

INFO 330 FINAL PROJECT

GROUP 15

*Lydia Sorbo, Vera Guber,
Helen Kuang, Sam
Sommerville, Jasmine
Yearian*

DATASET + DATA

- *Top 10000 Popular Movies (Main Dataset)*

- *Found on Kaggle*
- *Gathered from TMDB API*
- *Most of movie attributes*

- *IMDB Movies*

- *Found on Data.world*
- *Actors, directors, writers*

- *TMDB Movies*

- *Found on Kaggle*
- *Used to link IMDB and Top 10k*

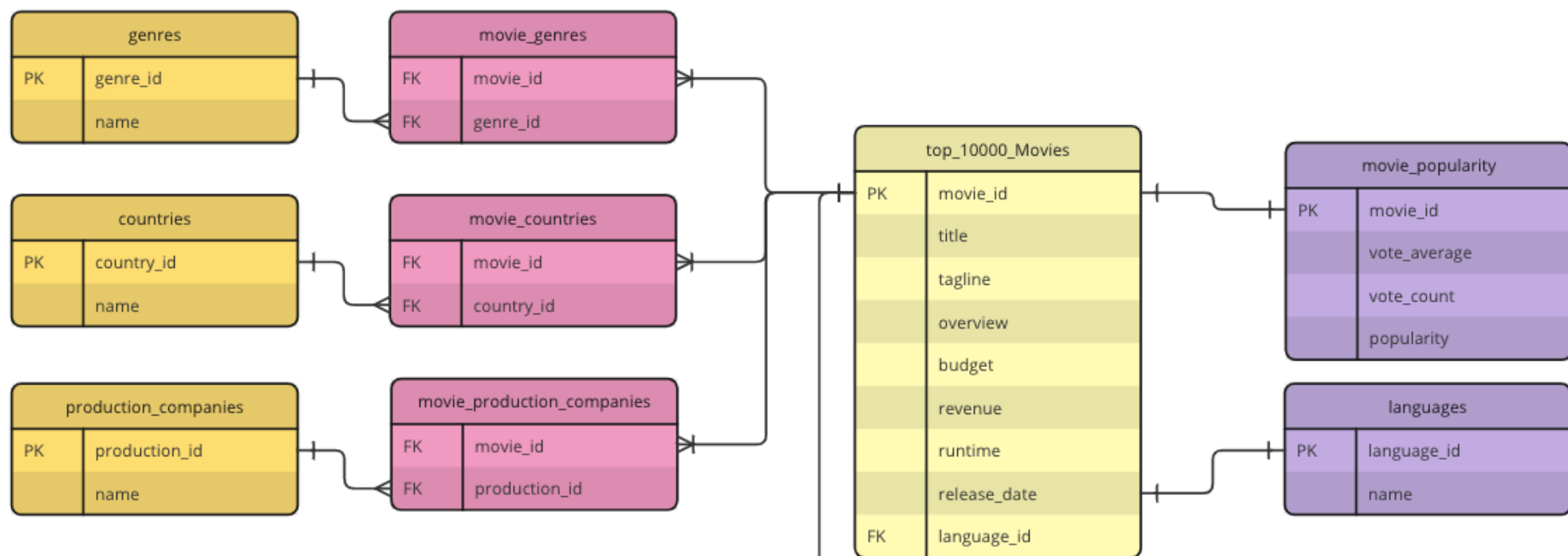
id	title	release_date	genres	original_language	vote_average	vote_count	popula
758323	The Pope's Exorcist	2023-04-05	['Horror', 'Mystery', 'Thriller']	English	7.4	619	5089.5
640146	Ant-Man and the Wasp: Quantumania	2023-02-15	['Action', 'Adventure', 'Science Fiction']	English	6.6	2294	4665.4
502356	The Super Mario Bros. Movie	2023-04-05	['Animation', 'Adventure', 'Family', 'Fantasy', 'Comedy']	English	7.5	1861	3935.5
868759	Ghosted	2023-04-18	['Action', 'Comedy', 'Romance']	English	7.2	652	2791.5
594767	Shazam! Fury of the Gods	2023-03-15	['Action', 'Comedy', 'Fantasy', 'Adventure']	English	6.8	1510	2702.5
76600	Avatar: The Way of Water	2022-12-14	['Science Fiction', 'Adventure', 'Action']	English	7.7	7853	2280.5
447365	Guardians of the Galaxy Volume 3	2023-05-03	['Science Fiction', 'Adventure', 'Action']	English	8.3	683	2520.5
934433	Scream VI	2023-03-08	['Horror', 'Mystery', 'Thriller']	English	7.3	1029	1862.5
677179	Creed III	2023-03-01	['Drama', 'Action']	English	7.3	1298	1894.5
493529	Dungeons & Dragons: Honor Among Thieves	2023-03-23	['Adventure', 'Fantasy', 'Comedy']	English	7.5	964	1655.5

imdb_title_id	title	original_title	year	date_published	genre	duration	country	lang
tt0000574	The Story of the Kelly Gang	The Story of the Kelly Gang	1906	1906-12-26	Biography, Crime, Drama	70	Australia	
tt0001892	Den sorte drøm	Den sorte drøm	1911	1911-08-19	Drama	53	Germany, Denmark	
tt0002101	Cleopatra	Cleopatra	1912	1912-11-13	Drama, History	100	USA	Eng
tt0002130	L'Inferno	L'Inferno	1911	1911-03-06	Adventure, Drama, Fantasy	68	Italy	Italia
tt0002199	From the Manger to the Cross; or, Jesus of Nazareth	From the Manger to the Cross; or, Jesus of Nazareth	1912	1913	Biography, Drama	60	USA	Eng
tt0002423	Madame DuBarry	Madame DuBarry	1919	1919-11-26	Biography, Drama, Romance	85	Germany	Ger
tt0002445	Quo Vadis?	Quo Vadis?	1913	1913-03-01	Drama, History	120	Italy	Italia
tt0002452	Independenta Romaniei	Independenta Romaniei	1912	1912-09-01	History, War	120	Romania	
tt0002461	Richard III	Richard III	1912	1912-10-15	Drama	55	France, USA	Eng
tt0002646	Atlantis	Atlantis	1913	1913-12-26	Drama	121	Denmark	Dan
tt0002844	Fantômas - À l'ombre de la guillotine	Fantômas - À l'ombre de la guillotine	1913	1913-05-12	Crime, Drama	54	France	Fre
tt0003014	Ingeborg Holm	Ingeborg Holm	1913	1913-10-27	Drama	96	Sweden	

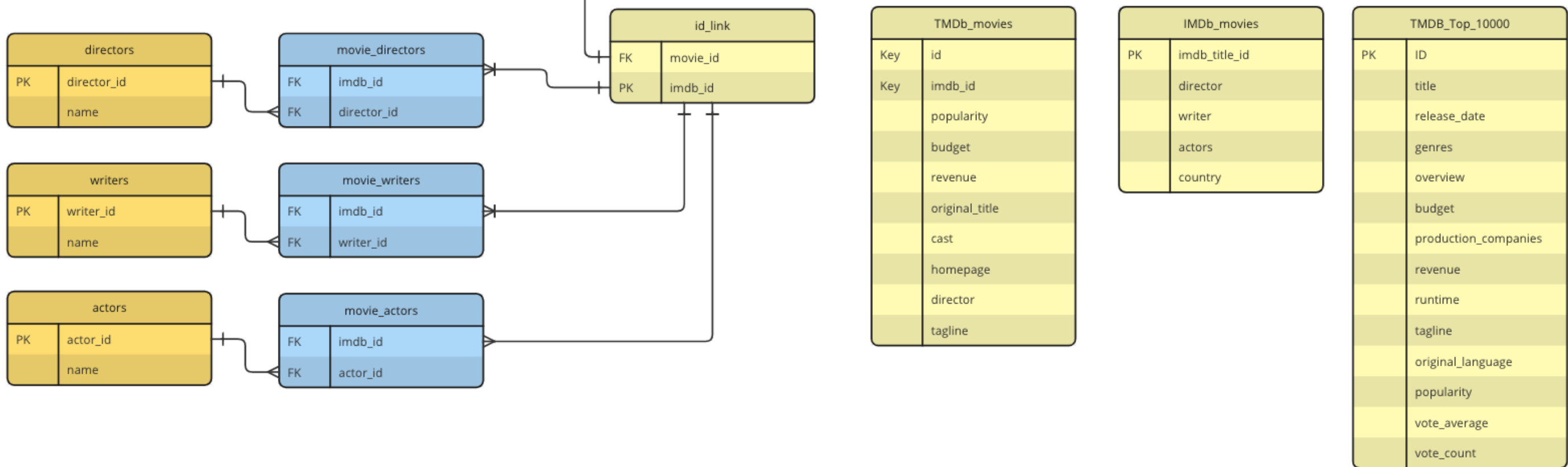
id	imdb_id	popularity	budget	revenue	original_title	cast
135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt Bryce Dallas Howard Irrfan Khan Vincent D'Onofrio Nick Robinson
76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy Charlize Theron Hugh Keays-Byrne Nicholas Hoult Josh Helman
262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley Theo James Kate Winslet Ansel Elgort Miles Teller
140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford Mark Hamill Carrie Fisher Adam Driver Daisy Ridley
168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel Paul Walker Jason Statham Michelle Rodriguez Dwayne Johnson
281957	tt1663202	9.1107	135000000	532950503	The Revenant	Leonardo DiCaprio Tom Hardy Will Poulter Domhnall Gleeson Paul Anderson
87101	tt1340138	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger Jason Clarke Emilia Clarke Jai Courtney J.K. Simmons
286217	tt3659388	7.6674	108000000	595380321	The Martian	Matt Damon Jessica Chastain Kristen Wiig Jeff Daniels Michael PeÄ±a
211672	tt2293640	7.404165	74000000	1156730962	Minions	Sandra Bullock Jon Hamm Michael Keaton Allison Janney Steve Coogan
150540	tt2096673	6.326804	175000000	853708609	Inside Out	Amy Poehler Phyllis Smith Richard Kind Bill Hader Lewis Black
206647	tt2379713	6.200282	245000000	880674609	Spectre	Daniel Craig Christoph Waltz LÃ©a Seydoux Ralph Fiennes Monica Bellucci
76757	tt1617661	6.189369	176000003	183987723	Jupiter Ascending	Mila Kunis Channing Tatum Sean Bean Eddie Redmayne Douglas Booth

ERD

topMovies.DB

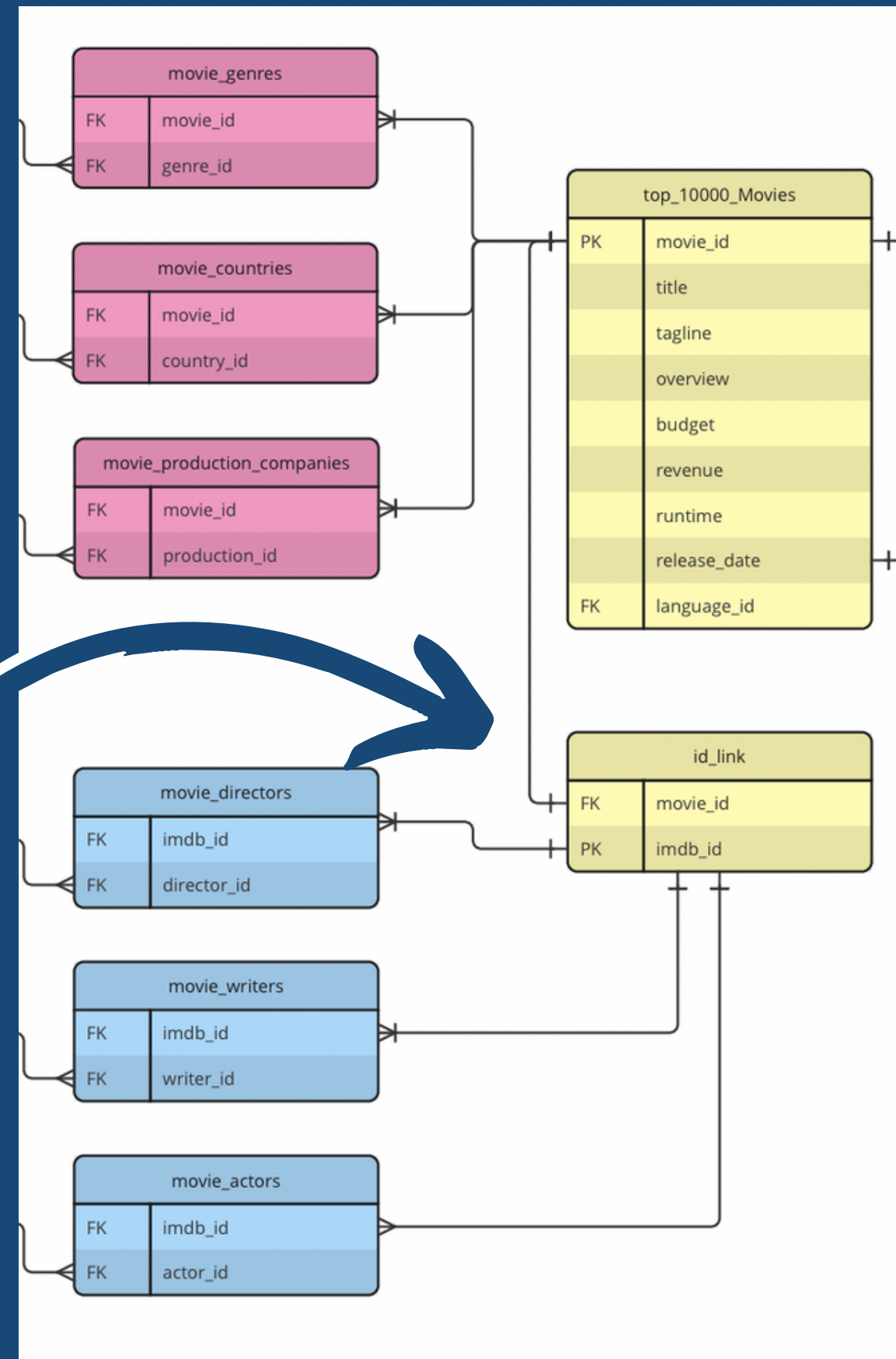


Reference tables (from datasets)



SCHEMA DESIGN

- Use of an id_link table
 - Able to expand movie attributes
- Use of other junction tables for many to many relationships
 - Ex: Genres, Directors, etc.



LYDIA

QUERY 1

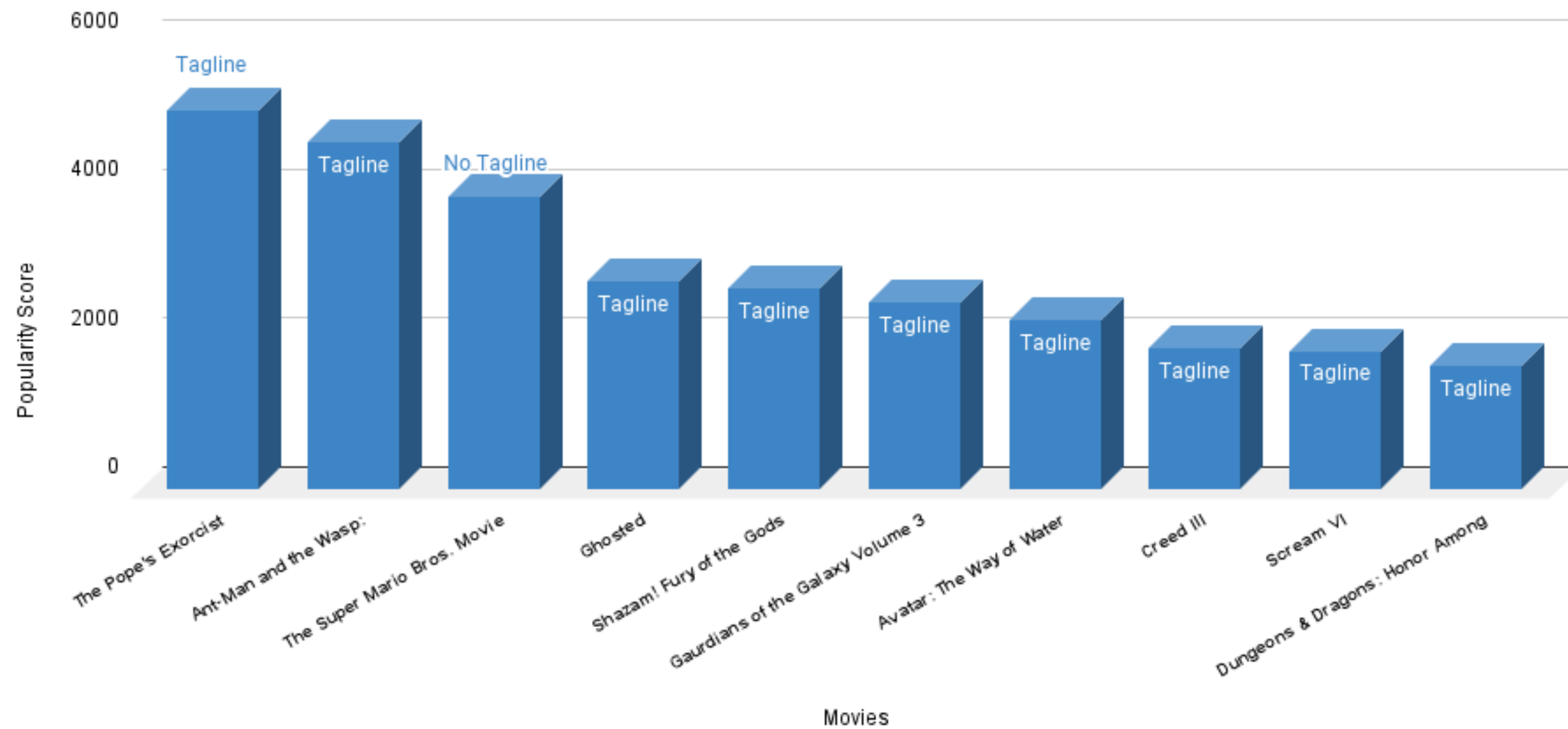
```
SELECT m.title, CASE WHEN m.tagline = '' THEN 'No' ELSE 'Yes' END AS has_tagline,  
p.popularity FROM top_10000_movies AS m JOIN movie_popularity AS p ON  
m.movie_id = p.movie_id ORDER BY p.popularity DESC LIMIT 10;
```

- Returns titles of 10 movies with the highest popularity and whether or not they have a tagline, also returns their popularity score.

title	has_tagline	popularity
The Pope's Exorcist	Yes	5089.969
Ant-Man and the Wasp: Quantumania	Yes	4665.438
The Super Mario Bros. Movie	No	3935.55
Ghosted	Yes	2791.532
Shazam! Fury of the Gods	Yes	2702.593
Guardians of the Galaxy Volume 3	Yes	2520.308
Avatar: The Way of Water	Yes	2280.912
Creed III	Yes	1894.044
Scream VI	Yes	1862.283
Dungeons & Dragons: Honor Among Thieves	Yes	1655.052

LYDIA VISUALIZATION

Top 10 Most Popular Movies



LYDIA

QUERY 2

```
SELECT d.name, AVG(t.revenue) AS avg_revenue, MAX(p.popularity) AS max_popularity,
t.title AS movie_with_highest_popularity FROM directors AS d JOIN movie_directors AS
md ON d.director_id = md.director_id JOIN id_link AS il ON md.imdb_id = il.imdb_id JOIN
top_10000_movies AS t ON il.movie_id = t.movie_id JOIN movie_popularity AS p ON
t.movie_id = p.movie_id WHERE p.popularity = (SELECT MAX(popularity) FROM
movie_popularity WHERE movie_id = t.movie_id) GROUP BY d.name ORDER BY
avg_revenue DESC LIMIT 10;
```

- Returns the top 10 directors with the highest average revenue and their most popular movie along with the popularity score.

name	avg_revenue	max_popularity	movie_with_highest_popularity
Colin Trevorrow	1671537444.0	62.284	Jurassic World
Jennifer Lee	1274219009.0	70.892	Frozen
David Yates	1041997255.5	130.573	Harry Potter and the Half-Blood Prince
Lee Unkrich	1003652619.5	68.898	Finding Nemo
Joss Whedon	987696224.333333	121.956	The Avengers
Pierre Coffin	887057440.666667	51.21	Despicable Me
Mike Thurmeier	881965799.5	66.637	Ice Age: Dawn of the Dinosaurs
Ronnie Del Carmen	857611174.0	62.145	Inside Out
Raman Hui	813367380.0	70.102	Shrek the Third
James Cameron	804804817.875	188.103	Avatar

HELEN

QUERY 1:

Compare the average movie rating between high budget and low budget movies

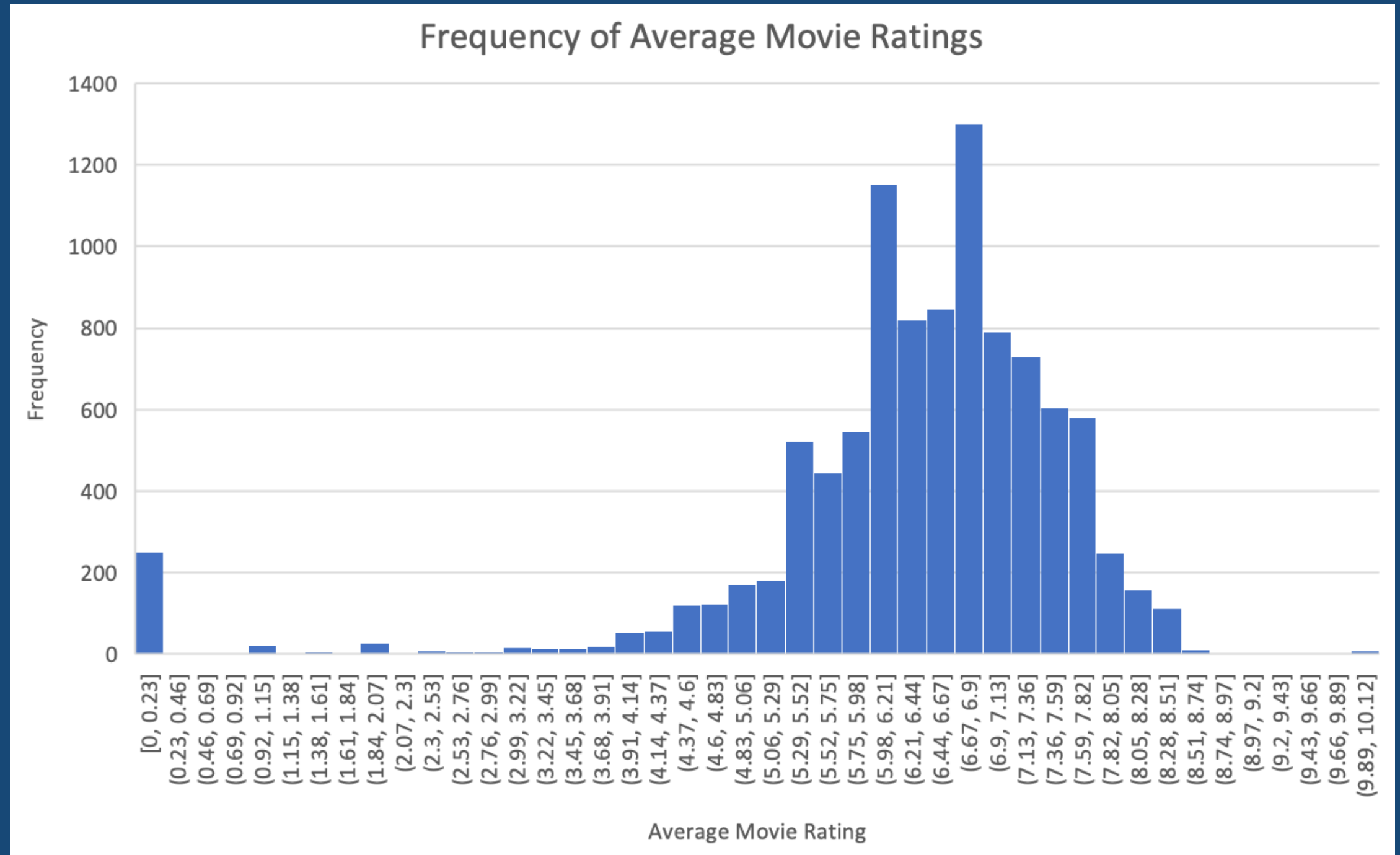
```
WITH high_budget AS (  
    SELECT movie_id  
    FROM top_10000_movies  
    WHERE budget > 100000  
)  
SELECT AVG(vote_average)  
FROM high_budget AS H, movie_popularity AS P  
WHERE H.movie_id = P.movie_id;  
-- Result: Approx. 6.58
```

```
WITH low_budget AS (  
    SELECT movie_id  
    FROM top_10000_movies  
    WHERE budget < 100000  
    AND budget != 0  
)  
SELECT AVG(vote_average)  
FROM low_budget AS L, movie_popularity AS P  
WHERE L.movie_id = P.movie_id;  
-- Result: Approx. 5.51
```

Conclusion: high-budget movies have a 1.07 higher average movie rating than low-budget movies

HELEN

VISUALIZATION



HELEN

QUERY 2:

Find the top 10 production companies with the highest average movie revenue

```
SELECT P.name, AVG(T.revenue)
FROM top_10000_movies AS T, movie_production_companies AS M,
production_companies AS P
WHERE T.movie_id = M.movie_id
AND M.production_id = P.production_id
GROUP BY P.production_id, P.name
ORDER BY AVG(T.revenue) DESC
LIMIT 10;
```

production_company	avg_revenue
Second Mate Productions	1013329906.0
3Foot7	956019788.0
Queen Films	903992901.0
Patalex IV Productions Limited	895921036.0
Beijing Century Media Culture	870322670.0
Beijing Dongfang International Cultural Communications Company	870322670.0
Chao Feng Pictures	870322670.0
Orange Image	870322670.0
Spring Era Film Company	870322670.0
Farewell Productions	791217826.0

VERA

QUERIES

SELECT popularity, vote_average FROM
movie_popularity ORDER BY popularity DESC limit 20;

- Returns popularity score and the associated vote_average of the top 20 movies.

popularity	vote_average
5089.969	7.4
4665.438	6.6
3935.55	7.5
2791.532	7.2
2702.593	6.8
2520.308	8.3
2280.912	7.7
1894.044	7.3
1862.283	7.3
1655.052	7.5
1436.725	7.3
1383.661	7.0
1358.468	5.9
1170.178	7.9
1107.559	7.0
1025.915	8.3
995.199	6.6
801.263	6.4
734.011	7.0
732.647	0.0

SELECT m.title AS movie_title, g.name AS genre_name, mp.popularity AS popularity_score
FROM top_10000_movies AS m JOIN movie_genres AS mg ON m.movie_id = mg.movie_id JOIN
genres AS g ON mg.genre_id = g.genre_id JOIN movie_popularity AS mp ON m.movie_id =
mp.movie_id ORDER BY mp.popularity DESC LIMIT 10;

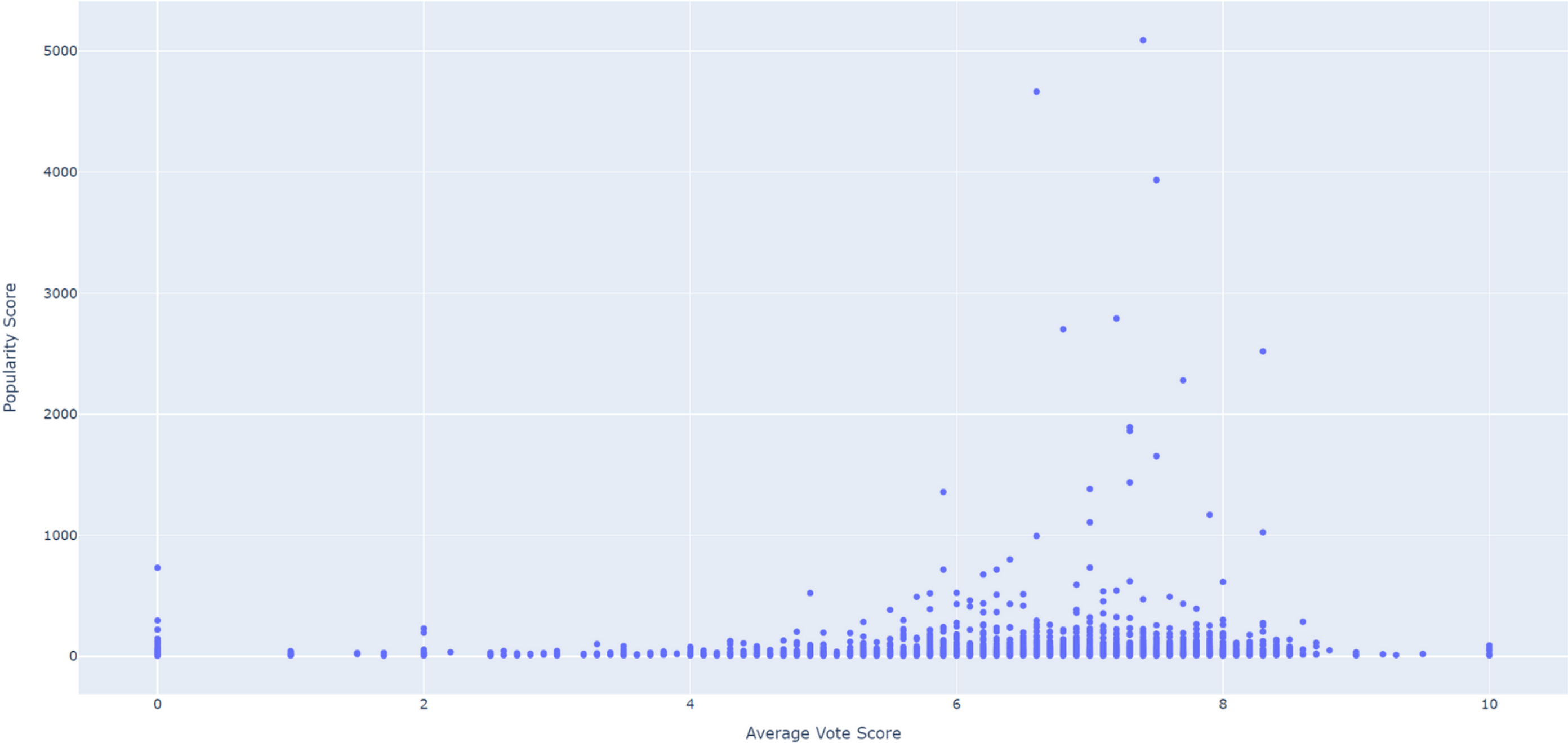
- Returns the titles and genres of movies with the highest popularity scores of the top 10 movies.

movie_title	genre_name	popularity_score
The Pope's Exorcist	Horror	5089.969
The Pope's Exorcist	Mystery	5089.969
The Pope's Exorcist	Thriller	5089.969
Ant-Man and the Wasp: Quantumania	Action	4665.438
Ant-Man and the Wasp: Quantumania	Adventure	4665.438
Ant-Man and the Wasp: Quantumania	Science Fiction	4665.438
The Super Mario Bros. Movie	Adventure	3935.55
The Super Mario Bros. Movie	Animation	3935.55
The Super Mario Bros. Movie	Comedy	3935.55
The Super Mario Bros. Movie	Family	3935.55

VERA

VIZUALIZATION

Average Ratings from TMDB Users and Popularity Scores





```
SELECT title, STRFTIME( '%m',  
release_date) AS release_month FROM  
top_10000_movies ORDER BY revenue  
DESC LIMIT 20;
```

SAM

QUERY 1

Returns the movies with the top 20
highest revenues and the month they
were released in

title	release_month
Avatar	12
Avengers: Endgame	04
Avatar: The Way of Water	12
Titanic	11
Star Wars: The Force Awakens	12
Avengers: Infinity War	04
Spider-Man: No Way Home	12
Jurassic World	06
The Lion King	07
The Avengers	04
Furious 7	04
Top Gun: Maverick	05
Frozen II	11
Avengers: Age of Ultron	04
Black Panther	02
Harry Potter and the Deathly Hallows: Part 2	07
Star Wars: The Last Jedi	12
Jurassic World: Fallen Kingdom	06
Frozen	11
Beauty and the Beast	03

SAM

QUERY 2

Returns 50 of the oldest and newest movies and their genres to make comparisons between which genres are most produced over time. There is a clear shift from a majority Drama or Comedy productions to a majority action productions today!

```
SELECT
top_10000_movies.title,top_10000_movies.release_date,
name AS genre FROM genres JOIN top_10000_movies
JOIN movie_genres ON movie_genres.movie_id =
top_10000_movies.movie_id AND movie_genres.genre_id
= genres.genre_id WHERE top_10000_movies.release_date
IS NOT NULL GROUP BY top_10000_movies.release_date
ORDER BY top_10000_movies.release_date ASC LIMIT 50;
```

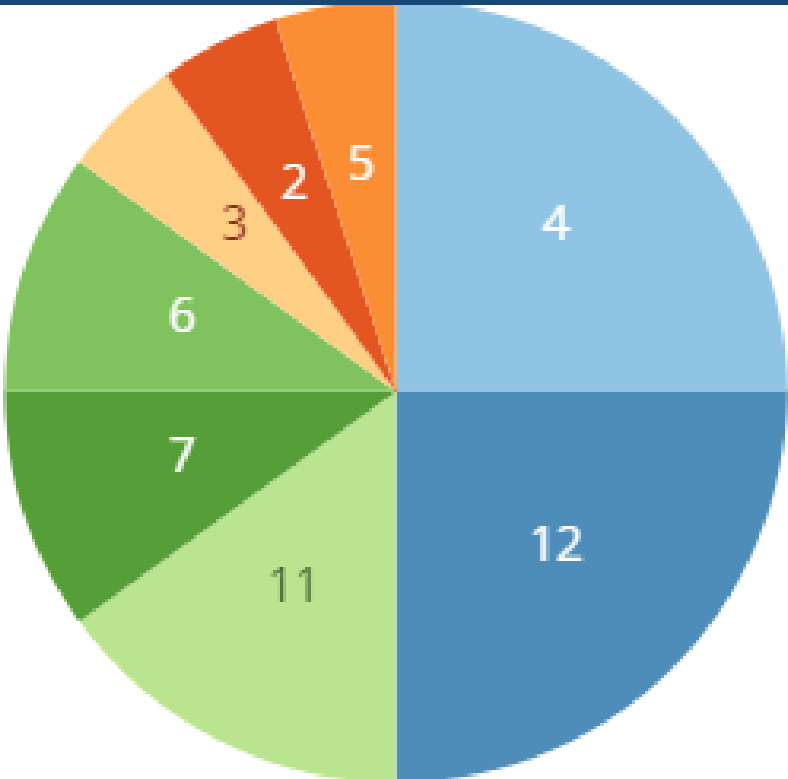
```
SELECT
top_10000_movies.title,top_10000_movies.release_date,
name AS genre FROM genres JOIN top_10000_movies
JOIN movie_genres ON movie_genres.movie_id =
top_10000_movies.movie_id AND movie_genres.genre_id
= genres.genre_id WHERE top_10000_movies.release_date
IS NOT NULL GROUP BY top_10000_movies.release_date
ORDER BY top_10000_movies.release_date DESC LIMIT
50;
```

SAM VISUALIZATION



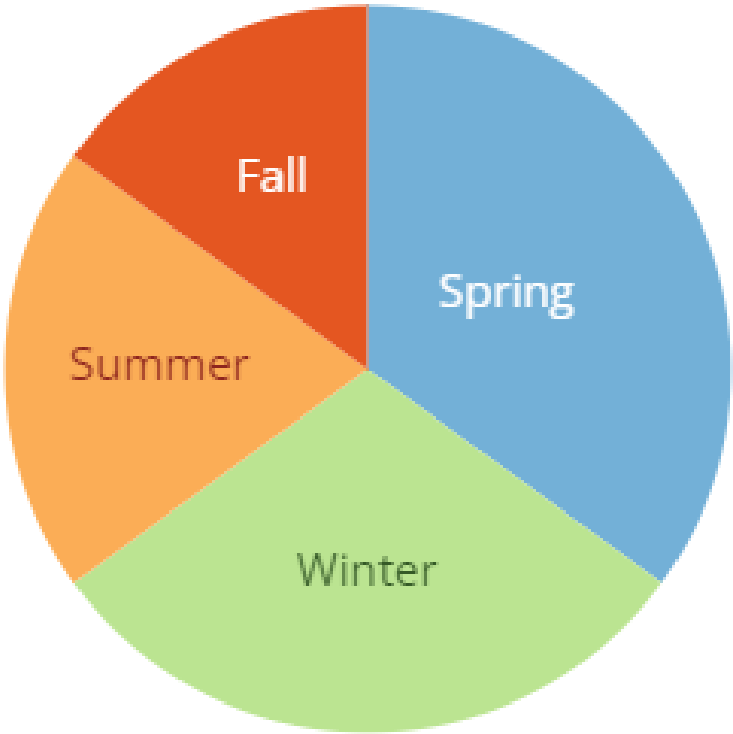
Movie Release By Month

4	5
12	5
11	3
7	2
6	2
3	1
2	1
5	1



Movie Release By Season

Spring	7
Winter	6
Summer	4
Fall	3



JASMINE QUERIES

```
SELECT G1.NAME AS GENRE1, G2.NAME AS GENRE2, MAX(T.BUDGET) AS HIGHEST_BUDGET
FROM MOVIE_GENRES AS MOVIE_GENRE_1
JOIN MOVIE_GENRES AS MOVIE_GENRE_2 ON MOVIE_GENRE_1.MOVIE_ID = MOVIE_GENRE_2.MOVIE_ID AND
MOVIE_GENRE_1.GENRE_ID < MOVIE_GENRE_2.GENRE_ID
JOIN GENRES AS G1 ON MOVIE_GENRE_1.GENRE_ID = G1.GENRE_ID
JOIN GENRES AS G2 ON MOVIE_GENRE_2.GENRE_ID = G2.GENRE_ID
JOIN TOP_10000_MOVIES AS T ON T.MOVIE_ID = MOVIE_GENRE_1.MOVIE_ID
GROUP BY G1.NAME, G2.NAME
ORDER BY HIGHEST_BUDGET DESC;
```

- Returns genre pairngs which have the highest budget

Number of movies per actor.

```
SELECT actors.name AS actor, COUNT(*) AS
movie_per_actor_count
FROM actors
JOIN movie_actors ON actors.actor_id =
movie_actors.actor_id
GROUP BY actors.name
ORDER BY movie_per_actor_count DESC;
```

- Returns number of movies per actor

```
SELECT actors.name AS actor, production_companies.name AS
production_company, COUNT(*) AS number_of_movies
FROM actors
JOIN movie_actors ON actors.actor_id = movie_actors.actor_id
JOIN id_link ON movie_actors.imdb_id = id_link.imdb_id
JOIN movie_production_companies ON id_link.movie_id =
movie_production_companies.movie_id
JOIN production_companies ON movie_production_companies.production_id =
production_companies.production_id
GROUP BY actors.actor_id, production_companies.production_id
ORDER BY number_of_movies DESC
```

- Returns the actors which have done the most movies with the same production comap

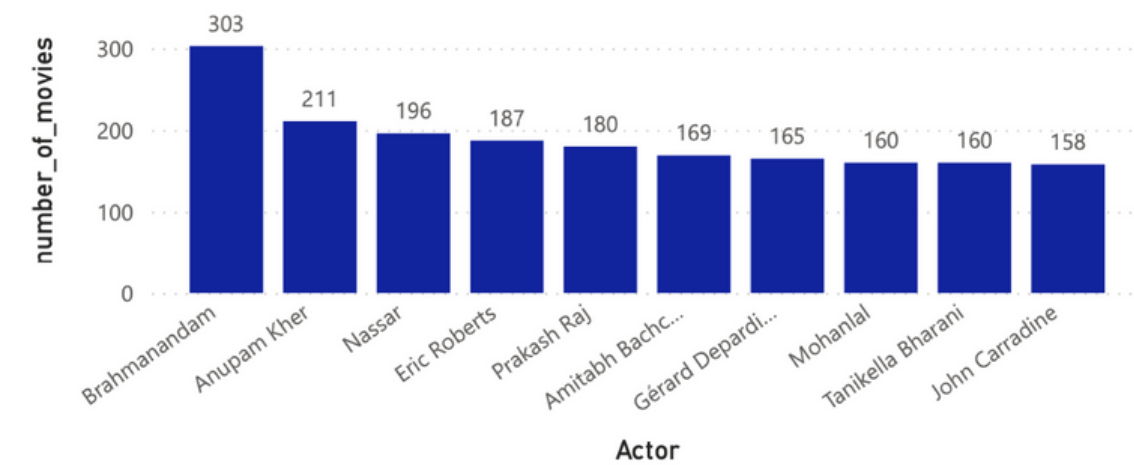
Action	Drama	579330426.0
Action	Thriller	579330426.0
Action	War	579330426.0
Drama	Thriller	579330426.0
Drama	War	579330426.0
Thriller	War	579330426.0
Action	Adventure	460000000.0
Action	Science Fiction	460000000.0
Adventure	Science Fiction	460000000.0
Action	Fantasy	379000000.0

	actor	movie_count
1	Brahmanandam	303
2	Anupam Kher	211
3	Nassar	196
4	Eric Roberts	187
5	Prakash Raj	180
6	Amitabh Bachchan	169
7	Gérard Depardieu	165
8	Tanikella Bharani	160
9	Mohanlal	160
10	John Carradine	158

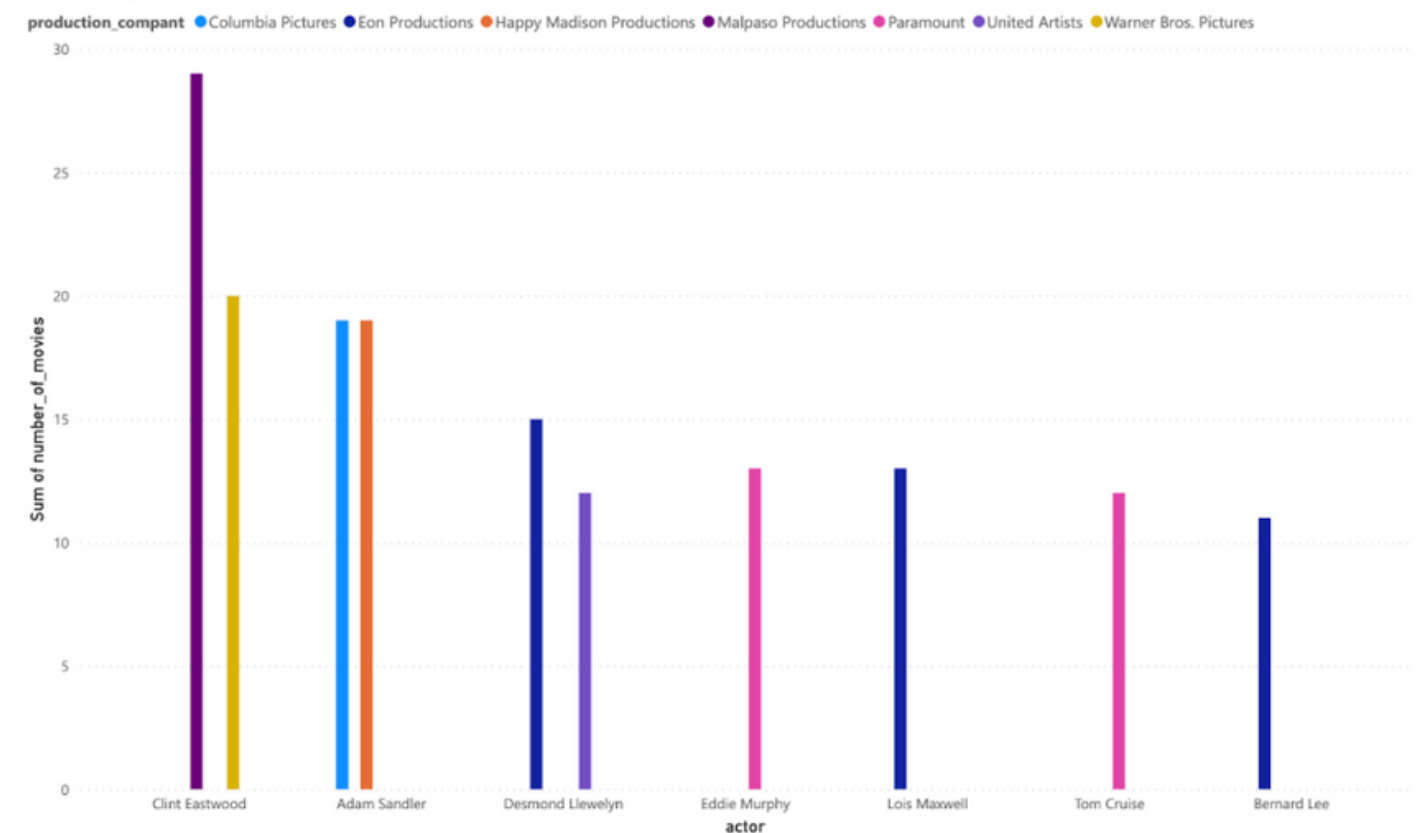
Clint Eastwood	Malpaso Productions	29
Clint Eastwood	Warner Bros. Pictures	20
Adam Sandler	Columbia Pictures	19
Adam Sandler	Happy Madison Productions	19
Desmond Llewelyn	Eon Productions	15
Eddie Murphy	Paramount	13
Lois Maxwell	Eon Productions	13
Desmond Llewelyn	United Artists	12
Tom Cruise	Paramount	12
Bernard Lee	Eon Productions	11

JASMINE VISUALIZATION

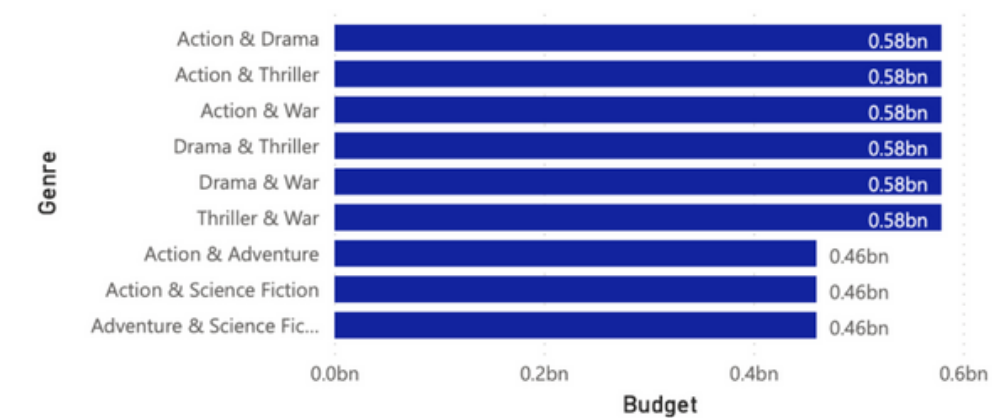
Number of Movies By Actor



Relationship Between Actors and Production Companies: Movie Count



Highest Budget Movies by Paired Genres



**THANK YOU FOR
WATCHING!**