

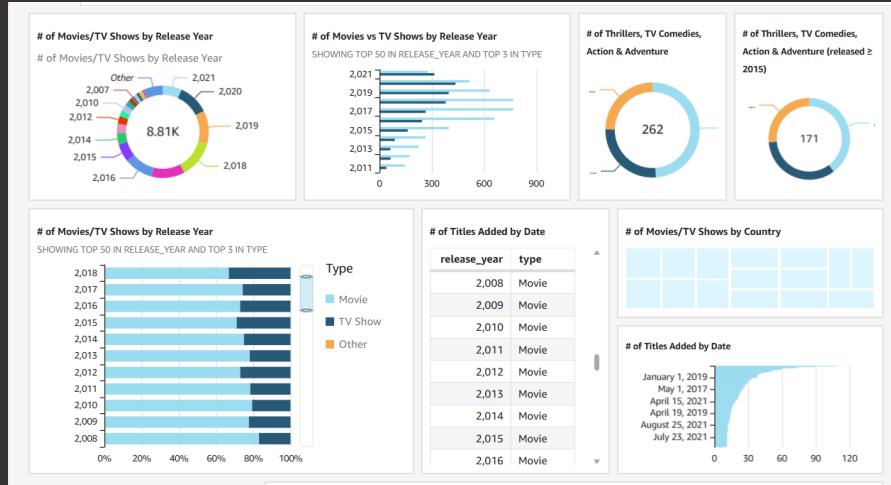


nextwork.org

Visualize data with QuickSight



Joba Amunigun





Joba Amunigun
NextWork Student

nextwork.org

Introducing Today's Project!

In this project, I will demonstrate how to use Amazon QuickSight to analyze data and generate visualizations and insights from a Netflix dataset. I'm doing this project to learn how to use data cloud services for data analysis

Tools and concepts

Services I used were QuickSight and Amazon S3. Key concepts I learnt includes editing in manifest.json files ,data visualization techniques (eg. charts, filters) and how to perform a data refresh in QuickSight

Project reflection

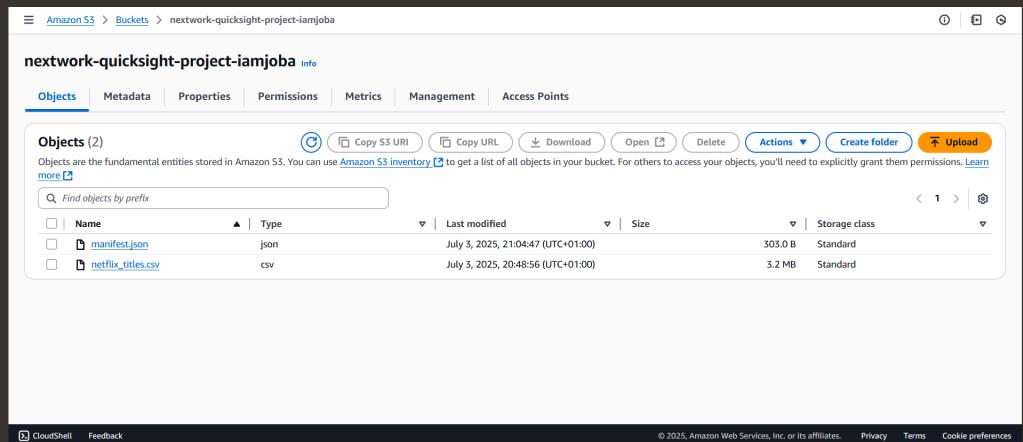
This project took me approximately 2 hours + .The most challenging part was connecting my S3 datasets on QuickSight. It was most rewarding to generate a PDF of my finished visualizations

After this project, I plan to work on Day 3 of thw AWS Challenge. I will start this project tommorow and it's about Cloud Security

Upload project files into S3

S3 is used in this project to store two files, which are manifest.json (which tells QuickSight about the format and structure of the data i am analyzing) and netflix_titles (which is the raw data that i am going to be analyzing)

I edited the manifest.json file by updating S3 URL that corresponded with my dataset's file location. It's important to edit this file because it's how QuickSight will find and analyze the data



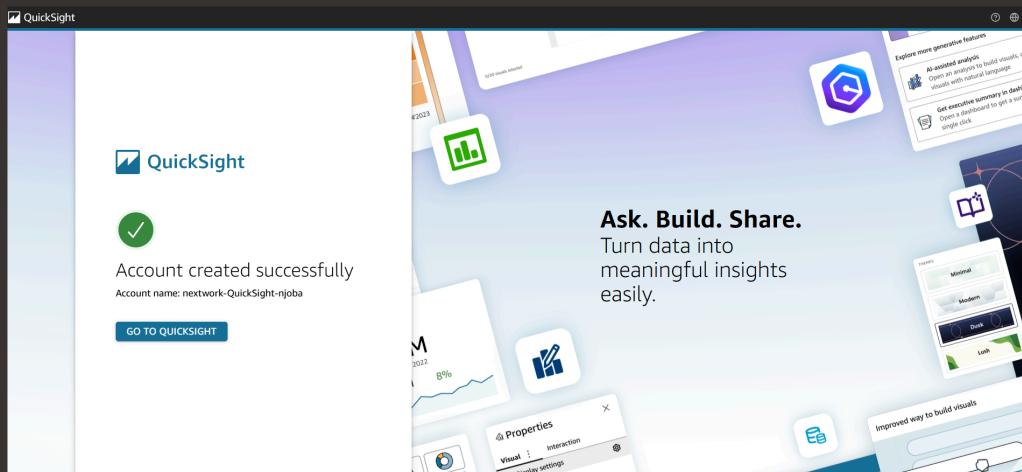
Joba Amunigun
NextWork Student

nextwork.org

Create QuickSight account

Creating a QuickSight account cost \$0 as it comes with a 30 days free trial. I will make sure to UNCHECK the add-on in the sign up flow called Pixel-Perfect Reports so that i don't get charged

Creating an account took me about 10-15 minutes including setting up my bucket permissions

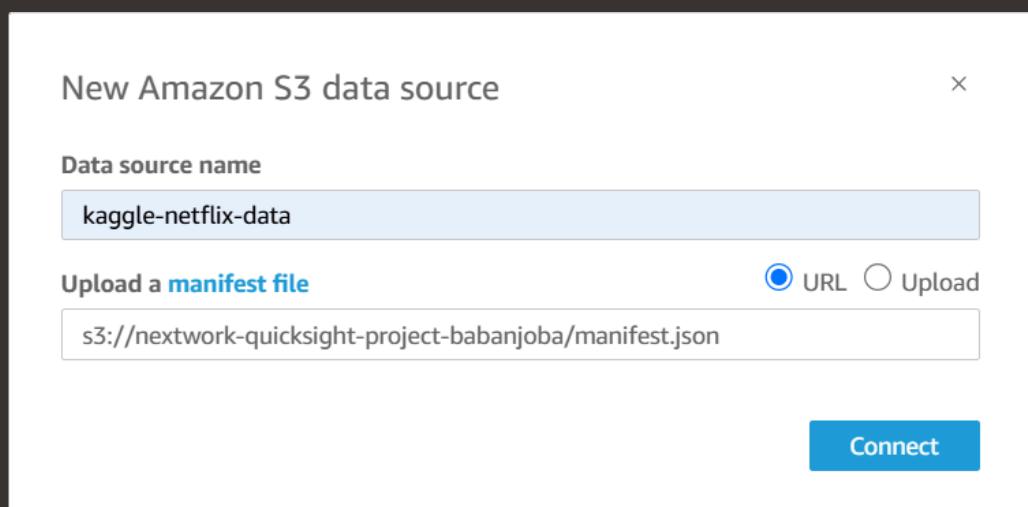




Download the Dataset

I connected the S3 bucket to QuickSight by visiting the datasets page. There were so many options for data sources i could connect to (databases and external tools/ platforms like salesforce) and i selected Amazon S3 because i uploaded my datasets in S3

The manifest.json file was important in this step because it tells QuickSight how to read the data - In this project it tells QuickSight that i uploaded a CSV file (spreadsheet) and the delimiters(i.e. commas) so that QuickSight knows how to break up the data for analysis .Otherwise, QuickSight might get confused and not show the data correctly





My first visualization

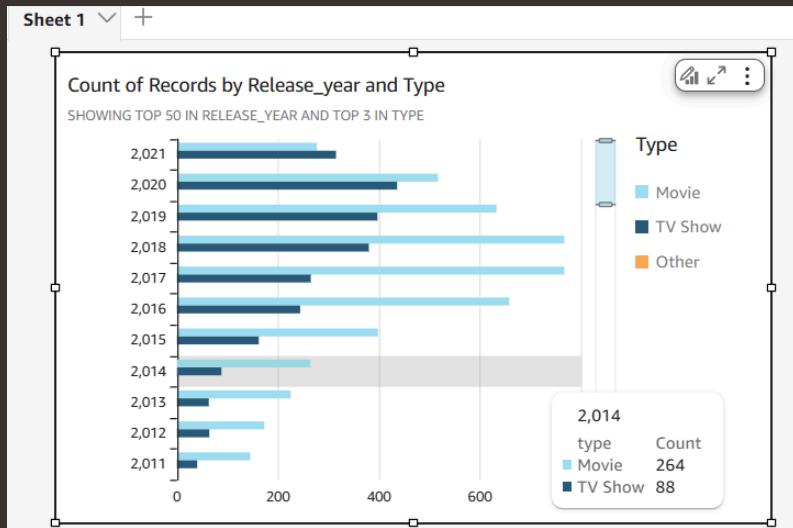
To create visualizations on QuickSight, I simply had to click on the data field (e.g release_year) and QuickSight automatically generates a visualization that best suits the data. I also dragged the data into the "Y Axis" heading to determine how my graphics should treat the data.

The graph shown here is a breakdown of contents inside Netflix i.e How many TV shows/movies is released On XYZ year ? You can see that there is a total of 8.8K+ content pieces and 2018 is the year that has the most content amount released

I created this graph by dragging and dropping the release_year label into the Y Axis heading and dragging the type label into the Group/Color heading

Joba Amunigun
NextWork Student

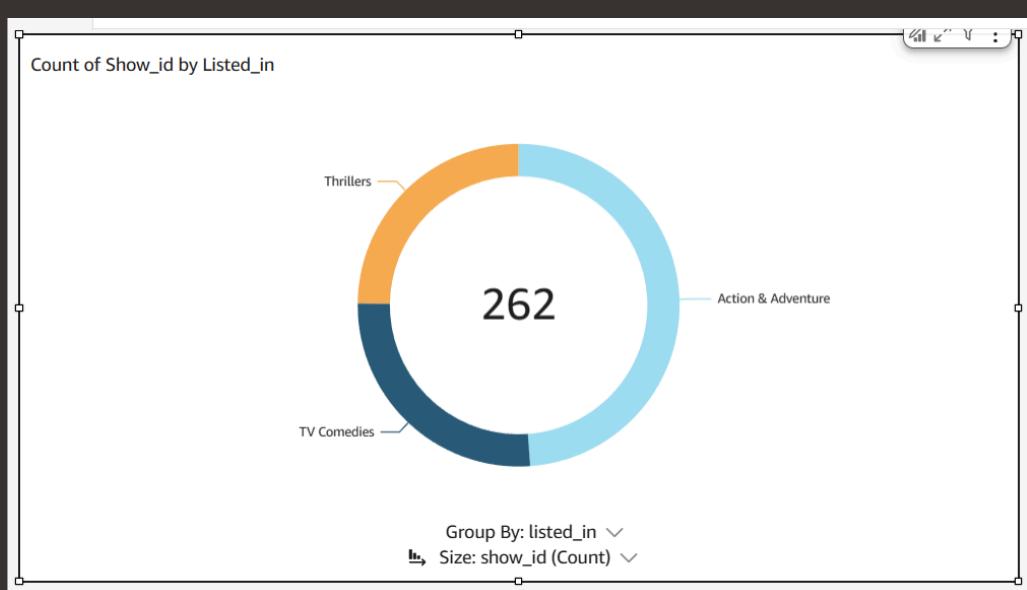
nextwork.org



Using filters

Filters are useful for narrowing down our data to the subsets that we want to focus on and in this case i used filters to focus on the specific categories that i want to analyze. For example, I used filters to only look at the release date from 2015 and beyond

This visualization is a breakdown of TV shows /Movies that belong in one of the three categories - 'Action & Adventure', 'TV Comedies', here 'Thrillers'. Here I added a filter based on the "listed -on" data label so that only these categories could pass the filter



Setting up a dashboard

As a finishing touch, I updated the titles of the charts on my dashboard so that they are easily readable -Compared to the old titles the default names will simply mention the data labels e.g. released_year but the new titles communicates the purpose of the charts e.g. # of Thrillers, TV Comedies, Action & Adventure

Did you know you could export your dashboard as PDFs too? I did this by selecting "publish" then "generate PDF" from the top right corner of the QuickSight Analysis





Refreshing source data

In this project's extension, I downloaded fresh data that's different from my original dataset because it had rows and rows of empty country data- Analyzing incomplete data brings the risk of generating wrong insights which can lead to wrong decisions that will cost the company time ,effort and money

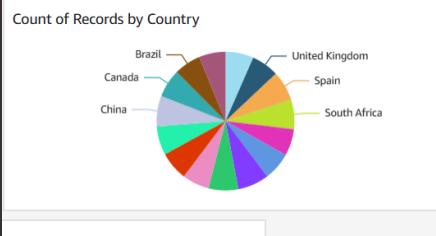
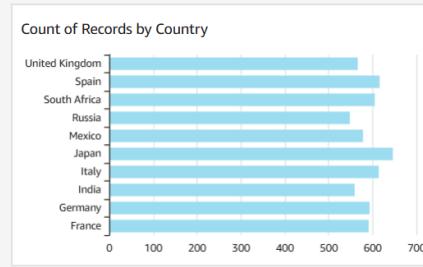
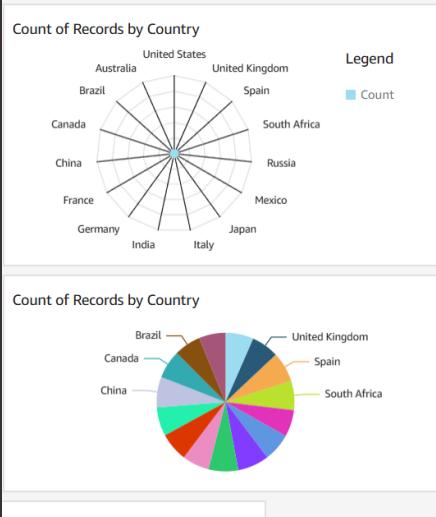
Once I downloaded new data, I had to update my S3 bucket because it is still storing the previous version of the data(i.e Column F containing the Country data). I also uploaded a new manifest.json file that points to my updated datasets name ("s3://nextwork-quicksight-project-babanjoba/netflix_titles_updated.csv) . This is to make sure that QuickSight is now pulling data from the new updated data

I initially couldn't see my updated data in QuickSight, so I had to visit the datasets page in Quicksight and perform a full reset of my data by simply refreshing - and Viola! my visualizations were drawing from the updated data



Joba Amunigun
NextWork Student

nextwork.org





nextwork.org

The place to learn & showcase your skills

Check out nextwork.org for more projects

