

## Introducing Project

In this project, I demonstrate how to use Amazon QuickSight to analyse a Netflix dataset and generate meaningful visualisations and insights. The goal of this project was to strengthen my understanding of cloud-based data services, particularly for analytical tasks.

By working through this exercise, I aimed to gain hands-on experience with tools commonly used in cloud data environments and develop a deeper understanding of the data analysis pipeline using AWS services.

---

### Tools and Concepts

The primary cloud services I utilised were Amazon S3 (Simple Storage Service) and Amazon QuickSight. Key concepts I became familiar with include:

- Using manifest.json files to define and configure dataset structures for QuickSight.
- Applying data visualisation techniques, such as generating charts and applying filters.
- Learning how to refresh data within QuickSight when the source data is updated.

---

### Project Reflection

This project took approximately 1.5 hours to complete. It offered a clear and practical introduction to working with AWS analytical tools. The experience was insightful and has encouraged me to explore more topics in cloud computing. My next area of focus will be cloud security, particularly using AWS Identity and Access Management (IAM).

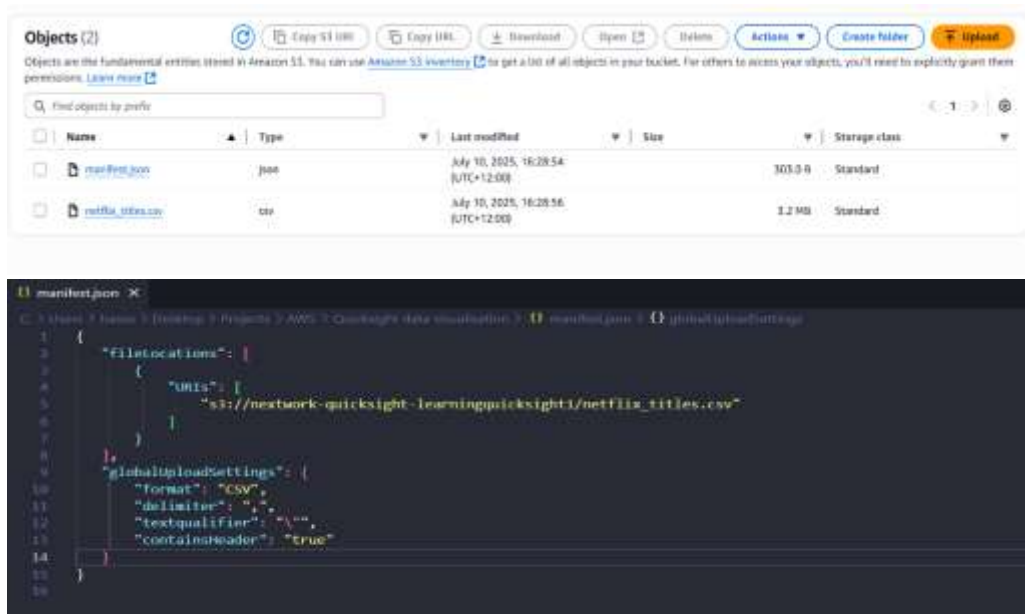
---

### Uploading Project Files into S3

To begin, I uploaded two essential files to my S3 bucket:

- netflix\_titles.csv – the raw dataset containing information on Netflix titles.
- manifest.json – a configuration file that informs QuickSight how to interpret the dataset.

The manifest.json file is crucial, as it tells QuickSight exactly how to process the incoming data file, including details such as the delimiter used to separate values, file type (CSV in this case), and the S3 URI from which QuickSight can retrieve the file. I edited this file by inserting the unique URI corresponding to my S3 bucket location.

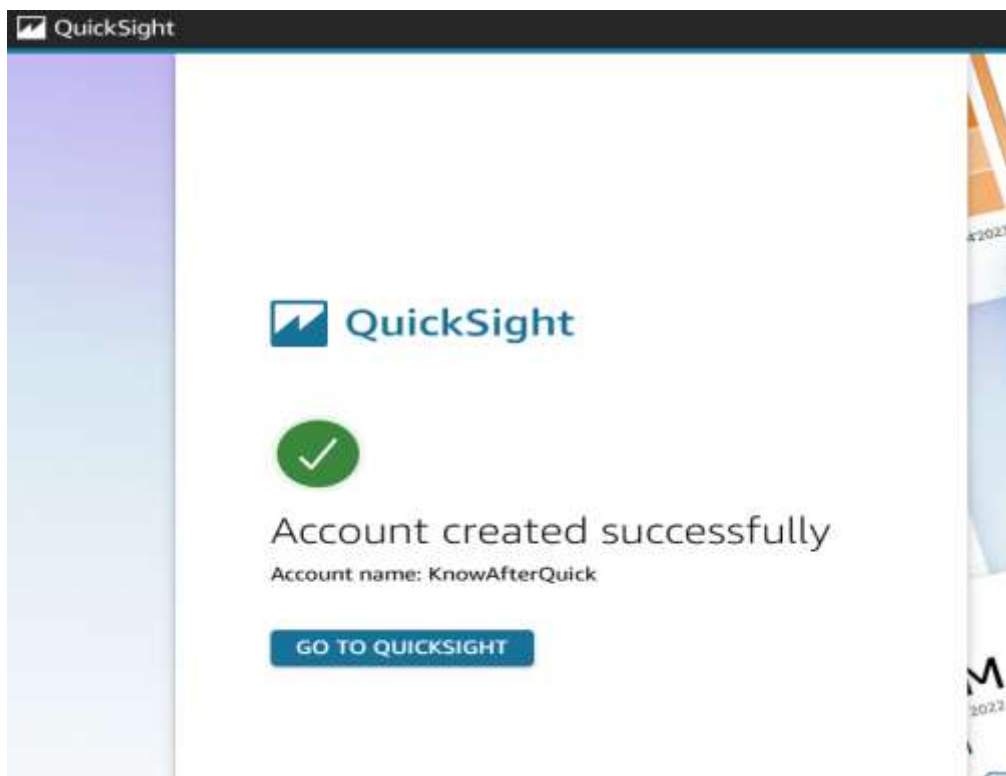


---

## Creating a QuickSight Account

Creating a QuickSight account was a quick and straightforward process. There was no cost involved, but it required attention to detail when configuring region settings and selecting the correct permissions.

Time taken: 2 minutes



## Downloading the Dataset

To connect QuickSight with my dataset stored in S3, I performed the following steps:

1. Navigated to the dataset creation page in QuickSight.
2. Selected S3 as the data source.
3. Named the dataset appropriately.



The screenshot shows a modal window titled "New Amazon S3 data source" with a close button (X) in the top right corner. Inside the modal, there is a section labeled "Data source name" with a text input field containing "kaggle-netflix-data". Below this, there is a section labeled "Upload a manifest file" with two radio buttons: "URL" (which is selected) and "Upload". Under the "URL" option, there is a text input field containing the S3 path "s3://nextwork-quicksight-learningquicksight1/manifest.json". At the bottom right of the modal is a blue button labeled "Connect".

---

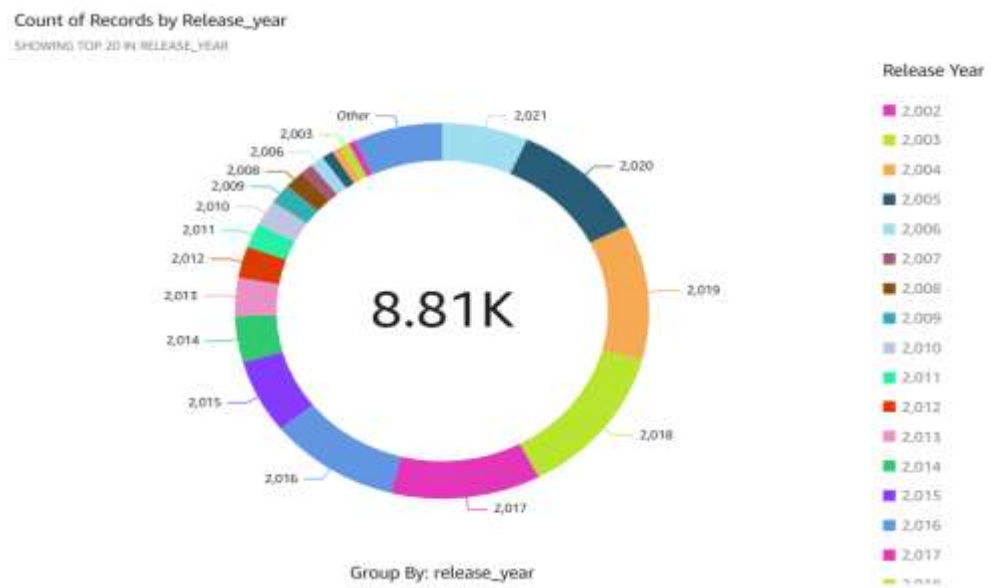
## My First Visualisation

My initial visualisation focused on analysing the release years of Netflix titles.

Using QuickSight's intuitive drag-and-drop interface, I selected the release year field and let QuickSight automatically generate a chart. I then changed the chart type to a donut chart for better visual clarity.

Insight:

The chart revealed a noticeable decline in releases post-2019, likely due to the global impact of the COVID-19 pandemic.



## Using Filters

Filters in QuickSight enable more targeted data analysis.

In this visualisation, I filtered the dataset to focus on the genres "TV Comedies", "Thrillers", and "Action & Adventure", and further narrowed it to titles released after 2015.

This allowed for a more refined view of content trends in specific genres during recent years.



## What Happens to Visualisations with Data Refresh

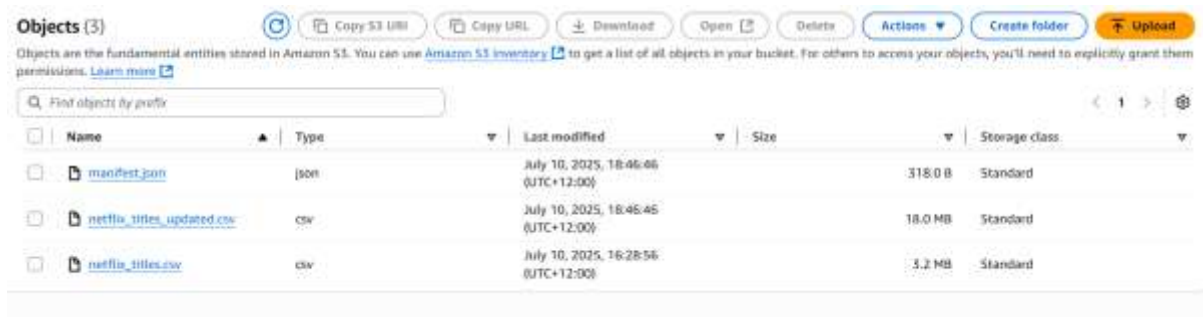
To simulate a scenario where source data changes, I updated the dataset as follows:

1. Filled missing entries in the 'country' column with "New Zealand (NZ)".
2. Saved the modified data as a new CSV file: Netflix\_titles\_update.csv.

3. Uploaded the updated file to a new folder in the same S3 bucket.
4. Updated the manifest.json file to reference the new data path.

Upon returning to QuickSight, I noticed that the visualisations had not updated automatically. This is because QuickSight caches data. To reflect the changes, I had to edit the data source in QuickSight and manually refresh the dataset.

This exercise was helpful in understanding how QuickSight handles data updates and reinforced the importance of refreshing datasets after modifying source files.



<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	<a href="#">manifest.json</a>	json	July 10, 2025, 18:46:06 (UTC+12:00)	318.0 B	Standard
<input type="checkbox"/>	<a href="#">netflix_titles_updated.csv</a>	csv	July 10, 2025, 18:46:45 (UTC+12:00)	18.0 MB	Standard
<input type="checkbox"/>	<a href="#">netflix_titles.csv</a>	csv	July 10, 2025, 18:28:56 (UTC+12:00)	3.2 MB	Standard

## Setting Up a Dashboard

As a final step, I assembled the various charts and visualisations into a cohesive dashboard within QuickSight. I adjusted layout sizes, updated titles for clarity, and ensured the visual elements aligned within a neat and structured box.

This final presentation offers a clear, interactive summary of insights derived from the Netflix dataset, and demonstrates the power of QuickSight as a business intelligence tool.

