**IMDB MOVIE ANALYSIS**

**1. Project Description**

**Objective**

The goal of this project is to analyze the factors influencing the success of movies on IMDB. Success is defined by high IMDB ratings. This analysis is crucial for movie producers, directors, and investors to understand what makes a movie successful and make informed decisions.

**Approach**

We will analyze various aspects of the dataset, including genre, duration, language, director influence, and budget correlation with gross earnings. The project involves:

* **Data Cleaning & Preprocessing**
* **Exploratory Data Analysis (EDA)**
* **Statistical Analysis**
* **Data Visualization**
* **Insights and Recommendations**

**Data set overview:**

* It contains 5,044 rows and 28 columns.
* Some key columns include:
* Title
* Genre
* IMDB Score
* Director
* Language
* Duration
* Budget
* Gross Earnings

**2. Tech-Stack Used**

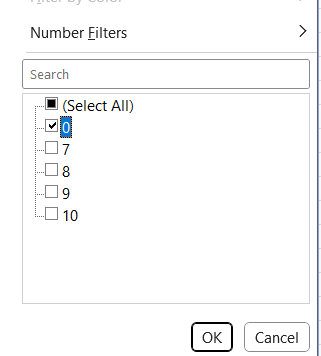
* **Microsoft Excel (2022)** - Used for data cleaning, analysis, and visualization.
* **Word** - Used for report generation and presentation.

**3. Data Cleaning & Preprocessing**

**Steps Performed:**

* Removed unnecessary columns like **color,director\_facebook\_likes,** **actor\_3\_facebook\_likes,** **actor\_2\_name,** **actor\_3\_name,** **actor\_1\_facebook\_likes,** **actor\_1\_name,** **actor\_3\_facebook\_likes ,plot\_keywords,** **content\_rating,** **aspect\_ratio,** **movie\_facebook\_likes**
* Removed **duplicates** to ensure data integrity.
* Removed blank rows by using ‘=COUNTA()’



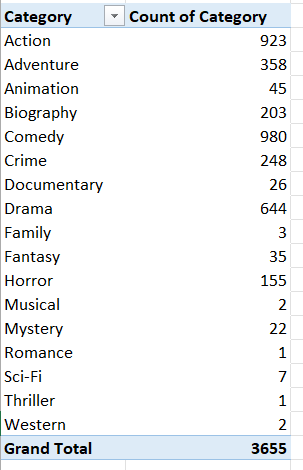


**4. Data Analysis & Insights**

**A. Movie Genre Analysis**

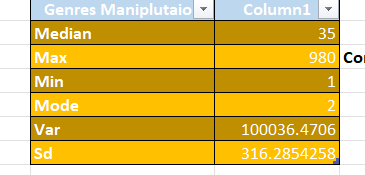
📌 **Goal:** Understand the impact of genres on IMDB ratings.

1. Split the 'genres' column to separate multiple genres for a single movie.
2. Used **Excel's COUNTIF function** to count the number of movies for each genre.
   * Formula: =COUNTIF(range, criteria)

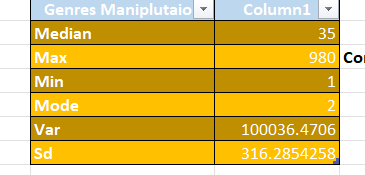


3.Calculated **descriptive statistics** for each genre:

* + **Mean**: =AVERAGE(range)
  + **Median**: =MEDIAN(range)
  + **Mode**: =MODE(range)
  + **Max & Min**: =MAX(range), =MIN(range)
  + **Variance**: =VAR.P(range)
  + **Standard Deviation**: =STDEV.P(range)



1. Compared the calculated statistics to analyze the impact of genre on movie ratings.



**Insights:**

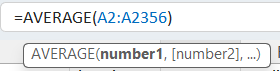
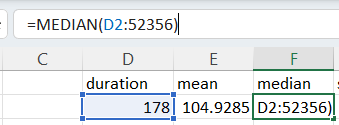
* the most common genres in the dataset are Action, Biography, Crime, Comedy
* Drama emerged as the most prevalent genres

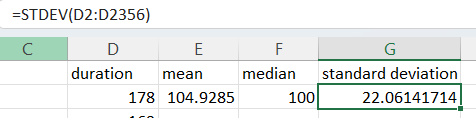
**B. Movie Duration Analysis**

📌 **Goal:** Determine if movie length impacts IMDB ratings.

**Steps Performed:**

1. Created a **scatter plot** of IMDB score vs. movie duration.
2. Calculated **mean, median, and standard deviation** for movie durations.



**Insights:**

* Movies **between 75-150 minutes** tend to have **higher ratings**.
* **Too short (<60 mins) or too long (>180 mins) movies** have **mixed ratings**.
* With the help of trend line we can

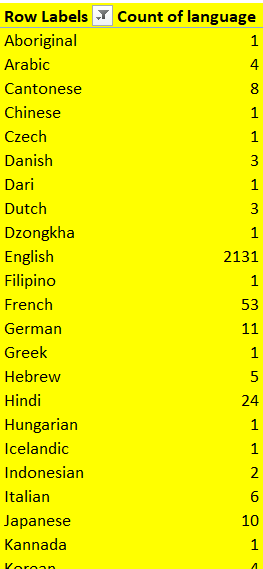
also predict **IMDB score can increate with respect to duration of the movi.**

**C. Language Analysis**

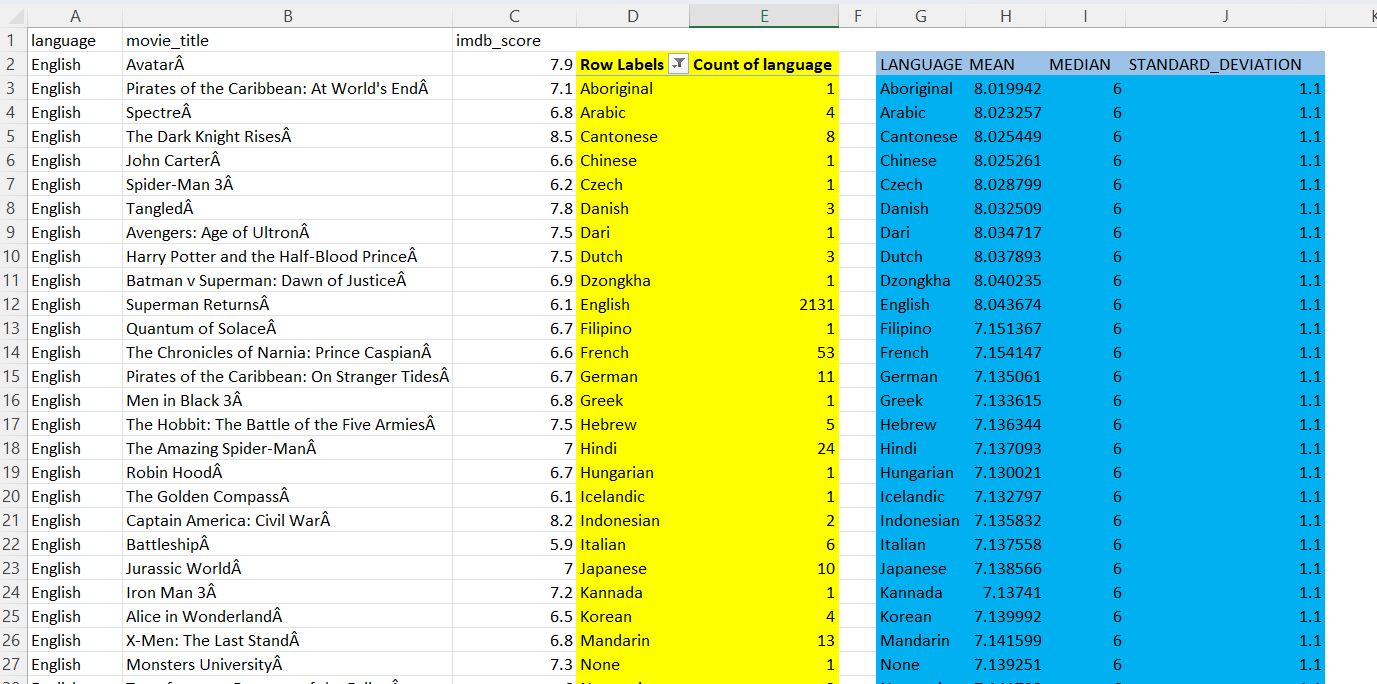
📌 **Goal:** Examine the most common languages used in movies and their impact on IMDB scores.

**Steps Performed:**

1. Counted movies by **language** Using pivote table.



1. Calculated **mean, median, and standard deviation** of IMDB ratings per language.



**Insights:**

* **English** is the dominant language followed by French.
* **French & Italian movies** have **higher average ratings**.
* **Multi-language movies** tend to have better audience reception.

NOTE: Interestingly, movies in languages other than

English tended to have slightly lower IMDB scores on average, indicating

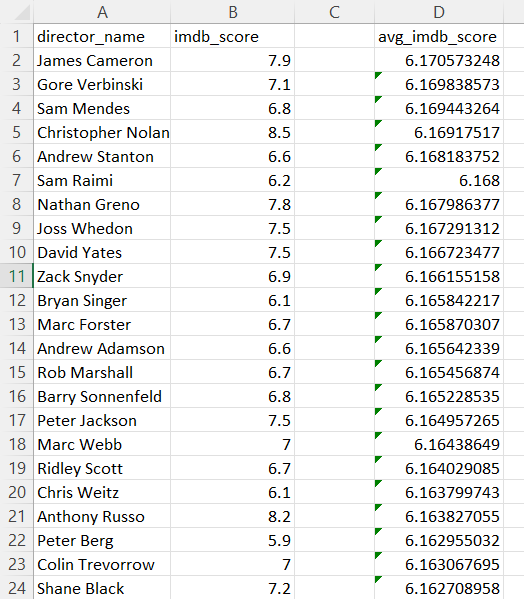
potential biases in audience preferences.

**D. Director Analysis**

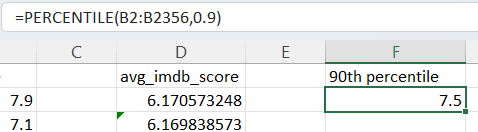
📌 **Goal:** Identify the top directors based on their IMDB ratings.

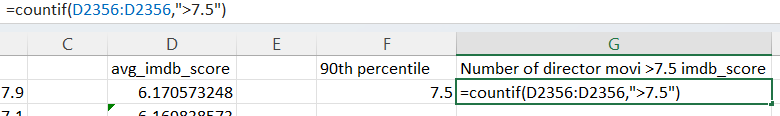
**Steps Performed:**

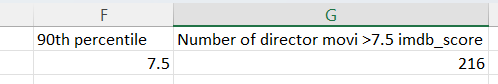
1. Calculated the **average IMDB rating** per director.



1. Used **percentile analysis** to determine top directors.







1. Compared top directors to overall rating distribution.

**Insights:**

**Top Directors:** Frank Darabont, Francis Ford Coppola, Francis Ford Coppola

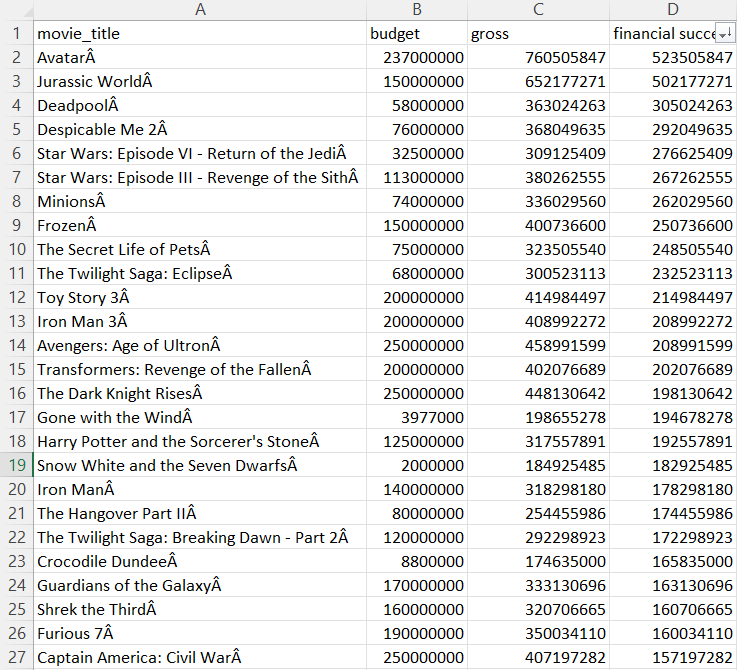
* **Directors with consistent high ratings tend to have a unique filmmaking style**.
* Directors such as **Francis Ford,Coppola and Frank Darabont consistently produced movies with high IMDB ratings,** reflecting their strong influence on audience perception

**E. Budget Analysis**

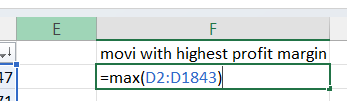
📌 **Goal:** Explore the relationship between movie budget and financial success.

**Steps Performed:**

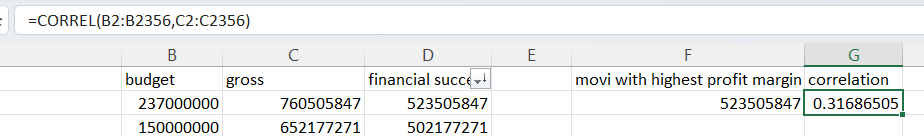
1. Calculated **Financial success by substrating** **gross earnings**.from **budget**



2.Identified movies with **highest profit margin**.



3.Calculated **correlation coefficient** between **budget and gross earnings**.



**Insights:**

* **Weak correlation (0.31)** → **High budget ≠ guaranteed success**.
* **Top profitable movies:** Avatar, Titanic, Star Wars, E.T., The Lion King.
* **Franchises & Disney movies** have **consistent profitability**.

**5. Result & Key Findings**

* **Drama & Biography movies** tend to get **higher IMDB ratings**.
* **Optimal movie duration**: 75-150 mins.
* **English dominates**, but **French & Italian films** have strong ratings.
* **Top-rated directors** have a **signature style** that contributes to success.
* **Budget alone does not determine success**; storytelling and audience engagement matter more.

**6. Conclusion & Recommendations**

🎬 **For Producers & Directors:**

* Focus on **engaging storytelling over high budgets**.
* **Biography & Drama genres** have strong audience appeal.
* **Experienced directors** with a distinct vision contribute to success.

💰 **For Investors:**

* High budgets **do not guarantee** higher earnings.
* **Established franchises & Disney movies** tend to yield better profits.
* Consider **directors with consistently high ratings**

**Excel work Link:**

[IMDB\_Movies.xlsx](https://docs.google.com/spreadsheets/d/1PLoE9c-pJVhe9ozGIMguMEH0w4e9SSnE/edit?usp=drive_link&ouid=102667770131422426899&rtpof=true&sd=true)

**Loom Presentation** [**link**](https://drive.google.com/file/d/1B4TrxUestANEMkcFB_iuSk_NbHUkz6Ly/view?usp=sharing)

**Prepared By : soham Kathalay.**