

Analysis of IMDb Database

ASHIN R
DA&DS
61224CBR016

Introduction to the Dataset

-
- Dataset contains information on movies, directors, genres, actors, ratings, and production details.
 - Tables include movie, ratings, genre, names, director_mapping, and role_mapping.
 - Objective: Analyze trends, data completeness, and key insights.

Counting Records in Tables

director_mapping	- 3867
genre	- 14662
movie	- 7997
role_mapping	- 15615
names	- 25735
ratings	- 7997

```
select 'director_mapping', count(movie_id) from director_mapping
union
select 'genre', count(genre) from genre
union
select 'movie' as table_name , count(title) as total_records from movie
union
select 'role_mapping', count(movie_id) from role_mapping
union
select 'names', count(name) from names
union
select 'ratings', count(movie_id) from ratings;
```

Identifying Null Values

MOVIE TABLE

- worldwide_gross_income - 3724
- languages - 194
- production_company - 528

NAMES TABLE

- height column - 17335
- birth column - 13431
- height column - 15226

`select`

```
count(*) - count(title) as title_nulls,
count(*) - count(year) as year_nulls,
count(*) - count(date_published) as date_published_nulls,
count(*) - count(duration) as duration_nulls,
count(*) - count(worldwide_gross_income) as worldwide_gross_income_nulls,
count(*) - count(languages) as languages_nulls,
count(*) - count(production_company) as production_company_nulls
from movie;
```

`select`

```
count(*) - count(id) as id_nulls,
count(*) - count(name) as name_nulls,
count(*) - count(height) as height_nulls,
count(*) - count(date_of_birth) as birth_nulls,
count(*) - count(known_for_movies) as movie_nulls
from names;
```

```

select
    (select count(distinct genre) from genre) as unique_genres,
    (select count(movie_id)
     from(select movie_id
           from genre
          group by movie_id
         having count(genre) = 1)as single_genre_movies) as exclusive_genre_movies;
|



select genre,count(movie_id) as count_movies
from genre
group by genre
order by count_movies desc
limit 1;

#7. Calculate the average movie duration for each genre.
select g.genre,avg(m.duration) as movie_duration
from movie m
join genre g on m.id=g.movie_id
group by g.genre
order by movie_duration desc;

from movie m
join ratings r on m.id=r.movie_id
join genre g on r.movie_id=g.movie_id
group by m.id
order by total_vote desc
limit 5;

#23. Identify the movie with the longest duration, along with its genre and prc
select m.title, m.duration, g.genre, m.production_company as production_company
from movie m
join genre g on m.id = g.movie_id
where m.duration = (select max(duration) from movie);

```

Genre and Movie Analysis

- 13 unique genres in the dataset and 3289 movies belong exclusively to one genre.
- Drama has the highest total number of movies produced (4825)
- Action genre has most average movie duration
- Sleeping Beauties is the oldest movie
- Avengers: Infinity War get highest vote
- La flor longest duration movie

Rating Statistics

- Kirket movie has top avg. rating
- 346 movie get median rating has 10
- April 1, 2018 - April 1, 2019 - 361

movies received median rating of 8

```
#10.Which are the top 10 movies based on their average rating?  
select m.title,r.avg_rating  
from movie m  
join ratings r on m.id=r.movie_id  
order by avg_rating desc  
limit 10;  
  
#11.Summarize the ratings table by grouping movies based on their median rating.  
select  
    median_rating,  
    COUNT(movie_id) as movie_count  
from ratings  
group by median_rating  
order by median_rating desc;  
  
select count(title) as movie_count  
from movie m  
join ratings r on m.id = r.movie_id  
where m.date_published between '2018-04-01' and '2019-04-01' and r.median_rating = 8;
```

Production Analysis

- Marvel Studios production company get highest votes for their movies(2656967)
- Pampero Cine production company produced highest duration movies
- English language movies is most produced
- 295 movies were produced in INDIA in 2019.
- 592 movies were produced in USA in 2019.

```
select languages, count(languages) as movie_count
from movie
group by languages
order by movie_count desc
limit 1;
```

```
select m.production_company, sum(r.total_votes) as total_votes
from movie m
join ratings r on m.id = r.movie_id
group by m.production_company
order by total_votes desc
limit 3;
```

```
select m.title, m.duration, g.genre, m.production_company as production_company
from movie m
join genre g on m.id = g.movie_id
where m.duration = (select max(duration) from movie);
```

Actor,Actress & Director Analysis

- 9 directors have worked on more than three movies
- average height of Actor-162.1818 , Actress -162.4715
- Savas Barutçu&Cihan Köse top actors whose movies have get median rating above 8
- Dean McKendrick is director oldest movies (Sleeping Beauties)

```

select rm.name_id,n.name,r.median_rating
from role_mapping rm
join names n on rm.name_id=n.id
join ratings r on rm.movie_id=r.movie_id
where rm.category like 'actor' and r.median_rating >= 8
order by median_rating desc
limit 2;

select count(name_id) as director_count
from (
    select name_id
    from director_mapping
    group by name_id
    having count(movie_id) > 3
) as director_count;

#20. Calculate the average height of actors and actresses
select rm.category, avg(n.height) as avg_height
from role_mapping rm
join names n on rm.name_id = n.id
where rm.category in ('actor', 'actress')
group by rm.category;

select m.title,m.country,n.name as director,m.date_published
from movie m
join director_mapping d on m.id=d.movie_id
join names n on d.name_id=n.id
order by m.date_published asc
limit 10;

```

Key Findings

- Highest null values in movie - worlwide_gross_income (3724)
- Highest null values in names table - Hight column (17335)
- Longest duration movie / produced - La flor(808min.)/Pampero Cine
- Oldest movie / Directed - Sleeping Beauties / Dean McKendrick
- Highest vote getting movie / produced - Avengers: Infinity War / Marvel Studios
- Top avg. rating movie - Kirket
- Average height of Actor & Actress - 162

Suggestions & Recommendations

- Handling NULL values effectively
- Optimizing database performance
- Investigate reasons for fluctuations in movie releases
- Explore correlations between ratings, genres, and success factors
- Use insights for market research and content production strategies

