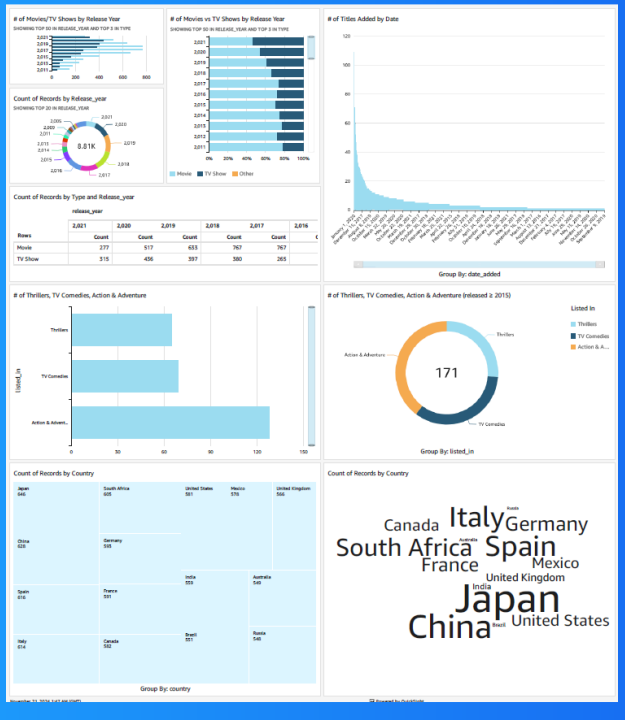
Visualize data with QuickSight

Unmilan Mukherjee



Introducing Today's Project!

What is Amazon QuickSight?

Amazon QuickSight is a tool provided by Amazon to easily use our datasets stored in AWS services and generate powerful and beautiful visualization from them. This is incredibly useful for Data Analysis of various datasets for companies and research.

How I used Amazon QuickSight in this project

I used Amazon QuickSight to create a dashboard of different tables/graphs and visuals to visualize Netflix's different data points from its dataset.

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

I did not expect how easy AWS QuickSight would make visualization of data. I previously did all this using Python and it was very messy, while this was simply dragging and dropping data and filtering by my needs.

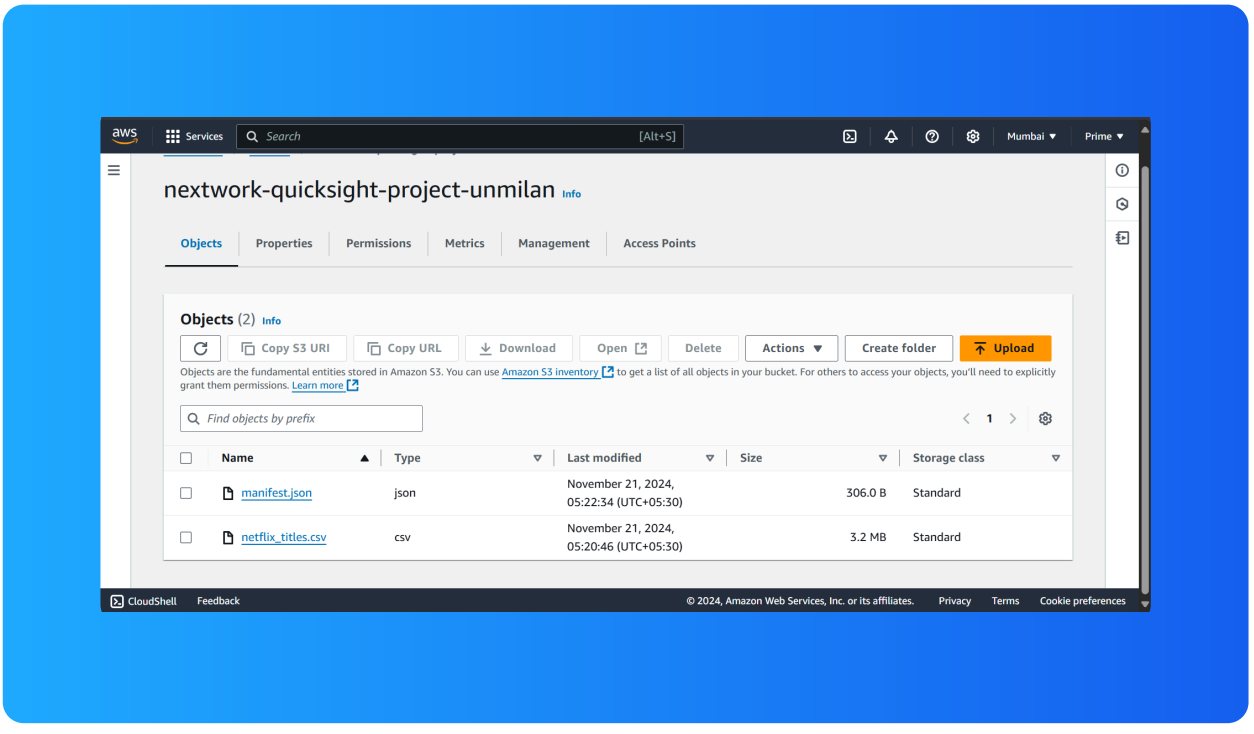
This project took me...

This project took me roughly 60 minutes to complete as it was my first time.

Upload project files into S3

S3 is used in this project to store two files, which are netflix_titles.csv and manifest.json.

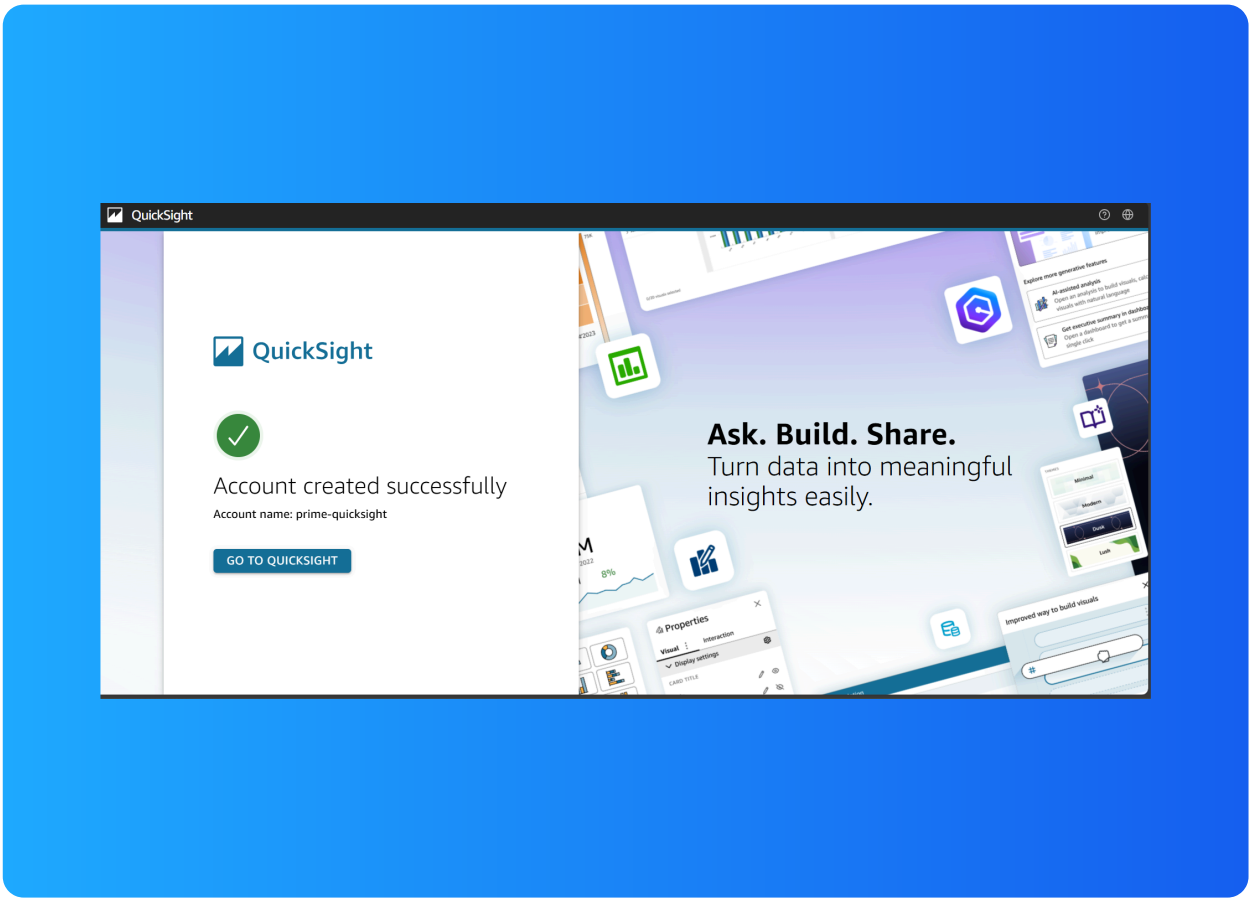
I edited the manifest.json file by changing the S3 URL in it to match the URL of the actual CSV that I uploaded to the bucket.



Create QuickSight account

Creating a QuickSight account cost me nothing for the first 30 days as it is a trial. But it will start billing you after that, so you have to delete your account if you are using it just for learning purposes.

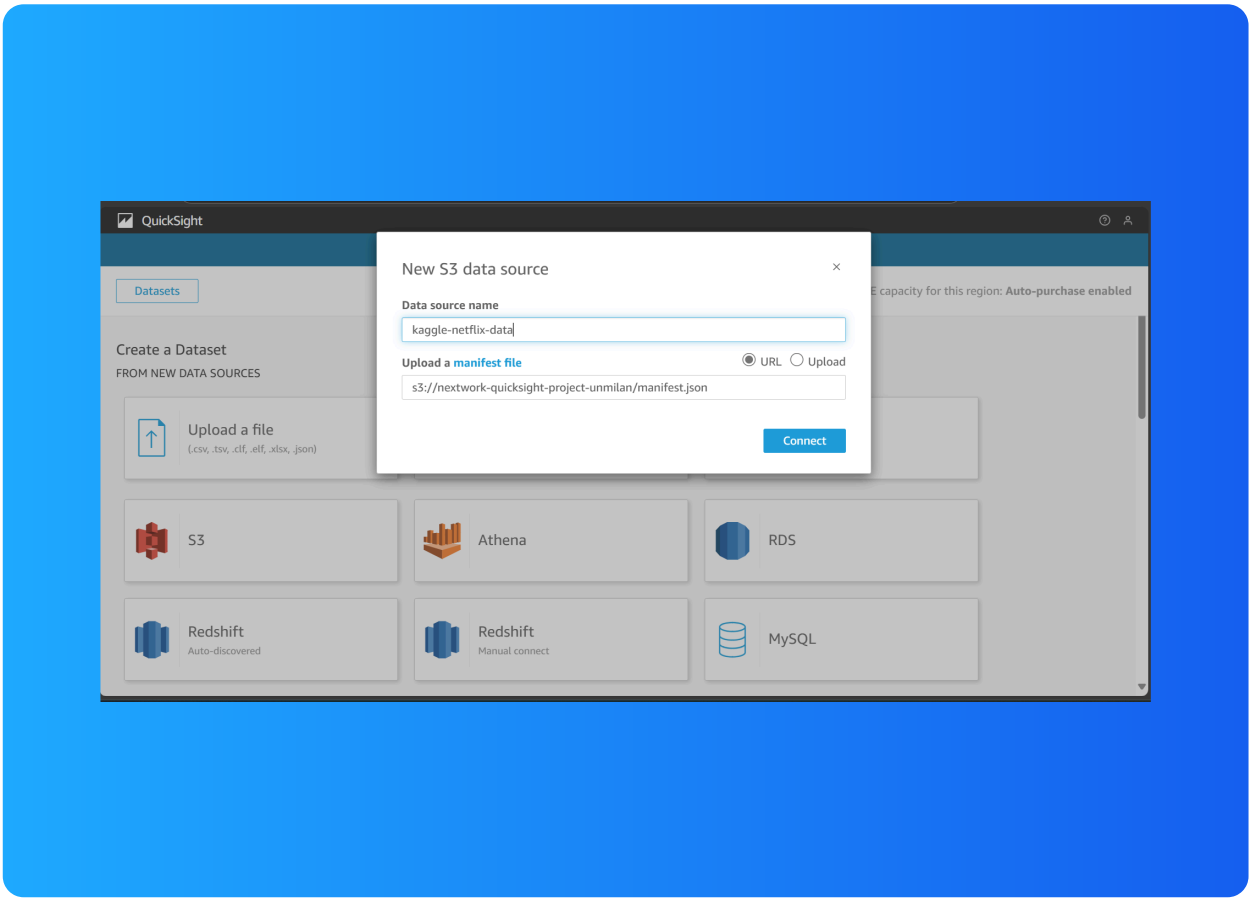
Creating an account took me a maximum of 10 minutes.



Download the Dataset

I connected the S3 bucket to QuickSight by visiting the datasets page in the Quicksight dashboard and clicking on New Dataset option in the top right corner.

The manifest.json file was important in this step because it is like a map for Quicksight that tells it where our data lives and how to read it. The manifest.json is required for QuickSight to properly display all the visuals from the dataset.

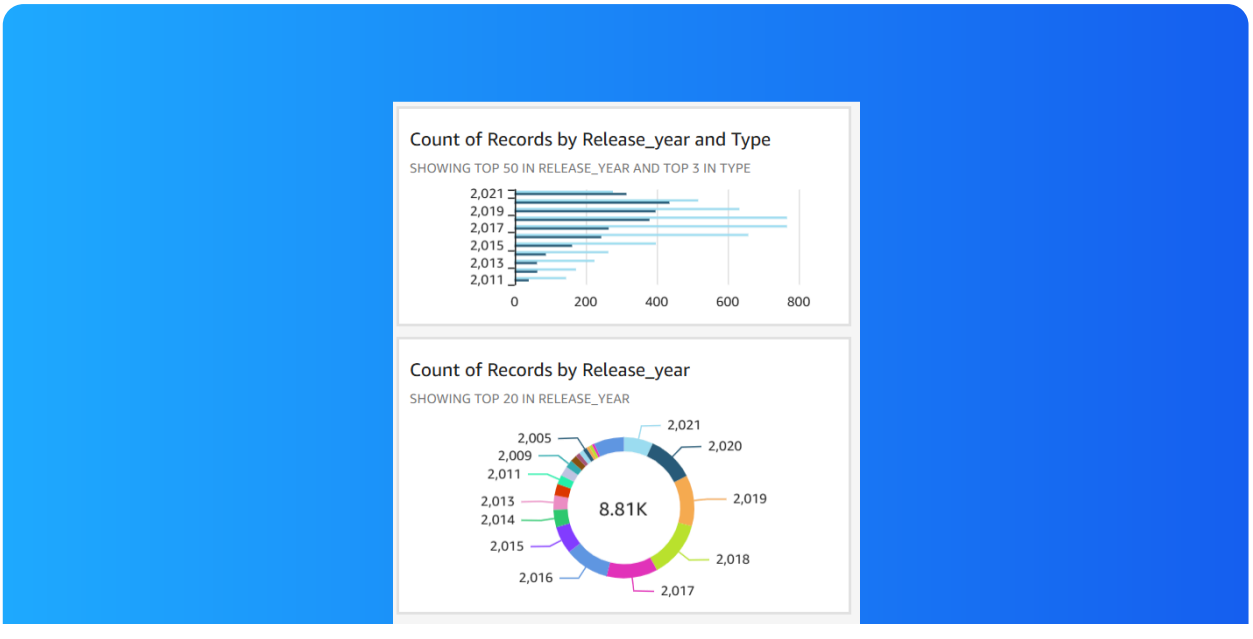


My first visualization

To create visualizations on QuickSight, I just had to drag one of my values detected from the CSV to the Sheet from the Visuals and voila! Our data is now beautifully presented, here we can add or remove different datas in different axes.

The charts/graphs shown here are breakdowns of the number of Netflix shows/movies that were released each year (as seen in the donut graph) and the top type of media(show/movie) released in each year (as seen in the horizontal bar graph).

I created this graph by dragging and dropping the release_year field into the donut chart and to create the horizontal bar chart I dragged and dropped the release_year in the y-axis and the type in the group/colour value.



Using filters

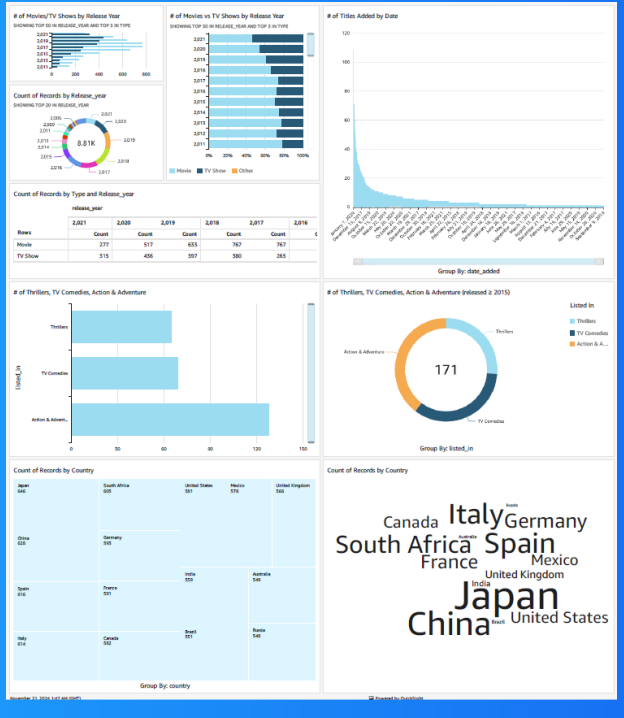
Filters are useful for sorting and organizing the data in our graphs in a meaningful or more precise way to get to the data points that matter to us the most. This is especially applicable when we have a huge amount of data and variables.

This visualization is a breakdown of TV shows and movies with the listing 'Action & Adventure', 'TV Comedies', or 'Thrillers', how many were released on 2015 or after. Here I added a filter by the listed_in and release_date variables.

Setting up a dashboard

As a finishing touch I took the country column from the dataset and visualized it as a beautiful Word Cloud.

Did you know you could export your dashboard as PDFs too? I did this by clicking the Publish button on the top right of the screen -> Give the dashboard a name -> Click export in the top right -> Click Generate PDF -> Wait a minute and download!



Refreshing source data

In this project's extension, I downloaded fresh data that's different from my original dataset because the original dataset had a missing country field. If our dataset has missing data or even worse entirely missing fields, it is detrimental to us.

I initially couldn't see my updates data in QuickSight, so I had to refresh the tab -> Go to the Datasets option -> Edit option on dataset -> Bottom right corner hit refresh -> Confirm -> Publish and Save.