



IMDB MOVIE ANALYSIS PROJECT REPORT...

EXCEL SHEET LINK:

[https://docs.google.com/spreadsheets/d/13JV2TJXDNIxFXUEWILWNAQ4ZR9K_7AGL/edit?
usp=sharing&ouid=104761672347413617130&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/13JV2TJXDNIxFXUEWILWNAQ4ZR9K_7AGL/edit?usp=sharing&ouid=104761672347413617130&rtpof=true&sd=true)

PRESENTED BY:

ADNAN ASHRAF

VIDEO PRESENTATION LINK:

https://drive.google.com/file/d/1DP59AXGzAcQR6u5t_jEy2CFDCpkxiwhY/view?usp=sharing

Project Description

THE IMDB MOVIE ANALYSIS PROJECT AIMS TO IDENTIFY KEY FACTORS INFLUENCING MOVIE SUCCESS ON IMDB, FOCUSING ON HIGH RATINGS.

THROUGH DATA CLEANING AND ANALYSIS IN MICROSOFT EXCEL, THE PROJECT EXPLORES GENRES, DURATIONS, LANGUAGES, DIRECTORS, AND BUDGETS' IMPACT ON IMDB SCORES. THE FIVE 'WHYS' APPROACH IS EMPLOYED TO DELVE DEEPER INTO IDENTIFIED PATTERNS. THE REPORT PROVIDES ACTIONABLE INSIGHTS FOR STAKEHOLDERS IN THE FILM INDUSTRY, ENHANCING UNDERSTANDING OF INFLUENTIAL ELEMENTS IN MOVIE SUCCESS.

THE TECH-STACK INVOLVES ADVANCED EXCEL FUNCTIONS, STATISTICAL MEASURES, AND DATA VISUALIZATION. THE PROJECT CULMINATES IN A DETAILED REPORT AND A LINKED EXCEL SHEET FOR FURTHER REFERENCE.



Agenda

DATA CLEANING:

01

- HANDLE MISSING VALUES, REMOVE DUPLICATES, AND CONVERT DATA TYPES.
- UTILIZE EXCEL FUNCTIONS FOR DATA PREPROCESSING.

DATA ANALYSIS:

02

- EXPLORE RELATIONSHIPS BETWEEN VARIABLES (RATINGS, GENRE, DIRECTOR, BUDGET, ETC.).
- USE STATISTICAL MEASURES AND VISUALIZATIONS FOR ANALYSIS.

FIVE 'WHYS' APPROACH:

03

- DIG DEEPER INTO IDENTIFIED PATTERNS TO UNCOVER ROOT CAUSES.

REPORT AND DATA STORY:

04

- CREATE A NARRATIVE WITH VISUALIZATIONS TO TELL THE STORY OF THE DATA.

Tech-Stack Used



DATA ANALYTICS TASKS 1

A. MOVIE GENRE ANALYSIS: ANALYZE THE DISTRIBUTION OF MOVIE GENRES AND THEIR IMPACT ON THE IMDB SCORE.

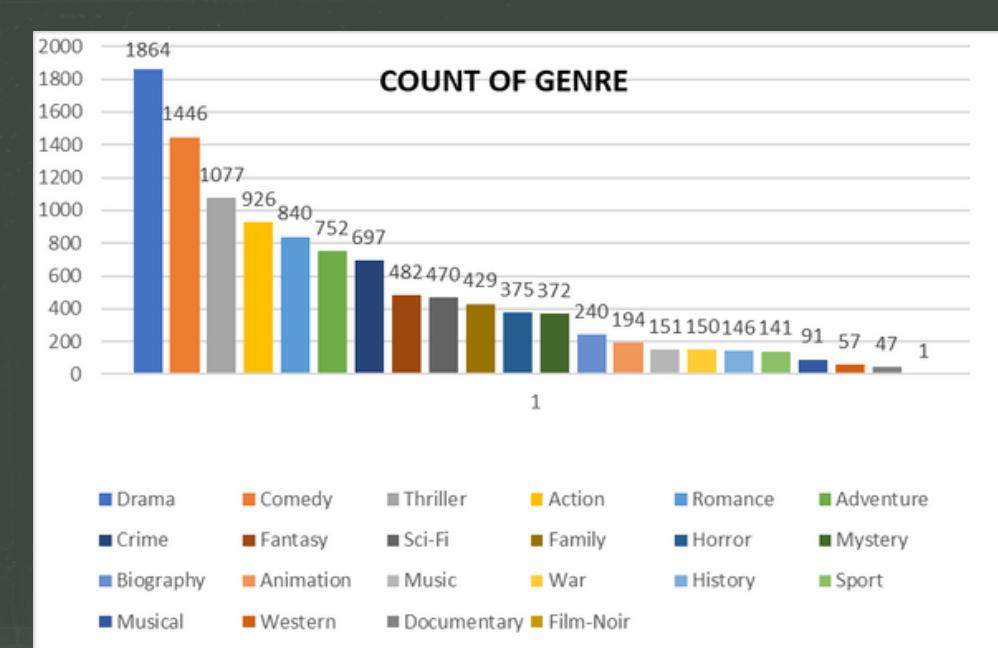
TASK: DETERMINE THE MOST COMMON GENRES OF MOVIES IN THE DATASET. THEN, FOR EACH GENRE, CALCULATE DESCRIPTIVE STATISTICS (MEAN, MEDIAN, MODE, RANGE, VARIANCE, STANDARD DEVIATION) OF THE IMDB SCORES.

Description of Insight:

The analysis reveals key insights into movie genres based on IMDb scores. Drama, with a mean of 6.47, emerges as a prevalent genre, closely followed by Comedy and Thriller.

Notably, Musical films boast the highest mean score of 6.76, while Film-Noir exhibits a singular entry with a mean score of 3.6. The standard deviation, ranging from 0.84 in History to 1.24 in Music, reflects the variability within each genre.

Drama, Adventure, and Fantasy encompass the broadest score range. Overall, this comprehensive examination provides a nuanced understanding of IMDb scores across diverse movie genres, offering valuable information for both cinephiles and industry professionals alike.



| Calculations | Mean | Median | Mode | Standard Deviation | Sample Variance | Range | Minimum | Maximum | Count |
|--------------|----------|--------|------|--------------------|-----------------|-------|---------|---------|-------|
| Drama | 6.471214 | 6.6 | 6.6 | 1.045812189 | 1.093723134 | 7.3 | 1.9 | 9.2 | 1598 |
| Comedy | 6.442576 | 6.5 | 6.7 | 1.081552265 | 1.169755301 | 7.3 | 1.6 | 8.9 | 1320 |
| Thriller | 6.472568 | 6.6 | 6.6 | 1.044429563 | 1.090833112 | 6.8 | 2.2 | 9 | 915 |
| Action | 6.460453 | 6.5 | 6.7 | 1.033598781 | 1.06832644 | 7.1 | 2.2 | 9.3 | 928 |
| Romance | 6.344228 | 6.4 | 6.3 | 1.075115105 | 1.155872488 | 7.3 | 1.6 | 8.9 | 771 |
| Adventure | 6.518284 | 6.6 | 6.7 | 1.060508984 | 1.124679305 | 7.1 | 1.9 | 9 | 711 |
| Crime | 6.439726 | 6.4 | 6.4 | 0.975455765 | 0.951513949 | 5.9 | 3 | 8.9 | 657 |
| Fantasy | 6.507407 | 6.6 | 6.7 | 1.052823291 | 1.108436882 | 6.9 | 2.4 | 9.3 | 432 |
| Sci-Fi | 6.50411 | 6.6 | 6.7 | 1.018470836 | 1.037282844 | 5.6 | 3.3 | 8.9 | 438 |
| Family | 6.476633 | 6.5 | 6.7 | 0.98904922 | 0.978218359 | 6.8 | 2.2 | 9 | 398 |
| Horror | 6.475562 | 6.6 | 5.9 | 1.084068221 | 1.175203909 | 6.2 | 2.7 | 8.9 | 356 |
| Mystery | 6.415489 | 6.5 | 7 | 0.981493207 | 0.963328915 | 6 | 2.5 | 8.5 | 368 |
| Biography | 6.357143 | 6.5 | 7.3 | 1.181293822 | 1.395455093 | 6.8 | 1.9 | 8.7 | 238 |
| Animation | 6.46911 | 6.6 | 7 | 1.02942945 | 1.059724993 | 7 | 2.3 | 9.3 | 191 |
| Music | 6.253333 | 6.4 | 7.4 | 1.236017998 | 1.527740492 | 6.4 | 1.9 | 8.3 | 150 |
| War | 6.377027 | 6.4 | 6.4 | 0.969682927 | 0.940284979 | 5.2 | 3.7 | 8.9 | 148 |
| History | 6.52069 | 6.6 | 6.1 | 0.845239656 | 0.714430077 | 4.4 | 4.1 | 8.5 | 145 |
| Sport | 6.303521 | 6.3 | 6.1 | 0.945644136 | 0.894242833 | 5.1 | 3.4 | 8.5 | 142 |
| Musical | 6.761702 | 6.85 | 7.4 | 0.936604067 | 0.877227179 | 4.2 | 4.8 | 9 | 94 |
| Western | 6.457895 | 6.6 | 6.7 | 1.163729744 | 1.354266917 | 6.6 | 2.3 | 8.9 | 57 |
| Documentary | 6.606383 | 6.8 | 6.9 | 1.030703486 | 1.062349676 | 4.2 | 3.9 | 8.1 | 47 |
| Film-Noir | 3.6 | 3.6 | #N/A | #DIV/0! | #DIV/0! | 0 | 3.6 | 3.6 | 1 |

DATA ANALYTICS TASKS 2

B. MOVIE DURATION ANALYSIS: ANALYZE THE DISTRIBUTION OF MOVIE DURATIONS AND ITS IMPACT ON THE IMDB SCORE.

TASK: ANALYZE THE DISTRIBUTION OF MOVIE DURATIONS AND IDENTIFY THE RELATIONSHIP BETWEEN MOVIE DURATION AND IMDB SCORE.

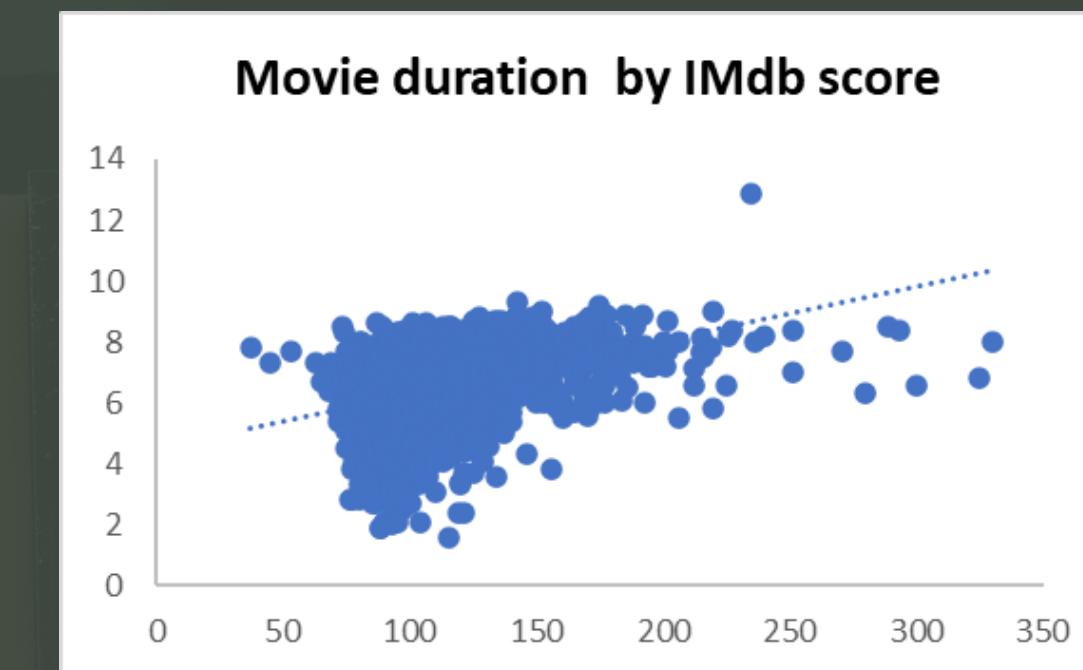
Movie Duration Analysis

Examining the distribution of movie durations uncovered intriguing insights into the relationship between duration and IMDB scores. Descriptive statistics, including mean (110.13), median (106), and mode (101), were computed for movie durations.

The scatter plot visualization further illustrated the correlation between movie duration and IMDB scores, revealing that most movies with high ratings fall within the 90 to 200 minutes range.

Notably, there is a clear trend indicating that, beyond this range, IMDB scores tend to decrease.

This nuanced understanding provides filmmakers with valuable information on optimal movie durations that resonate well with audiences, facilitating strategic decisions in film production.



| Row Labels | Sum of duration | Sum of imdb_score |
|--------------------------|-----------------|-------------------|
| The Host | 235 | 12.9 |
| The Shawshank Redem | 142 | 9.3 |
| The Godfather | 175 | 9.2 |
| The Dark Knight | 152 | 9 |
| The Godfather: Part II | 220 | 9 |
| Pulp Fiction | 178 | 8.9 |
| Schindler's List | 185 | 8.9 |
| The Good, the Bad and | 142 | 8.9 |
| The Lord of the Rings: 1 | 192 | 8.9 |
| Fight Club | 151 | 8.8 |
| Forrest Gump | 142 | 8.8 |
| Inception | 148 | 8.8 |
| Star Wars: Episode V - | 127 | 8.8 |
| The Lord of the Rings: 1 | 171 | 8.8 |
| City of God | 135 | 8.7 |
| Goodfellas | 146 | 8.7 |
| One Flew Over the Cuck | 133 | 8.7 |
| Seven Samurai | 202 | 8.7 |
| Star Wars: Episode IV - | 125 | 8.7 |
| The Lord of the Rings: 1 | 172 | 8.7 |
| The Matrix | 136 | 8.7 |
| American History X | 101 | 8.6 |
| Interstellar | 169 | 8.6 |
| Modern Times | 87 | 8.6 |
| Saving Private Ryan | 169 | 8.6 |
| Se7en | 127 | 8.6 |
| Spirited Away | 125 | 8.6 |

DATA ANALYTICS TASKS 3

**C. LANGUAGE ANALYSIS: SITUATION: EXAMINE THE DISTRIBUTION OF
MOVIES BASED ON THEIR LANGUAGE.**

**TASK: DETERMINE THE MOST COMMON LANGUAGES USED IN MOVIES AND
ANALYZE THEIR IMPACT ON THE IMDB SCORE USING
DESCRIPTIVE STATISTICS.**

Language Analysis - Descriptive Statistics and Impact on Movie Ratings

The language analysis provides a comprehensive overview of the impact of various languages on IMDb scores. English-language films dominate the dataset, showcasing a mean score of 6.42.

Notably, Filipino films stand out with an impressive mean of 7.36, while Mandarin films closely follow. The dataset includes diverse languages such as Aramaic, Dari, and Zulu, each contributing distinct mean scores. Interestingly, films with no specific language information also exhibit a high mean of 8.13. The standard deviation varies across languages, indicating the degree of dispersion in IMDb scores.

While some languages, like German, exhibit minimal variability, others like Portuguese and Filipino display a wider range. This analysis highlights the intriguing interplay between language and IMDb scores, offering valuable insights into the global cinematic landscape and audience preferences for films in different languages.

| English | Mean | Median | Mode | Standard Deviation | Range | Minimum | Maximum | Count |
|-------------|-----------|--------|------|--------------------|-------|---------|---------|-------|
| Mandarin | 6.4227015 | 6.5 | 6.7 | 1.05190093 | 7.7 | 1.6 | 9.3 | 3524 |
| Filipino | 7.3558824 | 7.3 | 7.2 | 0.519435111 | 2.6 | 5.8 | 8.4 | 34 |
| French | 7.0826087 | 7.2 | 5.9 | 0.860577065 | 3 | 5.2 | 8.2 | 23 |
| Aboriginal | 7.0214286 | 7.25 | 7.6 | 0.765786244 | 2.3 | 5.6 | 7.9 | 14 |
| Mongolian | 7.7636364 | 7.8 | 7.4 | 0.675681474 | 2.4 | 6.1 | 8.5 | 11 |
| Aramaic | 7.66 | 8 | #N/A | 0.990173947 | 2.7 | 6 | 8.7 | 10 |
| Japanese | 7.3428571 | 7.3 | 7.3 | 0.350509833 | 1.1 | 6.7 | 7.8 | 7 |
| Dutch | 7.1857143 | 7 | #N/A | 1.155318962 | 3.6 | 5.3 | 8.9 | 7 |
| Hungarian | 7.7 | 7.7 | #N/A | 0.570087713 | 1.4 | 7 | 8.4 | 5 |
| Danish | 7.22 | 7.4 | #N/A | 0.801249025 | 2 | 6 | 8 | 5 |
| Norwegian | 7.76 | 8 | #N/A | 0.978774744 | 2.6 | 6.1 | 8.7 | 5 |
| Czech | 7.15 | 7.3 | 7.6 | 0.574456265 | 1.2 | 6.4 | 7.6 | 4 |
| Dari | 7.5666667 | 7.8 | 7.8 | 0.404145188 | 0.7 | 7.1 | 7.8 | 3 |
| Bosnian | 6.6333333 | 6.6 | #N/A | 0.450924975 | 0.9 | 6.2 | 7.1 | 3 |
| Portuguese | 7.9 | 8.1 | #N/A | 0.529150262 | 1 | 7.3 | 8.3 | 3 |
| <i>NONE</i> | 8.1333333 | 8.4 | #N/A | 0.550757055 | 1 | 7.5 | 8.5 | 3 |
| Spanish | 6.95 | 6.95 | #N/A | 0.777817459 | 1.1 | 6.4 | 7.5 | 2 |
| German | 7.5 | 7.5 | #N/A | 0.141421356 | 0.2 | 7.4 | 7.6 | 2 |
| Romanian | 7.9 | 7.9 | #N/A | 0.424264069 | 0.6 | 7.6 | 8.2 | 2 |
| Maya | 6.7 | 6.7 | #N/A | #DIV/0! | 0 | 6.7 | 6.7 | 1 |
| Kazakh | 7.8 | 7.8 | #N/A | #DIV/0! | 0 | 7.8 | 7.8 | 1 |
| Cantonese | 6 | 6 | #N/A | #DIV/0! | 0 | 6 | 6 | 1 |
| Italian | 7.1 | 7.1 | #N/A | #DIV/0! | 0 | 7.1 | 7.1 | 1 |
| Thai | 7.3 | 7.3 | #N/A | #DIV/0! | 0 | 7.3 | 7.3 | 1 |
| Korean | 4.3 | 4.3 | #N/A | #DIV/0! | 0 | 4.3 | 4.3 | 1 |
| Hindi | 7.1 | 7.1 | #N/A | #DIV/0! | 0 | 7.1 | 7.1 | 1 |
| Russian | 7.4 | 7.4 | #N/A | #DIV/0! | 0 | 7.4 | 7.4 | 1 |
| <i>None</i> | 6.5 | 6.5 | #N/A | #DIV/0! | 0 | 6.5 | 6.5 | 1 |
| Zulu | 8.5 | 8.5 | #N/A | #DIV/0! | 0 | 8.5 | 8.5 | 1 |
| Hebrew | 7.3 | 7.3 | #N/A | #DIV/0! | 0 | 7.3 | 7.3 | 1 |
| Arabic | 8 | 8 | #N/A | #DIV/0! | 0 | 8 | 8 | 1 |
| Vietnamese | 7.2 | 7.2 | #N/A | #DIV/0! | 0 | 7.2 | 7.2 | 1 |
| Indonesian | 7.4 | 7.4 | #N/A | #DIV/0! | 0 | 7.4 | 7.4 | 1 |
| Persian | 7.9 | 7.9 | #N/A | #DIV/0! | 0 | 7.9 | 7.9 | 1 |

DATA ANALYTICS TASKS 4

**D. DIRECTOR ANALYSIS: INFLUENCE OF DIRECTORS ON MOVIE
RATINGS.**

**TASK: IDENTIFY THE TOP DIRECTORS BASED ON THEIR AVERAGE
IMDB SCORE AND ANALYZE THEIR CONTRIBUTION
TO THE SUCCESS OF MOVIES USING
PERCENTILE CALCULATIONS.**

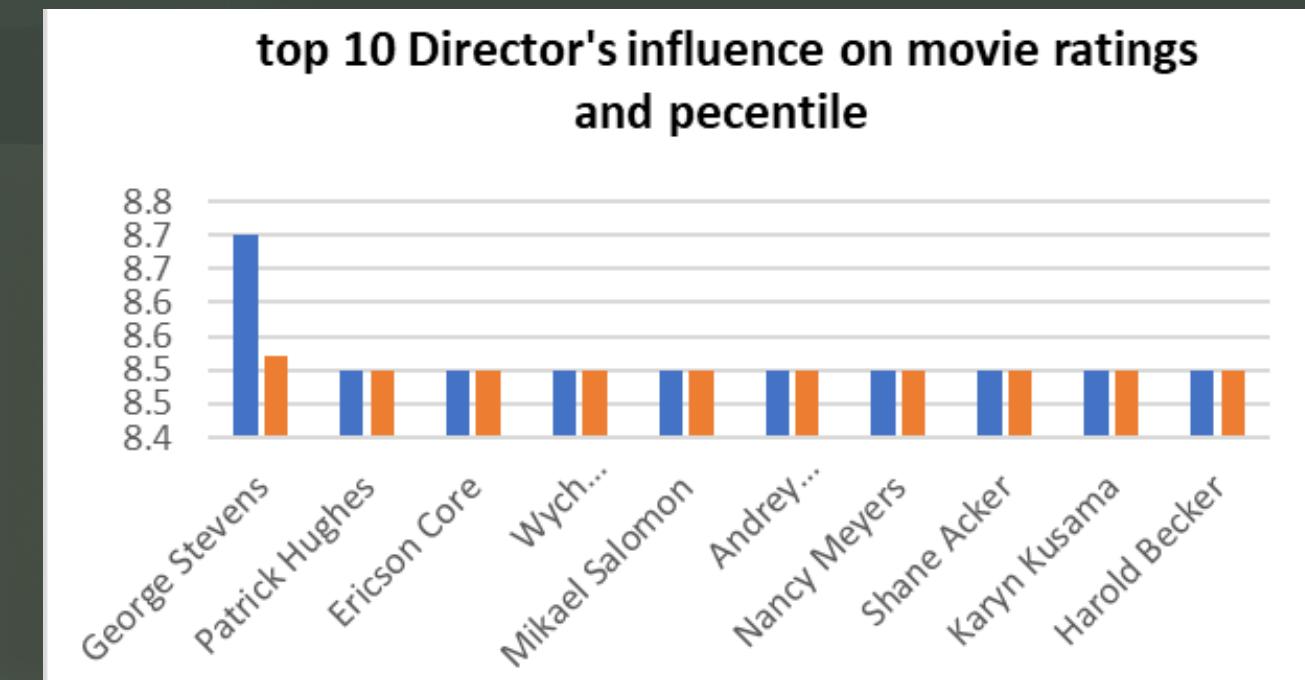
Director Analysis

The director analysis reveals intriguing patterns in the influence of directors on movie ratings.

Noteworthy directors such as George Stevens, Patrick Hughes, Ericson Core, and others consistently exhibit high average IMDB scores, with George Stevens leading with a mean score of 8.7.

The percentile calculations, indicating the position of each director within the dataset, are consistently around 7.5. This suggests that these directors consistently produce movies that perform exceptionally well in terms of audience ratings.

Filmmakers and producers can leverage this information to collaborate with renowned directors, understanding their historical impact on movie success.



| E | F | G |
|----------------------|------|------------|
| director_name | MEAN | percentile |
| George Stevens | 8.7 | 7.5 |
| Patrick Hughes | 8.5 | 7.5 |
| Ericson Core | 8.5 | 7.5 |
| Wych Kaosayananda | 8.5 | 7.5 |
| Mikael Salomon | 8.5 | 7.5 |
| Andrey Konchalovskiy | 8.5 | 7.5 |
| Nancy Meyers | 8.5 | 7.5 |
| Shane Acker | 8.5 | 7.5 |
| Karyn Kusama | 8.5 | 7.5 |
| Harold Becker | 8.5 | 7.5 |
| Britt Allcroft | 8.5 | 7.5 |
| Wallace Wolodarsky | 8.4 | 7.5 |
| Kevin Jordan | 8.4 | 7.5 |
| Mike Bruce | 8.4 | 7.5 |
| Ben Affleck | 8.3 | 7.5 |
| Mort Nathan | 8.2 | 7.5 |
| Peter Hastings | 8.2 | 7.5 |
| Mickey Liddell | 8.2 | 7.5 |
| Vincent Ward | 8.1 | 7.5 |
| Danny Steinmann | 8.1 | 7.5 |
| Michael Landon Jr. | 8.1 | 7.5 |
| William Wyler | 8.1 | 7.5 |

DATA ANALYTICS TASKS 5

**E. BUDGET ANALYSIS: EXPLORE THE RELATIONSHIP BETWEEN
MOVIE BUDGETS AND THEIR FINANCIAL SUCCESS.**

**TASK: ANALYZE THE CORRELATION BETWEEN MOVIE BUDGETS AND
GROSS EARNINGS, AND IDENTIFY THE MOVIES
WITH THE HIGHEST PROFIT MARGIN.**

Budget Analysis

The budget analysis unveils compelling insights into the relationship between movie budgets, gross earnings, and profit margins.

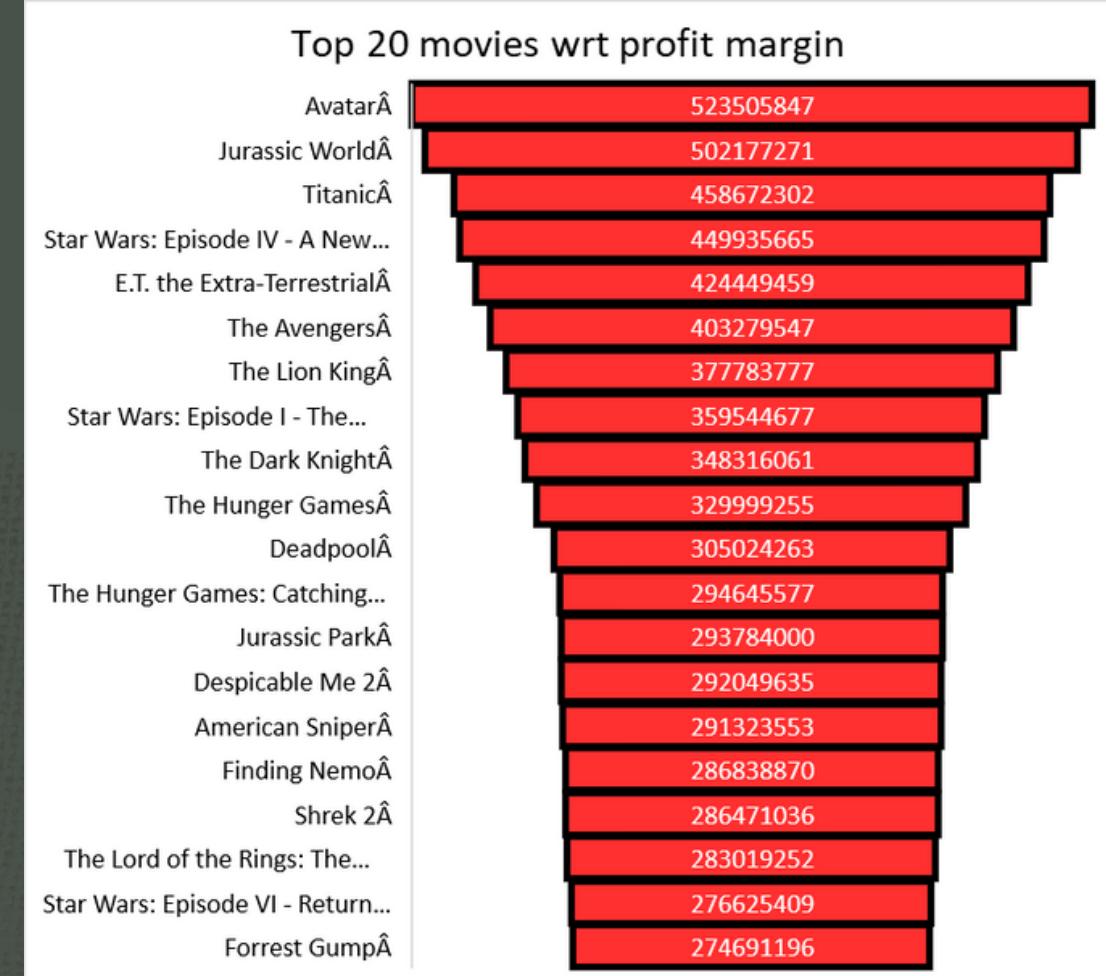
Notable movies like "Avatar," "Jurassic World," and "Titanic" stand out with substantial budgets and impressive profit margins, showcasing their financial success.

The correlation coefficients indicate a weak positive correlation between movie budgets and gross earnings, suggesting that higher budgets do not necessarily guarantee higher earnings.

Notably, movies such as "The Avengers" and "The Dark Knight" achieved remarkable profit margins despite their considerable budgets.

Filmmakers can utilize these findings to make informed decisions about budget allocation and maximize profit margins.

| Movies | Sum of budget | Sum of gross | Profit Margin | correlation |
|--|---------------|--------------|---------------|-------------|
| Avatar | 237000000 | 760505847 | 523505847 | 0.099104722 |
| Jurassic World | 150000000 | 652177271 | 502177271 | 0.098195268 |
| Titanic | 200000000 | 658672302 | 458672302 | 0.098148543 |
| Star Wars: Episode IV - A New Hope | 11000000 | 460935665 | 449935665 | 0.097583646 |
| E.T. the Extra-Terrestrial | 10500000 | 434949459 | 424449459 | 0.098346809 |
| The Avengers | 220000000 | 623279547 | 403279547 | 0.099044176 |
| The Lion King | 45000000 | 422783777 | 377783777 | 0.098262303 |
| Star Wars: Episode I - The Phantom Menace | 115000000 | 474544677 | 359544677 | 0.098704339 |
| The Dark Knight | 185000000 | 533316061 | 348316061 | 0.098740131 |
| The Hunger Games | 78000000 | 407999255 | 329999255 | 0.098253662 |
| Deadpool | 58000000 | 363024263 | 305024263 | 0.098456347 |
| The Hunger Games: Catching Fire | 130000000 | 424645577 | 294645577 | 0.098706784 |
| Jurassic Park | 63000000 | 356784000 | 293784000 | 0.098582142 |
| Despicable Me 2 | 76000000 | 368049635 | 292049635 | 0.098797336 |
| American Sniper | 58800000 | 350123553 | 291323553 | 0.098957128 |
| Finding Nemo | 94000000 | 380838870 | 286838870 | 0.099189639 |
| Shrek 2 | 150000000 | 436471036 | 286471036 | 0.099262389 |
| The Lord of the Rings: The Return of the King | 94000000 | 377019252 | 283019252 | 0.099012026 |
| Star Wars: Episode VI - Return of the Jedi | 32500000 | 309125409 | 276625409 | 0.099082956 |
| Forrest Gump | 55000000 | 329691196 | 274691196 | 0.099385928 |
| Star Wars: Episode V - The Empire Strikes Back | 18000000 | 290158751 | 272158751 | 0.099613923 |
| Home Alone | 18000000 | 285761243 | 267761243 | 0.099950269 |
| Star Wars: Episode III - Revenge of the Sith | 113000000 | 380262555 | 267262555 | 0.100278678 |
| Spider-Man | 139000000 | 403706375 | 264706375 | 0.100243697 |





DEXCEL SHEET LINK:

[https://docs.google.com/spreadsheets/d/13JV2TJXDNIxFXUEWILWNAQ4ZR9K_7AGL/edit?
usp=sharing&ouid=104761672347413617130&rtpof=true&sd=true](https://docs.google.com/spreadsheets/d/13JV2TJXDNIxFXUEWILWNAQ4ZR9K_7AGL/edit?usp=sharing&ouid=104761672347413617130&rtpof=true&sd=true)

PRESENTED BY:

ADNAN ASHRAF

VIDEO PRESENTATION LINK:

https://drive.google.com/file/d/1DP59AXGzAcQR6u5t_jEy2CFDCpkxiwhY/view?usp=sharing