IMDB MOVIE ANALYSIS FINAL PROJECT REPORT

Project Description

The goal of this project was to investigate the factors influencing the success of movies on IMDB. Success, in this context, was measured by high IMDB ratings and financial performance. Understanding these factors is essential for movie producers, directors, and investors seeking to make data-driven decisions to improve the likelihood of a movie's success.

The analysis was performed using the IMDB Movies dataset, exploring variables such as genre, movie duration, language, director influence, and budget.

Approach

The project was conducted in a structured, task-wise manner:

- Data Cleaning: Organized dataset in Google Sheets, handled missing values, created additional columns for derived insights.
  Task-wise Analysis: Each core factor was explored as an independent task with clear visualizations and descriptive statistics.
- Statistical Techniques: Used Excel/Google Sheets functions including `AVERAGE`, `MEDIAN`, `MODE`, `VAR`, `STDEV`, `CORREL`, `MAX`, `INDEX`, `MATCH`.
- Visualization: Created bar charts, scatter plots, and trendlines for effective data storytelling.
- Report Drafting: Summarized findings task-wise with actionable insights.

Tech-Stack Used

- **Google Sheets: Version June 2025 (Browser-based)**
- **Features Used: Functions, Pivot Tables, Data Sorting, Filtering, Charting Tools**
- **Purpose: Data analysis, visualization, and report preparation**

IMDB Movie Analysis — Task A Detailed Report (Genre Analysis)

Task A: Movie Genre Analysis

Objective:
  To analyze the distribution of movie genres and their impact on IMDB Scores, helping stakeholders understand which genres tend to perform better in terms of ratings.

Step-by-Step Work Performed:

1 Dataset Preparation

- Imported dataset into **Google Sheets**
- Identified columns:

    - **J to Q:** Contained multiple genres per movie
      AB: IMDB Score of each movie

2Extracting Unique Genres

Step:
 Created a list of all unique genres appearing across columns J to Q.

Formula Used in AC2:

=UNIQUE(FLATTEN(J2:Q))

- Column **AC:** Unique Genres list generated

3Calculating Movie Count per Genre

Formula in AD2:

=COUNTIF(FLATTEN(J2:Q), AC2)

- Column **AD:** Total number of movies for each genre

4Genre-wise Descriptive Statistics

| **AE:** Mean IMDB Score
| **AF:** Median IMDB Score
| **AG:** Mode IMDB Score
| **AH:** Range (Max - Min)
| **AI:** Variance
| **AJ:** Standard Deviation

Formulas Used:

- **Mean IMDB Score (AE2):**

=AVERAGEIF(FLATTEN(J2:Q), AC2, AB2:AB)

- **Median IMDB Score (AF2):**

=MEDIAN(FILTER(AB2:AB, (MMULT(N(J2:Q=AC2), TRANSPOSE(SIGN(COLUMN(J2:Q))))>0)))

- **Mode IMDB Score (AG2):**

=MODE(FILTER(AB2:AB, (MMULT(N(J2:Q=AC2), TRANSPOSE(SIGN(COLUMN(J2:Q))))>0)))

- **Range (AH2):**

=MAX(FILTER(AB2:AB, (MMULT(N(J2:Q=AC2), TRANSPOSE(SIGN(COLUMN(J2:Q))))>0))) - MIN(FILTER(AB2:AB, (MMULT(N(J2:Q=AC2), TRANSPOSE(SIGN(COLUMN(J2:Q))))>0)))

- **Variance (AI2):**

=VAR.P(FILTER(AB2:AB, (MMULT(N(J2:Q=AC2), TRANSPOSE(SIGN(COLUMN(J2:Q))))>0)))

- **Standard Deviation (AJ2):**
  =STDEV.P(FILTER(AB2:AB, (MMULT(N(J2:Q=AC2), TRANSPOSE(SIGN(COLUMN(J2:Q))))>0)))
- Dragged formulas down for all genres

Charts Created:

1. Bar Chart: Average IMDB Score by Genre

- **X-axis:** Genres
- **Y-axis:** Mean IMDB Score

Steps:

- Selected **AC (Genres)** and **AE (Mean IMDB Score)**
- Inserted Bar Chart
- Customized chart title to "Average IMDB Score by Genre"

2. Combo Chart: Movie Count vs IMDB Score

- **Bar:** Movie Count per Genre (AD)
- **Line:** Mean IMDB Score per Genre (AE)

Steps:

- Selected **AC to AE** range
  Inserted Combo Chart
- Assigned:

  - Movie Count → Bar type
  - Mean IMDB Score → Line type

3. INSIGHTS:

- Extracted unique genres from multi-genre movies.
- Calculated genre-wise statistics:
  - Mean, Median, Mode, Range, Variance, Standard Deviation of IMDB Scores.
- Created bar chart showing distribution and average IMDB score per genre.
- Top 5 Highest Rated Genres Identified.

Key Findings:

- Popular genres (Action, Adventure) dominate in quantity but not always in quality
  Genres like Documentary and Musical show consistently high average IMDB Scores
- Combo chart highlights imbalance between movie output and critical reception

Conclusion:

- Genre selection impacts movie ratings
- This analysis helps studios and investors align project choices with audience preferences
- Sets the foundation for deeper insights in subsequent tasks

IMDB Movie Analysis — Task B: Movie Duration Analysis

1 Objective

The goal of Task B is to analyze the distribution of movie durations and examine the relationship between movie length and its IMDB rating. This helps filmmakers and stakeholders understand whether the runtime of a movie impacts audience perception and success.

2Approach

- Dataset utilized from IMDB movie records
- A new dedicated sheet was created specifically for Task B
- Columns used:

  - **D:** Movie Duration (in minutes)
  - **Z:** IMDB Score
  - **AC to AE:** Empty columns reserved for statistical calculations

Specific Steps Taken:

- **Descriptive Statistics Calculated:**

  - Mean Duration
  - Median Duration
  - Standard Deviation of Duration

Formulas Used:

| AC1 (Mean Duration):

=AVERAGE(D2:D)

| AD1 (Median Duration):

=MEDIAN(D2:D)

| AE1 (Standard Deviation):

=STDEV.P(D2:D)

- **Scatter Plot Created:**

  - X-axis: Movie Duration
  - Y-axis: IMDB Score

- **Trendline Added:**

  - Type: Linear
  - Optional R^2 value displayed to assess strength of correlation

3 Tech-Stack Used

- **Google Sheets (Web version)**
- Functions Applied:

  - AVERAGE, MEDIAN, STDEV.P for descriptive statistics
  - Chart tools for visual analysis

4 Insights

- analyzed distribution of movie durations.
- Calculated Mean, Median, Standard Deviation of durations.
- Scatter plot of Duration vs IMDB Scores with Trendline.
- Insight: Moderate positive correlation between movie duration and ratings

5 Results

- No definitive evidence that longer movies automatically guarantee higher ratings
- Duration alone is an incomplete predictor of movie success
- Findings can guide filmmakers on optimal movie lengths but should be considered alongside other factors (genre, director, production quality, etc.)

Conclusion

This task reinforced the importance of runtime as a contributing factor, though not the sole determinant of success. Future tasks will explore other influences such as language, directors, and budgets to provide a holistic understanding of movie performance.

IMDB Movie Analysis — Task C: Language Analysis

---

1 Objective

The goal of Task C is to examine the distribution of movies based on language and analyze how language influences IMDB ratings. This provides insights for producers and distributors to understand language-specific trends and audience preferences.

2 Approach

- Created a new dedicated sheet for Task C

- Dataset Columns:

  - **T:** Language
  - **Z:** IMDB Score
  - **AC to AG:** Calculations

Specific Steps Taken:

**Extracted Unique Languages
| Formula in AC1: |**

=UNIQUE(T2:T)

**Movie Count per Language
| Formula in AD2: |**

=COUNTIF(T2:T, AC2)

Descriptive Statistics of IMDB Scores by Language
| Mean IMDB Score (AE2): |

=AVERAGEIF(T2:T, AC2, Z2:Z)

| Median IMDB Score (AF2): |

=MEDIAN(FILTER(Z2:Z, T2:T=AC2))

| Standard Deviation (AG2): |

=STDEV.P(FILTER(Z2:Z, T2:T=AC2))

Applied formulas for all unique languages

Charts Created:

- Bar Chart: Movie Count by Language

- Combo Chart: Movie Count vs Mean IMDB Score

3 Tech-Stack Used

- **Google Sheets**

- Functions:

  - UNIQUE, COUNTIF, AVERAGEIF, MEDIAN, STDEV.P, FILTER

- Chart Tools for visual interpretation

4 Insights

- Counted movie distribution by language.
- Calculated Mean, Median, Standard Deviation of IMDB scores per language.
- Bar chart visualizing language-wise performance.
- Finding: Certain languages consistently perform better on IMDB.

5Results

- Language significantly influences movie production volume

- Non-English language films, though fewer in number, may achieve competitive or superior ratings

- Insight helps studios target language markets effectively based on success potential

Conclusion

Language plays a critical role in both production strategy and audience perception. While English films dominate in number, opportunities exist in other language markets, especially where audience reception is stronger.

Task D — Director Analysis

Objective

Analyze how directors influence movie success by evaluating their average IMDB scores and ranking them using percentile calculations.

Approach & Steps Followed

Used a new sheet exclusively for Task D
Columns:

- **B:** Director Name
  Z: IMDB Score
- **AC:** Unique Director Names
- **AD:** Average IMDB Score per Director
- **AE:** Percentile Rank

Formulas Applied

Unique Director List:

=UNIQUE(B2:B)

Average IMDB Score per Director:

=IF(AC2="", "", AVERAGEIF(B:B, AC2, Z:Z))

Percentile Rank Calculation:

=IF(AD2="", "", PERCENTRANK(FILTER(AD$2:AD, AD$2:AD<>""), AD2))

Visualization

- Bar Chart created for **Top Directors by Average IMDB Score**
  Filtered Percentile Rank to show directors in the **top 5%**
- Highlighted their corresponding average IMDB scores

Insights

- Calculated average IMDB scores per director.
- Used percentile calculations to highlight top-performing directors.
- Identified directors significantly influencing movie success.

Tech Stack

- Google Sheets
  Functions used: UNIQUE, AVERAGEIF, PERCENTRANK, FILTER
- Chart Tools for Bar Chart

Conclusion

Director reputation plays a measurable role in movie success, as shown by their consistent IMDB scores.

Awesome! Here's your **Task E: Budget Analysis Report Draft** — ready to plug into your main report or presentation:

Task E — Budget Analysis

Objective

Analyze the relationship between movie budgets and their financial success by exploring the correlation between budget and gross earnings, as well as identifying the movie with the highest profit margin.

Approach & Steps Followed

Created a dedicated sheet for Task E analysis
Defined relevant columns:

- **W:** Budget
- **I:** Gross Earnings
  L: Movie Title

- **Z:** IMDB Score

   Added new calculated columns:

- **AC:** Budget-Gross Correlation
- **AD:** Profit Margin (Gross - Budget)
- **AE:** Highest Profit Margin
- **AF:** Movie with Highest Profit

Formulas Applied

✔️ **Budget-Gross Correlation:**

=CORREL(W2:W, I2:I)

✔️ **Profit Margin Calculation:**

=I2 - W2

✔️ **Highest Profit Margin:**

=MAX(AD2:AD)

✔️ **Movie with Highest Profit:**

=INDEX(L2:L, MATCH(AE2, AD2:AD, 0))

Visualization Created

 Scatter Plot:

- **X-axis:** Budget
- **Y-axis:** Gross Earnings
  Chart Title: "Relationship Between Movie Budget and Gross Earnings"
  Added Trendline with R^2 value to visually assess the correlation

   Insights

- The calculated correlation coefficient is approximately **0.10**, indicating a **very weak positive relationship** between movie budget and gross earnings.
- High budgets do not guarantee high earnings based on this dataset.
- The movie with the highest profit margin was identified using the dataset's calculations.

Tech Stack

- Google Sheets for data cleaning, calculations, and visualization

- Formulas: `CORREL`, `MAX`, `INDEX`, `MATCH`
- Chart Tools for Scatter Plot and Trendline

Conclusion

The analysis reveals that a movie's financial success is **not solely dependent on its budget**. Other factors such as content, marketing, star power, and audience reception may play significant roles. Investors should not rely on budget size as the only predictor of success.

Result

The project provided valuable insights for stakeholders in the film industry:

- Certain genres and languages correlate with higher IMDB ratings.
- Duration and director reputation have measurable influence.
- Budget alone does not guarantee higher gross earnings.
- Profit margins vary significantly, highlighting the need for strategic planning.

These insights enable data-driven decision-making for movie production and investment.