

Vignesh Netflix Data Analysis using AWS QuickSight

Step 1:

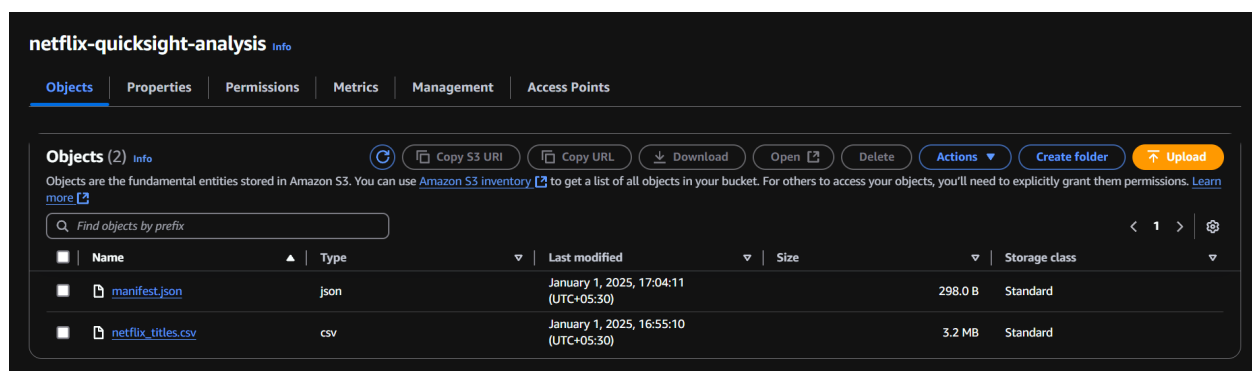
Download the Dataset

- Download netflix_titles.csv
- Download manifest.json

Step 2:

Store the Dataset in Amazon S3

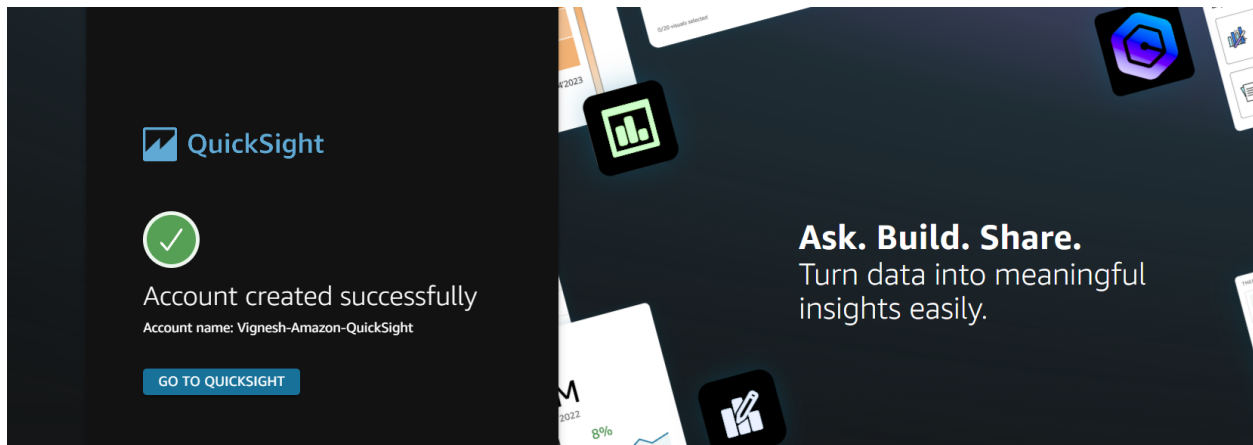
- Store the netflix_titles.csv in bucket.
- Copy the S3 URI of netflix_titles.csv.
- Open manifest.json in text editor and paste the copied URI and save it
- Upload manifest.json to the bucket



Step 3:

Create Amazon QuickSight Account

- Free tier (Enterprise) and type an account name and notification email id
- Set nearby region
- Enable quicksight's access to Amazon S3 Bucket → Select the specific bucket where data file is located.
- Be sure to uncheck the add-on as it is not required for this project.
- Click Finish
- The confirmation will look like this:

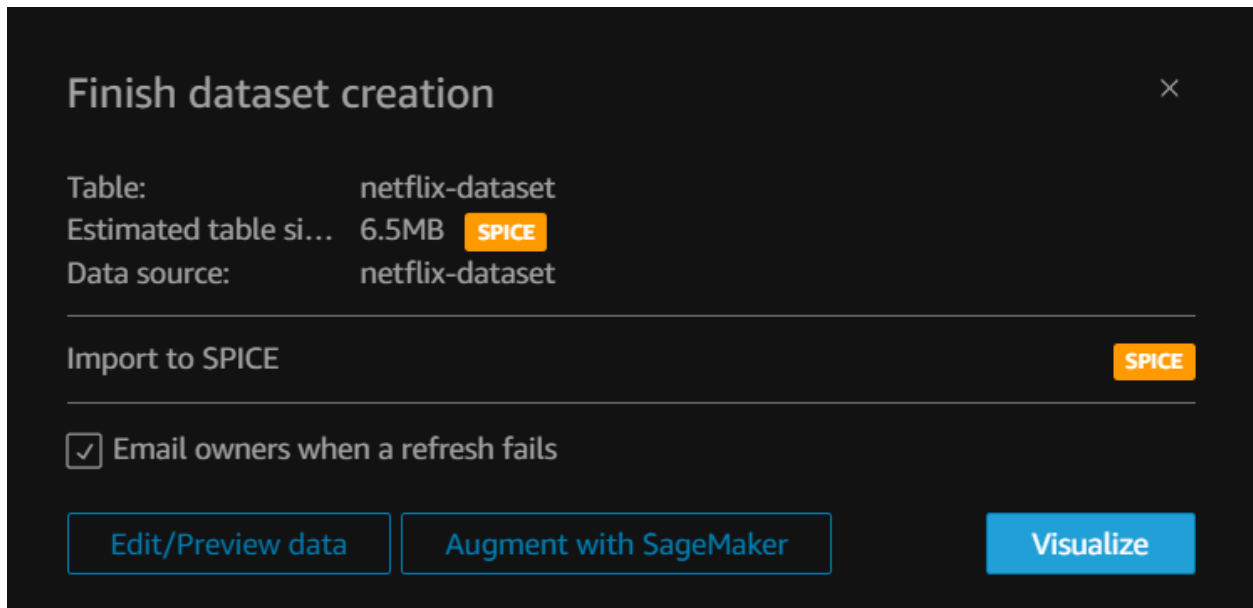


Step 4:

Connect S3 Bucket to Amazon QuickSight

- Click on **New analysis** → **New dataset** → Choose **S3**
- Give a name for data source (can be random name)
- Copy S3 URI of the manifest.json from the S3 bucket and paste it on the second text field.

- Click **Connect**
- Confirmation looks like this:



- Click **Visualize** and Click **Create**

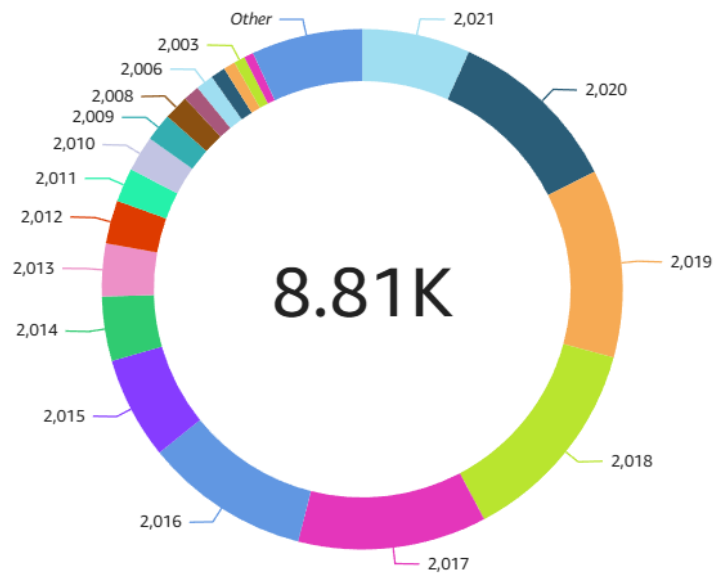
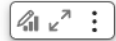
Step 5:

Visualization Starts!

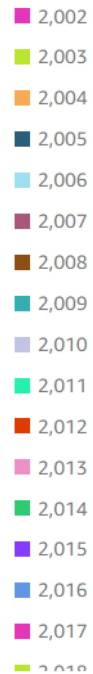
- The fields of the dataset are seen at the left side of the window
- First we can create a visualization for the **release_year** field. It contains counts of content (movies/tv shows) released for years.
- Click on **release_year** field → Change visual type to Donut Chart (can use any other too) and now you can see it's visualization like :

Count of Records by Release_year

SHOWING TOP 20 IN RELEASE_YEAR



Release Year

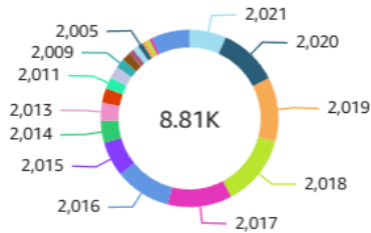


Group By: release_year ▾

- Resize it to add other visuals of the dashboard.
- Click on Add under Visuals to add new visuals.
- Next, a chart to show the number of tv shows and movies for a year.
- Select **release_year** and place it in y axis and select **type** and paste it in group

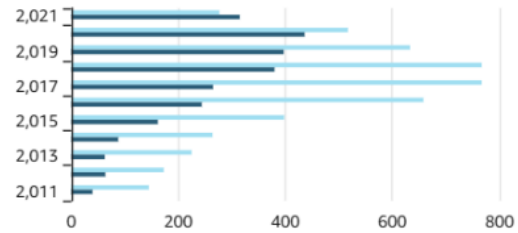
Count of Records by Release_year

SHOWING TOP 20 IN RELEASE_YEAR



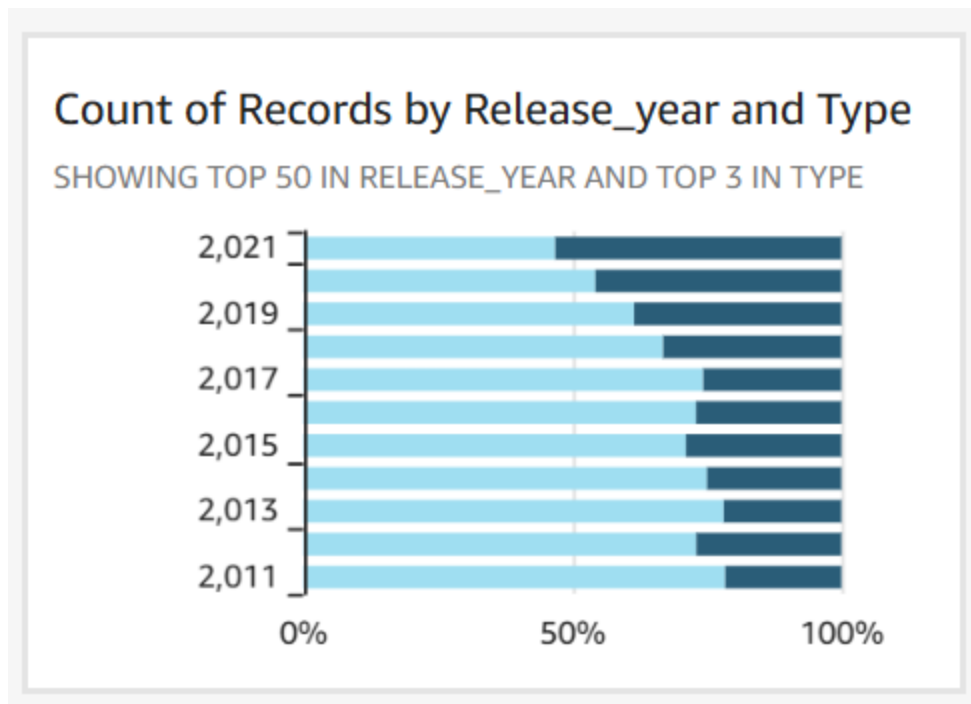
Count of Records by Release_year and Type

SHOWING TOP 50 IN RELEASE_YEAR AND TOP 3 IN TYPE



- New Charts:

3.The breakdown of TV shows/movies for each release year:



4.Show the number of movies vs TV shows per release year in a table:

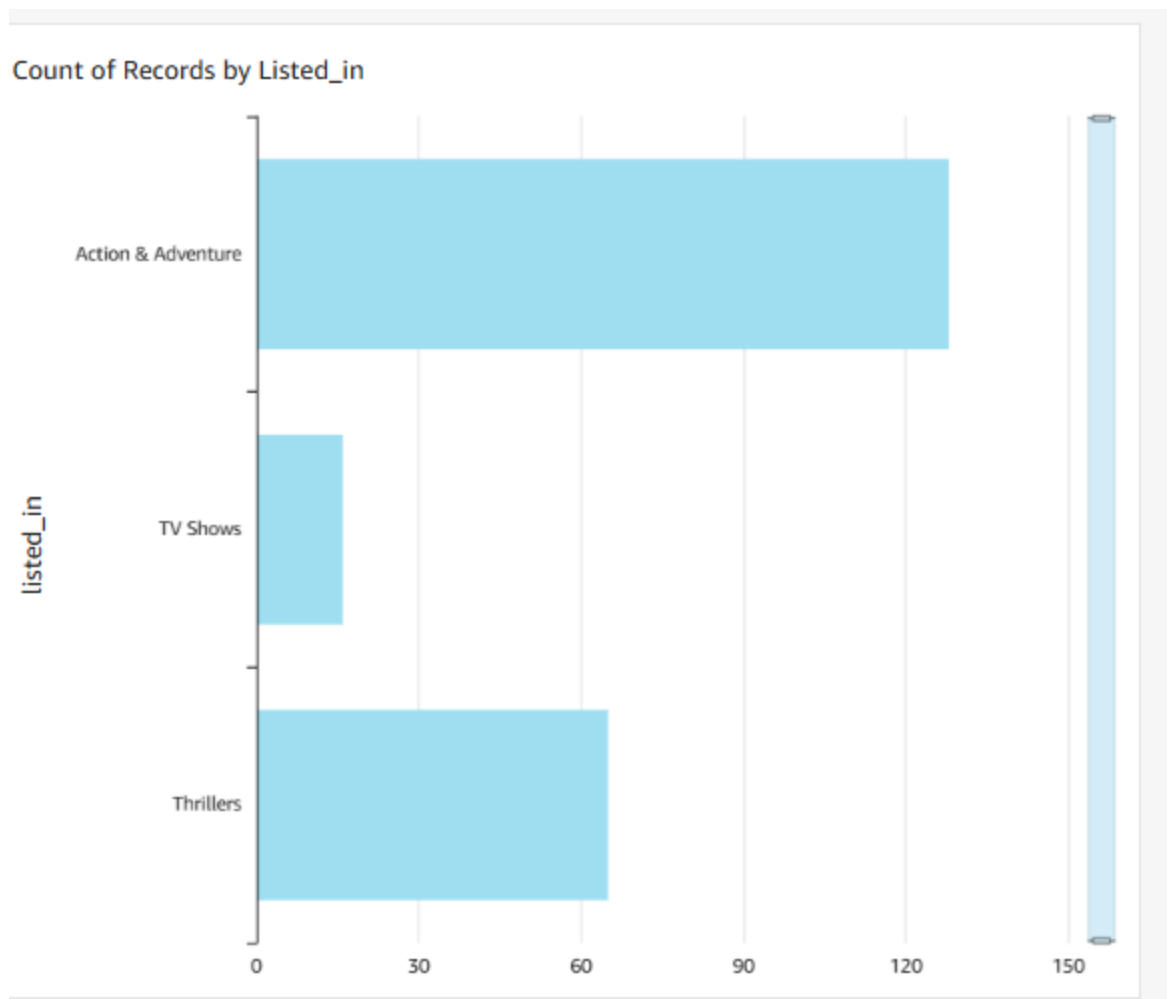
- **release_year** , **type** in group and **title** in values

Count of Title by Release_year and Type

release_y...	type	ti..
1,942	Movie	2
1,943	Movie	3
1,944	Movie	3
1,945	Movie	3
1,946	Movie	1
1,947	Movie	1
1,954	Movie	2
1,955	Movie	3
1,956	Movie	2
1,958	Movie	3
1,959	Movie	1
1,960	Movie	4
1,961	Movie	1
1,962	Movie	3
1,963	Movie	1
1,964	Movie	2
1,965	Movie	2
1,966	Movie	1
1,967	Movie	4
1,968	Movie	3
1,969	Movie	2
1,970	Movie	2
1,971	Movie	5
1,972	Movie	4
1.973	Movie	10

5.Of the TV shows and movies featured, how many were listed as 'Action & Adventure', 'TV Shows', or 'Thrillers'? For simplicity, ignore the TV shows and movies that have multiple listings/categories:

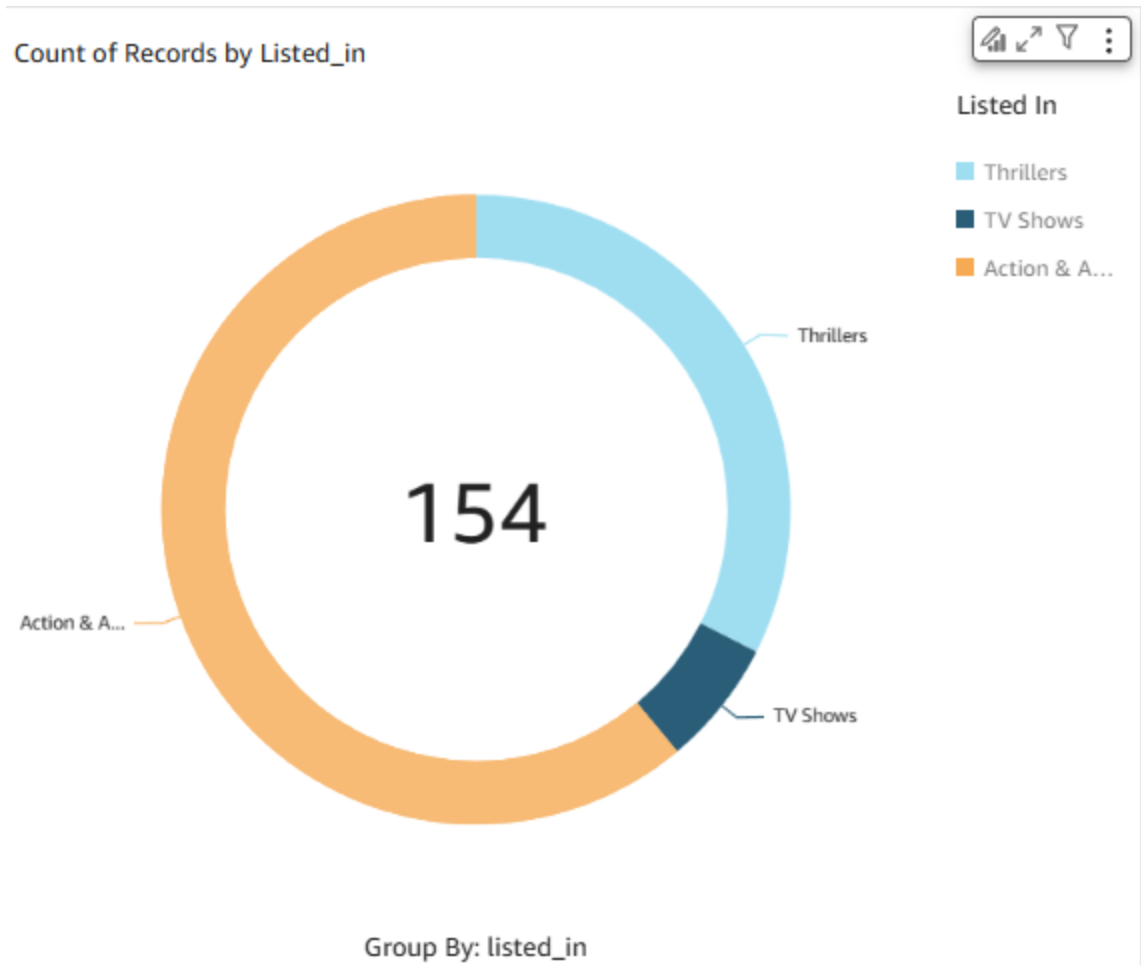
- Drag **listed_in** to a new space in Y Axis and then click of filter icon .
- Select **listed_in** for filtering and then check only the above mentioned types.
- Result looks like :



6.Of the TV shows and movies with the listing 'Action & Adventure', 'TV Shows', or 'Thrillers', list those that were released on 2010 or after:

- Duplicate previous chart

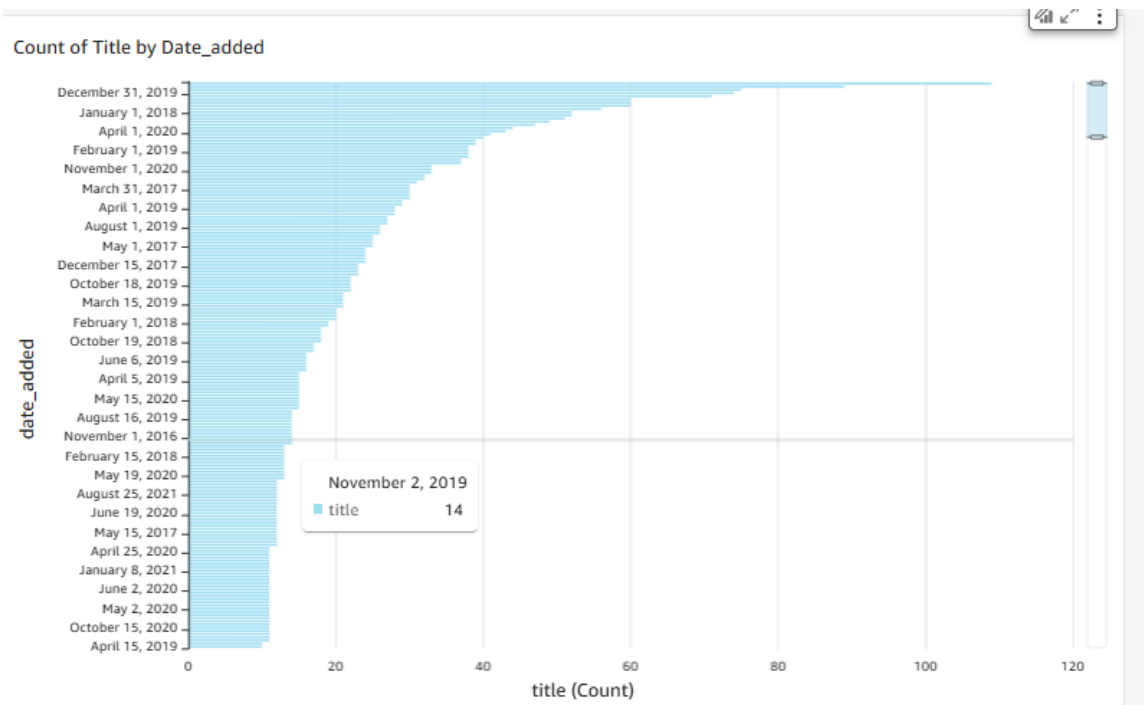
- Add new filter by choosing **release_year** and check the above mentioned years.
- Result:



(Donut Chart used for a change)

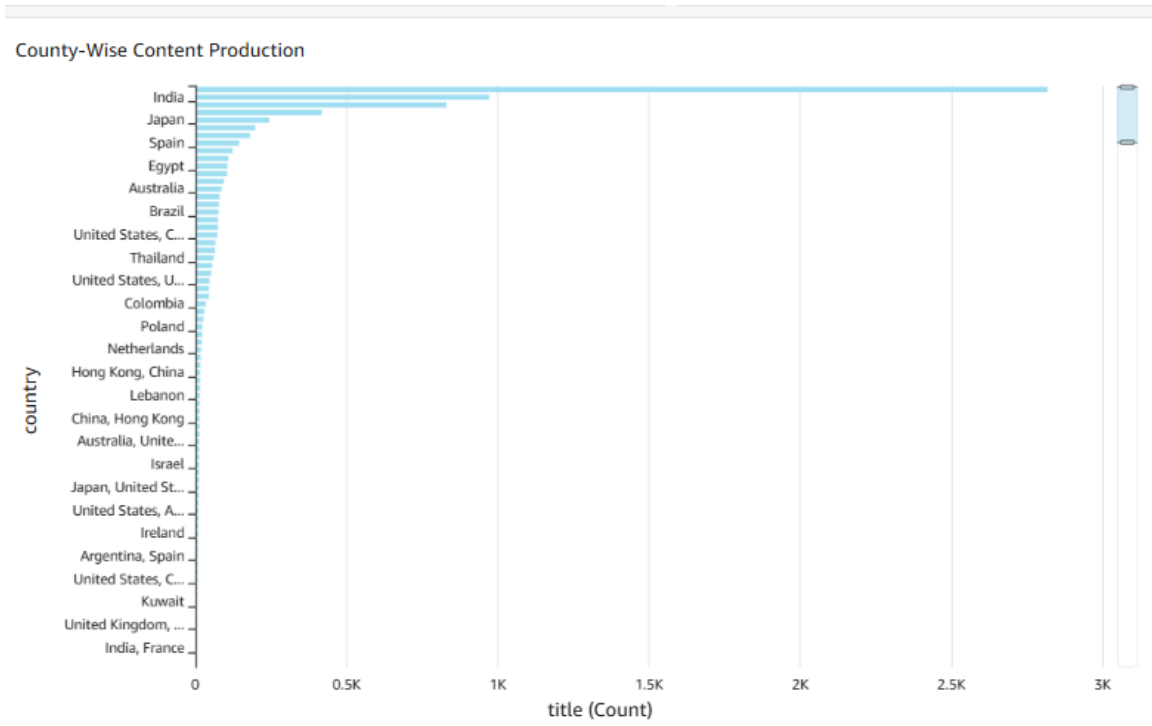
7. On what day did Netflix add the largest number of movies/TV shows to their catalogue?

- Drag **date_added** to new chart space in Y Axis and add **title** to Values
- Result:



8. Country-Wise Content Production:

- Horizontal Bar chart → **country** in Y axis, **title** in VALUES sort by descending
- Result:

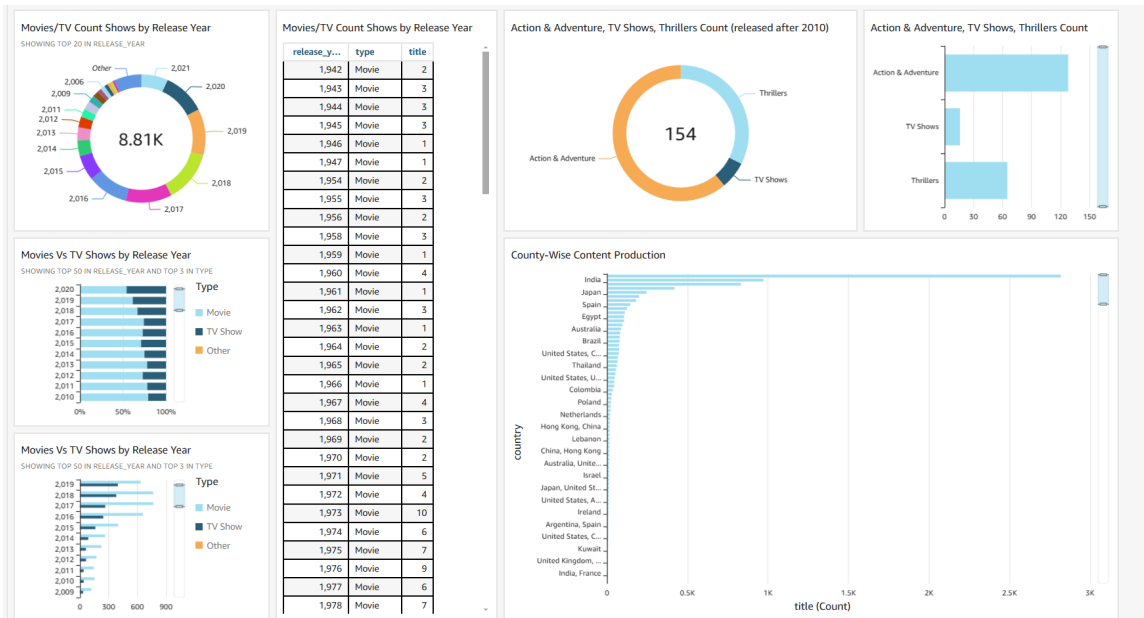


Finally change titles for charts accordingly !

Step 6:

- Click **Publish** and give a name for dashboard
- Click on Export to save as pdf!

The final output:



Final Step:

Terminate all services to avoid billing!

- Click on profile icon of quicksight
- Account settings→Manage→ uncheck and confirm delete
- Delete S3 Bucket.