📊 AWS QuickSight Data Visualization Project - Project Documentation

Hi there!
I've successfully completed a data visualization project using **Amazon QuickSight**, and this document is a complete walkthrough of everything I did. It covers how I set up the environment using IAM and S3, connected QuickSight, and used it to answer analytical questions. Screenshots of each step are included in the project repository for easy reference.

📘 What I Learned

While working on this project, I learned how to:

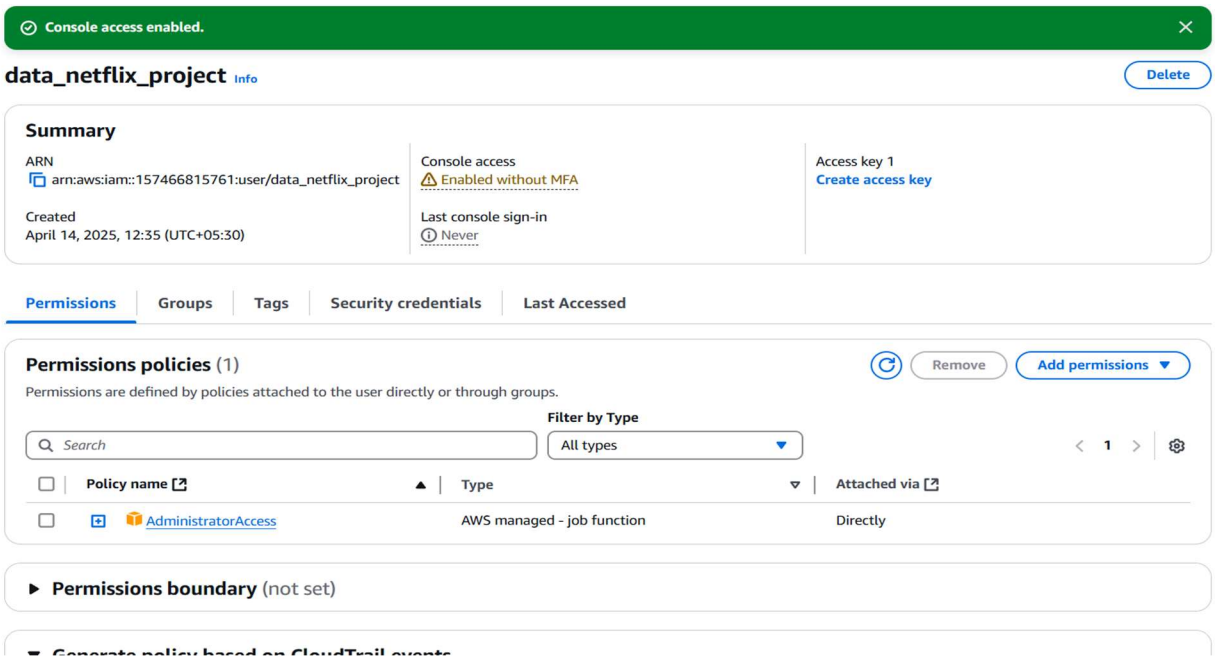
- Create and configure an **IAM user** with necessary permissions
- Set up an **S3 bucket** and upload data along with a manifest file
- Give QuickSight access to the S3 data source
- Build visuals and dashboards in **QuickSight**
- Use filtering, grouping, and calculations to draw insights from real-world data

---

🔐 IAM User and Permissions

I created an IAM user with **Administrator Access** for simplicity during the project setup. This user was used to:

- Create and manage the S3 bucket
- Upload the dataset and manifest file
- Grant QuickSight access to read data from the bucket

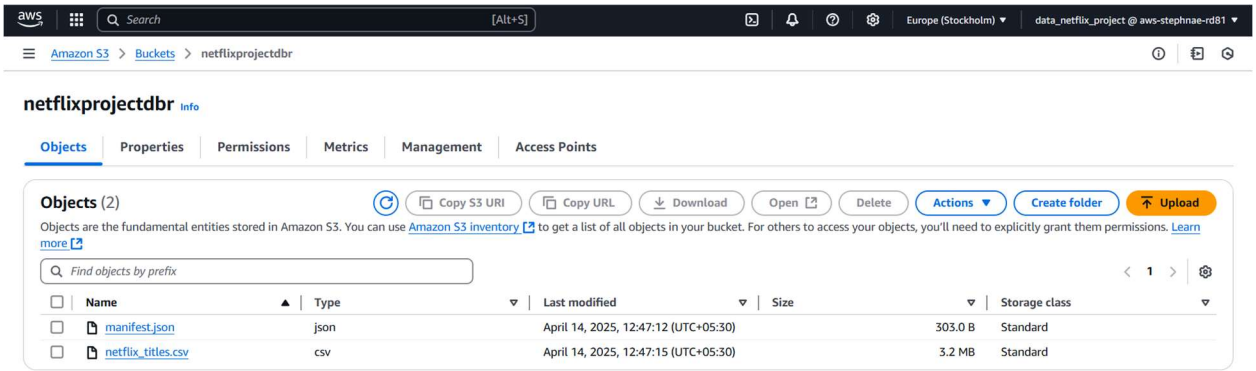


🗂️ S3 Bucket and Data Upload

Next, I created an **S3 bucket** and uploaded:

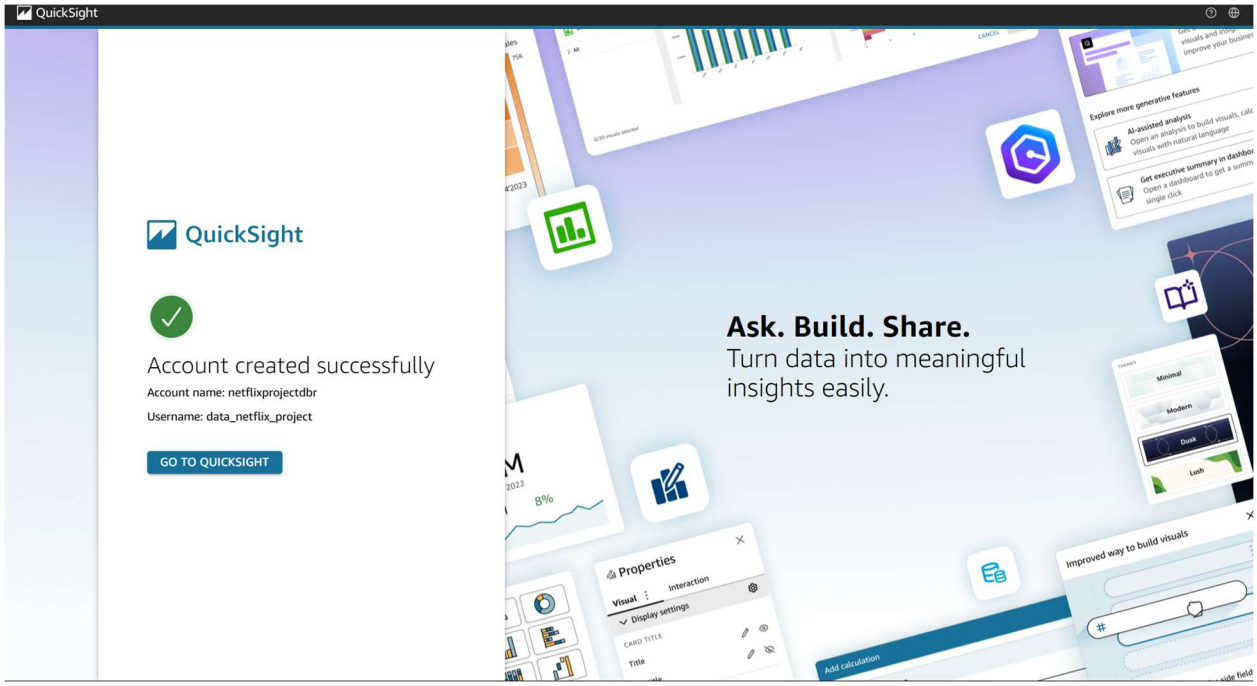
- A **JSON manifest file**
- The **Netflix dataset** containing information on movies and TV shows

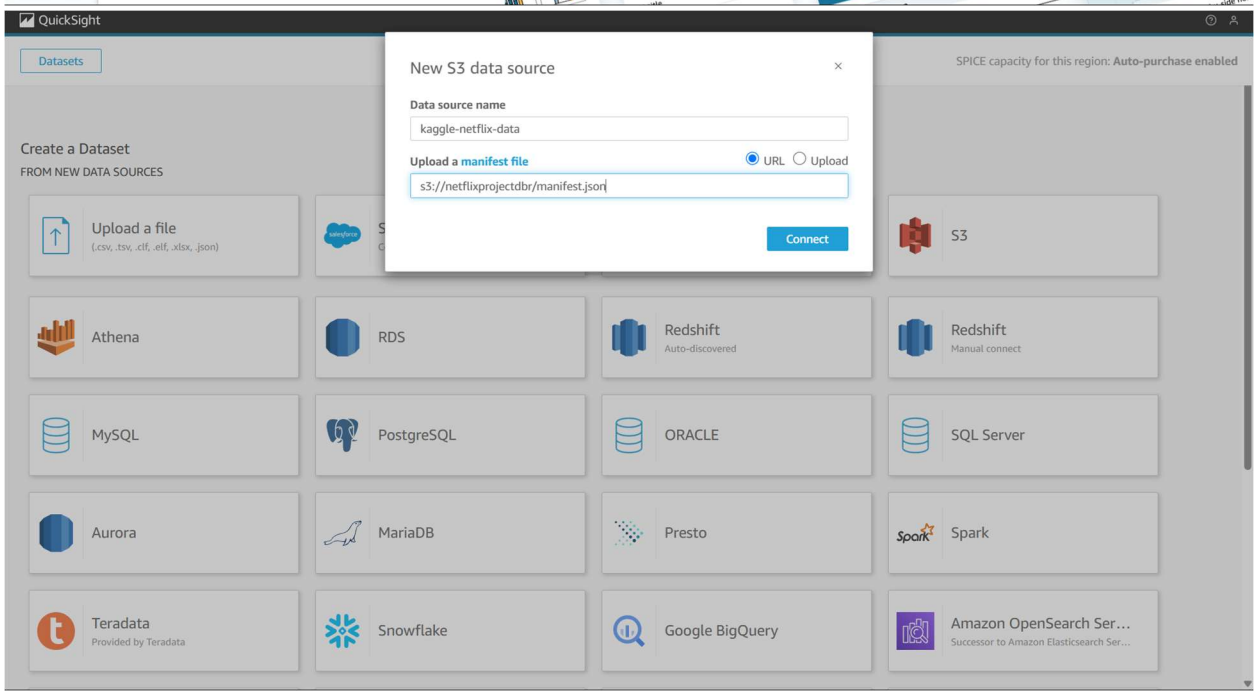
The manifest file was used to point QuickSight to the location of the dataset.



📊 QuickSight Setup

Then, I set up a **QuickSight account**, gave it access to the S3 bucket, and connected it to the dataset using the manifest file. After setting up the data source, I created visuals and dashboards to analyze the data.



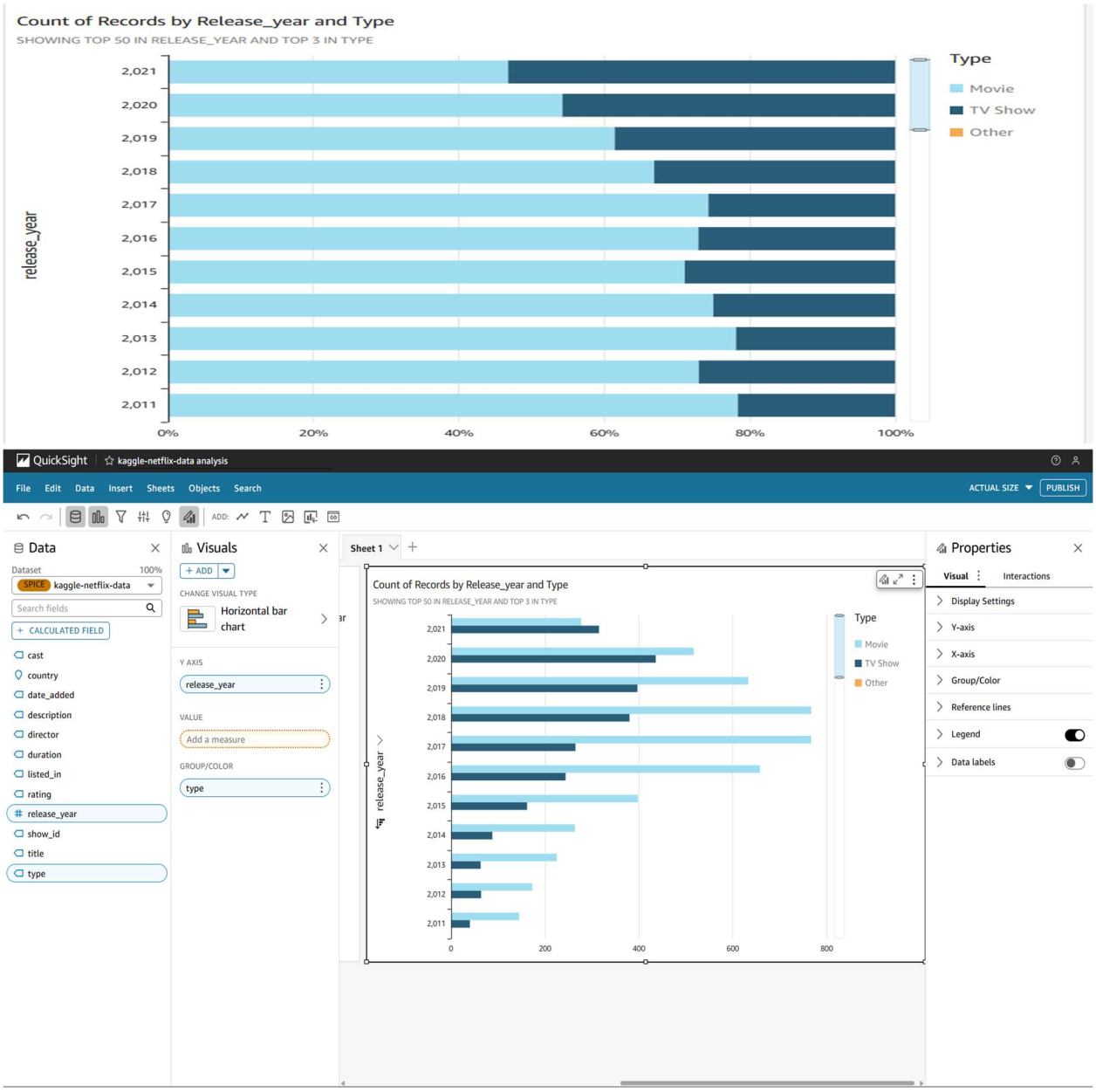


❓ Questions I Answered

These are the insights I extracted and visualized using QuickSight:

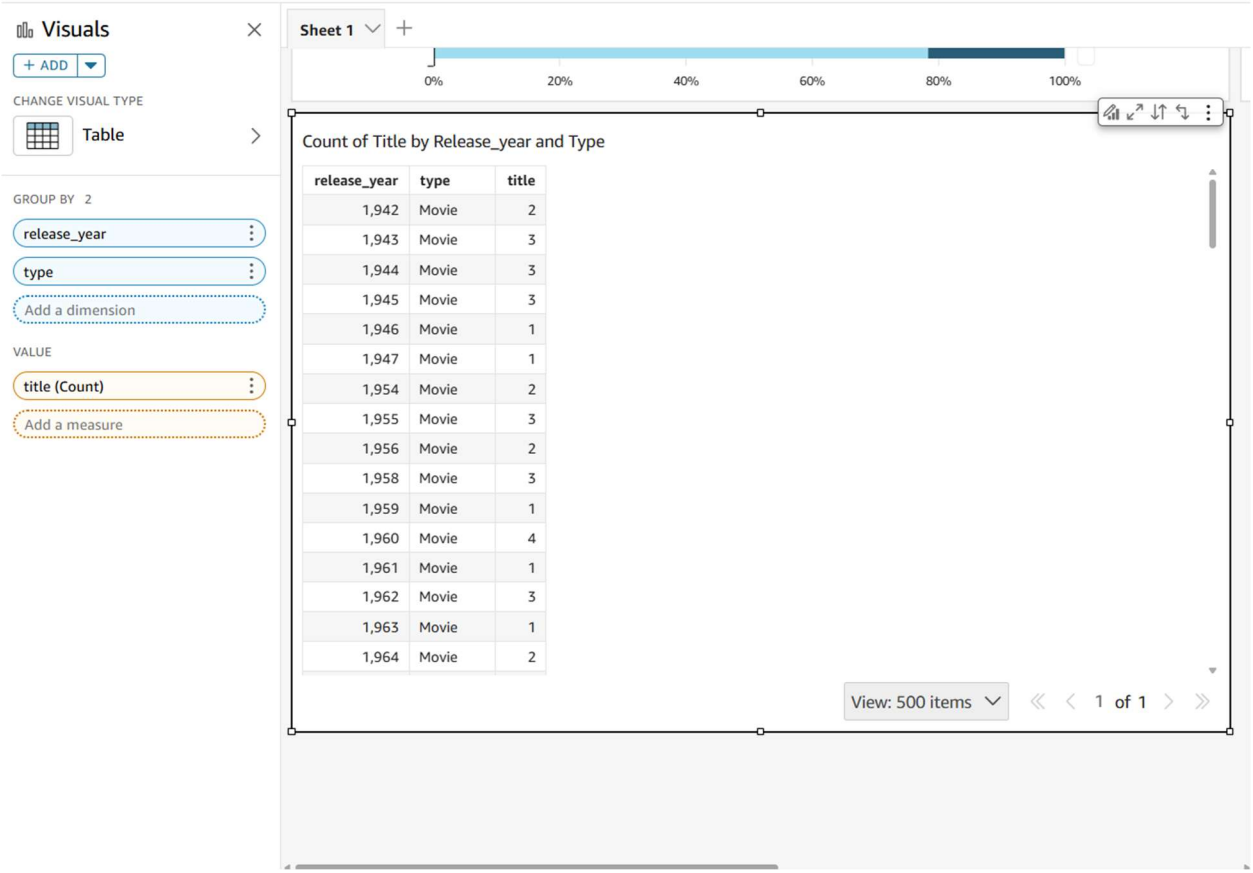
🔢 I created a stacked bar chart to visualize the percentage of movies and TV shows per release year.

I wanted to see the breakdown of content types per year and how they changed over time. I used a stacked bar chart and displayed the **% of movies vs TV shows** for each release year.



📋 I created a table to show the number of movies and TV shows per release year.

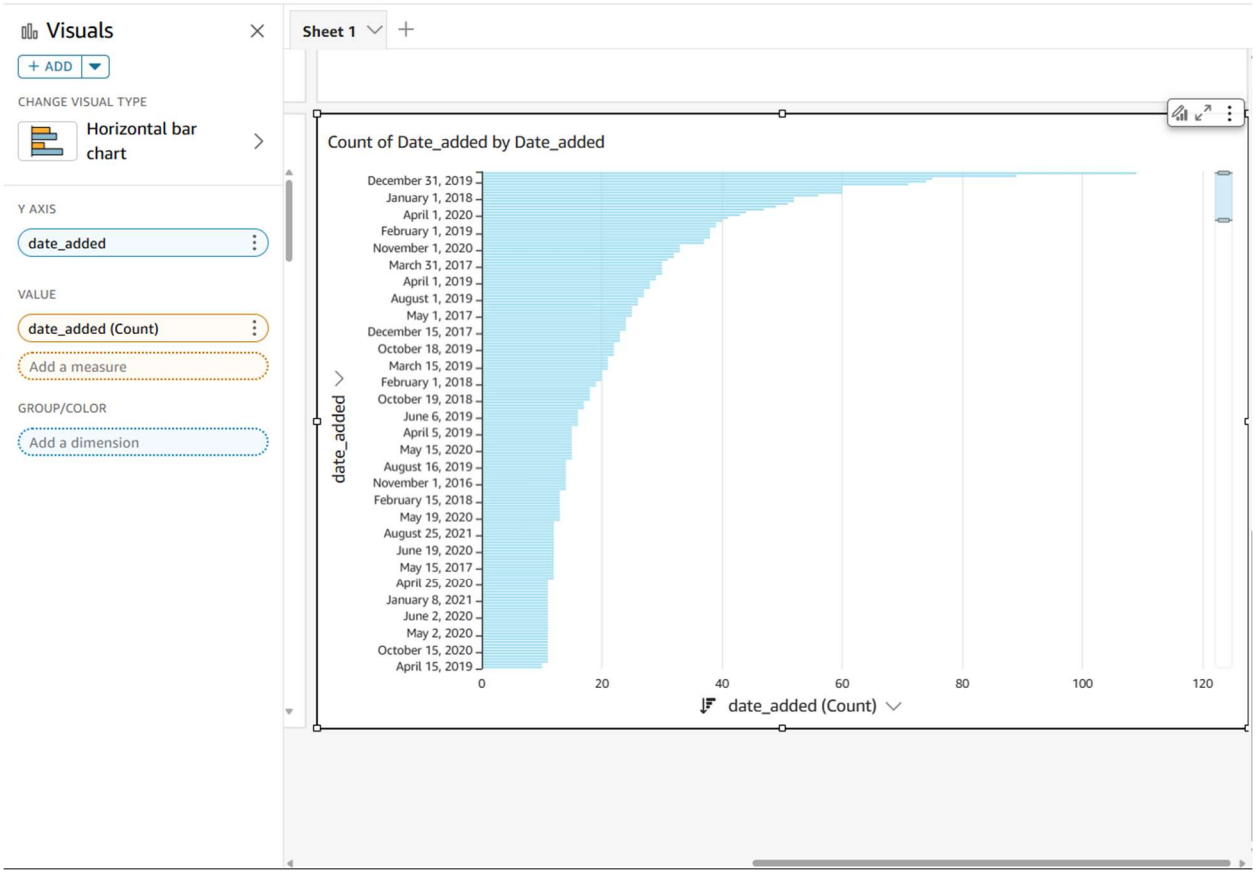
To go with the above visualization, I also wanted to see the actual numbers—how many movies and how many TV shows were released each year. I did this using a **pivot-style table** in QuickSight.



release_year	type	title
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

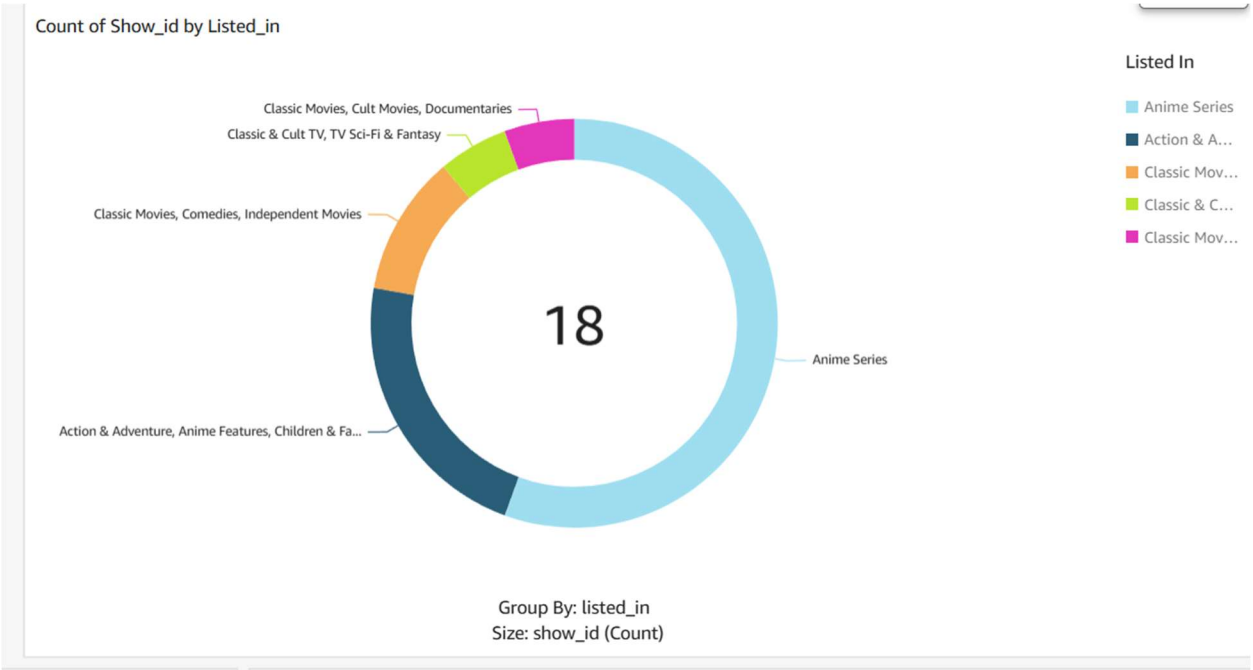
📅 I found the day Netflix added the highest number of movies or TV shows to their catalog.

By analyzing the "date_added" column, I was able to find the **single day** where the most content was added to the platform. I sorted the counts and used a bar chart to highlight that date.



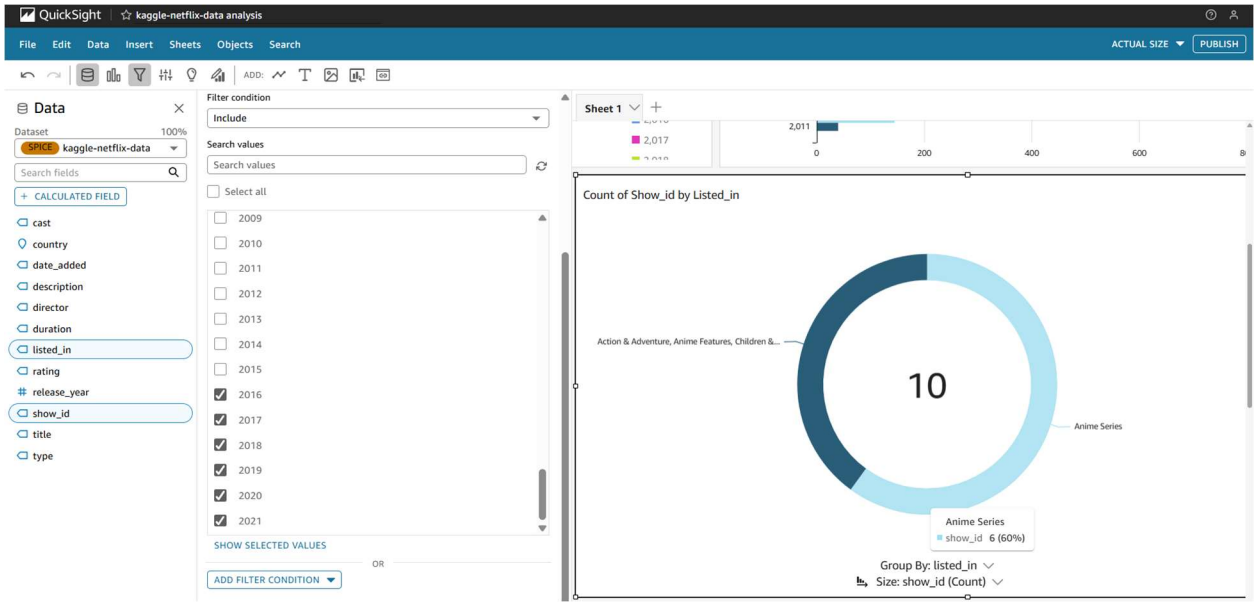
🎬 **I calculated how many items were listed as 'Action & Adventure', 'TV Comedies', or 'Thrillers'.**

I filtered the dataset to include only rows with a **single genre** and then counted how many entries were categorized under these specific genres. This gave me a clear picture of genre popularity.



📅 I filtered the above results to find how many of those titles were released in 2015 or later.

From the previous genre filter, I further refined it by filtering for **release_year ≥ 2015** to understand how many relevant titles were added in recent years.



💡 Conclusion

This project helped me understand how to:

- Use AWS services like IAM, S3, and QuickSight together
- Analyze data from scratch
- Answer real-world business questions using visuals

It's a hands-on example of how cloud tools can be used for **data analysis and storytelling**.

Feel free to explore the repo and screenshots!