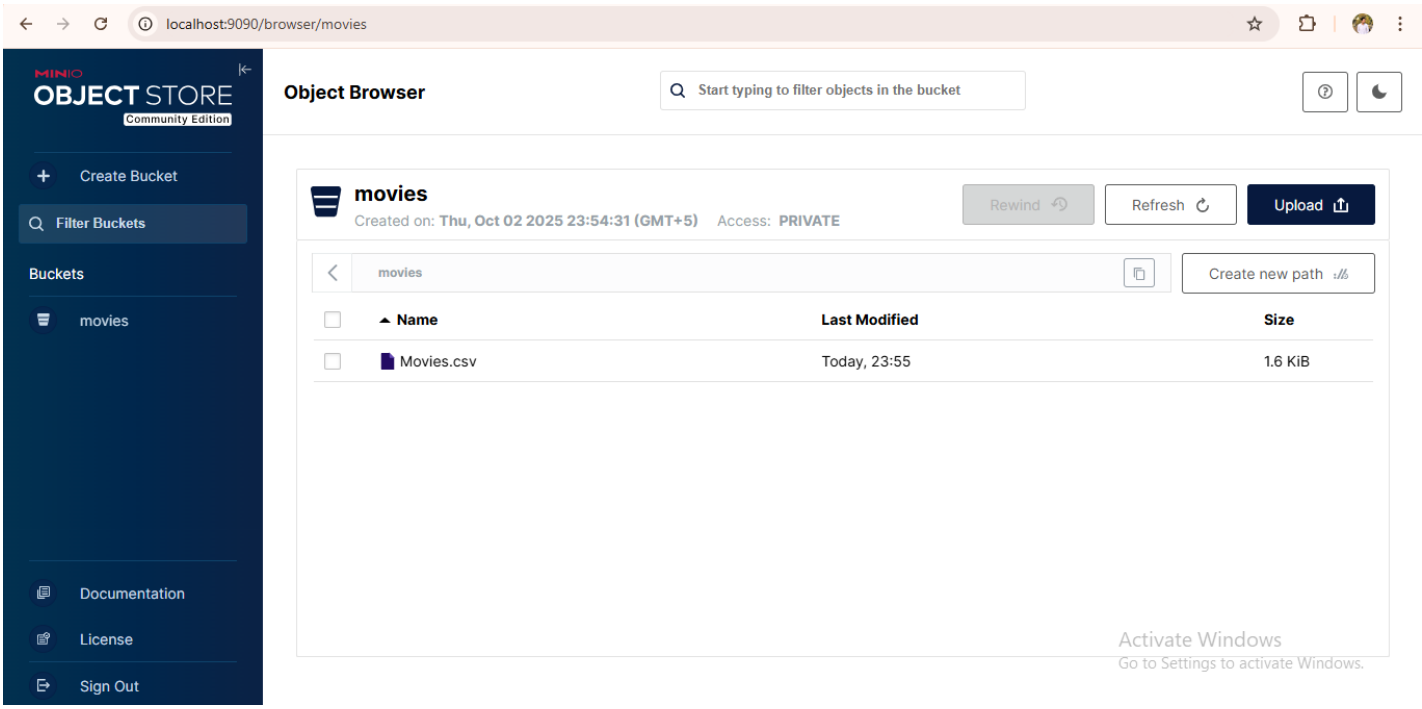
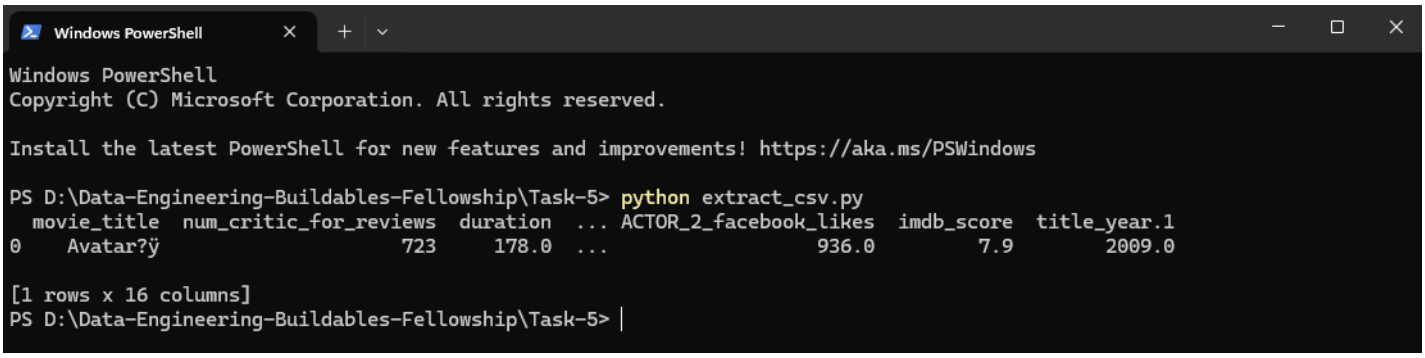
Buildables Task 5

Github Repo: https://github.com/afnank070/Data-Engineering-Buildables-Fellowship

MinIO bucket with uploaded CSV



Extracted CSV from MinIO



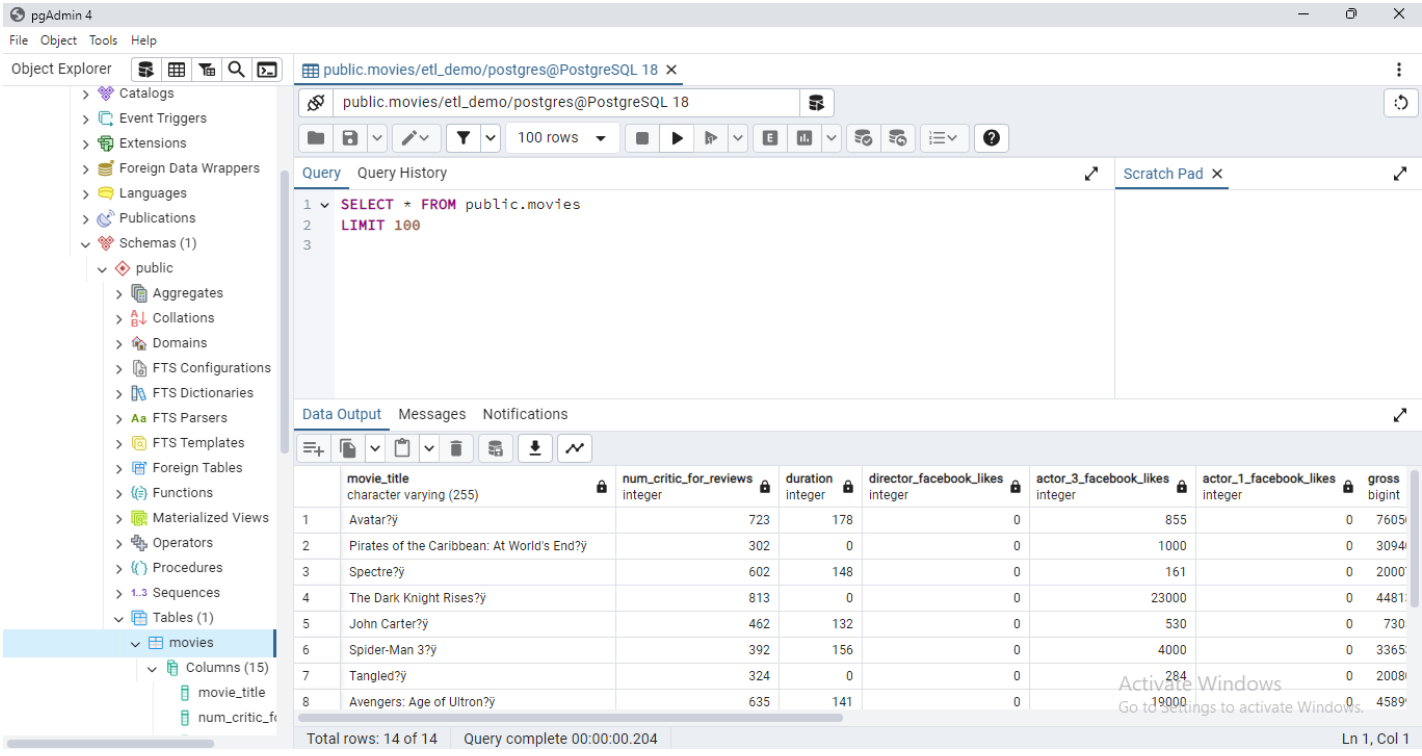
For now, in Transform step we will:

1. Fix missing values → fill empty numeric fields with 0.
2. Ensure numeric columns are numbers.
3. Remove duplicate column title_year.

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS D:\Data-Engineering-Buildables-Fellowship\Task-5> python transform_csv.py
Extracted data (first row):
  movie_title  num_critic_for_reviews  duration  ... ACTOR_2_facebook_likes  imdb_score  title_year.1
0    Avatar?ÿ                   723     178.0  ...                   936.0         7.9        2009.0

[1 rows x 16 columns]

Transformed data (first row):
  movie_title  num_critic_for_reviews  duration  ...  ACTOR_2_facebook_likes  imdb_score  title_year.1
0    Avatar?ÿ                   723     178.0  ...                   936.0         7.9        2009.0

[1 rows x 16 columns]

Transformed CSV saved as movies_transformed.csv
PS D:\Data-Engineering-Buildables-Fellowship\Task-5>
```

Loaded in PostgreSQL

```
PS D:\Data-Engineering-Buildables-Fellowship\Task-5> python load_csv.py
Data to load (first row):
  movie_title  num_critic_for_reviews  duration  ...  ACTOR_2_facebook_likes  imdb_score  title_year.1
0    Avatar?ÿ                   723     178.0  ...                   936.0         7.9        2009.0

[1 rows x 16 columns]

Data loaded into PostgreSQL successfully!
PS D:\Data-Engineering-Buildables-Fellowship\Task-5>
```



Airflow Dashboard: