Analyzing Film Data to Guide RSVP Movies

Presenter :
Gboyega Adenaike

Supervisor :
Prof. Aritri Debnath

About Me:

Name: Gboyega Adenaike
Contact: segedenny56@gmail.com
LinkedIn profile: linkedin.com/in/gboyega-adenaike-860b0644
Github: https://github.com/gboyeg

SKILLS

- Data Analysis Tools: Excel, SQL, Python (Pandas, Numpy)
- Data Visualization: Excel, Power BI, Matplotlib, seaborn
- Predictive Modeling
- Machine Learning: scikit learn, linear regression, clustering, deep learning algorithm
- Report writing
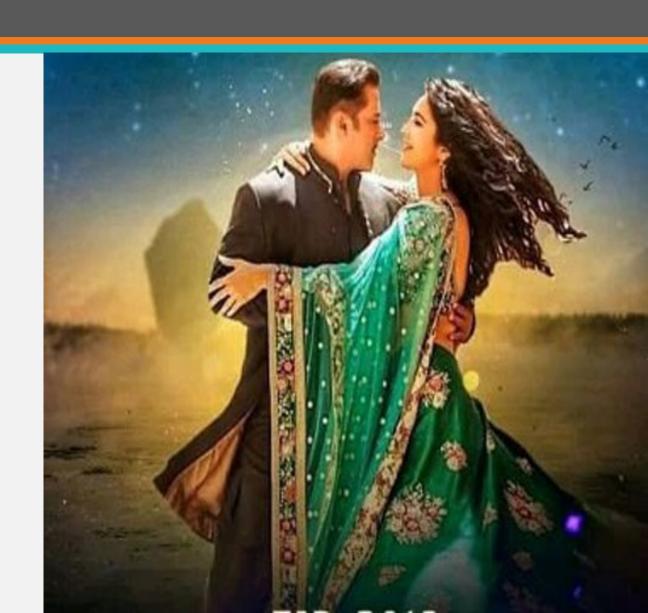- Critical thinking, communication, problem – solving, teamwork

AREA OF INTEREST:

- Data Analyst
- Business Analyst
- Data Science

Overview

- Introduction
- Project Objectives
- Data Description
- Data Exploration and Preparation
- Insights
- Recommendations
- Conclusion
- Appreciation

Introduction

RSVP Movies, a leading Indian film production house celebrated for its blockbuster films, is launching an exciting new project in 2022 aimed at a worldwide audience. To maximize the potential of this initiative, they are leveraging data analytics. In my role as a data analyst and SQL specialist, I have been assigned the responsibility of examining movie data from the last three years to deliver critical insights that will assist RSVP Movies in making informed strategic choices.

To achieve this set goal, our analytical journey is divided into four segments, each targeting specific aspects of the dataset to derive meaningful insights. These insights will inform RSVP Movies' decisions and ensure a data-driven approach to their new global project.
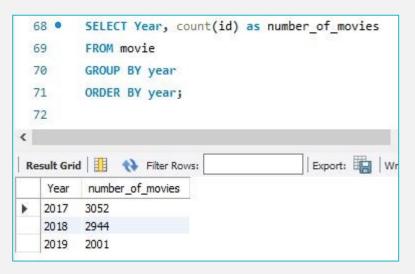
Data description

- The dataset comprises six distinct tables: Genre, Movies, Director_Mapping, Role_Mapping, Names, and Ratings. The breakdown of columns and rows for each table is as follows:

- - Genre: 2 columns and 14,662 rows

- - Movies: 9 columns and 7,997 rows

- - Names: 5 columns and 25,735 rows

- - Ratings: 4 columns and 7,997 rows

- - Director_Mapping: 2 columns and 3,867 rows

- - Role_Mapping: 3 columns and 15,615 rows

- Additionally, the following columns in the Movies table contain null values:

- - Country: 20, - Worldwide_gross_income: 3,724 - Languages: 194 , - Production_Company: 528.

Table	Column
movie	id
movie	title
movie	year
movie	date_published
movie	duration
movie	country
movie	worlwide_gross_income
movie	languages
movie	production_company
genre	movie_id
genre	genre
director_mapping	movie_id
director_mapping	name_id
role_mapping	movie_id
role_mapping	name_id
role_mapping	category
names	id
names	name
names	height
names	date_of_birth
names	known_for_movies
ratings	movie_id
ratings	avg_rating
ratings	total_votes
ratings	median_rating

Insights

Segment 1 - Understanding the Data:

▪We begin by comprehending the dataset's structure and content. Familiarity with the data is crucial before embarking on any analysis. This segment helps us get acquainted with what each column represents.

▪The number of movies over the years have significantly dropped, having 2001 movies in the year 2019 with the highest to be in 2017 as 3052.

▪The highest number of movies published was in March 824, having December to be the lowest month 438.

▪Drama takes the spotlight with the highest number of movies produced followed by Comedy, and Thriller are the most frequently produced genres, with Action and horror having low audience engagement.
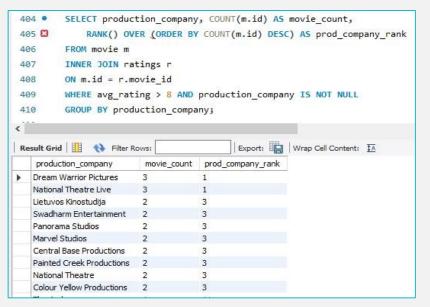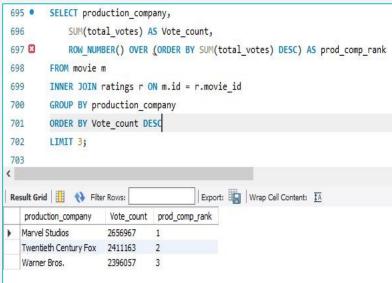


```
68  SELECT Year, count(id) as number_of_movies
69  FROM movie
70  GROUP BY year
71  ORDER BY year;
72
```

Year	number_of_movies
2017	3052
2018	2944
2019	2001

```
94  SELECT Month(date_published) as month_num, count(id) as num_of_movies
95  FROM movie
96  GROUP BY month_num
97  ORDER BY num_of_movies;
98
```

month_num	num_of_movies
12	438
7	493
6	580
11	625
5	625
2	640
8	678
4	680
10	801
1	804
9	809
3	824

```
154  SELECT GENRE,
155       COUNT(ID) AS TOTAL_MOVIES
156  FROM MOVIE M
157  INNER JOIN GENRE G ON M.ID = G.MOVIE_ID
158  GROUP BY GENRE
159  ORDER BY TOTAL_MOVIES DESC
160  LIMIT 5;
161
162
```

GENRE	TOTAL_MOVIES
Drama	4285
Comedy	2412
Thriller	1484
Action	1289
Horror	1208

Insights

Segment 2 - Box Office Performance Analysis :

The top production company is Dream Warrior Pictures and National Theatre Live; this will help RSVP Movies in finding potential partners for their next project.

Marvel Studios stands tall as one of the top global production houses, making it a potential partner for RSVP Movies and in term of revenue generation.

Production houses with fewer but high-quality releases (like Star Cinema) outperform those with high volumes of average-performing films while Thriller genre generated the highest gross income of about $1,236,005,118, followed by comedy for $1,034,799,409.


```
404 •    SELECT production_company, COUNT(m.id) AS movie_count,
405 ☒       RANK() OVER (ORDER BY COUNT(m.id) DESC) AS prod_company_rank
406    FROM movie m
407    INNER JOIN ratings r
408    ON m.id = r.movie_id
409    WHERE avg_rating > 8 AND production_company IS NOT NULL
410    GROUP BY production_company;
```

production_company	movie_count	prod_company_rank
Dream Warrior Pictures	3	1
National Theatre Live	3	1
Lietuvos Kinostudija	2	3
Swadharm Entertainment	2	3
Panorama Studios	2	3
Marvel Studios	2	3
Central Base Productions	2	3
Painted Creek Productions	2	3
National Theatre	2	3
Colour Yellow Productions	2	3

```
695 •    SELECT production_company,
696       SUM(total_votes) AS Vote_count,
697 ☒       ROW_NUMBER() OVER (ORDER BY SUM(total_votes) DESC) AS prod_comp_rank
698    FROM movie m
699    INNER JOIN ratings r ON m.id = r.movie_id
700    GROUP BY production_company
701    ORDER BY Vote_count DESC
702    LIMIT 3;
703
```

production_company	Vote_count	prod_comp_rank
Marvel Studios	2656967	1
Twentieth Century Fox	2411163	2
Warner Bros.	2396057	3

genre	year	movie_name	worldwide_gross_income_$	movie_rank
Thriller	2017	The Fate of the Furious	1236005118	1
Comedy	2017	Despicable Me 3	1034799409	2
Comedy	2017	Jumanji: Welcome to the Jungle	962102237	3
Drama	2017	Zhan lang II	870325439	4
Thriller	2017	Zhan lang II	870325439	5
Drama	2018	Bohemian Rhapsody	903655259	1
Thriller	2018	Venom	856085151	2
Thriller	2018	Mission: Impossible - Fallout	791115104	3
Comedy	2018	Deadpool 2	785046920	4
Comedy	2018	Ant-Man and the Wasp	622674139	5
Drama	2019	Avengers: Endgame	2797800564	1
Drama	2019	The Lion King	1655156910	2
Comedy	2019	Toy Story 4	1073168585	3
Drama	2019	Joker	995064593	4
Thriller	2019	Joker	995064593	5

Insights
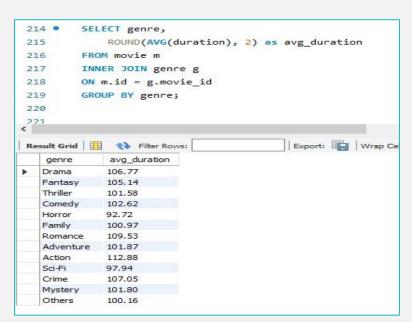
Segment 3 - Genre :

The top genre remains to be drama with 24 movies and followed by Comedy with 9 movies. This information can help filmmakers identify the genres that captivate audiences in the USA.

It turns out that the 'Thriller' genre ranks among the top 3 genres with the most movies. Filmmakers seeking an exciting genre should keep their eyes on 'Thriller.'

Movies of the 'drama' genre, which were produced in the highest numbers in 2019, have an average duration of approximately 106.77 minutes. Filmmakers can use this insight to plan their next project.
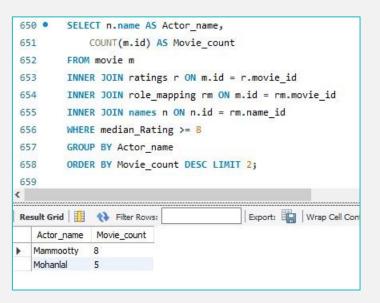
Insights

Segment 4 – Audience Ratings Assessment:

Actors like Mammootty, Mohanlal, and Vijay Sethupathi consistently appear in well-rated movies. The results may help identify acclaimed actors for upcoming regional projects.

Movies with a median rating of 7 seem to be the most numerous. This information can be invaluable for movie production planning.

The analysis reveals potential directors for RSVP Movies' next projects, based on the highest-rated genres. Directors like James Mangold, and Joe Russo have the highest number of successful movies.

```sql
650   SELECT n.name AS Actor_name,
651       COUNT(m.id) AS Movie_count
652   FROM movie m
653   INNER JOIN ratings r ON m.id = r.movie_id
654   INNER JOIN role_mapping rm ON m.id = rm.movie_id
655   INNER JOIN names n ON n.id = rm.name_id
656   WHERE median_Rating >= 8
657   GROUP BY Actor_name
658   ORDER BY Movie_count DESC LIMIT 2;
659
```

Result Grid

Actor_name	Movie_count
Mammootty	8
Mohanlal	5

```sql
363   SELECT
364       median_rating,
365       COUNT(m.id) AS movie_count
366   FROM ratings r
367   INNER JOIN movie m ON r.movie_id = m.id
368   GROUP BY median_rating
369   ORDER BY movie_count DESC;
```

Result Grid

median_rating	movie_count
7	2257
6	1975
8	1030
5	985
4	479
9	429
10	346
3	283
2	119
1	94

```sql
609   WITH Top_Three_Genre AS (
610       SELECT genre, COUNT(m.id) AS Movie_count
611       FROM movie m INNER JOIN genre g ON m.id = g.movie_id
612       INNER JOIN ratings r ON r.movie_id = m.id
613       WHERE avg_rating > 8
614       GROUP BY genre
615       ORDER BY Movie_count DESC LIMIT 3)
616   SELECT n.name AS director_name, COUNT(m.id) AS Movie_count
617   FROM movie m INNER JOIN director_mapping d ON m.id = d.movie_id
618   INNER JOIN names n ON n.id = d.name_id INNER JOIN genre g ON g.movie_id = m.id
619   INNER JOIN ratings r ON m.id = r.movie_id
620   WHERE g.genre IN (SELECT genre FROM Top_Three_Genre)
621   AND avg_rating > 8
622   GROUP BY director_name
623   ORDER BY Movie_count DESC LIMIT 3;
624
```

Result Grid

director_name	Movie_count
James Mangold	4
Joe Russo	3
Anthony Russo	3

Recommendations

1. RSVP Movies should analyze the strategies of Marvel Studios and Warner Bros. to gain insights into their success. Expand internationally by collaborating with leading production companies like Twentieth Century Fox.

2. Focus on investing in successful genres such as Action, Thriller, and Biopics to enhance audience engagement.

3. Broaden your reach by producing multilingual content to attract global viewers. Collaborate with esteemed actors like Mammootty, Vijay Sethupathi, and Mohanlal to build audience confidence.

4. Partner with acclaimed directors such as James Mangold and Joe Russo, known for their successful track records.

5. Increase audience interaction through pre-release marketing strategies and social media voting initiatives. Consider investing in franchise-based projects and explore partnerships with OTT platforms to maximize return on investment.

6. RSVP Movies should analyze the strategies of Marvel Studios and Warner Bros. to gain insights into their success. Expand internationally by collaborating with leading production companies like Twentieth Century Fox.

7. Utilize AI-driven data analysis to inform future film selection and investment strategies.

Conclusion

This analytical exploration using SQL provides RSVP Movies with essential insights.
By leveraging data-driven recommendations, they can effectively strategize their global audience-oriented film project. With a solid understanding of their data, RSVP Movies is well-positioned to craft cinematic experiences that connect with viewers around the globe.

By applying these insights and recommendations, RSVP Movies can mitigate financial risks, enhance production decisions, and position itself as a global industry leader in the years ahead.

Thank you !

Contact: segedenny56@gmail.com
LinkedIn profile: linkedin.com/in/gboyega-adenaike-860b0644