

MOVIES DATASET PROJECT

EXPLORATORY DATA ANALYSIS

KHOLOUD OTHMAN

MOVIES DATASET

The dataset provides a concise view of movie characteristics, with information on their identification, genre, release details, financial performance, popularity, user ratings, and runtime. These attributes can be used for analysis, comparison, and understanding trends in the movie industry.

Identification:

Id: A unique identifier for each movie.

Title , original language

Categorization:

Genres

Release:

Release date

Release year : extracted from release date

Financial:

Budget , revenue

Profit: calculated as revenue minus budget

rate:

Popularity

Vote average , vote count

Duration:

Runtime

DATA CLEANING

- Delete unused columns
- Check missing values
- Drop missing values
- Check duplicate data
- Drop duplicate data
- Check data types
- Change data types necessary

Q1 .

What is the most frequent genre of movies?

Top 10 genres

genres	Num of movies
Drama	7900
Comedy	6236
Action	3221
Documentary	2698
Horror	2611
Thriller	1645
Animation	1350
Romance	1306
Crime	1229
Adventure	892

Q2 .

What is genre has
the highest total
popularity score?

Top 10 genres

genres	total popularity
Drama	58115.94
Comedy	44302.24
Action	40560.73
Horror	35413.30
Animation	23571.17
Thriller	15442.64
Romance	12320.29
Documentary	11587.22
Adventure	10828.54
Crime	10741.99

Q3 .

What is genre has
the highest total
vote count?

Top 10 genres

genres	total vote count
Drama	1340443
Action	1252437
Comedy	1117562
Adventure	584689
Horror	484936
Animation	360231
Thriller	347602
Crime	345391
Science Fiction	311206
Fantasy	283823

Q4 .

What is the total revenue generated by movies in each genre?

Top 10 genres

genres	total revenue
Action	58491327615
Comedy	38221060960
Drama	32608163744
Adventure	27719903834
Animation	18320328249
Fantasy	13425063611
Horror	10873522183
Science Fiction	10062972452
Crime	8786216828
Thriller	8432519690

Q5.

What is the total profit each genre (revenue - budget)?

Top 10 genres

genre	profit
Action	35330886481
Comedy	22744071604
Drama	17867746053
Adventure	17349208397
Animation	12651580488
Fantasy	9250418851
Horror	7179675343
Science Fiction	5722838493
Crime	4649435523
Thriller	4340289341

Q6 .

**How Many movies
were released in
each year ?**

Top 10 years

release year	Num of movies
2019	1617
2018	1512
2021	1508
2017	1469
2020	1396
2016	1372
2015	1168
2022	1163
2014	1107
2013	1028

Q7 .

What is the average budget for movies released in each year?

Top 10 years

Release year	average budget
2023	7.495338e+06
1997	6.339903e+06
2001	6.079821e+06
1998	6.068609e+06
2004	5.555949e+06
2002	5.322581e+06
2000	5.240537e+06
2005	5.194781e+06
1996	5.183032e+06
2010	5.160979e+06

Q8 .

What is the total profit each year (revenue - budget)?

release year	Total profit
2017	7897879417
2019	7653602405
2016	7194549360
2015	7038349215
2018	6802055855
2011	6438243558
2022	6017638346
2013	5969613206
2014	5924710670
2010	5649541273

Q9 .

What is the average runtime of movies grouped by genre?

Top 10 genres

genres	average runtime
History	114.61
Music	108.20
War	106.90
Action	105.11
Drama	102.76
Crime	100.71
Romance	97.76
Adventure	97.59
Thriller	96.81
Mystery	95.05

Q10 .

What is the title of the movie with the longest runtime?

The Reality of Me

Q11 .

What is the title of the movie with the smallest runtime?

Quick Brown Fox and a Lazy Big Dog

Q12 .

What is the top 10 highest-grossing movies (revenue - budget)?

Top 10 movies

title	profit
Jurassic World	1521537444
The Lion King	1403075401
Furious 7	1325341399
Harry Potter and the Deathly Hallows: Part 2	1216511219
The Super Mario Bros. Movie	1178766975
The Lord of the Rings: The Return of the King	1024888979
Joker	1019458282
Aquaman	988528393
The Fate of the Furious	986005118
Captain Marvel	979416446

Q13 .

What is top 10 movies with the highest budget?

Top 10 movies

title	budget
DeAD	500000000
Mission: Impossible - Dead Reckoning Part One	290000000
The Lion King	260000000
Spider-Man 3	258000000
Harry Potter and the Deathly Hallows: Part 1	250000000
John Carter	250000000
Thor: Love and Thunder	250000000
Harry Potter and the Half-Blood Prince	250000000
Guardians of the Galaxy Vol. 3	250000000
The Hobbit: The Battle of the Five Armies	250000000

Q14.

What is the movies revenue with a vote average greater than 8 and a budget over \$100 million?

title	revenue
Guardians of the Galaxy Vol. 3	821801000
Harry Potter and the Deathly Hallows: Part 2	1341511219
Interstellar	701729206
Terminator 2: Judgment Day	520000000
Gladiator	465361176

Q15 .

What is languages
has the highest
popularity score for
movies?

Top 5 languages

original language	popularity
en	211303.717
ja	16309.972
es	14104.976
fr	12086.371
ko	7847.415

Q16 .

What is the average budget for movies in each original language based on budget?

Top 5 languages

Original language	Average budget
se	4.450000e+06
en	4.155203e+06
zh	2.947847e+06
kk	2.166667e+06
ca	2.106667e+06