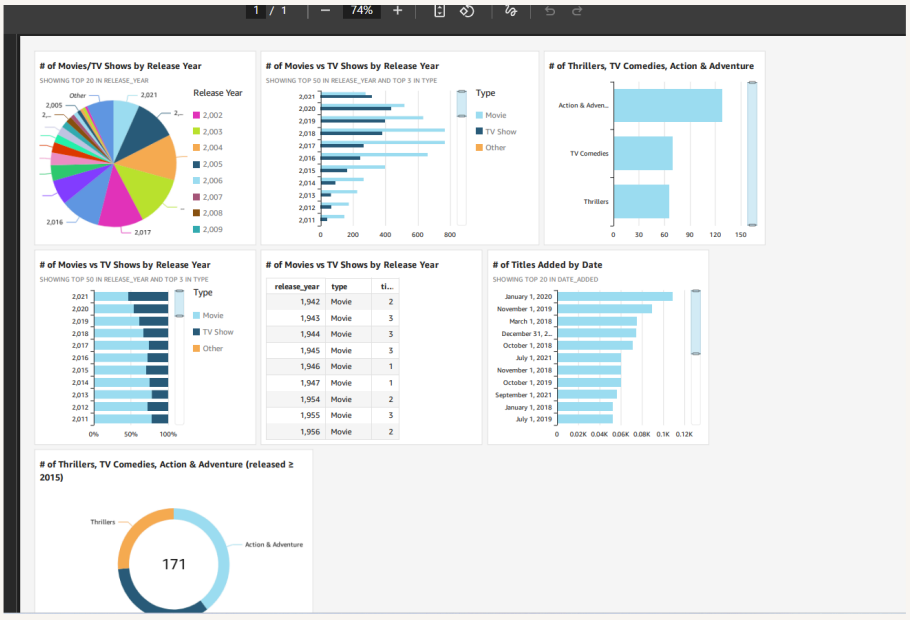
nextwork.org

Visualize data with QuickSight

shilpa kale



Introducing Today's Project!

In this project, we will demonstrate how to use Amazon QuickSight to analyze Netflix data and generate visualizations and insights! We are doing this project to learn how to use cloud data services for data analysis.

Tools and concepts

Services we used are Amazon S3, QuickSight. Key concepts learned manifest.json files, data visualiazation, filters.

Project reflection

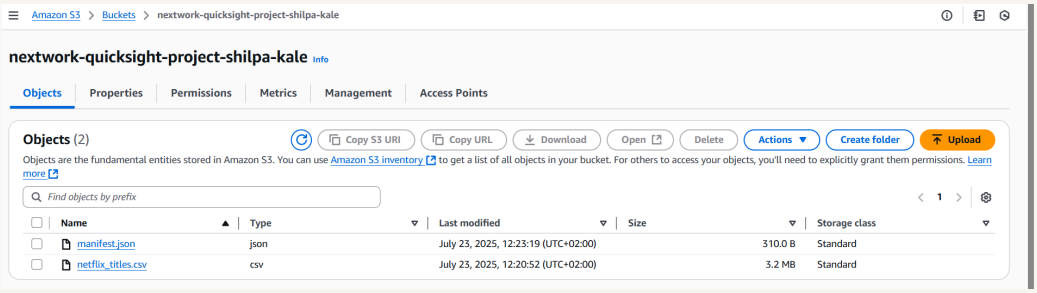
This project took me approximately 2 hours including demo parts. The most challenging part was understanding how manifest.json file works. It was most rewarding to see finished dashboard.

After this project, we plan to work on cloud security tomorrow.

Upload project files into S3

S3 is used in this project to store two files, which are manifest.json(which tells QuickSight about the structure and format of the data we are analysing) and the actual data set(csv file)

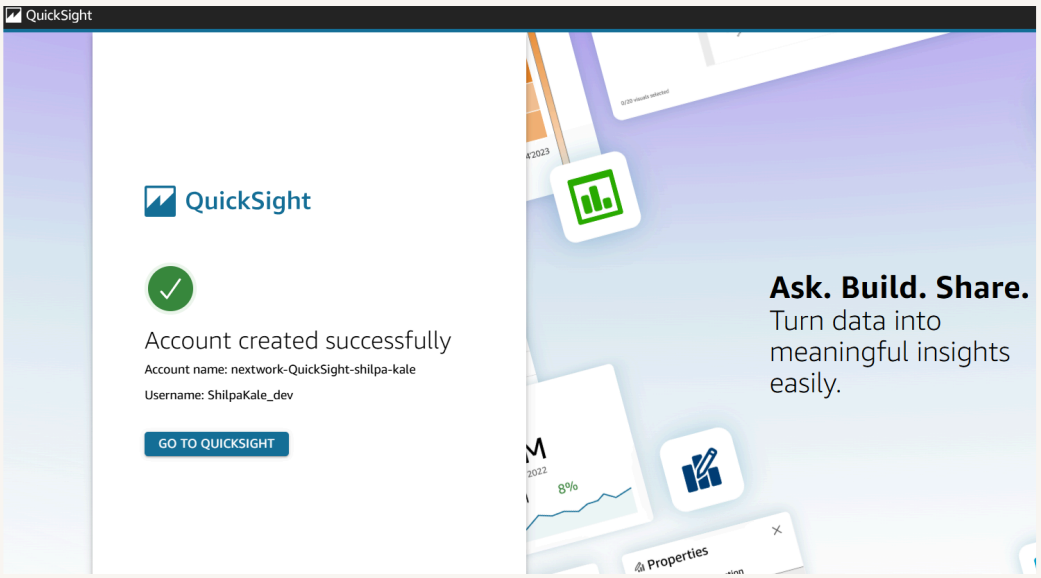
We edited the manifest.json file by updating the S3 URI that corresponds to our dataset file location. It's important to edit this file because its how QuickSight will find and analyse the data.



Create QuickSight account

Creating a QuickSight account cost $0 as it comes with a 30 day free trial! Make sure you UNCHECK an add-on in the sign up flow called Pixel-Perfect Report.

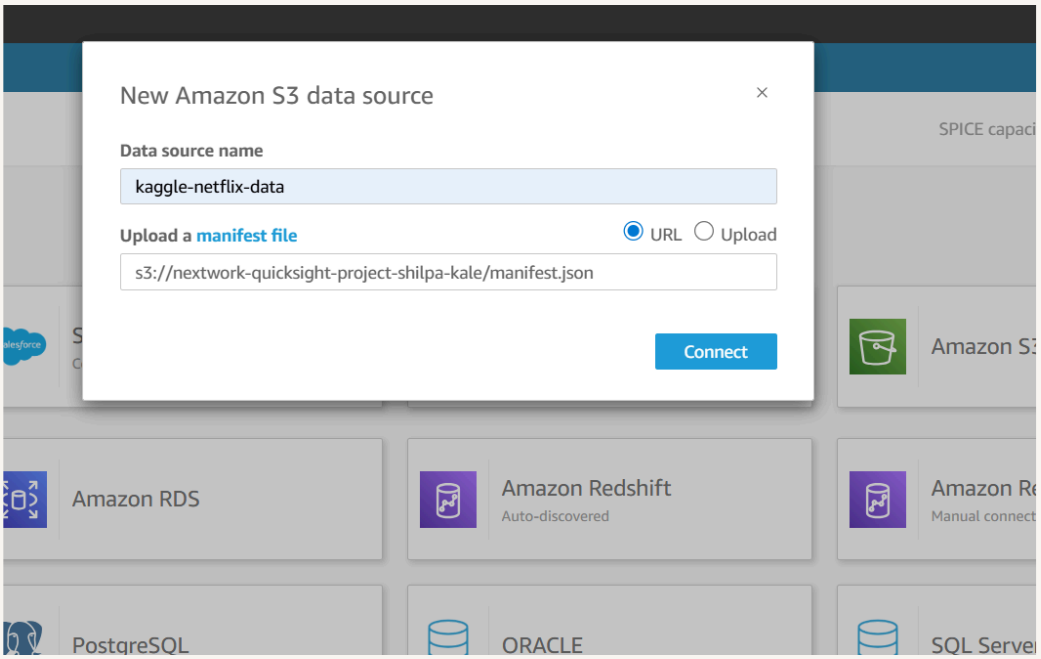
Creating an account took us about 5 mins including setting up our S3 bucket permissions.



Download the Dataset

We connected the S3 bucket to QuickSight by visiting the Datasets page. There were so many options for data sources we could connect to (databases, external tools/platforms like salesforce) and we selected S3.

The manifest.json file was important in this step because manifest.json tells QuickSight what your dataset looks like, so QuickSight knows how to understand the data and show it in charts or graphs. Without this map, QuickSight might get confused.

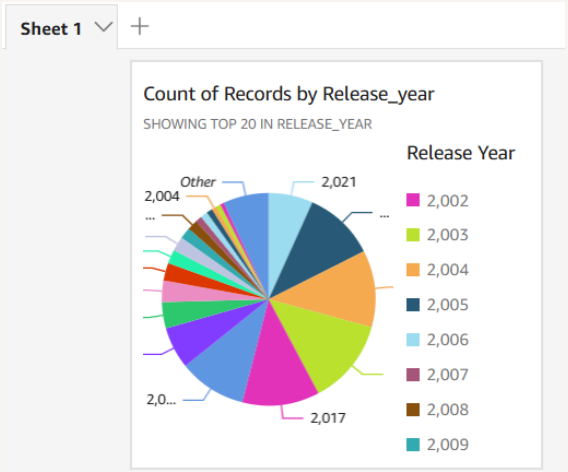


My first visualization

To create visualizations on QuickSight, we simply have to click on data fields/labels eg. release_year and QuickSight will automatically generate a graphic that best suits that type of data. You can also drag data labels into sections like group_by

The chart/graph shown here is a breakdown of release years of the content inside Netflix. i.e how many TV shows/movies were released on xyz year. You can see a total of 8800+ content pieces and year 2019 is year with the most of content released.

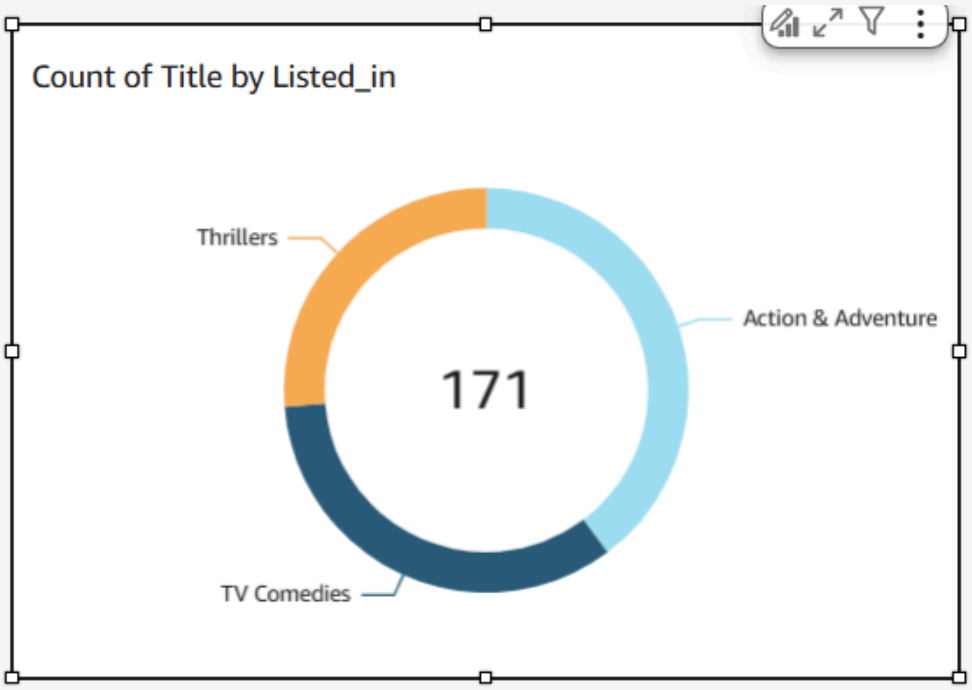
We created this graph by simply clicking on the release_year and changing the automatically geenerated chart from a bar chart to a donut chart.


Sheet 1
Count of Records by Release_year
SHOWING TOP 20 IN RELEASE_YEAR
Release Year
Other
2,004
...
2,021
...
2,0...
2,017
2,002
2,003
2,004
2,005
2,006
2,007
2,008
2,009

Using filters

Filters are useful for narrowing down our data for the subset that we want to focus on. We also use filters to only look at content which was released after 2015.

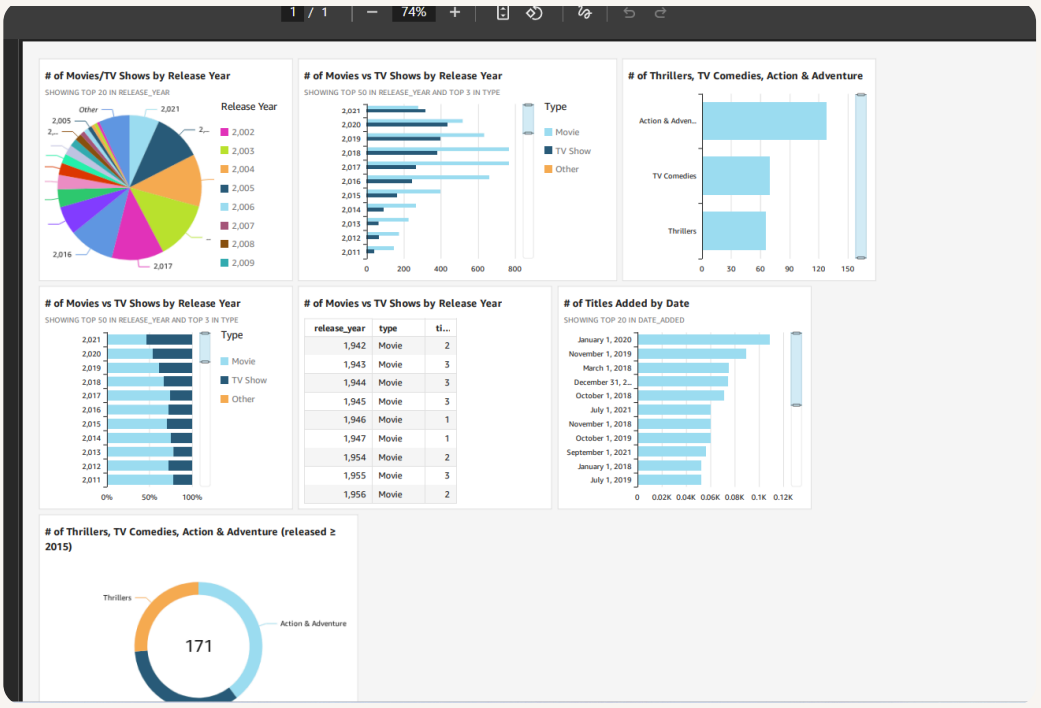
This visualization is a breakdown of TV shows and movies that belong in one of three categories, Action and Adventure, TV Comedies, Thrillers. Here we added a filter based on the listed_on label. i.e only these three categoreis could pass the filter.



Setting up a dashboard

As a finishing touch, we updated the titles of the chart so that they are easily readable. The new titles communicate the purpose of the chart clearly

Did you know you could export your dashboard as PDFs too? We did this by selecting publish and then generated pdf on the top right corner of our analysis



The place to learn & showcase your skills

Check out nextwork.org for more projects