

Visualize data with QuickSight

RA

raphaelmacadaeg0@gmail.com



Introducing Today's Project!

What is Amazon QuickSight?

Amazon QuickSight is a cloud-powered business intelligence (BI) service that enables users to visualize and analyze data quickly and efficiently. It provides an interactive and scalable solution for creating dashboards, reports, and visualizations without requiring deep technical expertise. QuickSight seamlessly integrates with AWS services like S3, RDS, and Redshift, as well as third-party data sources, making it highly versatile. Its AI-powered insights help users detect anomalies and uncover trends automatically, streamlining data analysis. Additionally, QuickSight supports easy collaboration by allowing users to share dashboards securely across teams. With a cost-effective, pay-per-session pricing model, businesses only pay for what they use, making it an ideal choice for organizations looking to make data-driven decisions efficiently.

How I used Amazon QuickSight in this project

In today's project, I used Amazon QuickSight to transform raw Netflix data into meaningful visual insights. I started by connecting my dataset from Amazon S3 and created various visualizations, including donut charts, bar charts, and tables, to analyze trends in movie and TV show releases. I utilized filters and formatting options to refine the data, making it more insightful and easier to interpret. Additionally, I explored advanced features like stacked bar charts to compare proportions and identified patterns in content additions. To ensure data completeness, I checked for missing values and considered best practices in data visualization. Finally, I structured my dashboard neatly, applied a finishing touch, and exported it as a PDF for easy sharing.

One thing I didn't expect in this project was...

One thing I didn't expect in this project was how easily missing data could impact the accuracy of insights. When I visualized country data, I noticed blank rows, which made me realize that incomplete datasets could lead to misleading conclusions. This highlighted the importance of data validation before analysis. It was a great reminder that even with powerful tools like Amazon QuickSight, the quality of insights still depends on the quality of the data itself!

 RA

raphaelmacadaeg0@gm...

NextWork Student

NextWork.org

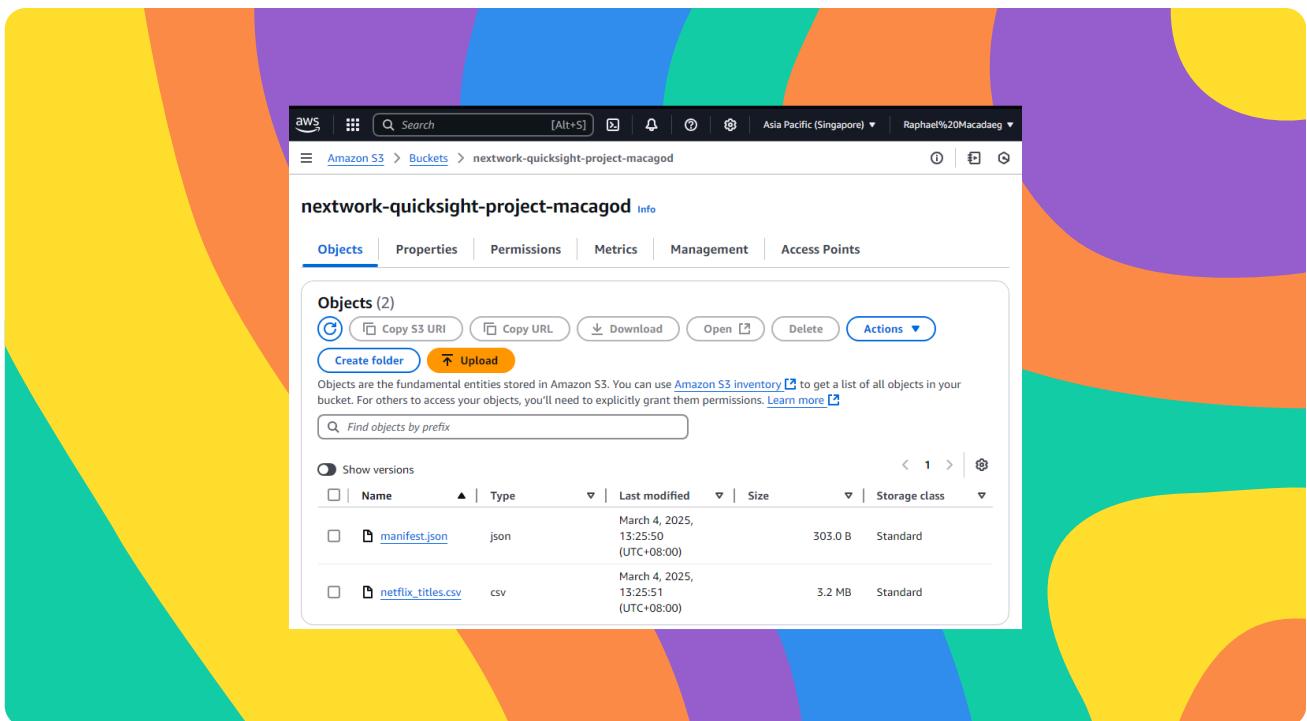
This project took me...

This project took me about 3-4 hours to complete. Most of the time was spent setting up visualizations, experimenting with different chart types, and ensuring the data was correctly represented. Refreshing the dataset and refining the dashboard layout also took some time, but overall, it was a smooth and insightful experience.

Upload project files into S3

S3 is used in this project to store two files, which are netflix_titles.csv and manifest.json. The CSV file contains the dataset for analysis, while the manifest.json file helps QuickSight locate and interpret the dataset stored in S3.

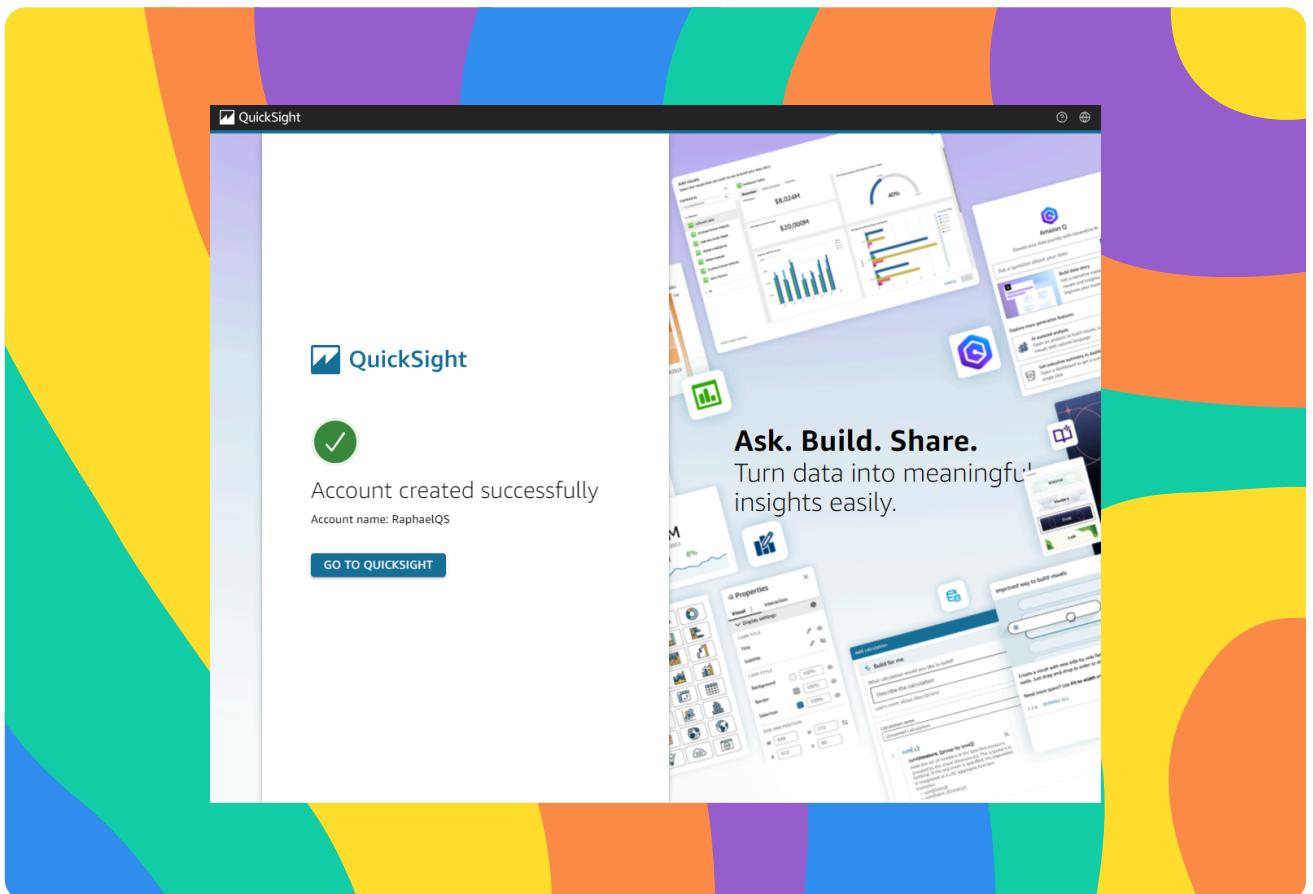
I edited the manifest.json file by replacing the URL with the S3 URL of my netflix_titles.csv file. It's important to edit this file because Amazon QuickSight uses it to locate and properly load the dataset from S3.



Create QuickSight account

Creating a QuickSight account costs nothing initially because it comes with a free trial of the Enterprise edition. However, to avoid charges after the trial period, it's important to delete your QuickSight account once you're done with the project.

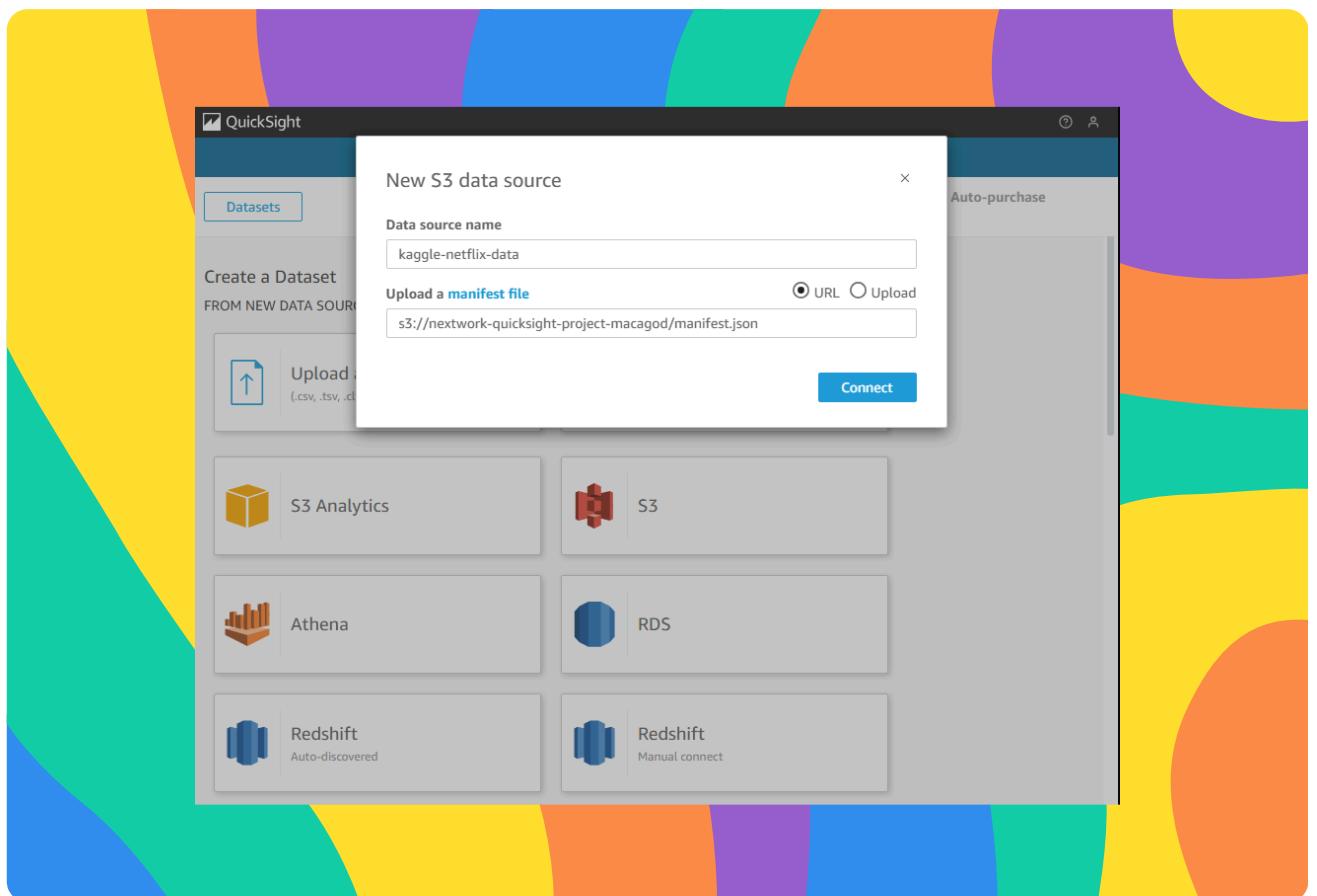
Creating an account took me a few minutes. The process was straightforward.



Download the Dataset

I connected the S3 bucket to QuickSight by visiting the Datasets page in the QuickSight menu and creating a new dataset using S3 data.

The manifest.json file was important in this step because it tells QuickSight where to find the dataset stored in S3. It acts as a map, specifying the exact S3 URL of the CSV file, so that QuickSight can properly locate and load the data for analysis

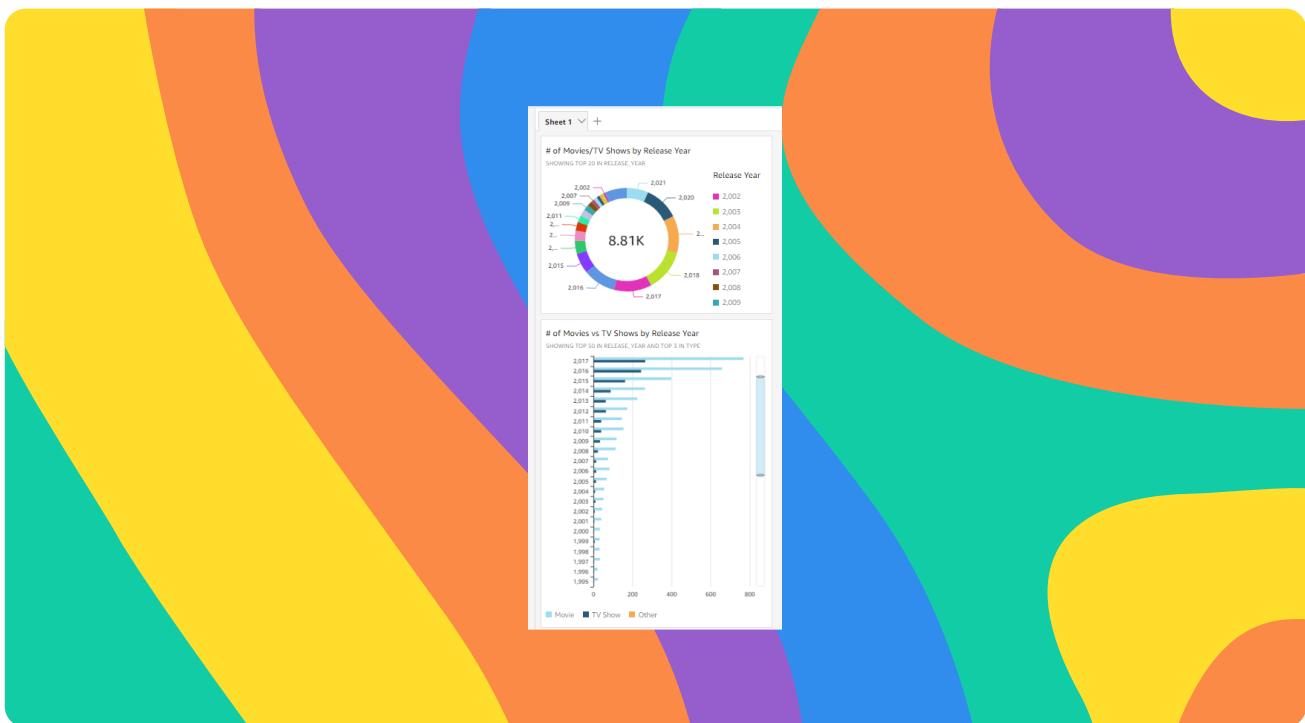


My first visualization

To create visualizations on QuickSight, I dragged the relevant dataset fields from the left-hand panel into the visualization area. First, I added the release_year field to the Y-axis and converted it into a donut chart to show the number of movies and TV shows by release year. Then, I created a horizontal bar chart comparing release_year with type to differentiate between movies and TV shows. I also resized the charts to ensure they fit well on the dashboard.

The chart/graph shown here is a breakdown of Netflix movies and TV shows by release year. The donut chart at the top represents the total number of movies and TV shows released each year, with different colors corresponding to different years. The number in the center (8.81K) likely represents the total count of titles analyzed. The horizontal bar chart below compares the number of movies vs. TV shows released over the years. The bars indicate how many titles were released in each year, with separate bars for movies and TV shows. This visualization helps identify trends, such as whether Netflix has been producing more content in recent years and how the balance between movies and TV shows has shifted over time.

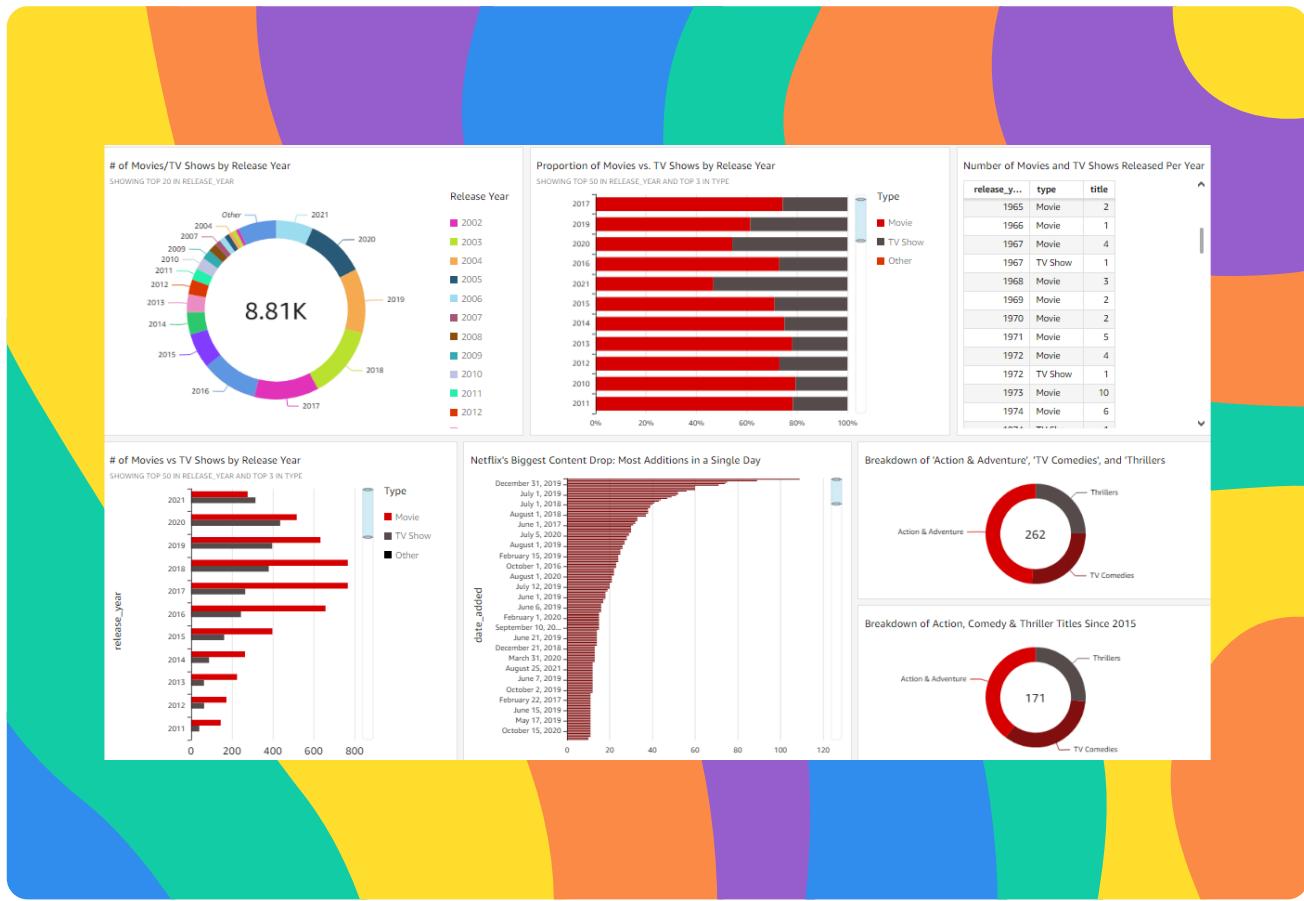
I created this graph by dragging and dropping the release_year field into the Y-Axis heading to represent the timeline of Netflix content. For the donut chart, I set release_year as the primary category to display the distribution of titles across different years. For the horizontal bar chart, I used release_year on the Y-axis and type (Movies/TV Shows) on the X-axis to compare the number of movies and TV shows released per year. I then adjusted the chart types by selecting a donut chart for the first visualization and a horizontal bar chart for the second, resizing them to fit neatly in the dashboard.



Using filters

Filters are useful for refining and narrowing down data to focus on specific insights. They help in eliminating irrelevant information, making it easier to analyze trends, compare categories, and extract meaningful conclusions from large datasets. In Amazon QuickSight, filters allow users to customize visualizations by selecting specific time periods, genres, or categories, ensuring more precise and relevant reporting.

This visualization breaks down Netflix's content trends. The donut chart shows the number of movies and TV shows by release year. The stacked bar chart compares the proportion of movies vs. TV shows over time. The table lists the number of movies and TV shows released per year. The horizontal bar chart further compares movie vs. TV show counts per year. The vertical bar chart highlights Netflix's biggest content drop days. The donut charts analyze the number of titles in 'Action & Adventure', 'TV Comedies', and 'Thrillers', with one chart filtering for titles from 2015 onward. I applied filters for content genre and release year to refine the analysis. These visualizations help in understanding content distribution patterns and trends over time.



Setting up a dashboard

As a finishing touch, I carefully reviewed the dashboard to ensure all visualizations were clear, accurate, and well-formatted. I refined chart titles to make them more intuitive, adjusted the layout for better readability, and ensured consistent color schemes for an aesthetically pleasing design. I also verified that all filters and sorting options were correctly applied, allowing users to interact with the data effortlessly. Finally, I double-checked the dataset integrity and confirmed that no additional updates were needed before publishing the dashboard.

Did you know you could export your dashboard as PDFs too? I did this by navigating to the Share option in Amazon QuickSight and selecting Export to PDF. From there, I adjusted the layout settings to ensure all visualizations fit neatly within the exported file. I also reviewed the preview to confirm that all charts, tables, and insights were clearly visible before finalizing the export. Once everything looked good, I downloaded the PDF and saved it for sharing and documentation purposes.

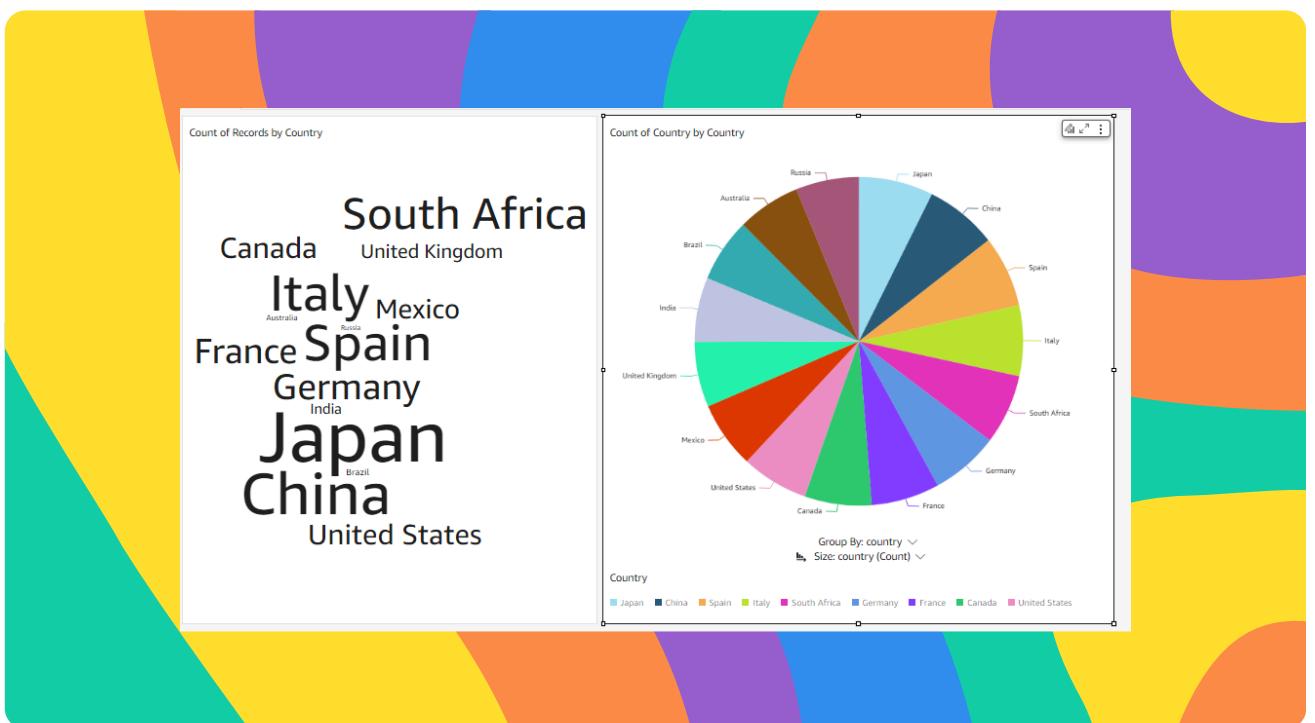


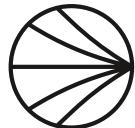
Refreshing source data

In this project's extension, I downloaded fresh data that's different from my original dataset because working with incomplete data can lead to misleading insights and flawed decision-making. Analyzing incomplete data brings the risk of drawing incorrect conclusions, especially when key attributes—such as country information—are missing. In this case, missing country data could result in misjudging where successful content is produced, which might lead Netflix to allocate resources inefficiently. By updating the dataset with complete country details, I ensured that my Amazon QuickSight visualizations were accurate, allowing for more reliable insights into content distribution and performance.

Once I downloaded new data, I had to update my S3 bucket because Amazon QuickSight pulls data directly from it, and using an outdated dataset would result in incomplete or inaccurate visualizations. I replaced the previous dataset with the new file to ensure all records, including the missing country data, were included. I also uploaded a new manifest.json file that correctly mapped the updated dataset, allowing QuickSight to recognize and process the new data structure.

I initially couldn't see my updated data in QuickSight, so I had to manually refresh the dataset. First, I refreshed my QuickSight tab and checked for the Import complete banner, confirming that the new data had been processed. However, my visualizations were still displaying outdated information due to cached data. To fix this, I navigated to the Dataset page, where I verified that the country data was updated. From there, I performed a manual refresh of the dataset to sync the latest changes with my analysis. After saving my changes and returning to the Analysis page, my visualizations successfully displayed the fresh data, with no more missing country values.





NextWork.org

Everyone should be in a job they love.

Check out nextwork.org for
more projects

