IMDB MOVIE ANALYSIS

REPORT

By Shivaani Dushyanth
shivzlantern09@gmail.com
Software Used: Microsoft Excel

PROBLEM STATEMENT:

The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

DATA CLEANING

Since the table has many unnecessary columns, empty values, null values, and duplicates, we use the Go To Special function to remove all the rows which are empty.

Then we remove all the duplicates from the table.

Finally, we drop all the unnecessary columns from the table.

https://docs.google.com/spreadsheets/d/1gWgHDLMc6vXZlp-jrnbMHfS3jP5CCmgn/edit?usp=drive_link&ouid=104301423844572907298&rtpof=true&sd=

MOVIE GENRE ANALYSIS

Analyze the distribution of movie genres and their impact on the IMDB score.

Process:

Use PivotTable to have Genres as Row Labels and Values as IMDB scores.
Manipulate imdb scores to get average,median,sum,mode,variance and standard deviation

Insight:

The most common movie Genre is Drama with an average IMDB score of 7.08.

Row Labels	Var of imdb_score3	StdDev of imdb_score	Sum of imdb_score2	Average of imdb_score5	Max of imdb_score4	Min of imdb_score3	Count of imdb_score
Drama	0.687754813	0.829309841	998.9	7.084397163	8.8	3.4	141
Comedy|Drama|Romance	0.572416364	0.756582027	953.5	6.486394558	8	4.3	147
Comedy|Drama	0.763474558	0.873770312	906.5	6.56884058	8.8	3.3	138
Comedy	1.502079763	1.225593637	807.2	5.849275362	8	1.9	138
Drama|Romance	0.54440122	0.737835497	802	6.973913043	8.1	4.1	115
Comedy|Romance	0.699147387	0.836150338	777.9	5.938167939	8.4	2.7	131
Crime|Drama|Thriller	0.612311051	0.782503068	562.5	6.859756098	8.5	5.1	82
Action|Crime|Thriller	0.39361039	0.627383766	359.2	6.414285714	7.6	4.4	56

MOVIE DURATION ANALYSIS

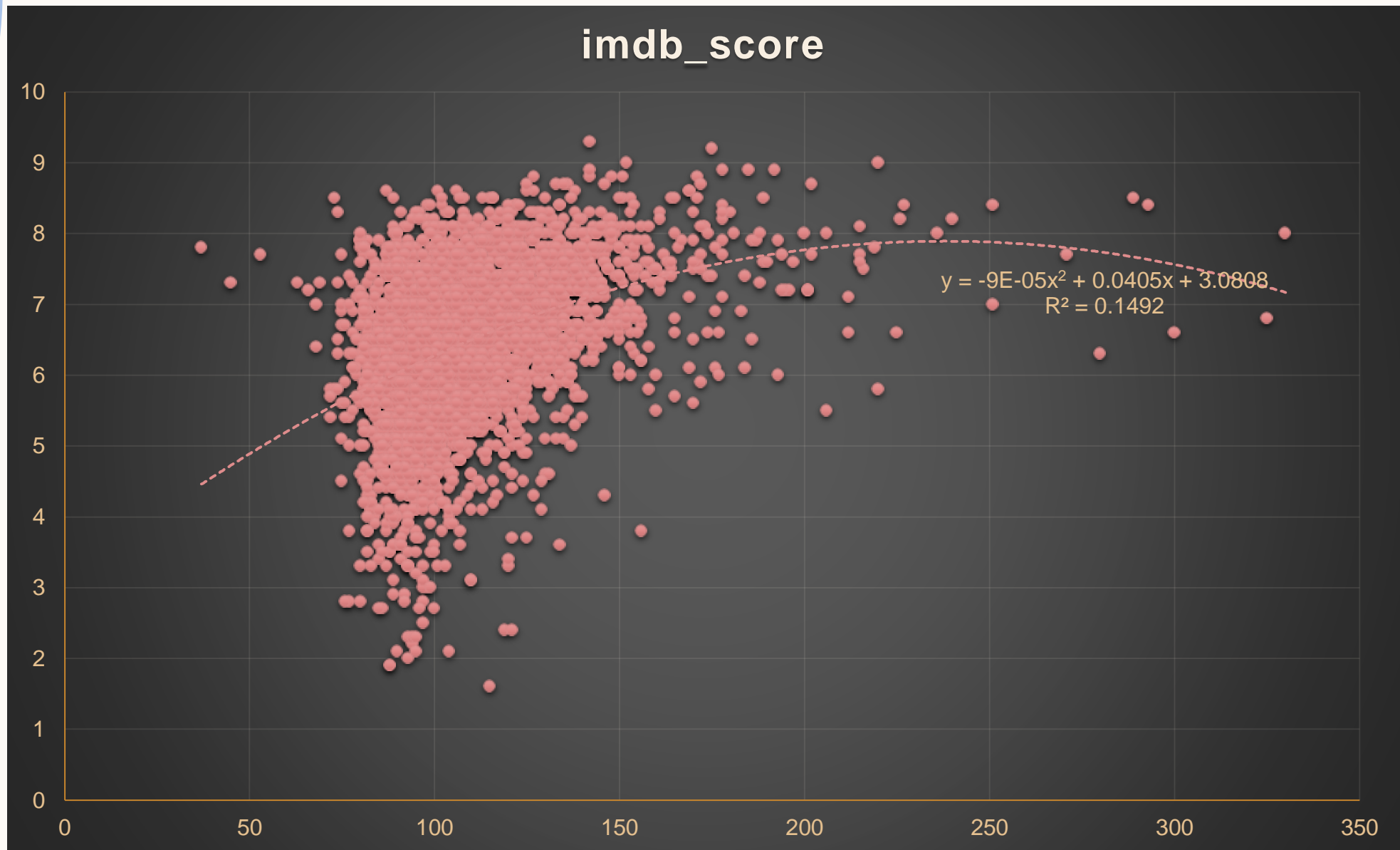Analyze the distribution of movie durations and its impact on the IMDB score.

Process:

Use MEDIAN(D2:D3757)
AVERAGE(D2:D3757)
MODE(D2:D3757)
VAR.S(D2:D3757)
STDEV.S(D2:D3757)

Insight:

Maximum number of movies have an IMDB score between 5 to 8 with duration between 90 to 120
The polynomial trendline shows the quadratic equation between the duration and imdb score.

imdb_score

$y = -9E\text{-}05x^2 + 0.0405x + 3.0808$
$R^2 = 0.1492$

Duration				
Average	Median	Mode	Variance	Std.Dev
110.258	106	101	512.8738	22.64672

LANGUAGE ANALYSIS

Examine the distribution of movies based on their language.

Process:

Use a PivotTable to get Language as Row Labels and Movie Titles as count in the values column.
Make the piechart of the same
Then use =AVERAGE(IF($T$2:$T$3757=$BA21,$Z$2:$Z$3757))
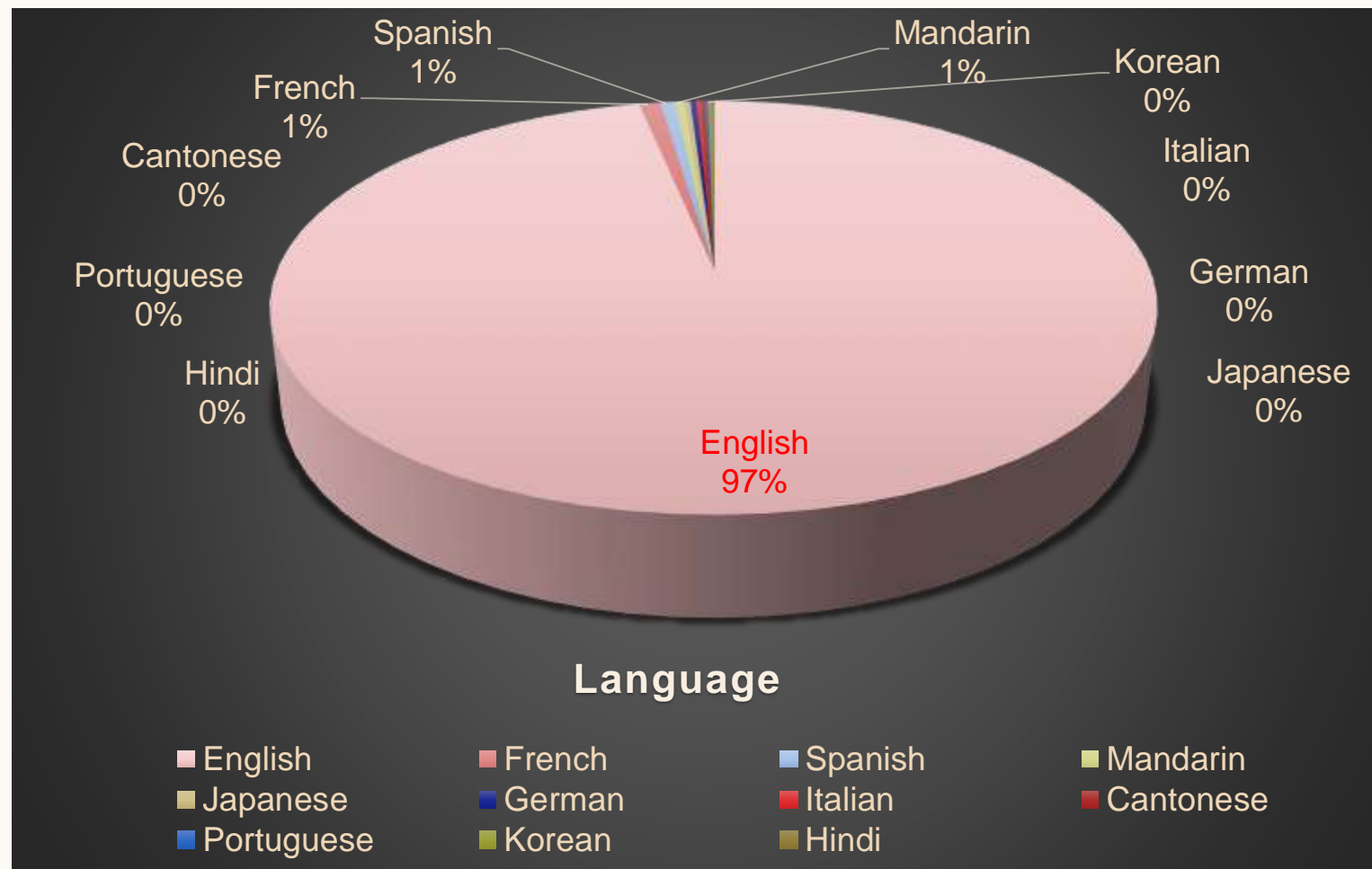=MEDIAN(IF($T$2:$T$3757=$BA21,$Z$2:$Z$3757))
=MODE(IF($T$2:$T$3757=$BA21,$Z$2:$Z$3757))
=STDEV.S(IF($T$2:$T$3757=$BA22,$Z$2:$Z$3757))
=VAR.S(IF($T$2:$T$3757=$BA21,$Z$2:$Z$3757))

Insight:

97% of Language used in movies is English followed by a 1% usage of French,Spanish and Mandarin

Language

IMDB	Row Labels	Count of movie_title	Average	Median	Mode	Std.Dev	Variance
23124.5	English	3598	6.427043	6.5	6.7	1.050538	1.10363
250.1	French	34	7.355882	7.3	7.2	0.519435	0.269813
162.9	Spanish	23	7.082609	7.2	5.9	0.860577	0.740593
106.2	Mandarin	15	7.08	7.4	7.9	0.77201	0.596
76.6	Japanese	10	7.66	8	#N/A	0.990174	0.980444

DIRECTOR ANALYSIS

Influence of directors on movie ratings

Process:

Use a Pivot Table to get the Directors names as Row Labels and their IMDB scores as Values.
Sort By Descending order
Create a chart for the following data

Insight:

Akira Kurosawa,Tony Kaye and Charles Chaplin are the top 3 directors with an IMDB score of 8.7,8.6 and 8.6 respectively.

Row Labels ⏷↓	Average of imdb_score5
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Alfred Hitchcock	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Richard Marquand	8.4

BUDGET ANALYSIS

Explore the relationship between movie budgets and their financial success.

Process:

Find the difference between the Gross and the Budget Columns as Gross Profit
Create a chart for Budget Analysis and Gross Profit Analysis
Using the CORREL(range1,range2)
range1-Gross and range2-Budget find the correlation coefficient

Insight:

Highest gross Profit is 523505847 for the Movie AvatarA.
The correlation coefficient brtween gross nd budgest is 0.09950.

Correlation
0.099496423

Row Labels	Max of gross	Max of budget	Profit
AvatarÂ	760505847	237000000	5.24E+08
TitanicÂ	658672302	200000000	4.59E+08
Jurassic WorldÂ	652177271	150000000	5.02E+08
The AvengersÂ	623279547	220000000	4.03E+08
The Dark KnightÂ	533316061	185000000	3.48E+08
Star Wars: Episode I - The Phantom MenaceÂ	474544677	115000000	3.6E+08
Star Wars: Episode IV - A New HopeÂ	460935665	11000000	4.5E+08
Avengers: Age of UltronÂ	458991599	250000000	2.09E+08
The Dark Knight RisesÂ	448130642	250000000	1.98E+08
Shrek 2Â	436471036	150000000	2.86E+08

Chart Area
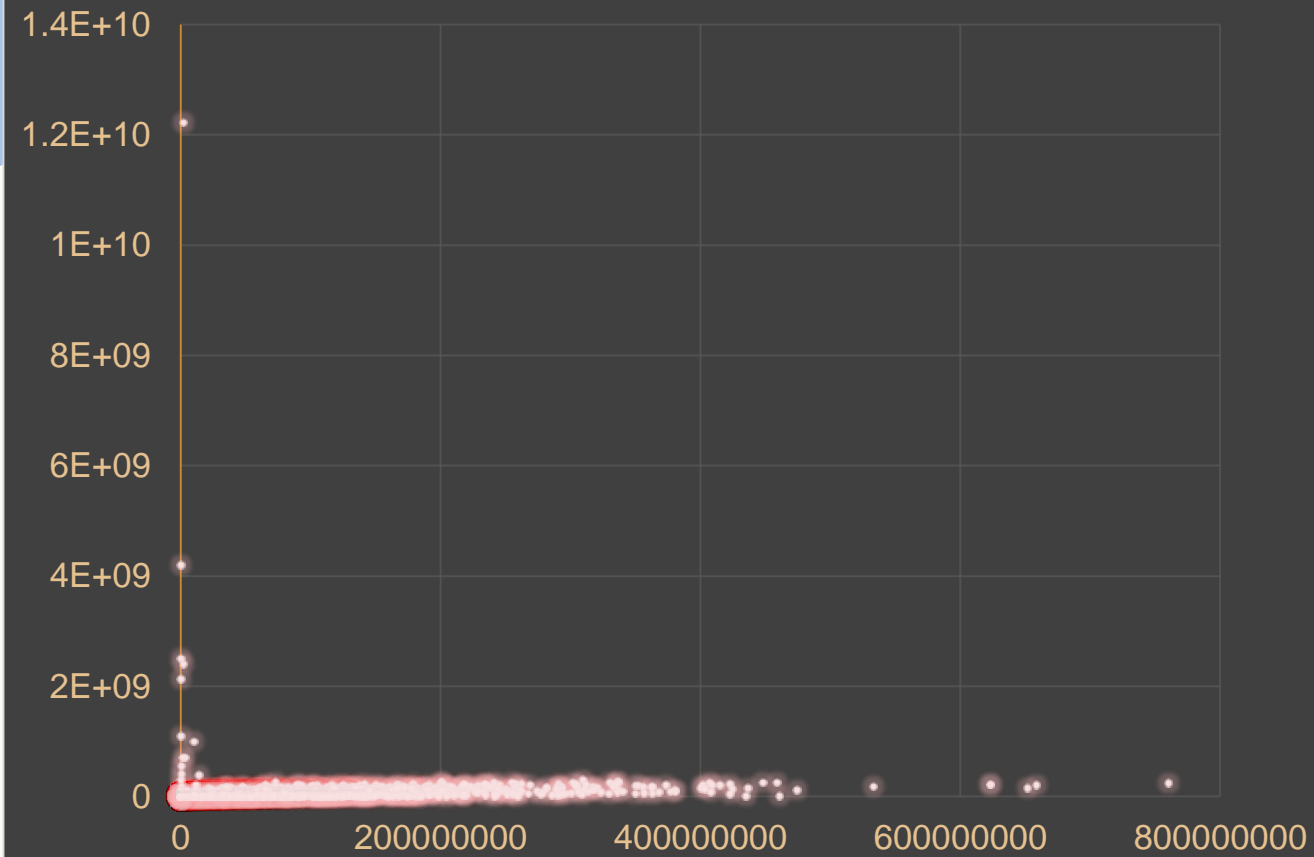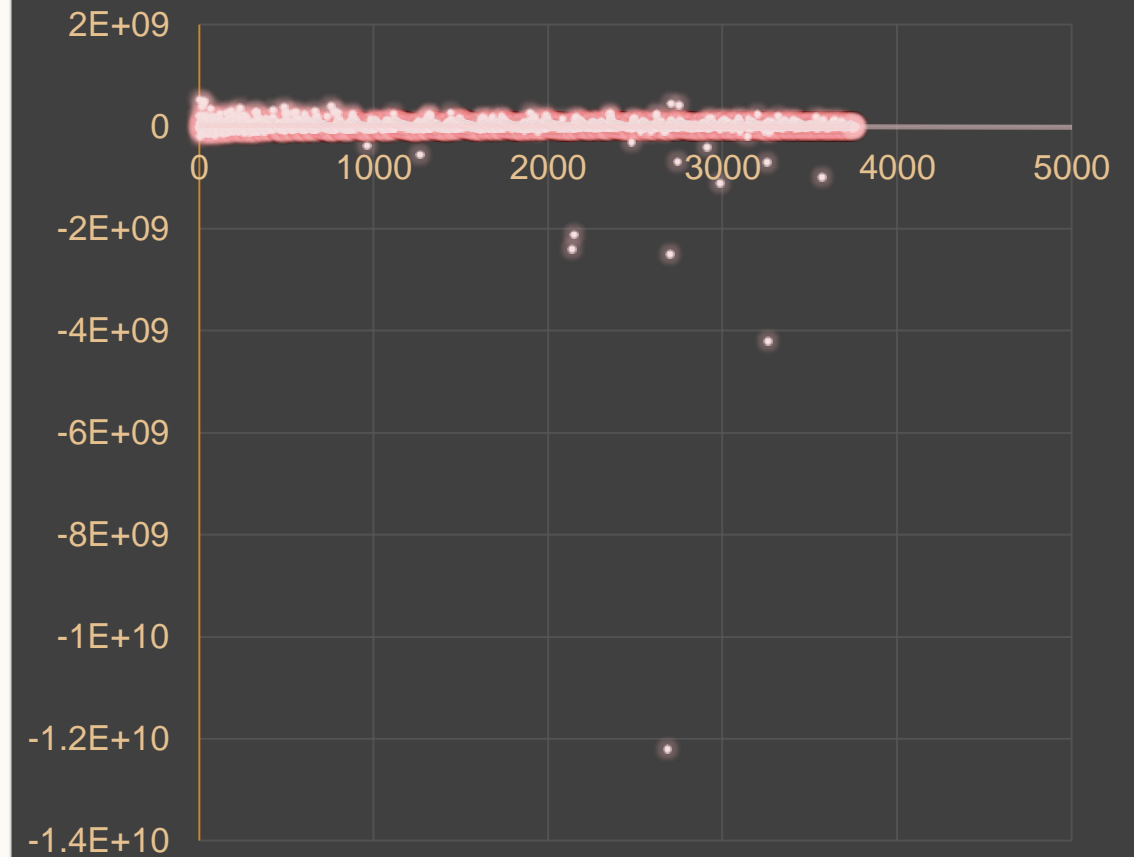
BEST MOVIE

Shawshank Redemption

IMDB:9.3

TOP COUNTRY

USA

STREAMED: 2987 times

PEOPLE'S DIRECTOR

Christopher Nolan

NUmber of voters: 1676169

BEST ACTOR2

Morgan Freeman

Facebook Likes:
220000

BEST ACTOR3

Anne Hathway

Facebook Likes:
77000

BEST ACTOR1

Johnny Depp

Facebook
likes:1560000
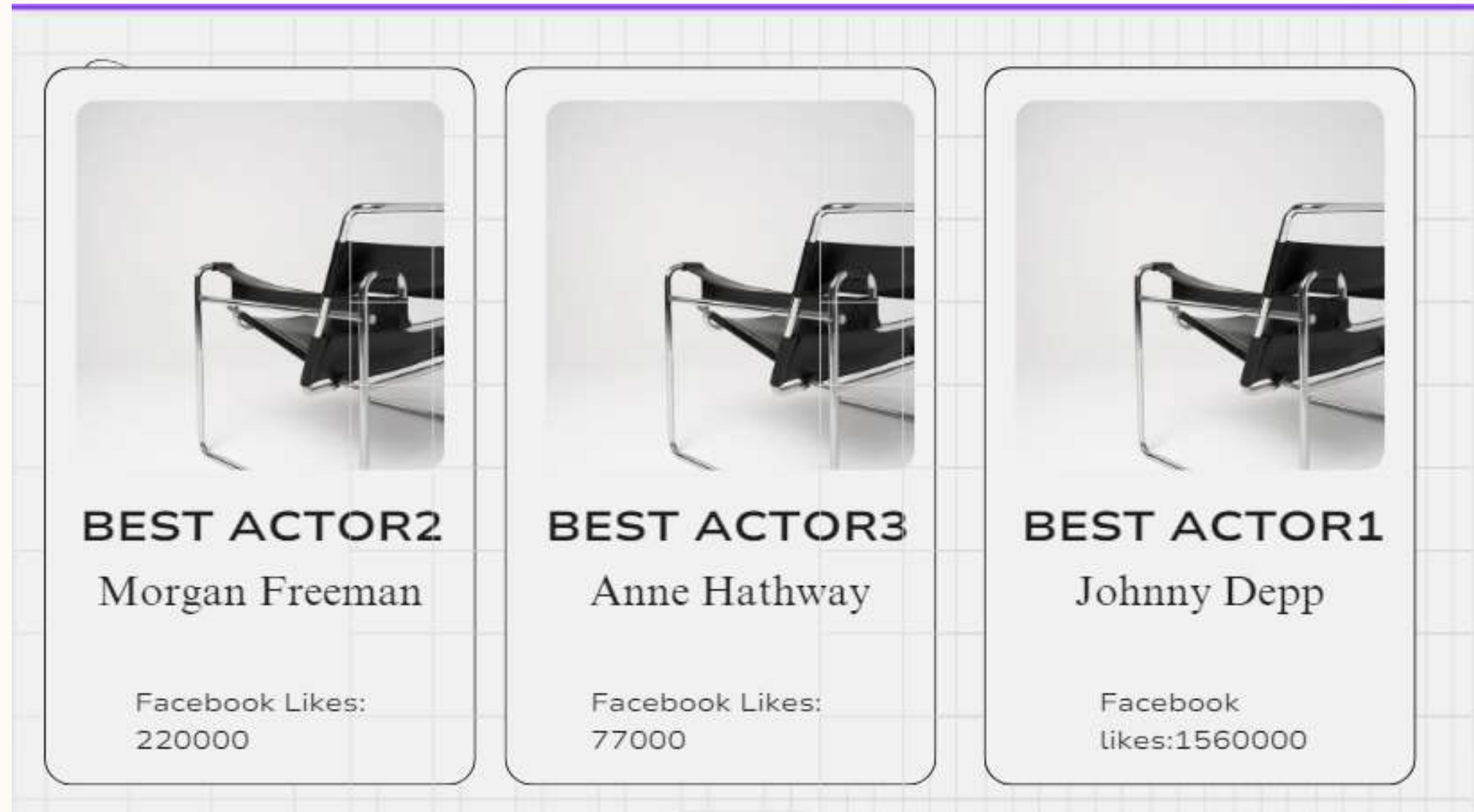
This project has helped me understand how IMDB filters and produces data to customers such that the reviewing process becomes easier and thus helps making decisions faster.

THANK
YOU