



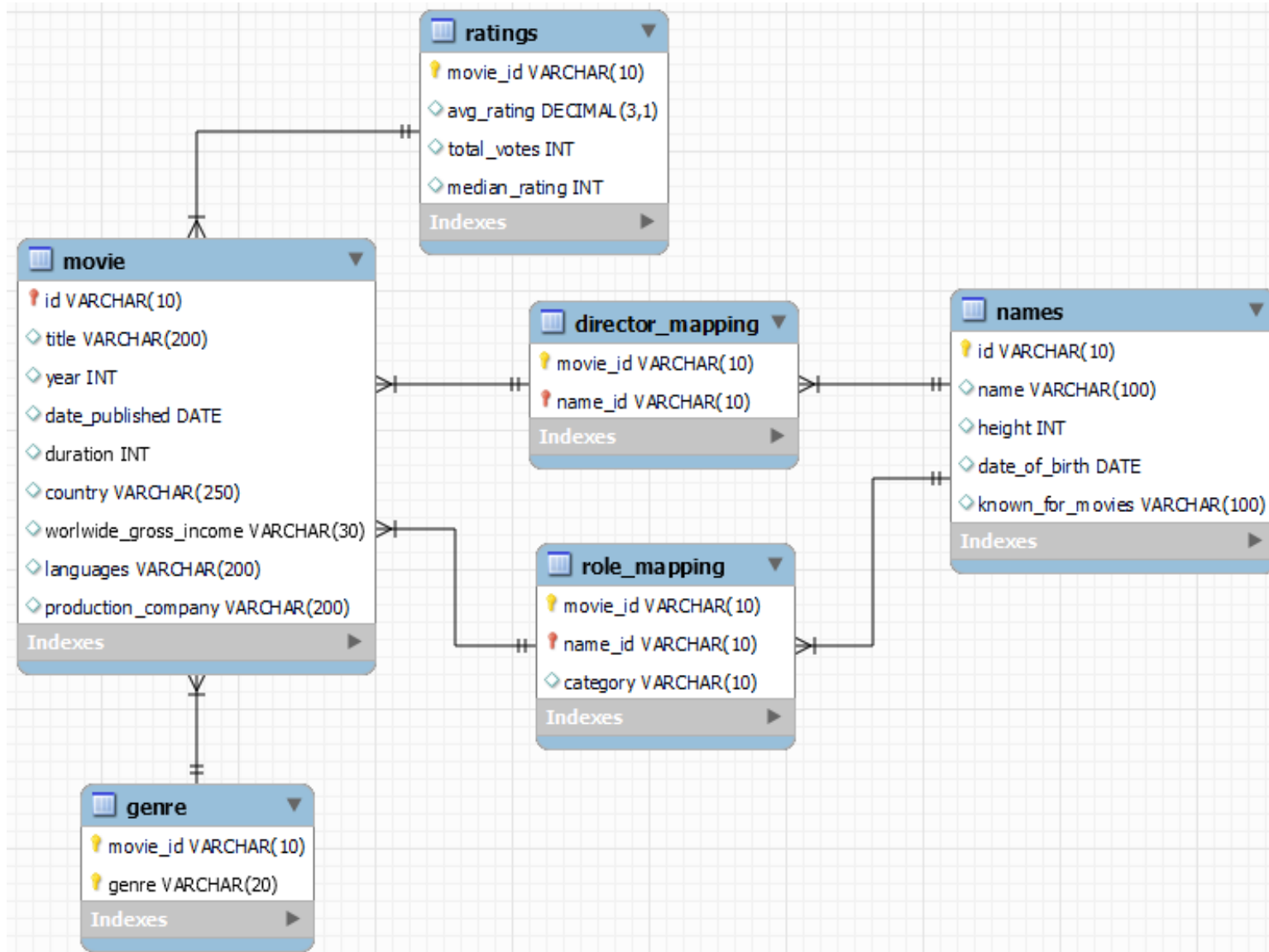
# IMDb Uncovered

Revealing Insights Through Data Analysis



Made by: Ahmad Shafique

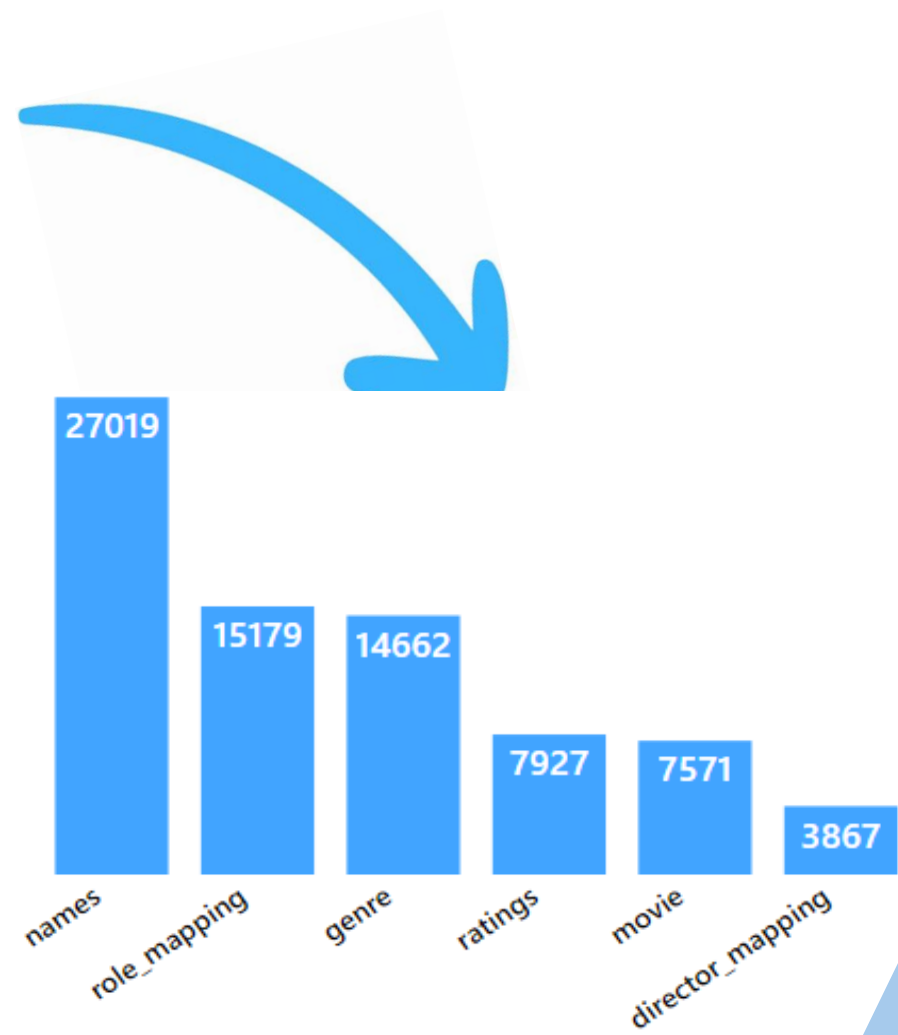
# Relational Schema



1

# Total rows in each table?

```
1 SELECT table_name,  
2      table_rows  
3 FROM INFORMATION_SCHEMA.TABLES  
4 WHERE TABLE_SCHEMA = 'imdb'  
5 ORDER BY table_rows DESC;
```



## Total Movies in Each Genre?

```
1 SELECT genre,  
2     COUNT(*) as total_movies  
3 FROM genre  
4 GROUP BY genre  
5 ORDER BY total_movies DESC;
```



## Number of Movies Released by Country



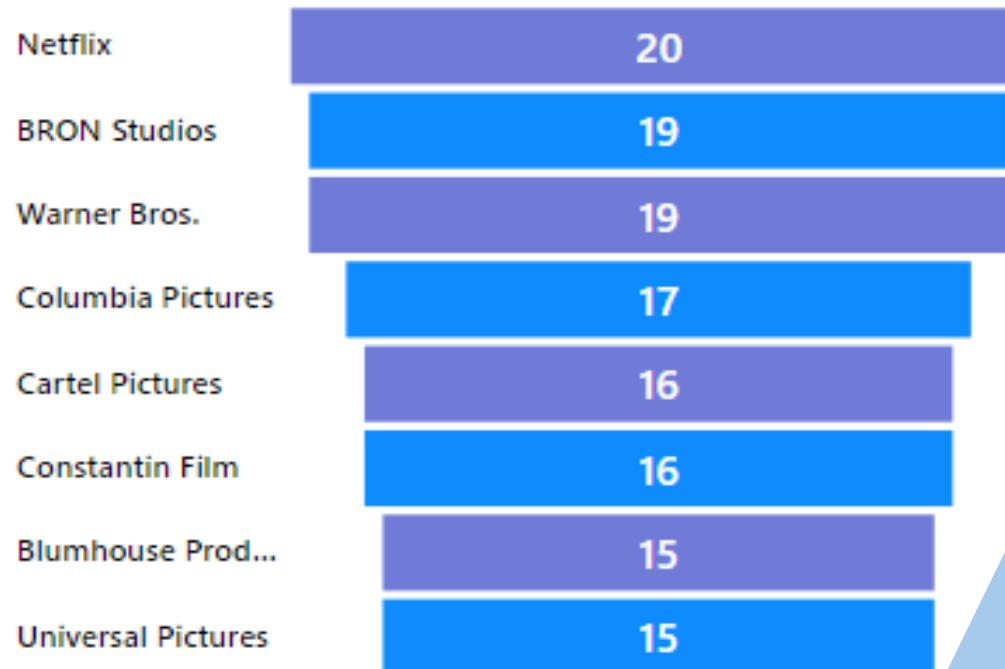
```

1  WITH numbering AS (SELECT (a.N + b.N * 10 + 1) AS n
2  FROM (SELECT 0 AS N
3         UNION ALL SELECT 1
4         UNION ALL SELECT 2
5         UNION ALL SELECT 3
6         UNION ALL SELECT 4
7         UNION ALL SELECT 5) AS a
8  CROSS JOIN (SELECT 0 AS N
9              UNION ALL SELECT 1) AS b),
10 country_table AS (SELECT SUBSTRING_INDEX(SUBSTRING_INDEX(m.country, ',', n.n), ',', -1) AS country
11                  FROM movie m
12                  INNER JOIN numbering n
13                  WHERE n.n ≤ 1 + (LENGTH(m.country) - LENGTH(REPLACE(m.country, ',', ''))))
14 SELECT country,
15        COUNT(*) AS total_movies
16 FROM country_table
17 GROUP BY country
18 ORDER BY total_movies DESC;

```

# Productions Houses with the Highest Number of Movies Released

```
1 SELECT production_company,  
2         COUNT(*) AS total_movies  
3 FROM movie  
4 WHERE production_company is NOT NULL  
5 GROUP BY production_company  
6 ORDER BY total_movies DESC  
7 LIMIT 8;
```



# Average Rating and Total Reviews per Genre

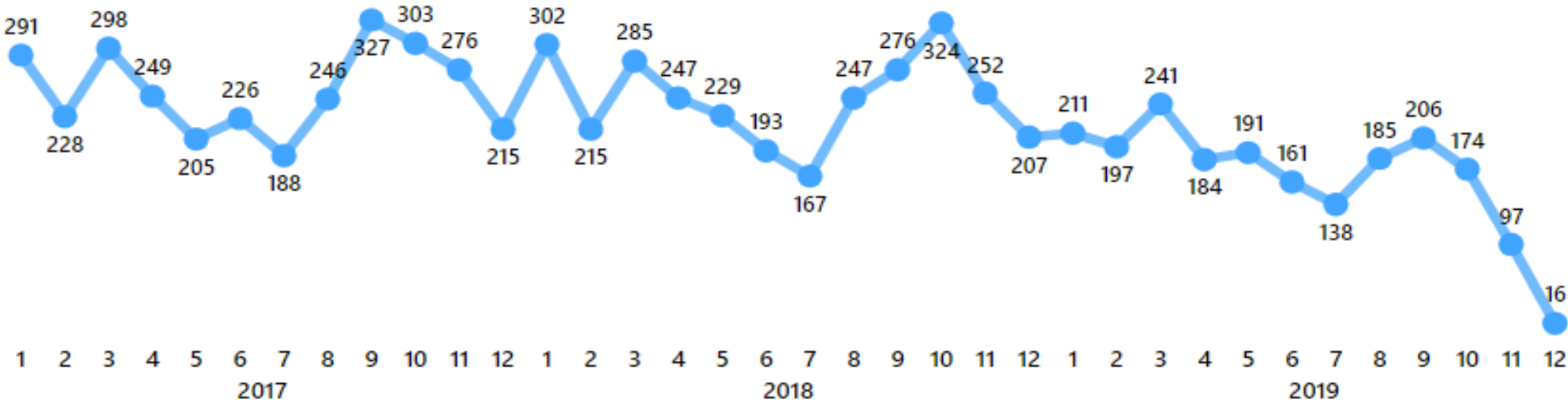
```
1 SELECT g.genre,  
2        ROUND(AVG(r.avg_rating),2) as average_rating,  
3        SUM(r.total_votes) AS total_reviews  
4 FROM ratings r  
5 INNER JOIN genre g  
6 ON r.movie_id=g.movie_id  
7 GROUP BY g.genre  
8 ORDER BY total_reviews DESC;
```



genre	average_rating	total_reviews
Drama	6.21	27170125
Action	5.62	22405065
Adventure	5.59	17048266
Comedy	5.71	13223552
Thriller	5.45	9824605
Sci-Fi	5.07	8337071
Horror	4.86	8052705
Crime	5.93	7806240
Fantasy	5.67	5881077
Mystery	5.65	5552778
Romance	5.99	3543952
Family	5.95	1751469
Others	6.17	42072

# Movies released every month?

```
1 SELECT EXTRACT(YEAR_MONTH FROM date_published) AS year_months,
2        COUNT(*) AS total_movies_released
3 FROM movie
4 GROUP BY EXTRACT(year_month from date_published)
5 ORDER BY year_months;
```





# How Many Directors Have Also Worked as Actors?

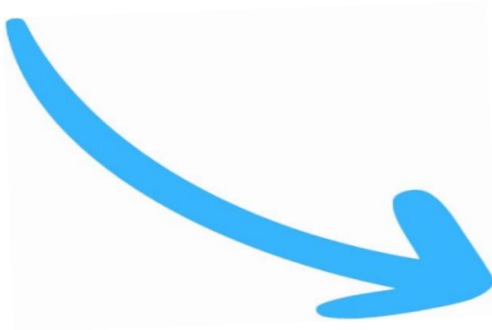
```
1 SELECT COUNT(DISTINCT d.name_id) AS total
2 FROM director_mapping d
3 INNER JOIN role_mapping r
4 ON d.name_id=r.name_id;
```



	total
▶	69

# Highest Rated Movie of Each Quarter Including Rating

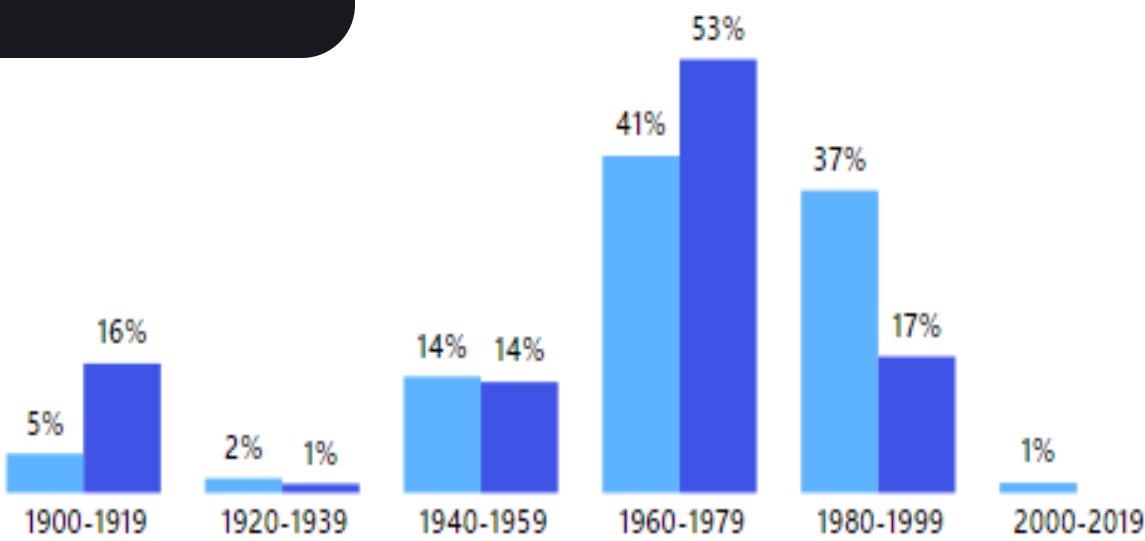
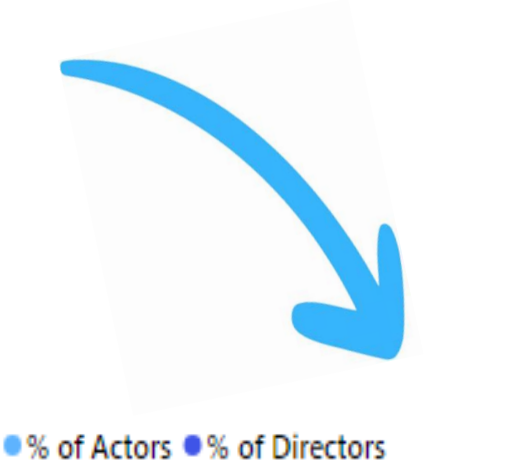
```
1 SELECT title,
2     quarter_year,
3     avg_rating
4 FROM(SELECT m.title,
5     m.date_published,
6     m.year,
7     CONCAT('Q', EXTRACT(QUARTER FROM date_published), ' ', m.year) AS quarter_year,
8     r.avg_rating,
9     DENSE_RANK() OVER(PARTITION BY CONCAT(EXTRACT(QUARTER FROM date_published), '-', m.year) ORDER BY avg_rating DESC) AS ranks
10 FROM movie m
11 INNER JOIN ratings r
12 ON m.id = r.movie_id) AS derived_table
13 WHERE ranks=1
14 ORDER BY year;
```



	title	quarter_year	avg_rating
▶	Wheels	Q1 2017	9.3
	The Colour of Darkness	Q2 2017	9.1
	National Theatre Live: Angels in America Part T...	Q3 2017	9.0
	Our Little Haven	Q4 2017	9.4
	Ekvtime: Man of God	Q1 2018	9.2
	AA BB KK	Q2 2018	9.2
	C/o Kancharapalem	Q3 2018	9.1
	Ananthu V/S Nusrath	Q4 2018	9.4
	Gini Helida Kathe	Q1 2019	9.8
	Love in Kilnerry	Q2 2019	10.0
	Fan	Q3 2019	9.6
	Kirket	Q4 2019	10.0

# Birth-year Ranges of Directors and Actors

```
1 WITH dir_dob AS (SELECT id,
2                   CONCAT(FLOOR(EXTRACT(YEAR FROM date_of_birth)/20)*20, '-', FLOOR(EXTRACT(YEAR FROM date_of_birth)/20)*20+19) AS year_range
3                   FROM names
4                   WHERE date_of_birth IS NOT NULL AND id IN(SELECT name_id
5                                                             FROM director_mapping)
6                   ),
7 act_dob AS (SELECT id,
8                CONCAT(FLOOR(EXTRACT(YEAR FROM date_of_birth)/20)*20, '-', FLOOR(EXTRACT(YEAR FROM date_of_birth)/20)*20+19) AS year_range
9                FROM names
10               WHERE date_of_birth IS NOT NULL AND id IN(SELECT name_id
11                                                         FROM role_mapping)
12               )
13 SELECT a.year_range,
14        d.dir_count,
15        a.act_count
16 FROM (SELECT year_range,
17             COUNT(id) AS act_count
18       FROM act_dob
19       GROUP BY year_range) AS a
20 LEFT JOIN (SELECT year_range,
21                 COUNT(id) AS dir_count
22            FROM dir_dob
23            GROUP BY year_range) AS d
24 ON a.year_range = d.year_range;
```



# Highest & Lowest Rated Movies

```
1 SELECT title AS lowest_rated_movies
2 FROM movie
3 WHERE id IN (SELECT movie_id
4              FROM ratings
5              WHERE avg_rating = (SELECT MIN(avg_rating)
6                                FROM ratings));
```



	lowest_rated_movies
▶	Ritoru Kyouta no bouken
	Pure Hearts: Into Chinese Showbiz
	Cumali Ceber: Allah Seni Alsin
	Badang

```
1 SELECT title AS highest_rated_movies
2 FROM movie
3 WHERE id IN (SELECT movie_id
4              FROM ratings
5              WHERE avg_rating = (SELECT MAX(avg_rating)
6                                FROM ratings));
```



	highest_rated_movies
▶	Kirket
	Love in Kilnerry

# 11 Average Duration Trends of Movies Over the Years

```
1 SELECT ROUND(AVG(duration)) AS duration_in_minutes
2 FROM movie;
```

```
1 SELECT year,
2         ROUND(AVG(duration)) AS duration_in_minutes
3 FROM movie
4 GROUP BY year
5 ORDER BY year;
```

Avg.  
Duration

104 minutes

2017

103 minutes

2018

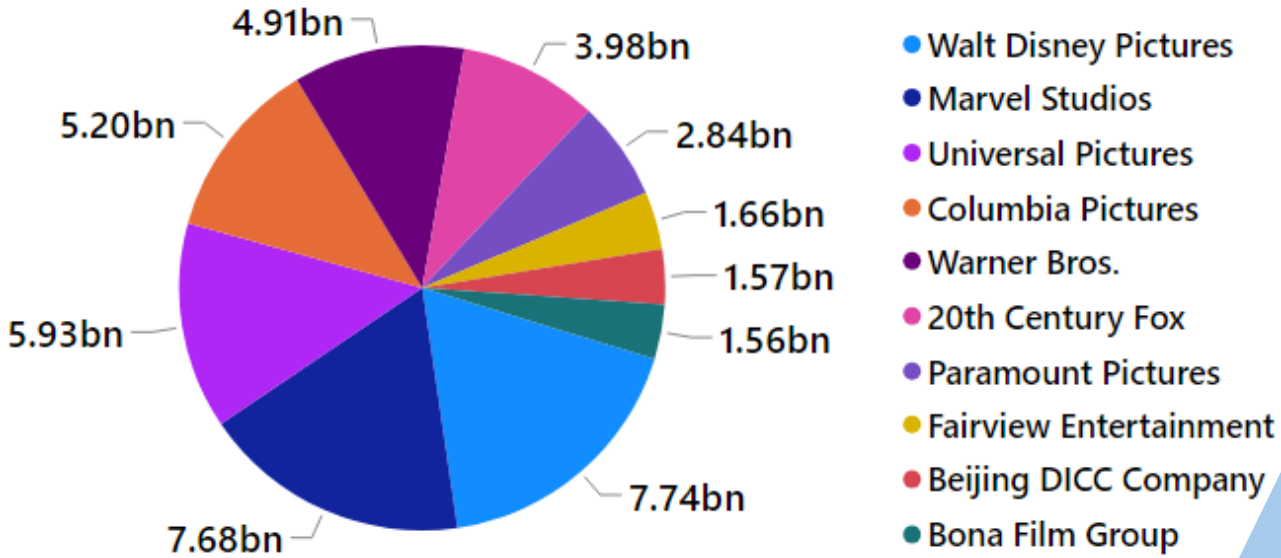
104 minutes

2019

106 minutes

# Leading Production Houses by Total Revenue

```
1 SELECT production_company,  
2        SUM(SUBSTRING(worldwide_gross_income, LOCATE(' ', worldwide_gross_income)+1)) AS total_worldwide_gross_income  
3 FROM movie  
4 WHERE worldwide_gross_income is NOT NULL  
5 GROUP BY production_company  
6 ORDER BY total_worldwide_gross_income DESC;
```



## Key Findings

- There were 6 tables in the dataset
- Drama is the most popular genre, followed by comedy and thriller
- Majority of movies were released in the USA, followed by Europe and India. Very few movies were released in Continent Africa.
- Netflix produced the most number of movies
- Movies from the drama genre had the highest average rating and the most reviews. Horror movies have the lowest average rating.
- Movies exhibit seasonality in release patterns, with the fewest releases in August and the highest in September and October.
- There were 69 directors who also acted as actors in one or more films.
- Top-rated movie of each quarter has an average rating greater than 9.
- Directors tend to be older, with many born before 1980, while actors are predominantly born after 1980.
- The average duration of a movie is 104 minutes. However, there is an increase of 1-2 minutes per year in the average duration of movie.
- Based on worldwide gross revenue, Walt Disney Pictures tops the chart with 7.74 bn, closely followed by Marvel Studios (7.68 billion).