nextwork.org

Visualize data with QuickSight

CH Arpita Chowdhury

CH
Arpita Chowdhury
NextWork Student

Introducing Today's Project!

In this project, I will demonstrate how to use Amazon QuickSight to analyze Netflix data and generate visualization and insights. I'm doing this project to learn how to use cloud for data service and data anaysis.

Tools and concepts

Services I used were QuickSight and S3. Key concepts I learnt include manifest.json, data visulization techniques i.e charts,filters and how to perform data reset in QuickSight.

Project reflection

This project took me approximately took 2hrs 30 min including demo time. The most challenging part was how manifest.json file works. It was most rewarding to see the end result that is the final dashboard and generating the PDF.

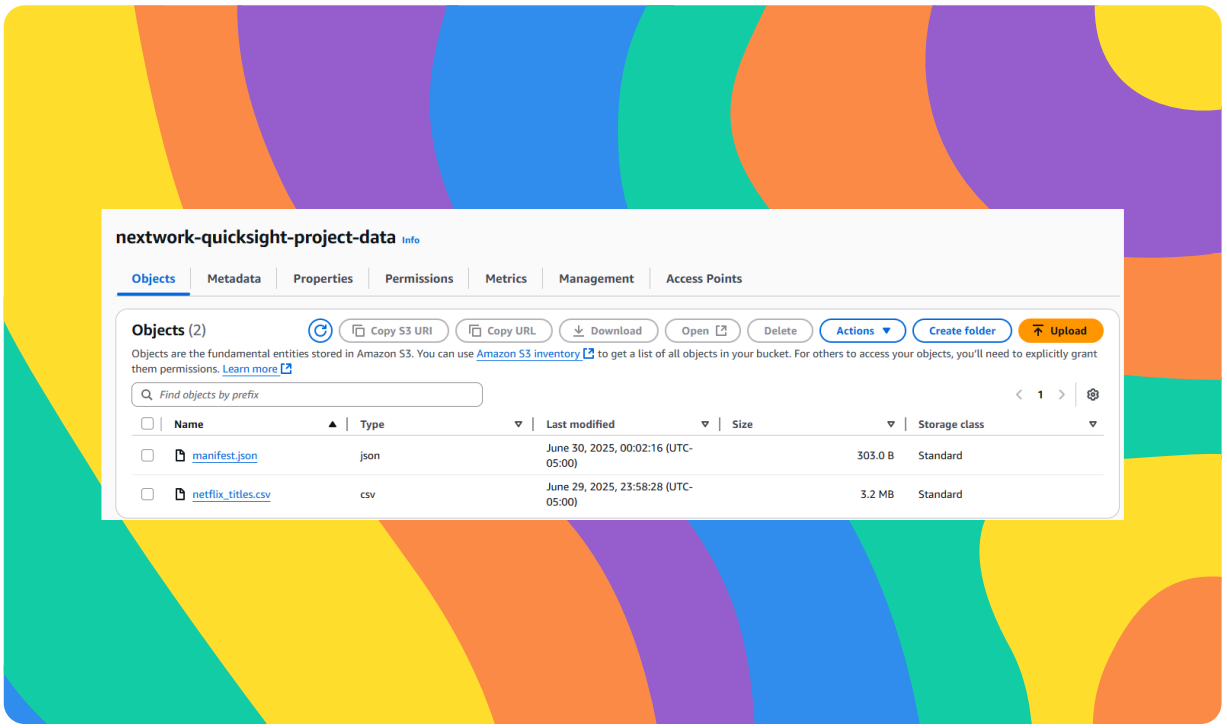
After this project, I plan to work on is Cloud Security with AWS IAM. I will start this project on right after completing this project.

CH
Arpita Chowdhury
NextWork Student

Upload project files into S3

S3 is used in this project to store two files, which are the netflix_titles.csv which is the dataset that is used for analyzing and visualization and manifest.json gives us the structure and the format of the data we are uploading.

I edited the manifest.json file by updating the S3 URL that corresponds to dataset file location It's important to edit this file because how Quicksight will find and analyze the data. If we didn't upload this file QuickSite wouldn't know the location of our dataset which will cause errors down the line.

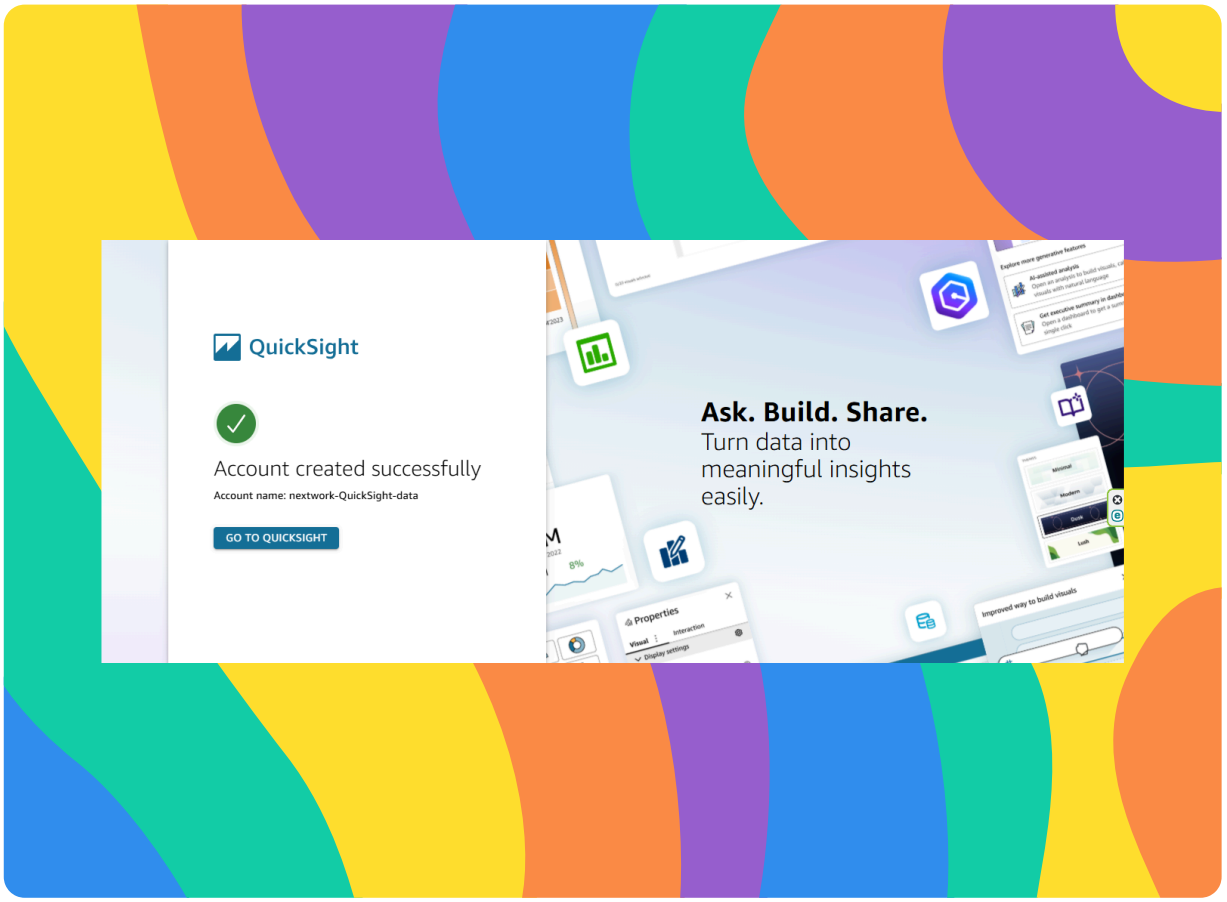


CH
Arpita Chowdhury
NextWork Student

Create QuickSight account

Creating a QuickSight account cost $0 as account gives a 30 day free trial. To do this make sure the uncheck the optional add-on option box beside the Add Pixel-Perfect-Report of $500.

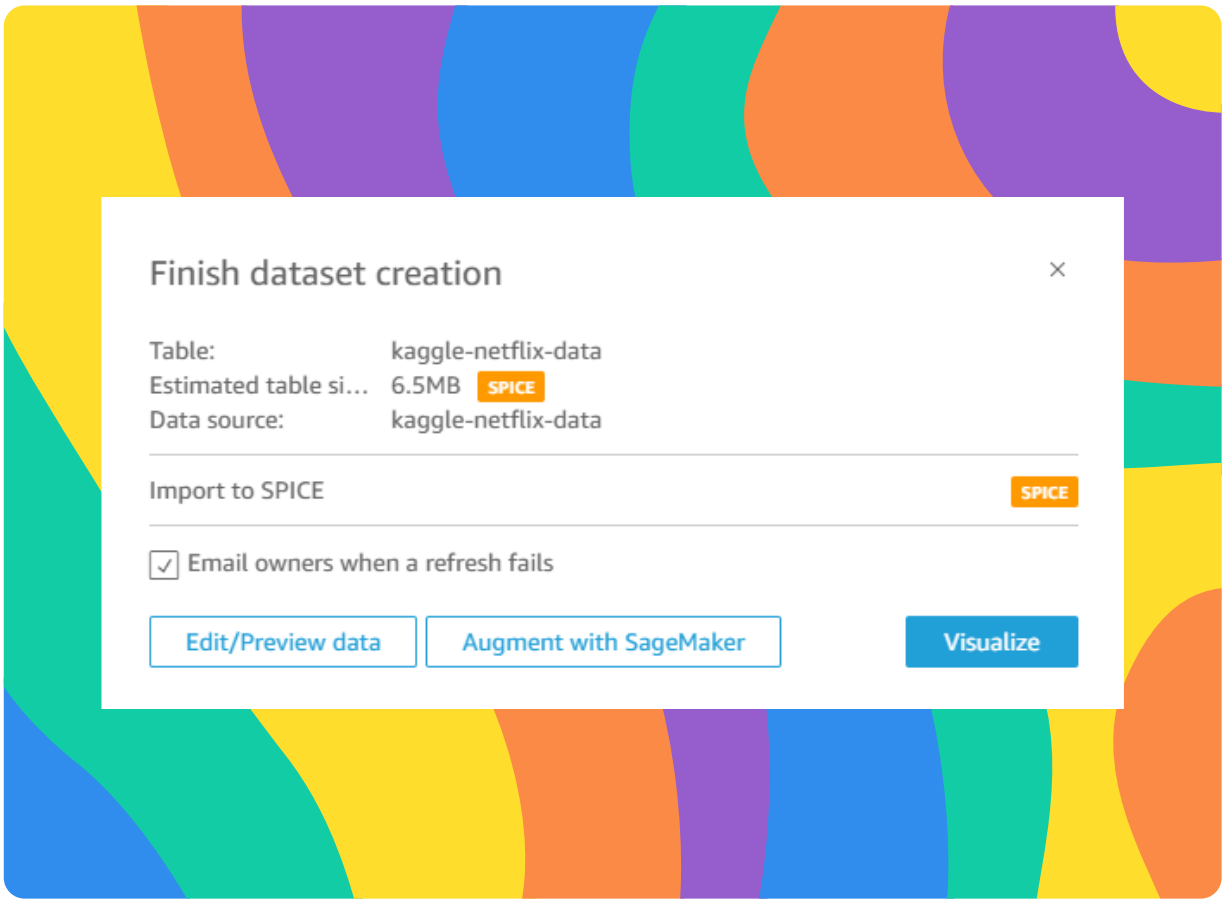
Creating an account took me around 3 minutes to set up the free version that included the Amazon S3 bucket services.



Download the Dataset

I connected the S3 bucket to QuickSight by visiting dataset page. I chose the Amazon S3 option.

The manifest.json file was important in this step because it tells QuickSight what the 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.



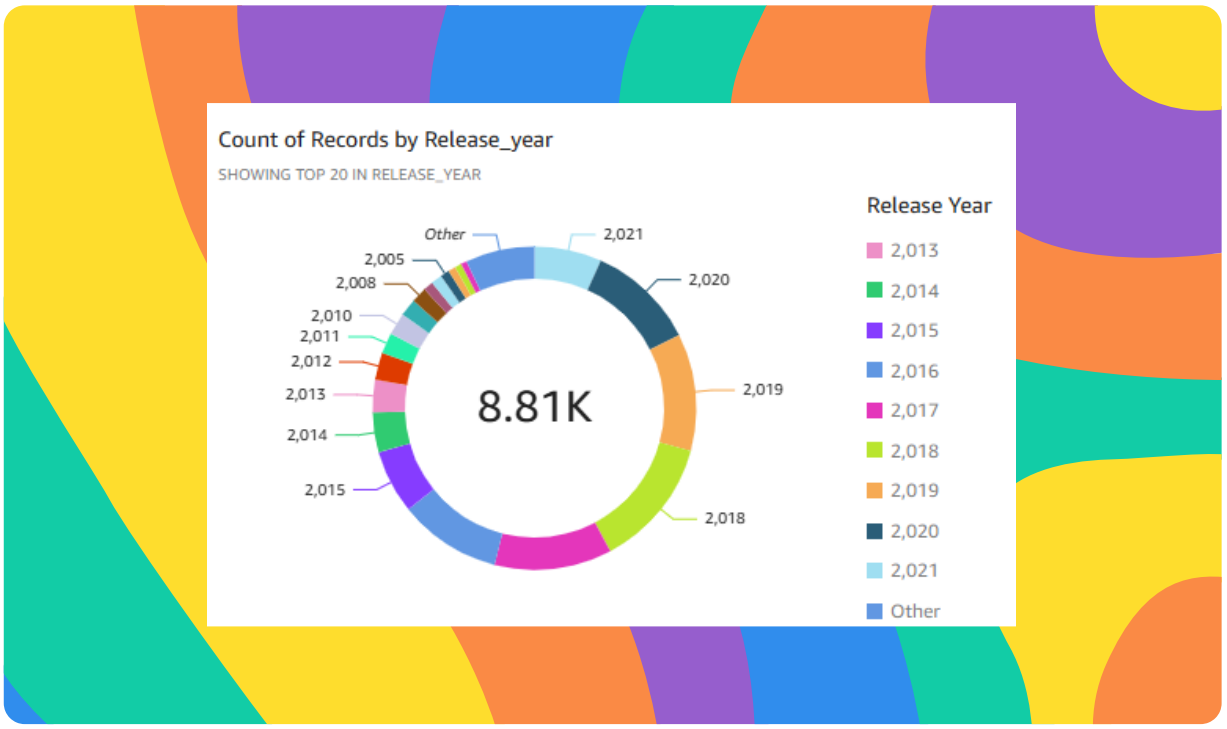
My first visualization

To create visualizations on QuickSight, I will simply click on data labels or field and quicksight will automatically generate a graphic that best suits that type of data. the data labels can also be dragged into section by group by or y-axis to determine how the graphic will trat the variable or data.

The chart/graph shown here is a breakdown of is a breakdown how much contents is released each year by netflix. The chart shows total of 8.81K data content with the biggest release year being 2018 with 1.15K content.

I created this graph by selecting the release year data lable from all the fields of dataset and changing the visual type to of donut chart from a bar chart.

CH
Arpita Chowdhury
NextWork Student

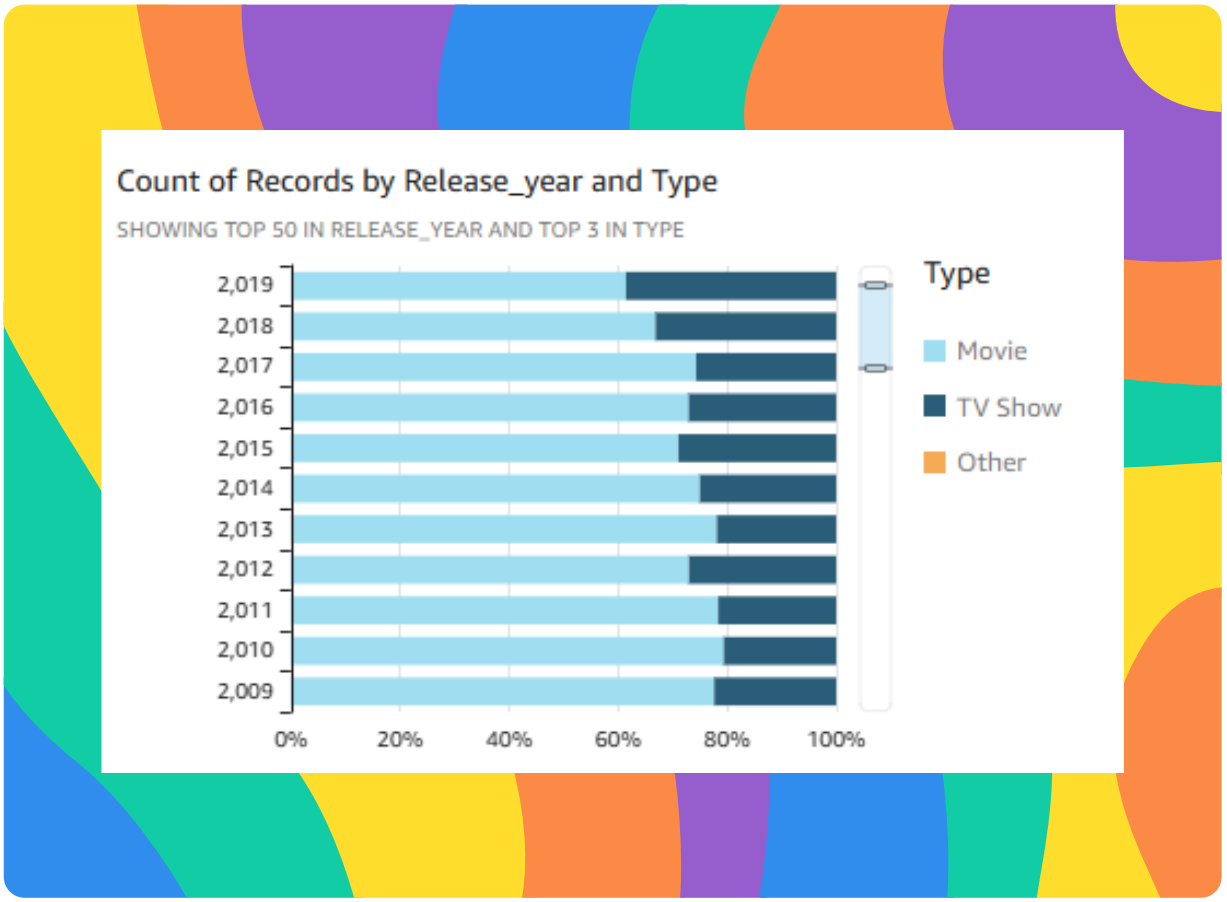


CH
Arpita Chowdhury
NextWork Student

Using filters

Filters are useful for narrowing down data to the subset we are focusing on the specific category to analyze.

This visualization is a breakdown of TV shows and movies that belong in one of three categories TV Comedies, Action and Thiller. Here I added a filter based listed_in data at lable.

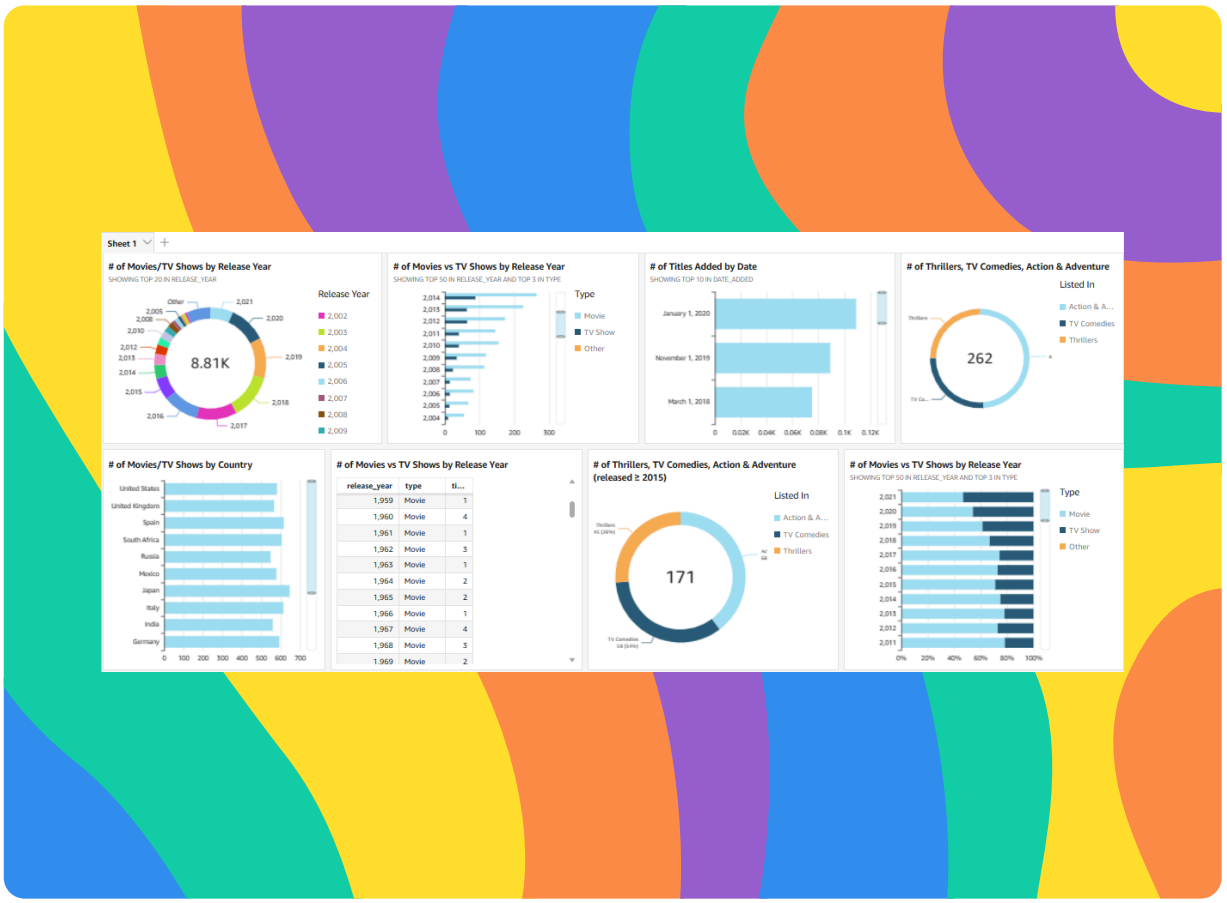


CH
Arpita Chowdhury
NextWork Student

Setting up a dashboard

As a finishing touch, I updated the title of the chart of the dashboard to make it easy readable. The default name just mentined the datalable used for analysis and purpose of chart.

Did you know you could export your dashboard as PDFs too? I did this by selescting publish and then select generate PDF after that select export to export and lastly download the PDF.



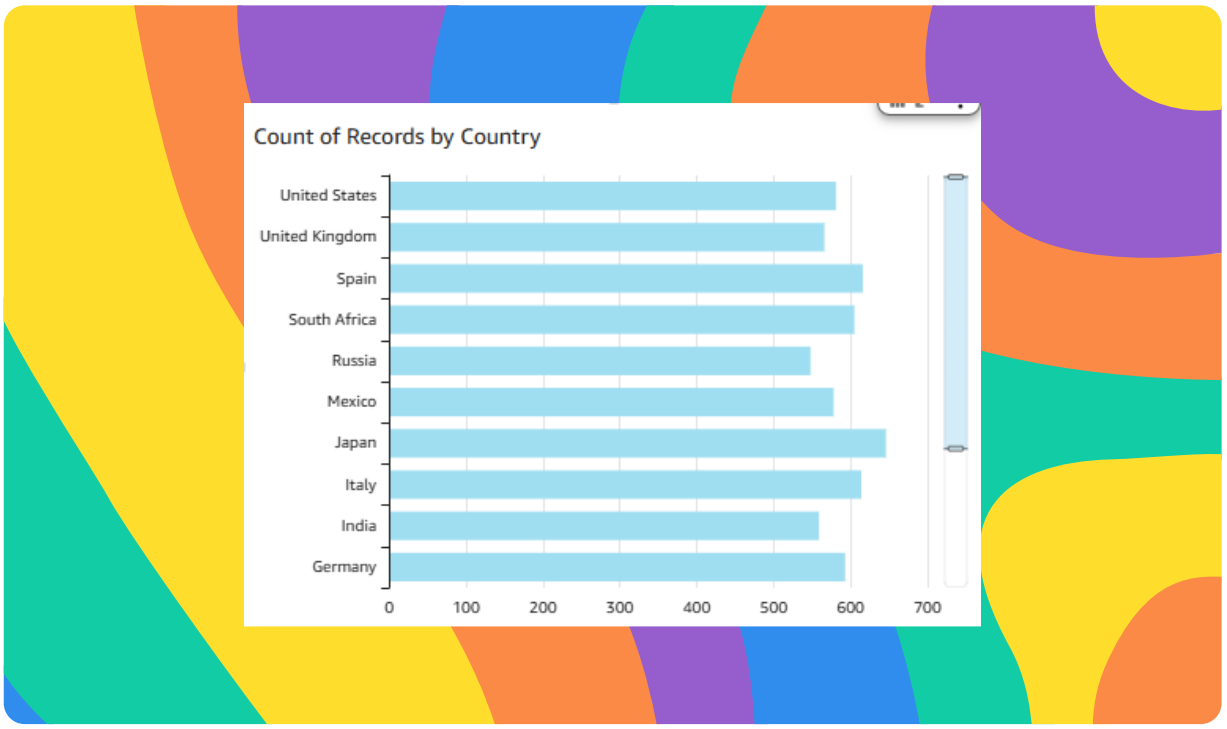
Refreshing source data

In this project's extension, I downloaded fresh data that's different from my original dataset because the original data had missing avlues for countries field where as the new dataset do not. Analysing incomplete data brings the risk of having decisions made based on incomplete data are more likely to be flawed because they're based on data that might not reflect the true scenario.

Once I downloaded new data, I had to update my S3 bucket because it was still storing previous dataset where country data was missing I also uploaded a new manifest.json file that points to updated dataset name to ensure that the updated dataset will be used by QuickSite.

I initally couldn't see my updated data in QuickSight, so I had to go to data and select dataset option where I clicked the edit option for the dataset that was already present then I selected the query modes refresh now option and refreshed the dataset after doing that I clicked save and publish.

CH
Arpita Chowdhury
NextWork Student



nextwork.org

The place to learn & showcase your skills

Check out nextwork.org for more projects