CIS5900 Term Project Tutorial

Authors: Sai Kiran Gontyala; Yoshitha Gattu

Instructor: Jongwook Woo

Date: 08/08/2019

Lab Tutorial

Saikiran Gontyala (sgontya@calstatela.edu)

Yoshitha Gattu (ygattu@calstatela.edu)

08/08/2019

International Energy Statistics Data Analysis using ElasticSearch

---

Objectives

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

- Upload dataset of International Energy Statistics into ElasticSearch server.
- Create Index Pattern
- Discover the data in Kibana
- Visualize the data using Kibana

Platform Specifications

- Dataset size: 59.7 MB
- Account: ElasticSearch/Cloud
- Kibana version: 7.2
- Memory size: 180 GB
- CPU Speed: 2.195 GHz
- Storage: 682 GB

Introduction

The Energy Statistics Database contains comprehensive energy statistics on the production, trade, conversion and final consumption of primary and secondary; conventional and non-conventional; and new and renewable sources of energy. The International Energy Statistics dataset has covered all about the growth of energy as an international trend. This dataset was published by the United Nations Statistics Division on the UNData site. The United Nations Statistics Division is

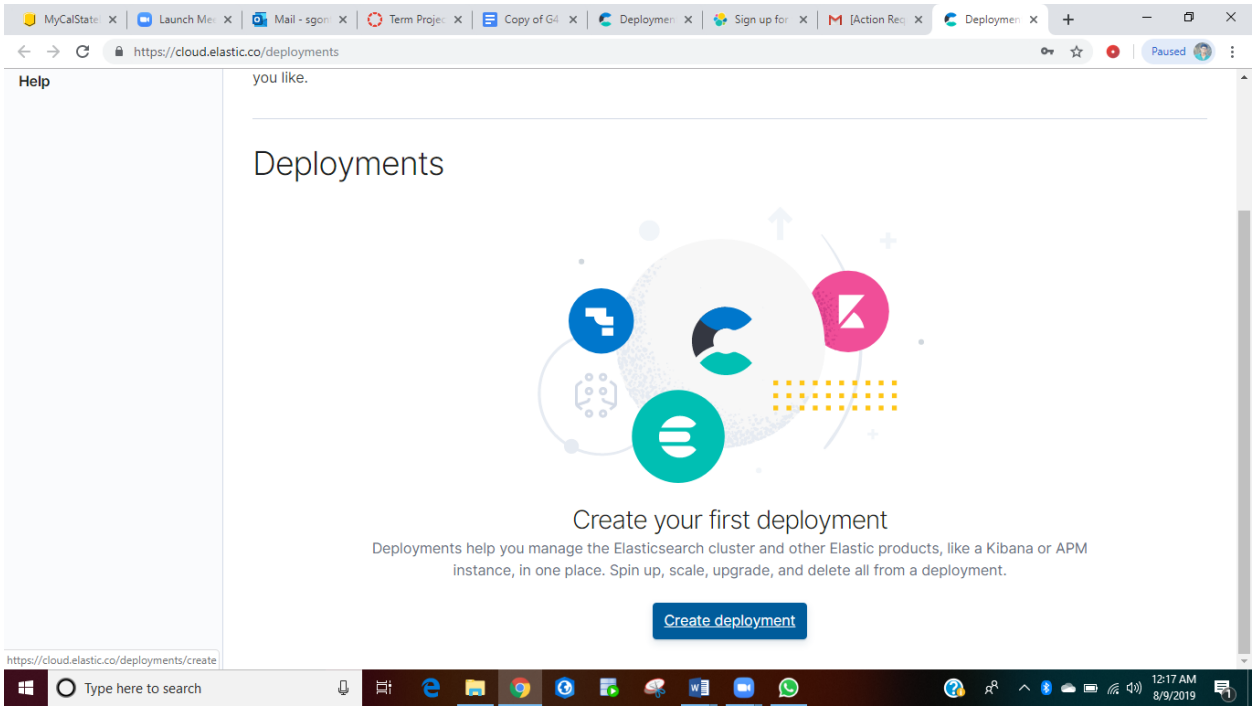
committed to the advancement of the global statistical system. This dataset contains data of energy trade and production from 1990-2014.

Step 1: Uploading data to ElasticSearch cloud

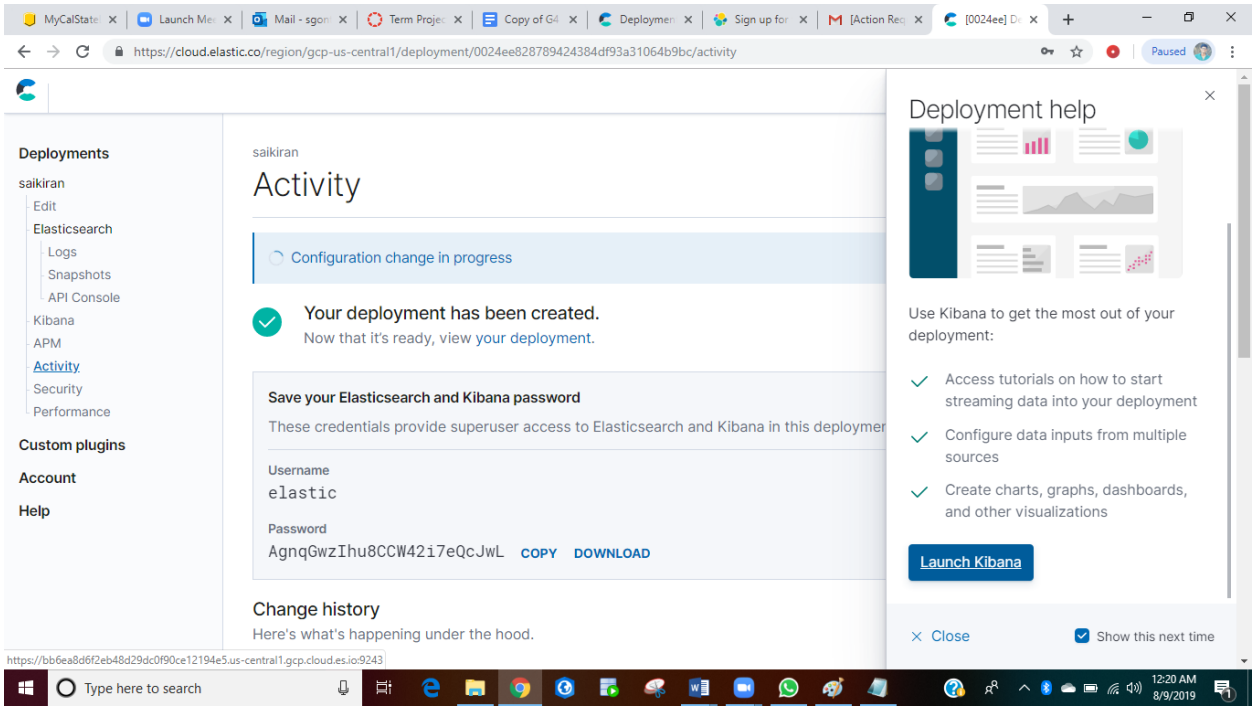
1. Download the dataset from the url:

https://www.kaggle.com/unitednations/international-energy-statistics#all_energy_statistics.csv

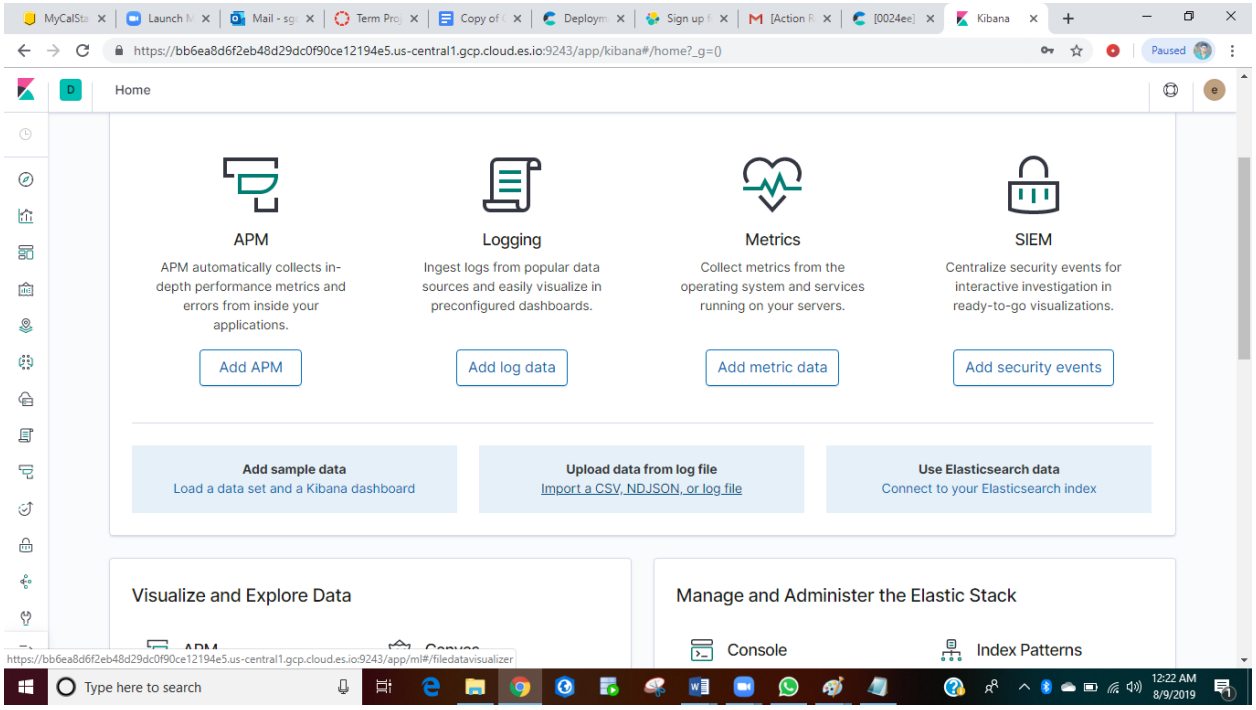
2. Create a deployment in Elasticsearch Cloud



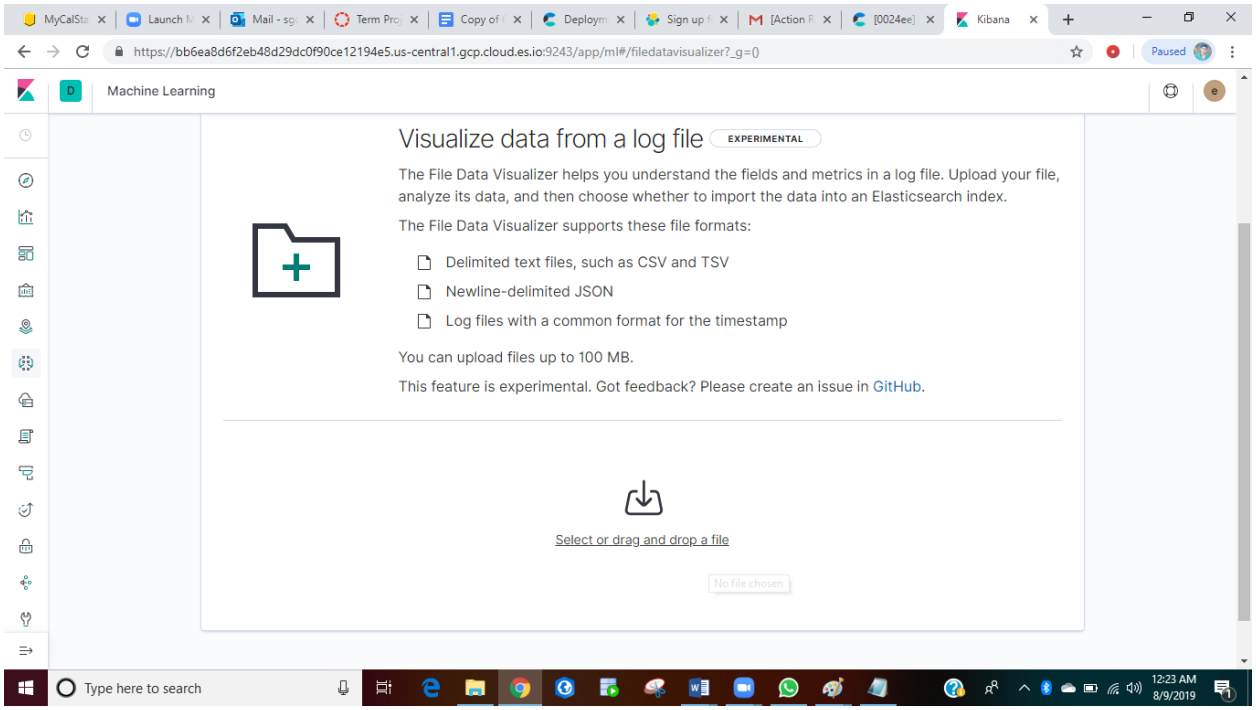
3. Launch Kibana



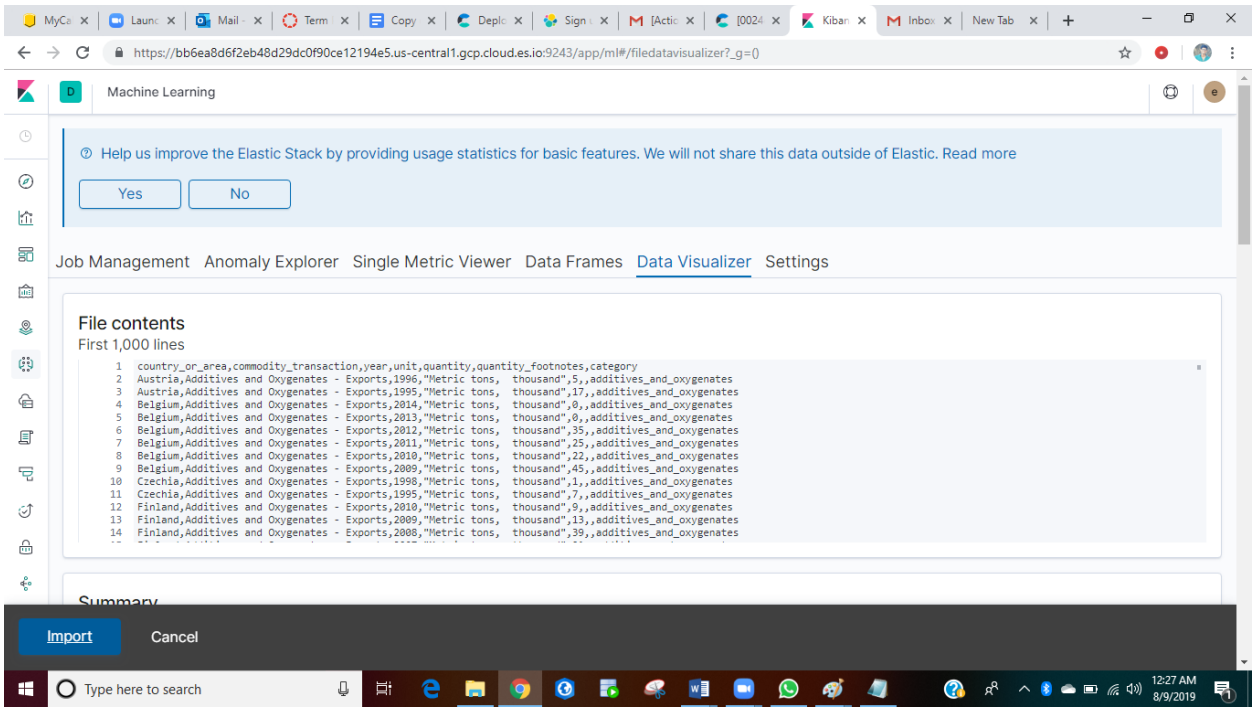
4. Import the data from the computer



5. Drag the data file to import it



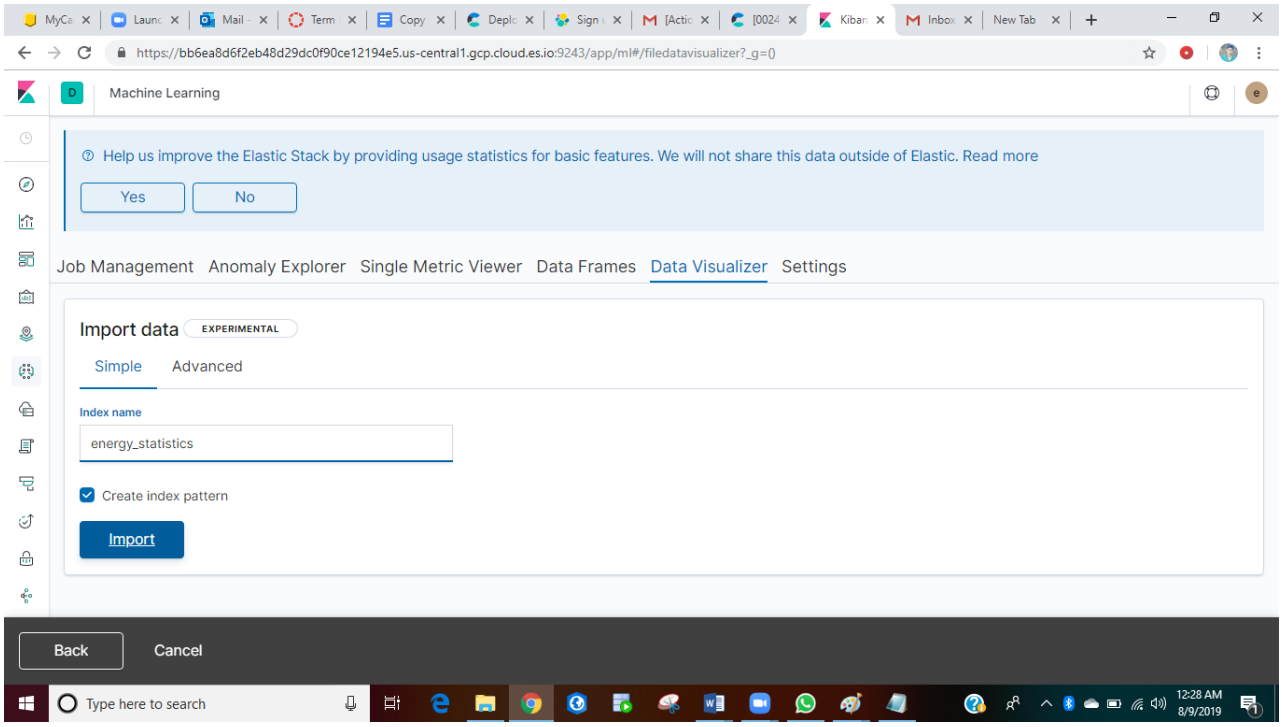
6. Click on "Import"



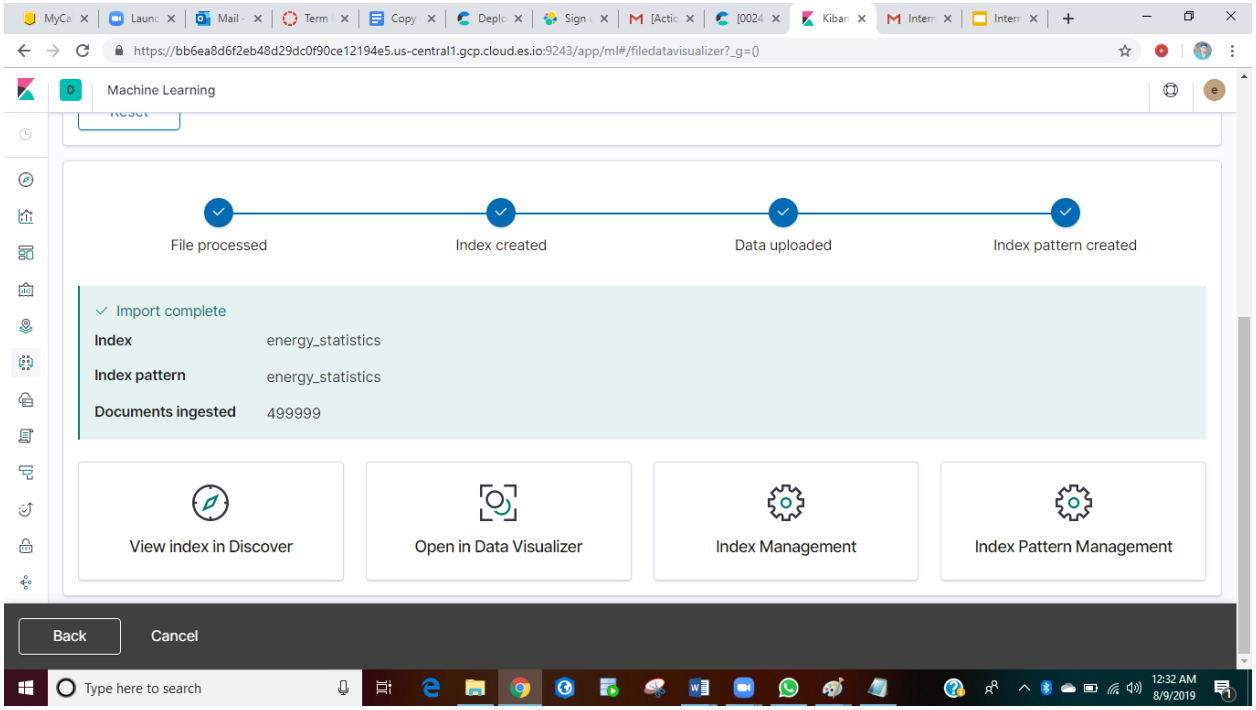
Step 2: Creating Index patterns

Index patterns help Kibana to know which Elasticsearch indices we want to explore. We created an index pattern for the Energy statistics data set..

1. Give the Index Pattern name as “energy_statistics” and then click “import”



2. Index patterns'energy_statistics' have been successfully created.

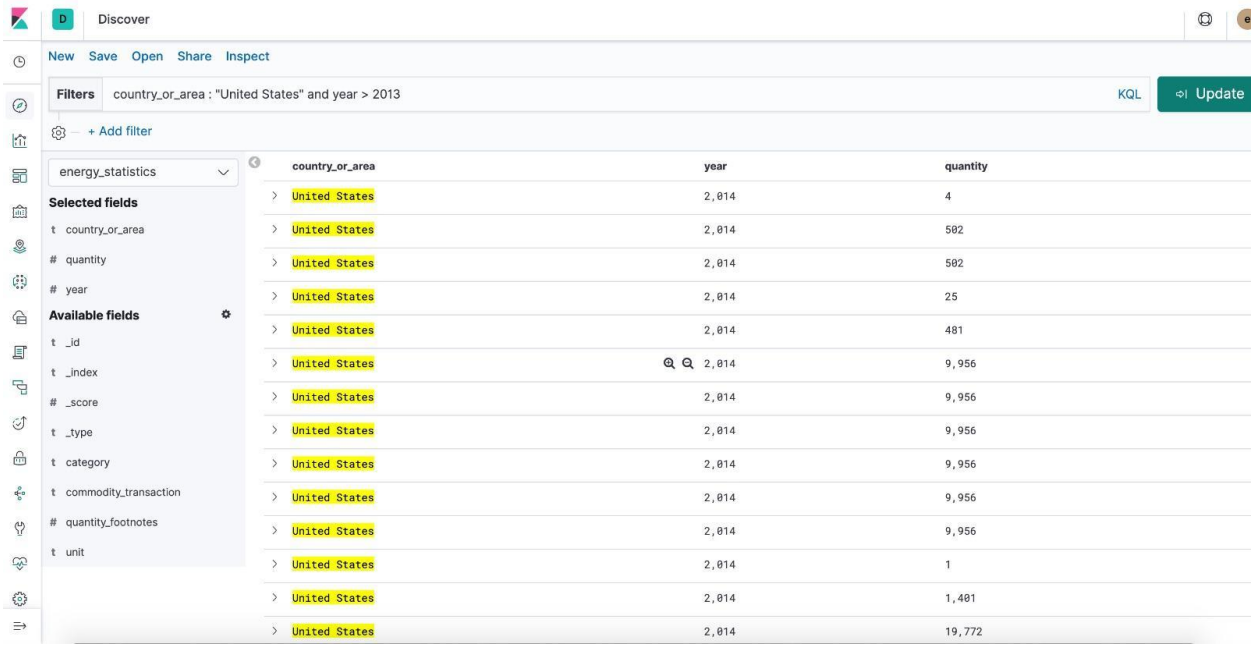


Step 3: Discovering data

Using the Discover application, we entered an Elasticsearch query to search our data and filter the results.

1. Open **Discover**. The current index pattern appears below the filter bar, in this case **erergy_statistics***. You might need to click **New** in the menu bar to refresh the data.
2. In the search field, enter the following string and select **Update/Refresh** button.
   For example:

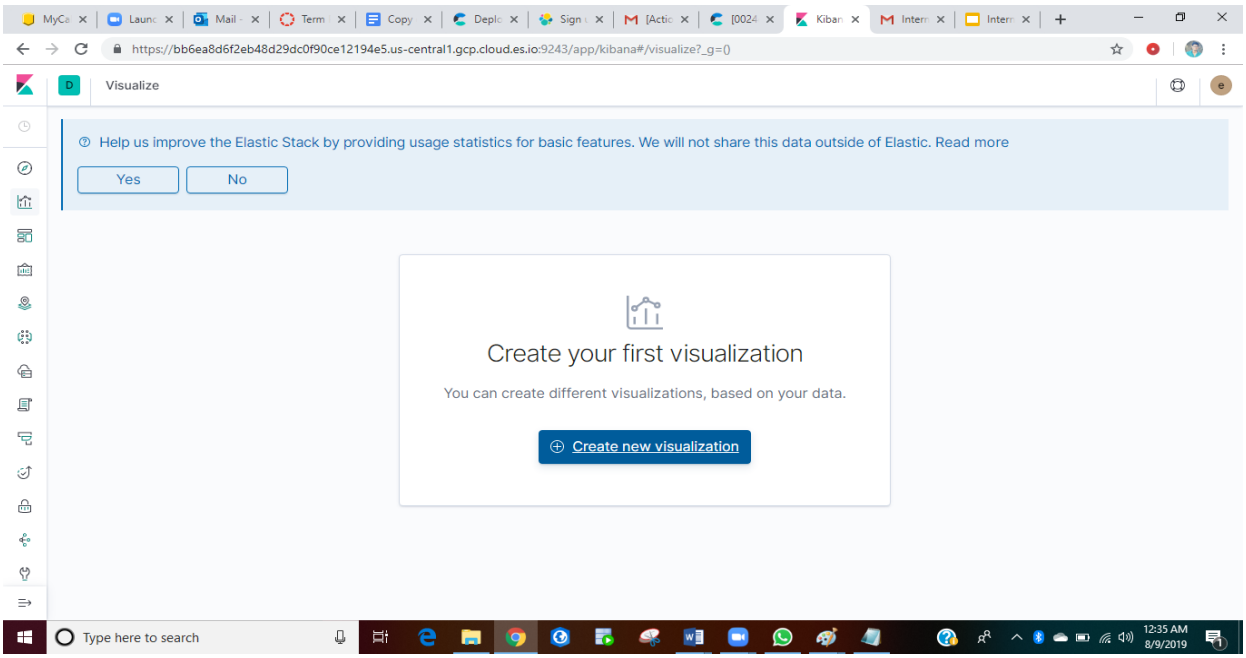
```
Country_or_area : "United States" and year > 2013
```



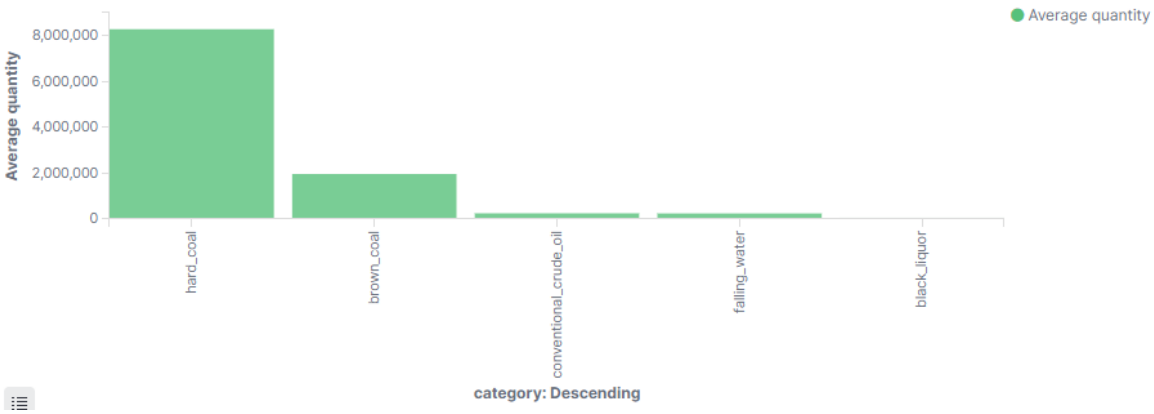
Step 4: Visualization

In the Visualize application, you can shape your data using a variety of charts, tables, and maps, and more.

In Kibana page, open **Visualize** and click **Create new Visualization**



1. Create a Vertical Bar chart that has :
    - Y-axis
        > Aggregation : Average
        > Field : quantity
    - X-axis
        > Aggregation : Terms
        > Field : category
        > Order by : Metric Average quantity
        > Order : Descending
        > Size : 5
    - Click **Apply Changes**
    - This Vertical Bar graph displays the average quantity of energy consumed by each energy source. By analyzing the bar chart, one can see the Hard_coal is utilized the most and black_liquor is consumed the least.



    - Click **Save** in the top menu bar and enter **VerticalBar1**

2. Create a Vertical Bar chart that has :
    - Y-axis
        > Aggregation : Count
    - X-axis
        > Aggregation : Terms
        > Field : Year
        > Order by : Metric Count

> Order : Descending

> Size : 5

- Split series

  > Sub aggregation : Terms

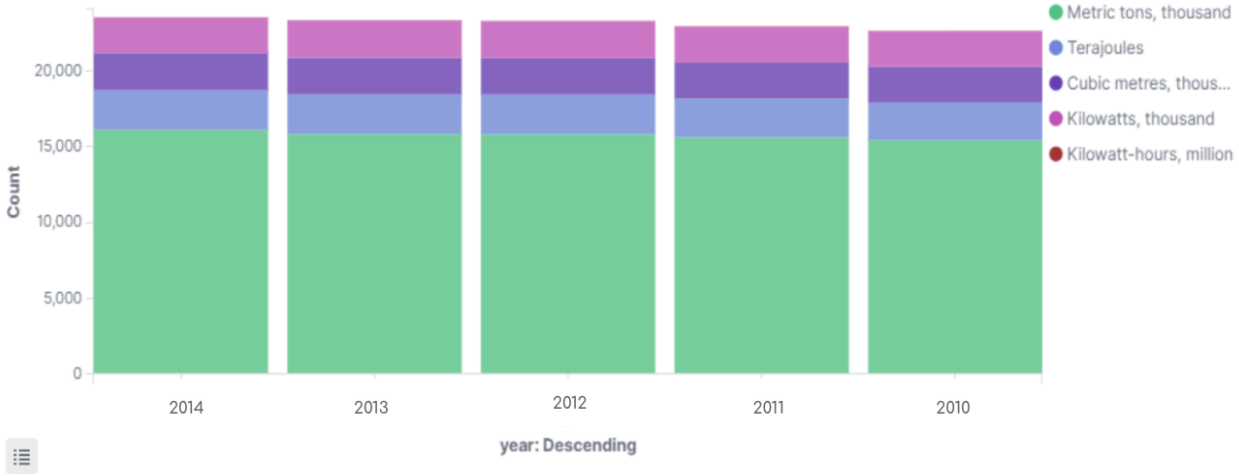
  > Field : Unit

  > Order by : Metric Count

  > Order : Descending

  > Size : 5

- Click **Apply Changes**
- The Vertical Bar Chart shows the units of measurement of energy used. In the graph are different units for energy: Metric tons, Cubic Meters, Tera Joules, Kilowatt and Kilowatt-hour which signify the energy usage of various sources of energy such as Electricity which is measured in Kilowatt-hours while energy usage of Natural gas is Cubic meters for various years.



- Click **Save** in the top menu bar and enter **VerticalBar2**

3. Create a Goal chart that has :

- Metrics

  > Aggregation : Count

- Split group

  > Aggregation : Terms

> Field : Year

> Order by : Metric count

> Order : Descending

> Size : 5

- Click **Apply Changes**
- The Goal Chart shows the percentage of Fuel consumed from the year 2010 to the year 2014. From the above graph, we notice that the fuel consumption is increasing each year with 235% being the highest in 2014 (data available till 2014)*.

- Click **Save** in the top menu bar and enter **GoalChart**

4. Create a Tag cloud chart that has :
   - Metrics

     > Tag size : Count
   - Tags

     > Aggregation : Terms

     > Field : country_or_area

     > Order by : Metric count

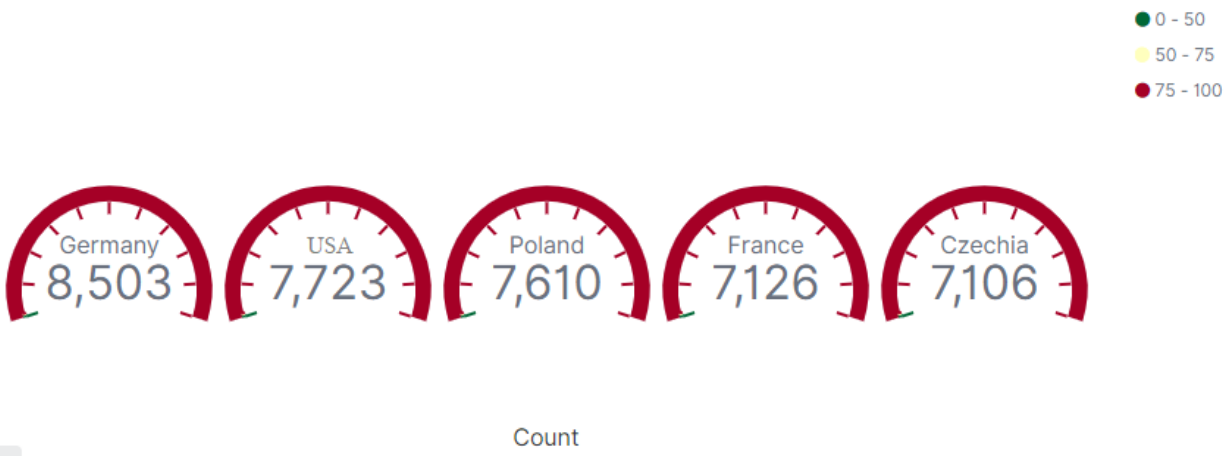
     > Order : Descending

     > Size : 15
   - Click **Apply Changes**

- This tag cloud visual representation which visualizes the information pertaining to the country which consumes the highest energy. According to the data we found out that Germany consumes the highest amount of energy followed by the United States and Poland.

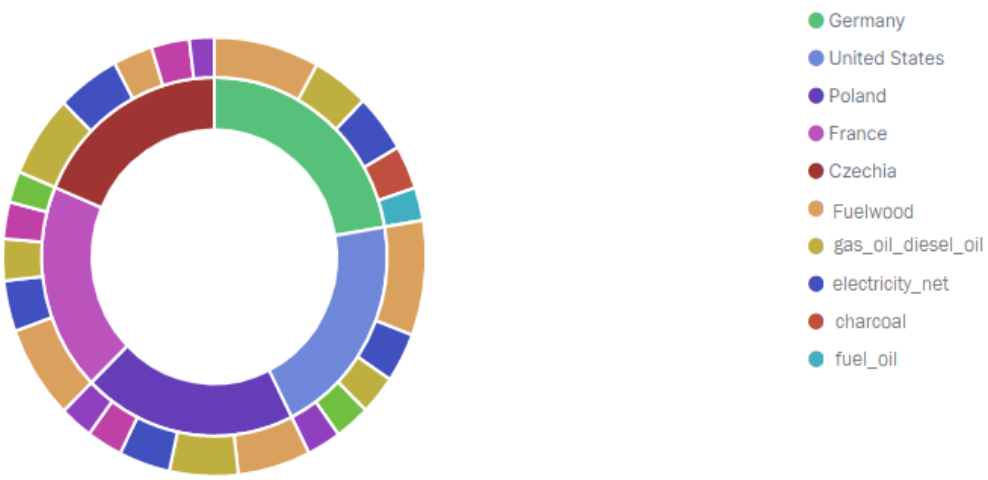
- Click **Save** in the top menu bar and enter **TagChart**

5. Create a Gauge chart that has :
   - Metrics
     > Metrics : Count
   - Split group
     > Aggregation : Terms
     > Field : country_or_area
     > Order by : Metric Count
     > Order : Descending
     > Size : 5
   - Click **Apply Changes**
   - The Gauge Chart demonstrates the percentage of the Fuel consumed to produced in each country. The graph shows that all countries utilize more than 75% of their energy produced respectively where Germany tops the list of countries producing the highest amount of energy.



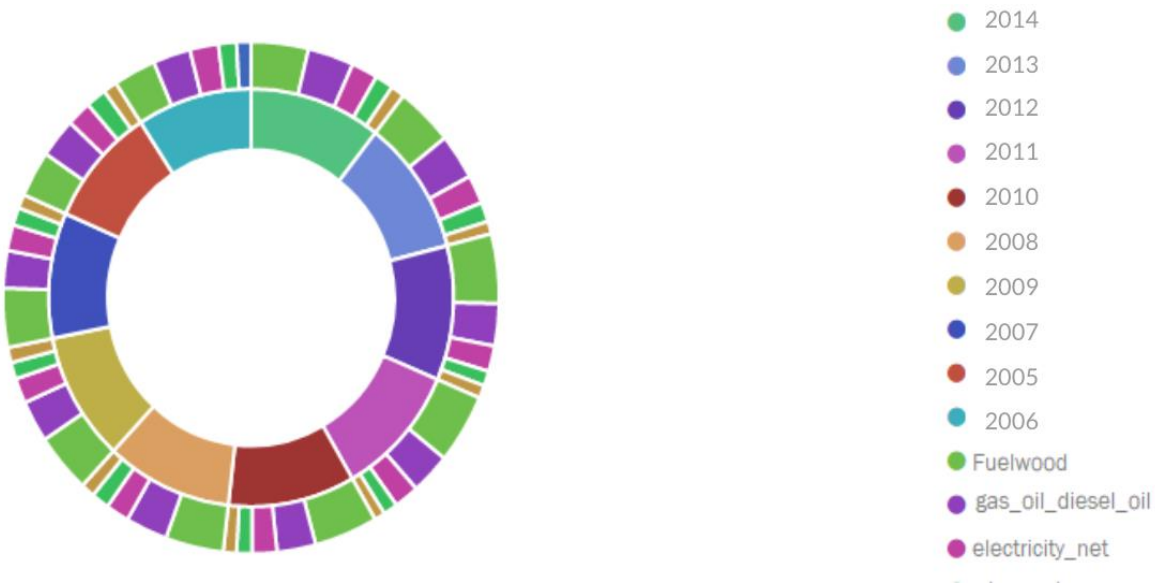
- Click **Save** in the top menu bar and enter **GaugeChart**

6. Create a Pie chart that has :
   - Slice size
     > Aggregation : Count
   - Split size
     > Aggregation : Terms
     > Field : country_or_area
     > Order by : Metric Count
     > Order : Descending
     > Size : 5
   - Split size
     > Aggregation : Terms
     > Field : quantity
     > Order by : Metric Count
     > Order : Descending
     > Size : 5
   - Click **Apply Changes**
   - The Pie Chart illustrates the Major source of energy used in each country. The analysis shows that fuelwood is widely used in Germany, United States, France, Poland where as gas_oil_diesel_oil is commonly used in Czechia.



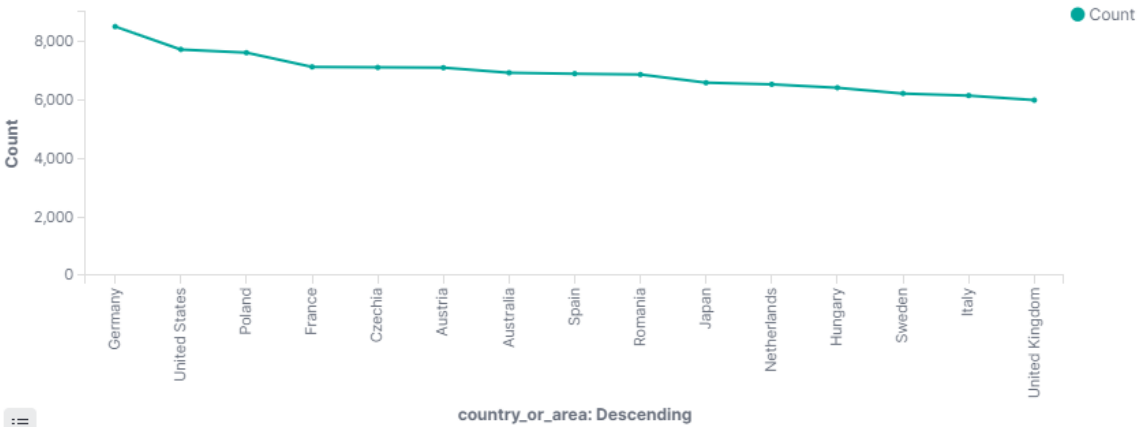
- Click **Save** in the top menu bar and enter **PieChart1**

7. Create a Pie chart that has :
   - Slice size
     > Aggregation : Count
   - Split size
     > Aggregation : Terms
     > Field : year
     > Order by : Metric Count
     > Order : Descending
     > Size : 10
   - Split size
     > Aggregation : Terms
     > Field : quantity
     > Order by : Metric Count
     > Order : Descending
     > Size : 5
   - Click **Apply Changes**
   - The Pie Chart gives the Major Source of energy produced and consumed from the year 2005-2014. We observe that Fuelwood is playing a vital role through all the years and it is highly produced and utilized.



- Click **Save** in the top menu bar and enter **PieChart2**

8. Create a Line chart that has :
    - Y-axis
        > Aggregation : Count
    - X-axis
        > Aggregation : Terms
        > Field : country_or_area
        > Order by : Metric Count
        > Order : Descending
        > Size : 15
    - Click **Apply Changes**
    - The line graph illustrates the amount of energy consumed by various countries. The line graph represents that Germany consumes more than 8000 Metric Tons of energy followed by The United States and United kingdom being the last.



- Click **Save** in the top menu bar and enter **LineChart**

9. Create a Data table chart that has :
    - Metric
        > Aggregation : Max
        > Field : quantity
    - Split rows
        > Aggregation : Terms
        > Field : year
        > Order by : Metric Max quantity
        > Order : Descending
        > Size : 10
    - Split rows
        > Aggregation : Terms
        > Field : country_or_area
        > Order by : Metric Count
        > Order : Descending
        > Size : 1
    - Click **Apply Changes**
    - The Data Table exhibits the countries which produced the highest amount of energy in each year from 2005-2014. The chart illustrates that The United States produced highest energy in 2014,2013 and 2012 where as, Germany produced highest energy from 2005-2011.

year: Descending	country_or_area: Descending	Max quantity
2014	United States	4,340
2013	United States	4,306
2012	United States	4,291
2011	Germany	3,885
2010	Germany	3,626
2009	Germany	3,223
2008	Germany	2,989
2007	Germany	2,817
2006	Germany	2,446
2005	Germany	2,126

- Click **Save** in the top menu bar and enter **DataTable**

References

https://www.kaggle.com/unitednations/international-energy-statistics#all_energy_statistics.csv

https://github.com/sgontya/InternationalEnergyStatisticsourGithub