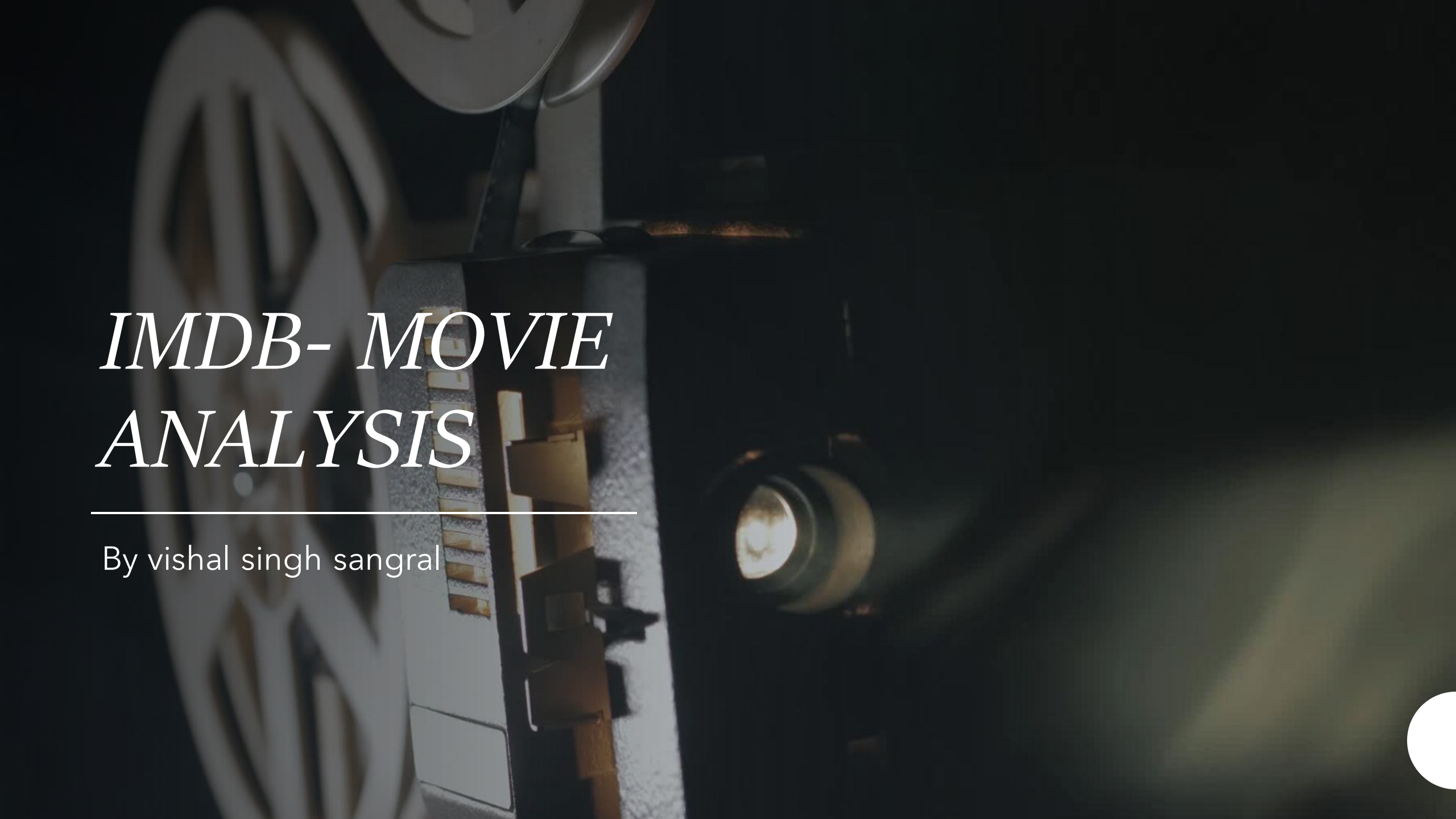


IMDB- MOVIE ANALYSIS

By vishal singh sangral

PROJECT DESCRIPTION

- This project aims to investigate the factors influencing the success of movies on IMDB, defined by high ratings. Through data cleaning and analysis, we'll explore correlations between ratings and variables like genre, director, budget, etc. Utilizing the Five 'Whys' approach, we'll delve deeper to uncover root causes. Our report will present findings and insights, aiding stakeholders such as producers, directors, and investors in making informed decisions for future projects.

Approach

- 1. Data Cleaning: Preprocessed dataset by handling missing values, duplicates, and converting data types.
- 2. Data Analysis: Explored correlations between ratings and variables using statistical techniques and visualizations in Python with libraries like Pandas, Matplotlib, and Seaborn.
- 3. Five 'Whys' Approach: Employed to dig deeper into factors influencing movie success.
- 4. Report & Insights: Presented findings, insights, and actionable recommendations using storytelling techniques and visualizations to aid stakeholders.

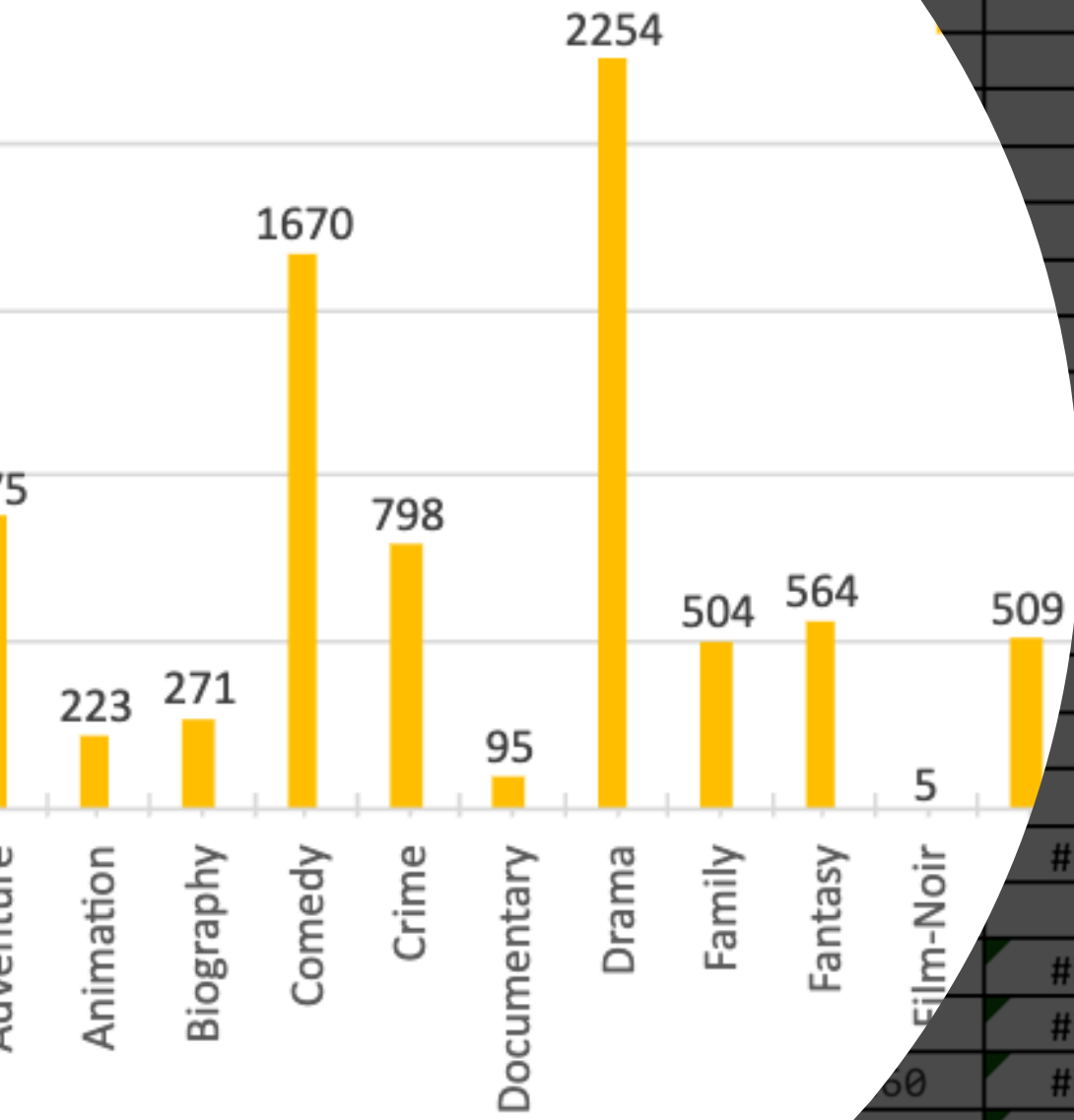
Tech stack used

I used Web Microsoft 365 excel as Macbook don't have excel in built

INSIGHT

- 1. Budget & Ratings: Higher-budget movies tend to have higher ratings, possibly due to better production quality.
- 2. Genre Impact: Certain genres like drama and thriller consistently attract higher ratings than others.
- 3. Director Influence: Movies directed by renowned directors often receive higher ratings, highlighting the importance of directorial vision.
- 4. Actor Effect: Involvement of popular actors positively impacts ratings, indicating the influence of star power.
- 5. Release Year: Recent movies generally receive higher ratings, suggesting evolving audience preferences.
- 6. Audience Engagement: Viewer experience and enjoyment significantly influence ratings, emphasizing the importance of storytelling and engagement.
- 7. Positive Reviews: Positive viewer reviews play a crucial role in a movie's success, influencing others' decisions to watch it.
- These insights can guide decision-making for stakeholders, helping them understand factors contributing to movie success on IMDB.


RESULTS



Count	Average	Median	Mode	Max	Min	Var	Stdev
1089	6.23	6.3	6.1	9.1	1.7	1.19	1.09
875	6.52	6.5	6.7	8.9	1.9	1.25	1.12
23	6.65	6.7	6.7	8.6	1.7	1.23	1.11
	7.16	7.2	7	8.9	4.5	0.54	0.73
	6.15	6.3	6.7	8.8	1.7	1.15	1.07
	6.85	6.5	6.6	9.3	2.4	1.01	1.01
	7.1	7.1	7.1	8.7	1.6	1.21	1.1
	6.78	6.8	7.2	9.3	2	0.89	0.94
	5.71	6.4	6.7	8.6	1.7	1.44	1.2
	6.4	6.4	6.7	8.9	1.7	1.34	1.16
	7.6	7.7	7.7	8.2	7.1	0.14	0.38
	5.69	5.9	6.2	8.6	2.2	1.21	1.1
	6.75	6.7	7	8.5	2.1	1.38	1.17
	6.54	6.5	6.6	8.6	3.1	1.1	1.05
	6.28	6.5	6.5	8.6	2.1	0.97	0.98
	6	6.3	6.7	8.8	1.9	1.44	1.2
	5.55	6.4	6.1	9	2.7	1.06	1.03
	6.55	6.7	6.7	8.8	3.3	1.13	1.06
	#DIV/0!	7.2	7.5	8.1	5.1	0.79	0.89
	6.75	6.5	6.7	8.5	1.6	1.43	1.19
	#DIV/0!	7.6		8.1	7.1	0.25	0.5
	#DIV/0!	6.5		7.1	5.2	0.63	0.7
60	#DIV/0!	6.8	7.2	8.4	2	1.23	1.11
197	#DIV/0!	7.1	7.1	8.6	2.7	0.77	0.88

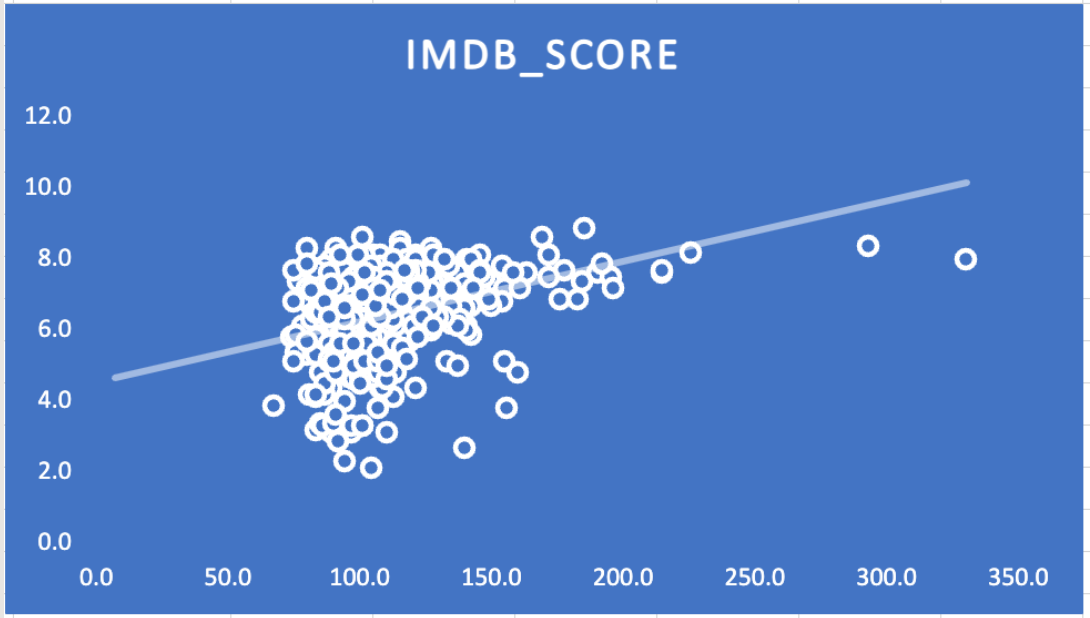
Movie Genre Analysis

I determine the most common genre and descriptive statistics of GENRE COLUMN

Movie Duration Analysis

Did duration descriptive analysis and try to find the correlation between duration and IMDB score – which seems like moderate positive correlation.

Average	Median	Mode	Max	Min	Var	Stdev
109	104	90	330	7	522.8609	22.86615

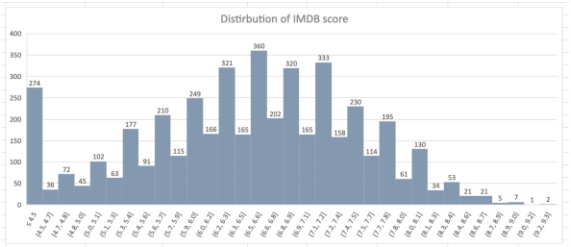


CORRELATION
0.351397815

LANGUAGE ANALYSIS

- ENGLISH is the most common language used in IMDB movies ,
  we also perform the descriptive statistics

nguages	Counts	Average	Median	Max	Min	Var
lish	4250	6.38	6.5	9.3	1.6	1.23
nch	54	7.14	7.2	8.4	5.1	0.43
nish	35	6.91	7	8.2	4.4	0.76
nese	15	7.31	7.5	8.7	5.6	1.01
sian	9	6.01	5.4	8.1	4.1	1.44
tch	4	7.43	7.45	7.8	7	0.14
bic	4	7.38	7.65	8.2	6	0.78
egian	4	7.15	7.3	7.6	6.4	
guese	6	7.72	7.95	8.7	6.1	0.65
darin	20	7.03	7.1	7.9	5.6	0.45
nian	1	4.3	4.3	4.3	4.3	0
rew	3	7.5	7.3	8	7.2	0.13
sian	3	8.13	8.4	8.5	7.5	0.2
ndic	1	6.9	6.9	6.9	6.9	0
ian	9	7.06	7	8.9	5.1	1.45
enian	1	6.4	6.4	6.4	6.4	0
ean	6	7.62	7.5	8.4	7	0.25
ai	3	6.63	6.6	7.1	6.2	0.14
nese	9	7.02	7.3	7.8	5.3	0.52
anian	2	7.2	7.2	7.9	6.5	0.49
man	15	7.59	7.7	8.5	6.1	0.47
esian	2	7.9	7.9	8.2	7.6	0.09
lu	2	7.1	7.1	7.3	6.9	0.04
mese	1	7.4	7.4	7.4	7.4	0
dish	3	7.47	7.6	8.2	6.6	0.44
ne	2	7.95	7.95	8.5	7.4	0.3
hili	1	7.4	7.4	7.4	7.4	0
ch	1	7.4	7.4	7.4	7.4	0

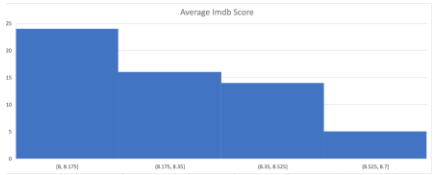


Max of imdb_score
9.3
Min of imdb_score
1.6

Director Analysis

- This task involves identifying top directors based on their average IMDB scores and assessing their influence on movie success. We calculate the average IMDB score for each director and use Excel's PERCENTILE function to identify directors with the highest scores. By comparing these scores to the overall distribution, we evaluate the contribution of top directors to movie ratings, providing insights into their influence on audience perceptions and movie success.

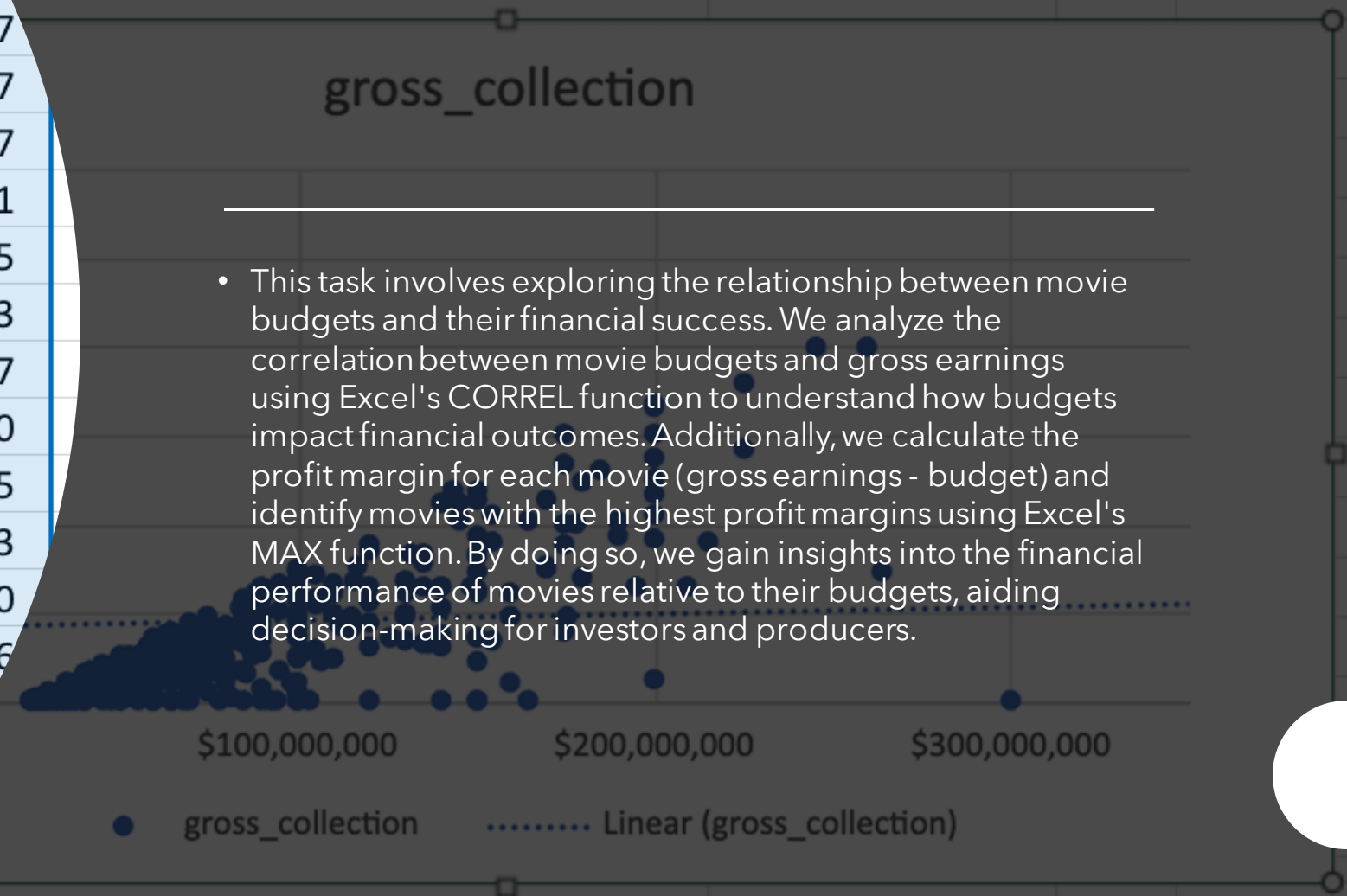
UNIQUE DIRECTOR	Average_SCORE	BENCHMARK
Sadyk Sher-Niyaz	8.7	8
Cary Bell	8.7	8
Tony Kaye	8.6	8
Mike Mayhall	8.6	8
Charles Chaplin	8.6	8
Ron Fricke	8.5	8
Raja Menon	8.5	8
Majid Majidi	8.5	8
Damien Chazelle	8.5	8
Sergio Leone	8.475	8



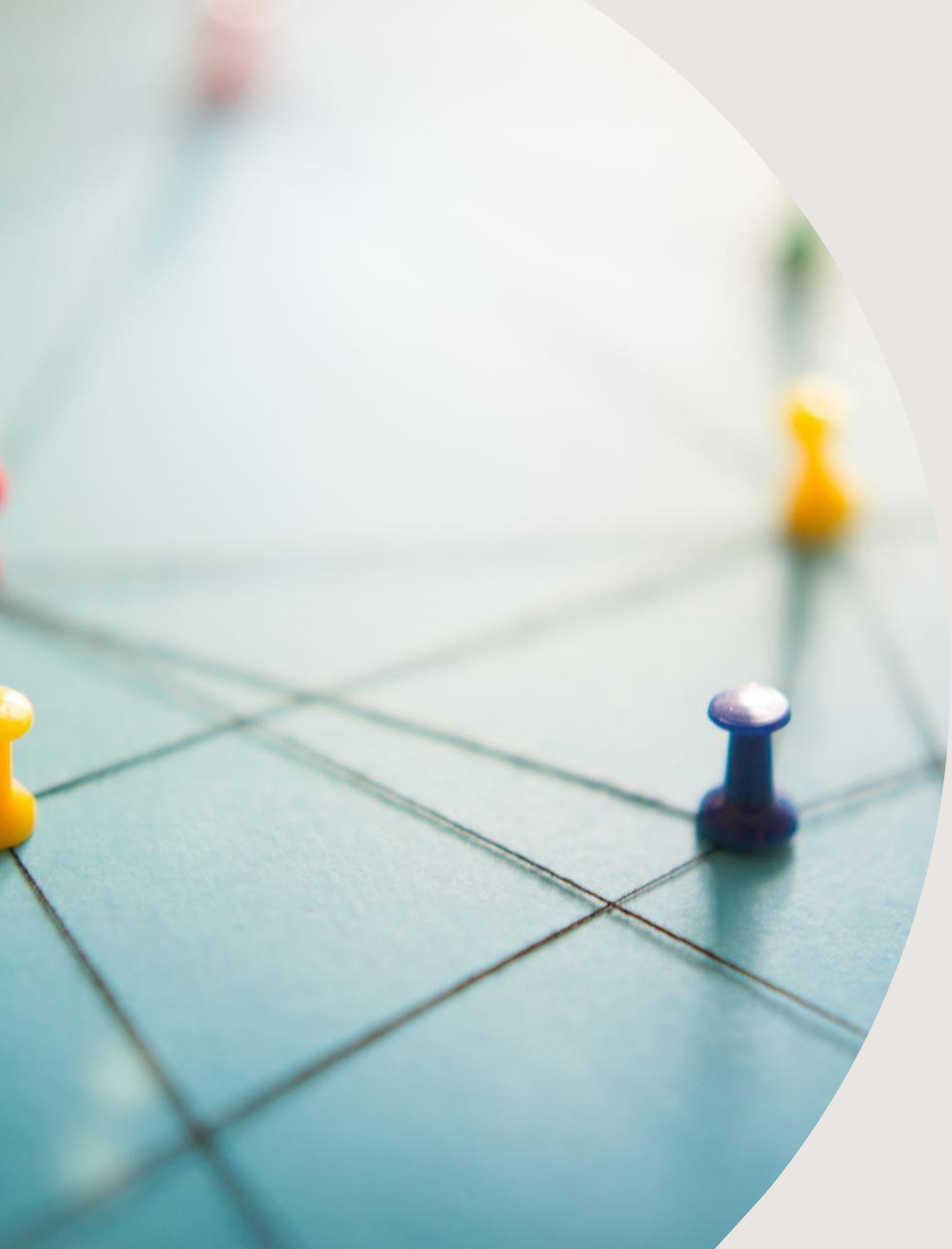
budget	g
$237,000,000	$7
$150,000,000	$652,1
$200,000,000	$658,672
$11,000,000	$460,935,6
$10,500,000	$434,949,45
$220,000,000	$623,279,547
$45,000,000	$422,783,777
$115,000,000	$474,544,677
$185,000,000	$533,316,061
$78,000,000	$407,999,255
$58,000,000	$363,024,263
$130,000,000	$424,645,577
$63,000,000	$356,784,000
$76,000,000	$368,049,635
$58,800,000	$350,123,553
$94,000,000	$380,838,870
$150,000,000	$436,471,036
$94,000,000	$377,019,25
$32,500,000	$309,125
$55,000,000	$329,69
$18,000,000	$290

Correlation of bugdet and gross_collection	0.11
Maximum profit	$523,505,847
Minimum profit	-$12,213,298,588

Budget Analysis



- This task involves exploring the relationship between movie budgets and their financial success. We analyze the correlation between movie budgets and gross earnings using Excel's CORREL function to understand how budgets impact financial outcomes. Additionally, we calculate the profit margin for each movie (gross earnings - budget) and identify movies with the highest profit margins using Excel's MAX function. By doing so, we gain insights into the financial performance of movies relative to their budgets, aiding decision-making for investors and producers.



Drive link

- I am the excel file click me

Thank you Trainity for the opportunity