CIS 3200 Term Project Tutorial Group 3

Authors: Vanessa Munoz; Shawn Tran; Hae-In Kim; David Montes

Instructor: Jongwook Woo

Date: 05/11/2021

Lab Tutorial

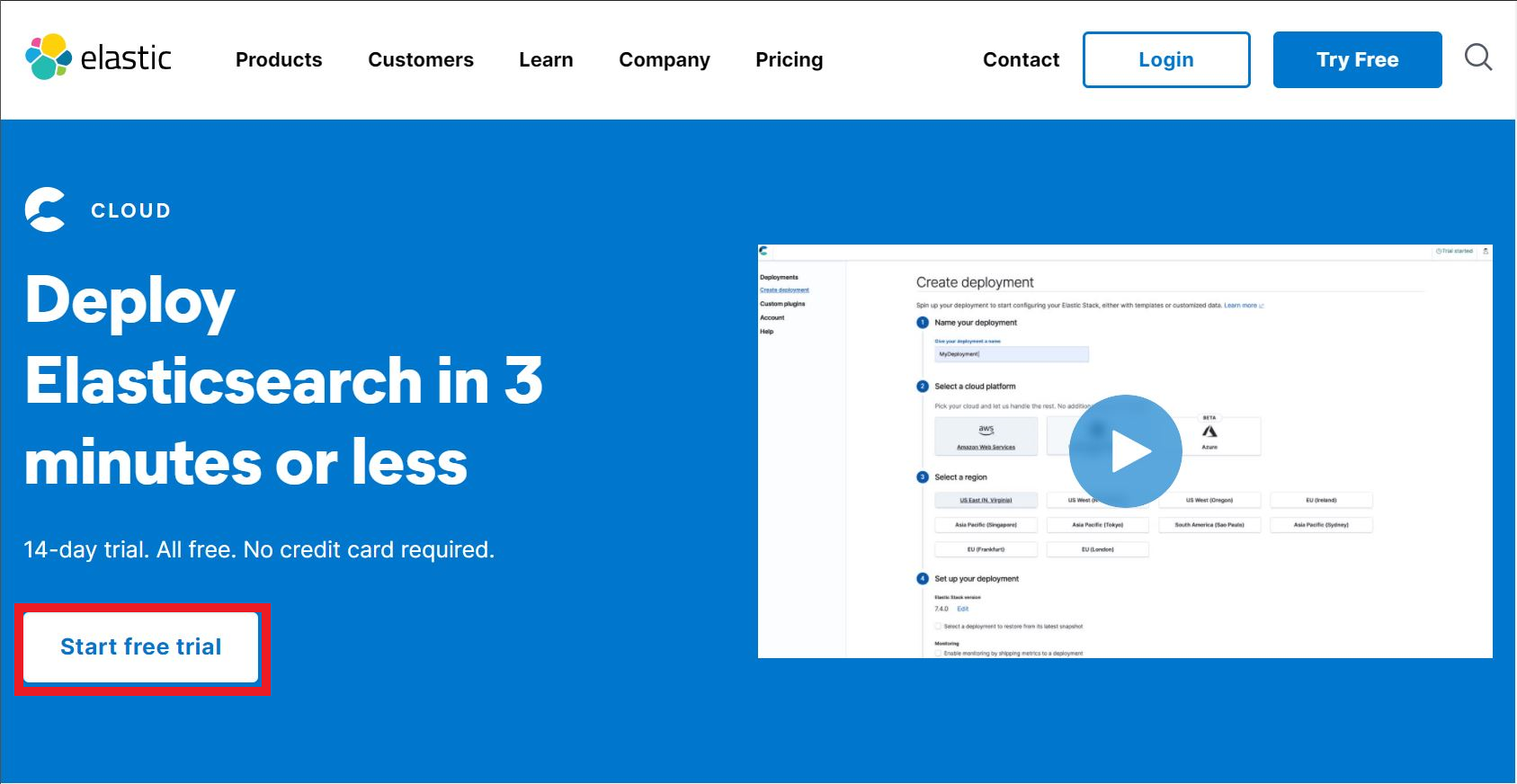
Objectives

In this Hands-on lab you will learn how to: Visualize a dataset through ElasticSearch and Kibana and create AzureML experiments to predict data.

Step 1: Creating an account on Elastic Cloud and logging in

In this first step we will go through signing up and setting up an Elastic Cloud

1. Visit<https://www.elastic.co/> and click the **Start free trial** button.



2. Create your account using an existing email.

3. Click on the verification link in the email you provided.

4. Follow the prompts and proceed to login to Elastic Cloud.

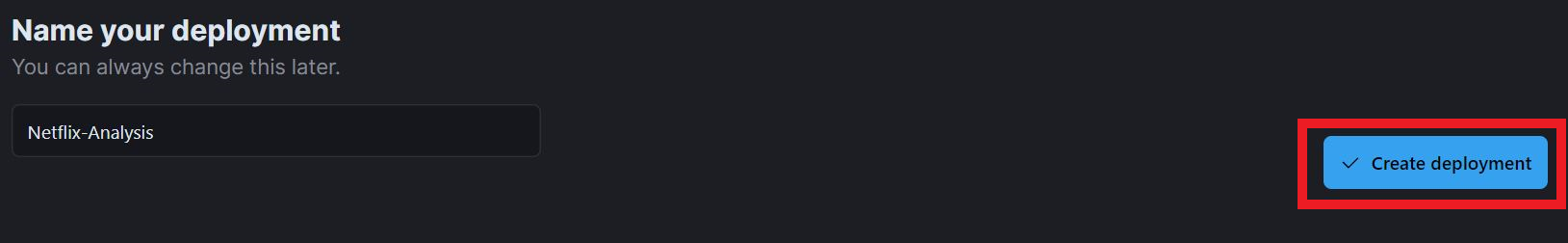
Step 2: Creating your first hosted Elastic Search Cluster

1. After signing in, click the **Start your free trial** button to begin creating a deployment.

2. Click **Select** under the General Purpose box.

3. Give your deployment a name.

a. In this case we have named our deployment *Netflix-Analysis.*

**

4. Select **Create deployment**.

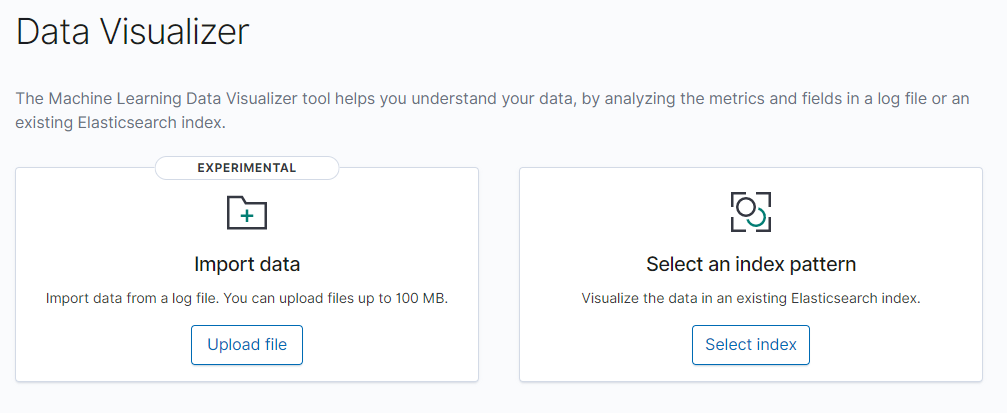
5. Take note of the username and password for this deployment.

6. After a few minutes once the deployment has been created click **Open Kibana**.

7. Click **Explore on my own**.

Step 3: Adding data to Elastic Cloud and index mapping

1. Click the data visualizer tab under the machine learning section of Kibana.

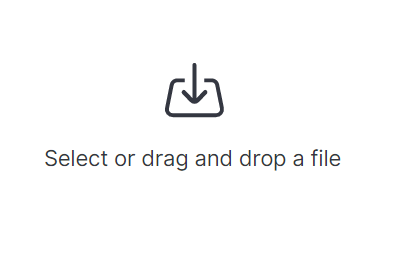


2. Open a new browser tab or new browser window and navigate to the following link:

https://github.com/stran13/CIS3200Netflix

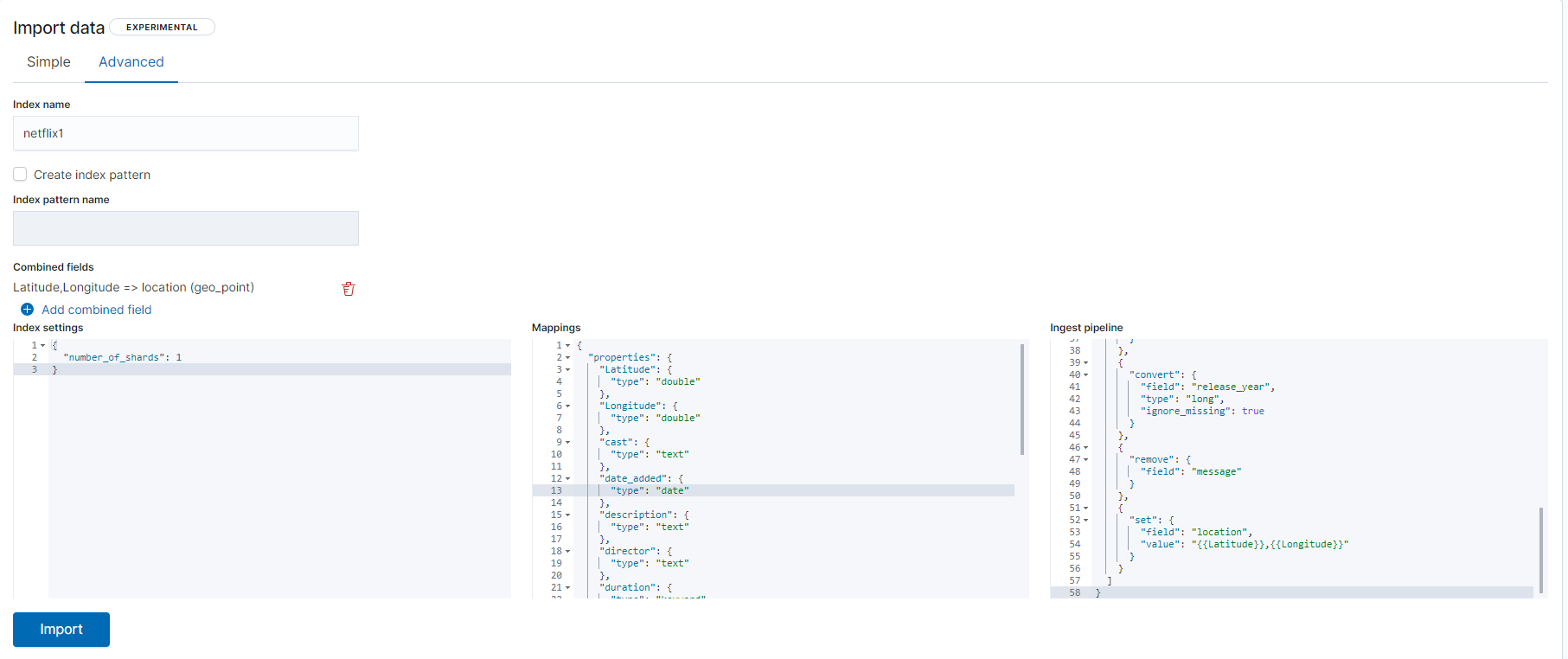
3. Click the netflix\_titles.csv file and download it.

4. Go back to the previous browser window, drag the **netflix\_titles.csv** to the icon shown below:

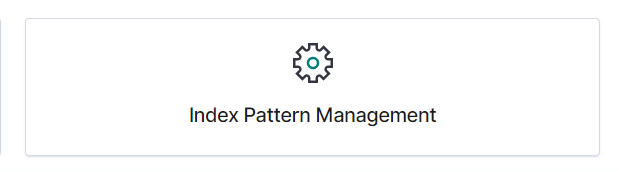


5. On the next screen, click **Import**.

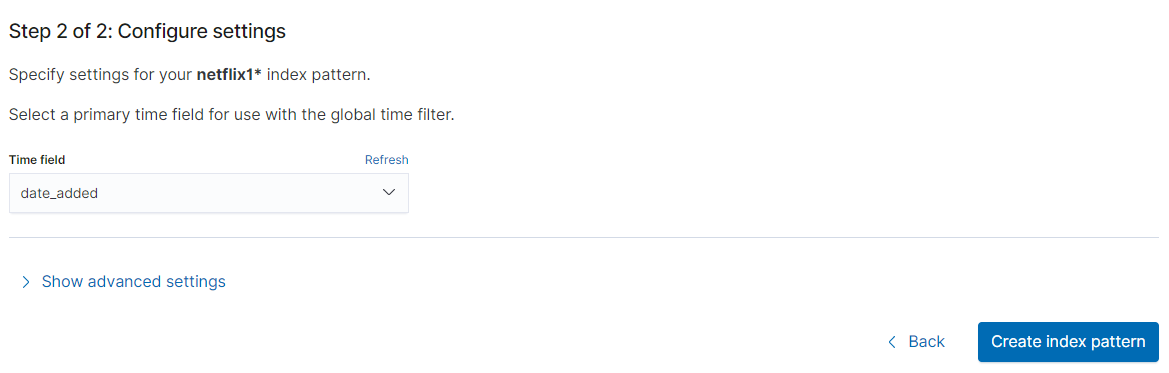
6. Select the advanced tab and uncheck the **Create index pattern** button. Then name the index anything you want, in this case we named it netflix1. Then under the mappings section, change the type of **date\_added** to date. Make sure your screen looks like this and then click import.



7. After the file has been imported click the index pattern management button at the bottom of the screen.



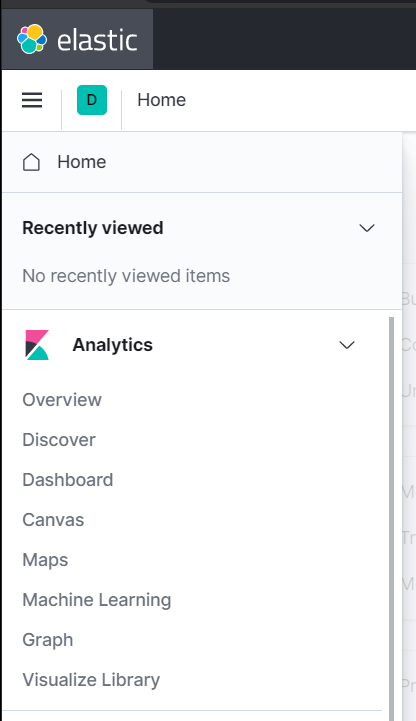
8. Once here, click the create index pattern button and type the name of the index you created previously. Then click next step. Select date\_added under the Time field drop down menu and select create index pattern.



Step 4: Creating a Bar Graph

1. Click on the 3 lines in the top left corner to bring up the side bar.

2. Click on **Visualize Library**.



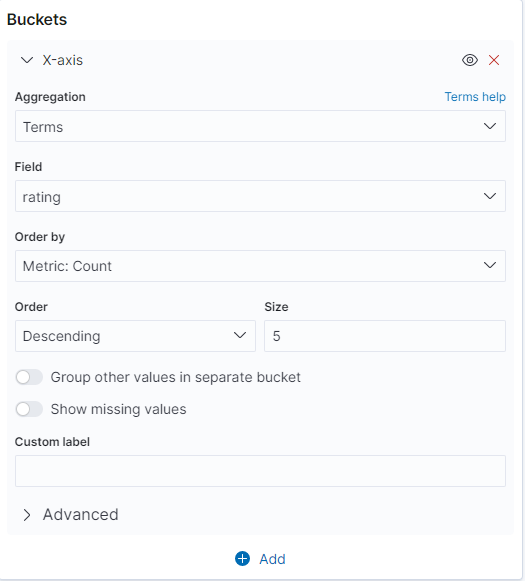
3. Click on **Create new visualization.**

4. Click on **Vertical bar** and type the name of the created index pattern and select it.

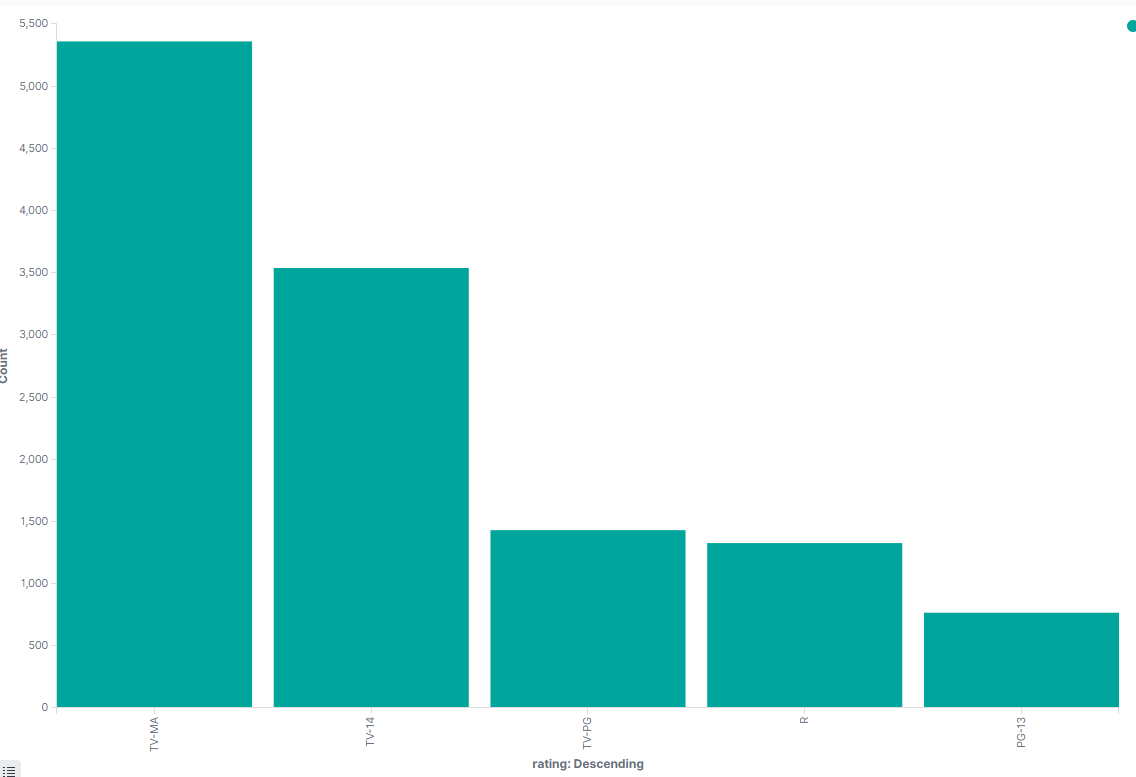
5. Change the date at the top to absolute from 1970 to today’s date and click update.



6. Click the add button under the buckets section on the right panel. Select X-axis and configure the bucket as so and click update.



7. The bar chart should look something like this.



8. Save the bar chart as **netflixbar**.

Step 5: Creating a Pie Chart

1. Click on the 3 lines in the top left corner to bring up the side bar.

2. Click on **Visualize** again.

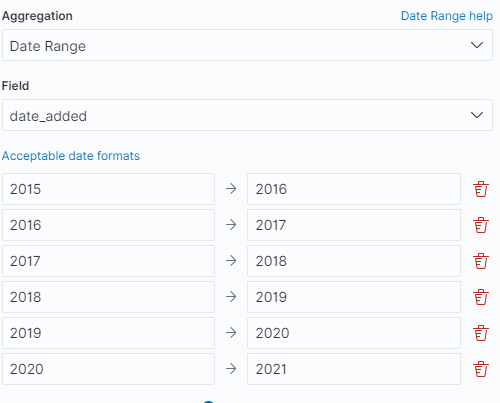
3. Click on **Create new visualization.**

4. Click on **Pie**.

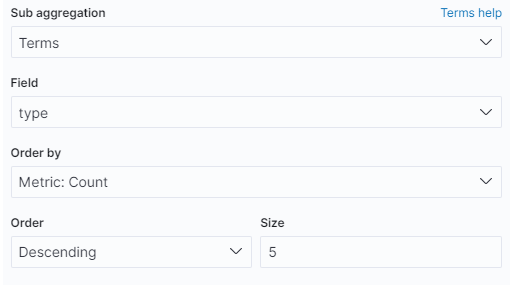
5. Search the index pattern you created and select it.

6. Make sure the date at the top right matches the previous dates on the bar graph.

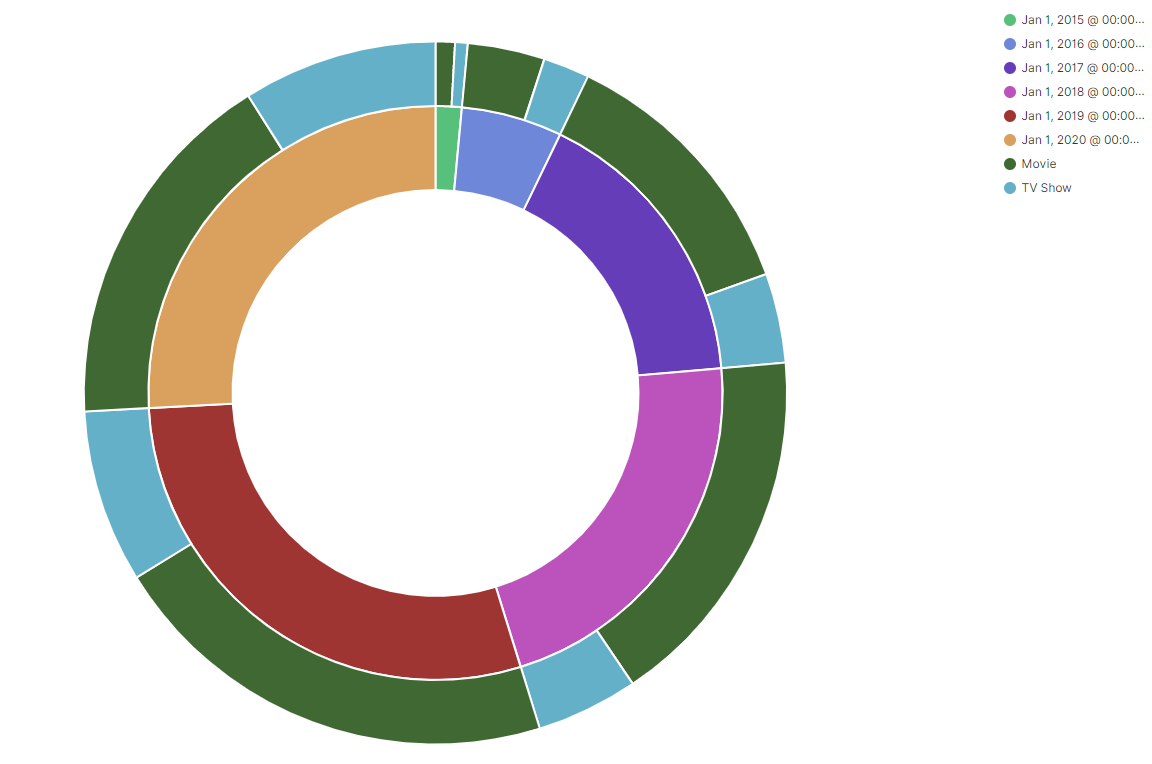
7. Click add buckets and select split slices. And configure the bucket as so and click update.



8. Then create another split slices and configure it as so and click update.



9. Once this is done your pie chart should look like this.



10. Hit save at the top right and save it as **netflixpie**.

Step 6: Creating a Geo-Map

1. Click on the 3 lines in the top left corner to bring up the side bar.

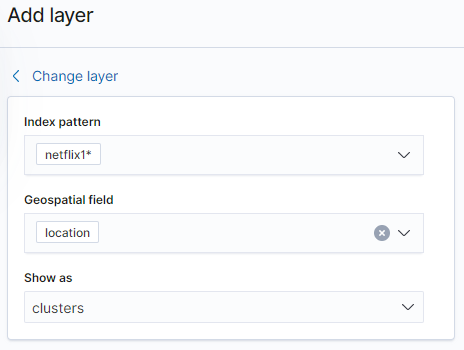
2. Click on **Maps** under the Kibana section.

3. Click on **Create map.**

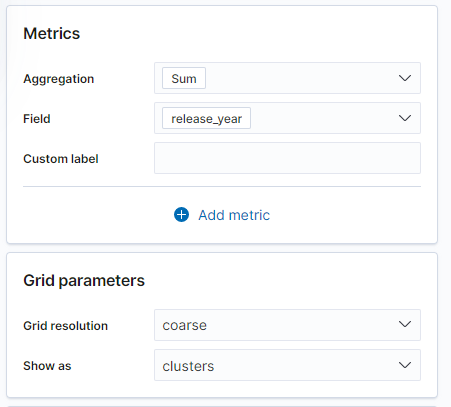
4. Ensure the dates on the top right matches the previous dates**.**

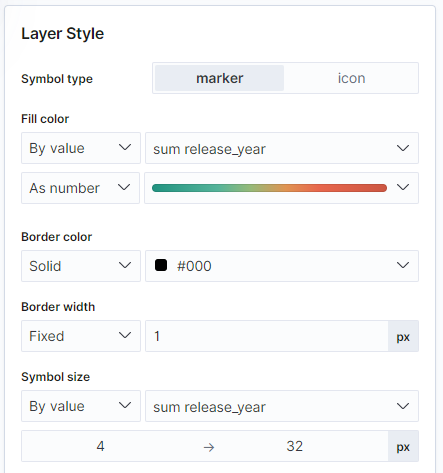
5. Select add layer and clusters and grids and select the index pattern you previously created.

6. Make sure the layer looks like this and then click add layer.

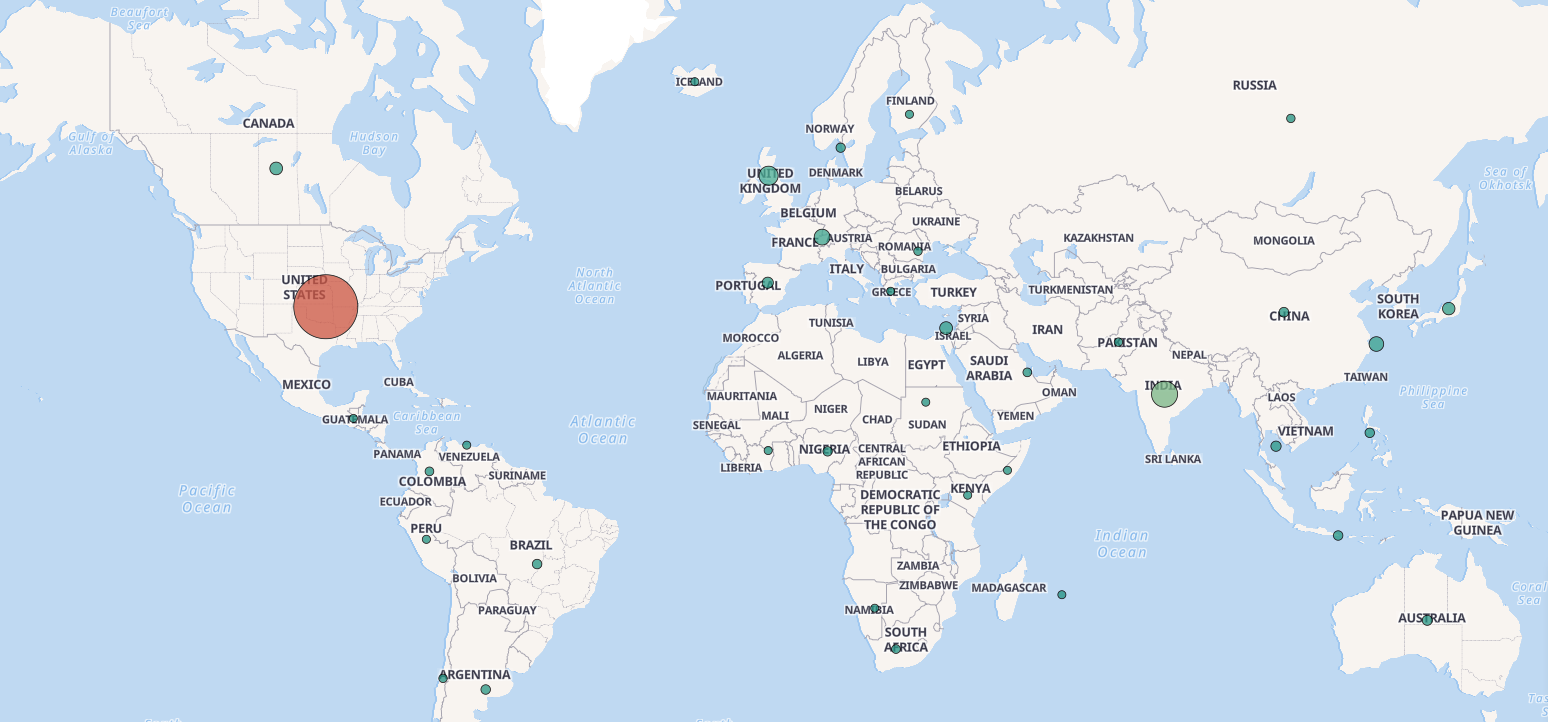


7. Match the source details for the layer as so.





8. Click save and close and save the map as **netflixmap**. It should look similar to this.

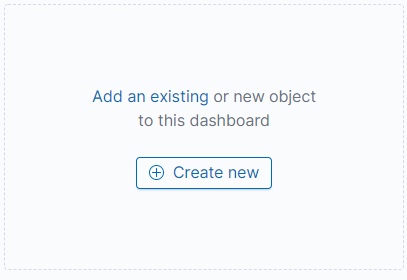


Step 7: Creating a Dashboard

1. Select **Dashboard** under the Kibana section on the left panel.

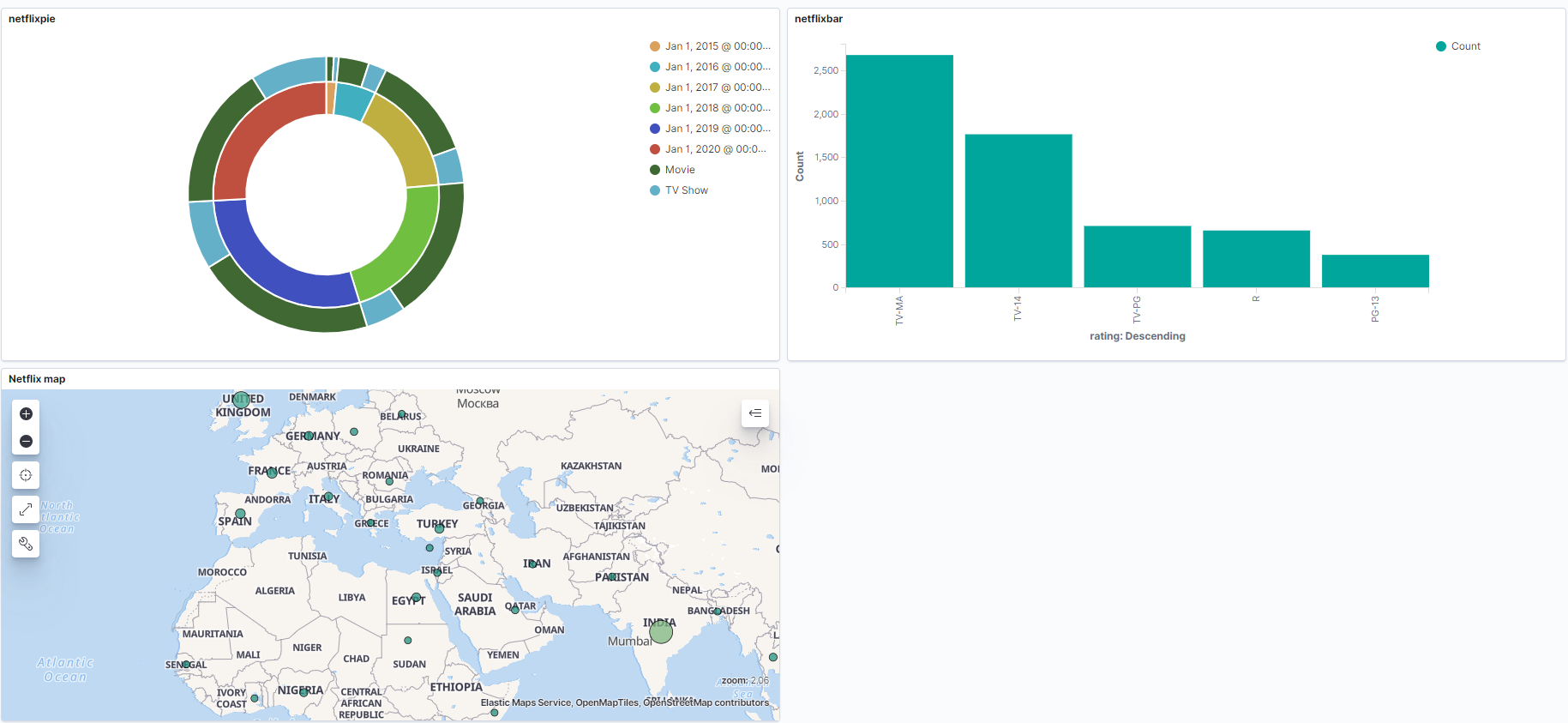
2. Click on create dashboard and ensure the top right dates match the previous visualizations.

3. Then click **add an existing** in the blank area.



4. Search for netflixbar, netflixpie, and netflixmap and select them.

5. The output should look like this.

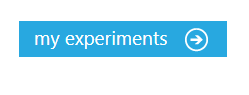


6. Hit save at the top right of the interface and save it as **netflixdashboard**.

Step 8: Creating AzureML Experiments

**Experiment 1:**

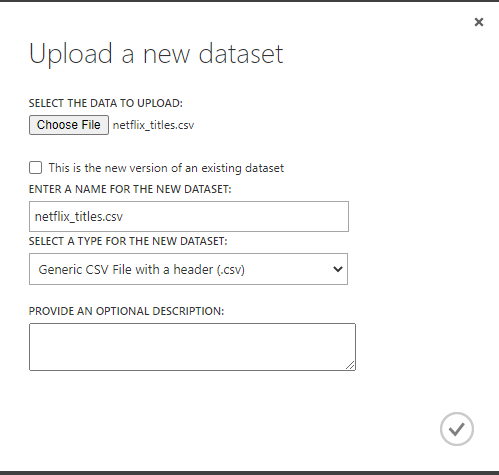
1. Create an AzureML account at studio.azureml.net, once you log in go to my experiments



1. Once here, hit new at the bottom left to add the dataset

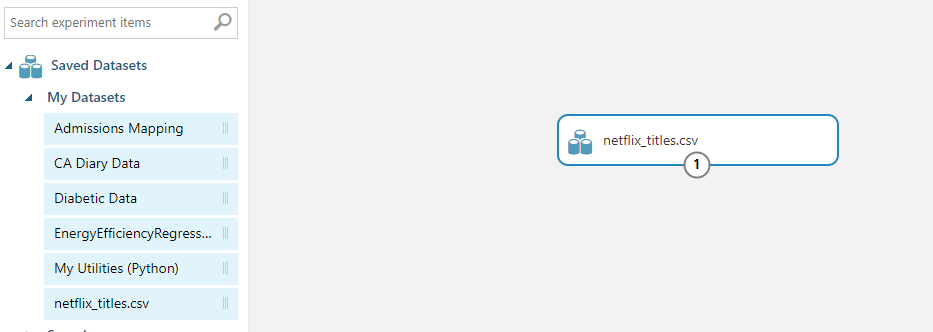


1. Select dataset on the left panel and select local file. Then choose the netflix titles csv from your local machine.

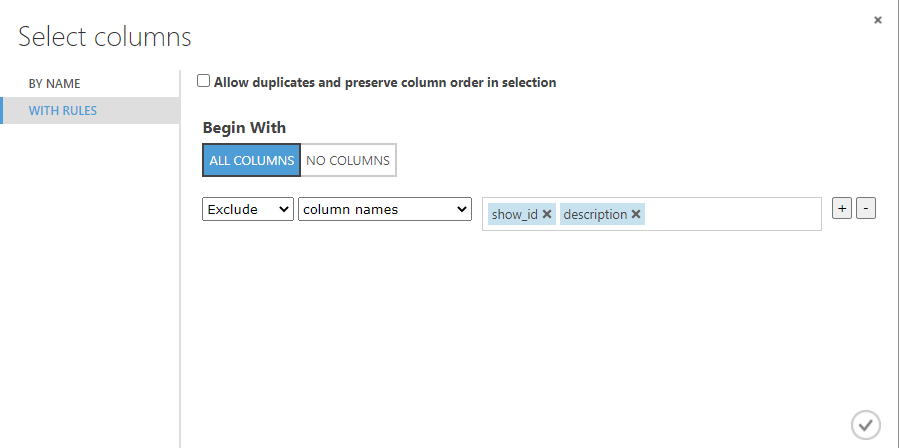


Click the checkmark button and wait for the csv to upload.

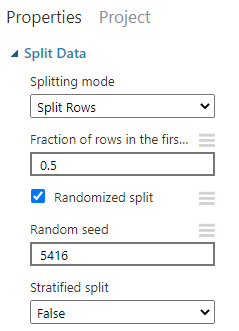
1. Once it has finished click the new button once again and create a blank experiment. Rename the experiment Netflix titles and drag the dataset onto the canvas.



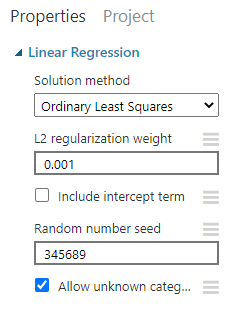
1. Search for the select columns in dataset module and drag it under the dataset. Then connect the output of netflix\_titles to the input of the select columns module.
2. Select the Select columns in dataset module and go to the properties pane, launch the column selector and configure it like this.



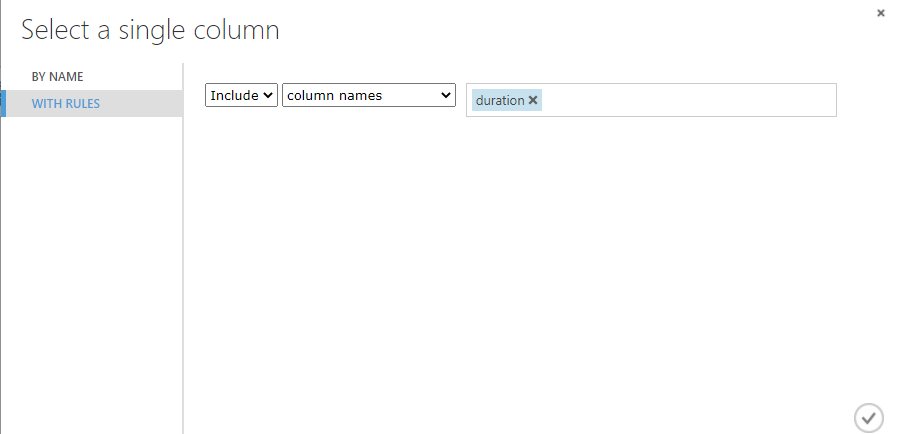
1. Then search for the split data module and drag it below the select columns in dataset module. Configure it as so.



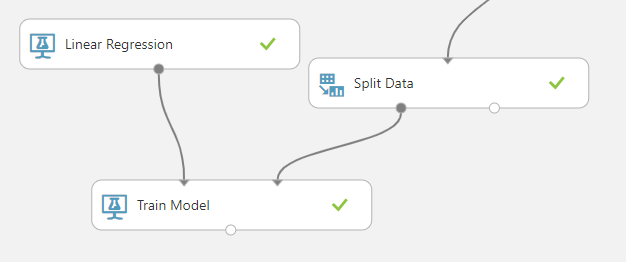
1. Search for the linear regression module and place it on the side of the split data module. Configure it like this.



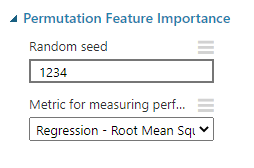
1. Search for the Train model module and place it under the linear regression and split data modules. Select the train model module and launch the column selector and select duration.



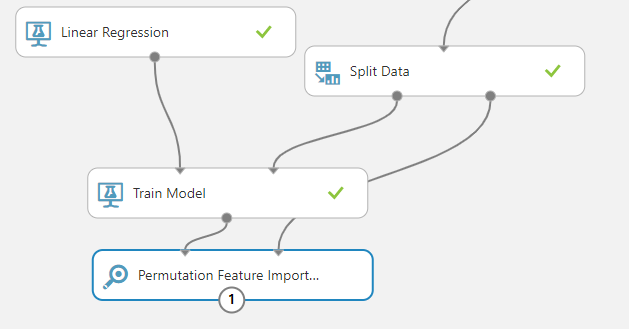
1. Connect the output of linear regression to the left input of train model and the left output of split data to the right input of train model as seen below.



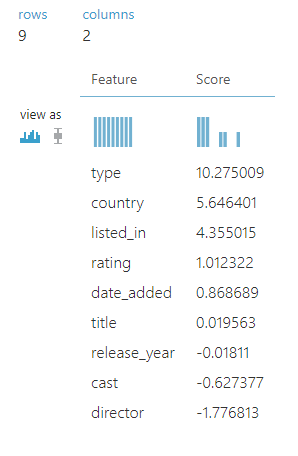
1. Search for the Permutation Feature Importance module and place it under the Train Model module with these properties.



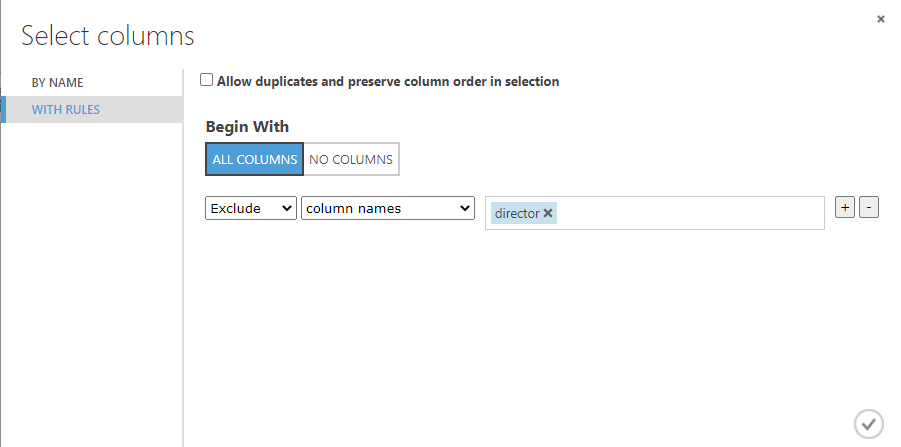
1. Connect the output of train model to the left input of permutation feature importance and connect the right output of split data to the right input as seen below. Then save and run the experiment.



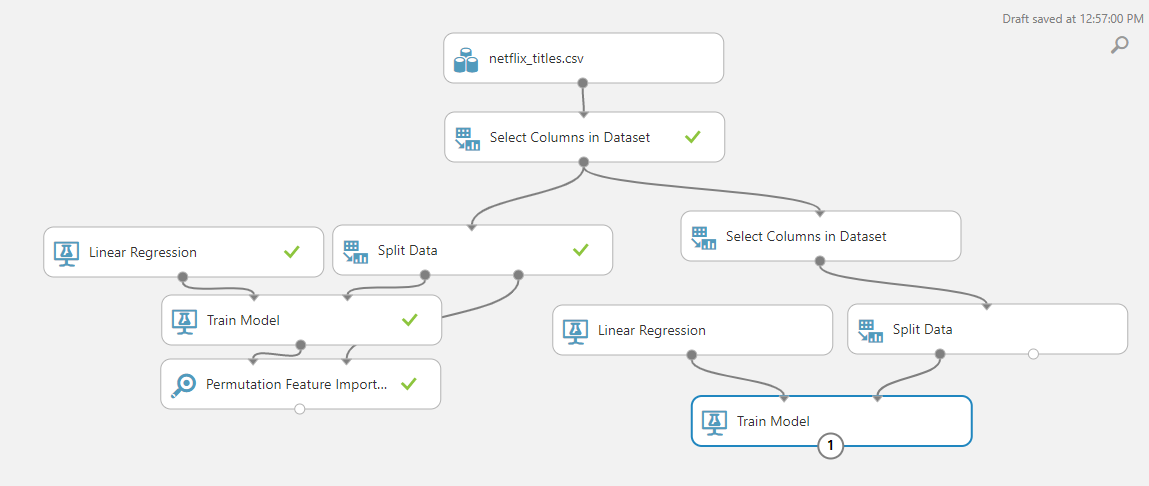
1. After running the experiment, visualize the permutation feature importance, it looks like this.



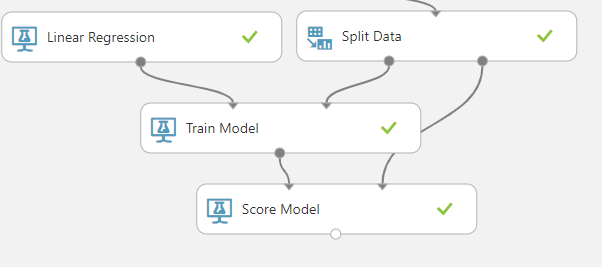
1. Add another Select columns in dataset module under the first and configure it like this.



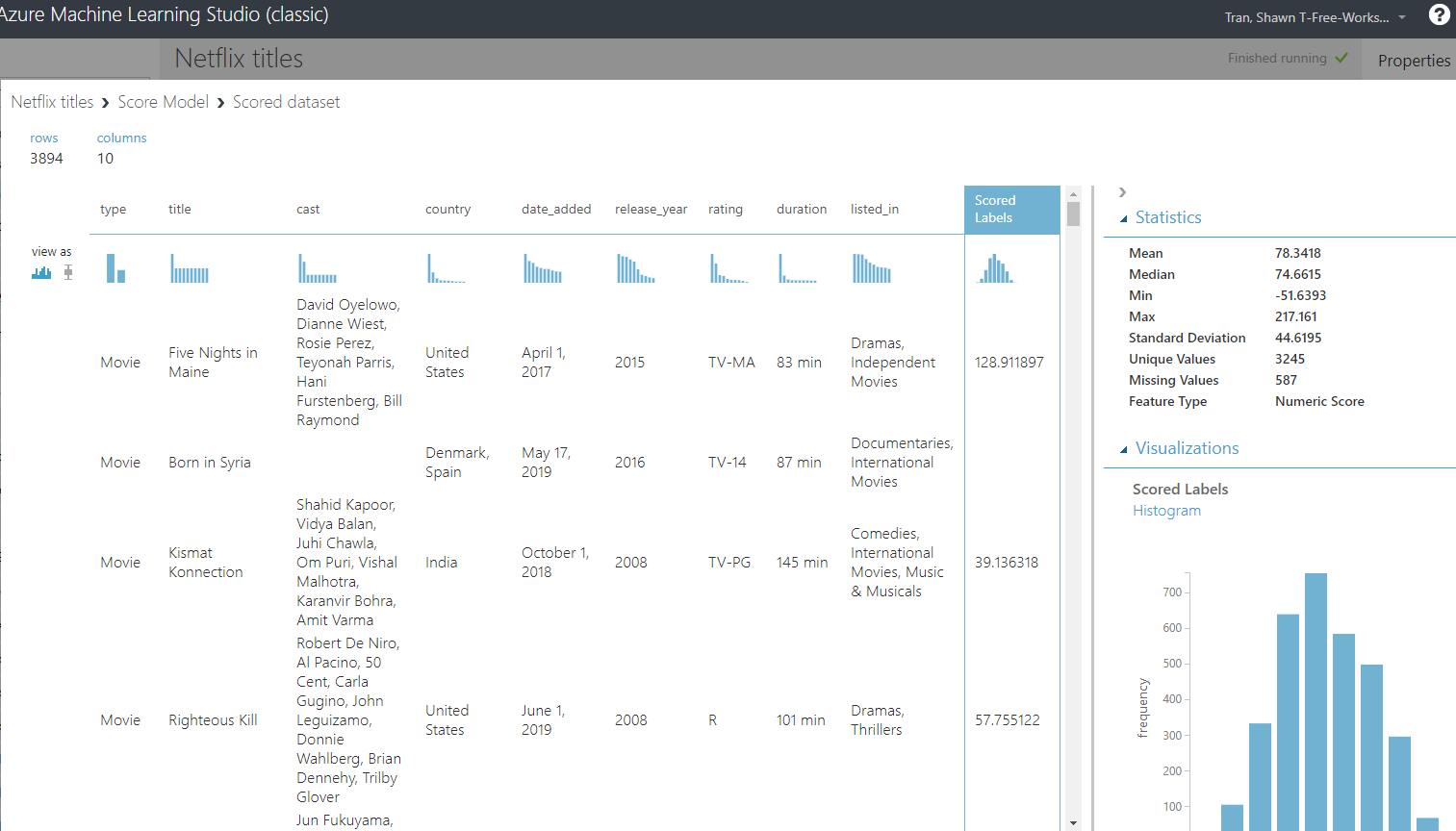
1. Click and drag around the Linear regression, split data, and train model modules and copy paste them under the new select columns module. Then connect the output of the new select columns in dataset to the new split data input. The experiment should look like this.



1. Search for the Score model module and add it under the train model module. Connect the output of the train model to the left input of score model and the right output of split data to the right input of score model as seen below.

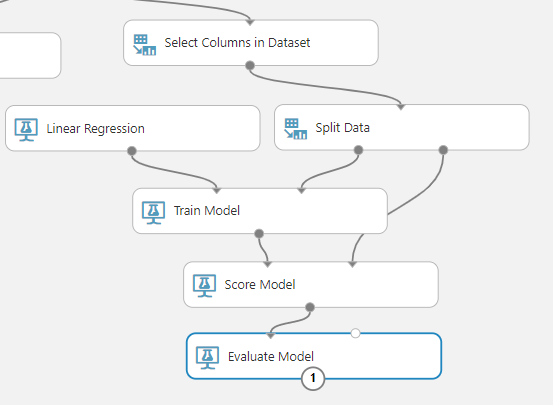


1. Save and run the entire experiment and visualize the scored dataset, it should look like this.

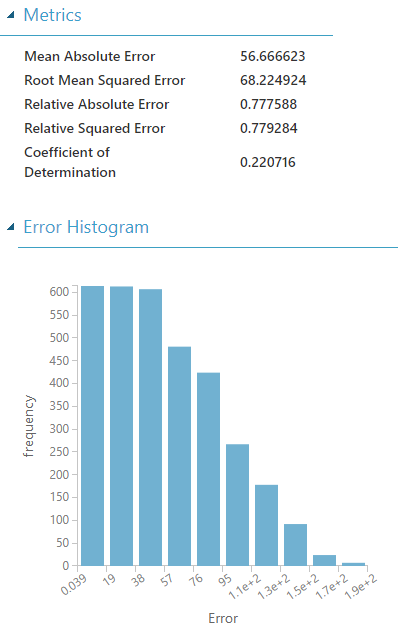


**Experiment 2:**

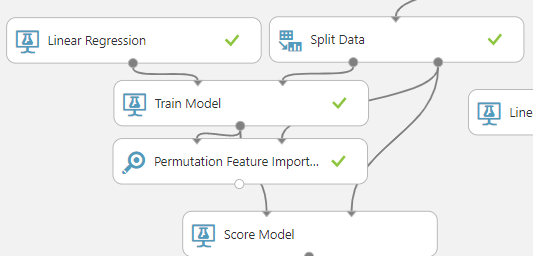
1. Go to studio.azureml.net and load the previous experiment netflix titles. Save it as a copy named netflix titles 2.
2. Search for Evaluate model and drag it under the right side set of modules like so.



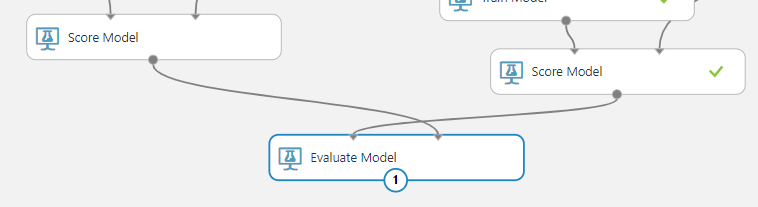
1. Save and run the experiment and visualize the output of the evaluate model module.



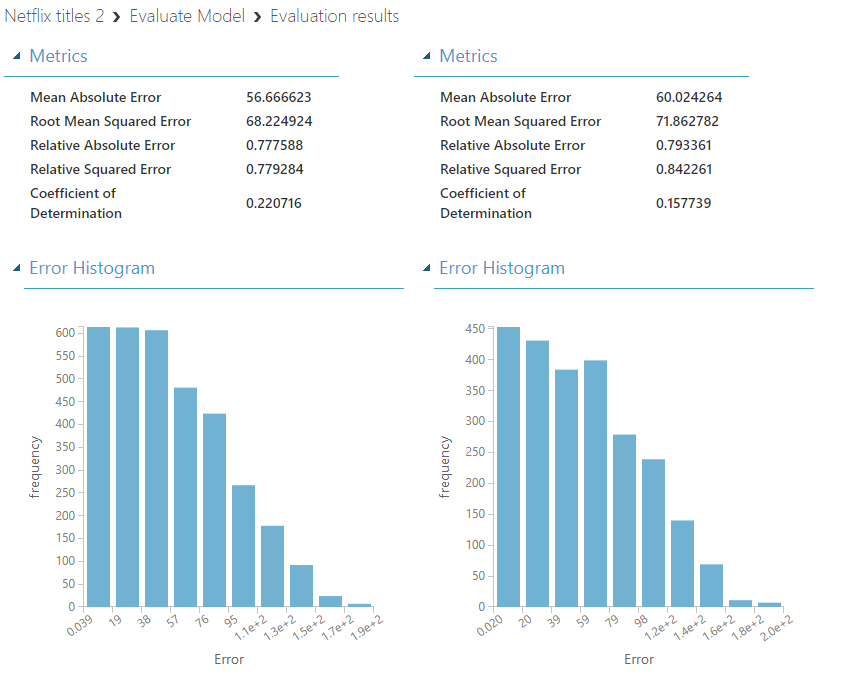
1. Search for the score model module and place it under the left set of modules under permutation feature importance. Connect the train model output to the left input of the new score model module and connect the right output of the split data module to the right input of the new score model module like so.



1. Then connect the output of the new score model to the right input of the evaluate model as seen below. Then save and run the experiment.



1. Visualize the results of the evaluate model module, it looks like this.



Here we see the first model we evaluated on the left has a better performance, but both models are still not ideal as the RMSE are both above 65 and the CoD are both under 25.

**This is the end of the lab.**