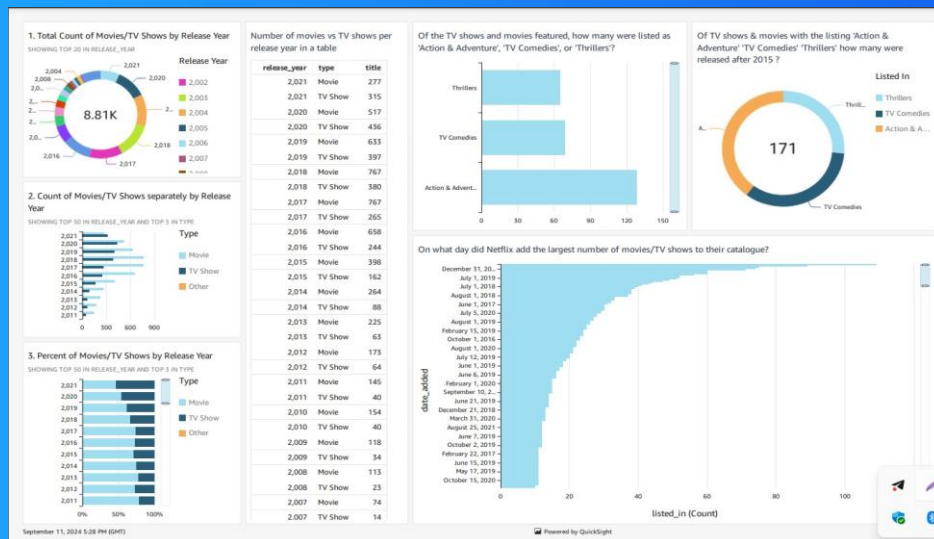


Visualize data with QuickSight



Rupam Patwari



Introducing Today's Project!

What is Amazon QuickSight?

Amazon QuickSight helps us to analyze data and create visualizations easily on huge datasets in a presentable format.

How I used Amazon QuickSight in this project

In this project, I analyzed a huge dataset of Netflix shows and movies to create a dashboard that extracts all the juicy insights with different filters.

This project took me...

90 mins

Upload project files into S3

S3 is used in this project to store two files, which are :

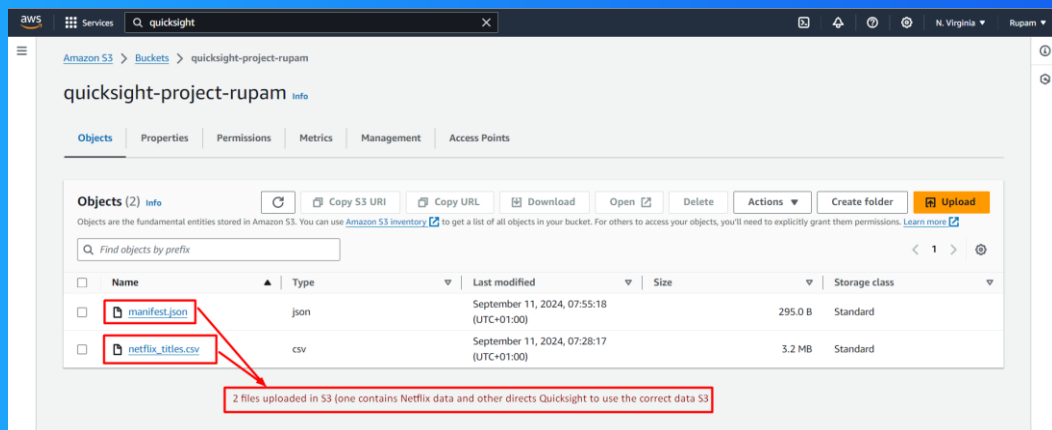
1. netflix_titles.csv
- 2.manifest.json

Note:

1. Netflix_titles.csv file contains the entire data of Netflix including movies,TV shows, etc.
2. Manifest.json file contains the path where the Netflix_titles.csv file is to be uploaded in S3 and also contains in which format the file is to be uploaded.

```
{
  "fileLocations": [
    {
      "URIs": [
        "s3://quicksight-project-rupam/netflix_titles.csv"
      ]
    }
  ],
  "globalUploadSettings": {
    "format": "CSV",
    "delimiter": ",",
    "textqualifier": "\"",
    "containsHeader": "true"
  }
}
```

Below ss shows the files uploaded in S3:



Why manifest.json is required ?

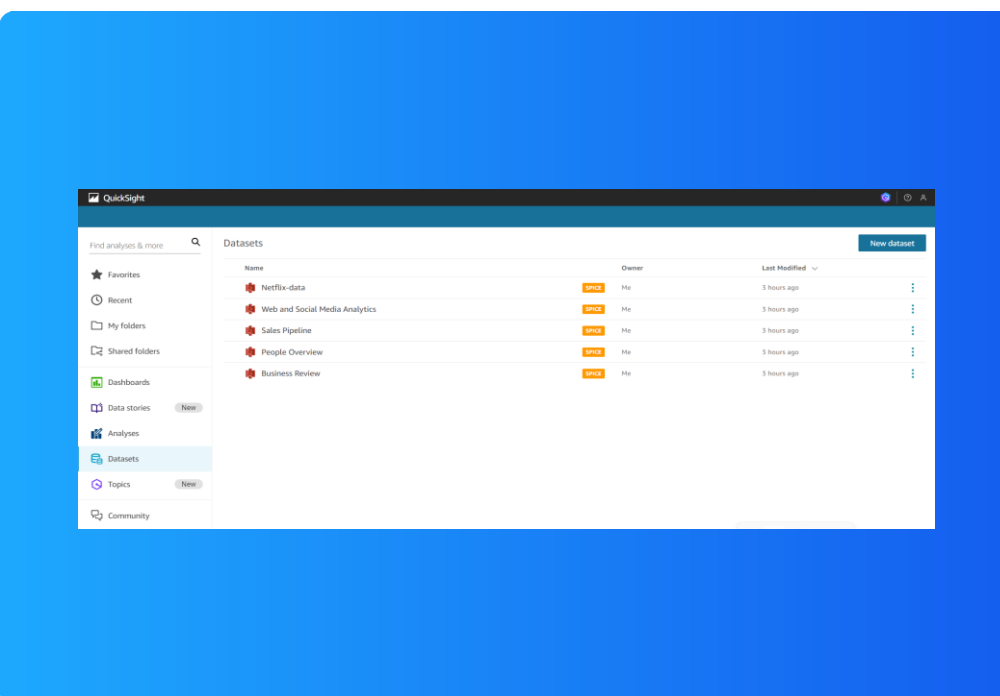
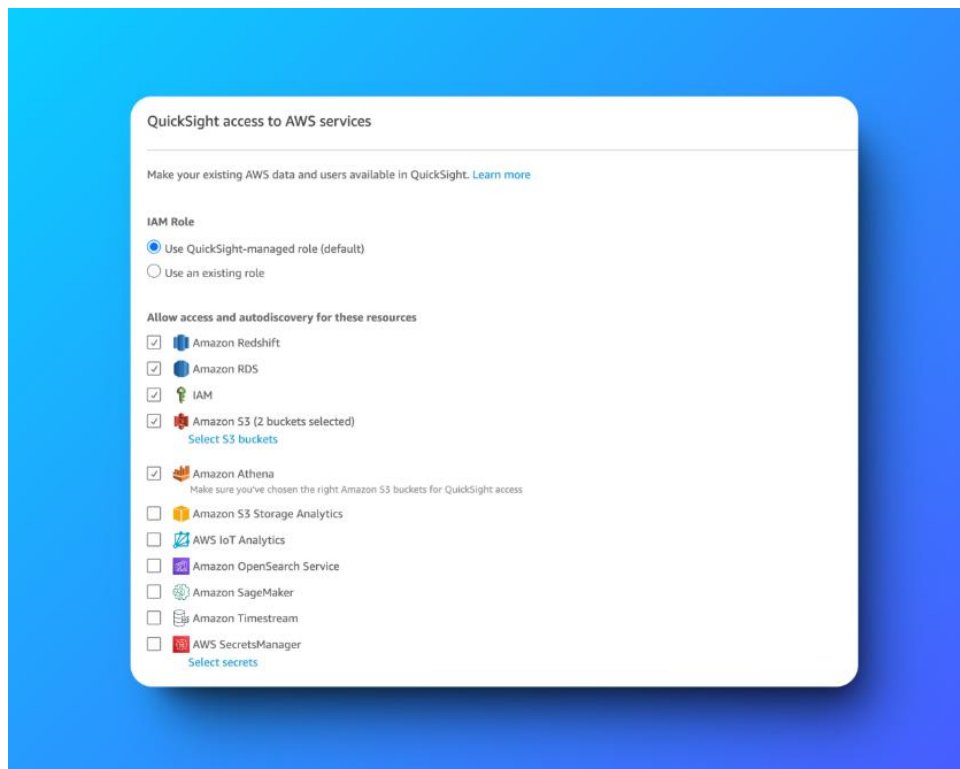
The **manifest.json** file is like a map that tells Amazon QuickSight where your data lives and how to read your data. Just like how people can speak in many different languages, data can come in all different kinds of forms!

manifest.json tells QuickSight what our 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 and not show your data correctly!

Create QuickSight account

Creating a QuickSight account is free of cost (free trial lasts for 30days)

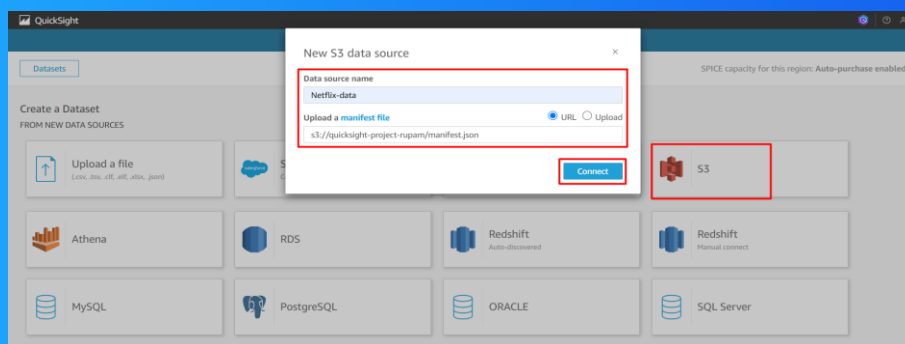
I also had to enable Quicksight access to S3 because my dataset is stored in S3 bucket and specific access is required for Quicksight to process that S3 data.



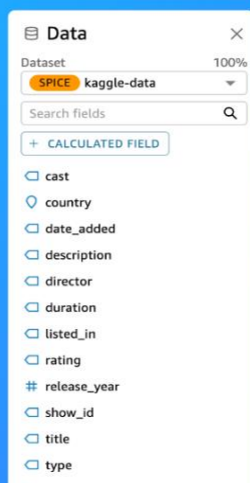
Download the Dataset

Connected the S3 bucket to QuickSight by visiting Quicksight AWS service.

There we needed to upload that manifest.json file to direct to the S3 path where our Netflix data is stored as shown below:



After connecting to the S3 bucket, Quicksight automatically imports all the column names from the dataset and lists it for our ease.

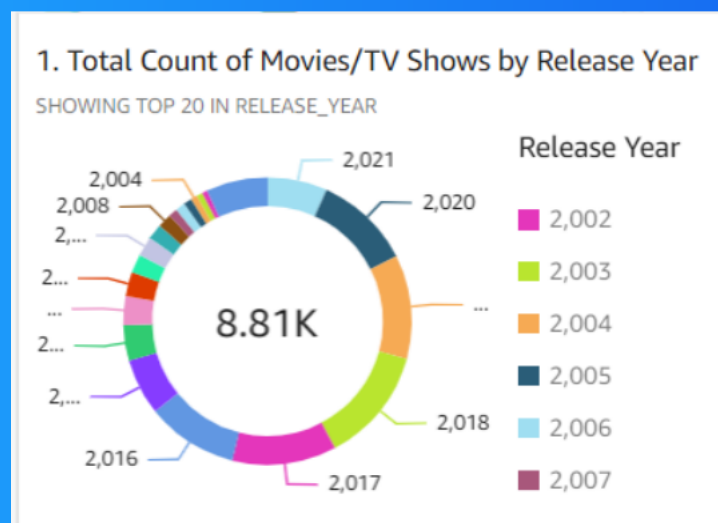


My first visualization

To create visualizations on QuickSight, dragged the relevant fields into Quicksight dashboard's autograph space.

The chart/graph shown below is a breakdown of movies vs tv shows for each release year.

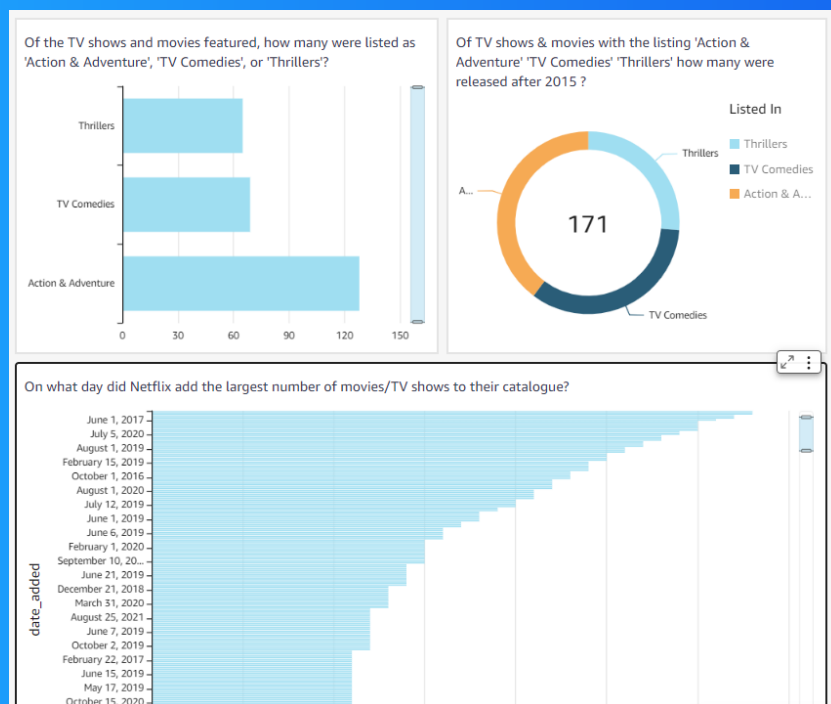
I created this graph by putting the release year on Y axis and making the type (i.e movie/tv show) the grouping variable



Using filters

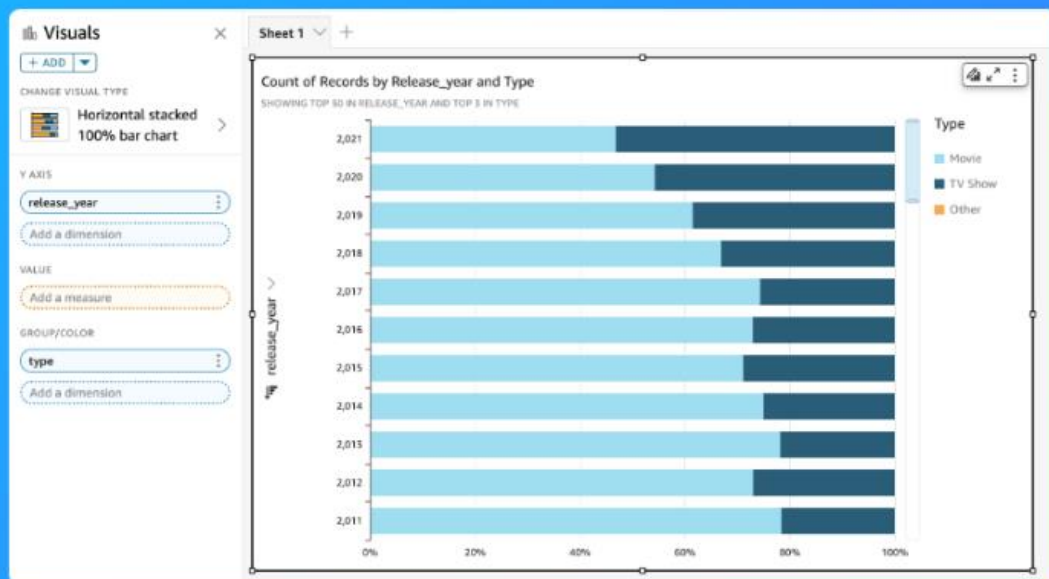
Filters are useful for specifying the exact subset of data that we want to analyze - effectively excluding any irrelevant data.

Here I added a filter by excluding movies and TV Shows that were released before 2015. This helped to create a visualisation on movies & TV shows, that were released from 2015 onwards.



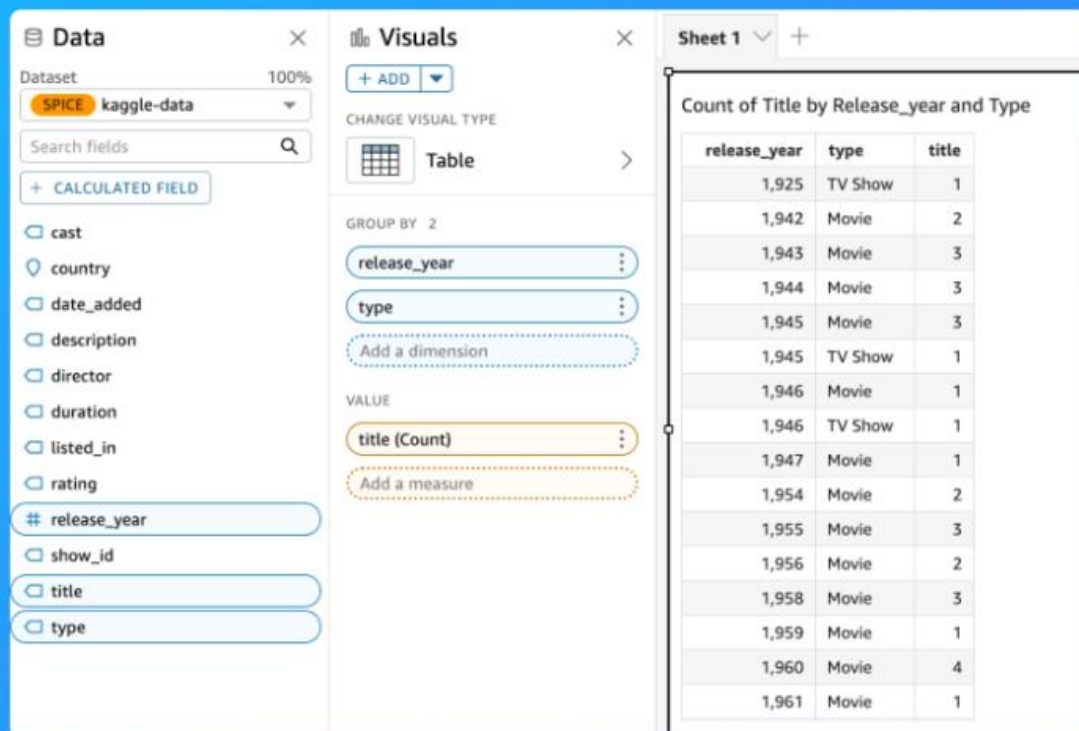
Few examples of visualizing data in Quicksight

- We already did the breakdown of TV shows/movies for each release year. Would it be possible to stack movies and TV shows in the same row, so we can visualise the % of each?



- Show the number of movies vs TV shows of these titles per release year..

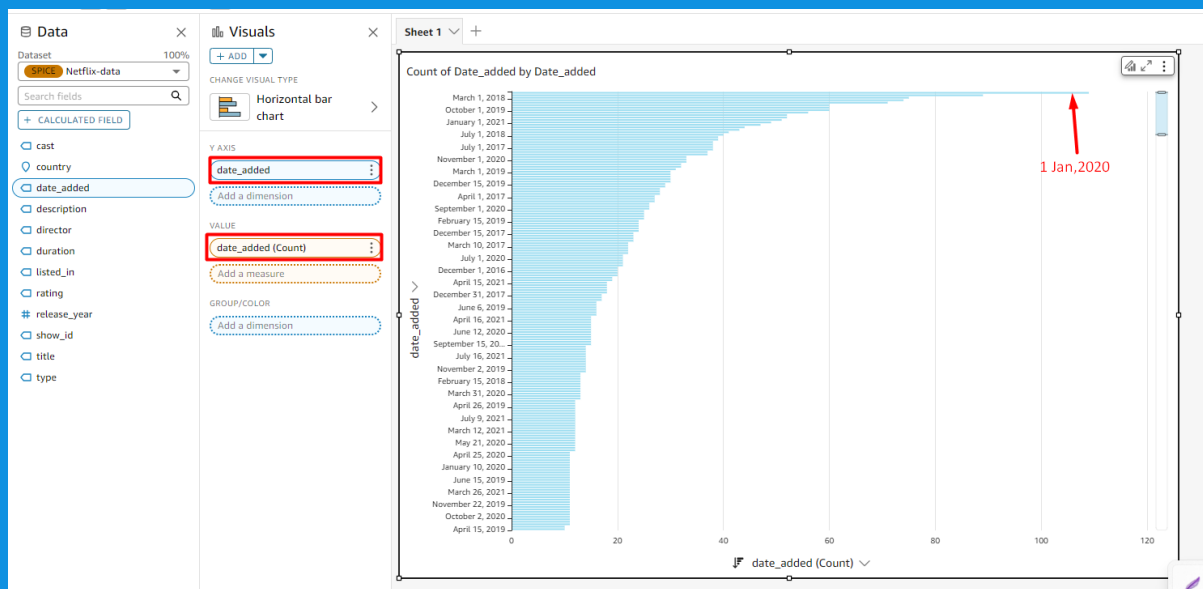
Sol: Change visual type to **Table**, then add release_year as **GroupBy** label.
Add title as **Value** metric.
Add the type label as a dimension!



- On what day did Netflix add the largest number of movies/TV shows to their catalogue?

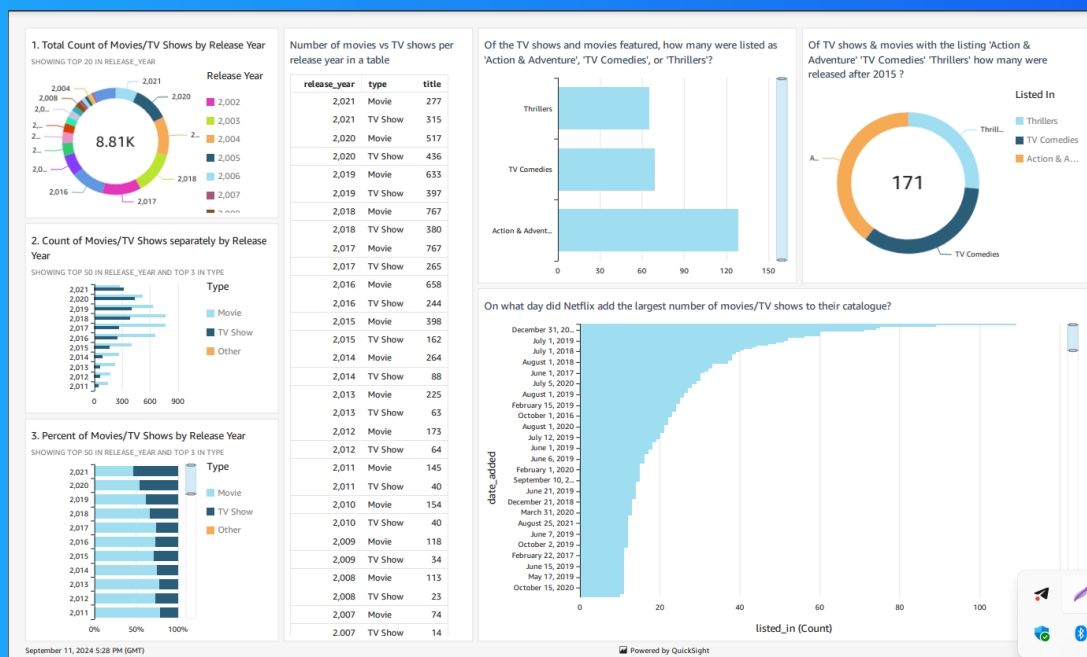
Answer: January 1, 2020. Netflix added 109 TV shows/movies to their catalogue.

To do this, move the date_added label to both the Y-Axis and Value headings.



Setting up a dashboard

As a finishing touch, I rearranged all the views and renamed them to get overview of relevant information and then using Publish button on top right corner, to make our dashboard public and also pdf downloadable.



Finally, did an export of the final dashboard in pdf format to keep it handy for use anytime. I will attach the pdf file in github for easy viewing.

Thanks and Regards,

Rupam Patwari