**IMDB Movie Analysis**

**1. Objective**

This project aims to **identify key factors that influence a movie's success** on IMDB. By analyzing patterns and correlations, we can provide actionable insights to help producers, directors, and investors make informed decisions for future film projects.

**2. Approach**

**Data Collection & Preprocessing:**

* Imported, cleaned, and transformed the dataset using Excel.
* Performed data transformation to ensure consistency and usability.

**Data Analysis:**

* Conducted statistical analysis and summarized data using Pivot Tables in Excel.
* Performed deeper insights and advanced visualizations using Power BI.

**Data Visualization:**

* Used Power BI’s robust charting and visualization tools to identify key trends influencing movie ratings.

**Reporting:**

* Created interactive dashboards in Power BI to present findings and share insights with stakeholders.
* Performed drill-through analysis to explore directors-specific insights.

**Documentation:**

* Documented all the processes, insights, and recommendations for better understanding.

**3. Tech Stack Used**

* **Excel:** Data cleaning, Pivot Tables, and basic statistical analysis
* **Power BI:** Advanced analytics, DAX calculations, and interactive dashboard creation.
* **MS Word:** Documenting the project and insights.

**4. Data Cleaning & Preprocessing**

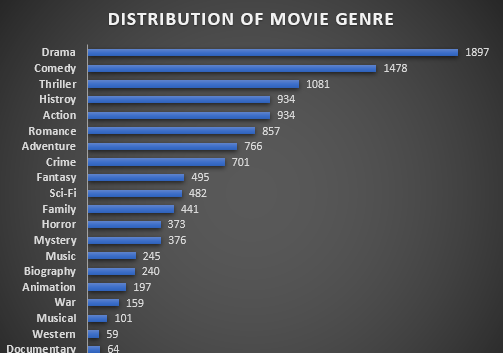
* **Removed Unnecessary Columns:** Dropped irrelevant columns to streamline the dataset.
* **Handled Missing Values:** Removed missing values from director name, budget, and gross as they had missing values that couldn't be reliably imputed.
* **Corrected Data Types:** Converted gross and budget from whole numbers to currency format.
* **Removed Duplicates:** Ensured data integrity by eliminating duplicate entries.
* **Handled Outliers:** Applied log transformation to budget and gross to reduce skewness.
* **Added Derived Column:** Created Profit Margin using the formula: Profit Margin=((gross−budget)/budget)\*100

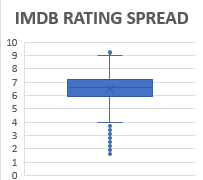
**5. Business Problem Addressed**

**A. Movie Genre Analysis**

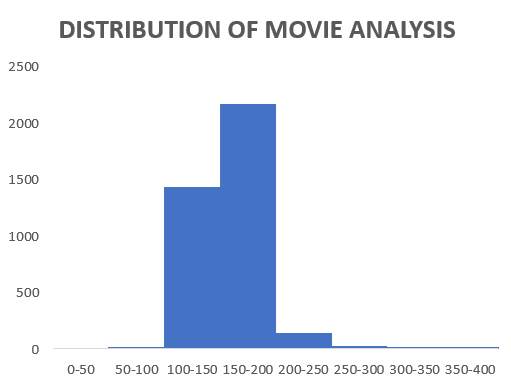
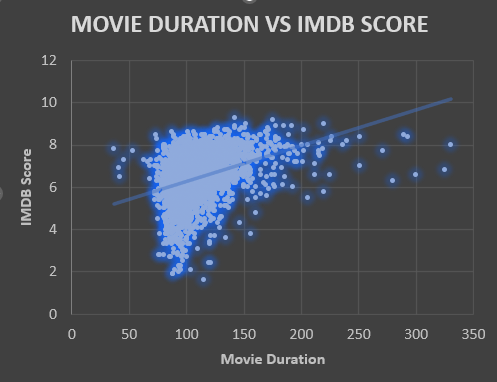
**Objective:** Examine genre distribution and its impact on IMDb scores.

* Used **COUNTIF()** to calculate the frequency of each genre.
* Created a **Bar Chart** to analyze genre distribution.
* Performed **descriptive analysis** to evaluate the impact of genres on IMDb scores.
* Used a **Box & Whiskers Plot** to understand IMDb rating spread.
* **Findings:** The most popular genres are **Drama, Comedy, and Thriller**.



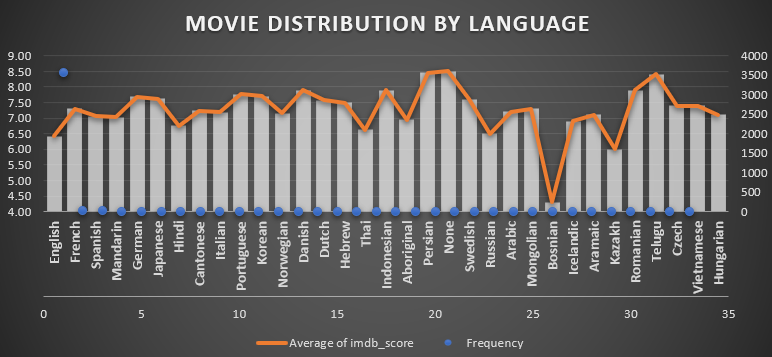


|  |  |  |  |
| --- | --- | --- | --- |
| **B. Movie Duration Analysis**  **Objective:** Evaluate how movie length influences ratings.   * Created a **Frequency Table** and plotted a **Histogram** to show movie duration distribution. * Created a **Scatter Plot** with a **trendline** to analyze movie duration vs. IMDb scores. * **Findings:**   + A **positive correlation** exists between movie duration and IMDb scores.   + **Longer movies tend to have higher IMDb scores on average.** | | |  |
|  |  |  |  |



**C. Language Analysis**

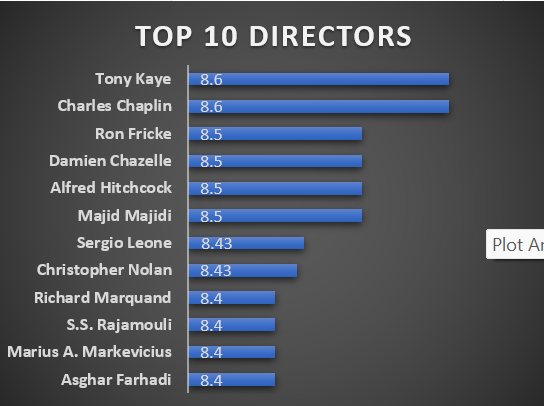
* **Objective:** Analyze the distribution of movies by language.
* Used **Pivot Tables** to calculate the frequency and average IMDb score of languages.
* Created a **Scatter Plot** to analyze common movie languages.
* Used a **Line Chart with error bars** to display language impact on IMDb scores.
* **Findings:**
  + **English** is the most common language used in movies.
  + However, its **average IMDb score is moderate**, indicating that **quality matters more than language**.



**D. Director Influence**

**Objective:** Assess the role of directors in movie ratings.

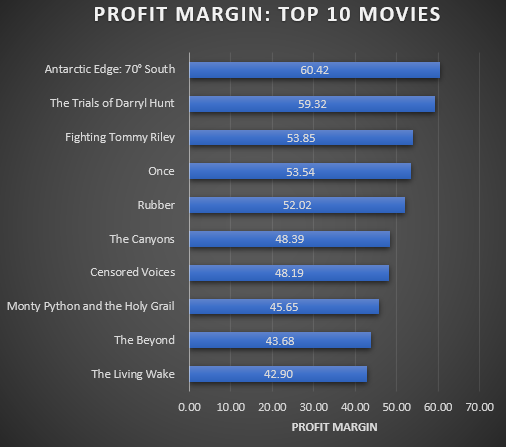
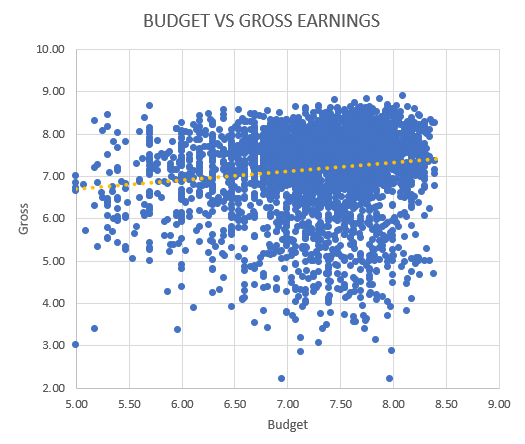
* Created **Pivot Charts** to identify top directors based on average IMDb scores.
* Used a **Bar Chart** to display the **top 10 directors**.



**E. Budget vs. Success**

**Objective:** Explore the relationship between budgets and financial performance.

* Created a **Scatter Plot** to show the correlation between movie budgets and gross earnings.
* **Findings:**
  + **Higher budgets tend to result in higher gross earnings.**
* Used a **Bar Chart** to analyze movie profit margins.



**6. Data Visualization & Reporting**

**A. Filters**

Users can refine the data using the following filters:

* **Filter by Country**: Select movies from a specific country.
* **Filter by Year**: Choose a time range (1920 - 2016) to analyze trends over different periods.

**B. Key Performance Indicators (KPIs)**

Displayed on the left side of the dashboard, these KPIs provide a quick summary of essential movie metrics:

* **Average Gross:** ₹50.54M (represents the average revenue generated per movie).
* **Average Budget:** ₹37M (shows the average production budget).
* **Average Rating:** 6.46 (IMDb rating based on audience reviews).
* **Director Count:** 1,734 (total number of unique directors in the dataset).
* **Genre Count:** 20 (total number of unique genres in the dataset).

**C. Visualizations & Insights**

1. **Total Revenue by Duration (Line Chart)**
   * Shows how movie duration affects revenue.
   * Revenue tends to peak around a certain duration (100-150 minutes).
   * Longer movies (>200 min) tend to generate lower revenue.
2. **IMDb Score vs Duration (Scatter Plot with Trend Line)**
   * Displays the correlation between a movie’s duration and its IMDb rating.
   * Trend line suggests that moderately long movies (80-150 min) generally receive higher ratings.
3. **Top Directors (Bar Chart)**
   * Lists directors with the highest IMDb ratings.
   * **Notable directors include**:
     + Tony Kaye (8.60)
     + Ron Fricke (8.50)
     + Christopher Nolan (8.43)
     + Charles Chaplin (8.60)
     + Alfred Hitchcock (8.50)
4. **Director vs Profit Margin (Bar Chart)**
   * Highlights the profitability of directors based on revenue vs budget.
   * Directors like **Dena Seidel, Ricki Stern, and Quentin D.** show high profit margins.
5. **Language vs Rating (Donut Chart)**
   * Displays how different languages impact IMDb ratings.
   * Top-rated languages include:
     + Spanish (21.68%)
     + German (20.56%)
     + Mandarin (19.88%)
     + English (18.09%)
6. **Rating: Top 5 Genre (Horizontal Bar Chart)**
   * Shows the most popular genres based on IMDb ratings and movie count.
   * **Top 5 genres include**:
     + Drama (358 movies)
     + Comedy (302 movies)
     + Adventure (258 movies)
     + Action (252 movies)
     + Thriller (202 movies)

**7. Insights & Results**

* **Popular Genres:** Drama, Comedy, and Thriller are the most preferred genres.
* **Market Trends:** Viewers prefer movies with a duration between 100 to 200 minutes.
* **Higher IMDB Scores:** Strong audience engagement and interactivity contribute to better ratings.
* **Quality Matters:** The quality of a movie is more significant than just its budget.
* **Budget & Earnings:** Higher-budget movies tend to generate higher gross earnings.

**8. Recommendation**

* Focus on producing **Drama, Comedy, and Thriller** films.
* Optimize movie durations between **100–200 minutes** to align with viewer preferences.
* Allocate budgets effectively, ensuring a balance between **quality and financial returns**.

**9. Future Enhancements**

* **Expand Timeframe:** Include data beyond 2016 for more recent trends.
* **Advanced Analytics:** Machine learning predictions for box office performance.