Visualising Netflix Data Using Amazon QuickSight

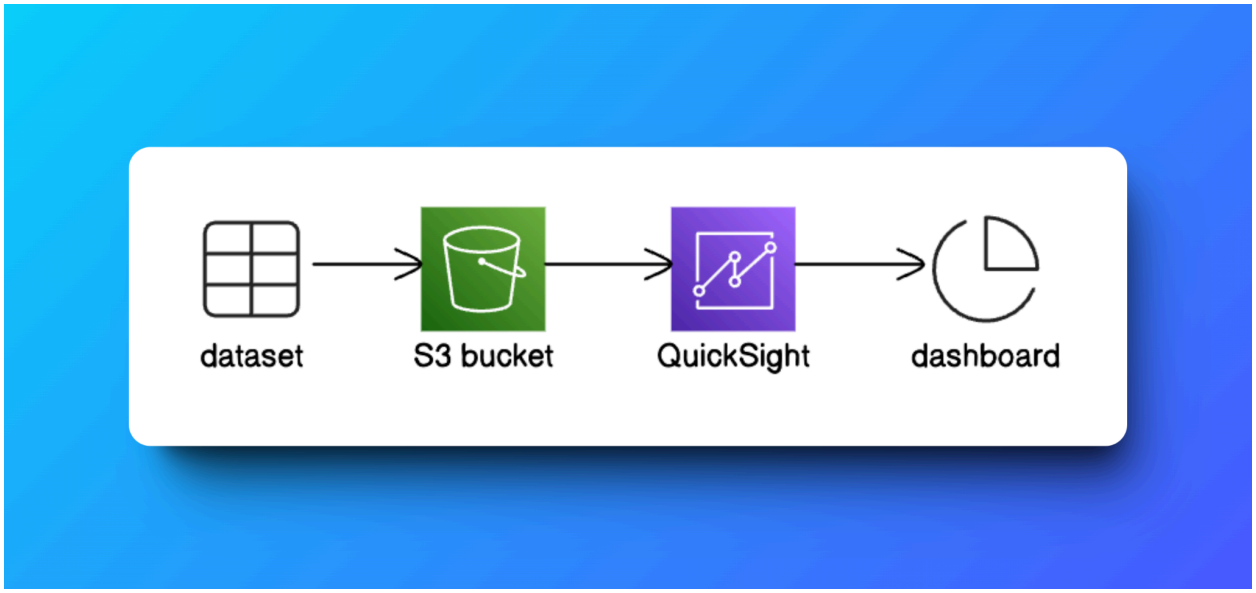
Netflix titles Analysis

Let's use Amazon QuickSight to visualize Netflix data and trends
AWS services : Amazon S3, Amazon QuickSight

Amazon QuickSight helps us to analyse data and create visualizations easily. Today we're going to analyse a huge dataset of Netflix shows and movies to create a dashboard that extracts all the insights.

In this project, we will:

1. 🪣 Upload a dataset into an S3 bucket.
2. 🆕 Create an account on Amazon QuickSight.
3. 🔗 Connect our dataset (in the S3 bucket) to Amazon QuickSight.
4. 📊 Create a variety graphs, charts and analysis using QuickSight.
5. 🏆 Publish a dashboard full of insights into our dataset!



Step 1: Download the Dataset

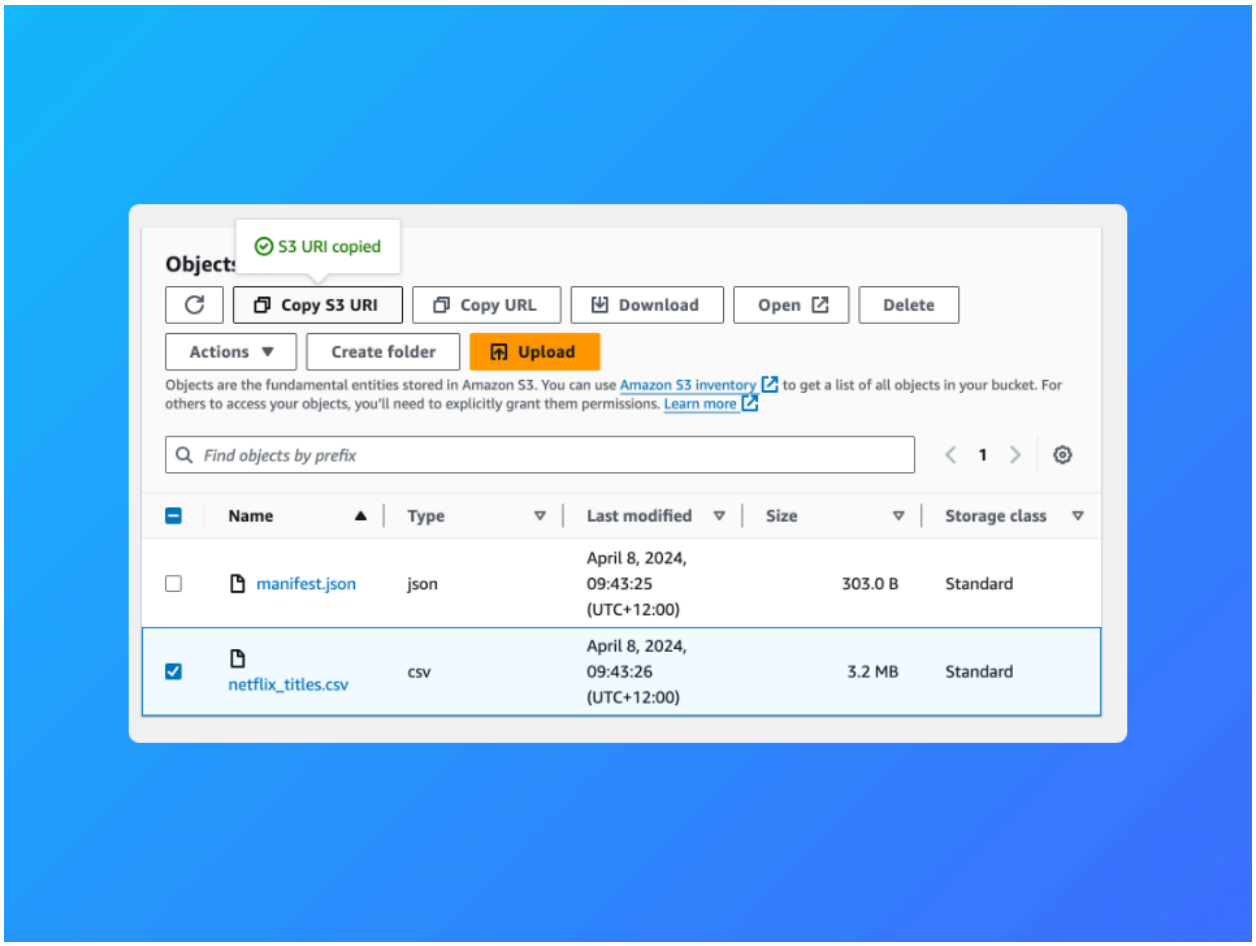
- Download netflix_titles.csv. (right click, and select Save Link As). This file contains all the data we're analysing!
- Download manifest.json. (right click, and select Save Link As). We'll explain why we need it in a minute!

Step 2: Store the Dataset in Amazon S3

1. Login into AWS account
2. Open our S3 Bucket
3. Select Create Bucket
4. Name the bucket quicksight-project-name(replace name with your name)
5. Select the Region closest to you
6. Keep the rest of the settings as default, and select **create bucket**.
7. Upload the CSV file and the **manifest.json** file into the bucket.
8. Copy the S3 URL of our **netflis_titles.csv** file.

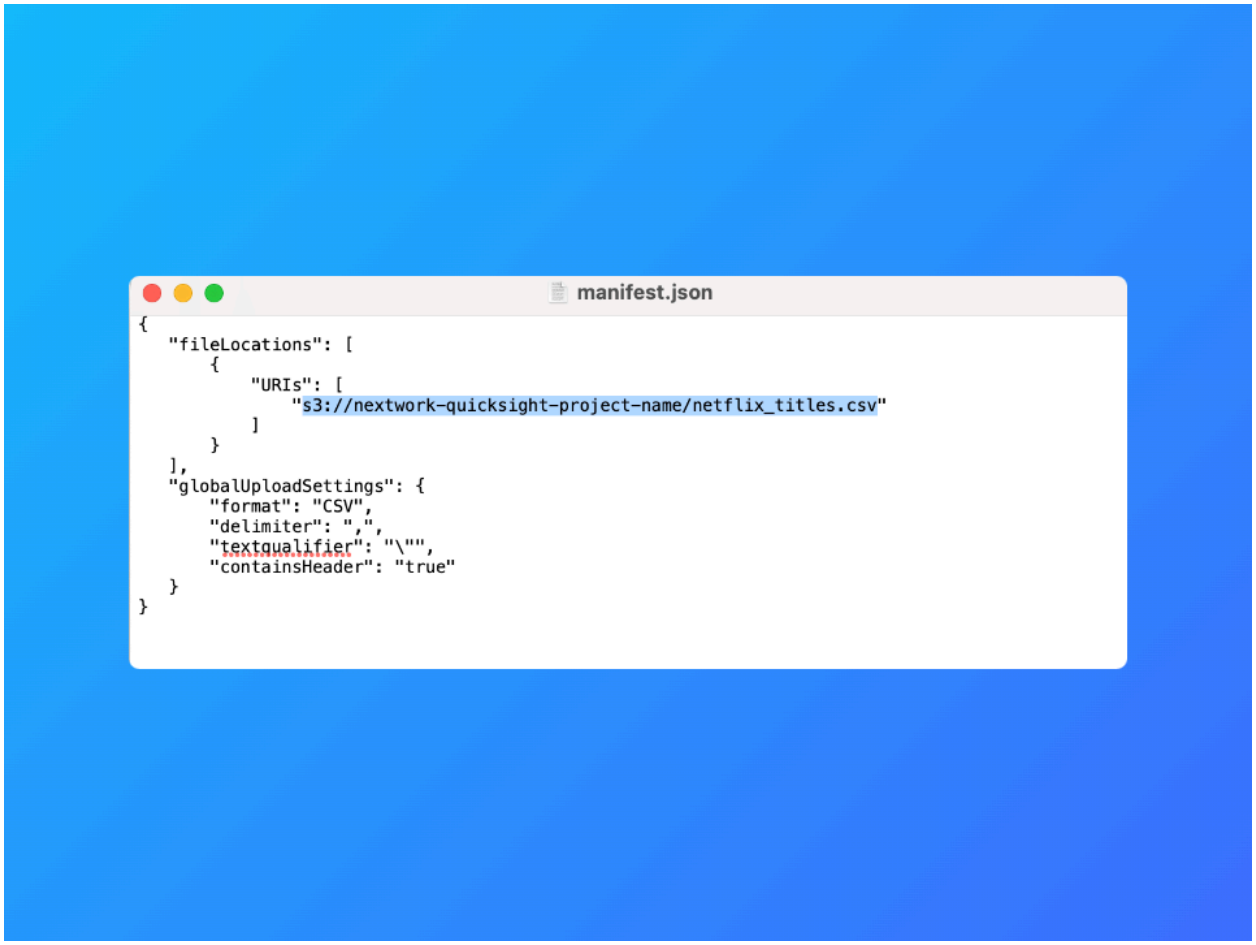
Why are we using Amazon S3? Isn't this project about QuickSight?

In this step, we're **storing** our dataset in an Amazon S3 bucket. Later, QuickSight will connect to S3 to use the data in this bucket and create visualisations.



Step 3: This is what our console should look like once we've copied our csv file's S3 URL.

- Open our **manifest.json** file in our laptop's text editor - for example, notepad
- Replace the URL in the **manifest.json file** with the S3 URL of our dataset

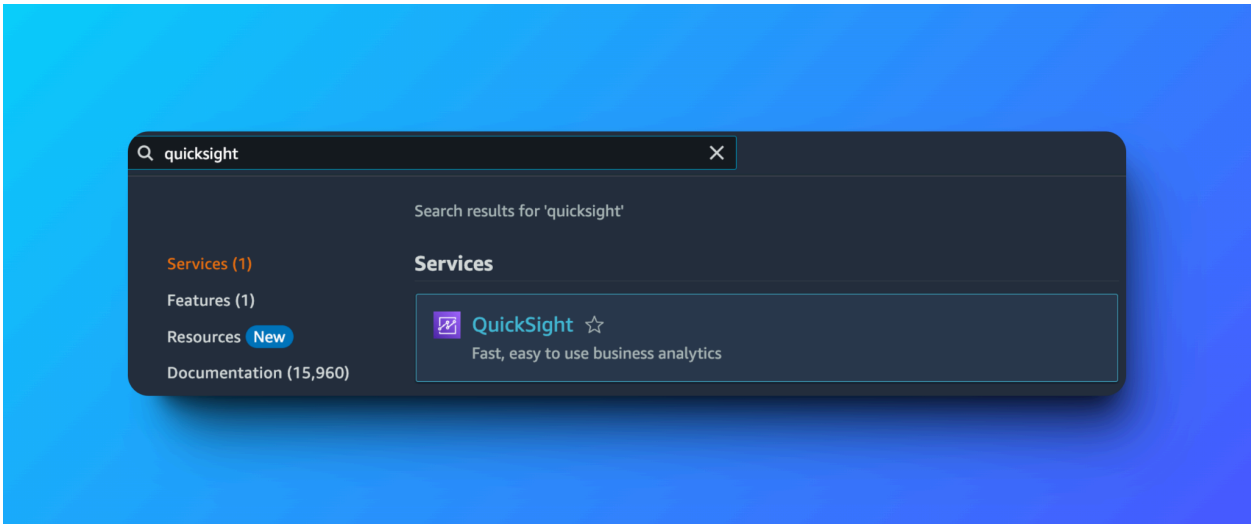


Step 2: Make sure name is now replaced with our name

- Re-upload the edited **manifest.json** file into our bucket, which we'll notice automatically replaces the existing one.

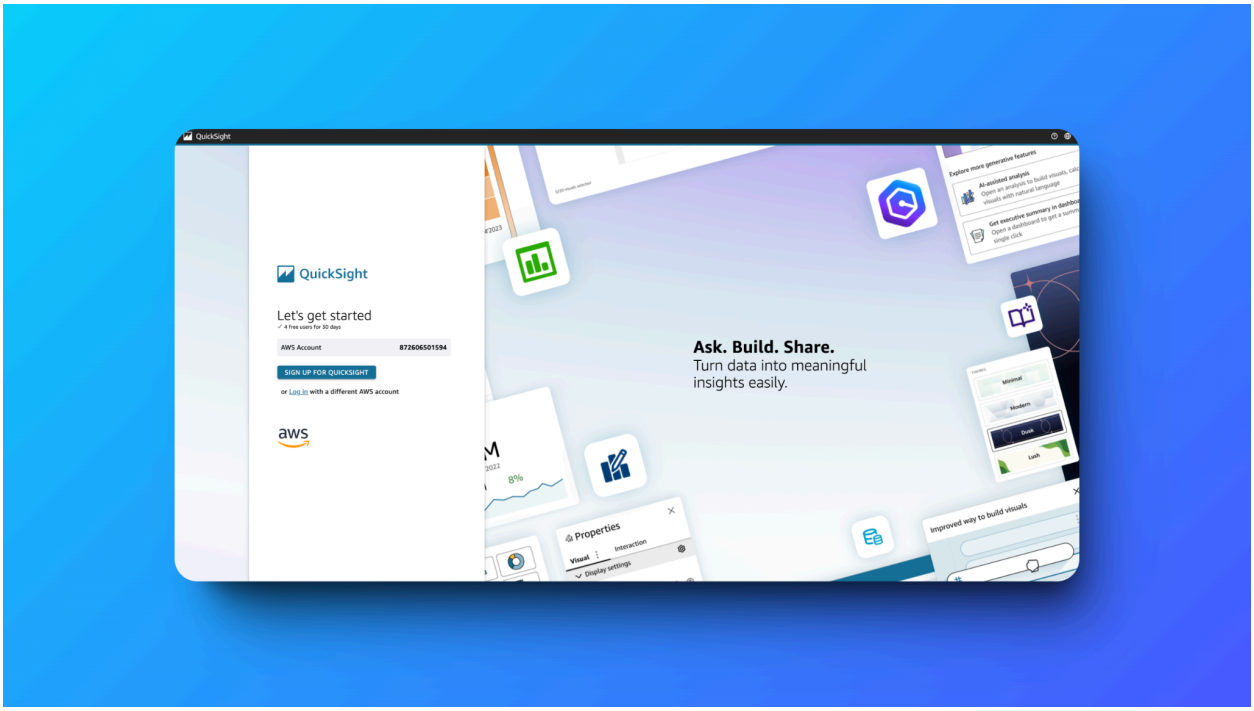
Step #3 Create our Amazon QuickSight Account

- Search for **Amazon QuickSight.**



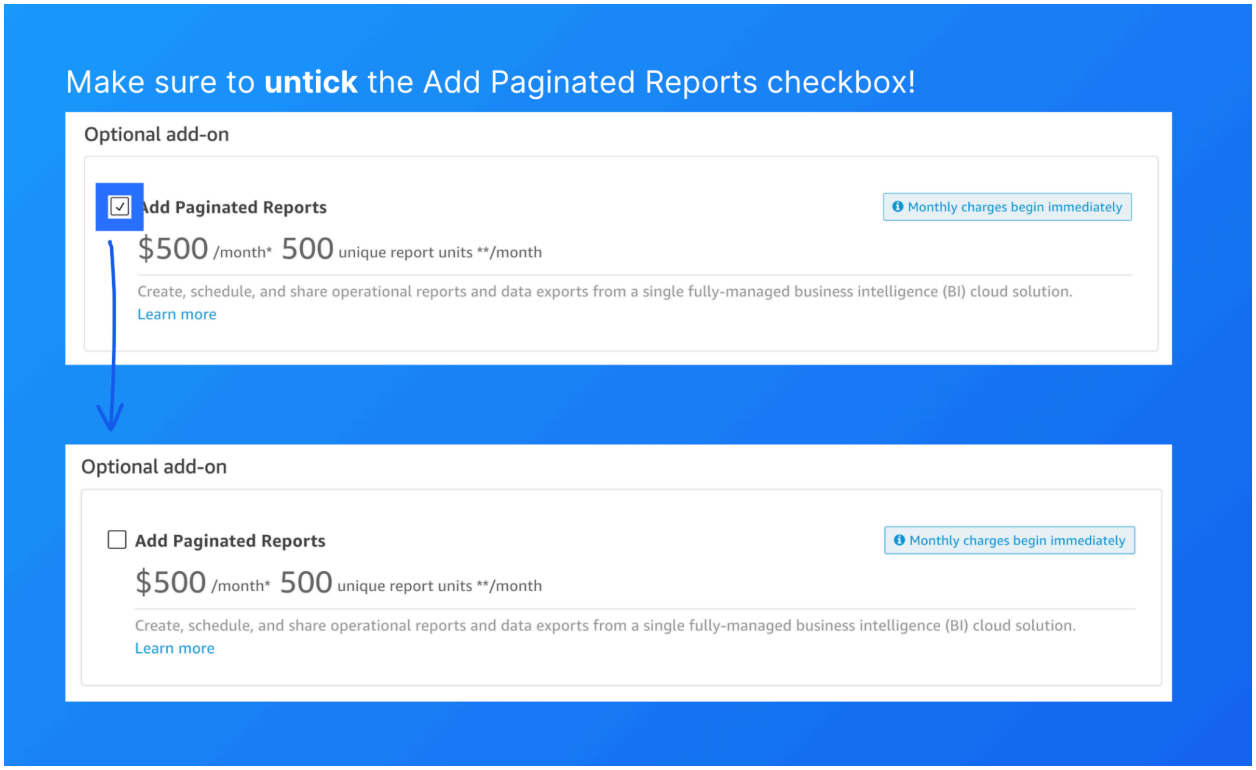
Step 3: Use that handy search bar to help ya out!

- Select **Sign up for QuickSight.**



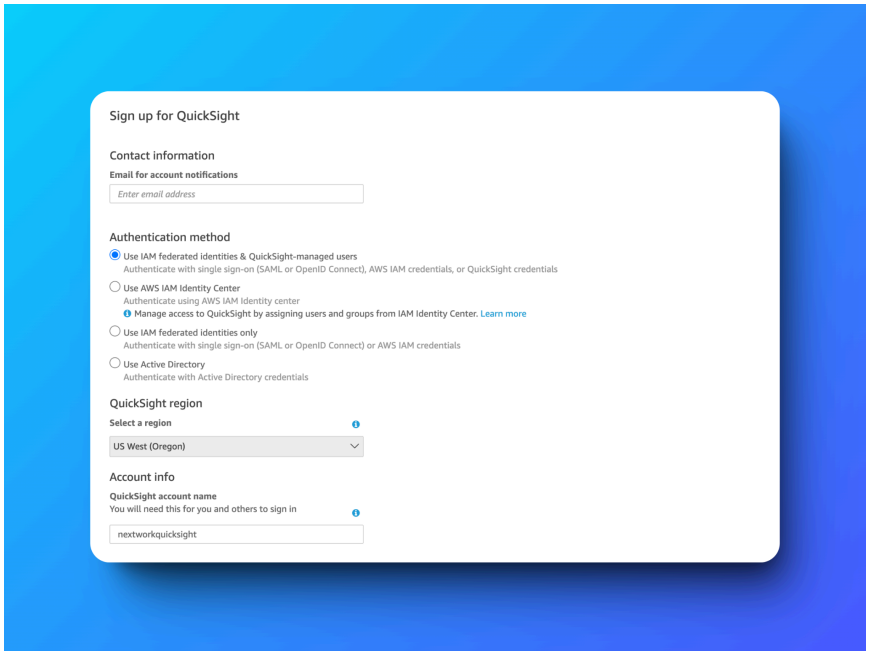
Step 3: Sign up for the free QuickSight trial

- Sign up for a free trial of the **Enterprise** edition if you don't have an account.
- 🚨 **IMPORTANT STEP** 🚨
  Make sure you uncheck this offer to upgrade.



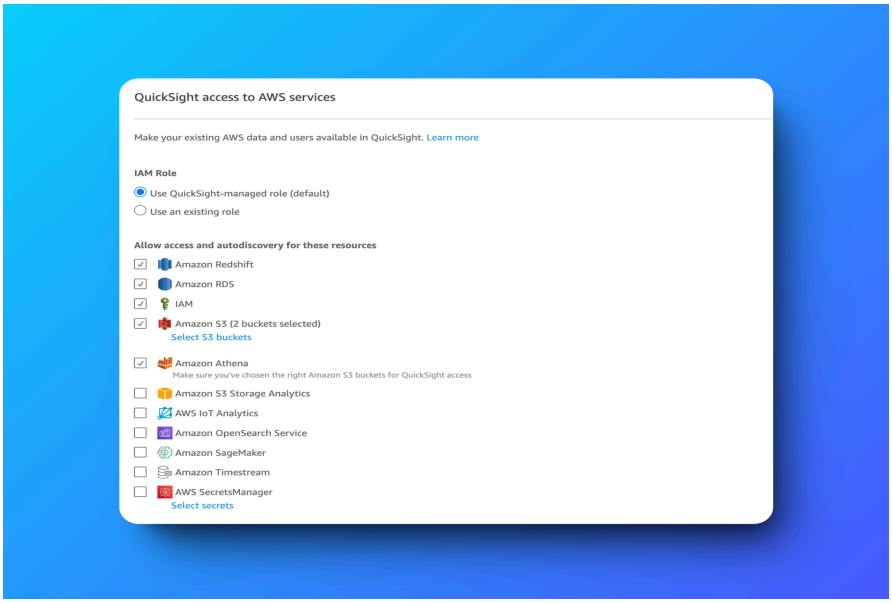
Step 3: Make sure we uncheck the Add Paginated Reports Option

- Enter our details for our QuickSight account - **make sure the email we use is the same email for your AWS account.** Or else you'll run into an error later!

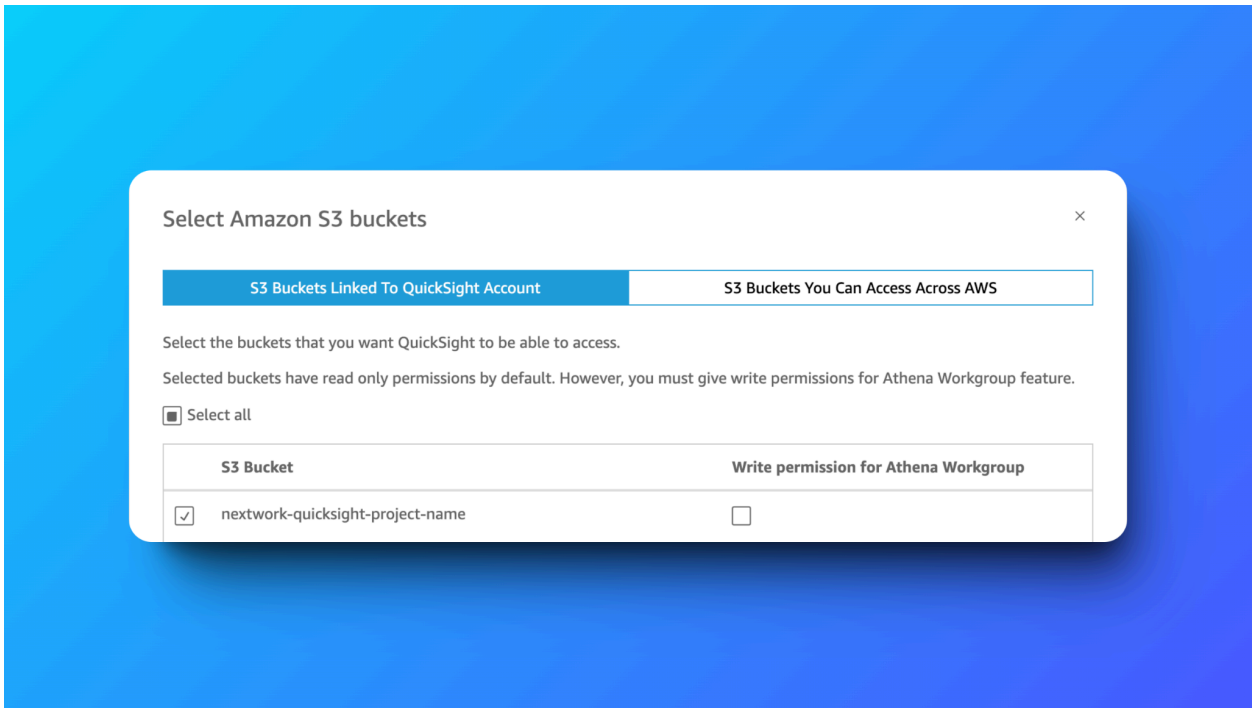


Step 3: Creating our QuickSight Account

- Select **Amazon S3.**
- Select **S3 buckets.**



- Tick the box for the S3 bucket we created.

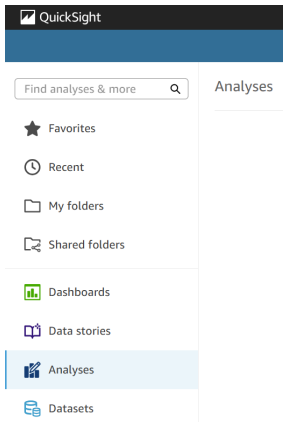


Step 3: Tick the S3 bucket we've just created!

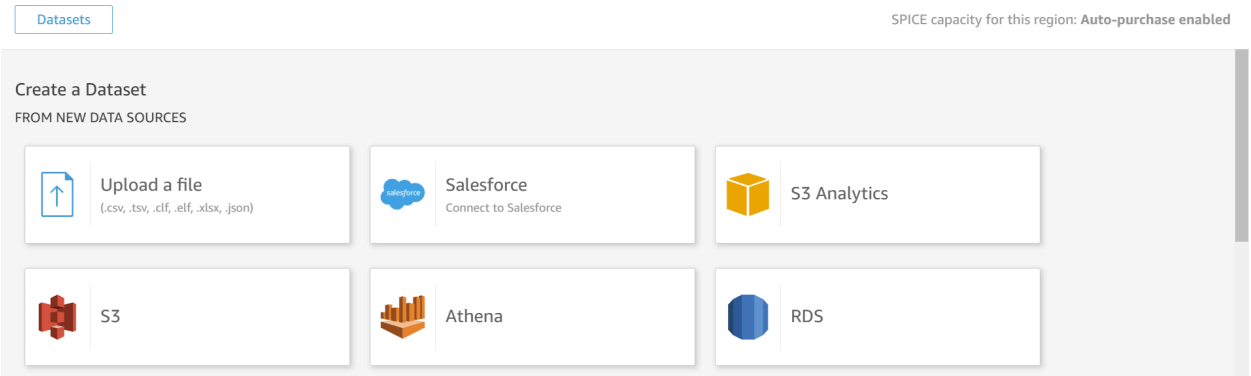
- Select **Finish.**
- We've created an Amazon QuickSight account!

Step 4: Connect our S3 bucket to Amazon QuickSight

From the Left hand navigation bar, select **datasets → New dataset**.



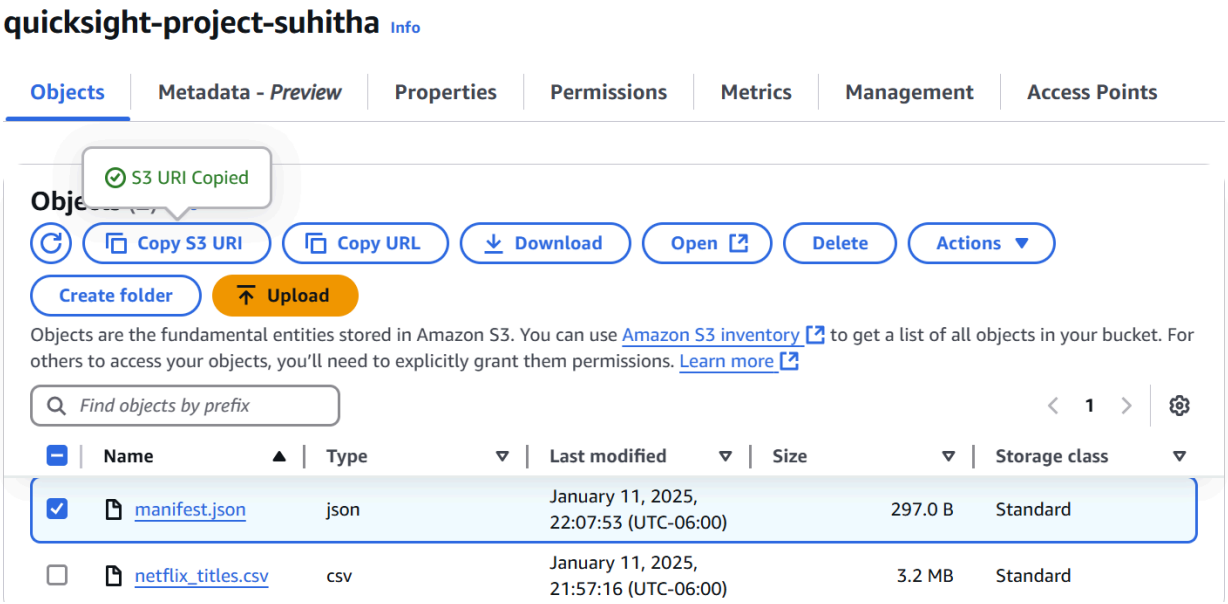
- Select **S3.**
- For the first field (source name), enter kaggle-netflix-data



What is the source name for?

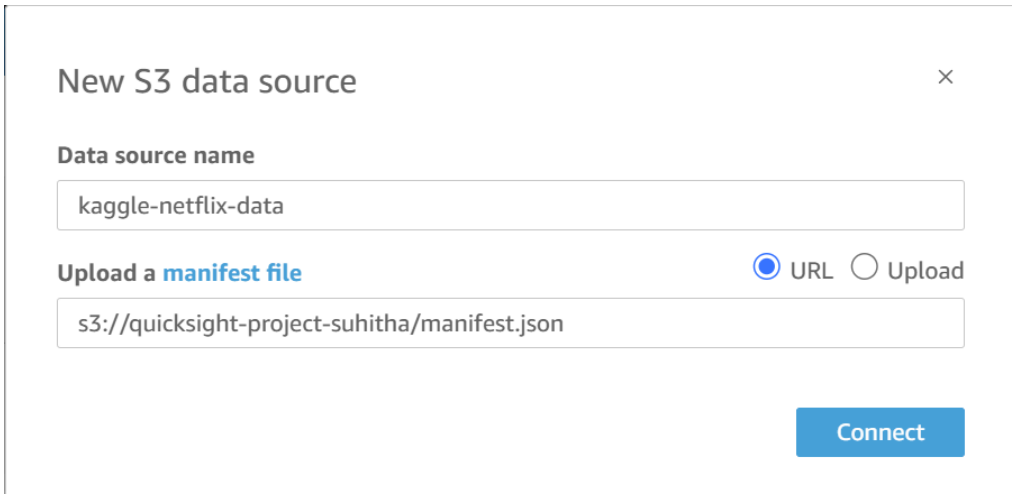
This source name is just a label to remind us what this piece of data is about! In this case, we sourced the csv field from a database called Kaggle, and the file is all about Netflix TV shows and movies.

- Ooo there's also a second field called **manifest.json URL** - does manifest.json found familiar?
- Open a **new tab** to open your AWS Management Console again. Head back to your S3 bucket.
- Select the checkbox next to **manifest.json**, then select **Copy S3 URL**.



Step 4: Copy the S3 URL

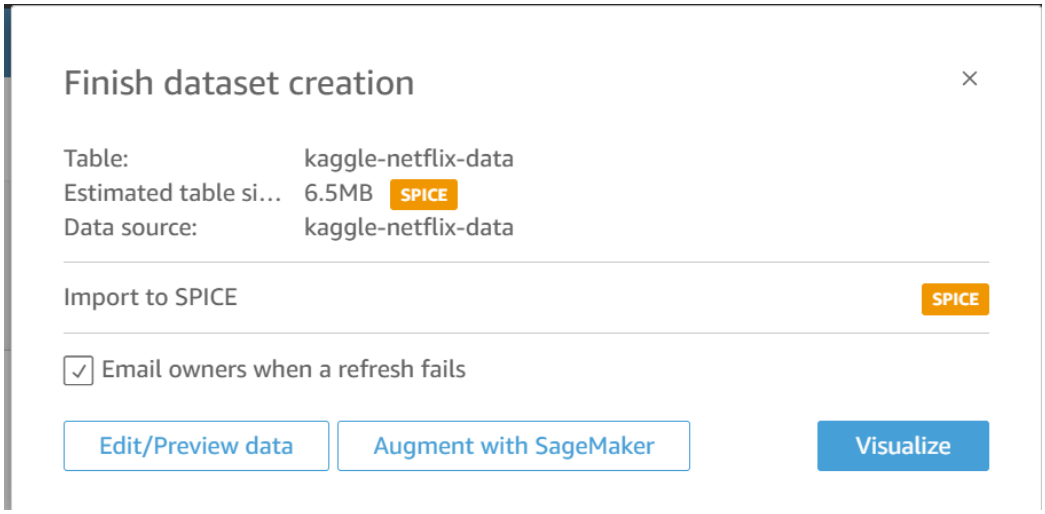
Enter the S3 URL to our **manifest.json** file.



Why do you need the manifest.json file? Check the tips below if you're stuck!

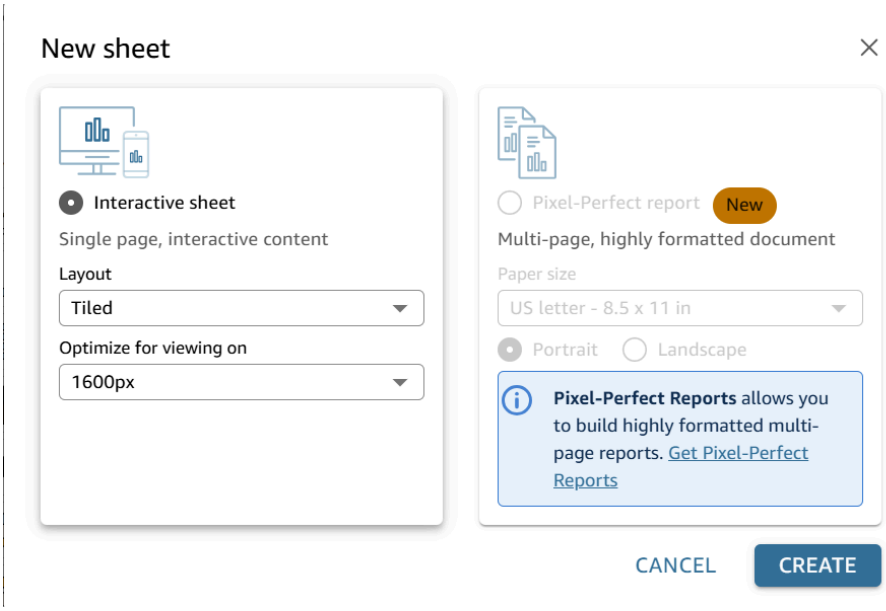
manifest.json tells QuickSight what our dataset looks like, so QuickSight knows how to understand the data and show it in charts or graphs. Without this map, QuickSight might get confused and not show your data correctly!

- Select **Connect**.
- Success!



Step 4: Time to visualize your data - exciting!

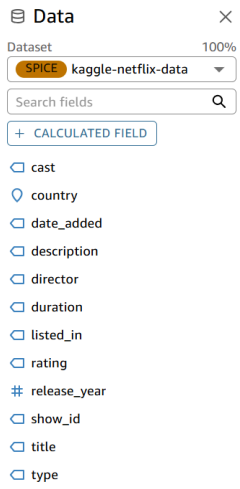
- Select **Visualize**.
- Select **Interactive sheet** to start creating visualisations.
- Select **Create**.



Step #5 Create our First QuickSight Visualization

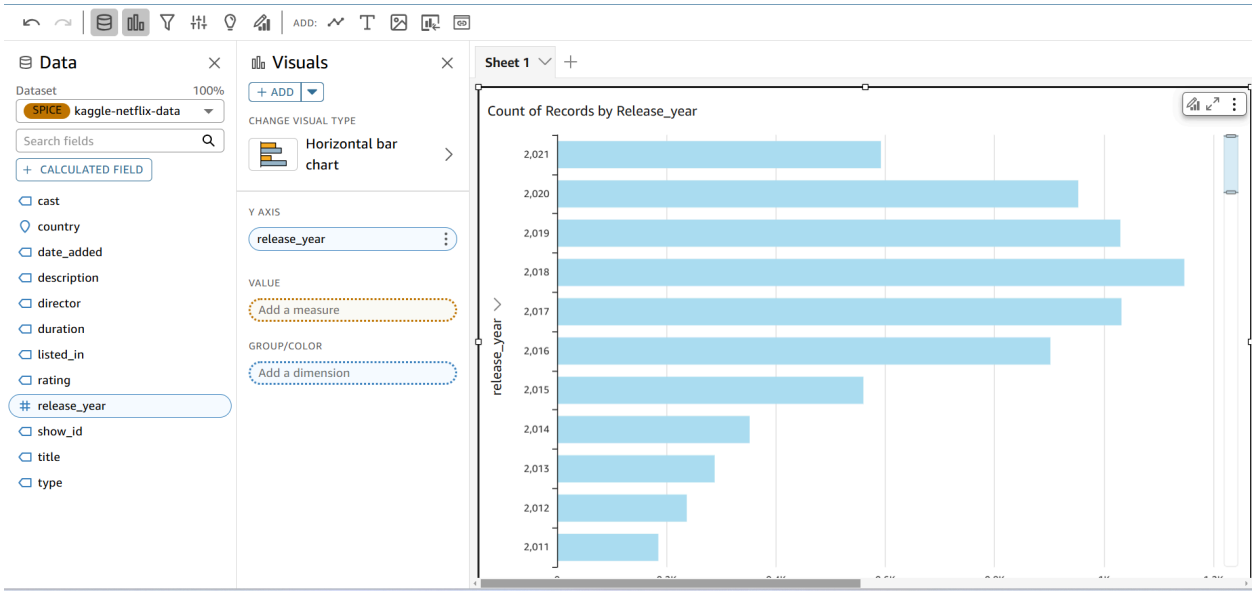
With QuickSight we can sort, filter and customize our data to create top visualizations. We can also experiment with different types of graphs like bar charts, pie charts, line graphs etc.

We can see on the left hand panel that the **dataset's fields are already imported.**



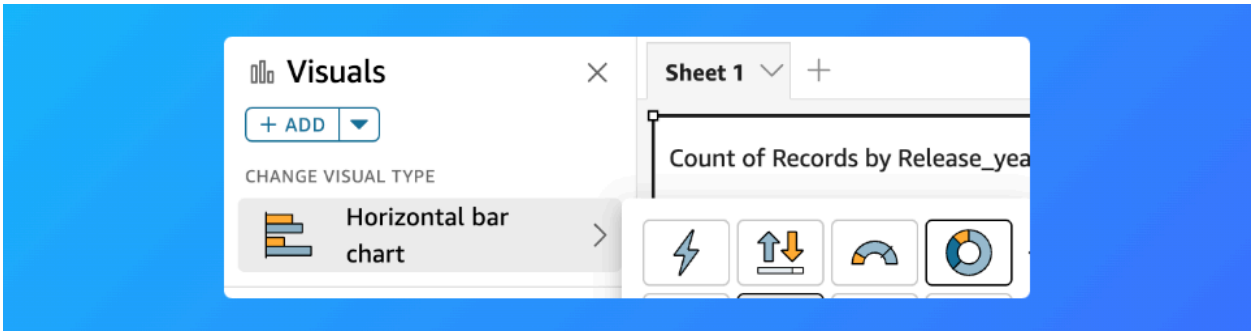
Step 5: Our dataset fields in the left hand panel

- Now, drag fields into the graph to create visualizations.
- Drag **release_year** into the Y-axis heading. Now we can see a breakdown on the year that these Netflix-featured TV shows and movies were released.



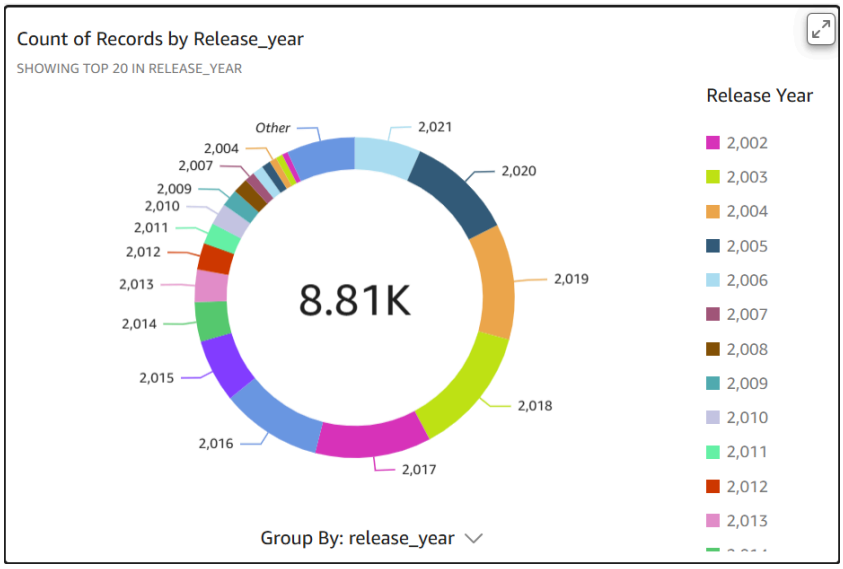
Step 5: Create our first visualization

We can create different types of charts too. See what happens when we change our chart to a donut chart!

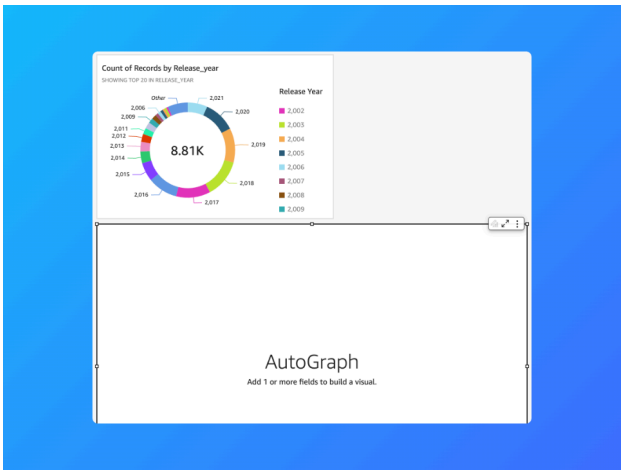


Step 5: Different types of charts available

- Let's save this in a dashboard. Click on the frame surrounding, and click on the white boxes at the edges to resize it.



- Now let's create a new visual, select + **ADD** under the **Visuals** heading on our middle navigation bar, and we'll see another blank frame pop out.

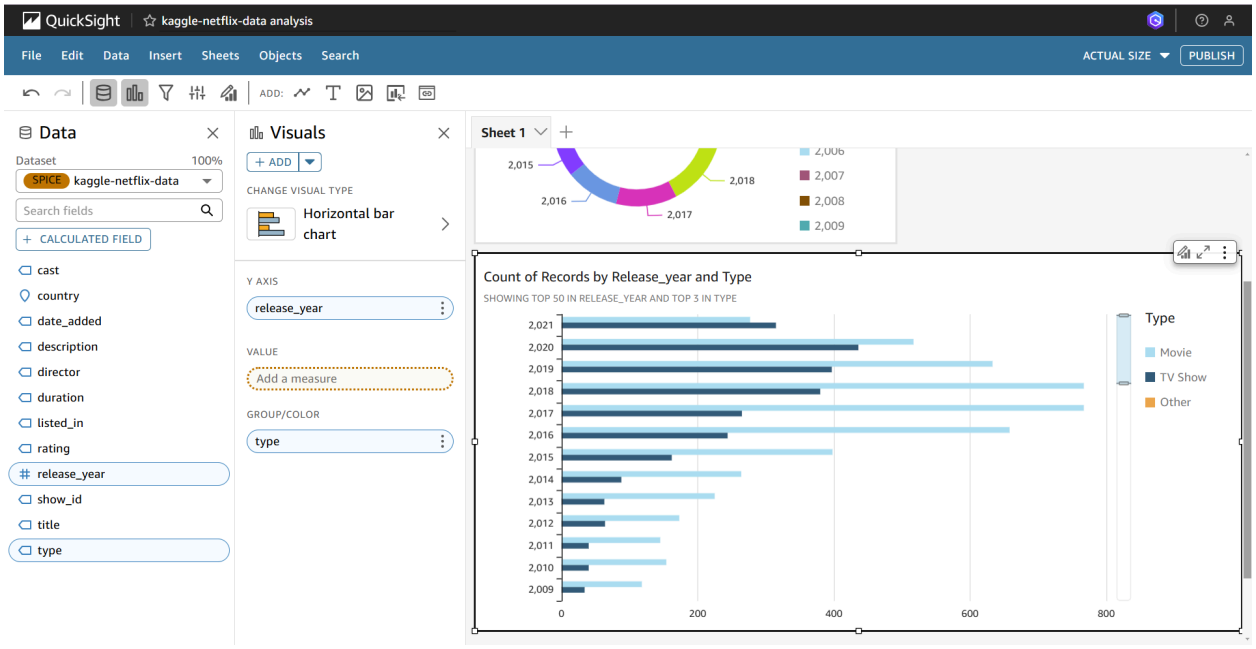


Step 5: new frame!

Based on the above visualizations, we know the overall count of Tv shows and movies. But what if someone asked out of 953 how many of those were movies and how many of those were Tv shows.

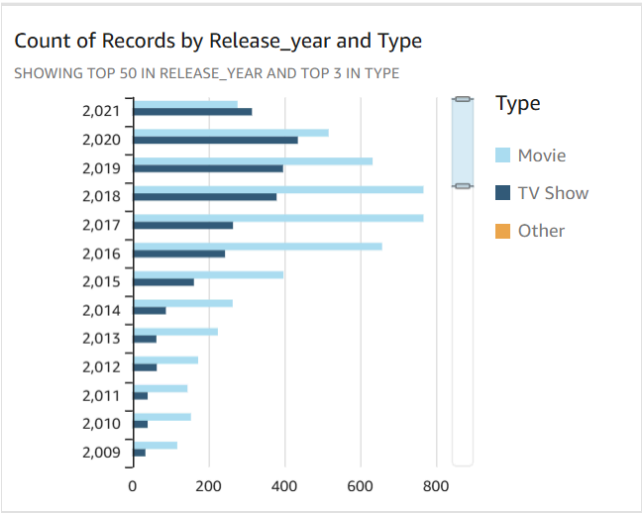
Now, what would we do to see a breakdown of TV shows vs movies for every year?

- Drag the **release_year label** into the **Y axis** heading.
- Next, drag the **type** label into the **Group/Color** heading.



Step 5: That's a nice looking chart!

- Resize your bar chart so it's the same width as the pie chart. If you click on the frame and hover over the black edges, you can also start to move your chart around the screen.

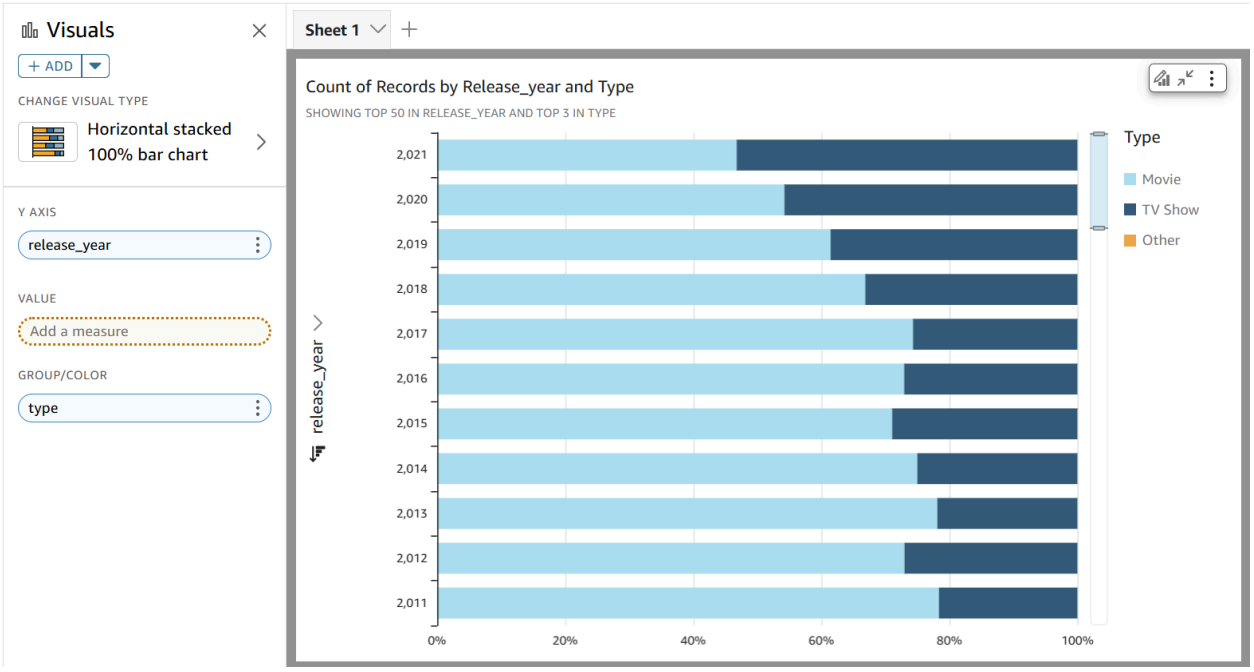


Let's say we've just presented our first two QuickSight charts to Netflix's data team lead, and they're impressed and would like to see what else we can find! They give us a brief with 5 questions that we should be able to answer with QuickSight.

Continue adding to our dashboard (make sure not to replace your existing graph for each task).

[Add a new chart instead of replacing the current one!]
1. "I quite like 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 you can visualise the % of each?"**

Answer: Picking up where we left off in the previous task, change the graph type to a Horizontal stacked 100% bar chart!

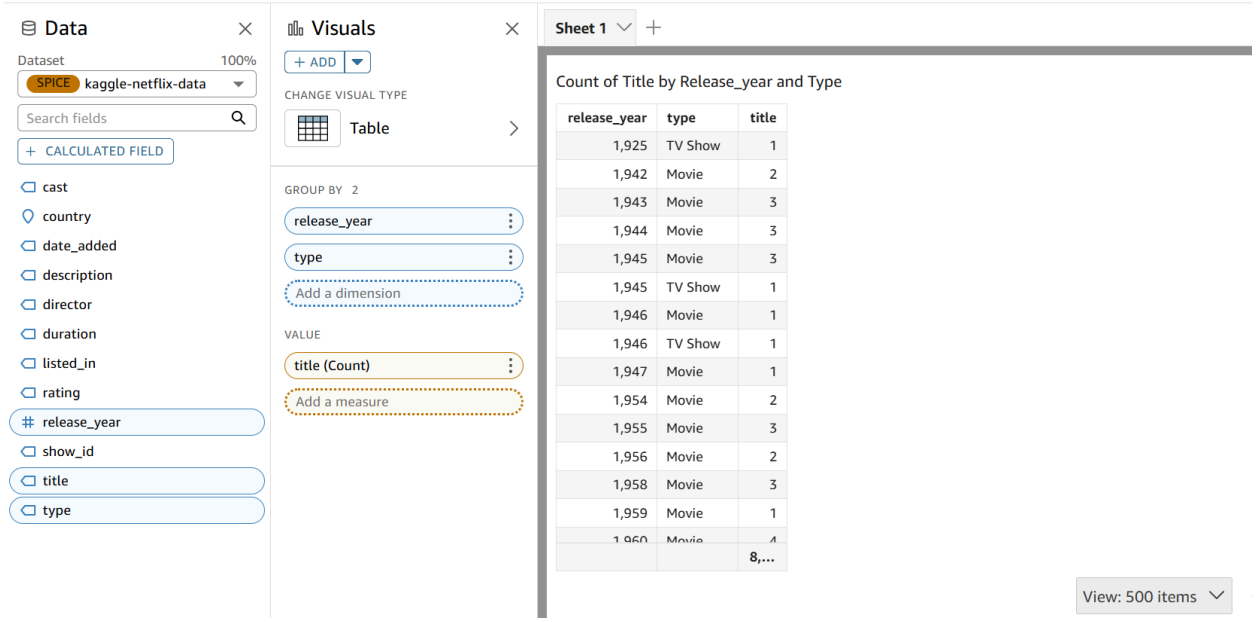


Step 6: Horizontal stacked 100% bar chart

[Add a new chart instead of replacing the current one!]
2. "Now can you show me the same thing on a table? i.e. **please show me the number of movies vs TV shows of these titles per release year in a table if possible."**

Answer: Change your visual type to **Table**, then add **release_year** as your **Group By** label. Add **title** as your **Value** metric. Add the **type** label as a dimension!

Here, we got each group but as per the given question we need to figure out how many there are in each group we need to drag something in the value field → add title means every tv show should have title and movie as well. Now, we can see the count in the year 1944 → movie → 3 released.

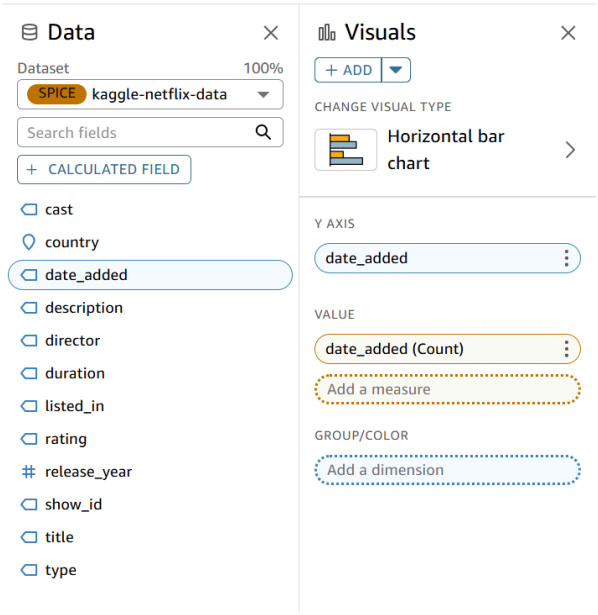


Step 6: Table grouped by release_year

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

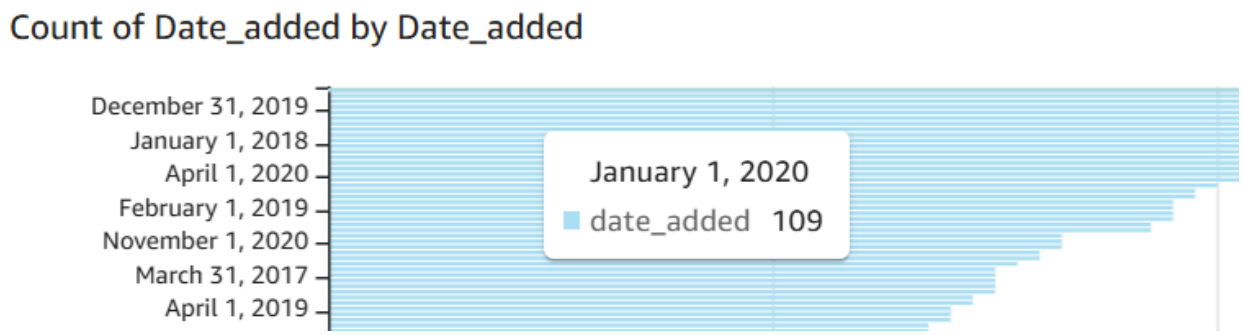
Answer: January 1, 2020. Netflix added 109 TV shows/movies to their catalogue.

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



Step 6: Adding 'date_added' to Y Axis and Value headings

- Then, next to the **Value** heading, click on the three dots and select **Sort order: Descending**.
- Use your mouse to scroll to the top bar - it's January 1, 2020!

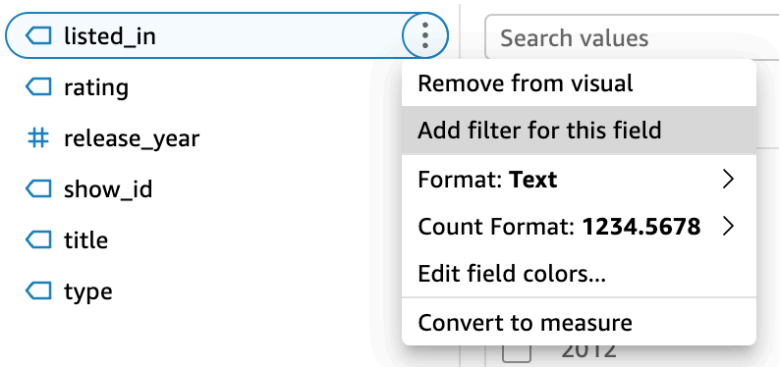


Step 6: Sort date_added by descending to find the largest number of movies/TV shows.

5. Of the TV shows and movies featured, how many were listed as 'Action & Adventure', 'TV Comedies', or 'Thrillers'? For simplicity, ignore the TV shows and movies that have multiple listings/categories."

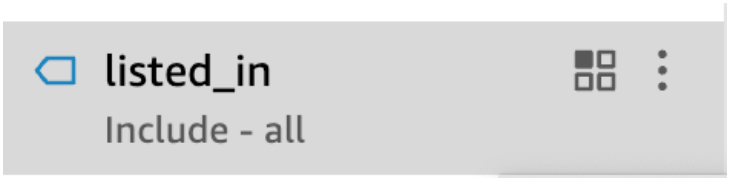
Answer: 65 + 69 + 128 = 262

- To do this, add a filter for the listed_in field.



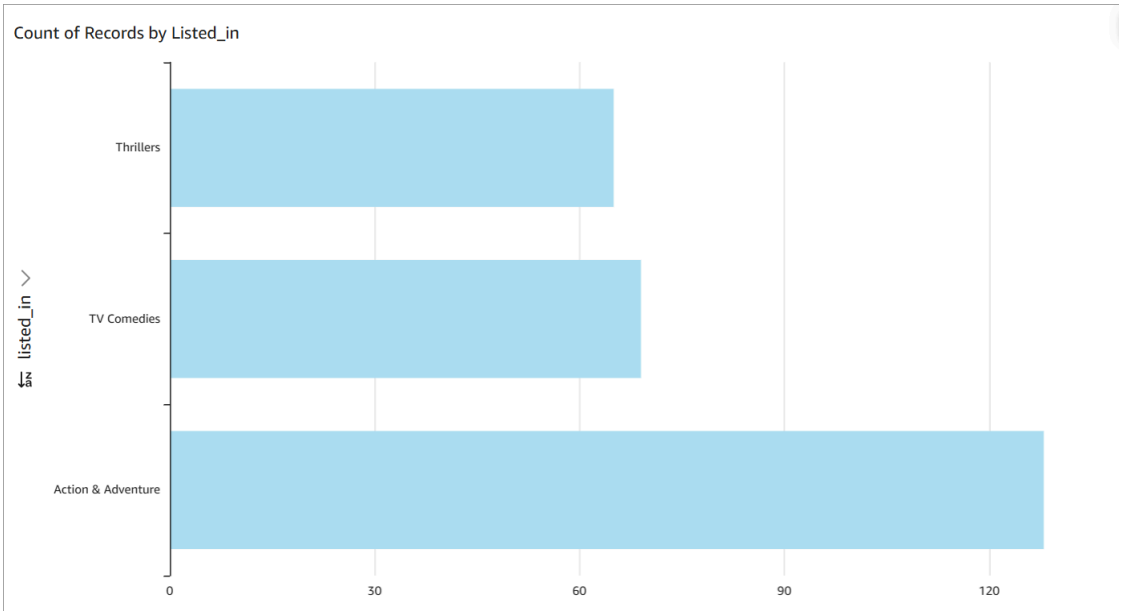
Step 6: Filter by 'listed_in'

- Click on the created filter to edit it:

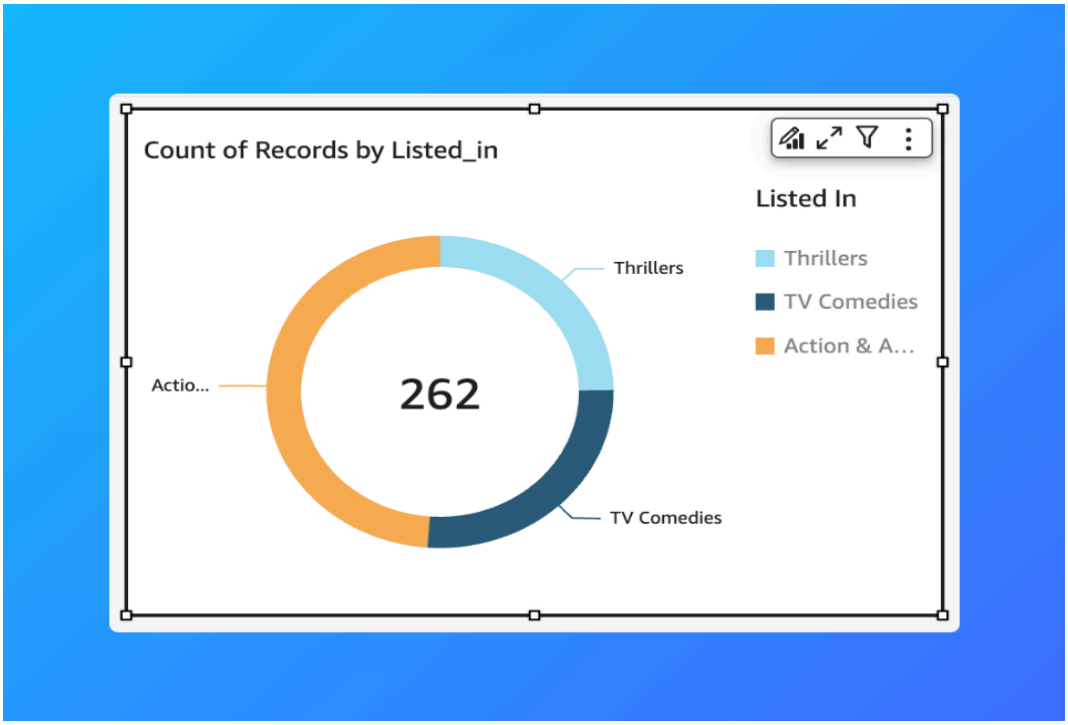


Step 6: Edit filter

- Deselect the Select all checkbox, then select these three tags (use the Search values search box to help you):
    - Action & Adventure
    - Thrillers
    - TV Comedies
- Select Apply when you're done. Ta-da!



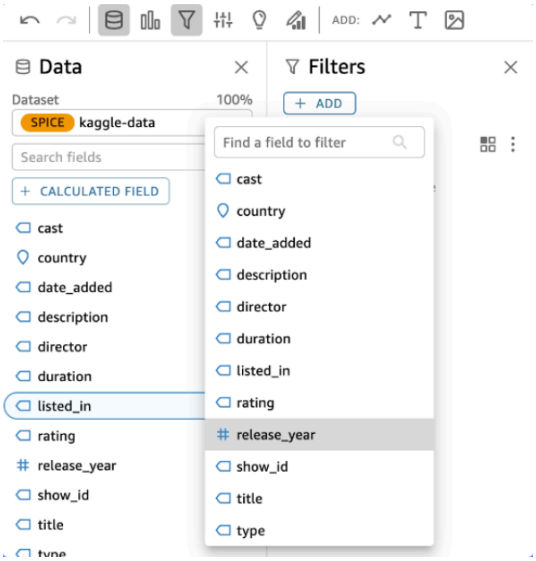
Tip: If we prefer, we could even display this as a donut chart to see the total count at a glance.



6. Of the TV shows and movies with the listing, how many were released in 2015 or after?

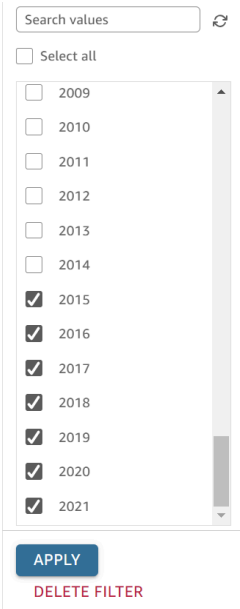
Step 6: How many TV shows and movies were released on or after 2015?.

- Next, let's make sure this is filtered to items that were released from 2015 or after. Select the Filter icon and add a **release_year** filter.

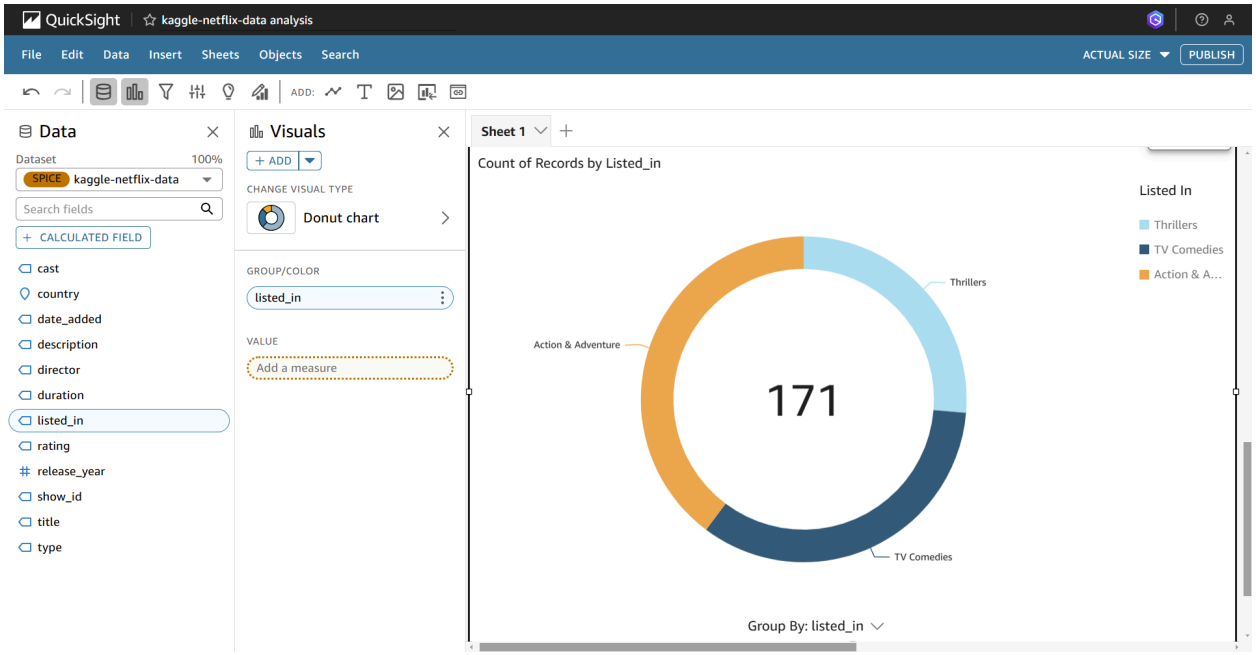


Step 6: Apply the filter

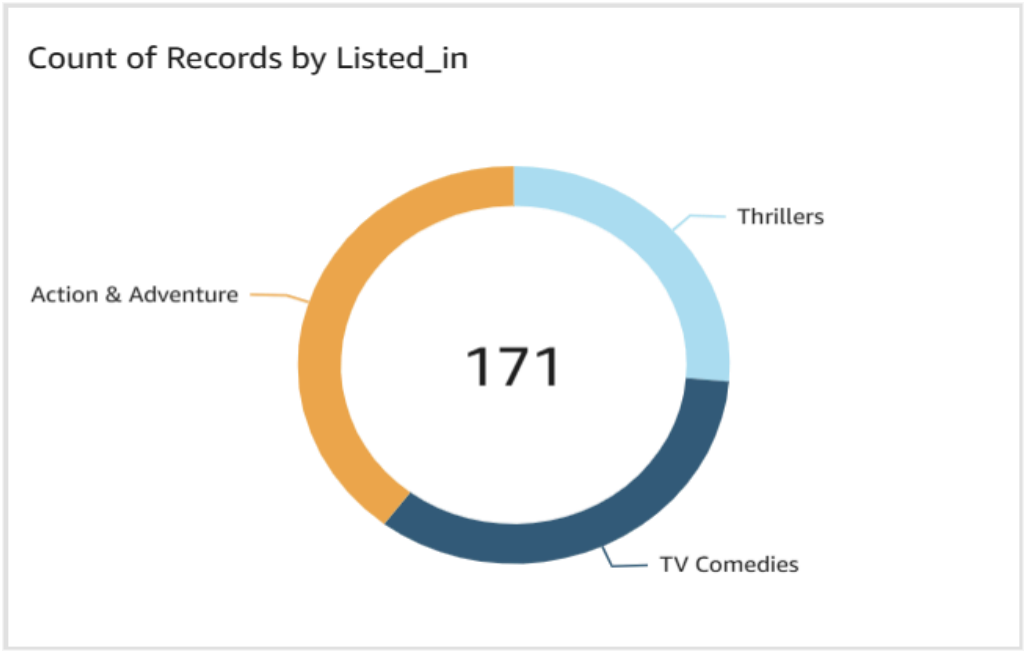
- Make sure only the years 2015 and over are selected, then select **Apply**.



Step 6: Only apply the years 2015 and over are selected.

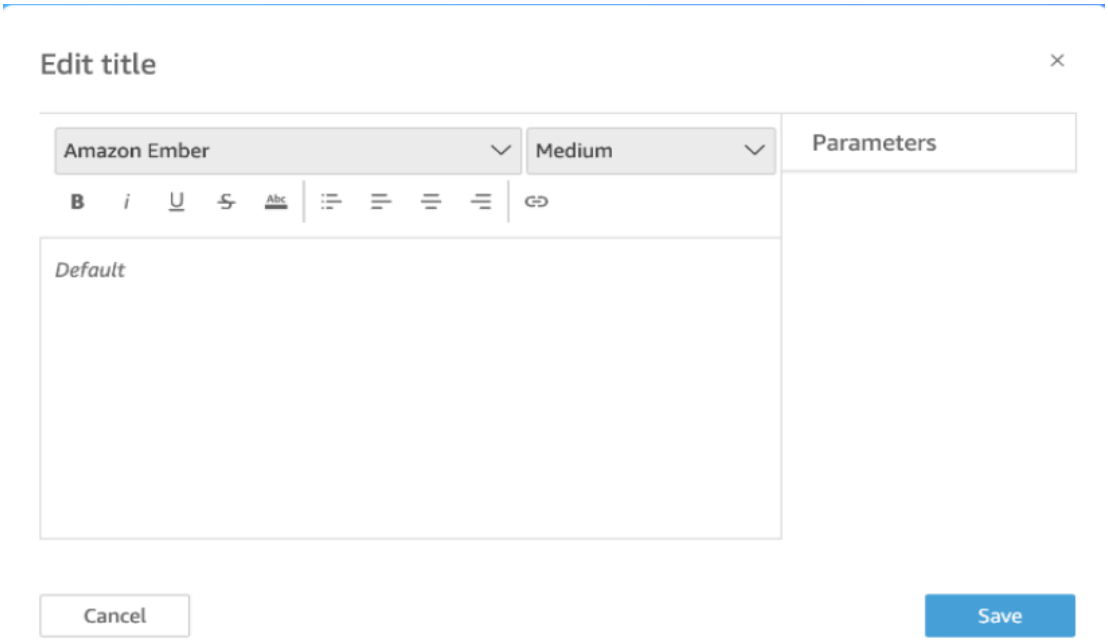


Step 6: Count of Records in Listed_in.



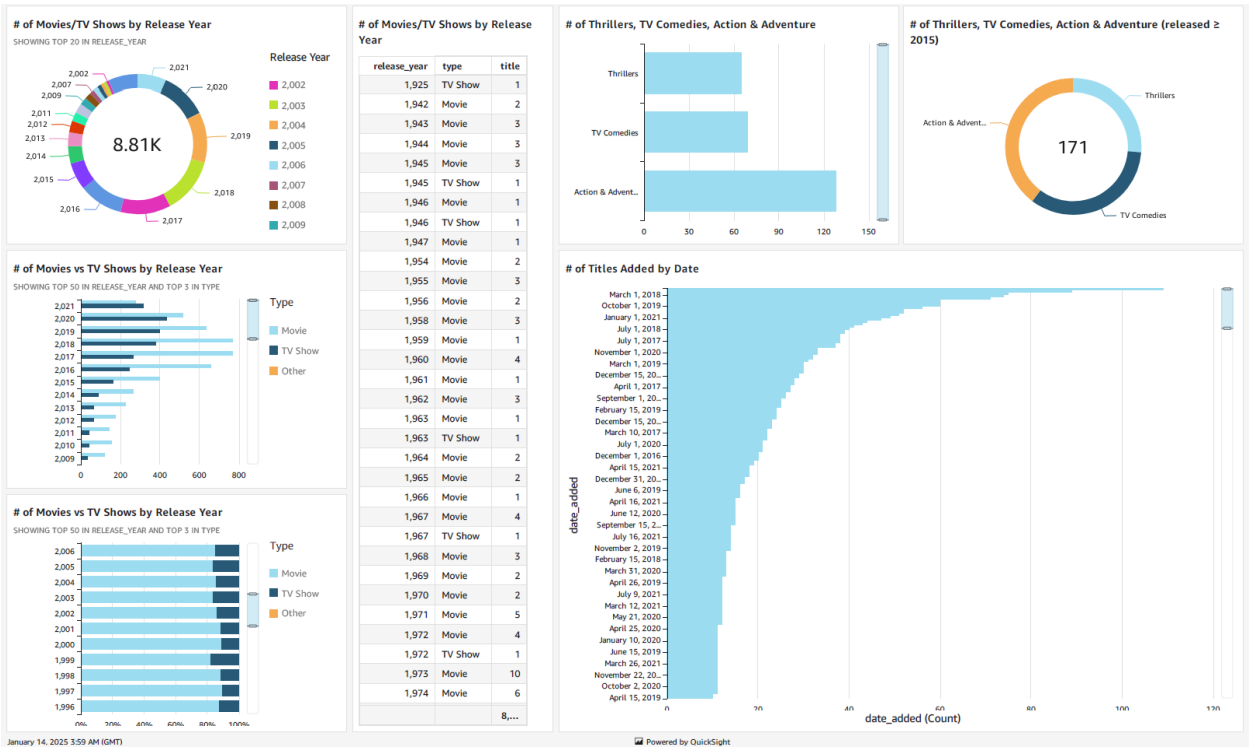
To round it all off, edit the titles in your charts so that anyone can understand them at a glance.

- Double click on the titles of all the charts you see in front of you.
- This **Edit title** panel will pop up:



Step 7: Add titles to your charts

Final Dashboard



Summary

In this project, we created a dashboard using Amazon QuickSight.

I…..

- 🪣 Uploaded my dataset into an S3 bucket.
- 🔗 Connected my dataset to Amazon QuickSight.
- 📊 Created graphs, charts and analysis using QuickSight.
    - Visualized a catalogue of TV/movies by release year, listing category and added date.
    - Experimented with different visualisation styles - for example, donut, table or bar graphs.
- 💄 Formed a dashboard that I published and can export!