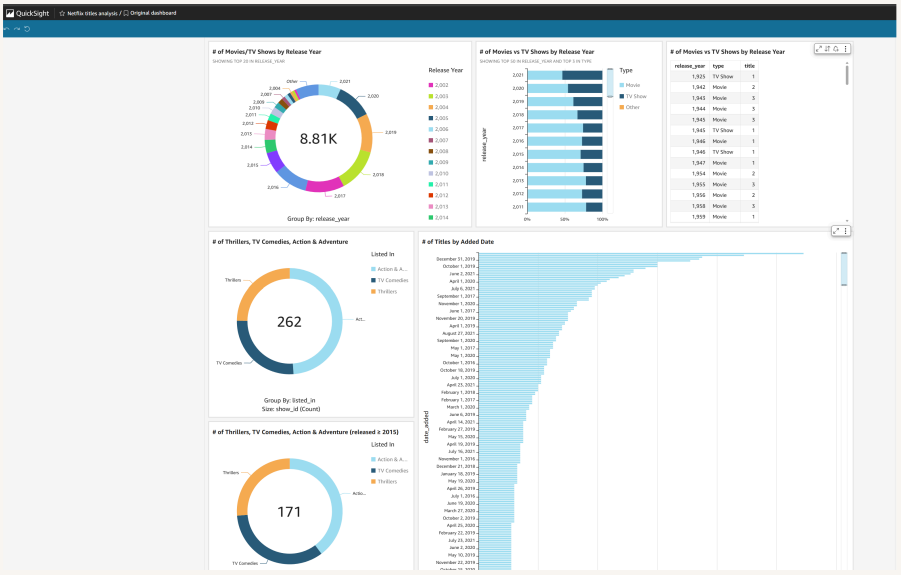
nextwork.org

Visualize data with QuickSight

GS gsweaponx@gmail.com



Introducing Today's Project!

In this project, I will demonstrate how to use Amazon QuickSight to analyse Netflux data and generate visulaizations and insights. I'm doing this project to learn cloud services, cloud data services for data analysis.

Tools and concepts

Services I used were QuickSight and S3. Key concepts I learnt include manifest.json, data visualization techniques and how to perform data refresh in QuickSight.

Project reflection

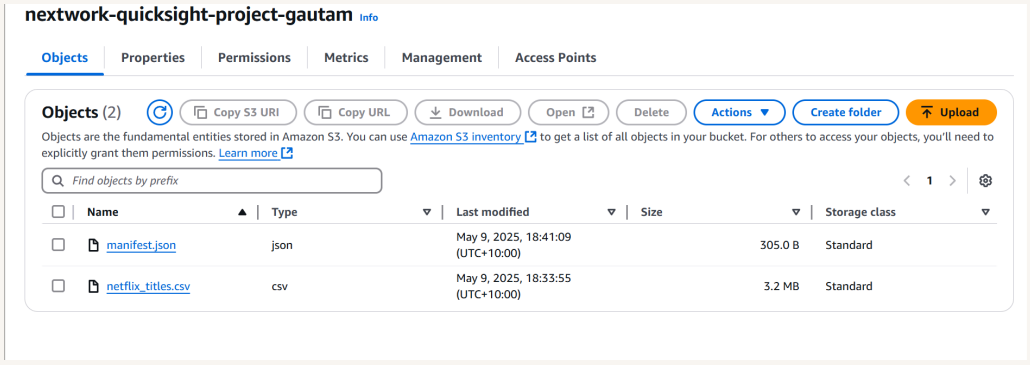
This project took me approximately 2 hours including demo time. It was most rewarding to generate a PDF of our finished visualization.

After this project, I plan to work on Day 3 of beginners AWS challenge which is about Cloud Security.

Upload project files into S3

S3 is used in this project to store two files, which are manifest.json(contains metadata about the data) and netflix_titles.csv (raw data we are analysing)

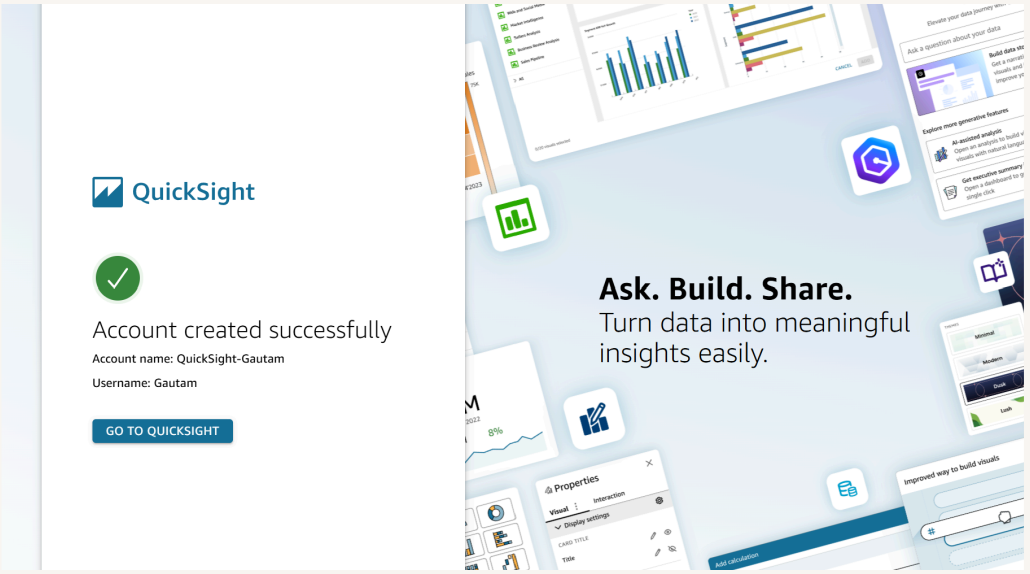
I edited the manifest.json file by updating the S3 uri that corresponds to our dataset's file location. It's important to edit this file because it is how QuickSight will find and analyse the data otherwise it wouldn't know the correct location.



Create QuickSight account

Creating a QuickSight account cost $0 as it comes with 30 days free trial. NOTE : Please uncheck add on at the bottom.

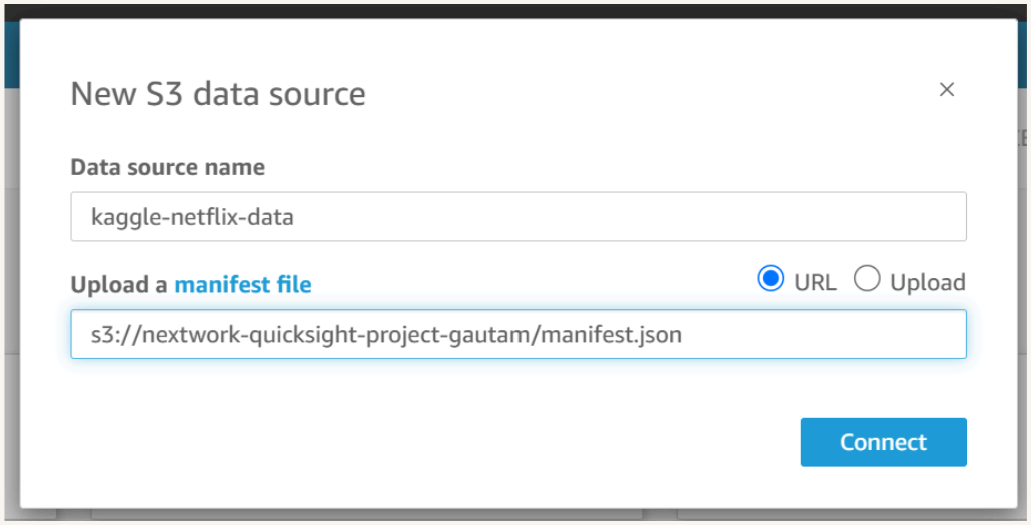
Creating an account took me around 5 minutes, including setting up my S3permission set up



Download the Dataset

I connected the S3 bucket to QuickSight by visiting Dataset tab and we selected S3 as our data source from the options.

The manifest.json file was important in this step because it tells QuickSight how to read the data. In this project it tells QuickSight we are uploading a .CSV file and the delimiter so that QuickSight knows how to break up the data.

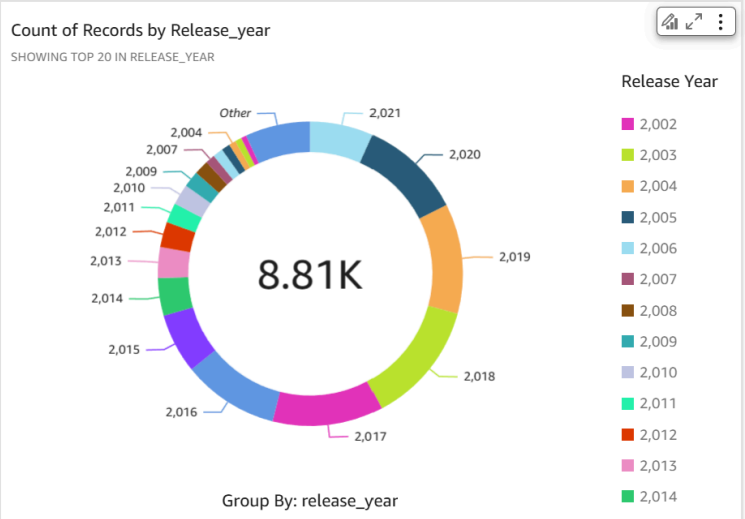


My first visualization

To create visualizations on QuickSight, I clicked on data fields/labels. And QuickSight automatically generate graphics to best suit that data, which you can also change according to your own preference.

The chart/graph shown here is a breakdown of the release years of the content on Netflix, i.e. how many TV shows or movies were release on each year. According to my analysis 2018 was the year with most amount of release.

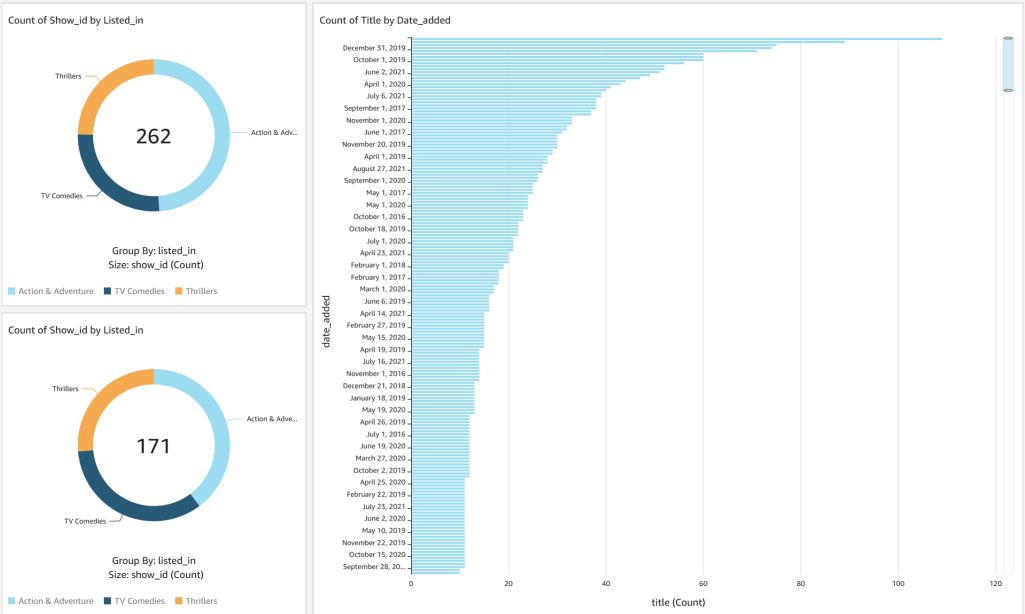
I created this graph by clicking on the release year data label and changing the automatically generated chart to a donut chart.


Count of Records by Release_year
SHOWING TOP 20 IN RELEASE_YEAR
Release Year
Other
2,021
2,020
2,019
2,018
2,017
2,016
2,015
2,014
2,013
2,012
2,011
2,010
2,009
2,007
2,004
8.81K
2,002
2,003
2,004
2,005
2,006
2,007
2,008
2,009
2,010
2,011
2,012
2,013
2,014
Group By: release_year

Using filters

Filters are useful for narrowing down our data to the subsets we want to focus on. Here the focus was on number of titles released in a year, count of titles by the type and titles released after 2015 for selected 3 types

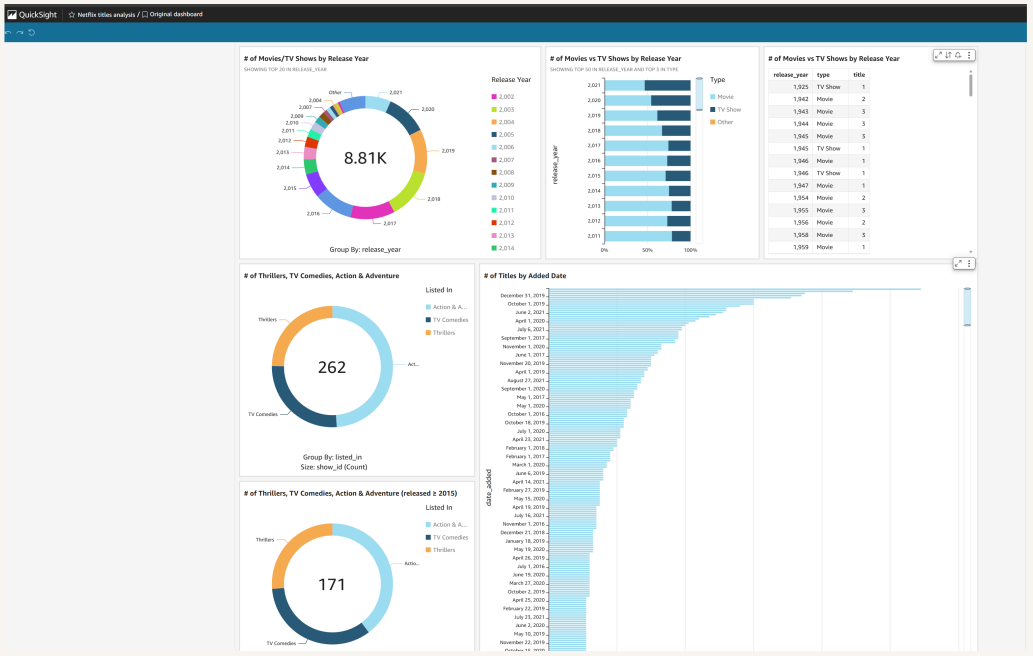
This visualization is a breakdown of TV Shows and movies that belong to one of 3 categories (Action & Adventure, Thriller and TV Comedies), the other one shows the number of titles for these 3 categories which were released after 2015.



Setting up a dashboard

As a finishing touch, I fixed the titles of the chart so that they are easliy readables as default names simply mentioned the data labels, not the purpose of the charts.

I published my dashboard by selecting Publish on top righht corner of the QuickSight window and then clicking "Generate PDF".



nextwork.org

The place to learn & showcase your skills

Check out nextwork.org for more projects