

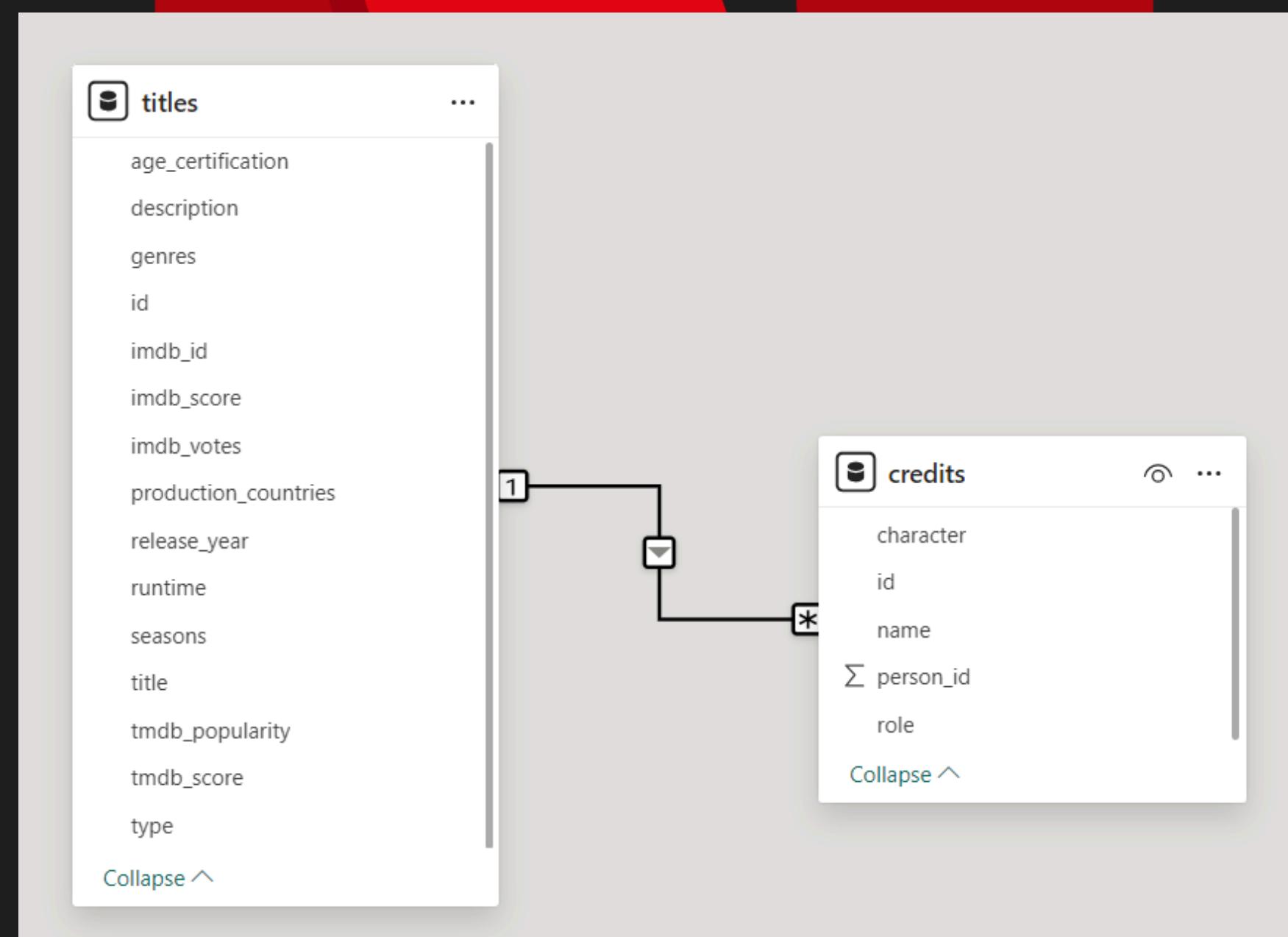


NETFLIX ANALYSIS

LET'S GET INTO WORLD OF ANALYSIS

Abhay Rajpal

Data Model



LET'S GO THROUGH QUESTIONS

HERE, I HAVE QUESTIONS REGARDING NETFLIX 'S MOVIES AND SHOWS TO SOLVE
FOR DATA ANALYSIS BY USING SQL.

-- Which were the top 10 movies based on IMDB score? --

```
SELECT TITLE, IMDB_SCORE  
FROM TITLES  
WHERE TYPE = "MOVIE"  
ORDER BY IMDB_SCORE DESC  
LIMIT 10;
```

	TITLE	IMDB_SCORE
▶	GoodFellas	8.7
	Full Metal Jacket	8.3
	Once Upon a Time in America	8.3
	Taxi Driver	8.2
	Monty Python and the Holy Grail	8.2
	The Land	8.1
	Richard Pryor: Live in Concert	8.1
	Parrot Sketch Not Included: Twenty Years of M...	8.1

-- What were the top 10 shows according to IMDB score? --

```
SELECT TITLE, TYPE, IMDB_SCORE  
FROM TITLES  
WHERE TYPE = "SHOW"  
ORDER BY IMDB_SCORE DESC  
LIMIT 10;
```

	TITLE	IMDB_SCORE
▶	Seinfeld	8.9
	Monty Python's Flying Circus	8.8
	Monty Python's Fliegender Zirkus	8.1
	Danger Mouse	7.4
	Saved by the Bell	7.1
	Knight Rider	6.9
	Wheel of Fortune	6.7
	Thomas & Friends	6.5

-- What were the bottom 10 movies according to IMDB score? --

```
SELECT TITLE, TYPE, IMDB_SCORE  
FROM TITLES  
WHERE TYPE = "MOVIE"  
ORDER BY IMDB_SCORE ASC  
LIMIT 10;
```

	TITLE	IMDB_SCORE
▶	Bill Hicks: Sane Man	
	Dostana	2.1
	Bandie	4.4
	Chadi Jawani Budhe Nu	4.6
	Endless Love	4.9
	Khoon Khoon	5.1
	Aakhri Adaalat	5.1
	Mujrim	5.4

-- What were the bottom 10 shows according to IMDB score?

```
SELECT TITLE, TYPE, IMDB_SCORE  
FROM TITLES  
WHERE TYPE = "SHOW"  
ORDER BY IMDB_SCORE ASC  
LIMIT 10;
```

	TITLE	IMDB_SCORE
▶	Five Came Back: The Reference Films	
	Fireman Sam	6.1
	Thomas & Friends	6.5
	Major Dad	6.5
	Wheel of Fortune	6.7
	Knight Rider	6.9
	Saved by the Bell	7.1
	Danger Mouse	7.4

-- What were the average IMDB and TMDB scores for shows and movies? --

```
SELECT TYPE ,ROUND(AVG(IMDB_SCORE), 2)AS AVG_IMDB_SCORE,  
ROUND(AVG(TMDB_SCORE), 2) AS AVG_TMDB_SCORE  
FROM TITLES  
GROUP BY TYPE  
ORDER BY AVG_IMDB_SCORE DESC;
```

	TYPE	AVG_IMDB_SCORE	AVG_TMDB_SCORE
▶	MOVIE	6.82	6.61
	SHOW	6.64	6.69

-- Count of movies and shows in each decade --

```
SELECT CONCAT(FLOOR(RELEASE_YEAR/10)*10, "s") AS DECADE, COUNT(*) AS TOTAL_CONTENT
FROM TITLES
GROUP BY DECADE
ORDER BY DECADE ASC;
```

	DECADE	TOTAL_CONTENT
▶	1940s	1
	1950s	5
	1960s	8
	1970s	18
	1980s	41
	1990s	6

-- What were the most common age certifications for movies?

```
SELECT AGE_CERTIFICATION ,COUNT(*) AS TOTAL  
FROM TITLES  
WHERE TYPE ="MOVIE" AND AGE_CERTIFICATION != ""  
GROUP BY AGE_CERTIFICATION  
ORDER BY TOTAL DESC;
```

	AGE_CERTIFICATION	TOTAL
▶	R	21
	PG-13	4
	PG	3

-- Who were the top 10 actors that appeared the most in movies/shows?

```
SELECT DISTINCT C.NAME, COUNT(*) AS NO_OF_APPEARANCE  
FROM TITLES T  
JOIN CREDITS C  
ON C.ID = T.ID  
WHERE C.ROLE = "ACTOR"  
GROUP BY C.NAME  
ORDER BY NO_OF_APPEARANCE DESC  
LIMIT 10;
```

	NAME	NO_OF_APPEARANCE
▶	Terry Jones	4
	Graham Chapman	3
	Michael Palin	3
	Terry Gilliam	3
	Eric Idle	3
	Carol Cleveland	2
	John Young	2
	John Cleese	2
	Neil Innes	2
	Charles Scorsese	1

```
-- Who were the top 5 directors that directed the most movies/shows?
```

```
SELECT DISTINCT C.NAME, COUNT(*) AS NO_OF_APPEARANCE  
FROM TITLES T  
JOIN CREDITS C  
ON C.ID = T.ID  
WHERE C.ROLE = "DIRECTOR"  
GROUP BY C.NAME  
ORDER BY NO_OF_APPEARANCE DESC  
LIMIT 5;
```

	NAME	NO_OF_APPEARANCE
▶	Terry Jones	2
	Martin Scorsese	1
	John Boorman	1
	Terry Gilliam	1
	Robert Aldrich	1

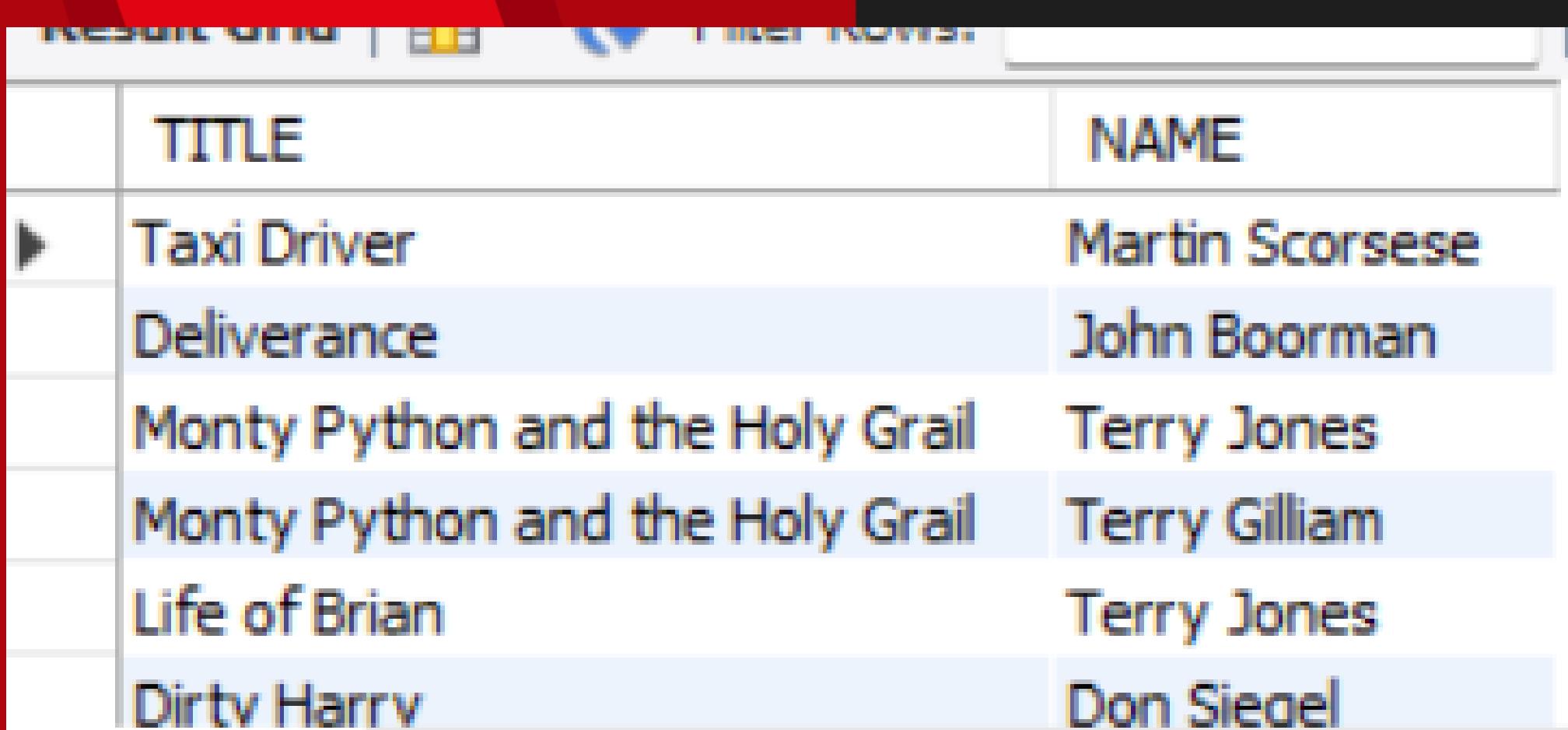
```
-- Calculating the average runtime of movies and TV shows separately
```

```
SELECT TYPE AS CONTENT_TYPE, ROUND(AVG(RUNTIME),2) AS AVG_RUNTIME  
FROM TITLES  
GROUP BY TYPE  
ORDER BY AVG_RUNTIME DESC;
```

	CONTENT_TYPE	AVG_RUNTIME
▶	MOVIE	115.28
	SHOW	27.82

-- Finding the titles and directors of movies released on or after 1970

```
SELECT T.TITLE, C.NAME  
FROM TITLES T  
JOIN CREDITS C  
ON C.ID = T.ID  
WHERE C.ROLE = "DIRECTOR" AND RELEASE_YEAR >= 1970;
```



The screenshot shows the MySQL Workbench interface with the results of the SQL query. The results are displayed in a table with two columns: 'TITLE' and 'NAME'. The table contains six rows, with the second and third rows being identical.

	TITLE	NAME
▶	Taxi Driver	Martin Scorsese
▶	Deliverance	John Boorman
▶	Monty Python and the Holy Grail	Terry Jones
▶	Monty Python and the Holy Grail	Terry Gilliam
▶	Life of Brian	Terry Jones
▶	Dirty Harry	Don Siegel

-- Which shows on Netflix have the most seasons?

```
SELECT TITLE, SEASONS  
FROM TITLES  
WHERE TYPE = "SHOW"  
ORDER BY SEASONS DESC;
```

	TITLE	SEASONS
▶	Seinfeld	9.0
	Saved by the Bell	5.0
	Monty Python's Flying Circus	4.0
	Knight Rider	4.0
	Major Dad	4.0
	Wheel of Fortune	39.0

-- Which genres had the most movies?

```
SELECT GENRES, COUNT(*) AS MOST  
FROM TITLES  
WHERE TYPE = "MOVIE"  
GROUP BY GENRES  
ORDER BY MOST DESC  
LIMIT 5;
```

	GENRES	MOST
▶	[drama]	7
	[comedy', 'documentation']	5
	[comedy]	4
	[drama', 'crime']	3
	[romance', 'comedy']	2

-- Which genres had the most shows?

```
SELECT GENRES, COUNT(*) AS MOST  
FROM TITLES  
WHERE TYPE = "SHOW"  
GROUP BY GENRES  
ORDER BY MOST DESC  
LIMIT 5;
```

	GENRES	MOST
▶	[comedy]	2
	[documentation]	1
	[comedy, 'european']	1
	[scifi, 'action', 'crime', 'drama']	1
	[animation, 'family', 'comedy', 'fantasy', 'drama'...]	1

-- Titles and Directors of movies with high IMDB scores (>7.5) and high TMDB popularity scores (>8)

```
SELECT T.TITLE, C.ROLE  
FROM TITLES T  
JOIN CREDITS C  
ON C.ID = T.ID  
WHERE T.TYPE = "MOVIE"  
AND C.ROLE = "DIRECTOR"  
AND T.IMDB_SCORE > 7.5  
AND T.TMDB_SCORE > 8;
```

	TITLE	ROLE
▶	Taxi Driver	DIRECTOR

-- What were the total number of titles for each year?

```
SELECT RELEASE_YEAR, COUNT(*) AS TOTAL  
FROM TITLES  
GROUP BY RELEASE_YEAR  
ORDER BY TOTAL DESC;
```

	RELEASE_YEAR	TOTAL
▶	1989	12
	1990	6
	1979	5
	1988	5
	1987	4
	1984	4
	1982	4

-- Actors who have starred in the most highly rated movies or shows

```
SELECT C.NAME, COUNT(*) AS MOST
FROM TITLES T
JOIN CREDITS C
ON C.ID = T.ID
WHERE C.ROLE = "ACTOR" AND T.IMDB_SCORE >= 7.5 AND T.TMDB_SCORE >= 7.5
GROUP BY C.NAME
ORDER BY MOST DESC;
```

	NAME	MOST
▶	Terry Jones	4
	Graham Chapman	3
	Eric Idle	3
	Terry Gilliam	3
	Michael Palin	3
	John Cleese	2
	Carol Cleveland	2

--- What were the top 3 most common genres? ---

```
SELECT GENRES, COUNT(*) AS NO_OF_COUNTS  
FROM TITLES  
GROUP BY GENRES  
ORDER BY NO_OF_COUNTS DESC  
LIMIT 3;
```

	GENRES	NO_OF_COUNTS
▶	[drama]	7
	[comedy]	6
	[comedy', 'documentation']	5

```
-- Average IMDB score for leading actors/actresses in movies or shows --
```

```
SELECT C.NAME ,ROUND(AVG(T.IMDB_SCORE), 2) AS AVG_IMDB
FROM TITLES T
JOIN CREDITS C
ON C.ID = T.ID
WHERE C.ROLE ="ACTOR" OR C.ROLE ="ACTRESS" AND C.CHARACTER ="LEADING ROLE"
GROUP BY C.NAME
ORDER BY AVG_IMDB DESC;
```

	NAME	AVG_IMDB
▶	Michael Palin	8.33
	Terry Gilliam	8.33
	Eric Idle	8.33
	Graham Chapman	8.33
	Terry Jones	8.3
	Charles Scorsese	8.2
	Catherine Scorsese	8.2



**THANK YOU FOR
YOUR ATTENTION**