# 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 | 7466 |
| Comedy | 6076 |
| Documentary | 3121 |
| Action | 3081 |
| Horror | 2466 |
| Thriller | 1535 |
| Animation | 1517 |
| Romance | 1264 |
| Crime | 1161 |
| Adventure | 861 |

# Q2.

**What is genre has the highest total popularity score?**

**Top 10 genres**

| genres | total popularity |
|---|---|
| Drama | 56339.73 |
| Comedy | 43573.54 |
| Action | 39728.54 |
| Horror | 34622.66 |
| Animation | 24236.00 |
| Thriller | 14845.72 |
| Documentary | 13346.56 |
| Romance | 12040.09 |
| Adventure | 10450.94 |
| Crime | 10378.99 |

# Q3.
## What is genre has the highest total vote count?

## Top 10 genres

| genres | total vote count |
|---|---|
| Drama | 1308126 |
| Action | 1240029 |
| Comedy | 1098021 |
| Adventure | 564996 |
| Horror | 465478 |
| Animation | 357825 |
| Crime | 336705 |
| Thriller | 330134 |
| Science Fiction | 307331 |
| Fantasy | 274753 |

# Q4.

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

**Top 10 genres**

| genres | total revenue |
|---|---|
| Action | 58038546233 |
| Comedy | 37551646964 |
| Drama | 32159914974 |
| Adventure | 2589187046 |
| Animation | 18047307067 |
| Fantasy | 12966617312 |
| Horror | 10486890643 |
| Science Fiction | 10009852894 |
| Crime | 8411574600 |
| Thriller | 8350670046 |

# Q5.
**What is the total profit each genre (revenue - budget)?**

**Top 10 genres**

| genre | profit |
|---|---|
| **Action** | 35382262957 |
| **Comedy** | 22398719637 |
| **Drama** | 17774665256 |
| **Adventure** | 16053172609 |
| **Animation** | 12478343903 |
| **Fantasy** | 9054162552 |
| **Horror** | 6903717428 |
| **Science Fiction** | 5708318935 |
| **Crime** | 4382649285 |
| **Thriller** | 4338495270 |

# Q6 .
## How Many movies were released in each year ?

## Top 10 years

| release year | Num of movies |
|:---:|:---:|
| 2019 | 1614 |
| 2018 | 1517 |
| 2021 | 1504 |
| 2017 | 1451 |
| 2020 | 1406 |
| 2016 | 1366 |
| 2022 | 1213 |
| 2015 | 1161 |
| 2014 | 1085 |
| 2013 | 1022 |

# Q7 .
## What is the average budget for movies released in each year?

## Top 10 years

| Release year | average budget |
|---|---|
| 2023 | 6.997594e+06 |
| 1997 | 6.723333e+06 |
| 1998 | 6.138050e+06 |
| 2001 | 5.976343e+06 |
| 2000 | 5.480705e+06 |
| 2002 | 5.368375e+06 |
| 1996 | 5.287380e+06 |
| 2004 | 5.240576e+06 |
| 2007 | 5.201597e+06 |
| 2005 | 5.187259e+06 |

# Q8 .
## What is the total profit each year (revenue - budget)?

| release year | Total profit |
|:---:|:---|
| 2017 | 7607809389 |
| 2018 | 6907990896 |
| 2015 | 6547823980 |
| 2011 | 6421240957 |
| 2016 | 6416248580 |
| 2019 | 6118030801 |
| 2022 | 6021627345 |
| 2013 | 5991482874 |
| 2014 | 5953195217 |
| 2010 | 5619198887 |

# Q9 .
## What is the average runtime of movies grouped by genre?

## Top 10 genres

| genres | average runtime |
|---|---|
| History | 113.68 |
| War | 108.42 |
| Music | 106.34 |
| Action | 104.87 |
| Drama | 102.75 |
| Documentary | 102.07 |
| Crime | 100.84 |
| Romance | 97.56 |
| Adventure | 97.19 |
| Thriller | 96.76 |

# Q10 .

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

**Logistics**

# Q11 .

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

**The strange world of Gumball**

# Q12 .
**What is the top 10 highest-grossing movies (revenue - budget)?**

Top 10 movies

| title | profit |
|-------|--------|
| Jurassic World | 1521537444 |
| 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 |
| Zootopia | 873784195 |

# Q13.
## What is top 10 movies with the highest budget?

## Top 10 movies

| title | budget |
|-------|--------|
| DeAD | 500000000 |
| Mission: Impossible - Dead Reckoning Part One | 290000000 |
| Spider-Man 3 | 258000000 |
| The Hobbit: The Battle of the Five Armies | 250000000 |
| The Hobbit: The Desolation of Smaug | 250000000 |
| Harry Potter and the Deathly Hallows: Part 1 | 250000000 |
| Harry Potter and the Half-Blood Prince | 250000000 |
| John Carter | 250000000 |
| The Fate of the Furious | 250000000 |
| Guardians of the Galaxy Vol. 3 | 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 | 208354.806 |
| ja | 16438.958 |
| es | 14248.018 |
| fr | 11868.650 |
| ko | 7730.656 |

# Q16.

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

**Top 5 languages**

| Original language | Average budget |
|---|---|
| se | 4.450000e+06 |
| en | 4.095738e+06 |
| zh | 2.727207e+06 |
| kk | 2.166667e+06 |
| ca | 2.106667e+06 |