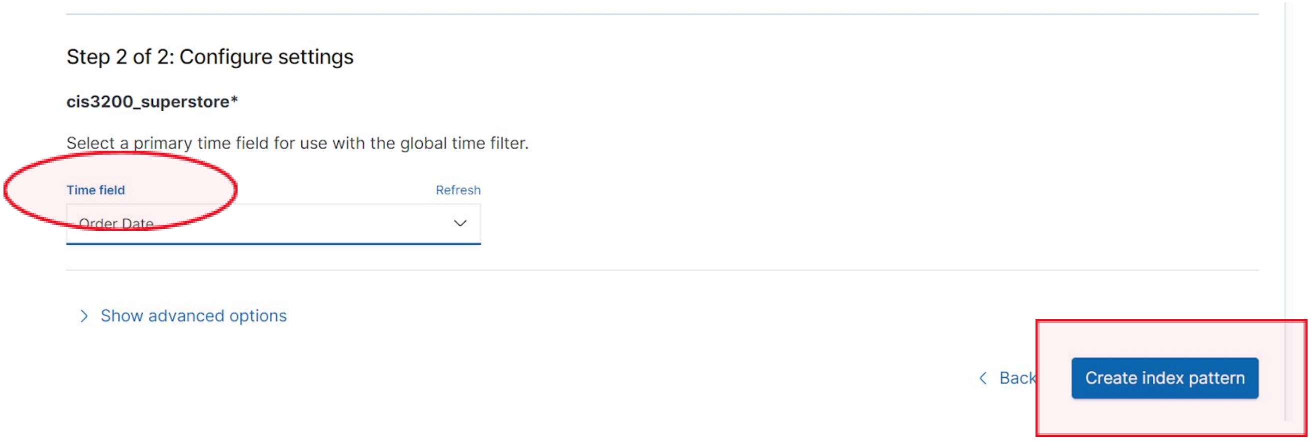
Use an asterisk (*) to match multiple indices. Spaces and the characters \, /, ?, ", <, >, | are not allowed.

Include system and hidden indices

✓ Your index pattern matches 1 source.

cis3200_superstore	Index

18. After creating the name, you can continue to click on “Time field” to choose one. This feature is used to filter the data based on time. The drop down fields will display all the time and data related fields from the index.



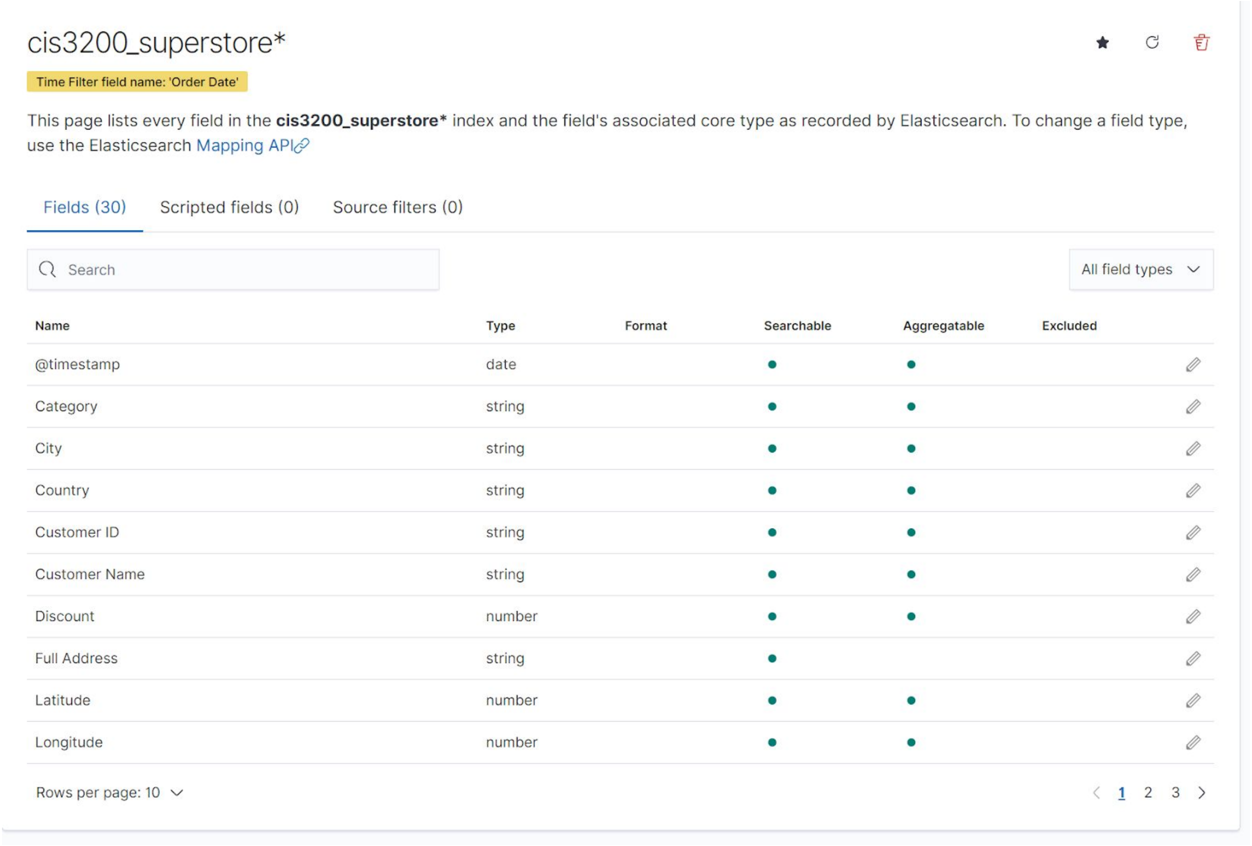
Step 2 of 2: Configure settings

cis3200_superstore*

Select a primary time field for use with the global time filter.

Time field: Order Date

Now, your new index pattern will look like the image below:



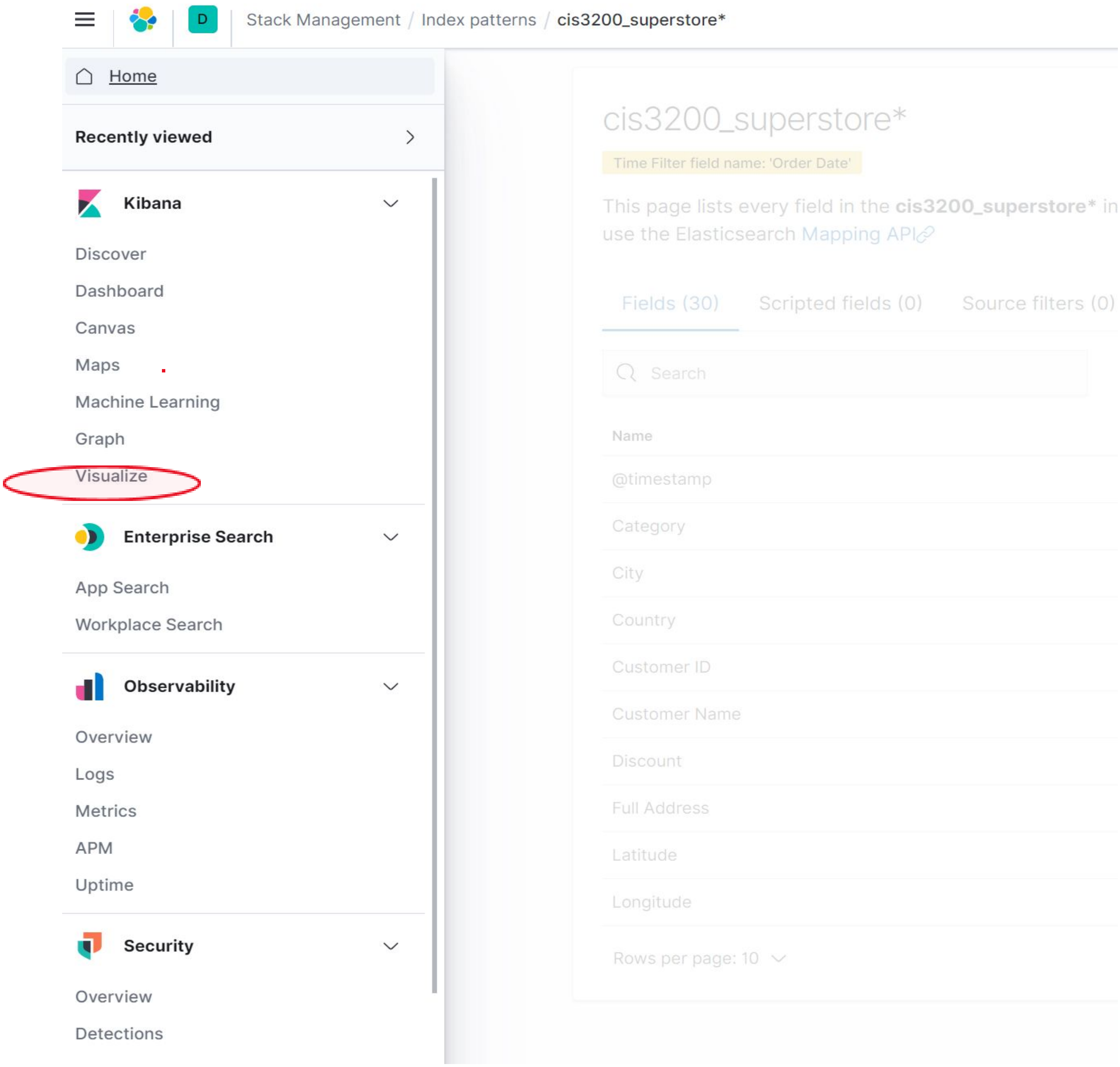
cis3200_superstore*

Time Filter field name: 'Order Date'

This page lists every field in the **cis3200_superstore*** index and the field's associated core type as recorded by Elasticsearch. To change a field type, use the Elasticsearch Mapping API

Name	Type	Format	Searchable	Aggregatable	Excluded
@timestamp	date		●	●	
Category	string		●	●	
City	string		●	●	
Country	string		●	●	
Customer ID	string		●	●	
Customer Name	string		●	●	
Discount	number		●	●	
Full Address	string		●		
Latitude	number		●	●	
Longitude	number		●	●	

19. Click on “Visualize” tab to start creating the Visualization for the data as follow:



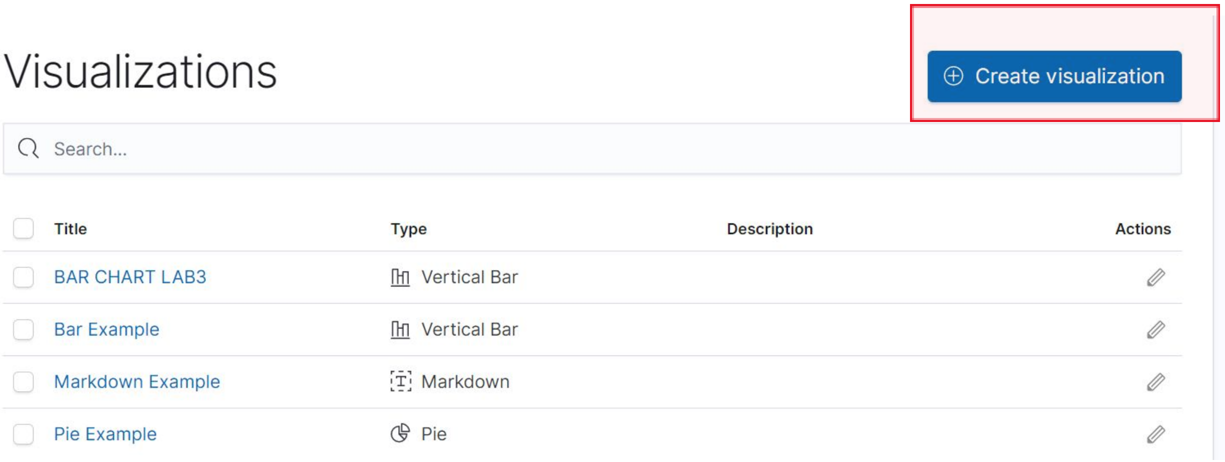
cis3200_superstore*

Time Filter field name: 'Order Date'

This page lists every field in the **cis3200_superstore*** in
use the Elasticsearch Mapping API

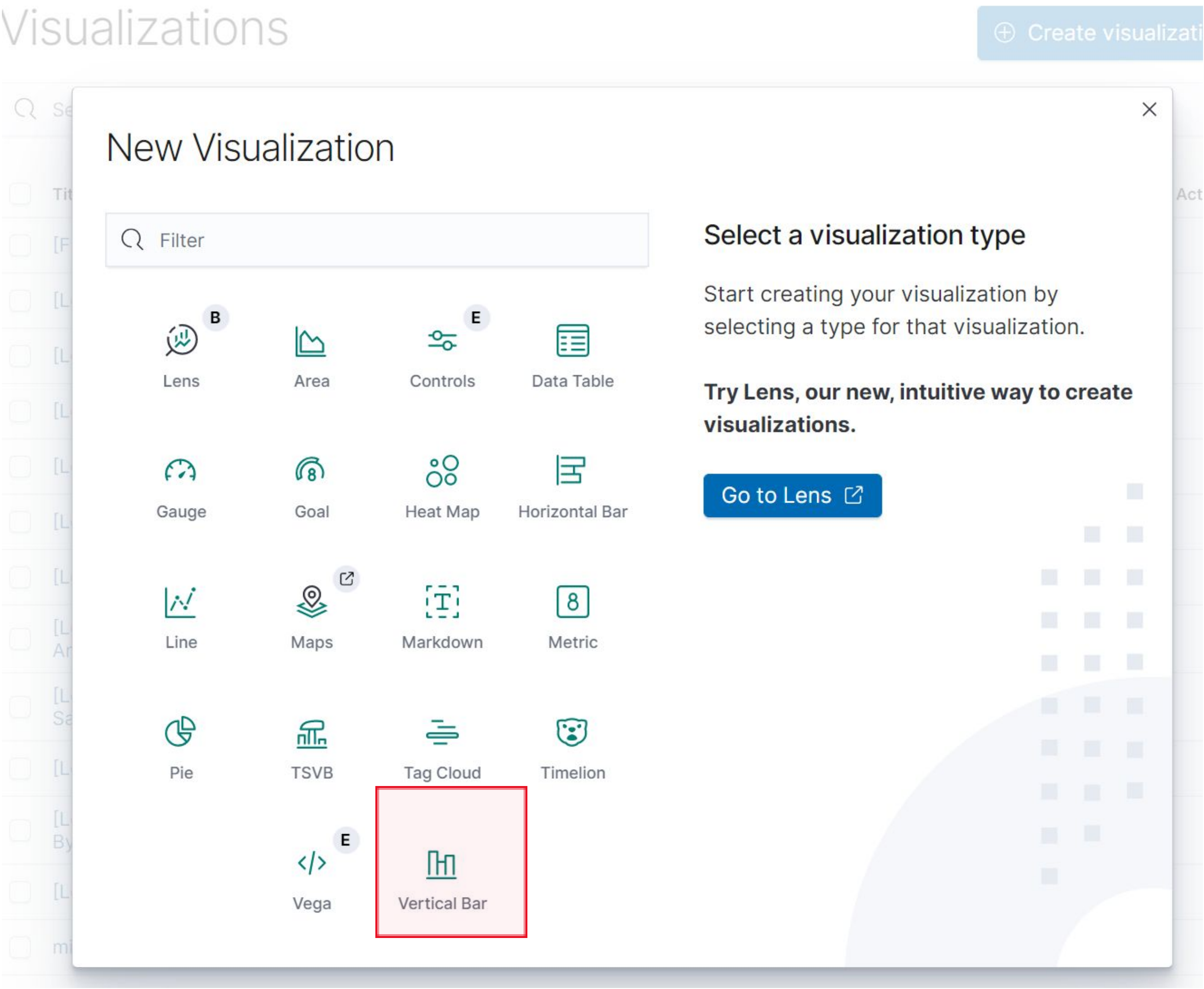
Name
@timestamp
Category
City
Country
Customer ID
Customer Name
Discount
Full Address
Latitude
Longitude

20. Continue to click on the tab “Create visualization” and choose “Vertical Bar” as follow:



Visualizations

□ Title	Type	Description	Actions
□ BAR CHART LAB3	Vertical Bar		✎
□ Bar Example	Vertical Bar		✎
□ Markdown Example	Markdown		✎
□ Pie Example	Pie		✎

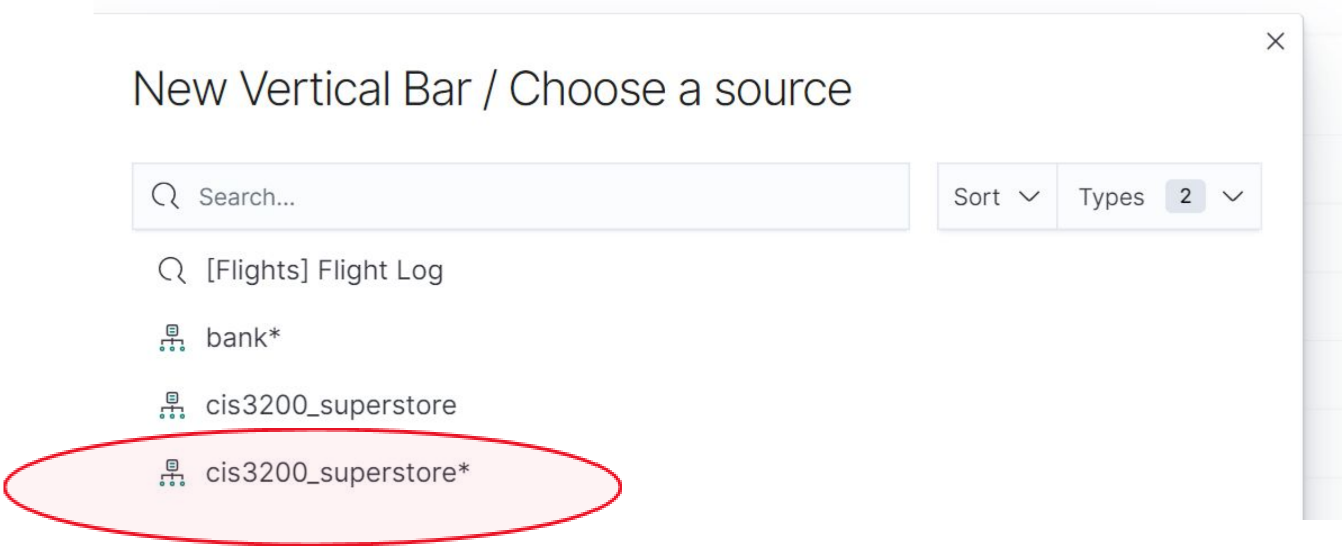


Select a visualization type

Start creating your visualization by selecting a type for that visualization.

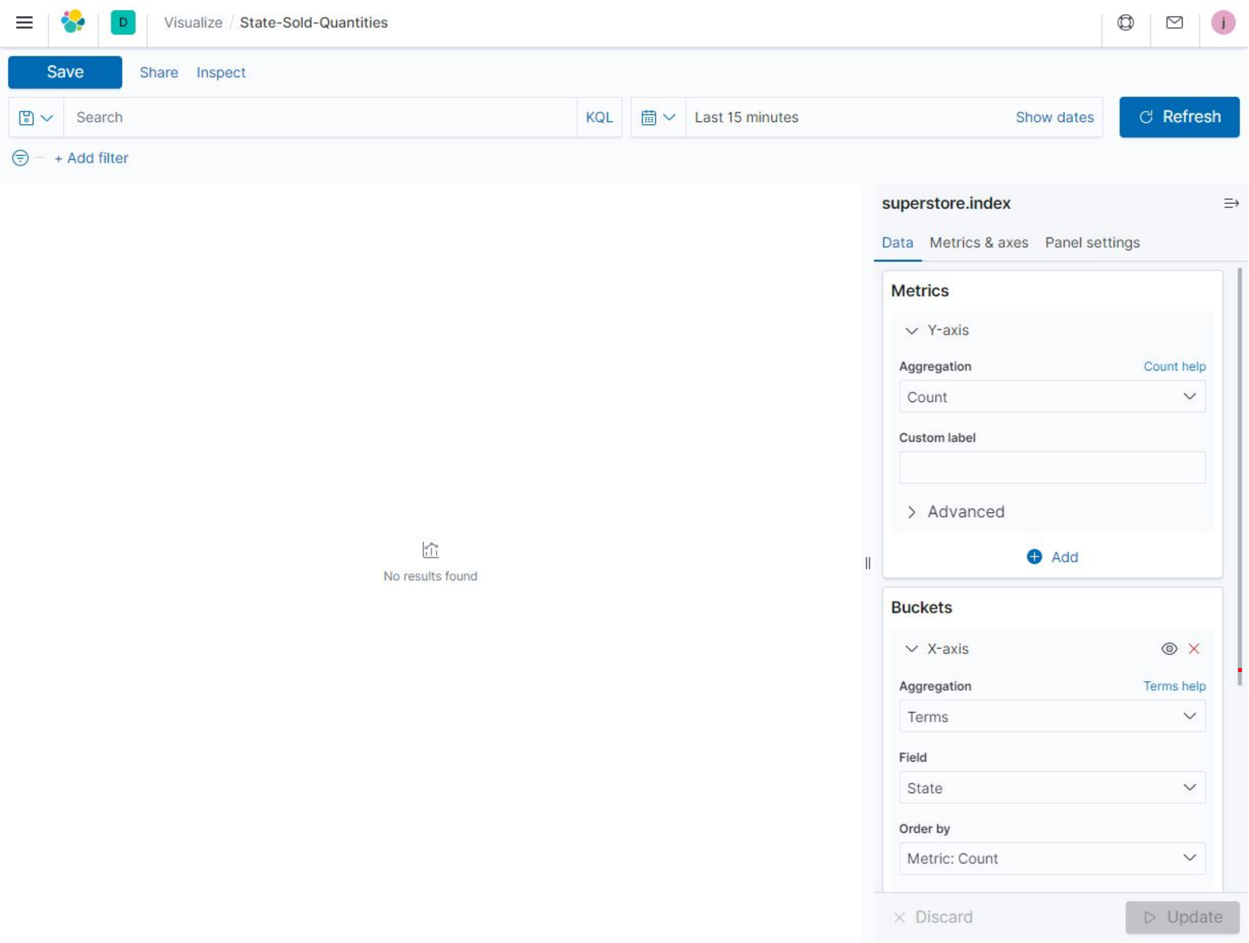
Try Lens, our new, intuitive way to create visualizations.

21. At the “New Vertical Bar,” click on the index pattern name which you just created in the previous step, “cis3200_superstore*,” which is the one we should click to start creating the bar chart:



New Vertical Bar / Choose a source

22.Next, you will need to set the time range, please select from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named “Refresh”.



Metrics

Y-axis

Aggregation: Count (Count help)

Custom label:

Advanced

Buckets

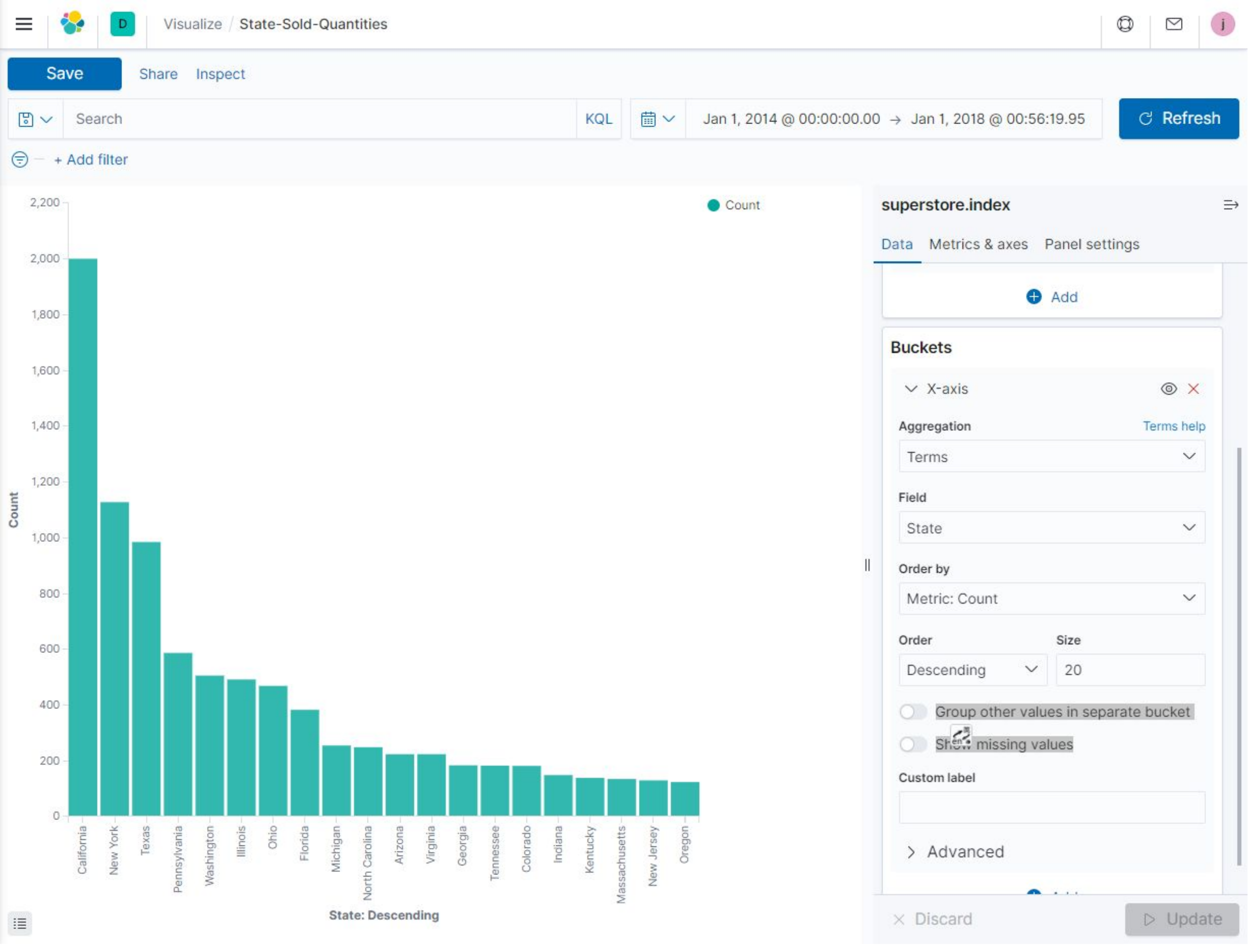
X-axis

Aggregation: Terms (Terms help)

Field: State

Order by: Metric: Count

23. After that, follow the graph's right menu setting, make sure the "Group other values in separate bucket" and "Show missing values" both default and they should be off. You should see a graph of each state's sales count. **Remember to save it!**



Aggregation: Terms

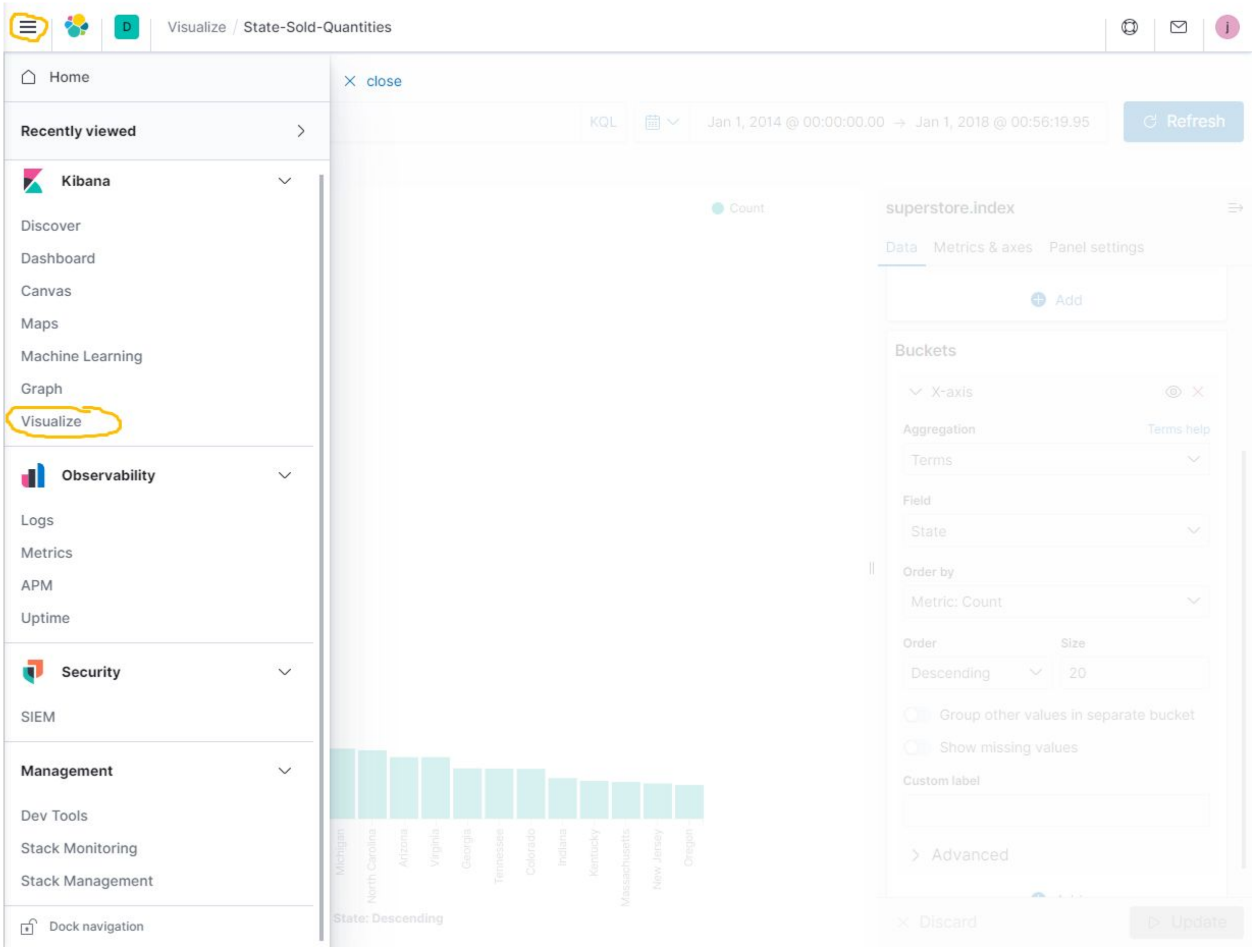
Field: State

Order by: Metric: Count

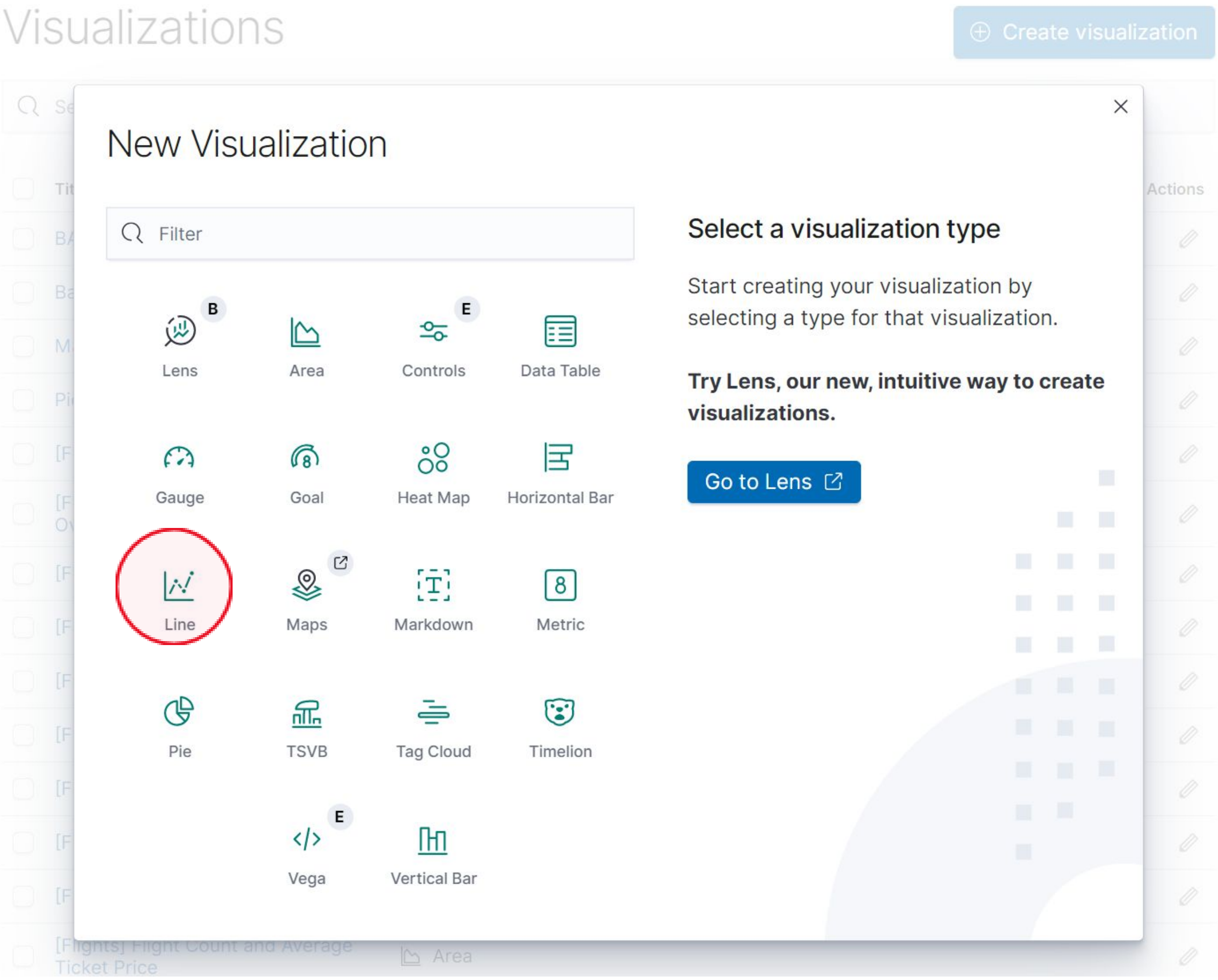
Order: Descending

Size: 20

24. Click the left menu and click to visualize. We will do a line chart for this step.(see the yellow circle)



25.Then select Line in the red circle.



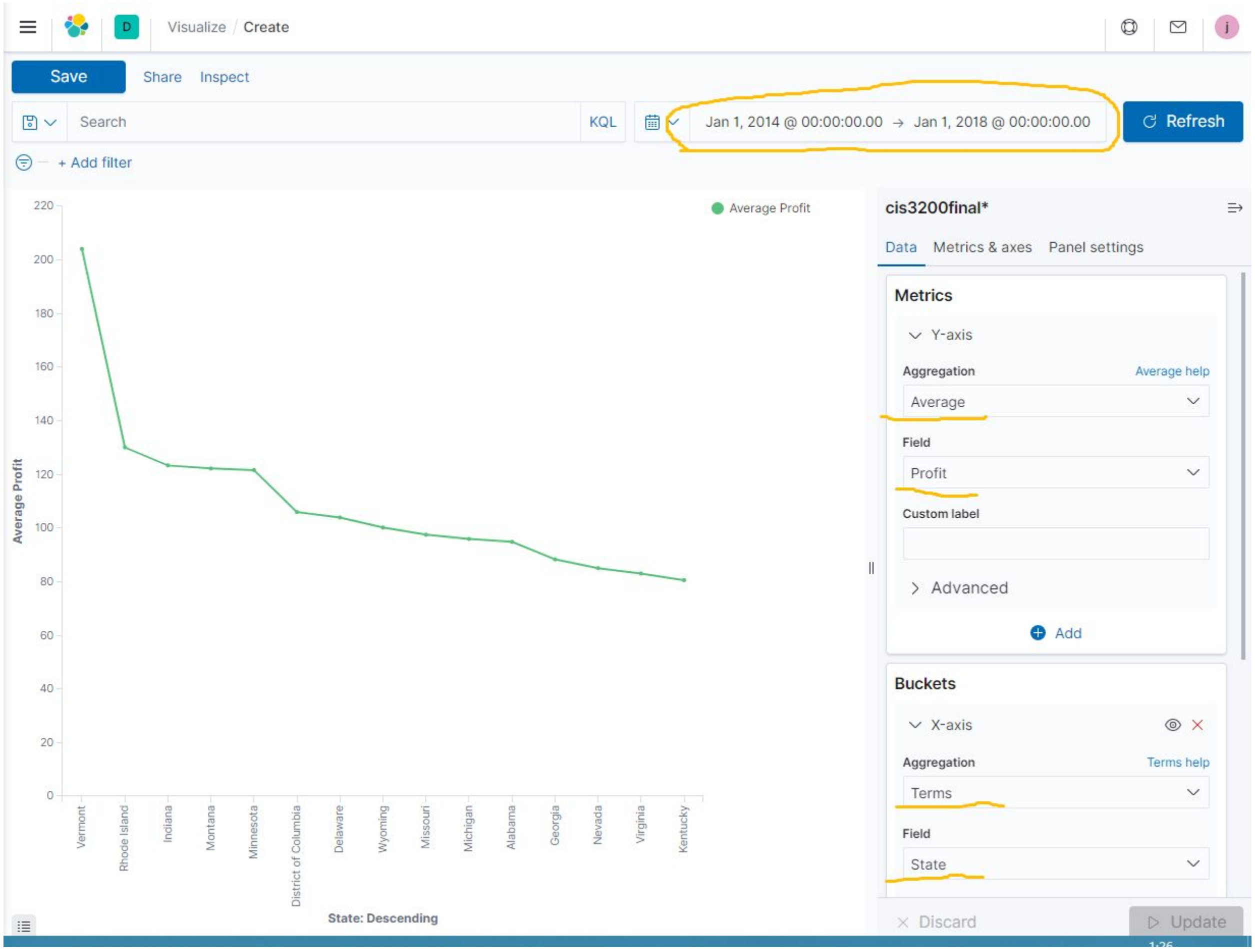
Select a visualization type

Start creating your visualization by selecting a type for that visualization.

Try Lens, our new, intuitive way to create visualizations.

26.Select the time range from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named “Refresh”. Aggregation choose Average, Field “profit”. Add a X-axis , For Aggregation select “Terms”, for the Field select “State’ and **Save it.**

you should have a graph similar to the one at the next.



Jan 1, 2014 @ 00:00:00.00 → Jan 1, 2018 @ 00:00:00.00

Metrics – Y-axis

Aggregation: Average

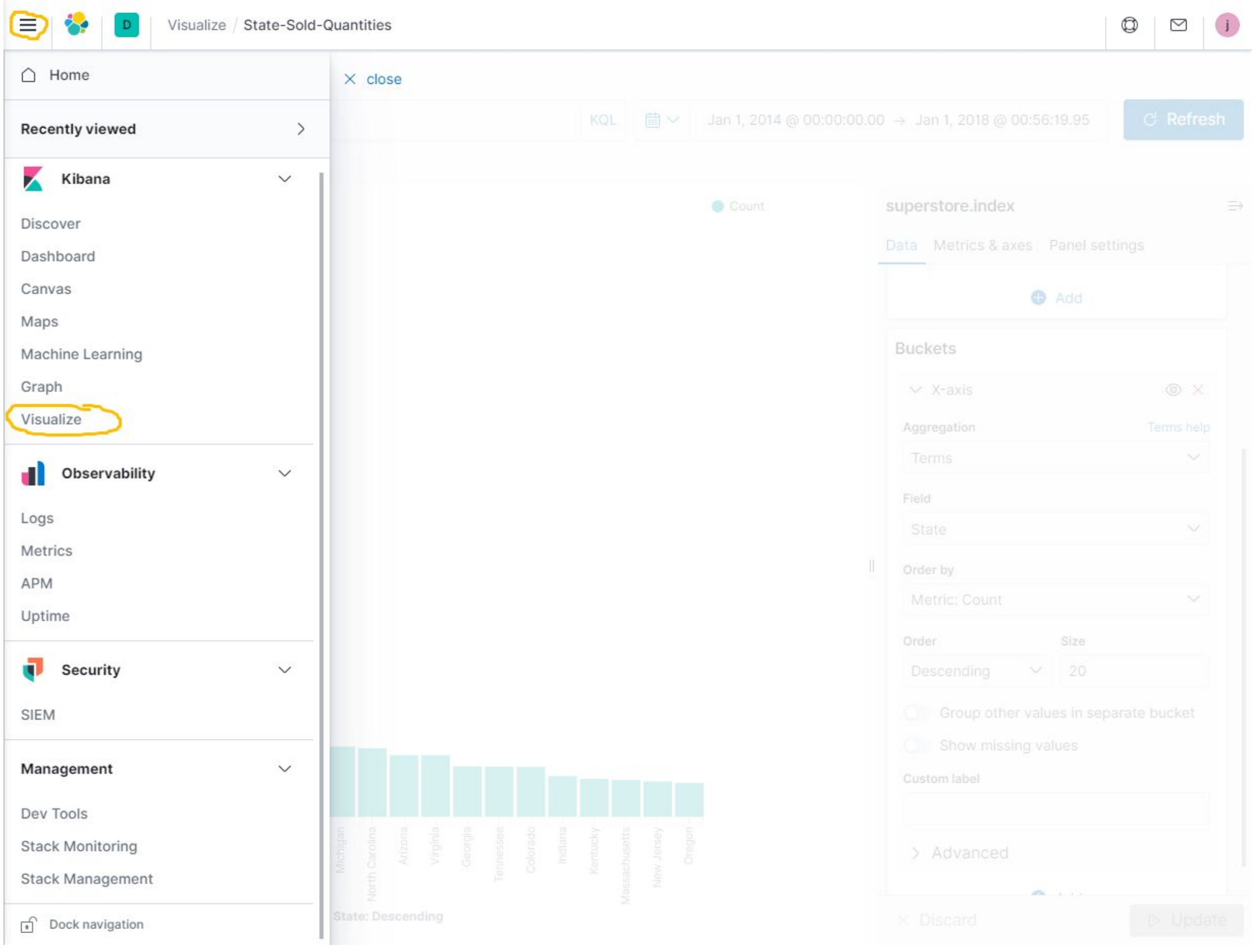
Field: Profit

Buckets – X-axis

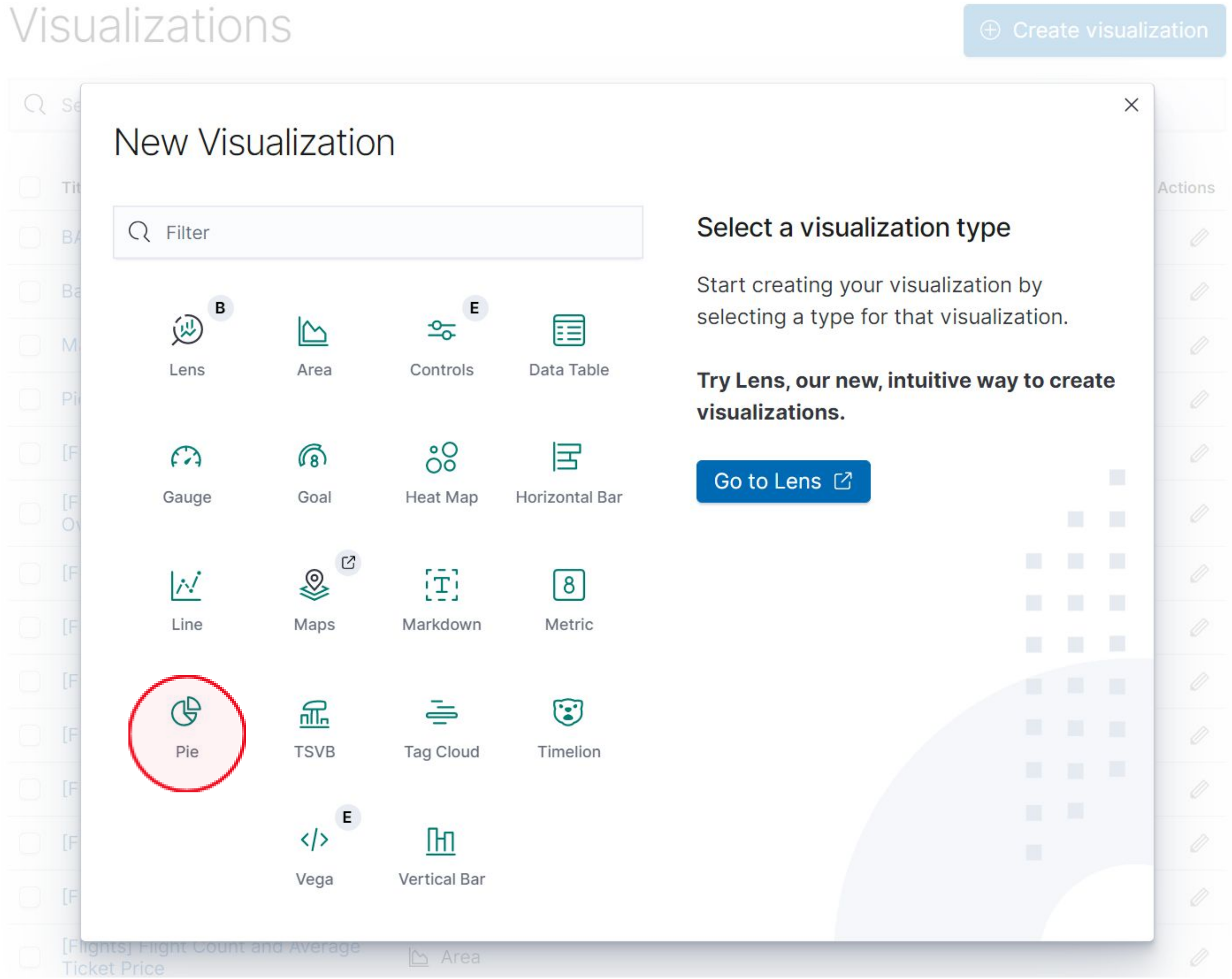
Aggregation: Terms

Field: State

26.Click the left menu and click to visualize. We will do a pie chart for this step.(see the yellow circle)



27.Then select the Pie In the red circle.



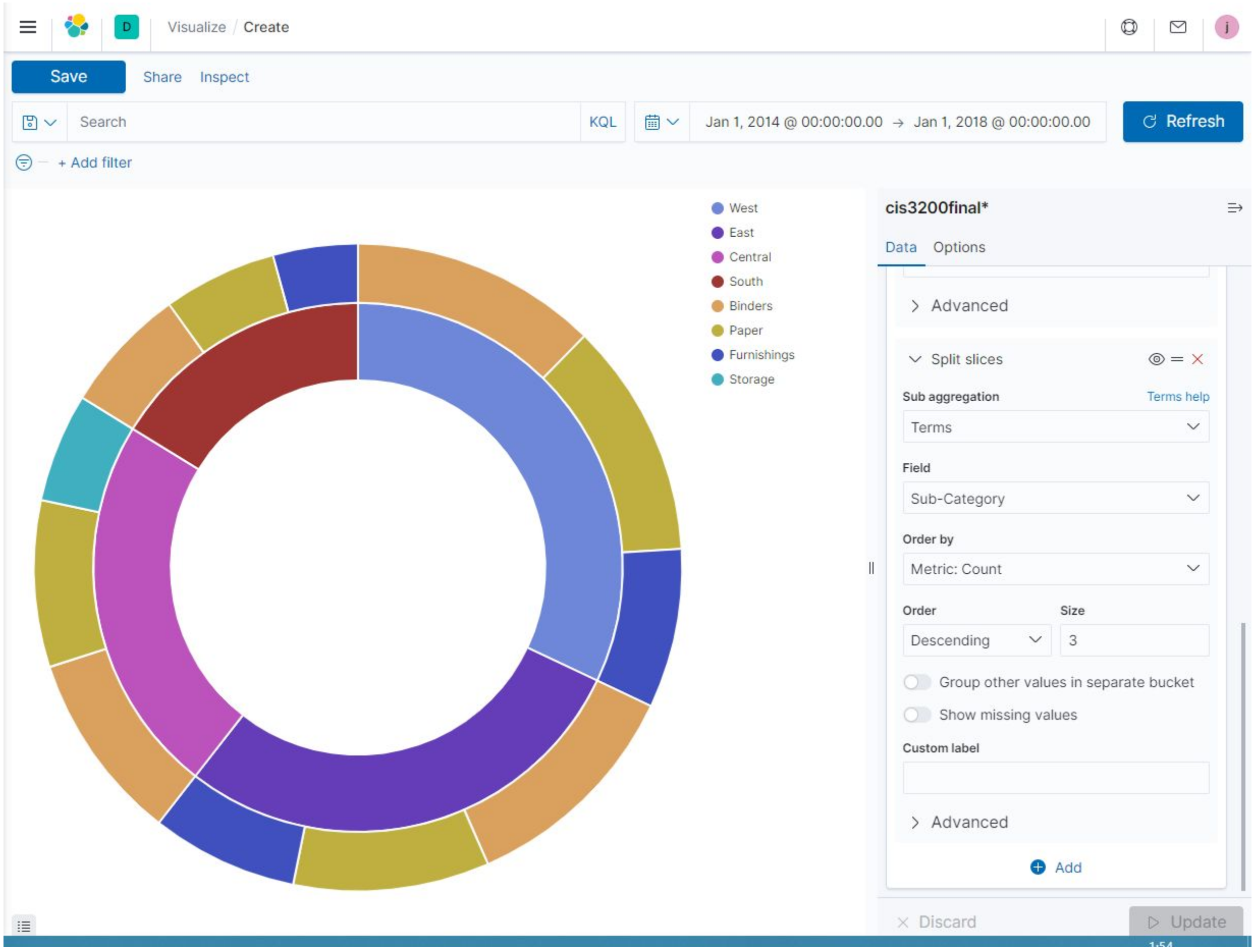
Select a visualization type

Start creating your visualization by selecting a type for that visualization.

Try Lens, our new, intuitive way to create visualizations.

28.Select the time range from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named "Refresh".

For Aggregation choose Count, At Buckets add a Split Slices,Aggregation choose Term Filed choose Region , order by Metric:Count . Order Descending size 5. others keep default. You should have the next graph shown on your screen, **Save it.**



Jan 1, 2014 @ 00:00:00.00 → Jan 1, 2018 @ 00:00:00.00

cis3200final*

Split slices

Sub aggregation: Terms

Field: Sub-Category

Order by: Metric: Count

Order: Descending

Size: 3

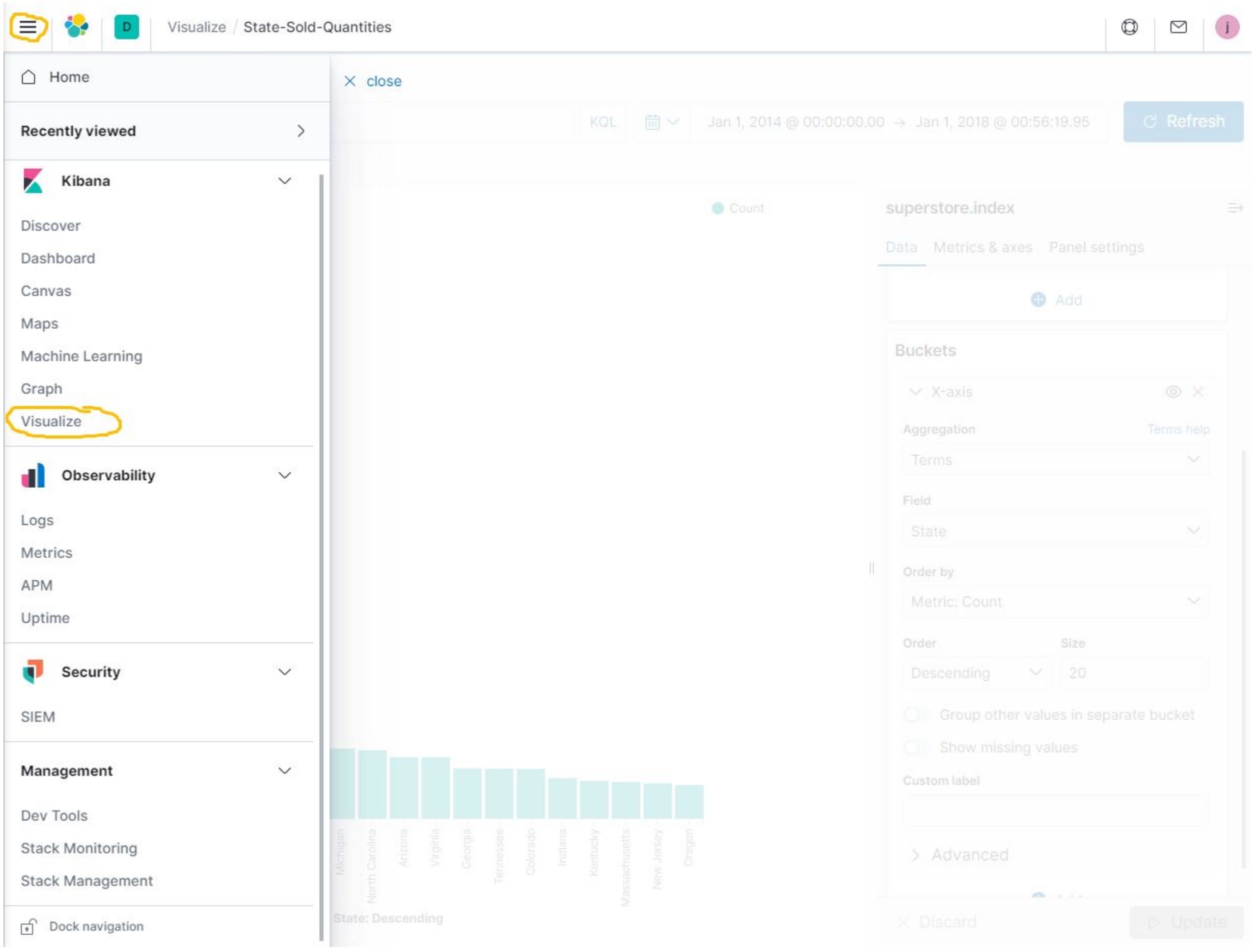
Group other values in separate bucket

Show missing values

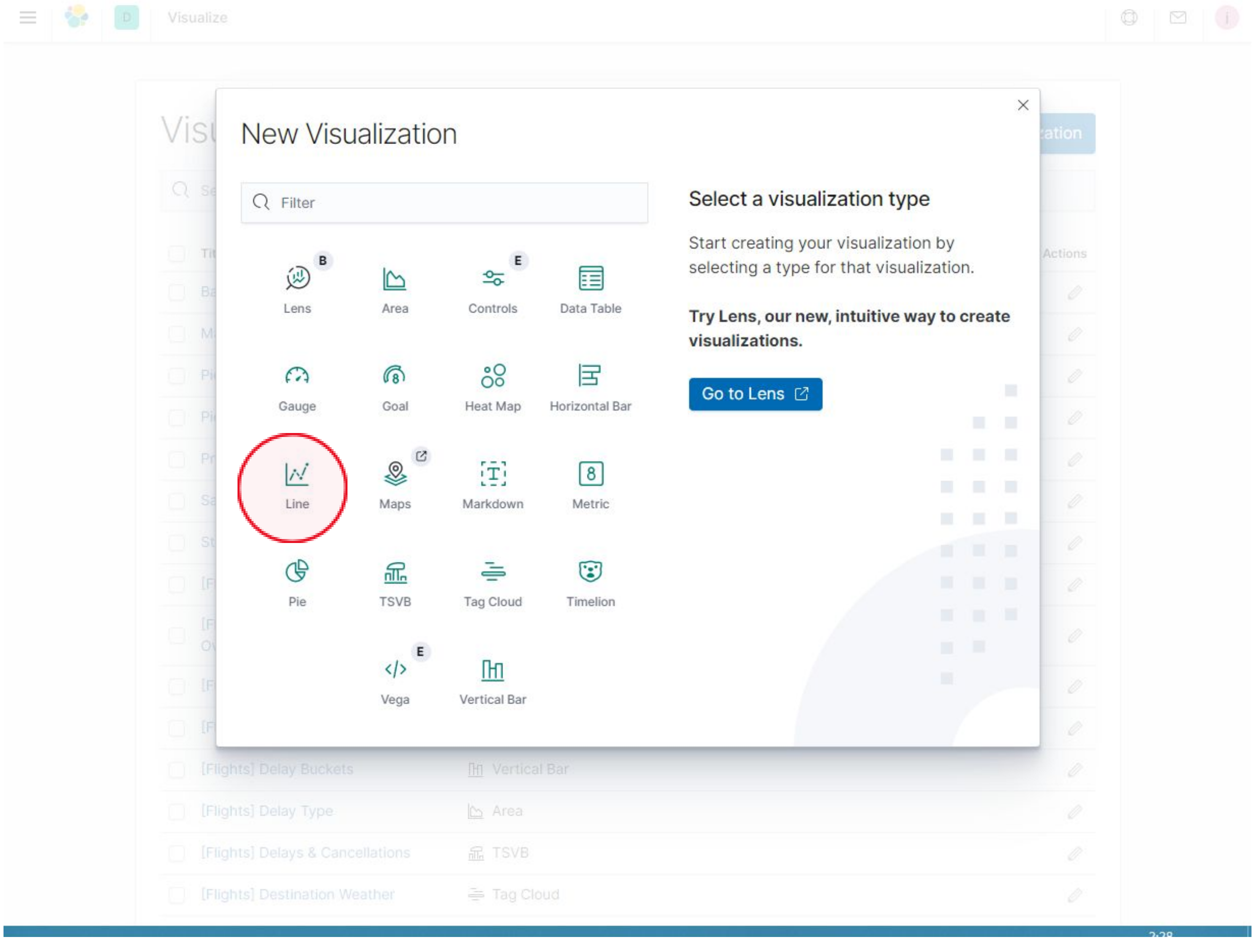
Custom label

29.Next we will do a quarter sales report at a 4 years’s range

Click the left menu and click to visualize.



30.Select Line graph in red circle.

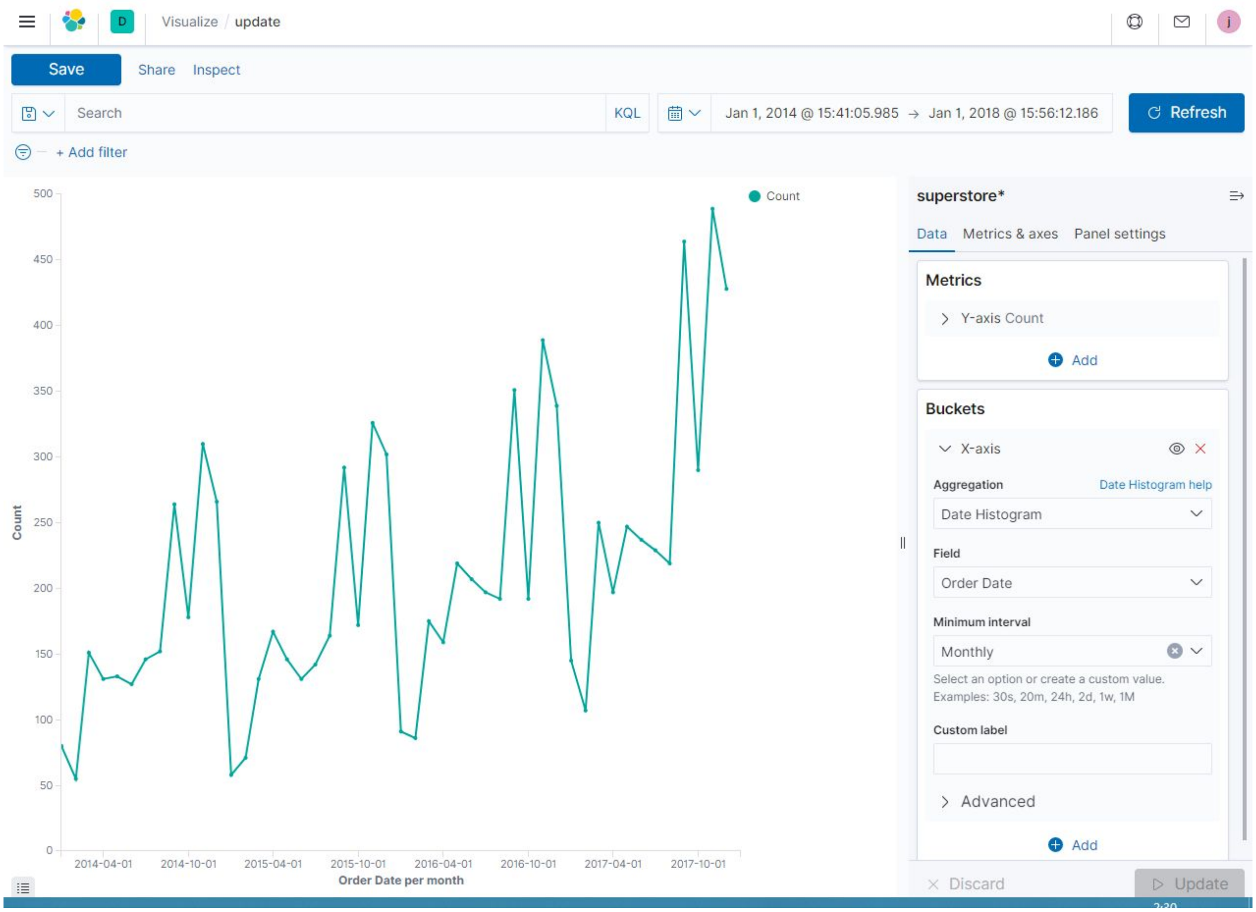


Select a visualization type

Start creating your visualization by selecting a type for that visualization.

Try Lens, our new, intuitive way to create visualizations.

31.Select the time range from Jan.1st 2014, 00:00:00, to Jan 1st 2018, 00:00:00. and then click the blue button named “Refresh”. Follow the left menu, you will be able to see this quarter report from the time range Jan 1. 2014 to Jan 1. 2018. Save it.



Aggregation: Date Histogram

Field: Order Date

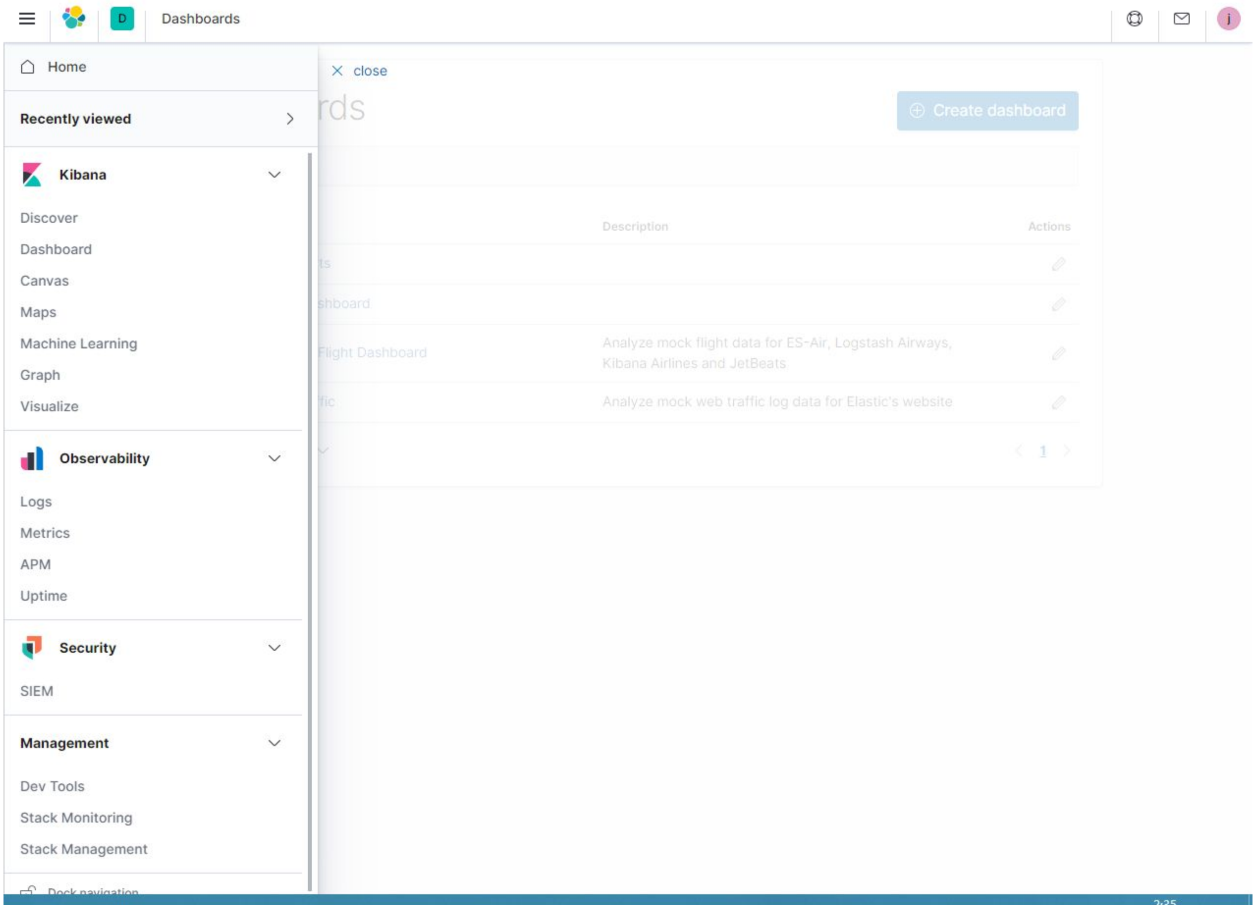
Minimum interval: Monthly

Select an option or create a custom value. Examples: 30s, 20m, 24h, 2d, 1w, 1M

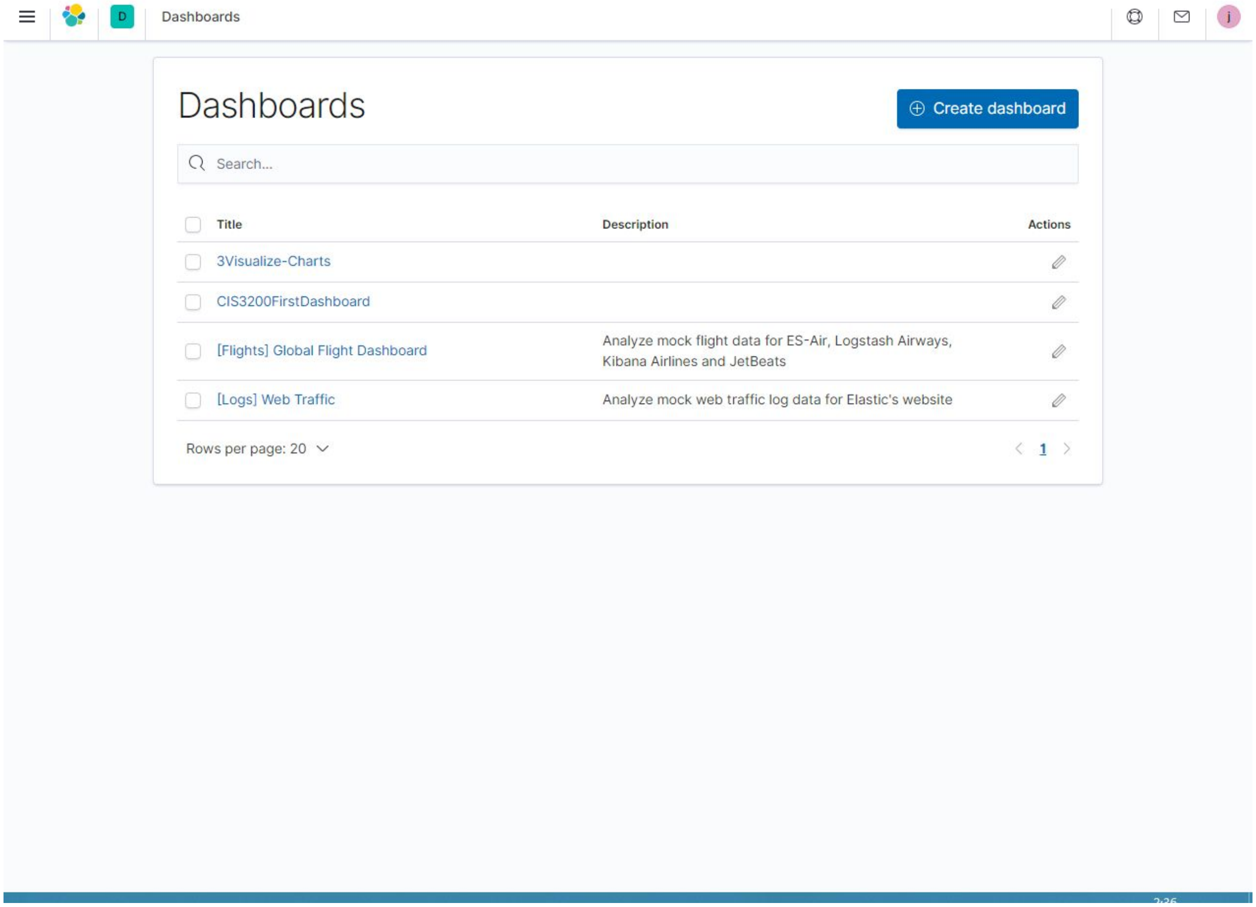
Custom label

Advanced

32.Click the left menu and click the dashboard.

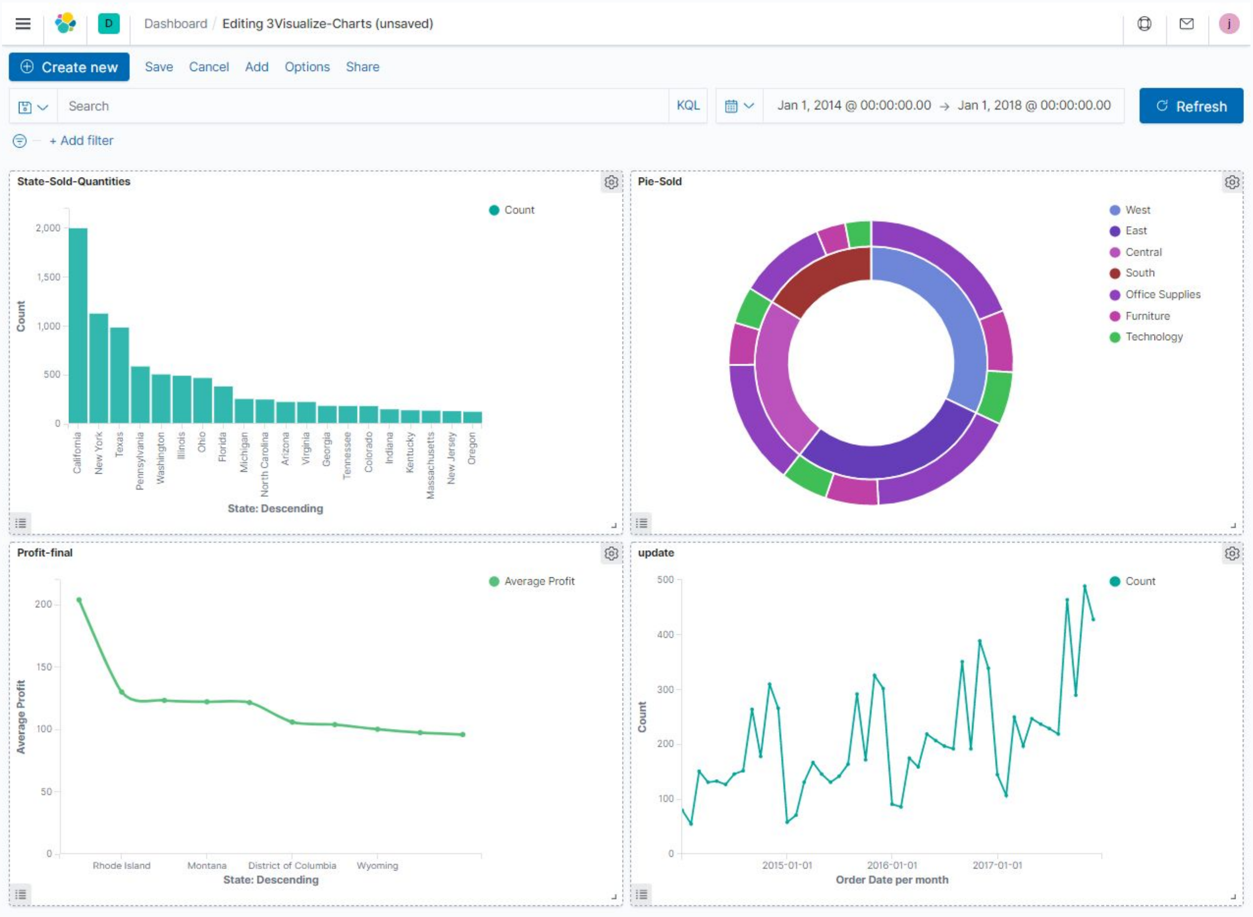


33.Click the Create dashboard in the blue button.



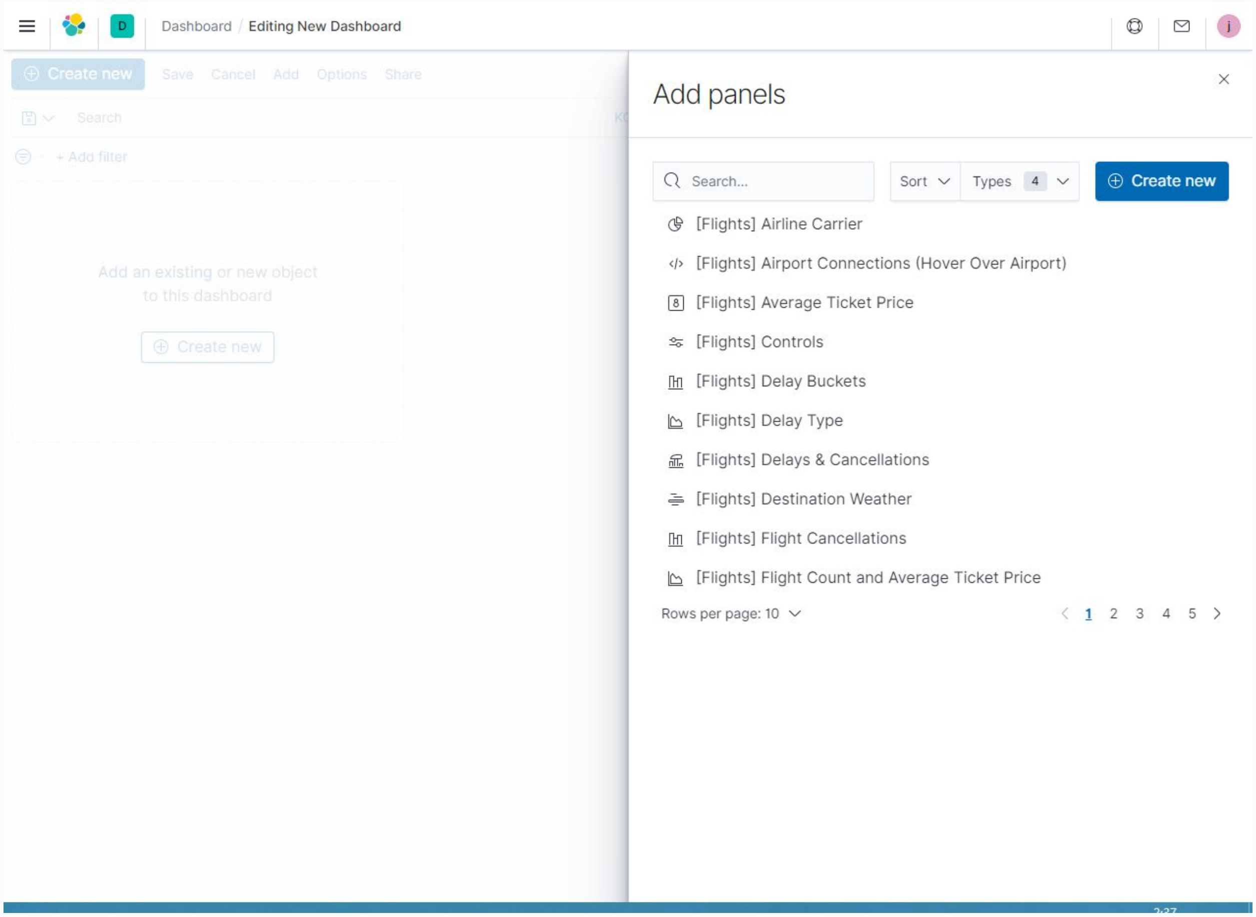
□	Title	Description	Actions
□	3Visualize-Charts		✎
□	CIS3200FirstDashboard		✎
□	[Flights] Global Flight Dashboard	Analyze mock flight data for ES-Air, Logstash Airways, Kibana Airlines and JetBeats	✎
□	[Logs] Web Traffic	Analyze mock web traffic log data for Elastic's website	✎

34.Then, Click the Add an existing option. You should be able to add all 4 we have done before. It would look like the one next page.



Dashboard / Editing 3Visualize-Charts (unsaved)

Jan 1, 2014 @ 00:00:00.00 → Jan 1, 2018 @ 00:00:00.00



Dashboard / Editing New Dashboard

Create new Save Cancel Add Options Share

Add an existing or new object to this dashboard

Create new

Add panels

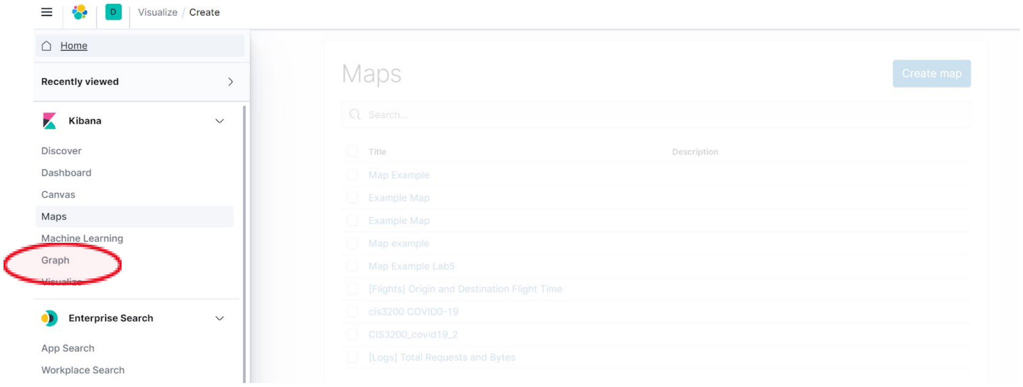
Search... Sort Types 4 Create new

- [Flights] Airline Carrier
- [Flights] Airport Connections (Hover Over Airport)
- [Flights] Average Ticket Price
- [Flights] Controls
- [Flights] Delay Buckets
- [Flights] Delay Type
- [Flights] Delays & Cancellations
- [Flights] Destination Weather
- [Flights] Flight Cancellations
- [Flights] Flight Count and Average Ticket Price

Rows per page: 10 1 2 3 4 5

Create GEO Map Visualization

1. Access Kibana and select *Maps* on the left side scroll bar or go to *Visualize* to create a map visual.



Visualize / Create

Home

Recently viewed

Kibana

Discover
Dashboard
Canvas
Maps
Machine Learning
Graph
Visualize

Enterprise Search

App Search
Workplace Search

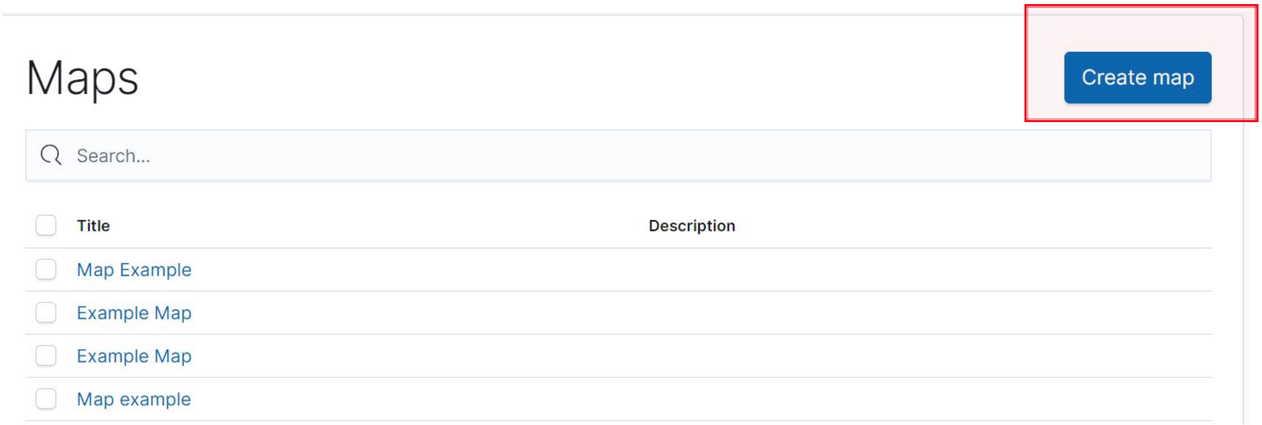
Maps

Create map

Search...

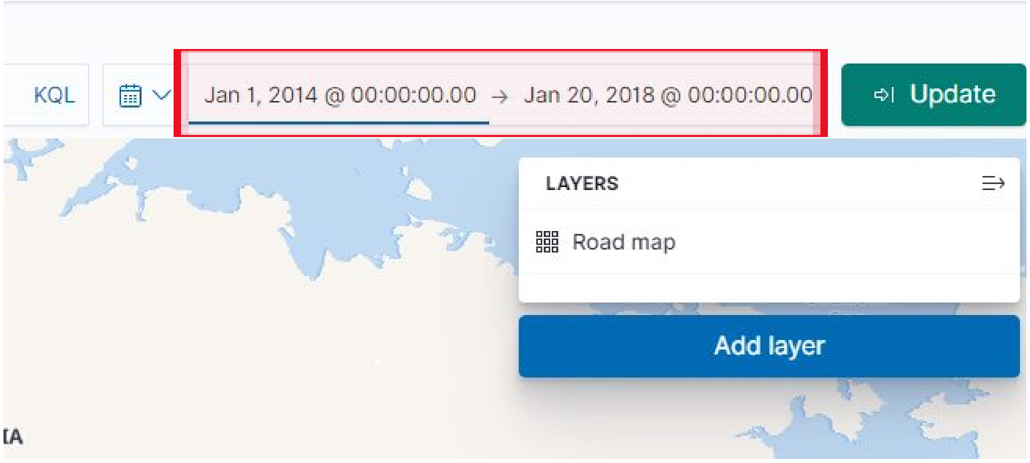
Title	Description
Map Example	
Example Map	
Example Map	
Map example	
Map Example Lab5	
[Flights] Origin and Destination Flight Time	
cis3200 COVID0-19	
CIS3200_covid19_2	
[Logs] Total Requests and Bytes	

2.If creating a brand new map, select Create Map to start a brand new visualization



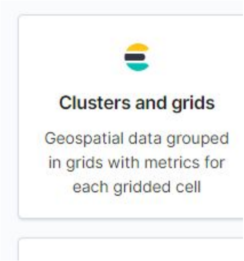
Title	Description
Map Example	
Example Map	
Example Map	
Map example	

3. Set the time frame from 1/1/14 - 1/1/2018 to ensure all of the data within the dataset is accounted for. You want to make sure this is done by selection *Absolute* and manually entering the correct date range.



Jan 1, 2014 @ 00:00:00.00 → Jan 20, 2018 @ 00:00:00.00

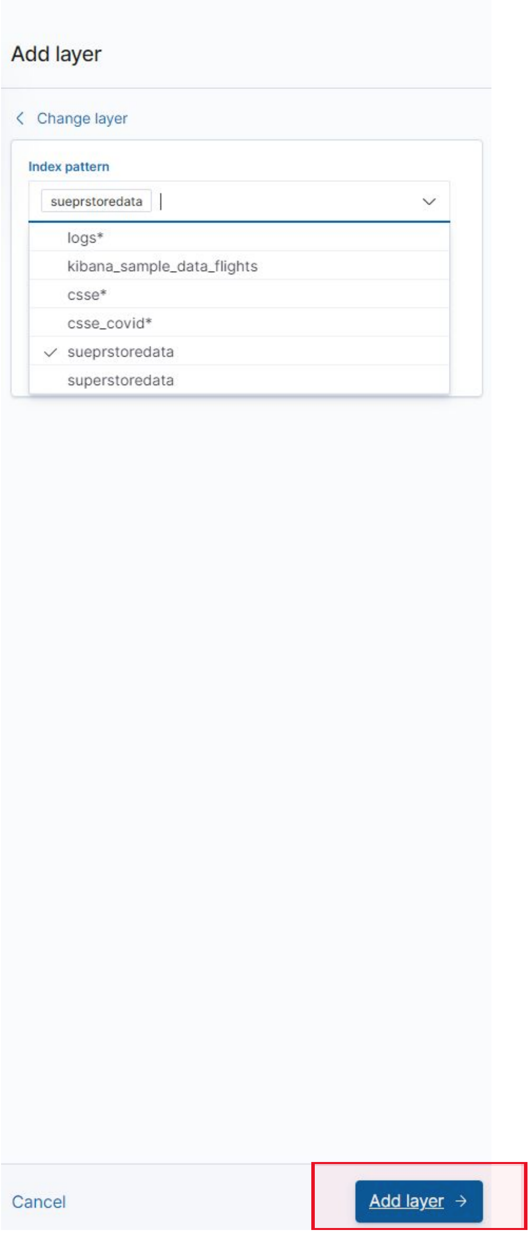
4. Select the Clusters and Grids map option from the menu on the right



Clusters and grids

Geospatial data grouped in grids with metrics for each gridded cell

5. Select *Add Layer* and select the index you created in the previous steps. This particular index is named *superstoredata.* Once selected, make sure to click on *Add Layer*



Add layer

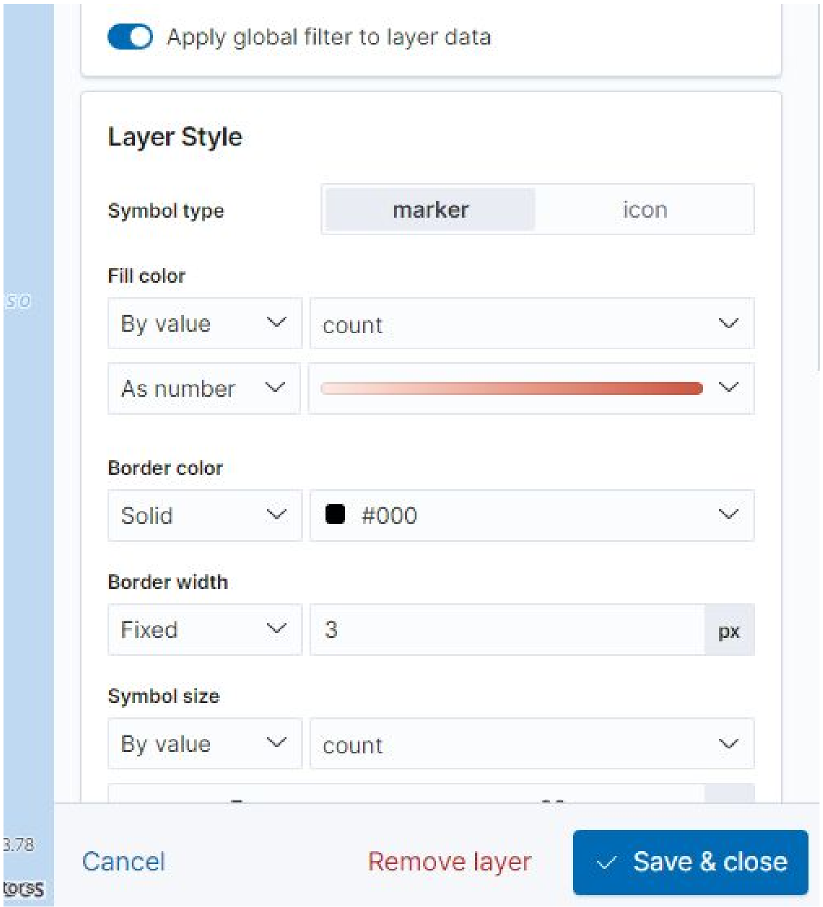
Change layer

Index pattern

sueprstoredata

- logs*
- kibana_sample_data_flights
- csse*
- csse_covid*
- ✓ sueprstoredata
- superstoredata

6. In this step we can add some changes to the *Layer Style,* some examples include changing the fill color gradient, border color, and border thickness.



Apply global filter to layer data

Layer Style

Symbol type: marker | icon

Fill color: By value — count

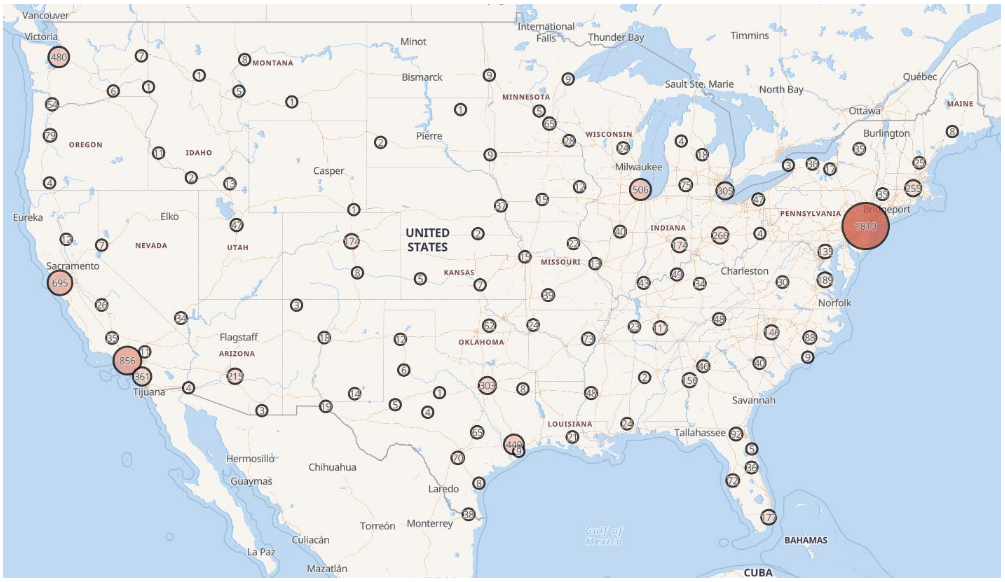
As number

Border color: Solid — ■ #000

Border width: Fixed — 3 px

Symbol size: By value — count

7. Make sure you click *Save and Close* before exiting the layer. Your new geo spatial map is ready for viewing as seen below



References

1. URL of Data Source

https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls

2. URL of your Github

xEqualsToY/Kibana-Data-Set: Data Visuliztion (github.com)

3. URL of References

Elastic Cloud Tutorial: Getting Started with a sample dataset | Elastic Blog