**Fantasy Premier League Data Analytics Project**

* Ashutosh Kulkarni 2548010
* Darren D’souza 2548016

Fantasy Premier League (FPL) is a popular fantasy sports game with millions of users worldwide, where participants manage virtual football teams and score points based on real match performances of players in the English Premier League.

This project focuses on building an end-to-end data analytics pipeline using AWS and Power BI to analyse Fantasy Premier League (FPL) player performance. The objective is to clean, transform, and visualize football performance data to assist in selecting well-performing and undervalued players.

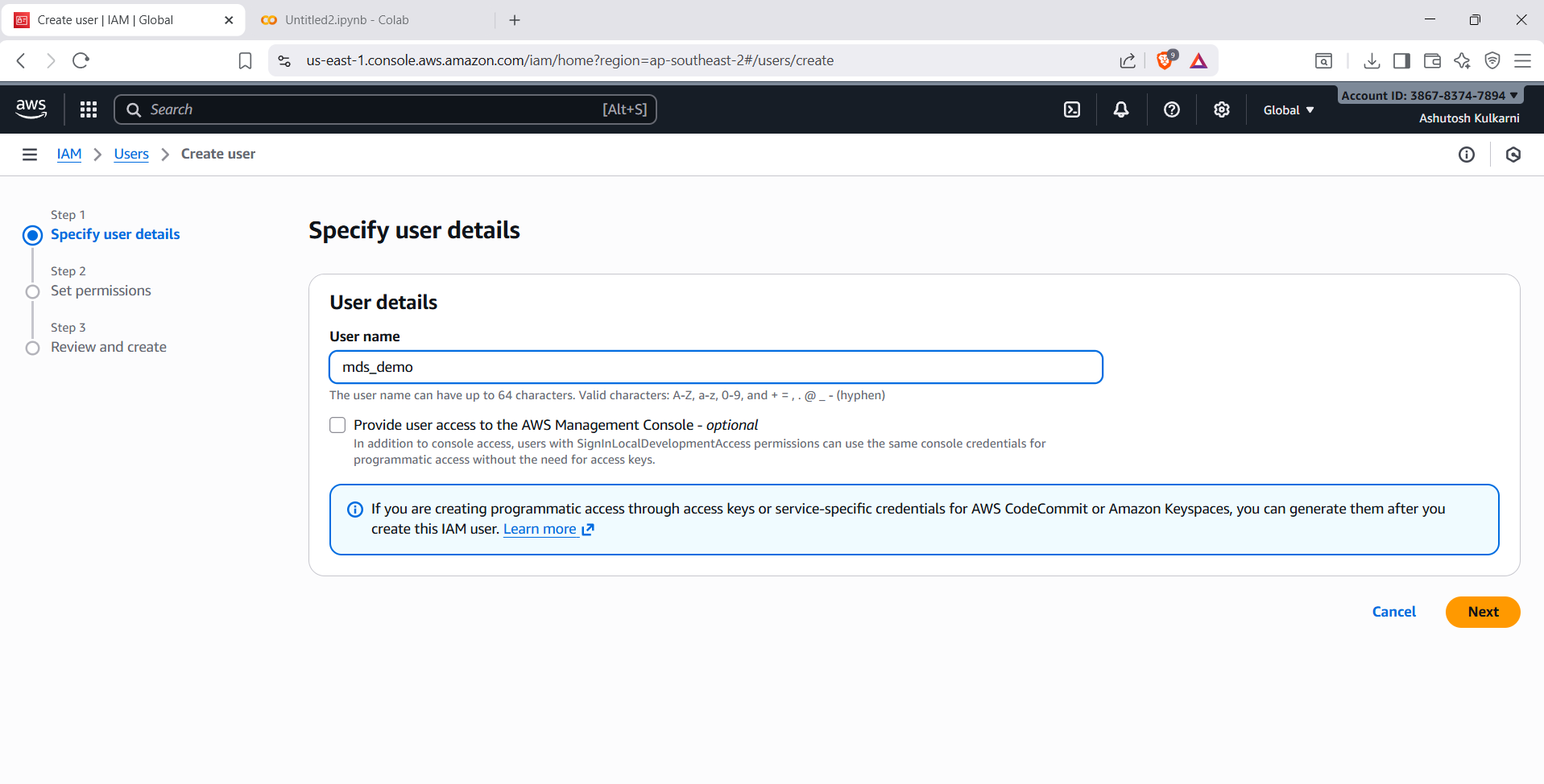
**Project Flow**



**Project Initialization**

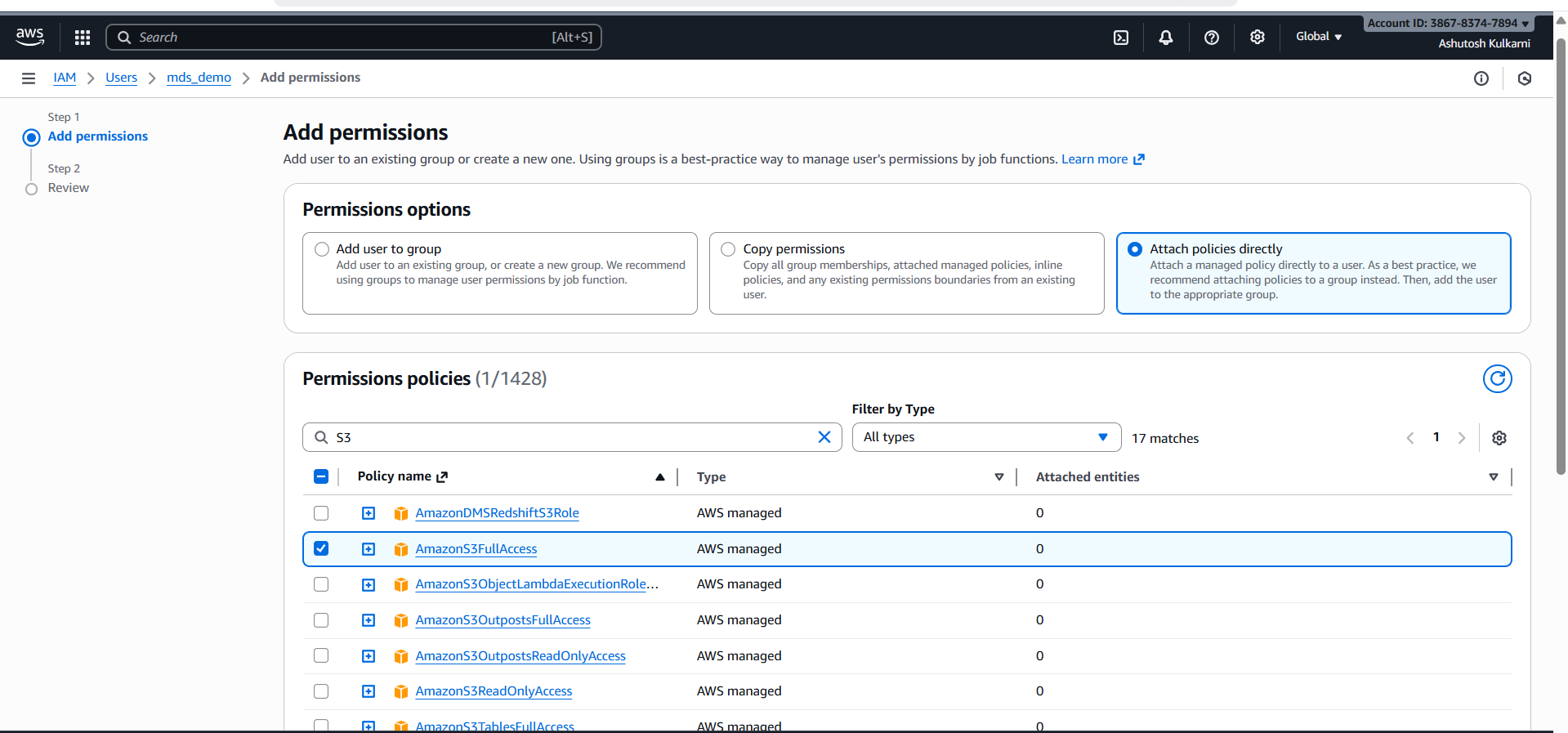
1. AWS IAM User Creation

An IAM (Identity and Access Management) user was created to securely access AWS services. This ensures proper authentication and controlled access instead of using the root account.

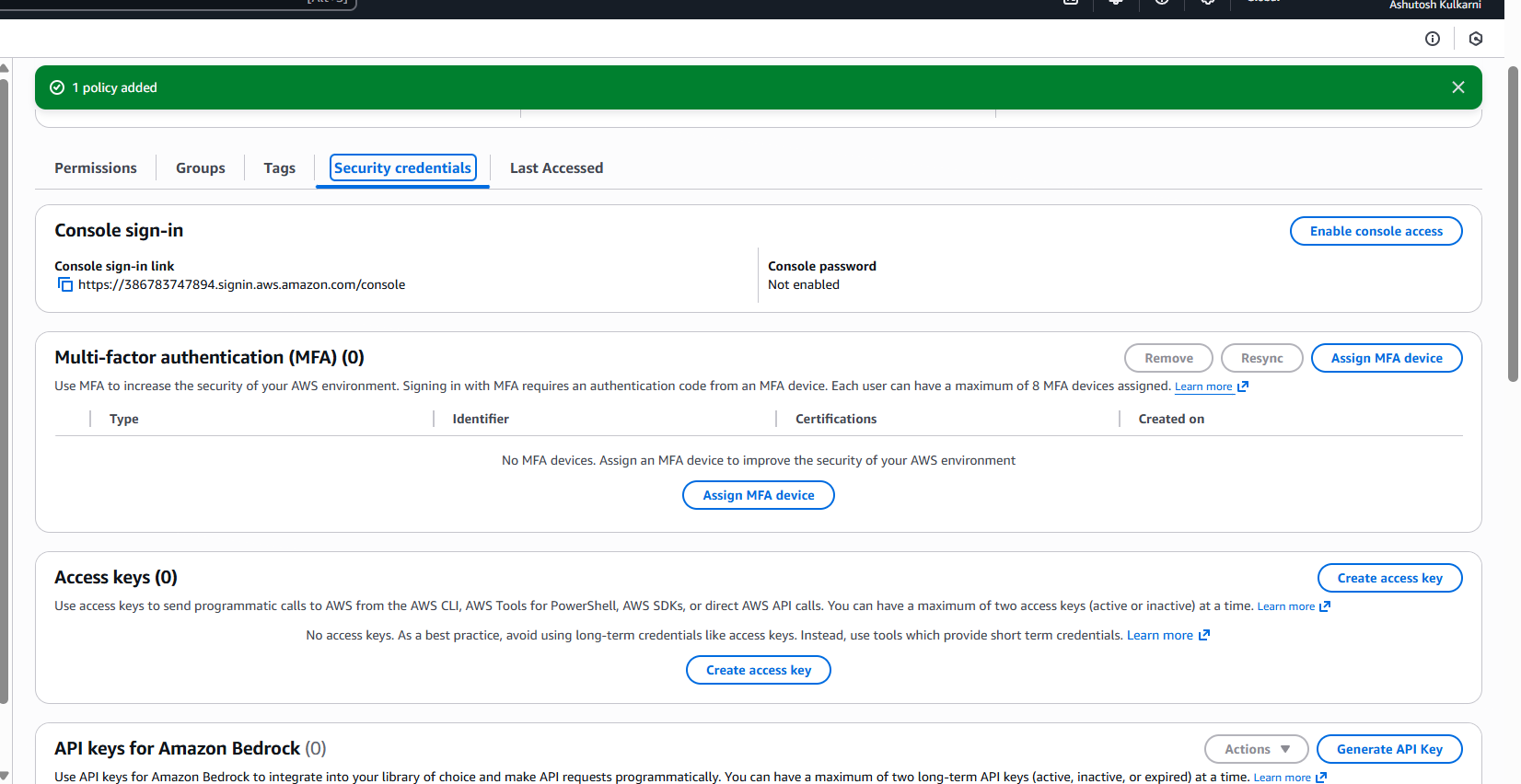


1. Add permissions

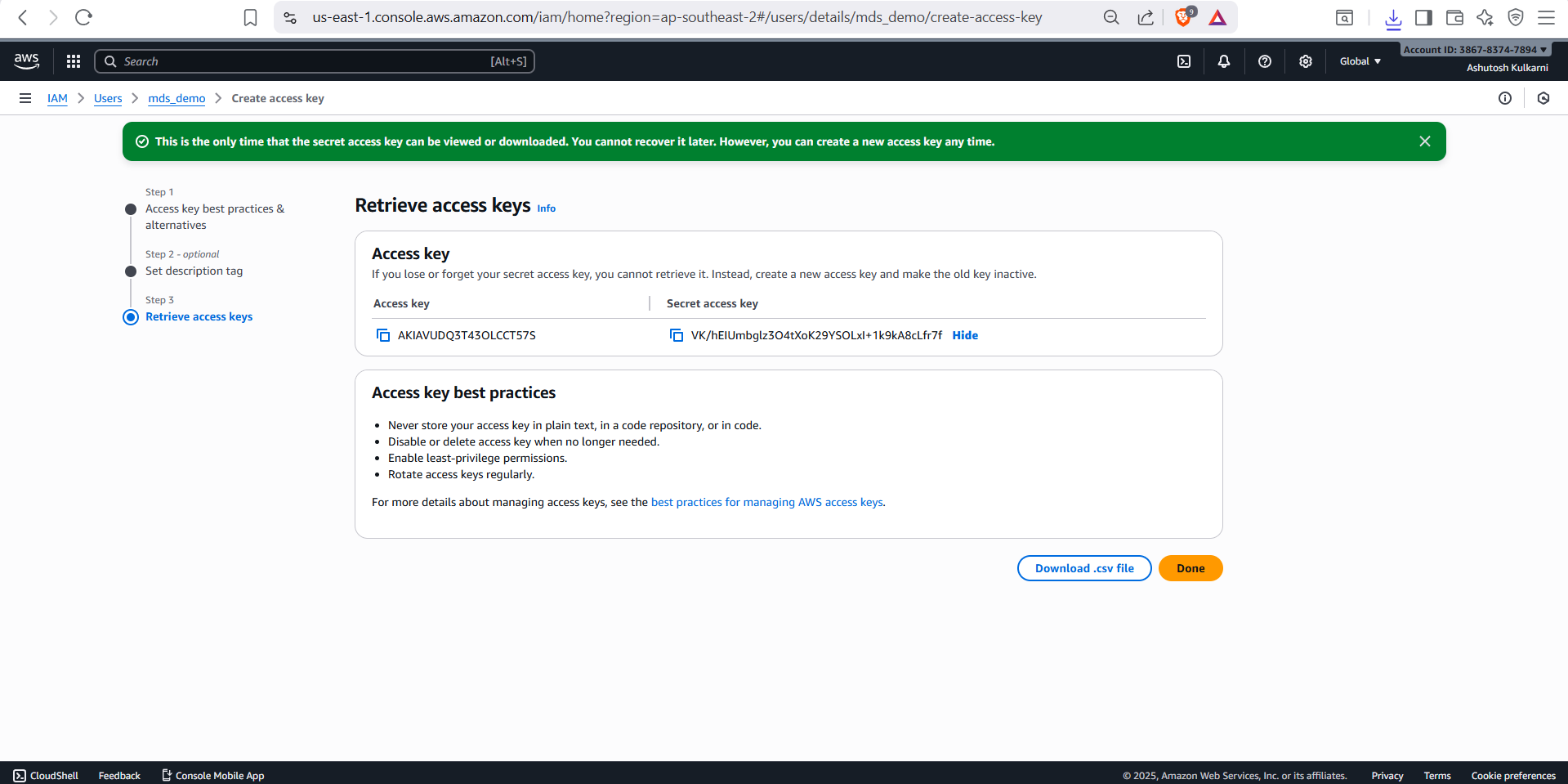
Appropriate IAM policies were attached to the user to allow access to services such as S3, Athena, and Lambda. Permissions followed the principle of least privilege to enhance security.



1. Generate Security Key

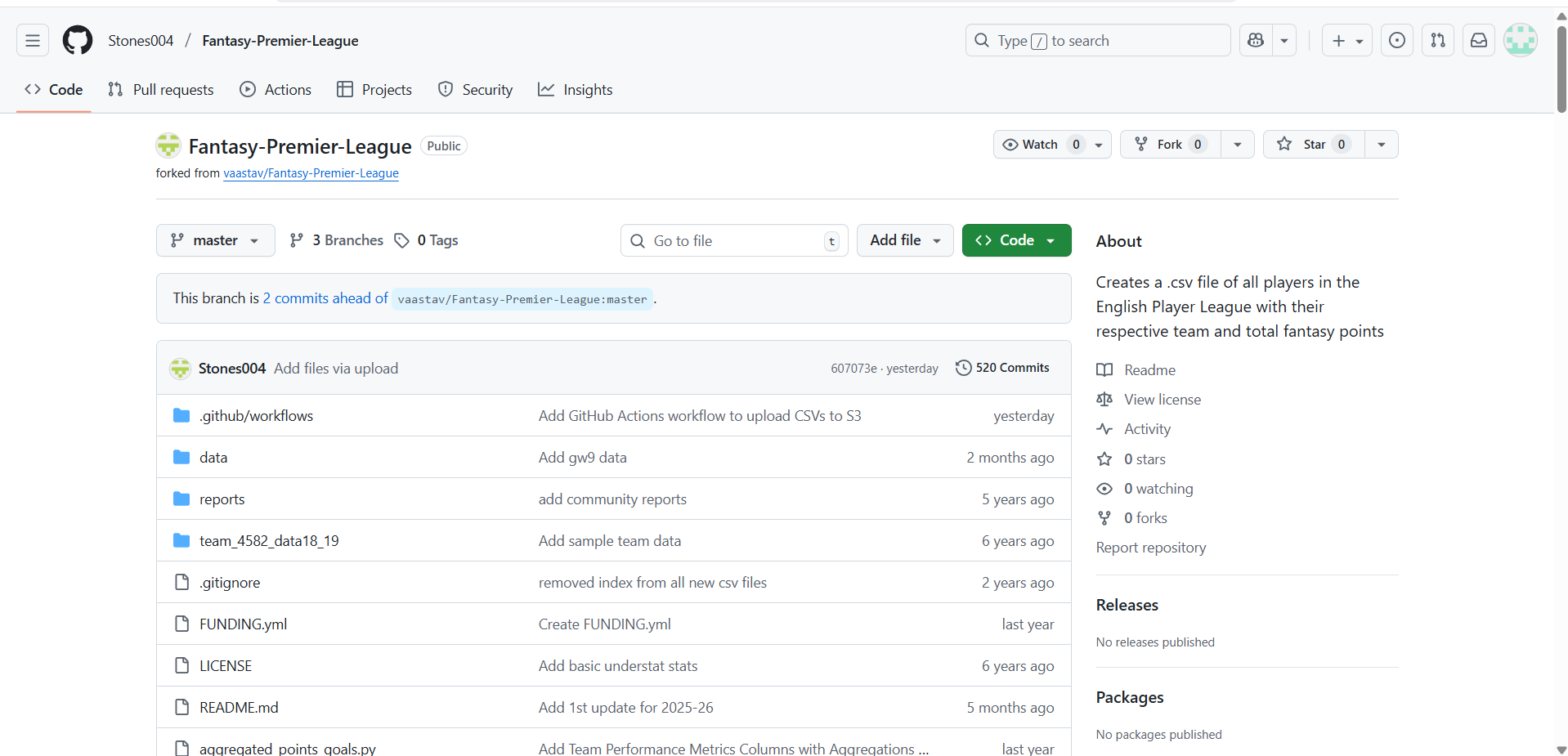


1. Retrieving and saving access keys



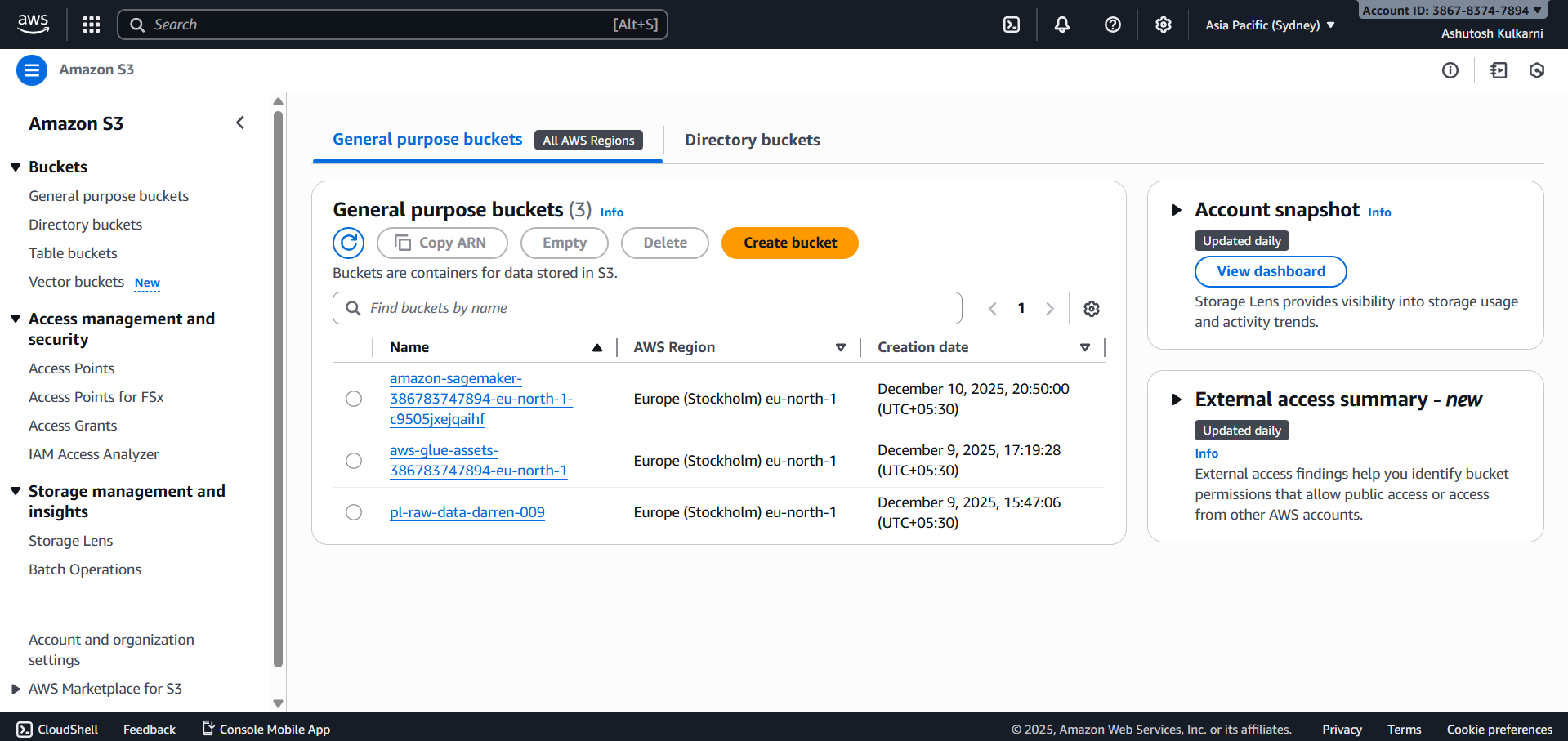
Data Generated

The Fantasy Premier League dataset used in this project was sourced from a public GitHub repository containing structured CSV files with event-level player performance data. Once validated, the data was downloaded and programmatically transferred into an Amazon S3 bucket.

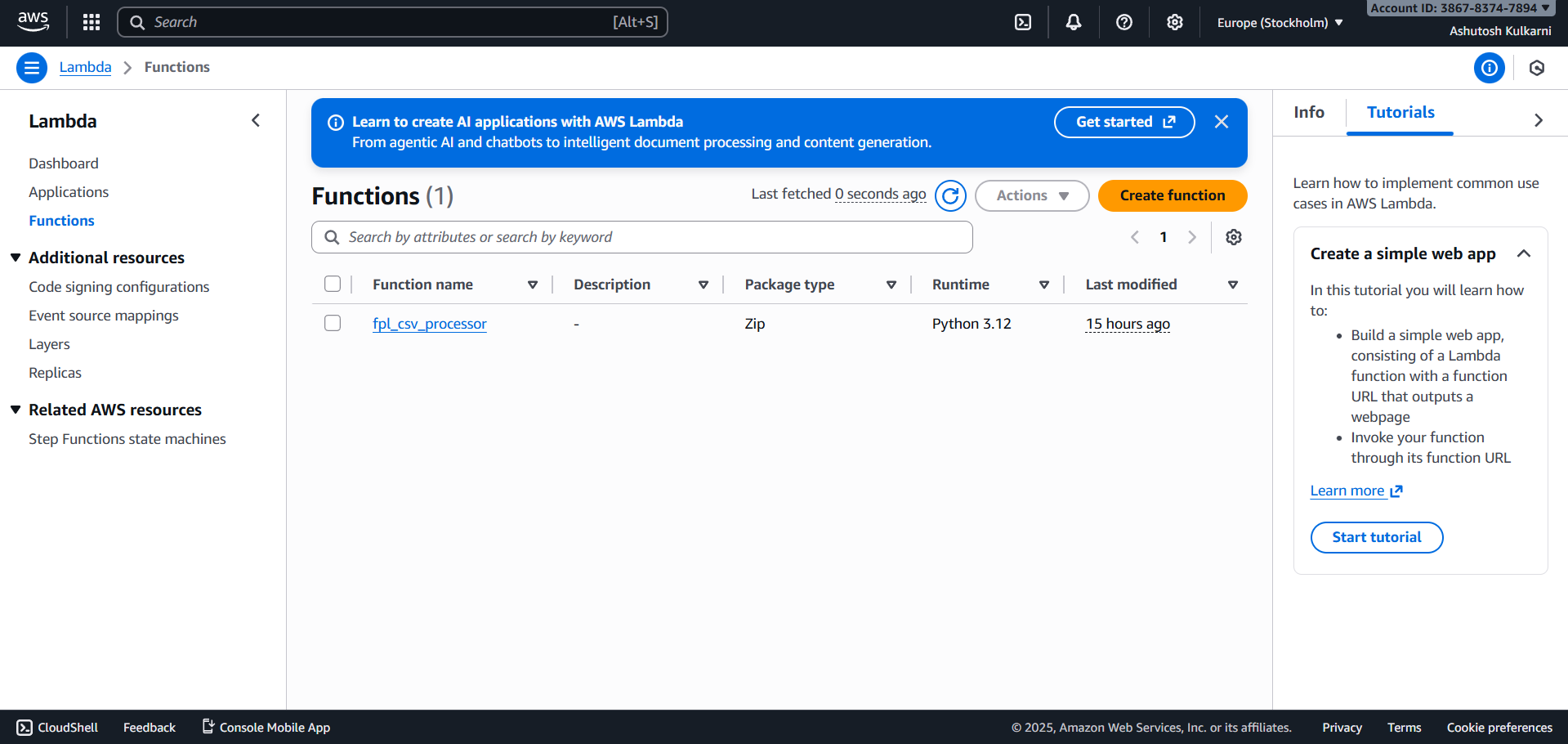


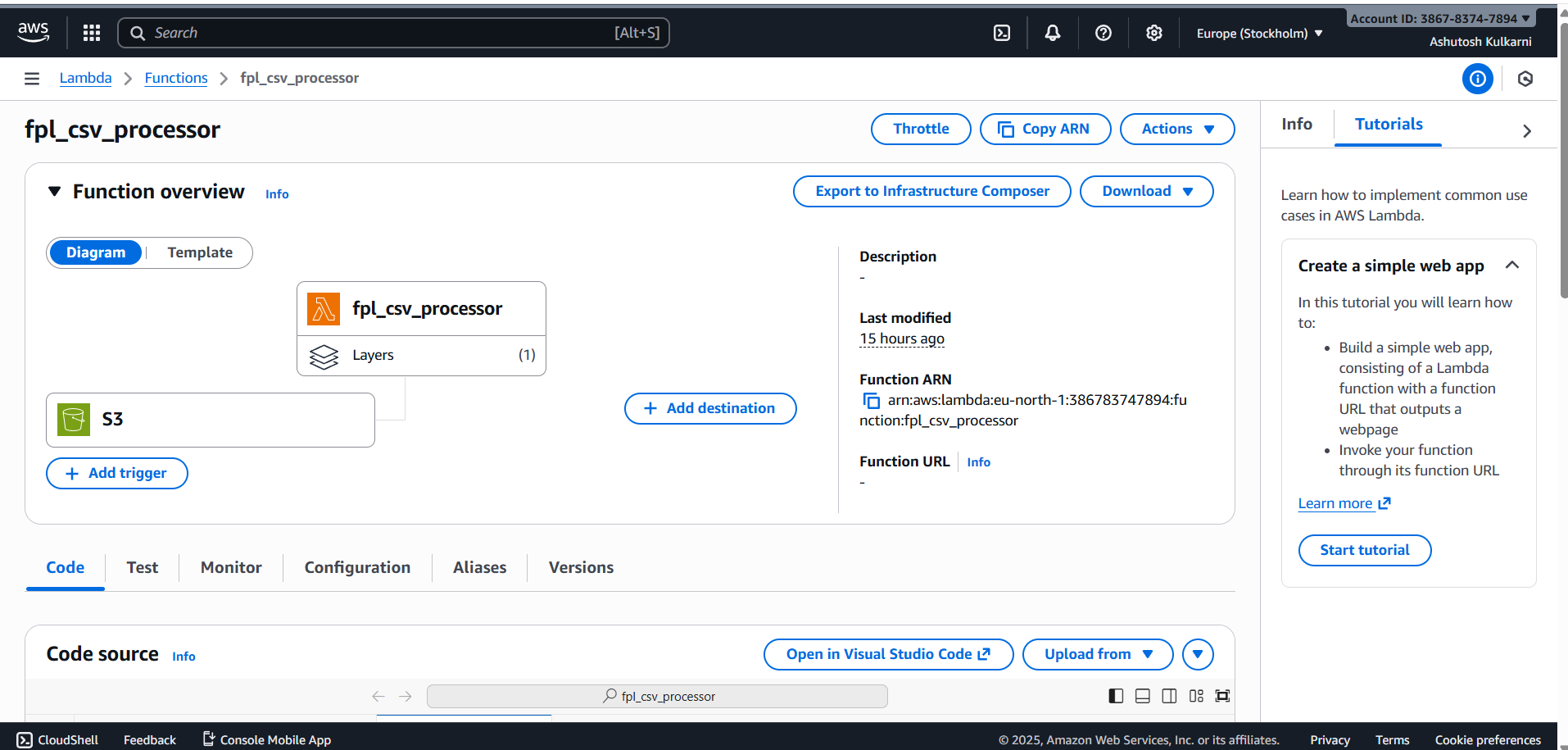
Amazon S3 bucket

Amazon S3 was used to store raw and processed CSV files. Separate folders were maintained for raw data and cleaned data to ensure data integrity and traceability. The last bucket was the one used in this project



AWS lambda

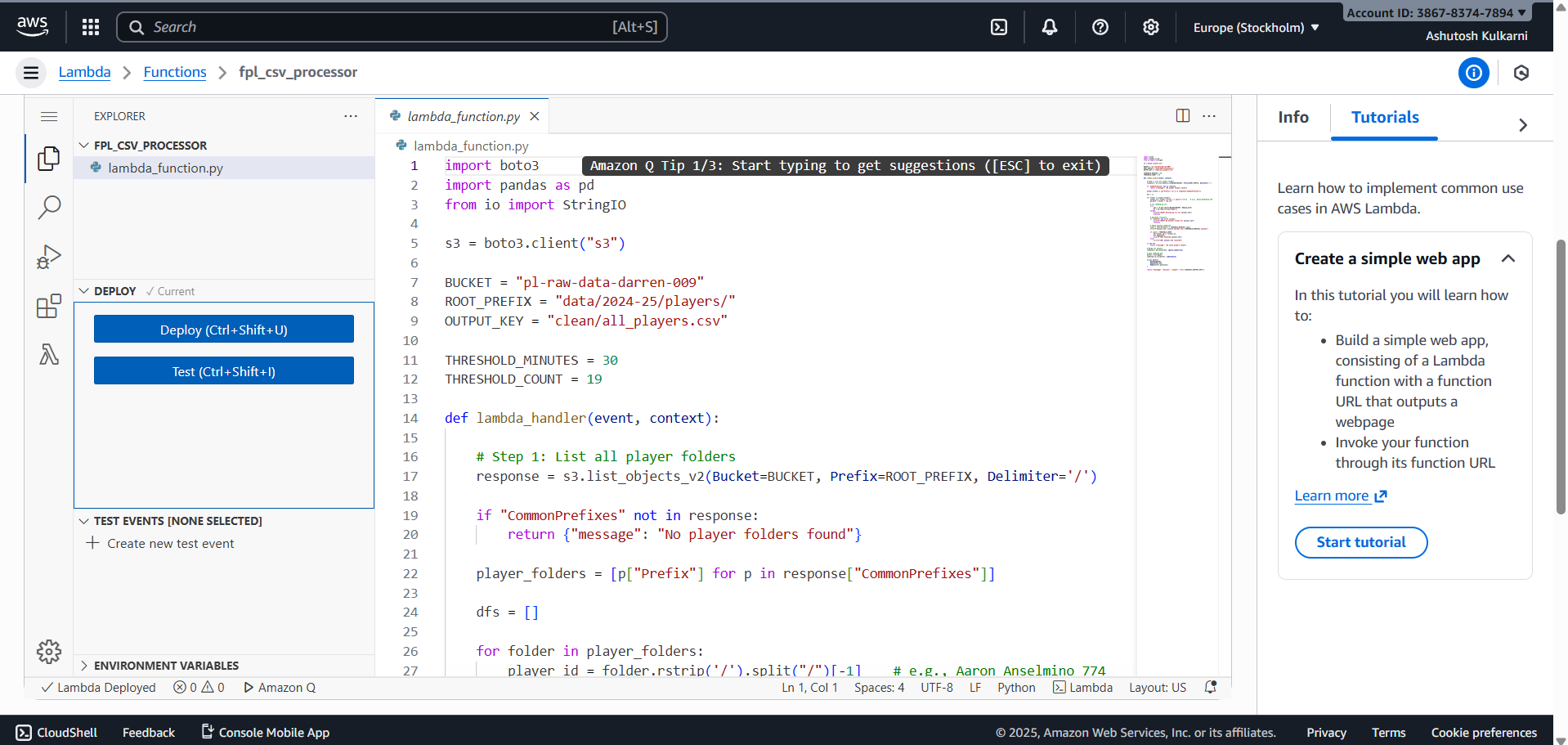


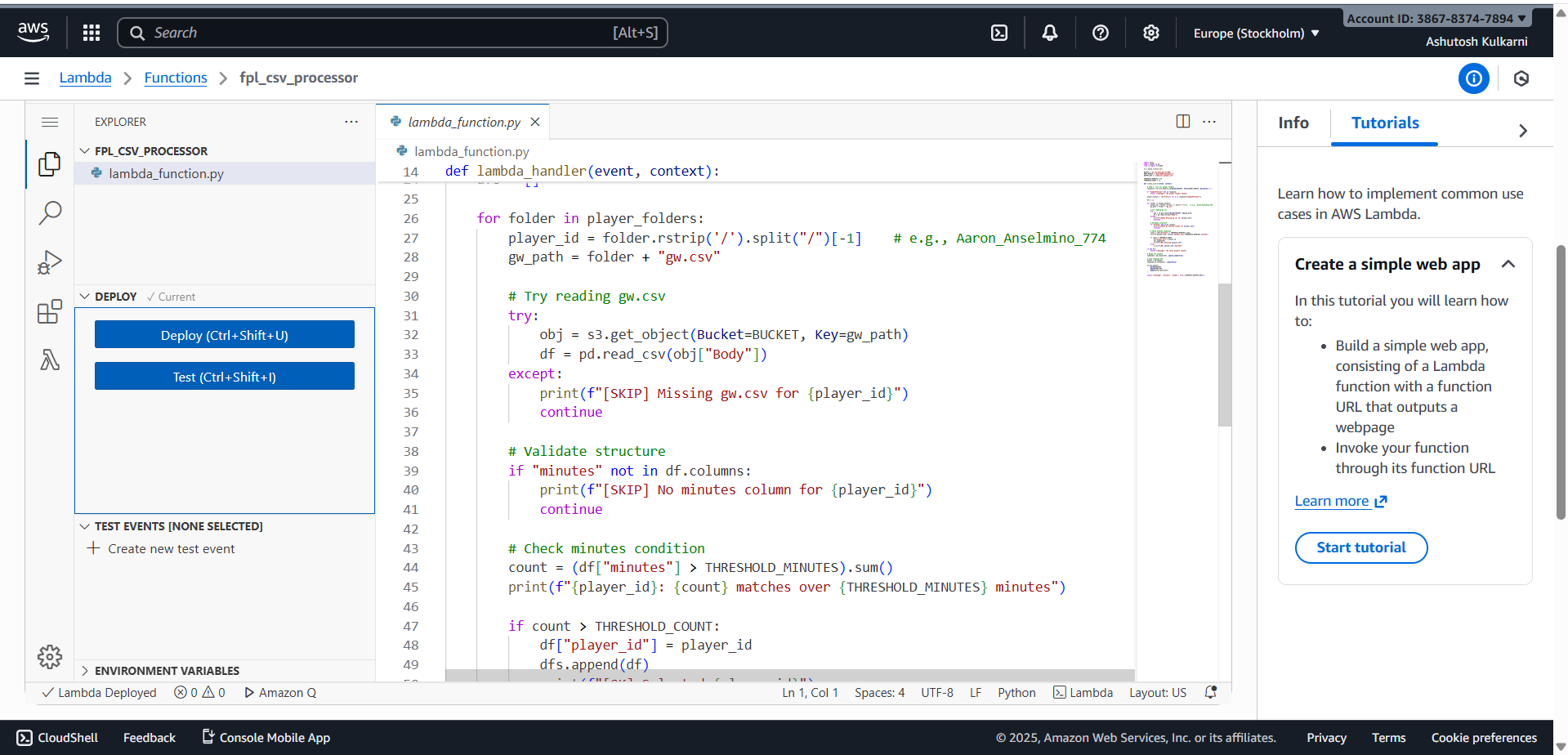


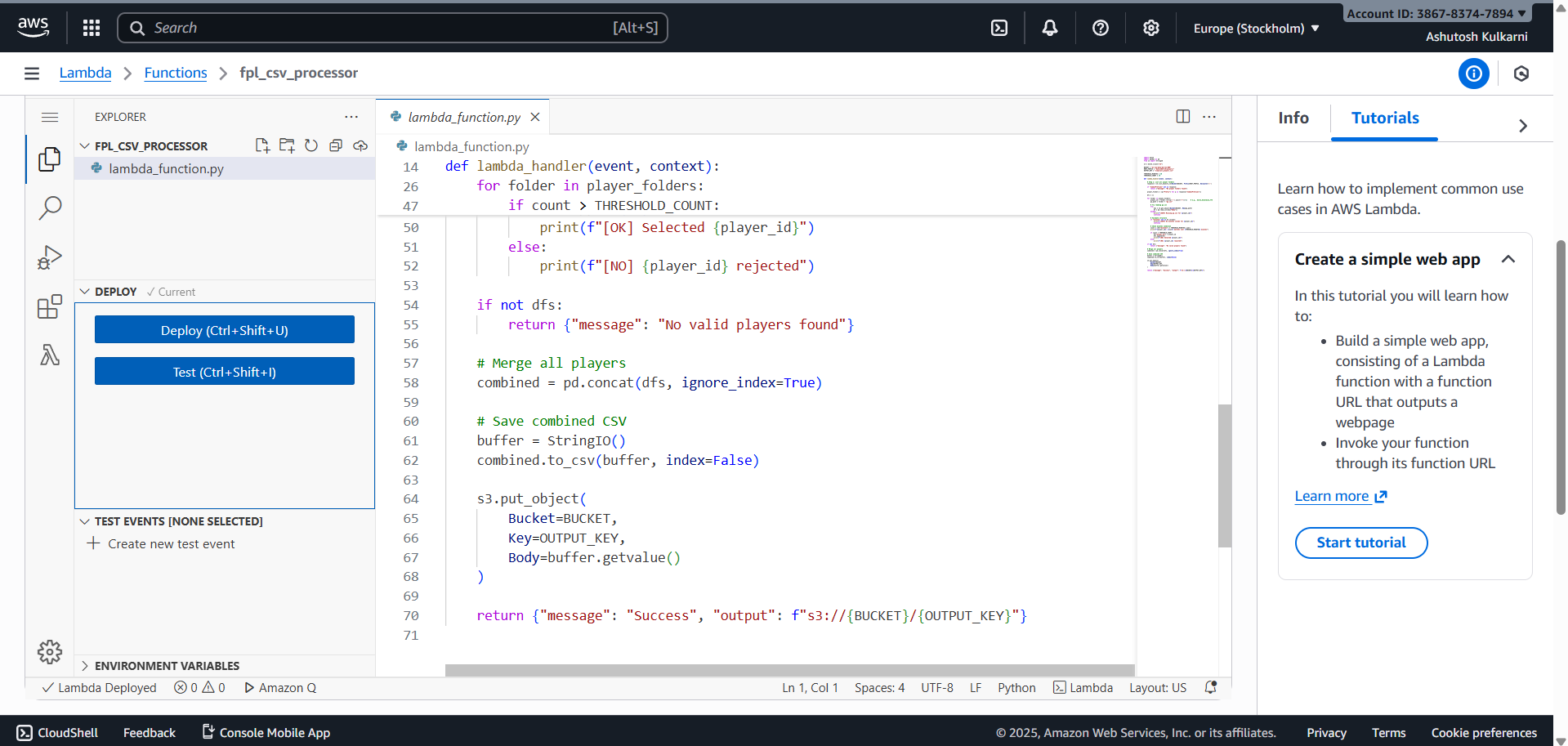
Data filtering

The function applies a logical filtering mechanism based on predefined thresholds, selecting only those players who meet minimum performance criteria. Specifically, players are included only if they have participated in a sufficient number of matches exceeding a minimum minutes threshold. This ensures that fringe or inactive players are excluded from the final analysis, improving data reliability and analytical relevance.

Once eligible players are identified, the Lambda function merges all valid player CSV files into a single consolidated dataset. The cleaned and combined CSV file is then written back to Amazon S3 under a separate output path, maintaining a clear separation between raw and processed data.

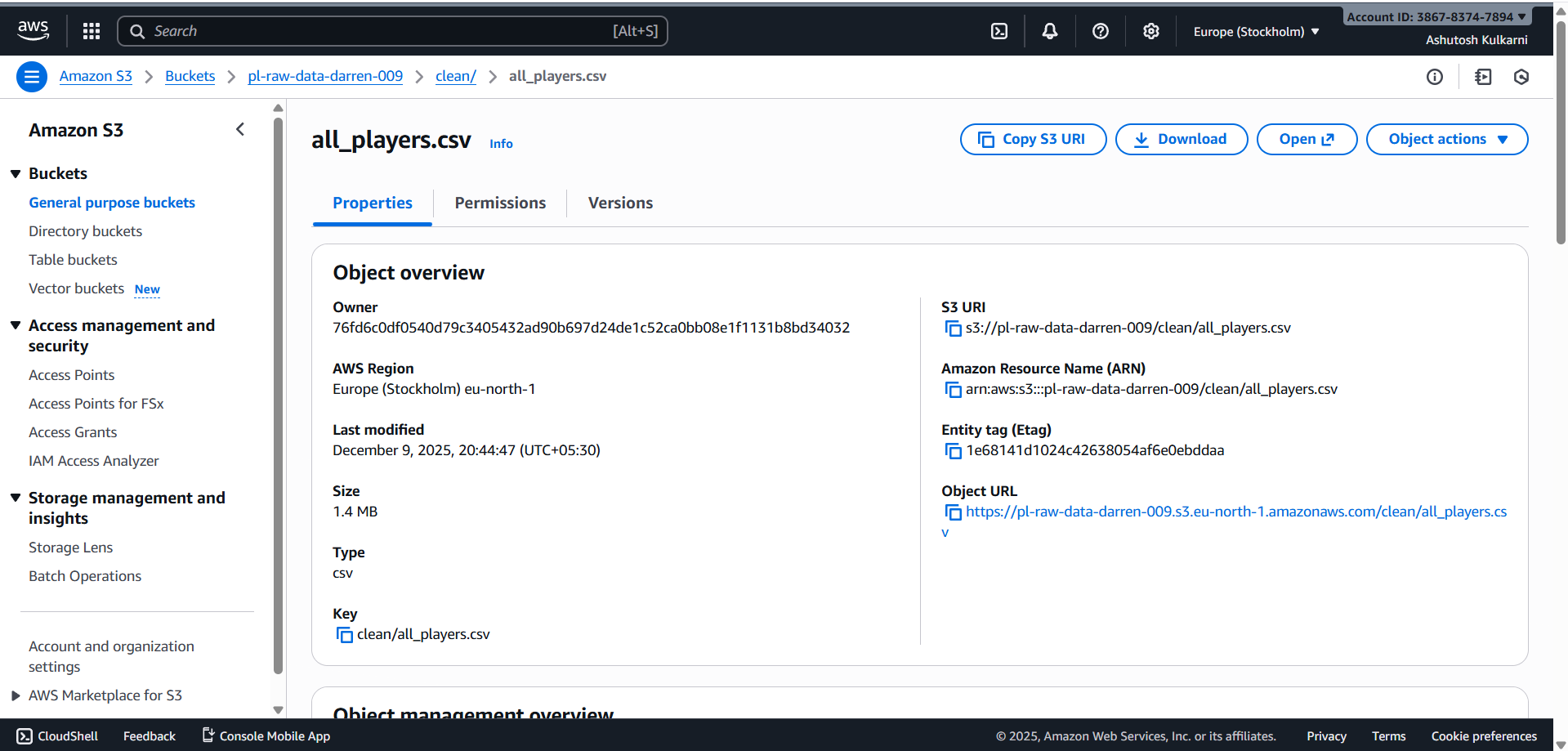


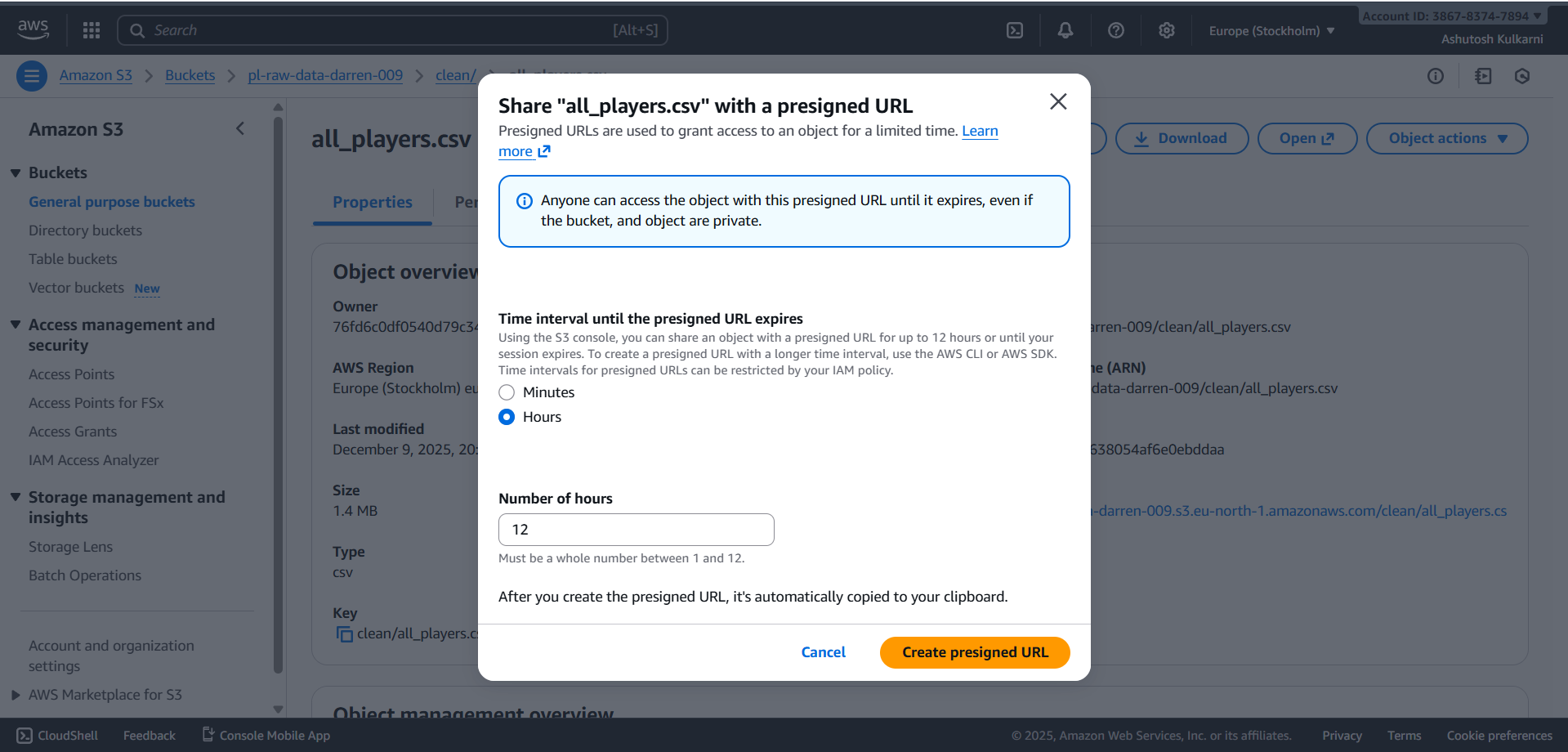




Creating pre-signed url

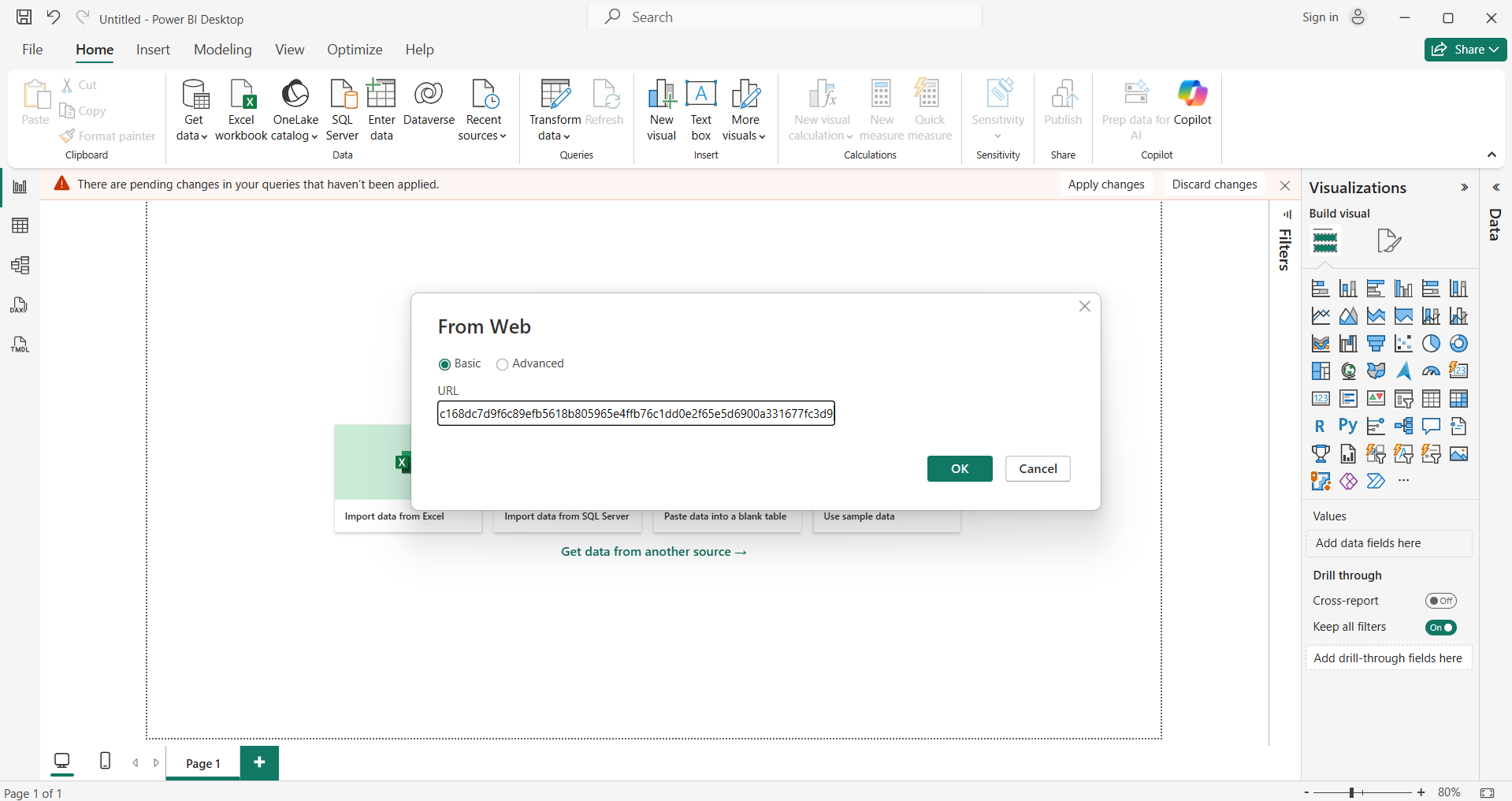
A secure pre-signed S3 URL is used to connect the cleaned dataset to Power BI, enabling seamless and credential-free data refreshes

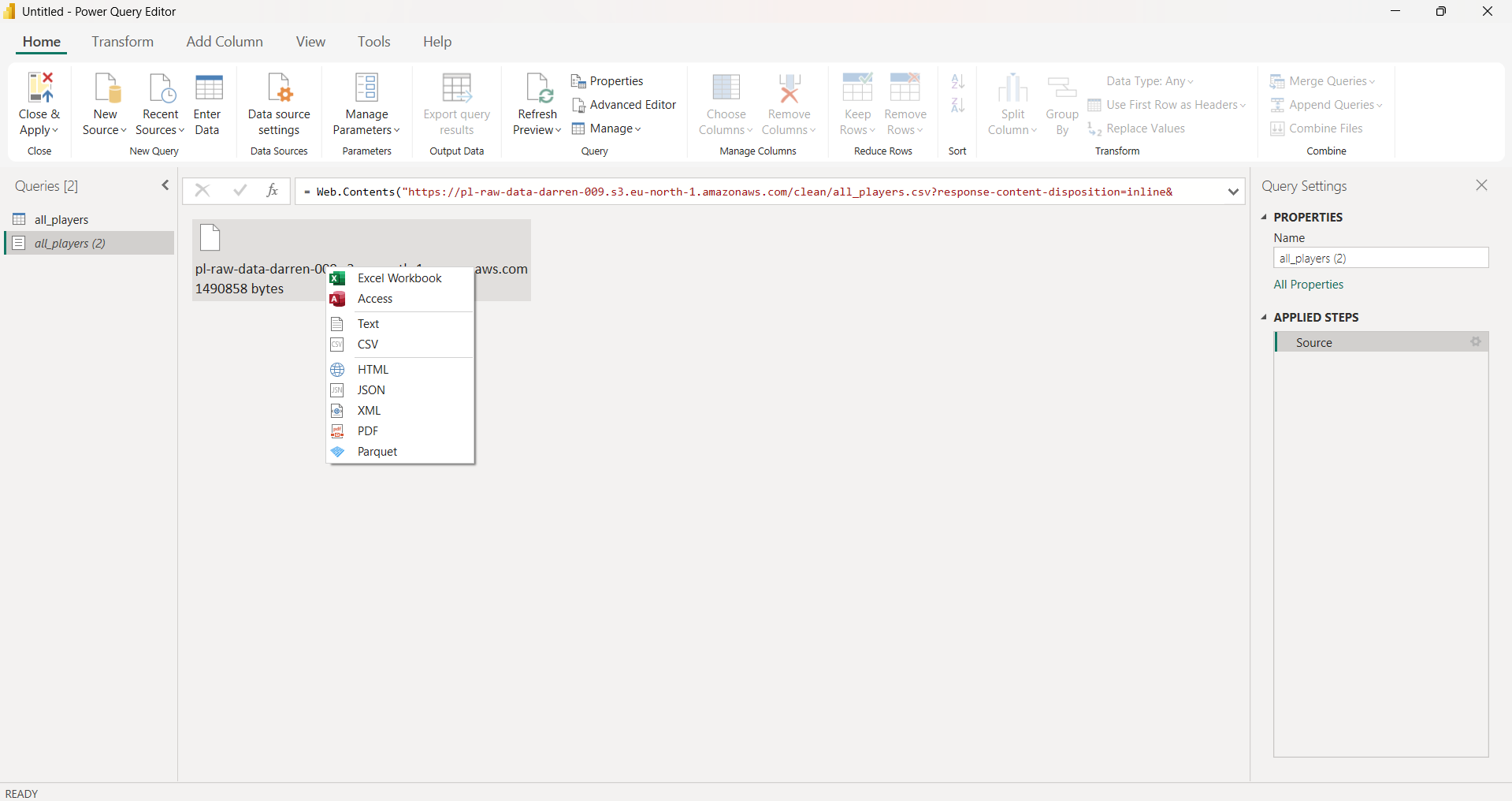


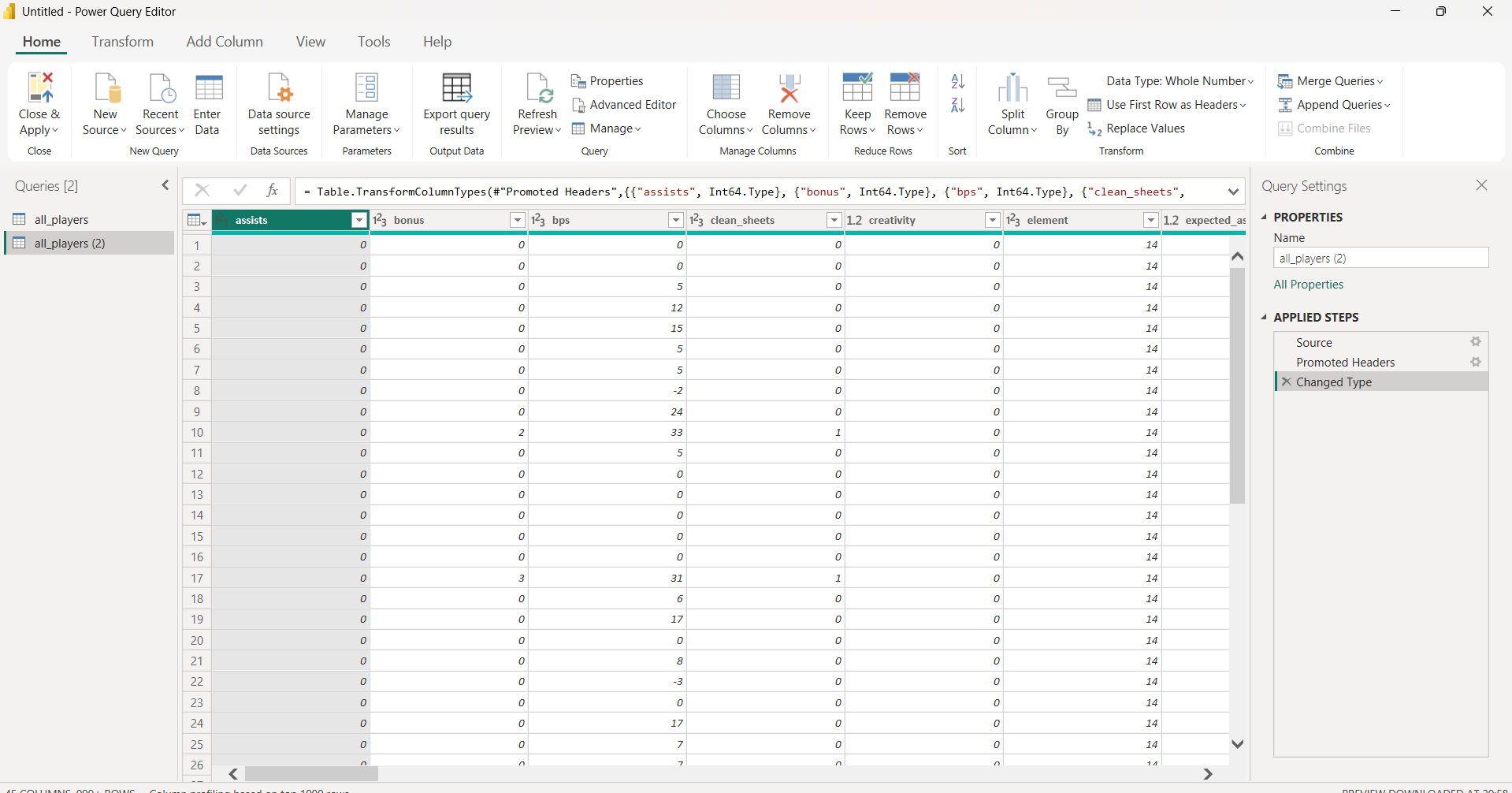


Connecting power BI to the cleaned csv

Through the generated URL from s3, we use the URL to get data in power BI from web.







Visualization

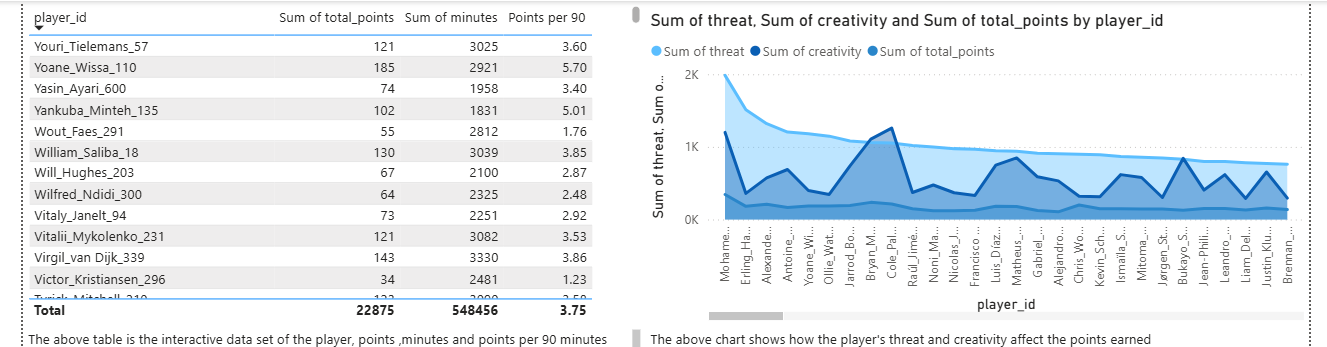
Within Power BI, advanced metrics such as points per 90 minutes, expected goal involvement, value efficiency, and form trends are calculated and visualized through interactive dashboards.

**1. Player Performance Summary Table**

This table provides an interactive summary of individual player performance metrics. It displays the total points scored by each player, total minutes played, and the calculated points per 90 minutes.

**2. Threat, Creativity, and Total Points by Player**

This visualization acts as a foundation for all player-level analysis in the dashboard showing how the creativity index and the threat index affect the total points scored by a player.



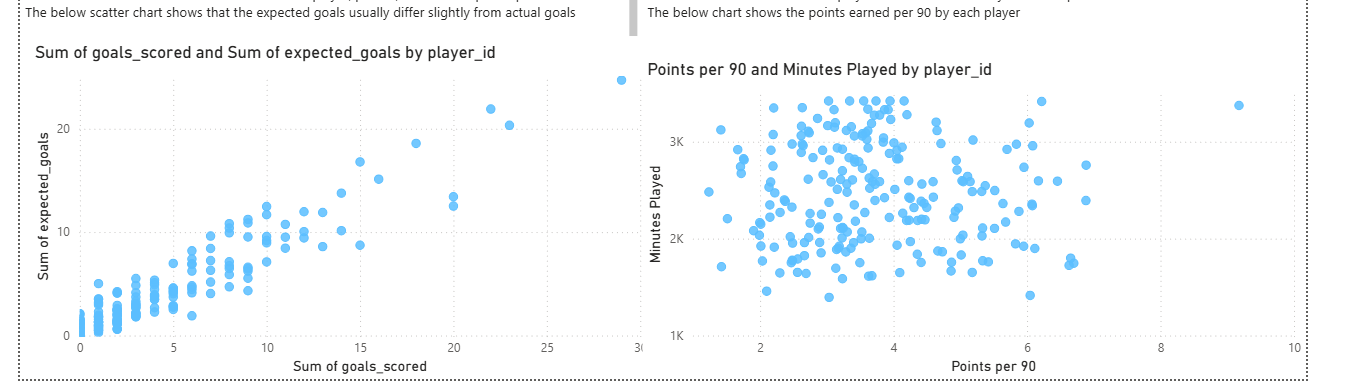
**3. Goals Scored vs Expected Goals (xG)**

This scatter plot compares the total goals scored by each player against their expected goals (xG). The chart highlights deviations between actual and expected performance.

Players scoring significantly more than their xG may be overperforming, while those scoring less than expected may be due for future improvement. This visualization is essential for identifying regression trends and avoiding reliance on short-term luck.

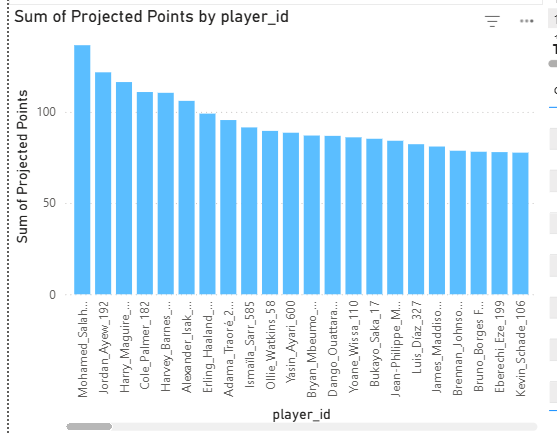
**4. Points per 90 Minutes vs Minutes Played**

This scatter plot analyzes player efficiency relative to playing time. By plotting points per 90 minutes against total minutes played, the chart helps distinguish between rotation players with high impact in limited minutes and regular starters with consistent contributions.



**5.Projected Points by Player**

This bar chart displays the **sum of projected fantasy points for each player**, ranked from highest to lowest. The projection is derived from calculated performance indicators such as expected goals, expected assists, historical points per 90 minutes, and playing time, providing an estimate of future returns rather than relying solely on past performance.



**Conclusion**

This project successfully demonstrates the design and implementation of an end-to-end data analytics solution for Fantasy Premier League performance analysis using cloud computing and business intelligence technologies. By leveraging AWS services such as IAM, Amazon S3, and AWS Lambda.

The integration of Power BI enabled the transformation of processed data into meaningful visual insights. Advanced performance metrics such as points per 90 minutes, expected goal involvement, value efficiency, and projected points were calculated and visualized through interactive dashboards. These insights allow users to evaluate player form, compare expected versus actual performance, analyse fixture influence, and make informed transfer and captaincy decisions.