

JULY 2024

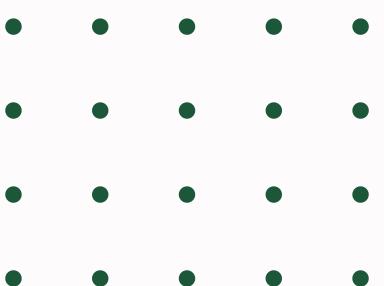
MODULE 5

VIDEO SUBMISSION

IMDB MOVIES ANALYSIS

GUNISHA CHOPRA

DATA ANALYST



ABOUT ME

My name is Gunisha Chopra. I am a data analyst trainee at TRAINITY. I am a graduate. I got my bachelor's in economics and mathematics from Delhi University which focuses on statistical and data analysis using several statistical programming tools like R, SQL, EXCEL etc. Current I'm exploring my interest in data science and analytics on my way to specialize in statistical and business analytics.

OTHER SKILLS:

- PYTHON
- TABLEAU
- MYSQL



Content

- 01** PROJECT DESCRIPTION
- 02** APPROACH
- 03** TECH-STACK USED
- 04** ANALYSIS
- 05** INSIGHTS AND TASKS
- 06** CONCLUSION
- 07** FURTHER ANALYSIS



• • • • •

PROJECT DESCRIPTION

In this project we are required to access the data from IMDB and generate insights on the questions asked by the team. The process is:

1

What's happening

2

Filter the data

3

Identifying the problem

4

Generate Insights



TECH-STACK USED



MICROSOFT EXCEL

We used Microsoft Excel to read and convert the dataframe according to our needs



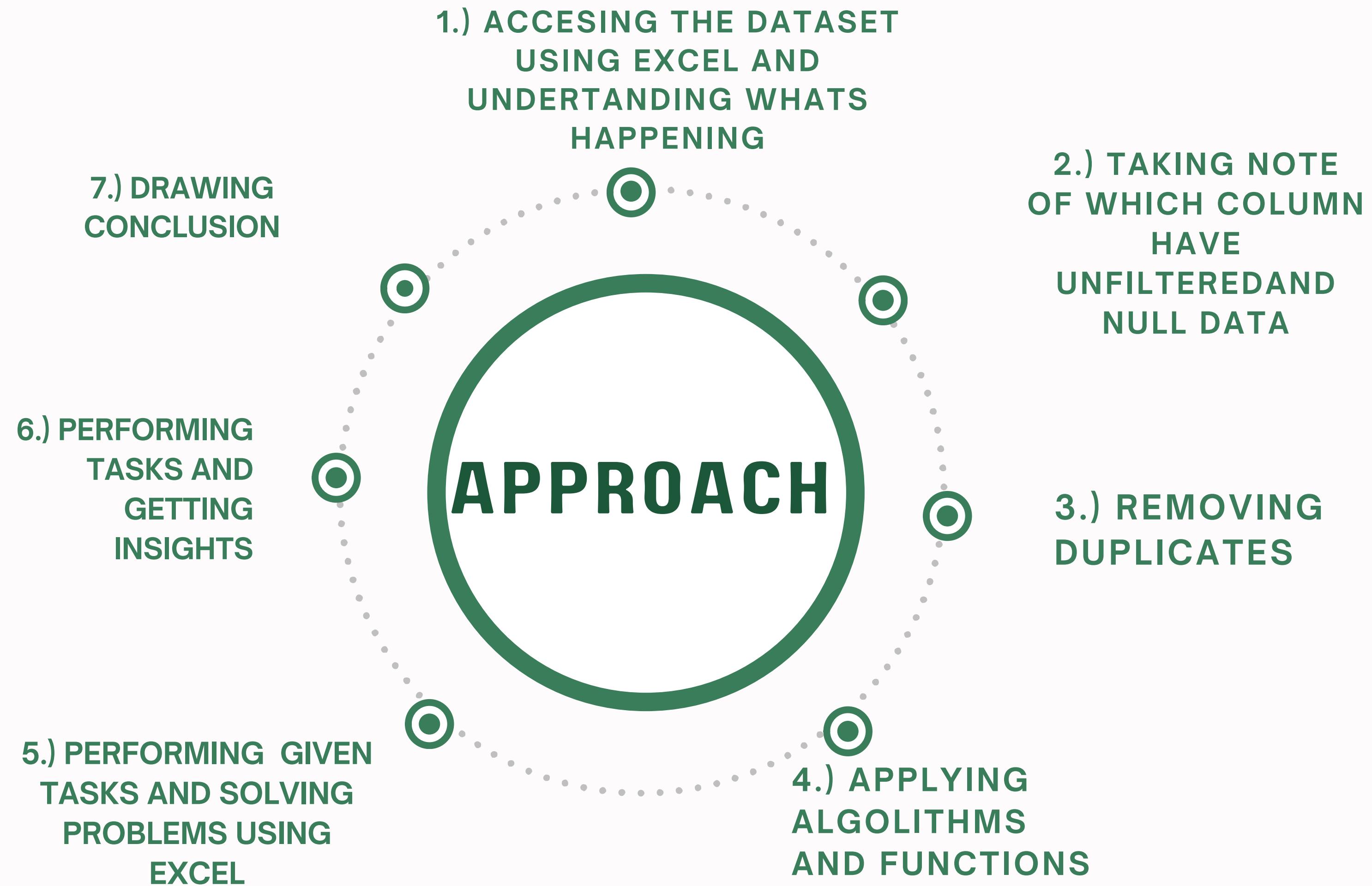
CANVA

Used to make the report and charts to represent the project report clearly.



MICROSOFT WORD

We use Microsoft word to write a report to submit to the team.

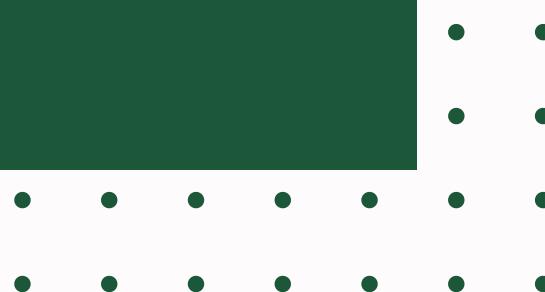


FIVE WHYS APPROACH

Once we have the problem better defined, we can use the 5 Whys technique to determine its root cause by repeatedly asking “Why”.

It is also known as the Root Cause Analysis was developed by Sakichi Toyoda, founder of Toyota Industries. Here's an example of how this technique could be used to figure out the cause of the following problem: A business went over budget on a recent project.

**ROOT
CAUSE
ANALYSIS**



TASKS

01

MOVIE GENRE
ANALYSIS

03

MOVIE DURATION
ANALYSIS

02

LANGUAGE
ANALYSIS

05

DIRECTOR
ANALYSIS

04

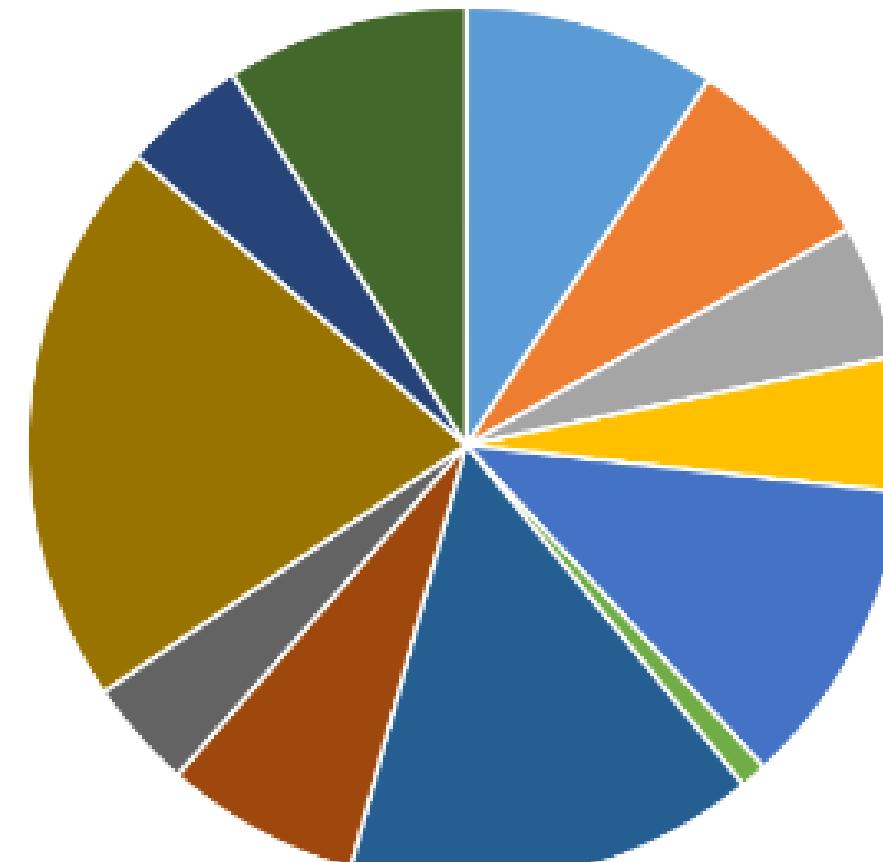
BUDGET ANALYSIS

MOVIE GENRE ANALYSIS

Total	Genres	Genres_Name	Mean	Mode	median	variance	stdev	Max	Min
951	Action		6.29	6.1	6.3	1.064	1.031626	9	2.1

Popular Genres :

Genres



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.

- The analysis shows the frequency of each genre in the dataset.
- There are 951 action movies making it the most common genre.

FORMULAS USED:

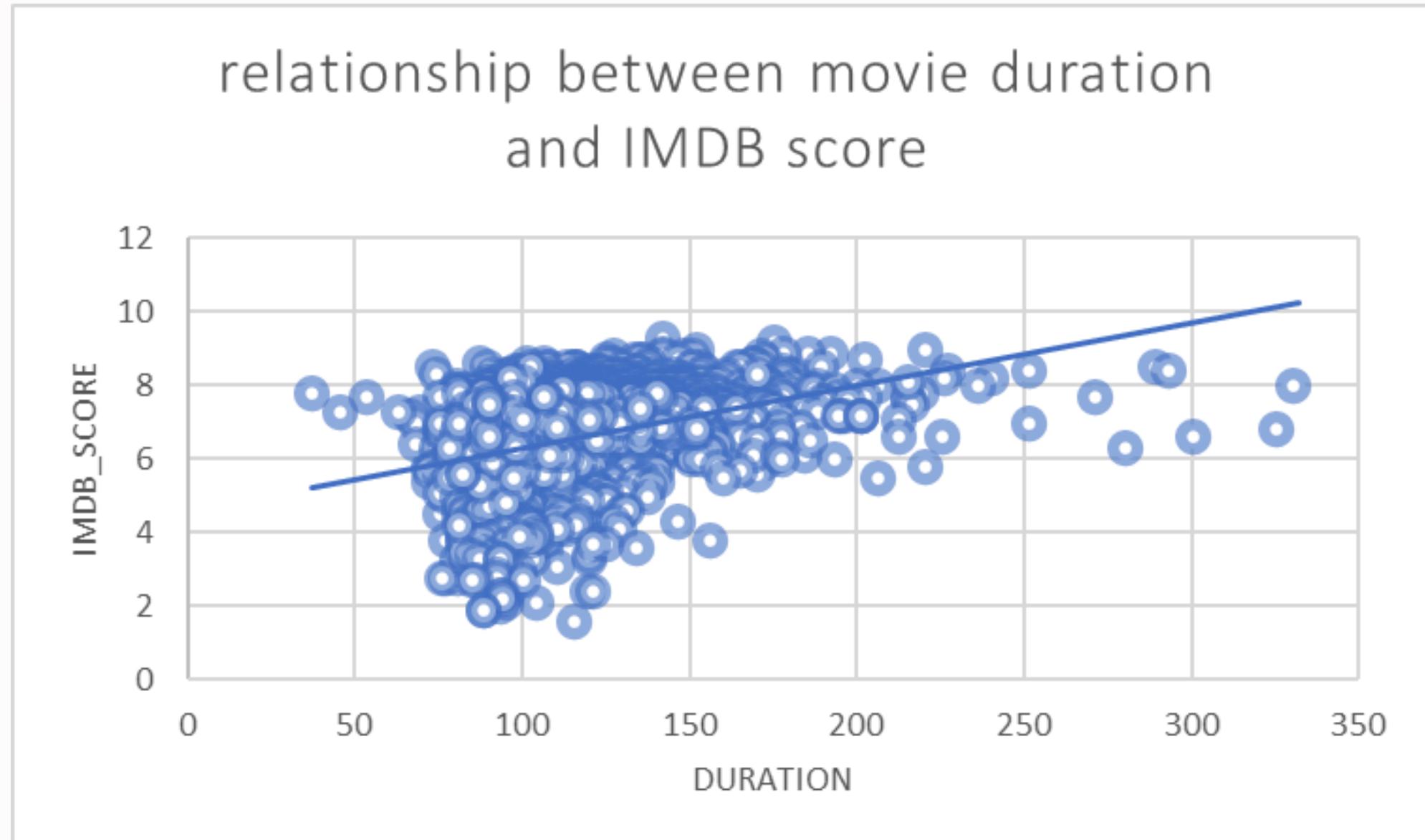
- COUNTIF
- VAR.P
- AVERAGEIF
- STDEV.P
- MODE
- MAX
- MEDIAN
- MIN

Action	Adventure	Fantasy	Sci_Fi
Thriller	Documentary	Comedy	Crime
Mystery	Drama	Horror	Romance

MOVIE DURATION ANALYSIS

Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

- The scatter plot shows a positive correlation between movie duration (x-axis) and IMDB score (y-axis).
- As movie duration increases, there seems to be a tendency for IMDB scores to also rise.
- Longer movies tend to receive higher IMDB ratings, which suggests that audiences appreciate well-crafted, immersive storytelling in lengthier films.



LANGUAGE ANALYSIS

Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

ANALYSIS:

NUMBER OF MOVIES: English is the most common language with 3566 movies, followed by French with 34 movies.

AVERAGE IMDb SCORE: The mean IMDB score for movies in each language indicates the overall rating for movies in that language.

STANDARD DEVIATION: The standard deviation of IMDB scores for movies in each language indicates the variability of ratings. A higher standard deviation suggests more variability in ratings,

IMPACT OF LANGUAGE ON IMDb SCORE: By comparing the average IMDB scores and standard deviations across languages, you can analyse the impact of language on the IMDB score.

DISTRIBUTION OF RATINGS: The median IMDB score for each language indicates the central tendency of ratings.

DIRECTOR ANALYSIS

Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

TOP 10 DIRECTORS:	Top_director_Name	Average of imdb_score	Percentile
• • • • • • •	Akira Kurosawa	8.7	8.1
• • • • • • •	Alfred Hitchcock	8.5	8.1
• • • • • • •	Asghar Farhadi	8.4	8.1066666667
• • • • • • •	Billy Wilder	8.3	8.137222222
• • • • • • •	Charles Chaplin	8.6	8.1633333333
• • • • • • •	Christopher Nolan	8.425	8.2
• • • • • • •	Damien Chazelle	8.5	8.2
• • • • • • •	David Sington	8.1	8.2
• • • • • • •	Elia Kazan	8.2	8.2
• • • • • • •	Fritz Lang	8.3	8.2108333333

BUDGET ANALYSIS

fx	=INDEX(C2:C3724, MATCH(MAX(D2:D3724), D2:D3724, 0)) & " - Profit: " & MAX(D2:D3724)
	E F G H I J
Corelation_Coefficient	Highest Profit Margin

The screenshot shows a portion of a Microsoft Excel spreadsheet. The formula bar at the top contains the formula =INDEX(C2:C3724, MATCH(MAX(D2:D3724), D2:D3724, 0)) & " - Profit: " & MAX(D2:D3724). Below the formula bar is a table with two columns. The first column has headers 'Corelation_Coefficient' and 'Highest Profit Margin'. The second column contains the value 'Avatar - Profit: 523505847'. The cell containing 'Avatar - Profit: 523505847' is highlighted with a yellow background.

In this case, the positive correlation coefficient suggests a weak positive relationship. As budgets increase, gross earnings tend to increase slightly, but not significantly.

Highest Profit Margin Movie: “Avatar”:

- The movie “Avatar” achieved the highest profit margin. Its profit is **five hundred twenty-three million five hundred five thousand eight hundred forty seven** (523,505,847).
- The profit margin is calculated as follows:

Profit Margin = (Gross Earnings - Budget)

The high-profit margin indicates that “Avatar” was highly successful financially.

CONCLUSIONS



OVERALL ANALYSIS

- **English Dominance**: English is the most common language used in movies, with a significantly higher number of movies compared to other languages. This indicates the dominance of English-language cinema in the dataset.
- **IMDB Scores**: The average IMDB scores vary across different languages, with some languages like Danish and Norwegian having relatively high average scores, while others like Russian and Hungarian have lower average scores.
- **Variability in Ratings**: There is variability in IMDB scores within each language, as indicated by the standard deviations. Languages with higher standard deviations, such as Portuguese and Thai, have more variability in ratings compared to languages with lower standard deviations.
- **Impact of Language**: The language of a movie appears to have an impact on its IMDB score, as evidenced by the differences in average scores across languages. However, other factors such as the quality of the movie, its storyline, and the performance of the actors may also play a significant role in determining the IMDB score.



THANK YOU

GUNISHA CHOPRA

+91 9718628433

Cgunisha@gmail.com

[linkedin.com/in/gunisha-chopra-b16768262](https://www.linkedin.com/in/gunisha-chopra-b16768262)

New Delhi, India

