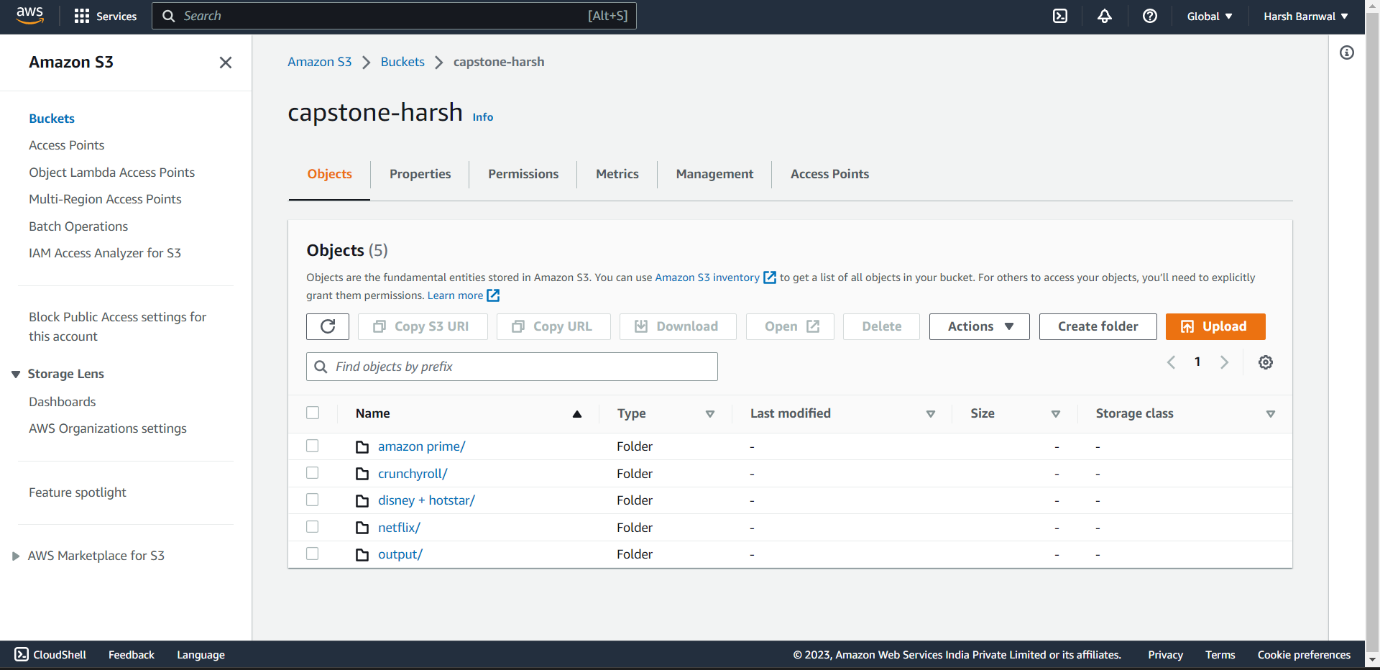
I have a data folder which consists of multiple datasets related to TV shows & movies across different OTT platforms.

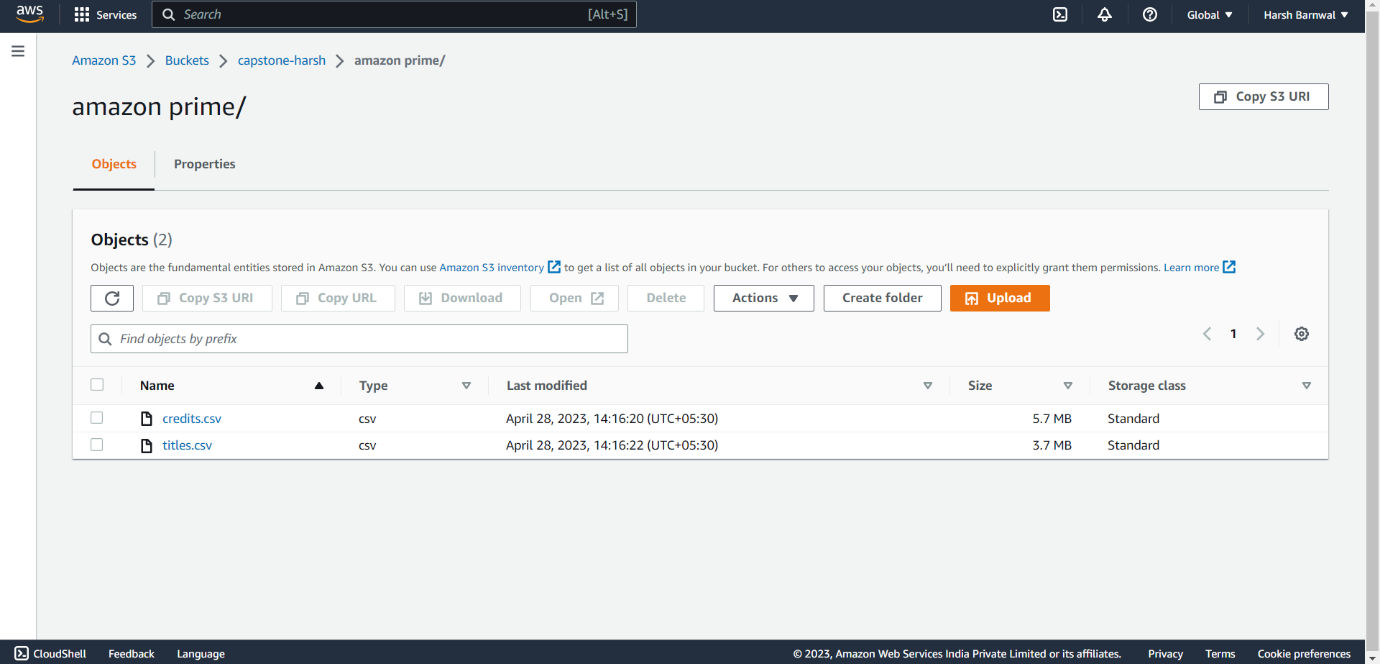
**Objective of Capstone Project –**

1. Stored all these data files in an AWS S3 bucket. The entire data profiling is done to understand the trends in data. Built QC checklist based on the data profiling done and created data catalogs using glue.
2. Built custom data processing scripts to clean, process & aggregate data. The script sends failure alerts report over email.
3. The processed can be loaded in S3 to be directly consumed by Tableau. A live dashboard is built to show the various trends across different OTT platforms.

The entire infrastructure is developed on AWS and visualization tool used is Tableau.

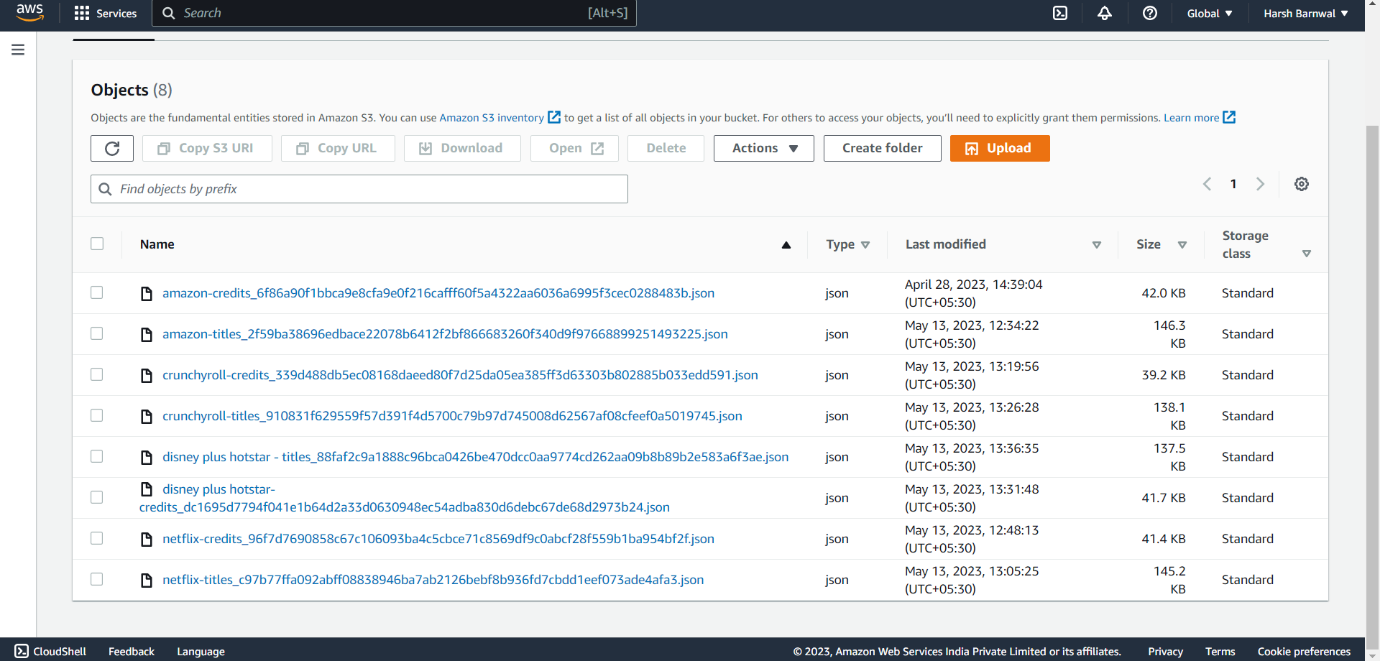
**1) Storing all the data files in an AWS S3 bucket –**

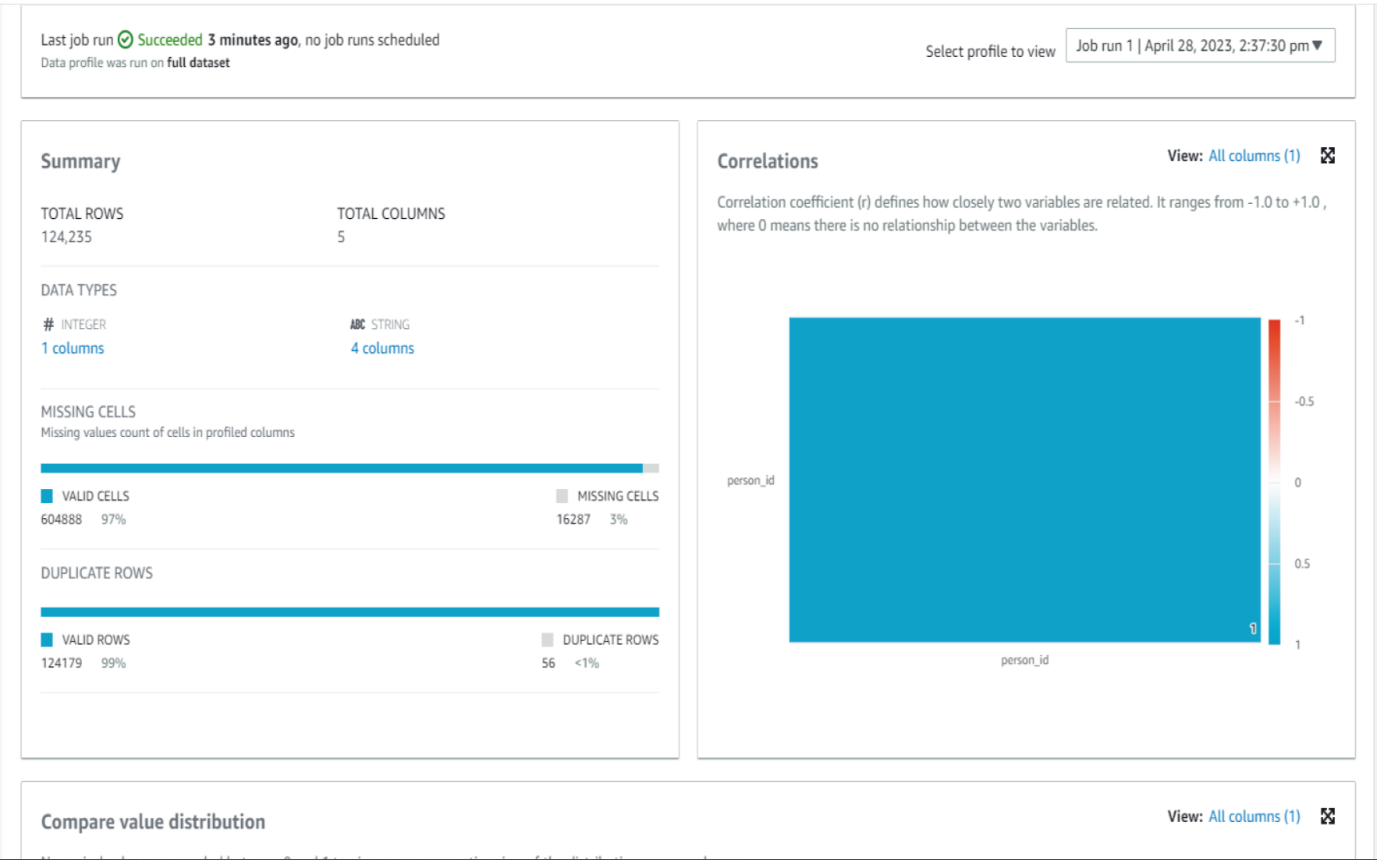


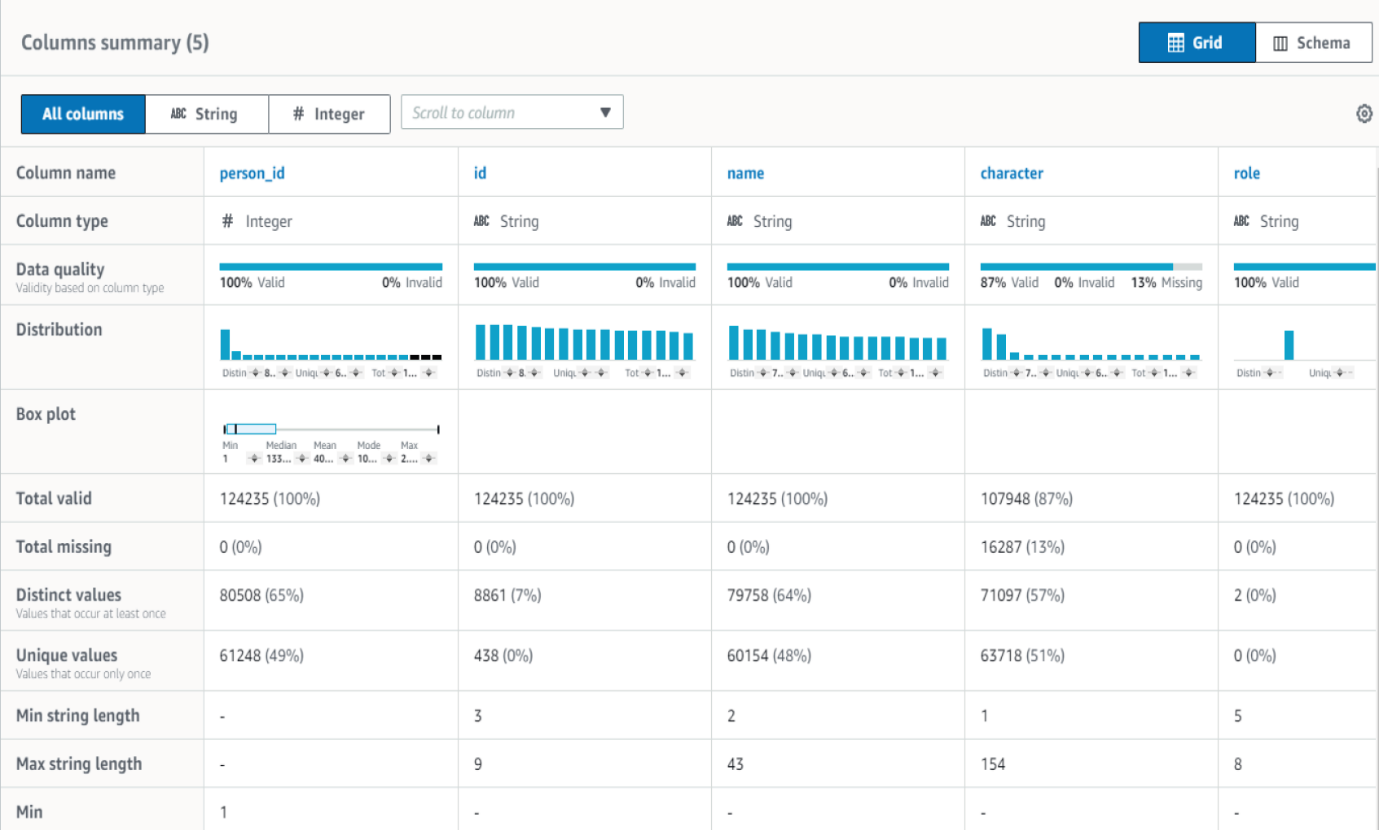


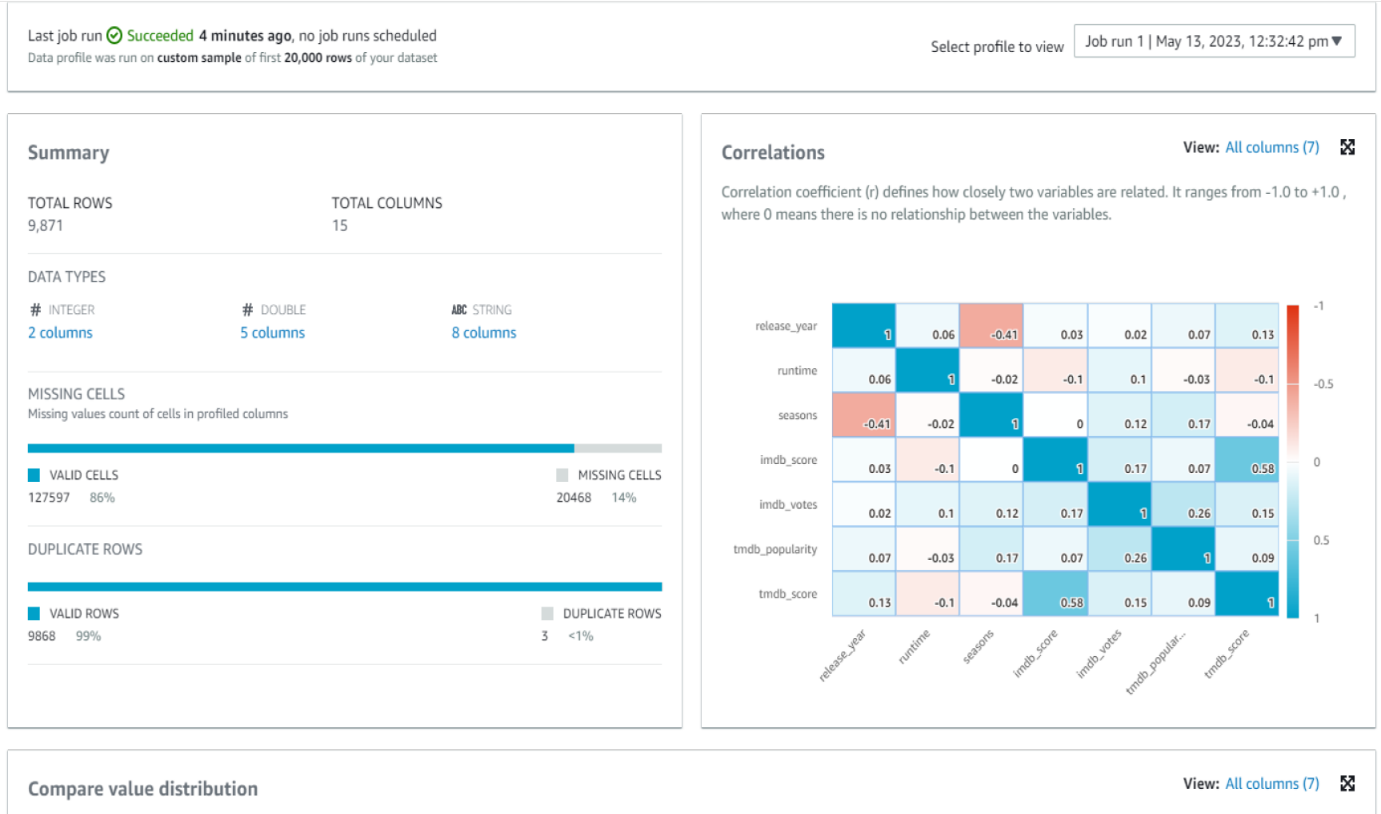
**Data profiling is done in AWS Glue DataBrew for all datasets and their output is received both as pdf and JSON file –**

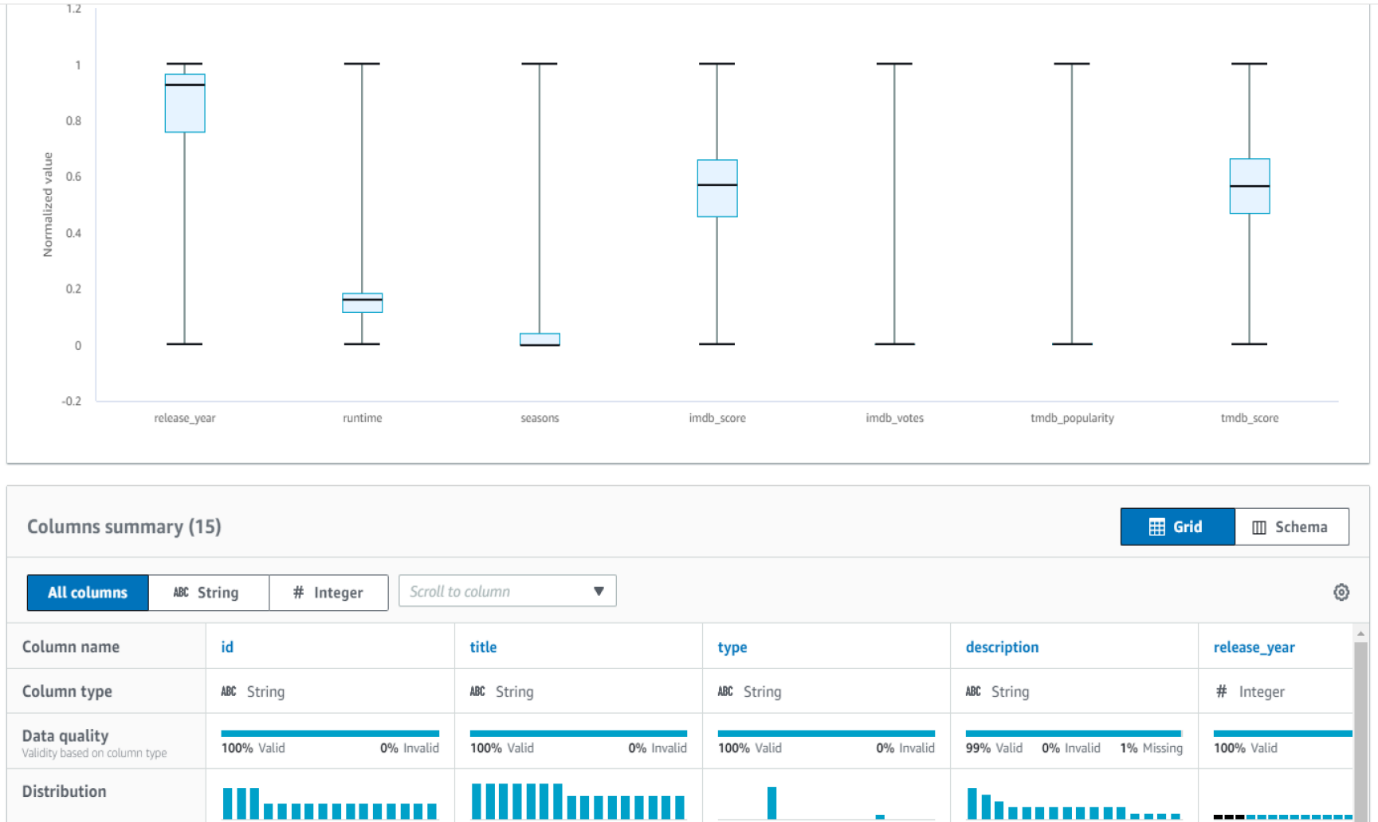






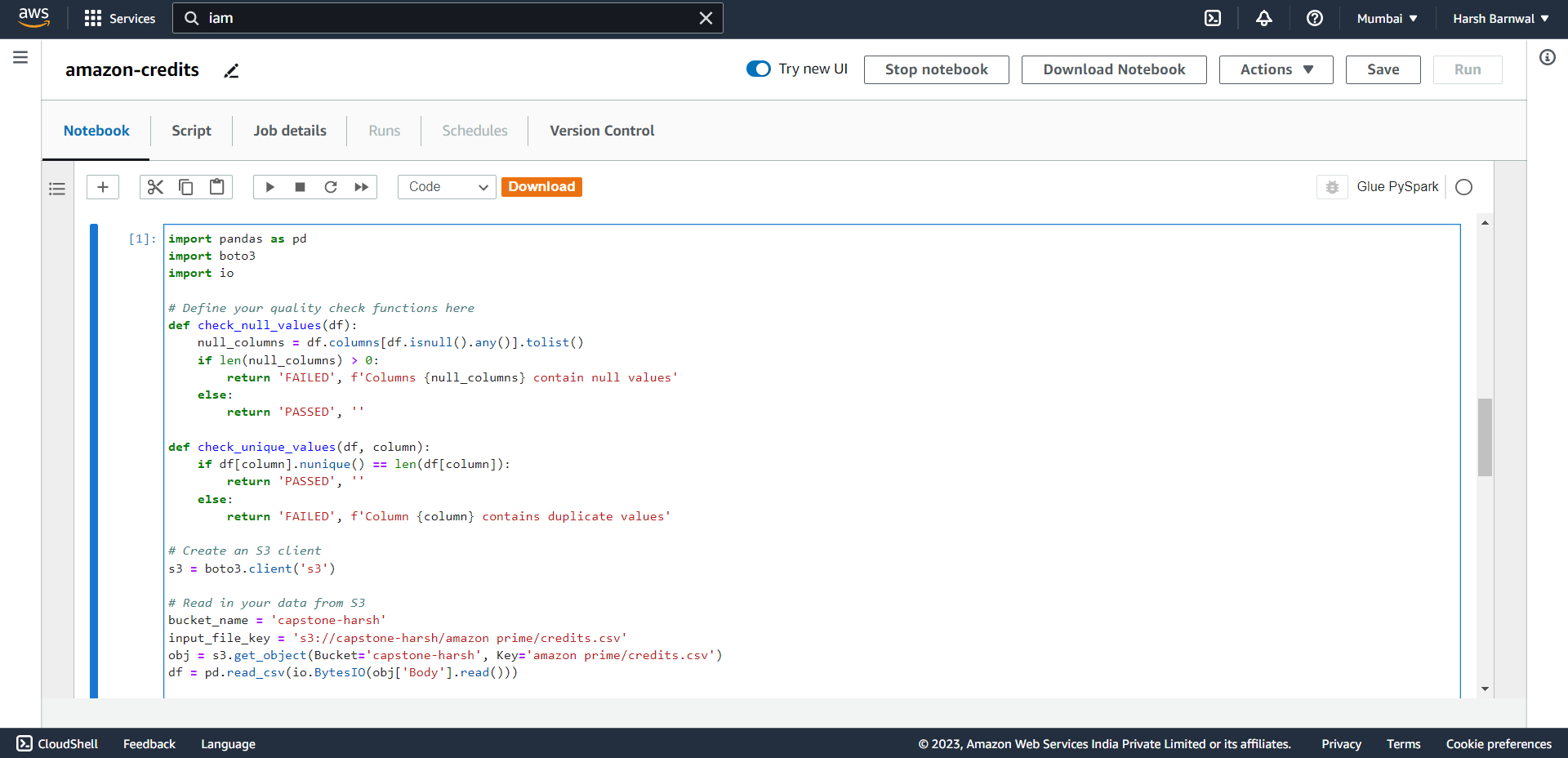


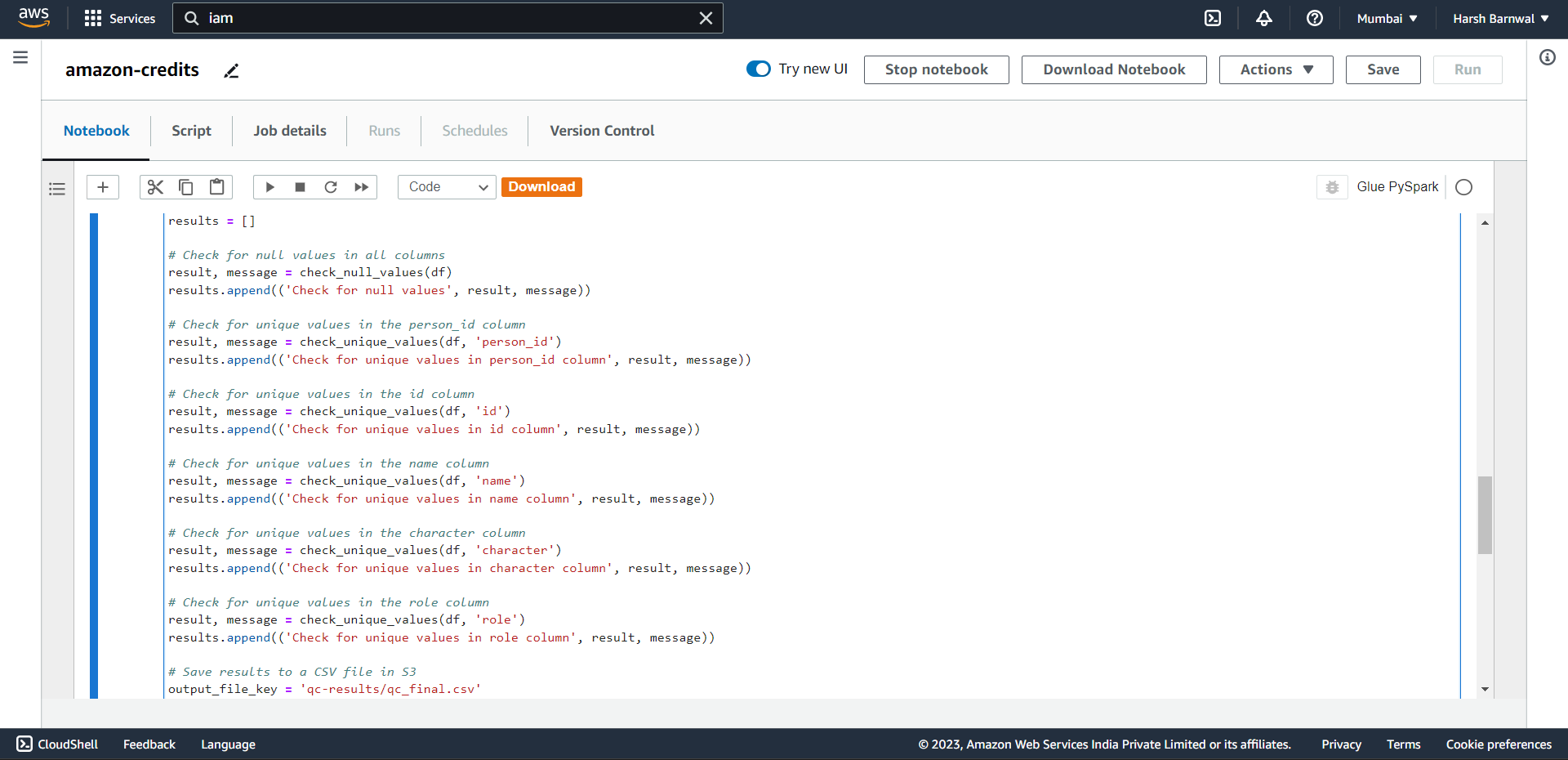


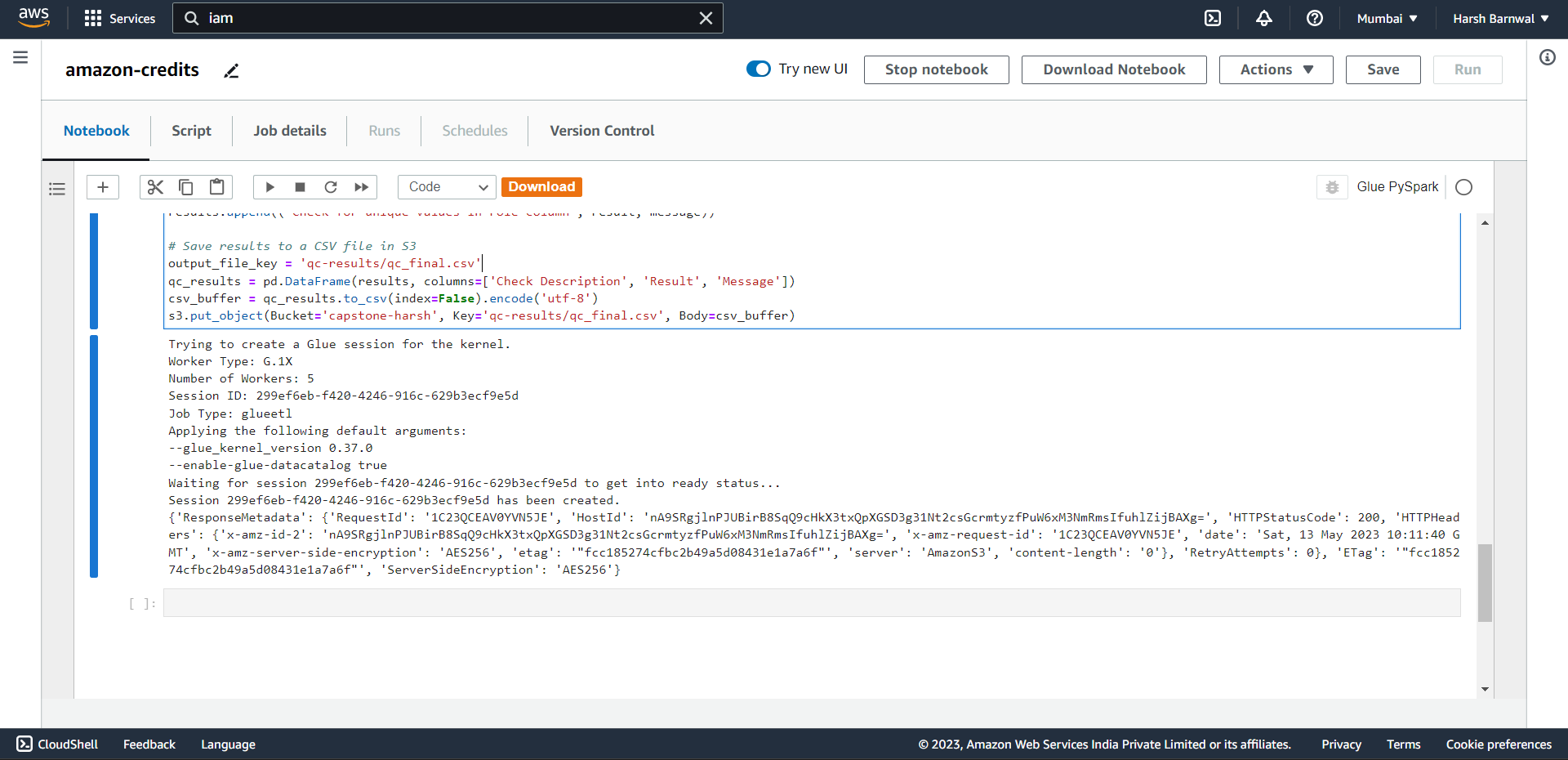


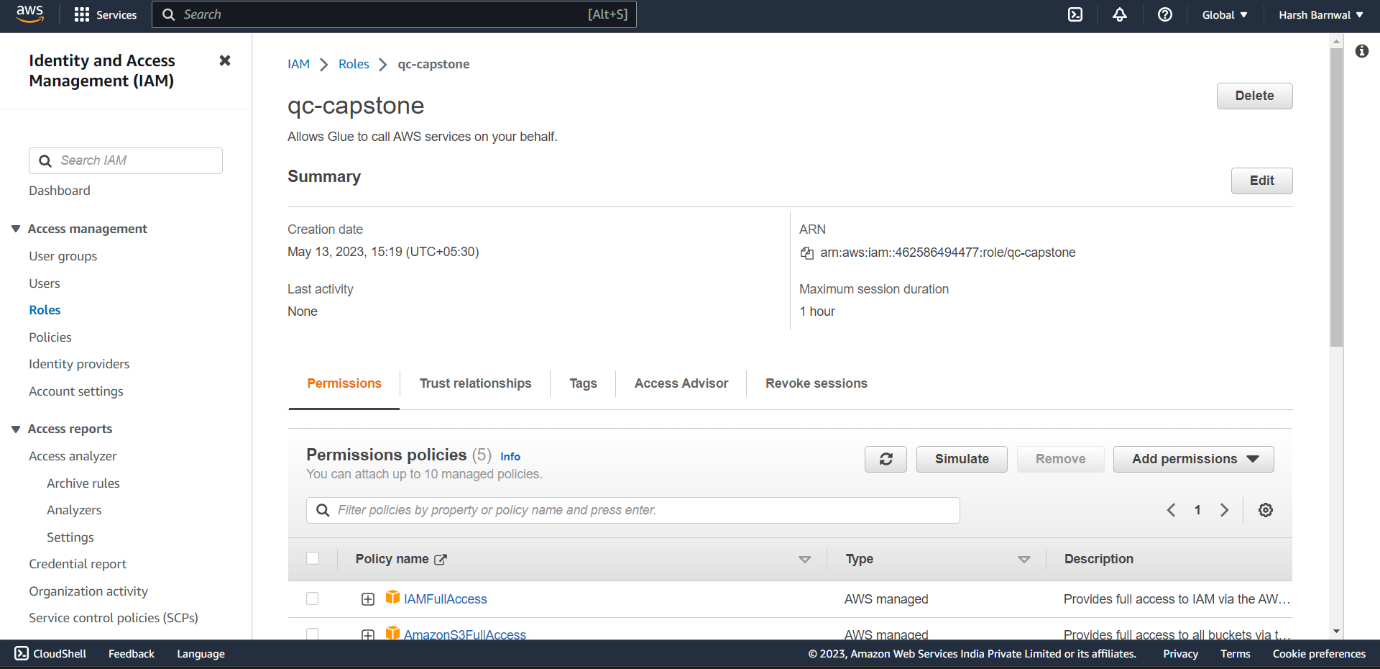
**Building QC checklist using Jupyter Script in AWS Glue and creating a job. The data file is taken and the results are received in a csv file, showcasing the findings.**

**Meanwhile, an IAM role is created with appropriate permissions and 1 inline permission, with proper JSON code to verify the technicalities.**

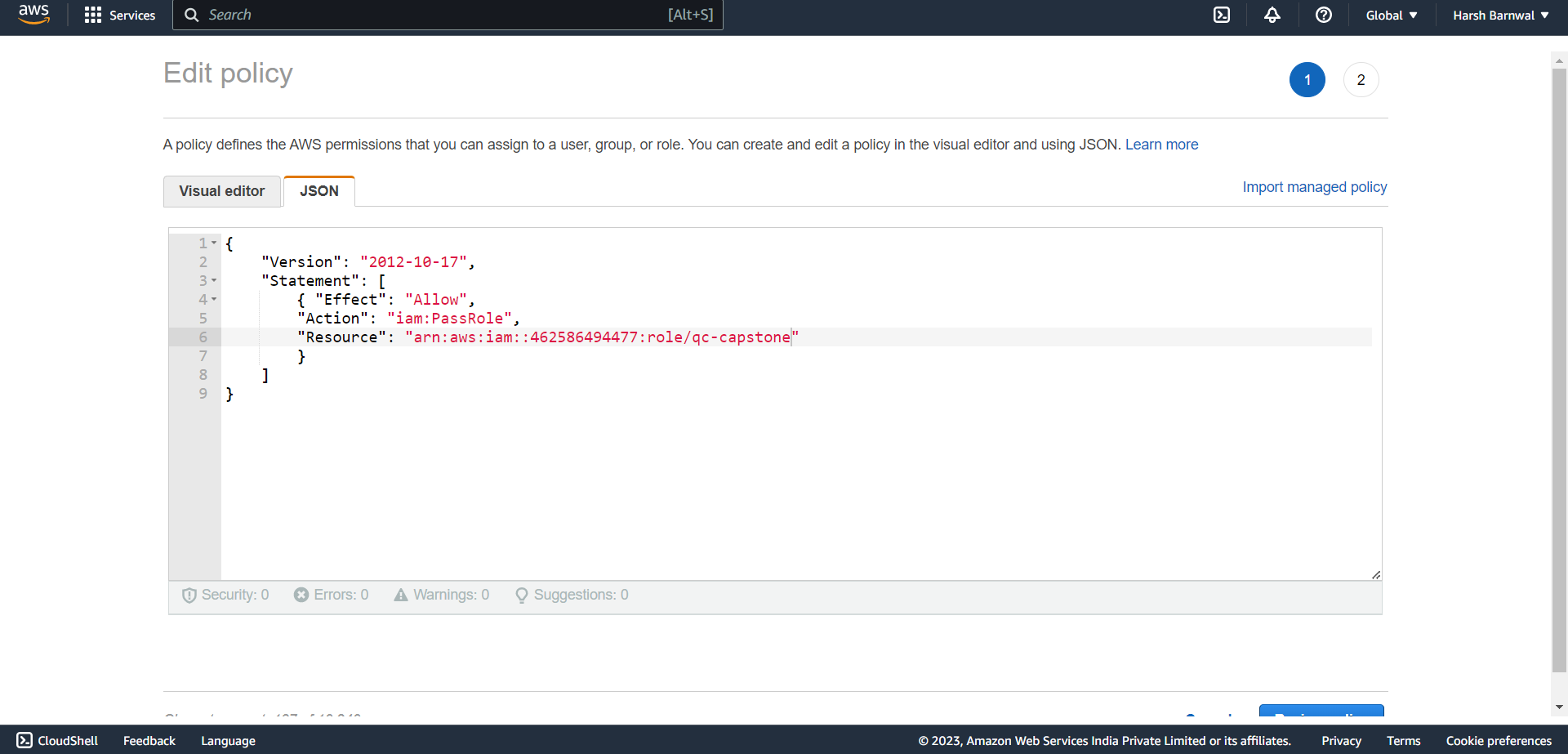


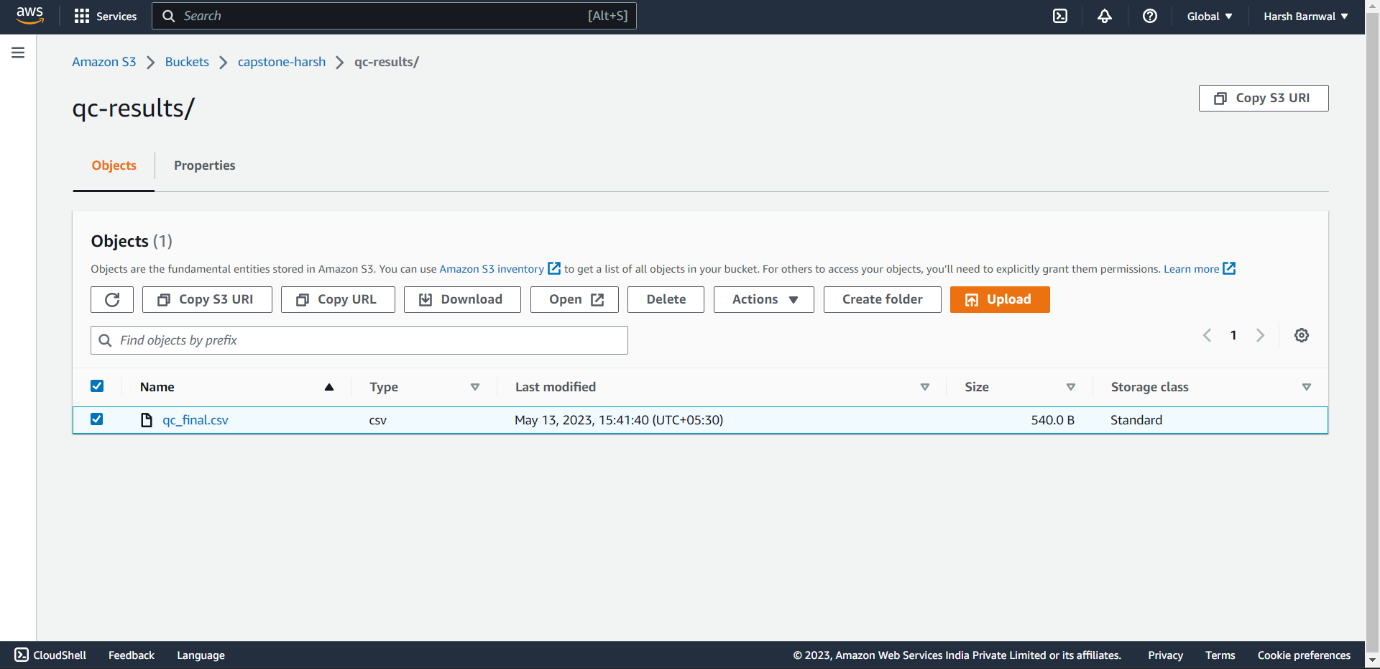


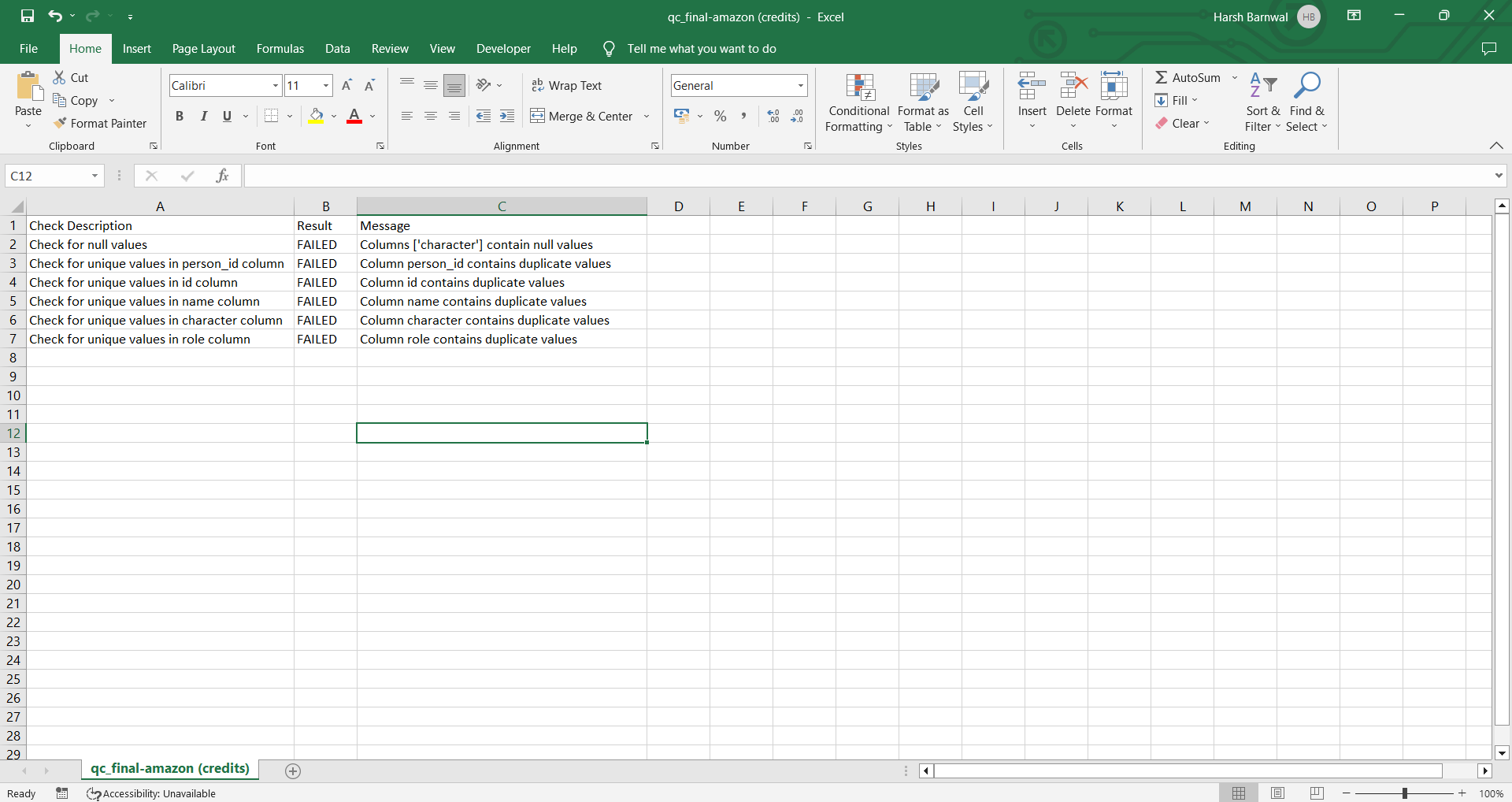


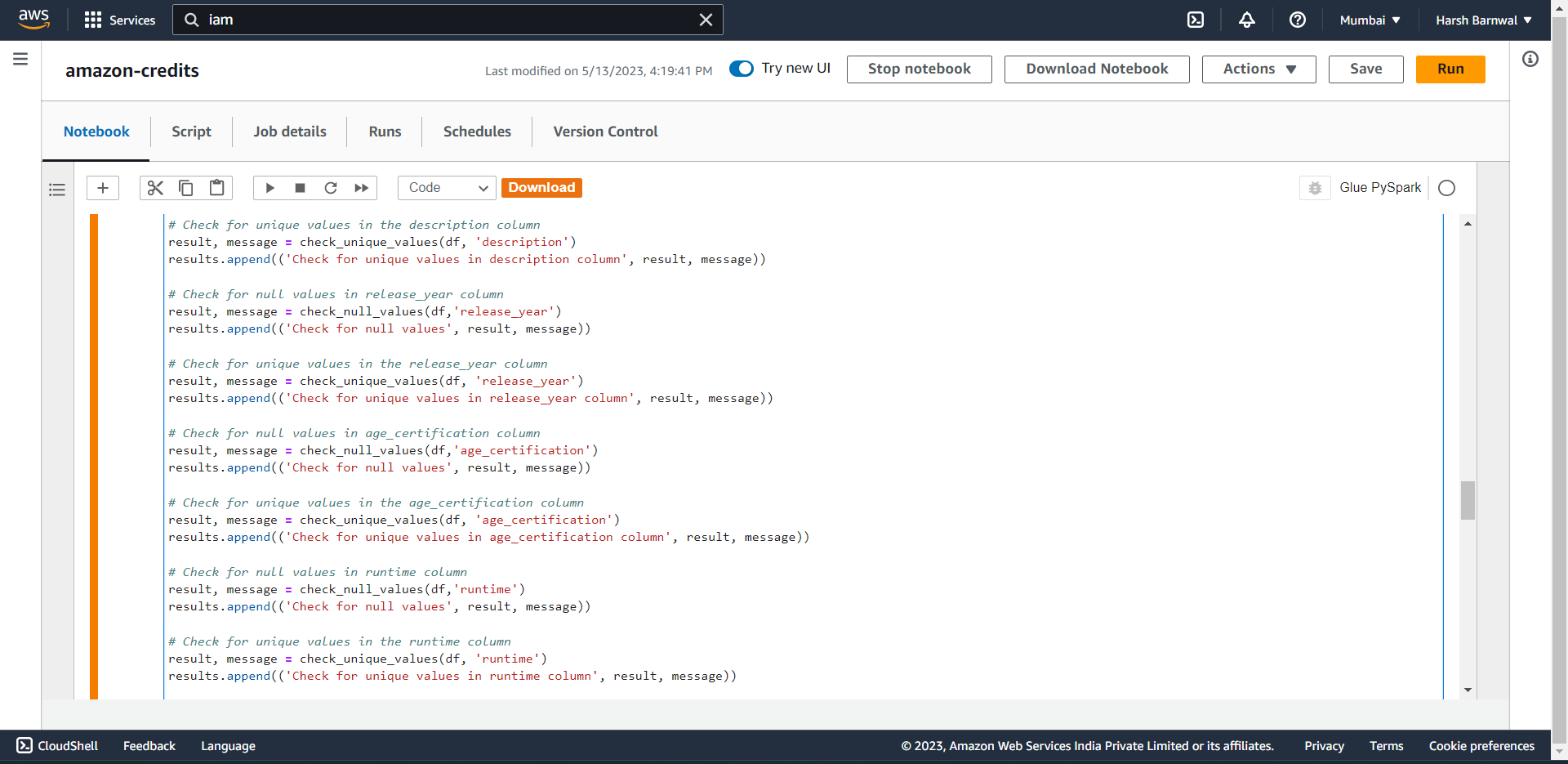
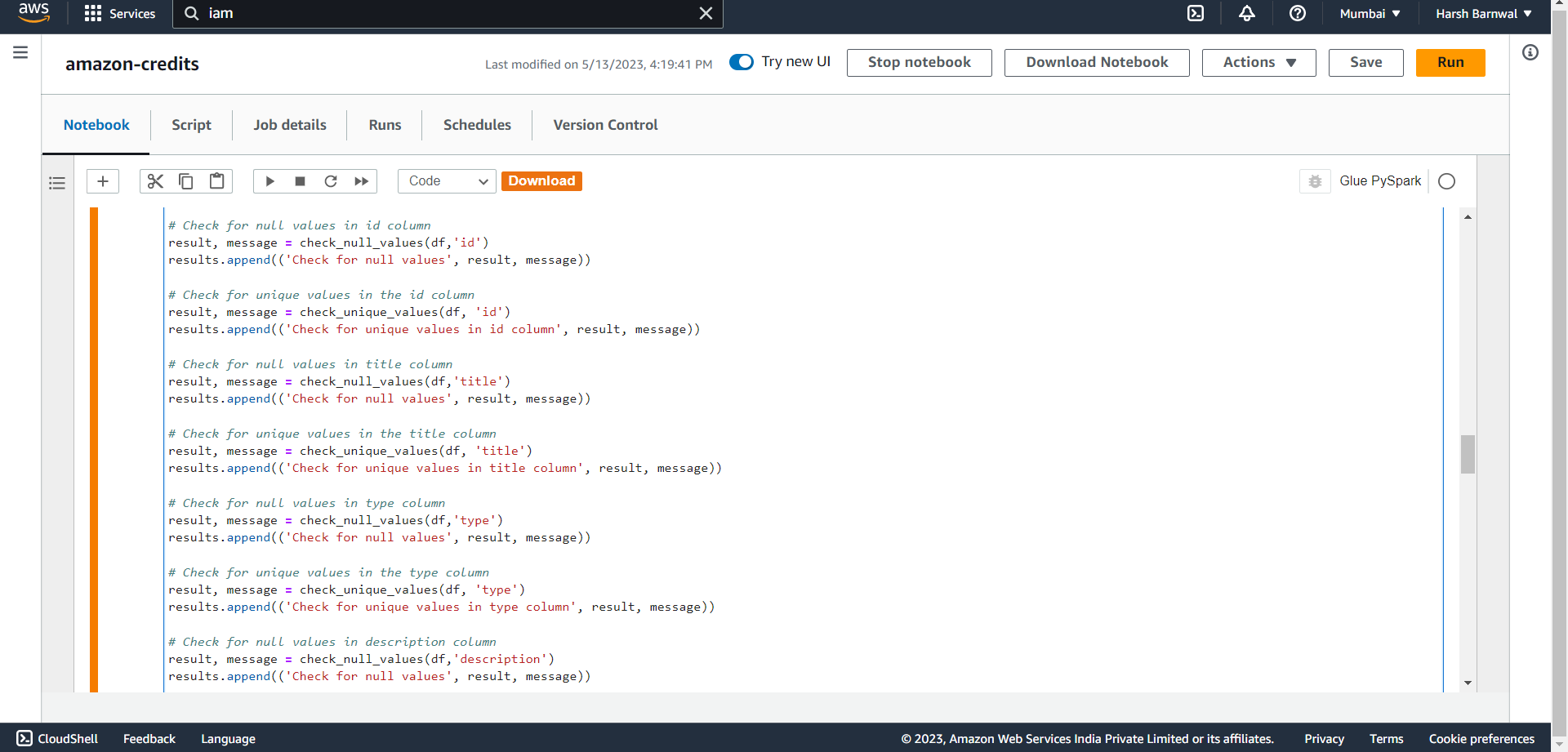
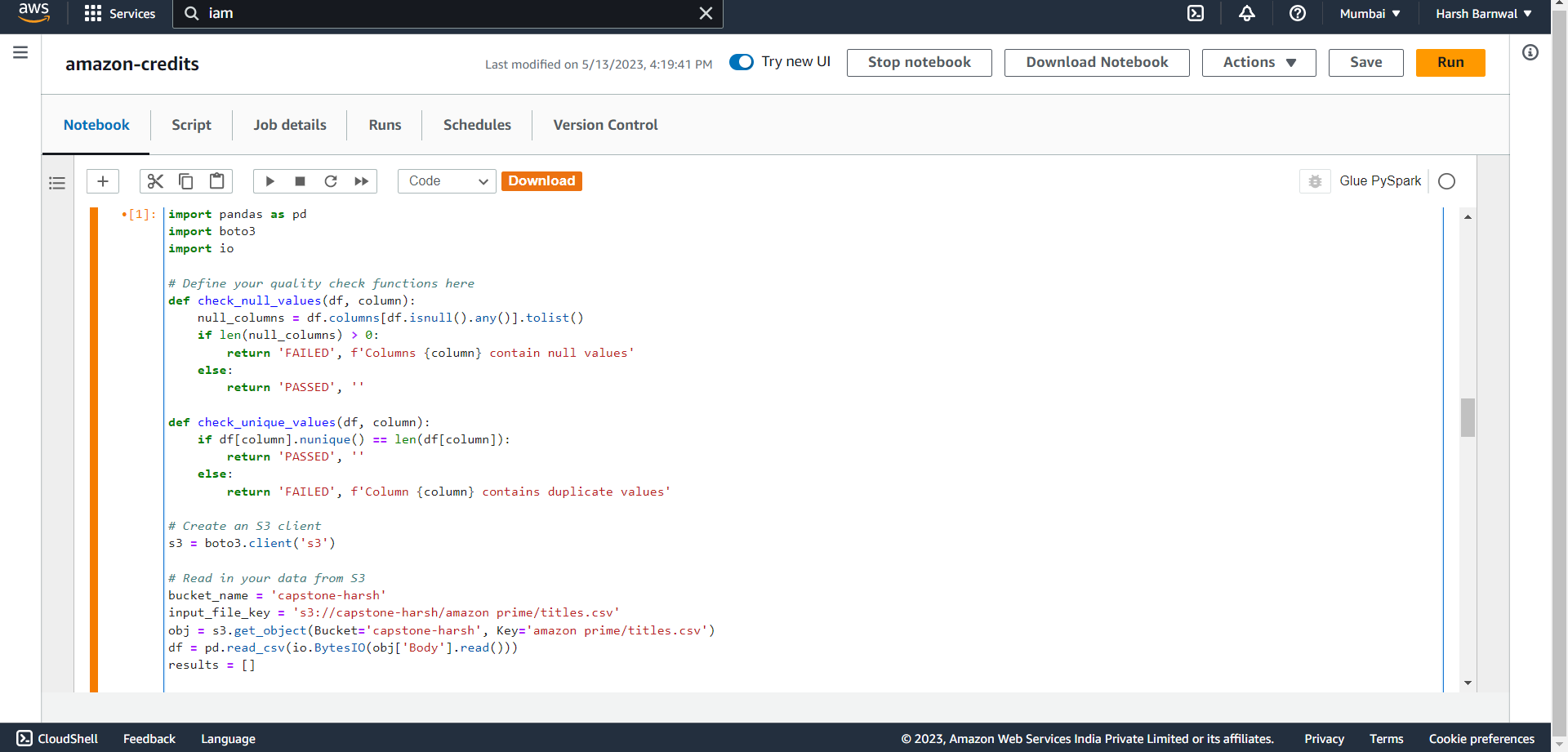


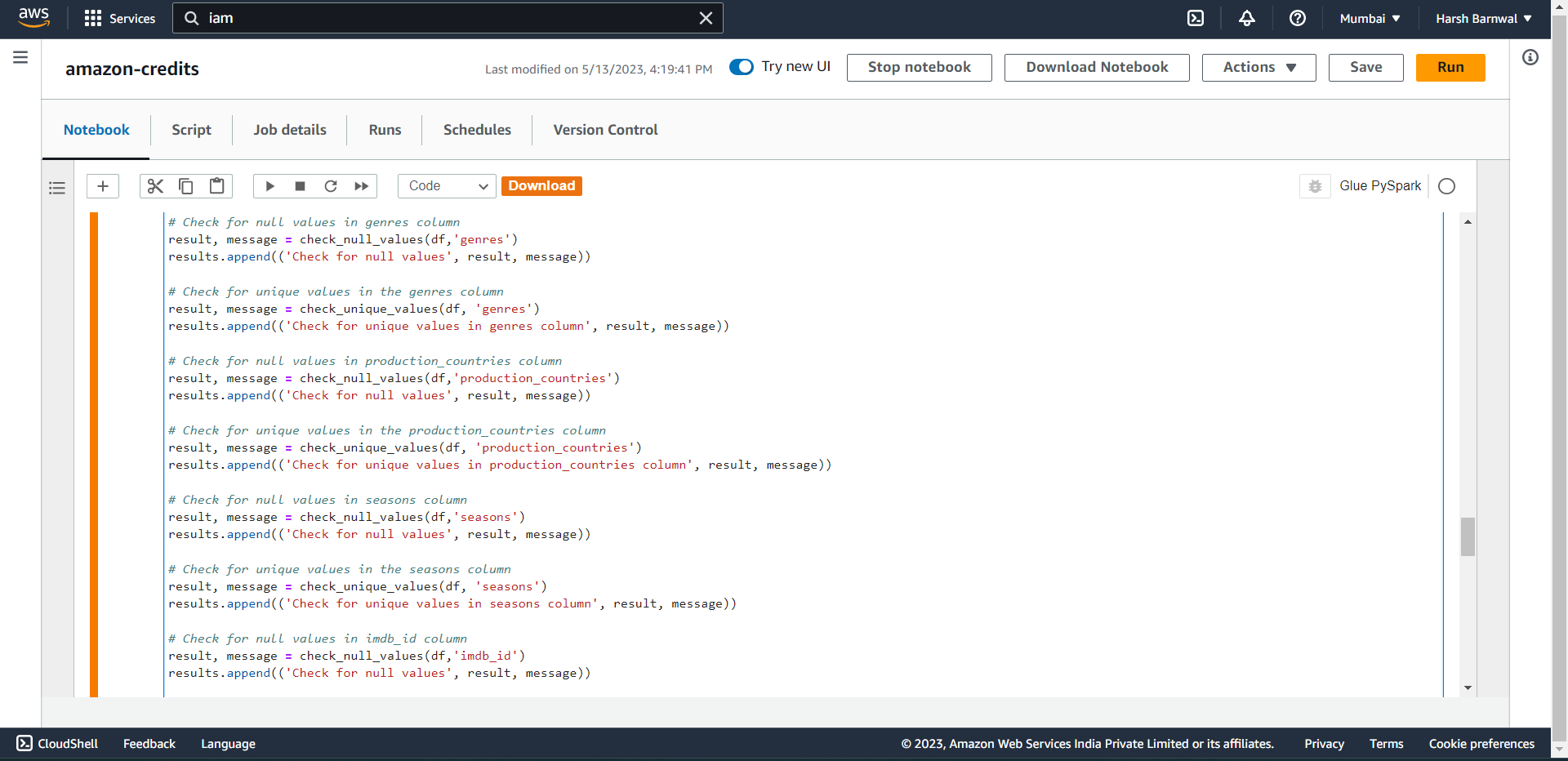


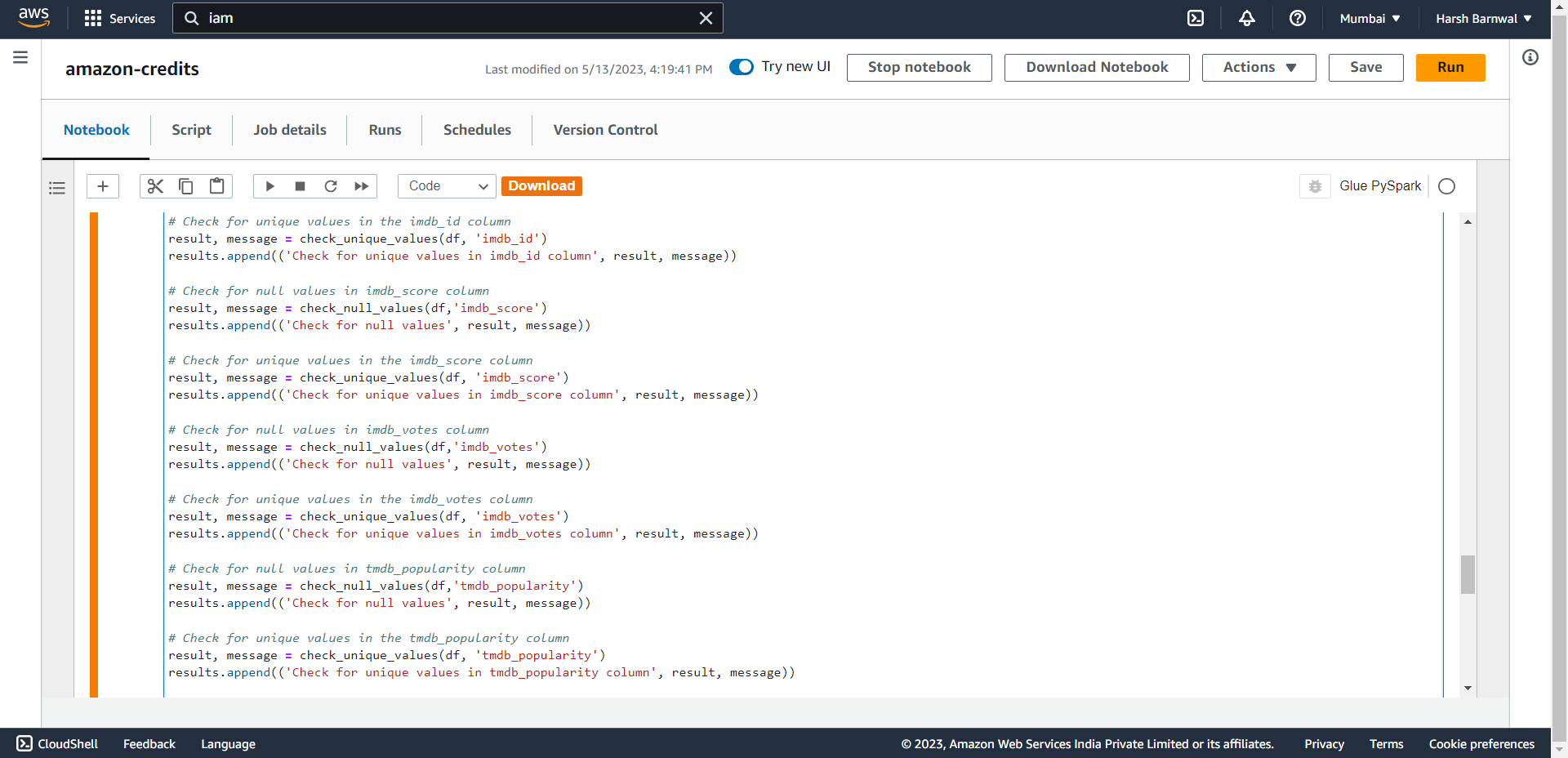


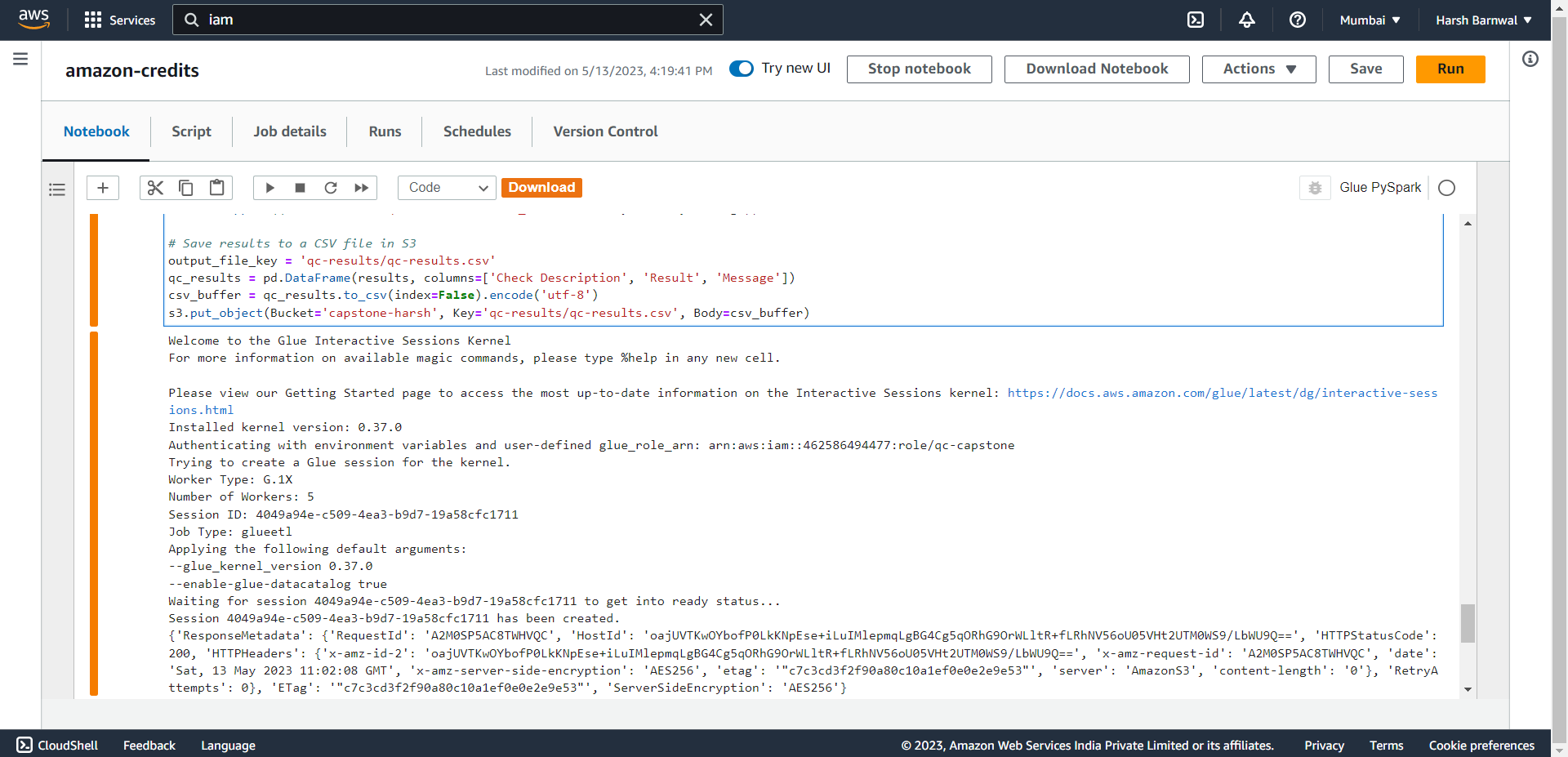


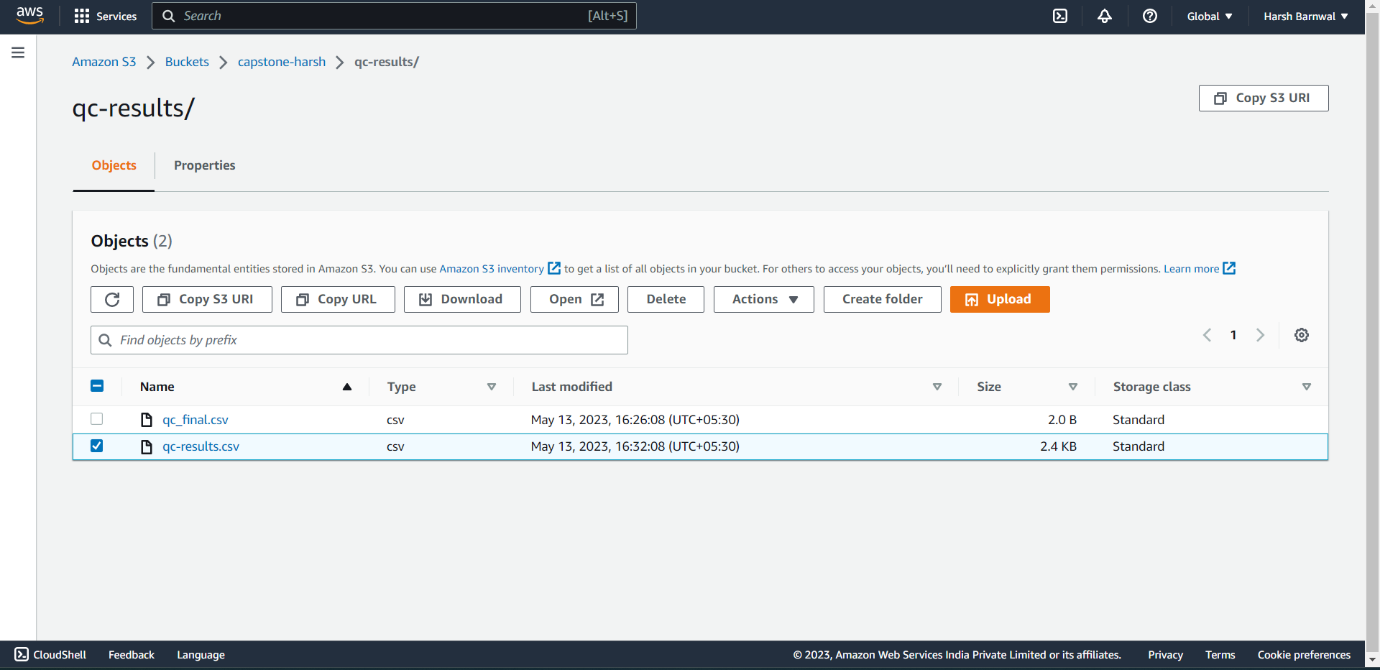


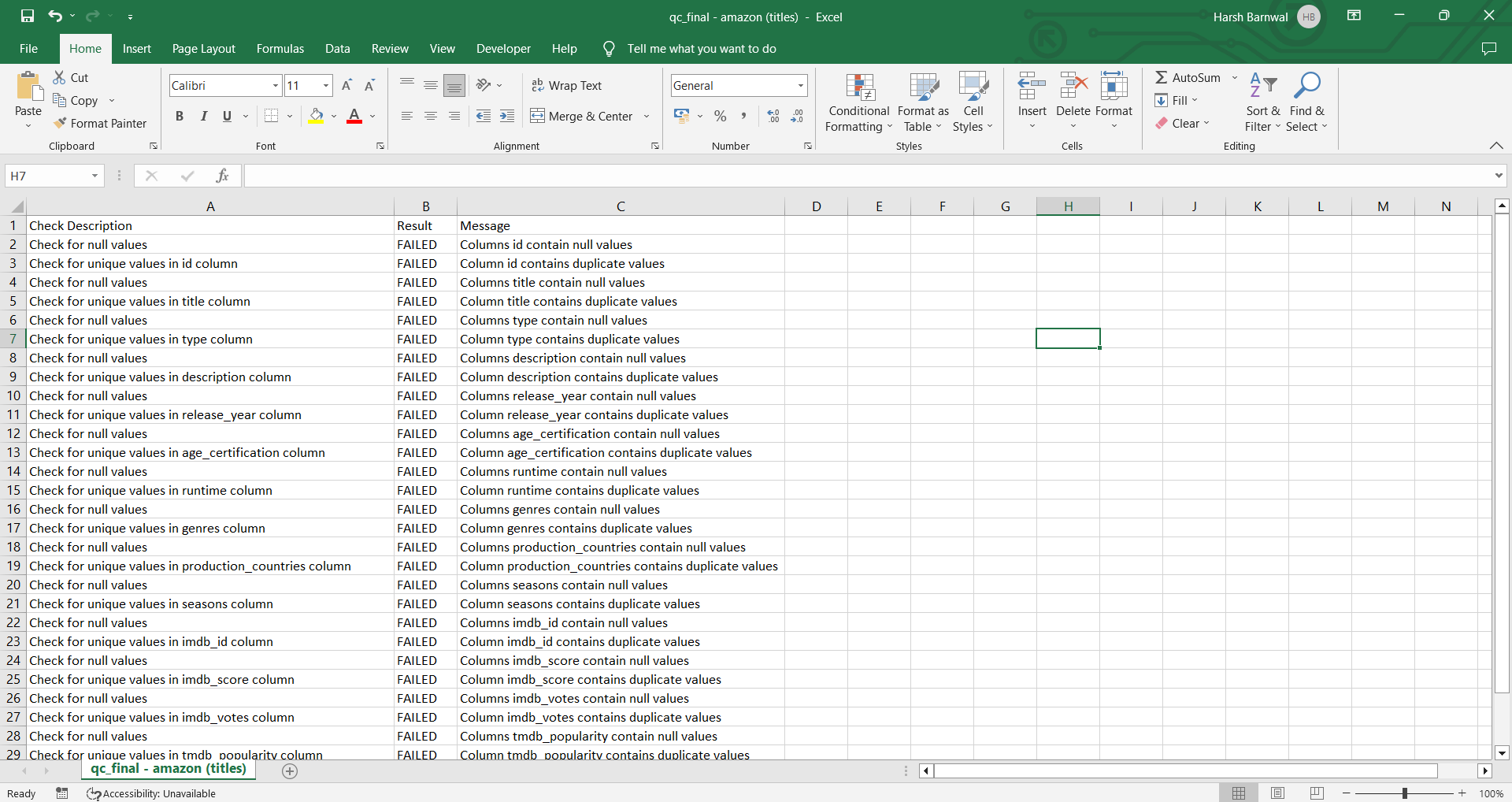






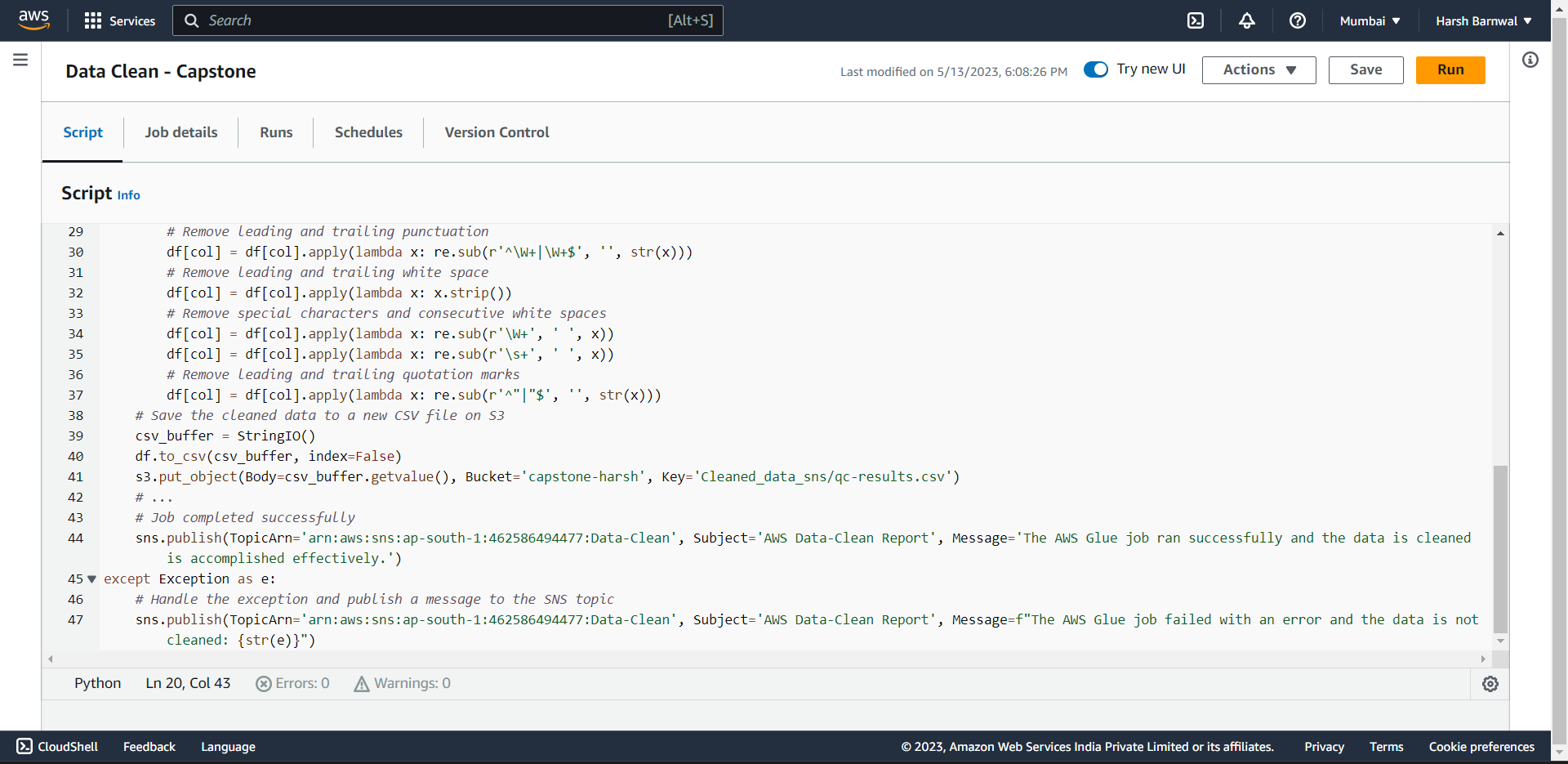


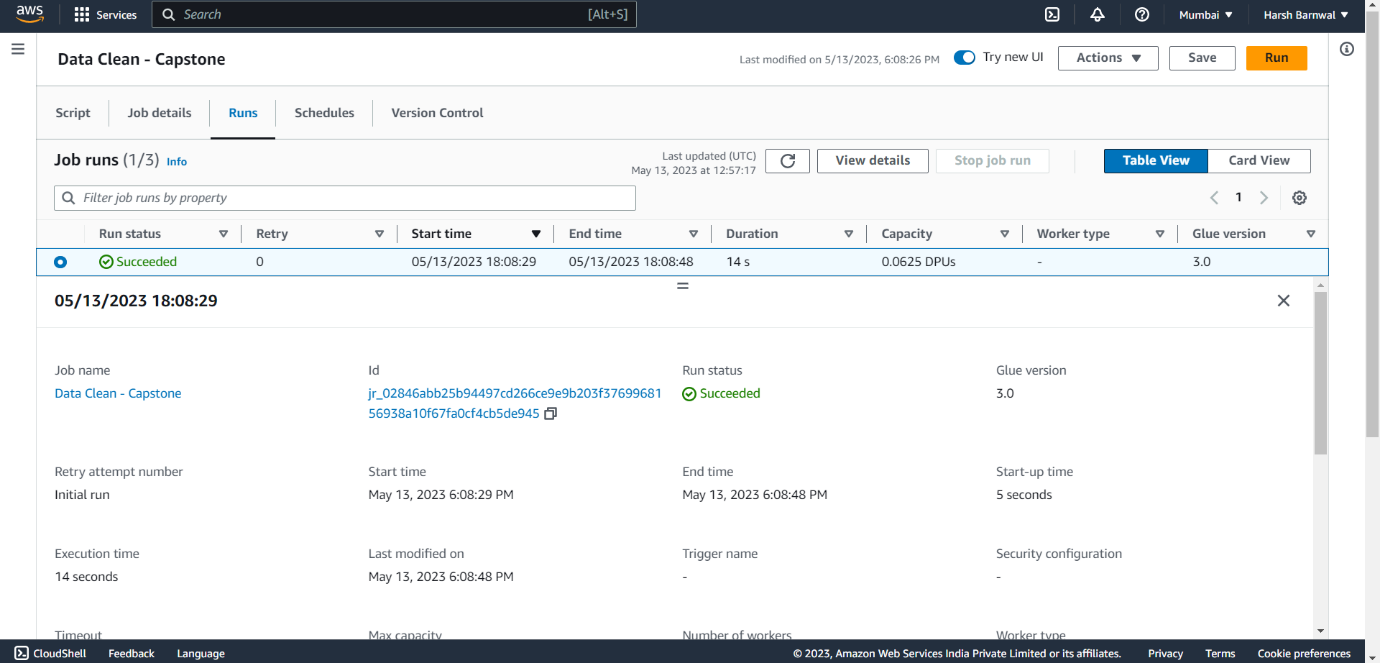
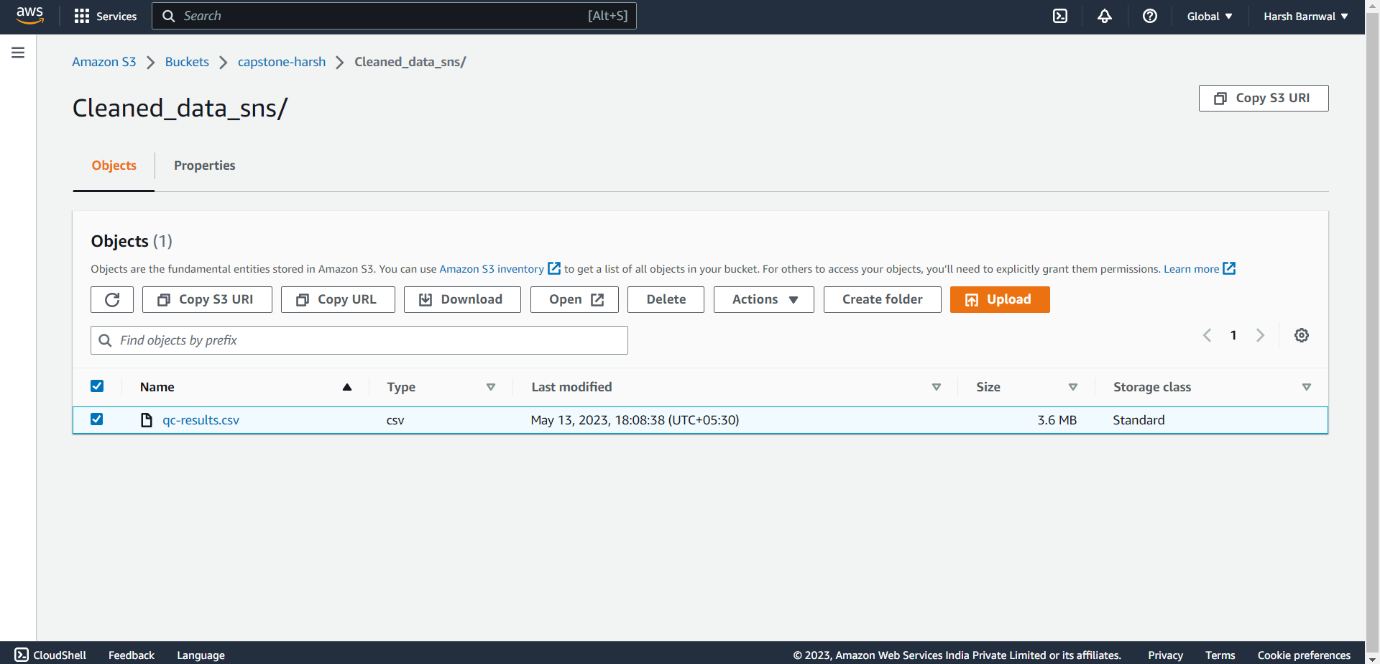


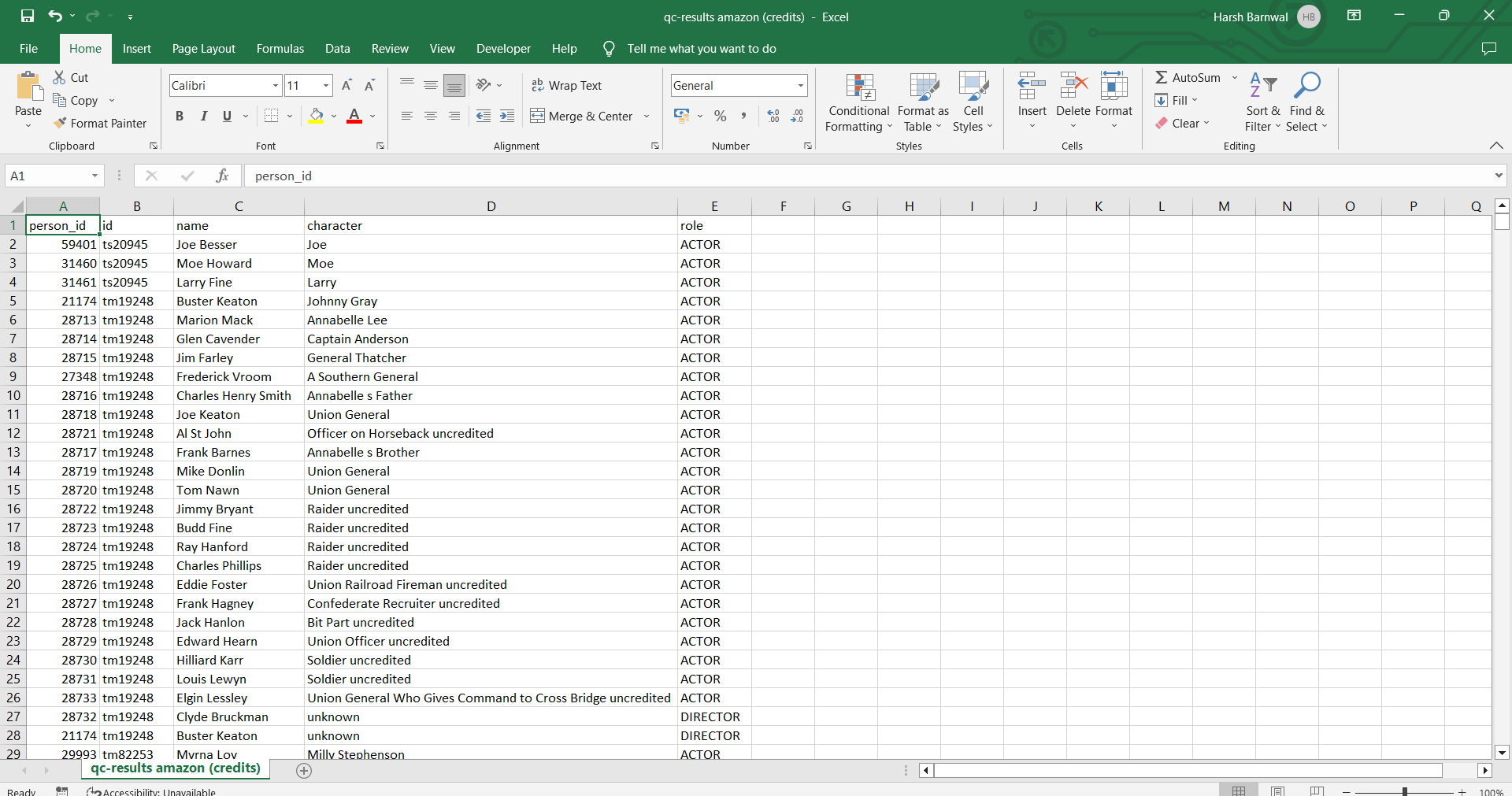
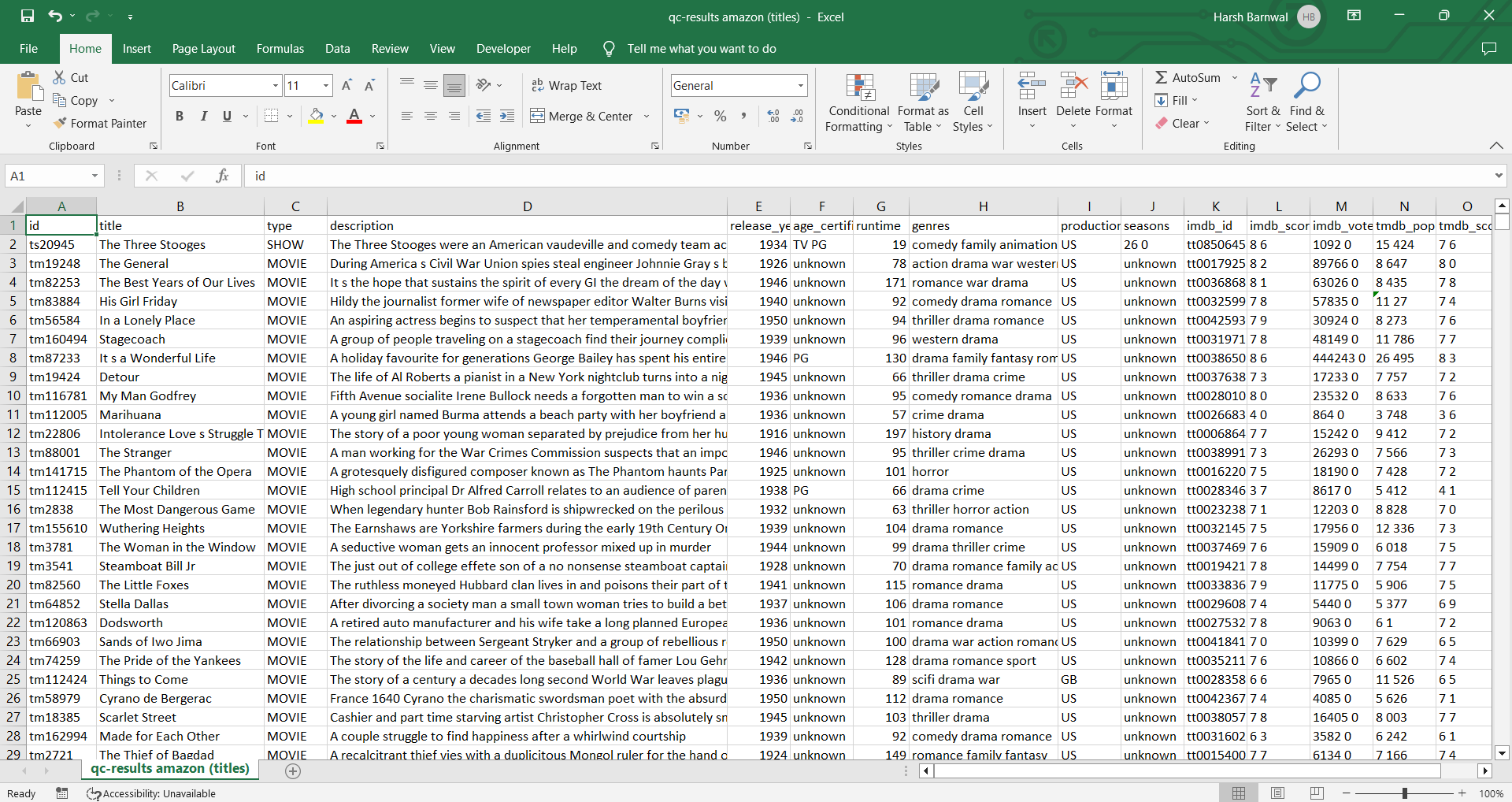


**2) Data Cleaning process using Python script in AWS Glue is performed as per the QC checklist. The script also involves sending a mail using AWS SNS when the job is successfully run and hence, data is cleaned.**

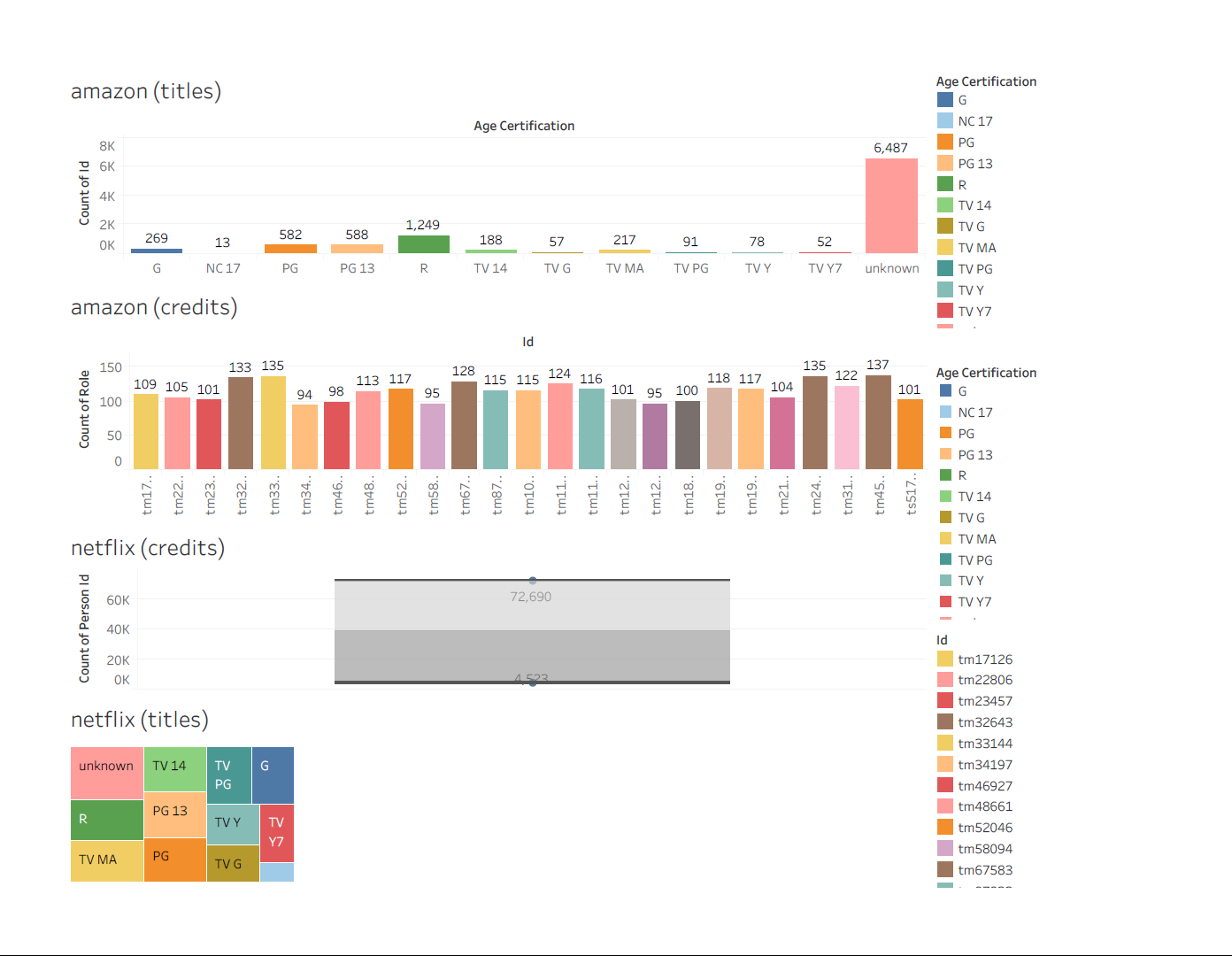


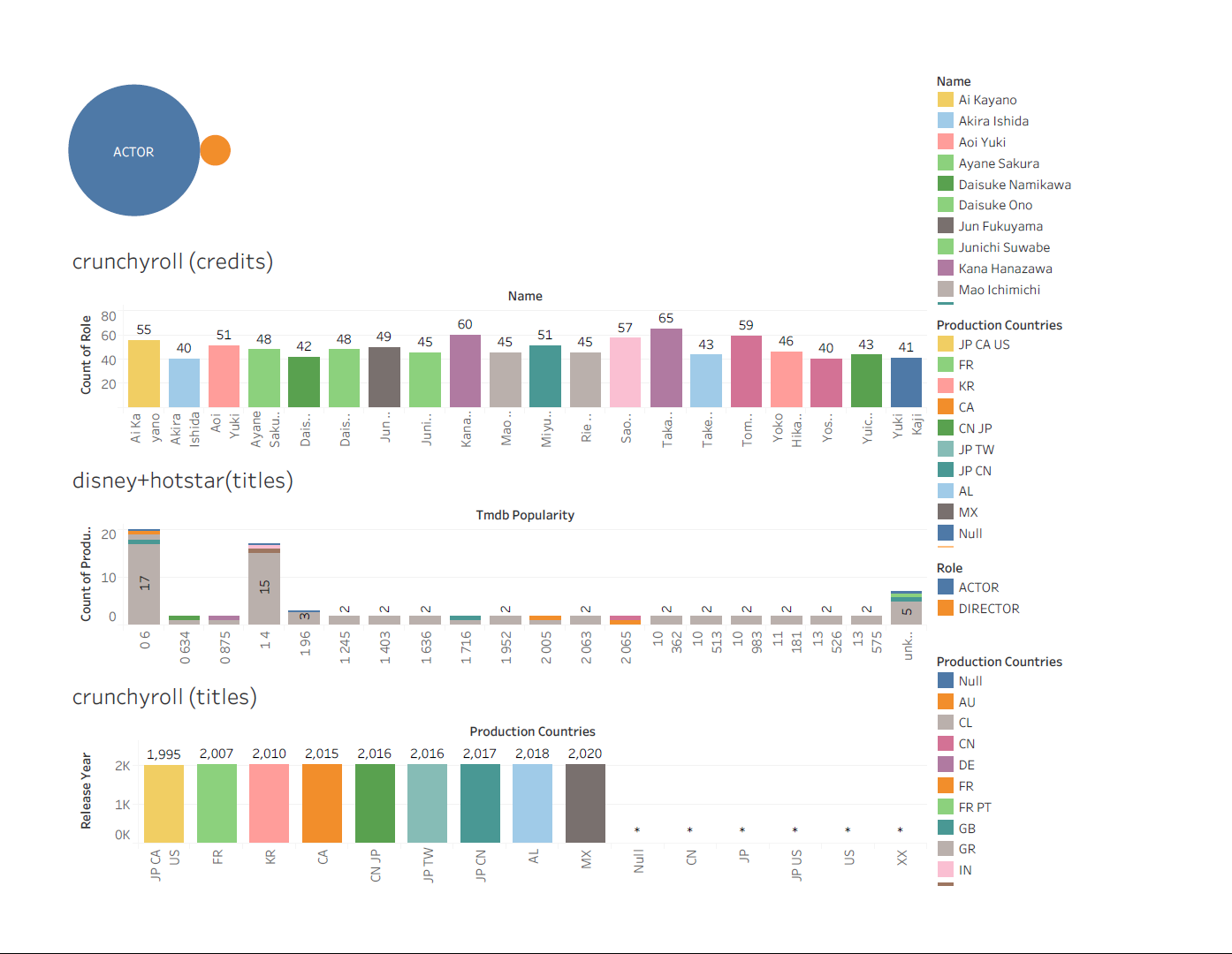


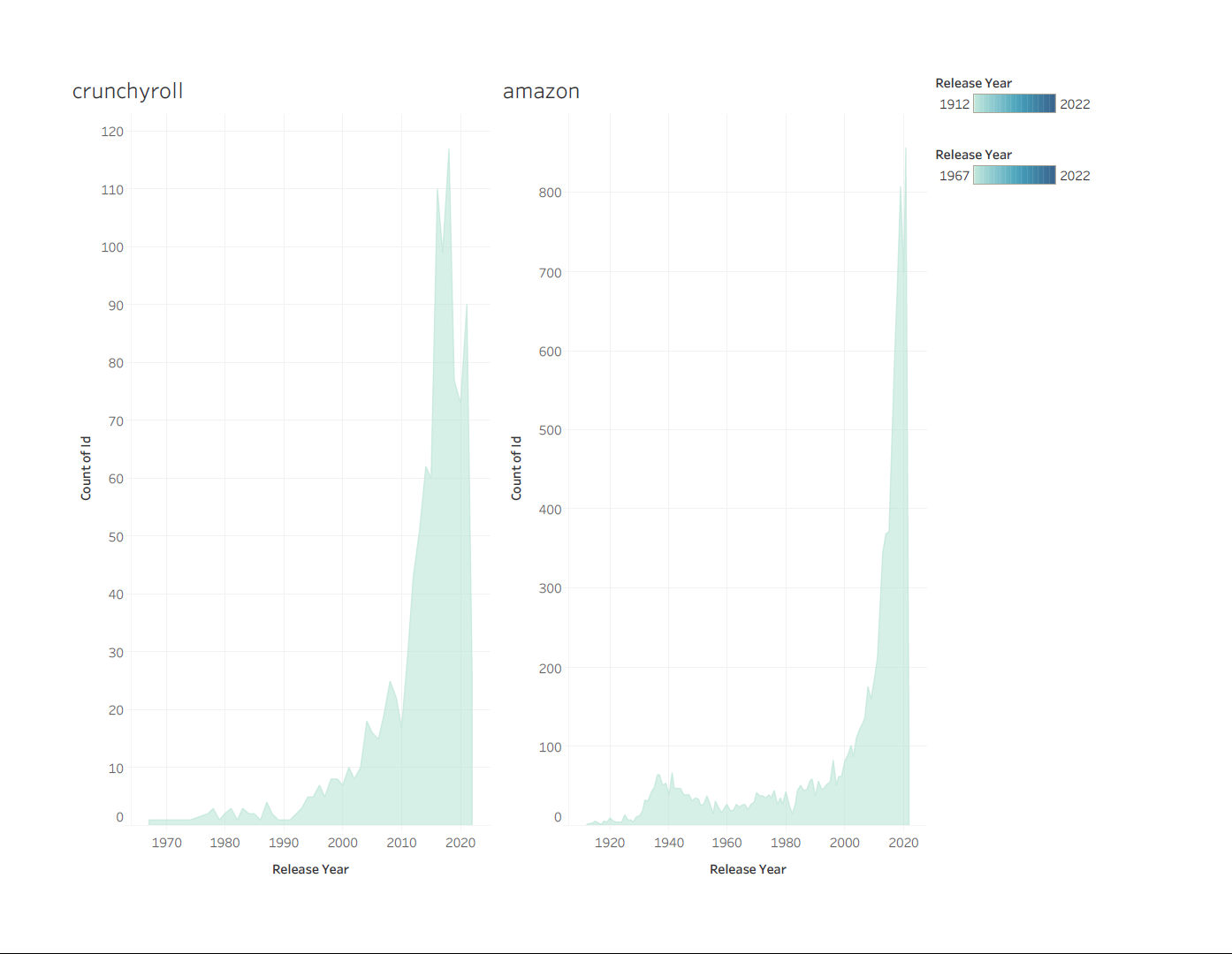
****

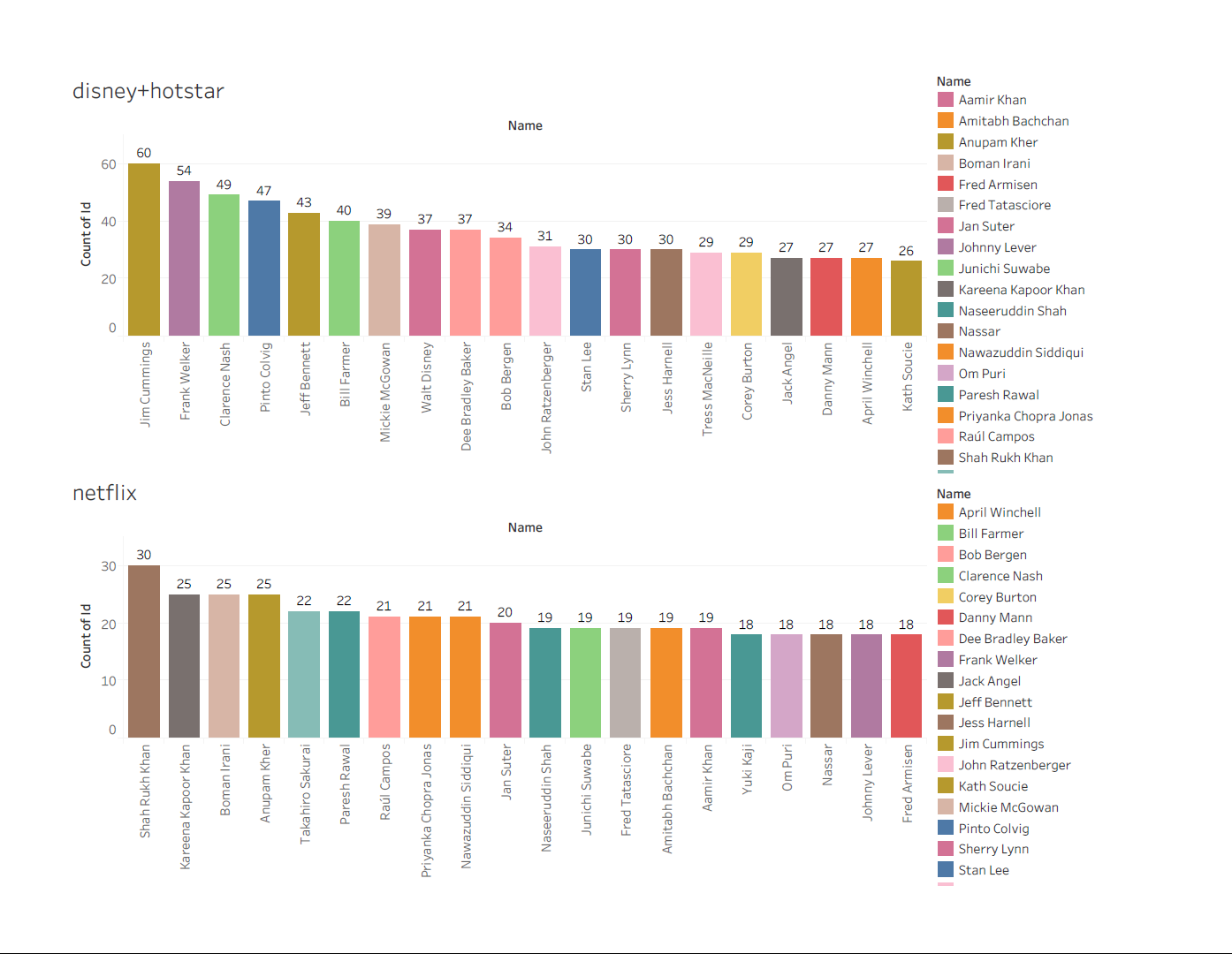
****

**3) Visualization of the cleaned data in Tableau in a dashboard -**

****

****

****

****