DATA VALIDATION AND QUALTY REPORT:

Query1: SELECT id, title FROM netflix_shows_prod GROUP BY id, title HAVING COUNT(*) > 1;

OUTPUT:

id	title
tm127384	Monty Python and the Holy Grail
tm70993	Life of Brian
tm84618	Taxi Driver

Query2: SELECT stg.age_certification, AVG(stg.imdb_votes) AS avg_imdb_votes_stage, AVG(prod.imdb_votes) AS avg_imdb_votes_prod FROM netflix_shows_stg AS stg JOIN netflix_shows_prod AS prod ON stg.id = prod.id GROUP BY stg.age_certification;

OUTPUT:

age_certification	avg_imdb_votes_stage	avg_imdb_votes_prod
	5096.375774916547	5096.375774916547
G	7752.71	7752.71
NC-17	16151.153846153846	16151.153846153846
PG	38758.4678111588	38758.4678111588
PG-13	68007.20243902439	68007.20243902439
R	81615.9737335835	163231.947467167
TV-14	24896.16229116945	24896.16229116945
TV-G	4836.7	4836.7
TV-MA	25059.47382198953	25059.47382198953
TV-PG	10858.173913043478	10858.173913043478
TV-Y	756.3932584269663	756.3932584269663
TV-Y7	7231.60396039604	7231.60396039604

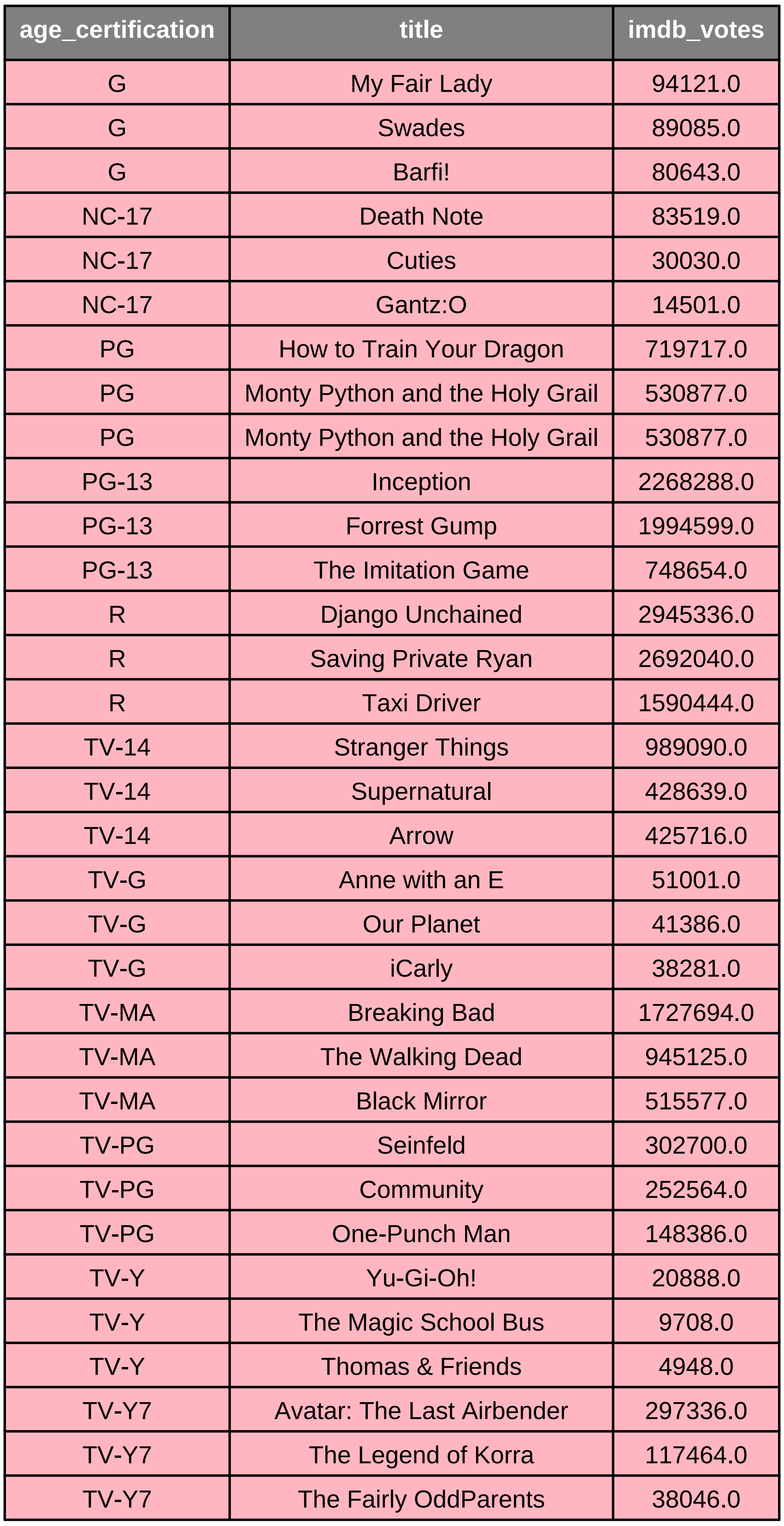
Query3: SELECT 'stage' AS environment, COUNT(*) AS total_records FROM netflix_shows_stg UNION ALL SELECT 'production' AS environment, COUNT(*) AS total_records FROM netflix_shows_prod;

OUTPUT:

environment	total_records
stage	5283
production	5003

```
Query4: SELECT age_certification, title, imdb_votes FROM ( SELECT age_certification, title, imdb_votes,
ROW_NUMBER() OVER (PARTITION BY age_certification ORDER BY imdb_votes DESC) AS rn FROM
netflix_shows_prod WHERE age_certification IS NOT NULL AND imdb_votes IS NOT NULL ) AS subquery
WHERE rn <= 3 ORDER BY age_certification, imdb_votes DESC;
```

OUTPUT:



age_certification	title	imdb_votes
G	My Fair Lady	94121.0
G	Swades	89085.0
G	Barfi!	80643.0
NC-17	Death Note	83519.0
NC-17	Cuties	30030.0
NC-17	Gantz:O	14501.0
PG	How to Train Your Dragon	719717.0
PG	Monty Python and the Holy Grail	530877.0
PG	Monty Python and the Holy Grail	530877.0
PG-13	Inception	2268288.0
PG-13	Forrest Gump	1994599.0
PG-13	The Imitation Game	748654.0
R	Django Unchained	2945336.0
R	Saving Private Ryan	2692040.0
R	Taxi Driver	1590444.0
TV-14	Stranger Things	989090.0
TV-14	Supernatural	428639.0
TV-14	Arrow	425716.0
TV-G	Anne with an E	51001.0
TV-G	Our Planet	41386.0
TV-G	iCarly	38281.0
TV-MA	Breaking Bad	1727694.0
TV-MA	The Walking Dead	945125.0
TV-MA	Black Mirror	515577.0
TV-PG	Seinfeld	302700.0
TV-PG	Community	252564.0
TV-PG	One-Punch Man	148386.0
TV-Y	Yu-Gi-Oh!	20888.0
TV-Y	The Magic School Bus	9708.0
TV-Y	Thomas & Friends	4948.0
TV-Y7	Avatar: The Last Airbender	297336.0
TV-Y7	The Legend of Korra	117464.0
TV-Y7	The Fairly OddParents	38046.0

```
Query5: SELECT column_name, SUM(null_count) AS total_nulls FROM ( SELECT 'tbl_index' AS
column_name, SUM(CASE WHEN tbl_index IS NULL THEN 1 ELSE 0 END) AS null_count FROM
netflix_shows_prod UNION ALL SELECT 'id' AS column_name, SUM(CASE WHEN id IS NULL THEN 1 ELSE
```

0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'title' AS column_name, SUM(CASE WHEN title IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'type' AS column_name, SUM(CASE WHEN type IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'description' AS column_name, SUM(CASE WHEN description IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'release_year' AS column_name, SUM(CASE WHEN release_year IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'age_certification' AS column_name, SUM(CASE WHEN age_certification IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'runtime' AS column_name, SUM(CASE WHEN runtime IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'imdb_id' AS column_name, SUM(CASE WHEN imdb_id IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'imdb_score' AS column_name, SUM(CASE WHEN imdb_score IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod UNION ALL SELECT 'imdb_votes' AS column_name, SUM(CASE WHEN imdb_votes IS NULL THEN 1 ELSE 0 END) AS null_count FROM netflix_shows_prod) AS unpivot GROUP BY column_name ORDER BY column_name;

OUTPUT:

column_name	total_nulls
age_certification	2103
description	5
id	0
imdb_id	0
imdb_score	0
imdb_votes	13
release_year	0
runtime	0
tbl_index	0
title	0
type	0

Query6: SELECT stg.type, AVG(stg.imdb_votes) AS avg_imdb_votes_stg, prod.avg_imdb_votes_prod FROM netflix_shows_stg AS stg JOIN (SELECT type, AVG(imdb_votes) AS avg_imdb_votes_prod FROM netflix_shows_prod GROUP BY type) AS prod ON stg.type = prod.type GROUP BY stg.type, prod.avg_imdb_votes_prod;

OUTPUT:

type	avg_imdb_votes_stg	avg_imdb_votes_prod
Movie	26683.217045119432	42408.32466311501
Show	17485.558102345414	18190.690383546415