EXPLORATORY DATA ANAYSIS REPORT ON THE MOVIE INDUSTRY

TABLE OF CONTENTS

BACKGROUND

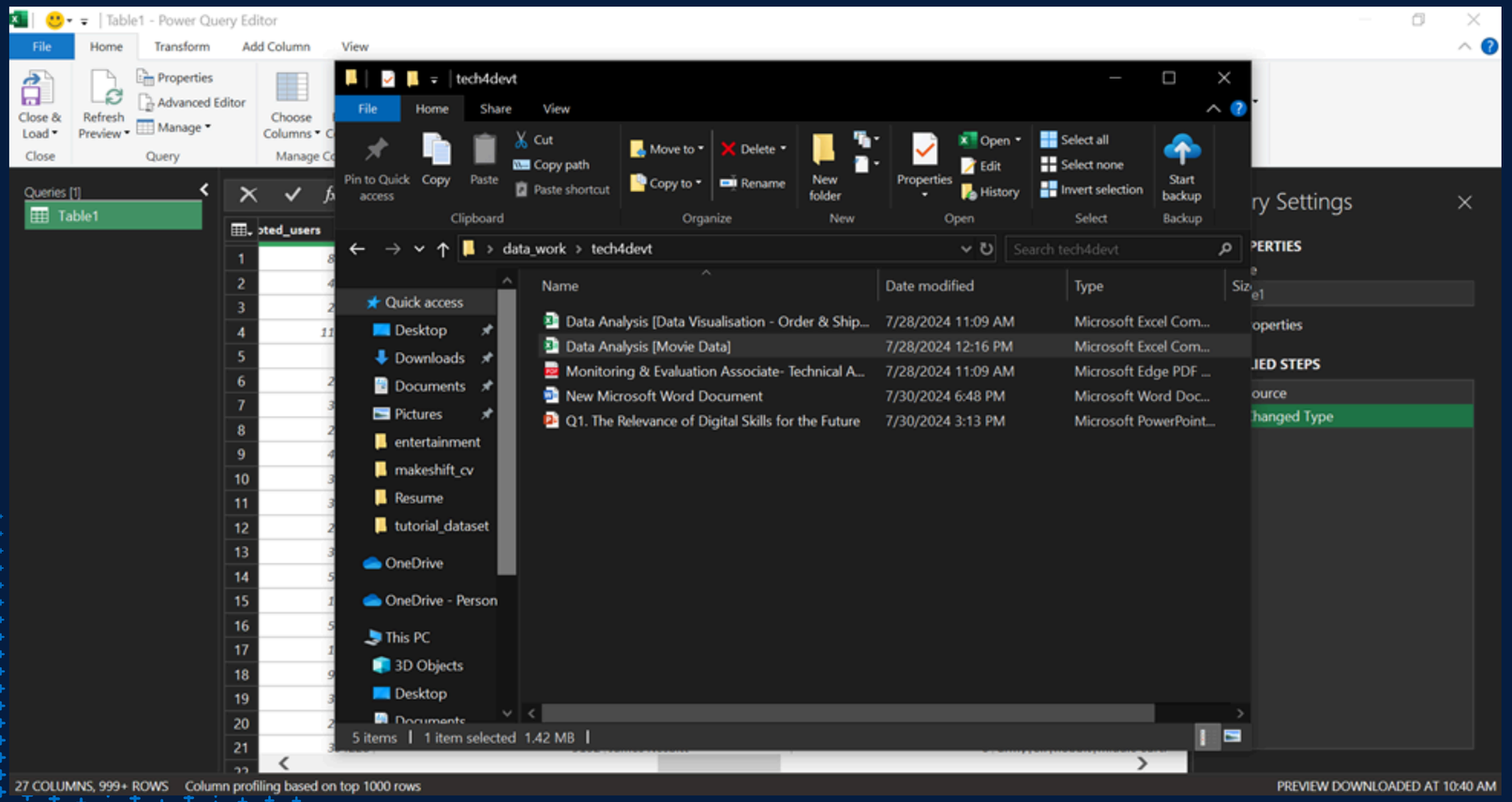
Cinema is an art form that weaves a narrative into a visual and auditory experience, thus imprinting an indelible mark on our collective consciousness. There is no single "most important" element as all film components — story, characters, acting, direction, production quality, pacing, emotional impact, relatability, and originality — contribute to its overall quality. But what makes a good movie? This question has stirred up numerous debates among film critics and movie buffs. Let's embark on a data analytics journey through the various key factors that the dataset at our disposal is made up of. To analyze and visualize the given data, we can focus on extracting insights such as the distribution of movie budgets, IMDB scores, genres, gross earnings, and other factors like the relationship between budget and gross, or the impact of an actor's or director's social media presence on movie success.

Here’s a step-by-step breakdown of how we can approach this: data cleaning and preparation, descriptive statistics, correlation analysis, visualization, specific information extraction and data insights.

DATA PREPARATION & CLEANING

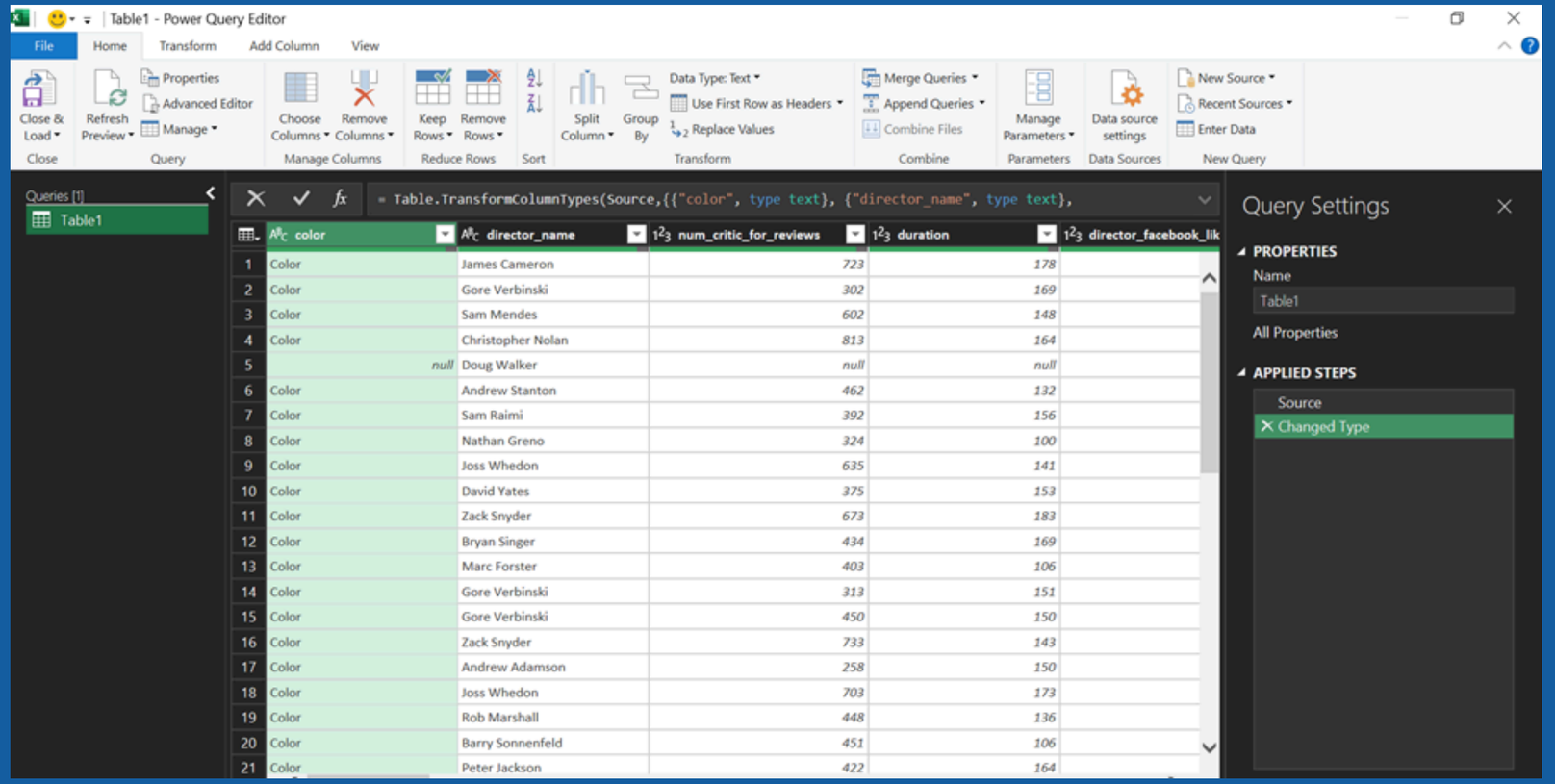
Data collection

The dataset for this project was sourced from the following:
C:\Users\USER\Desktop\data_work\tech4devt



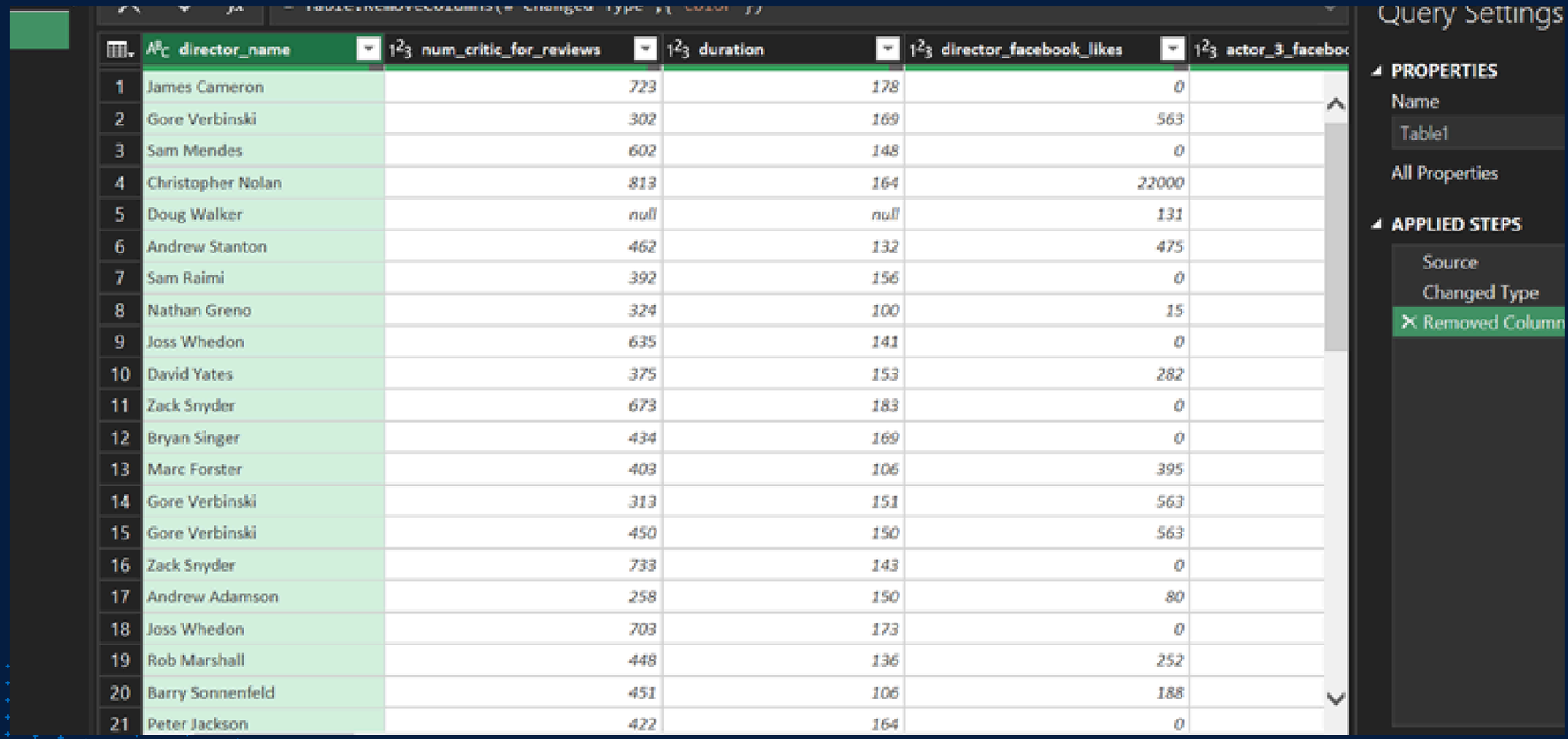
Dataset migration

The dataset was then imported from CSV file format into Excel Power Query



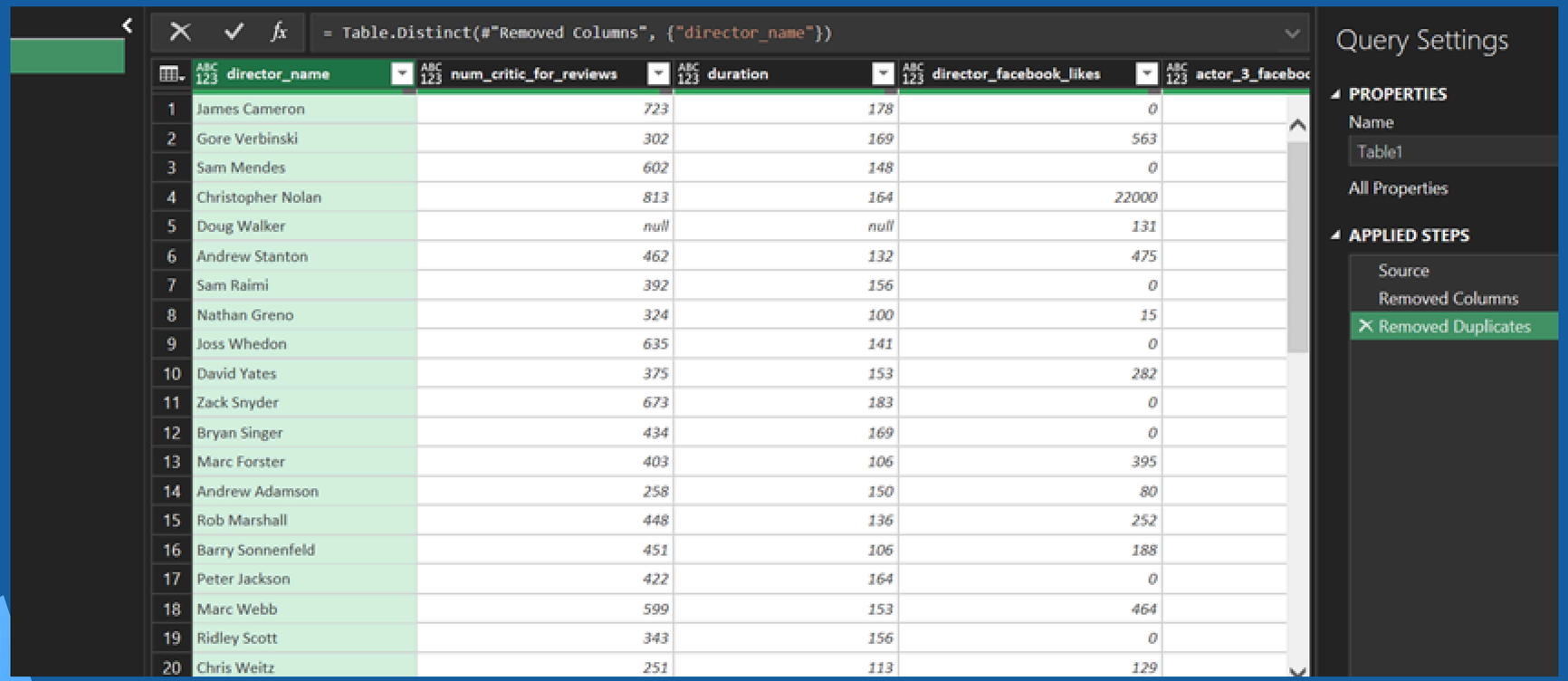
Data Transformation

Removal of unwanted columns

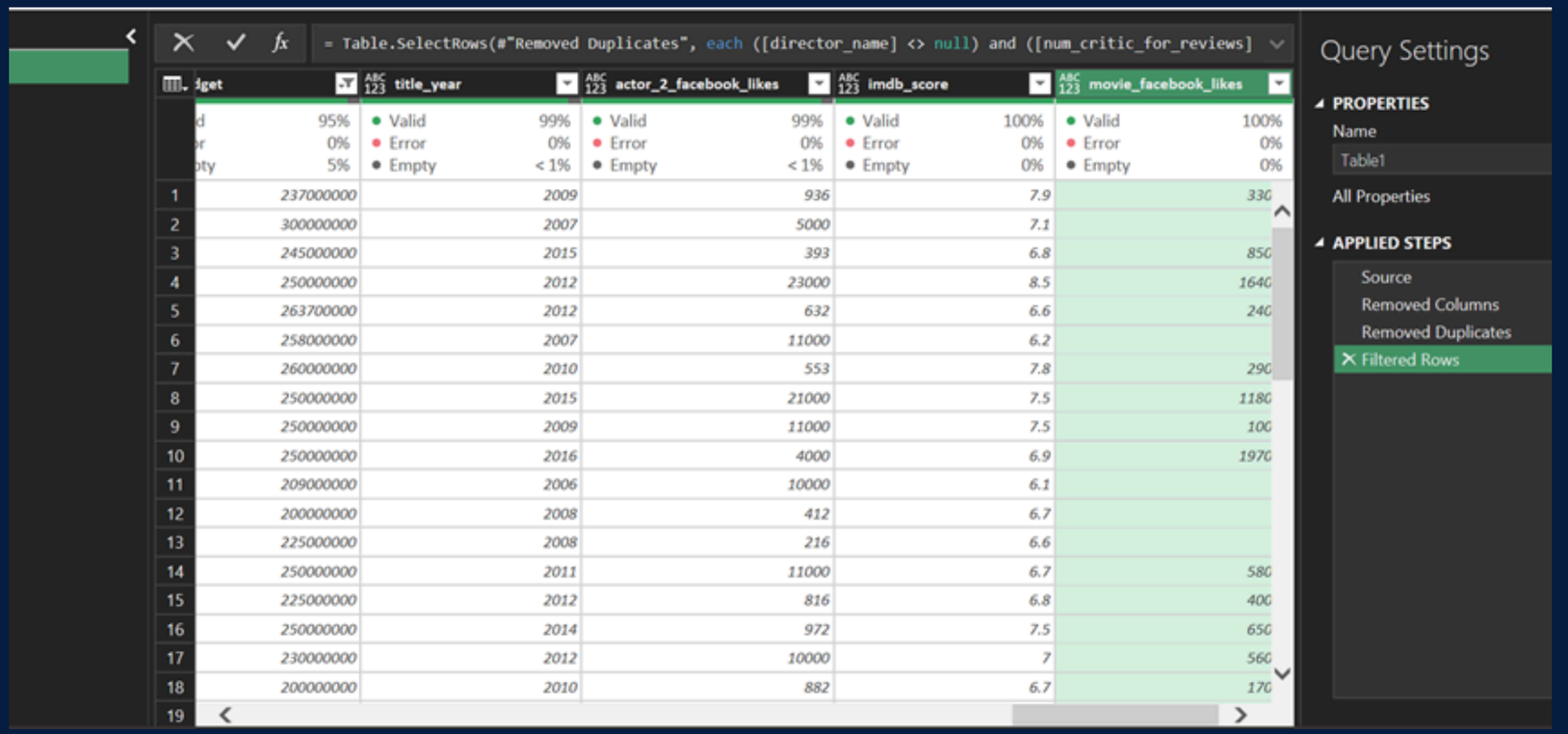


	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_faceboo
1	James Cameron	723	178	0	
2	Gore Verbinski	302	169	563	
3	Sam Mendes	602	148	0	
4	Christopher Nolan	813	164	22000	
5	Doug Walker	null	null	131	
6	Andrew Stanton	462	132	475	
7	Sam Raimi	392	156	0	
8	Nathan Greno	324	100	15	
9	Joss Whedon	635	141	0	
10	David Yates	375	153	282	
11	Zack Snyder	673	183	0	
12	Bryan Singer	434	169	0	
13	Marc Forster	403	106	395	
14	Gore Verbinski	313	151	563	
15	Gore Verbinski	450	150	563	
16	Zack Snyder	733	143	0	
17	Andrew Adamson	258	150	80	
18	Joss Whedon	703	173	0	
19	Rob Marshall	448	136	252	
20	Barry Sonnenfeld	451	106	188	
21	Peter Jackson	422	164	0	

Removal of duplicates

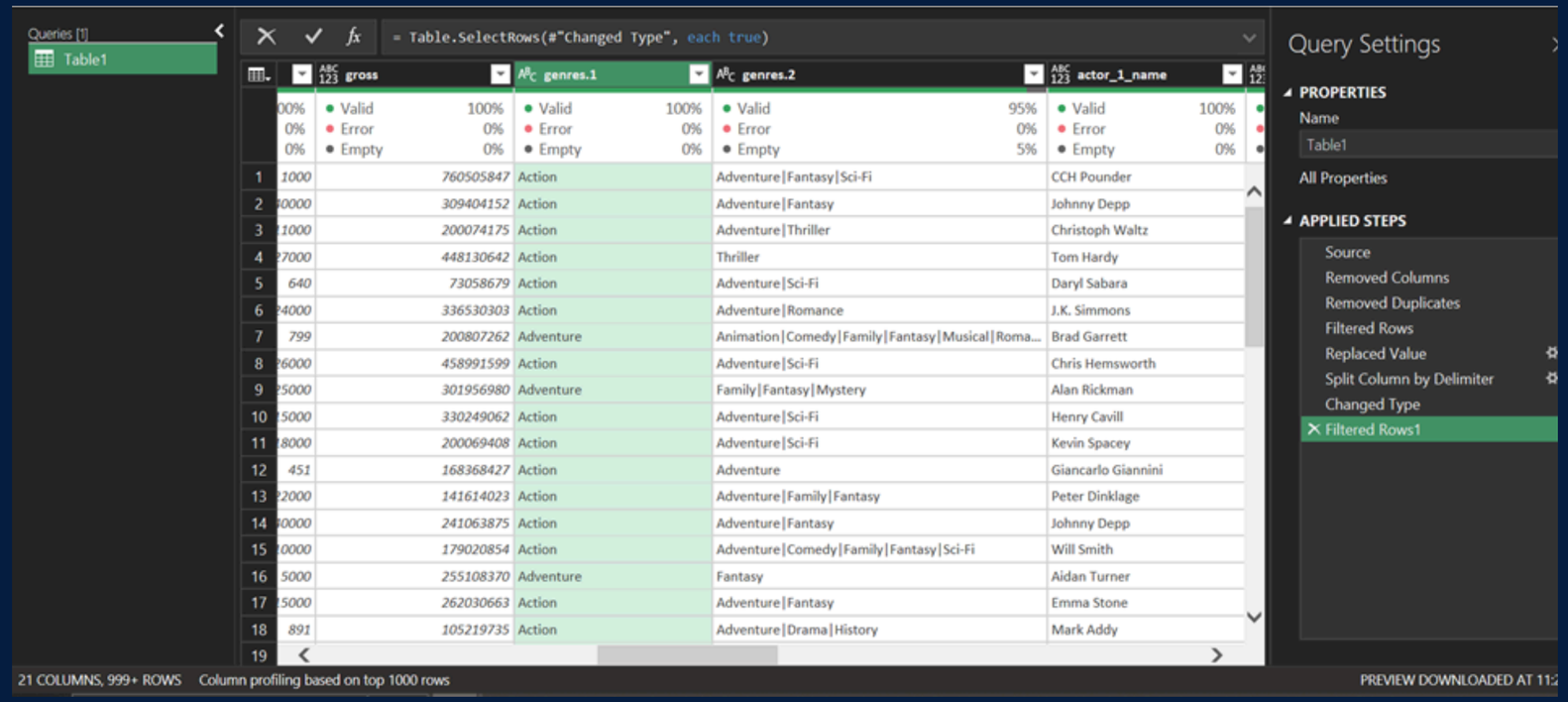


Removal of rows with null values

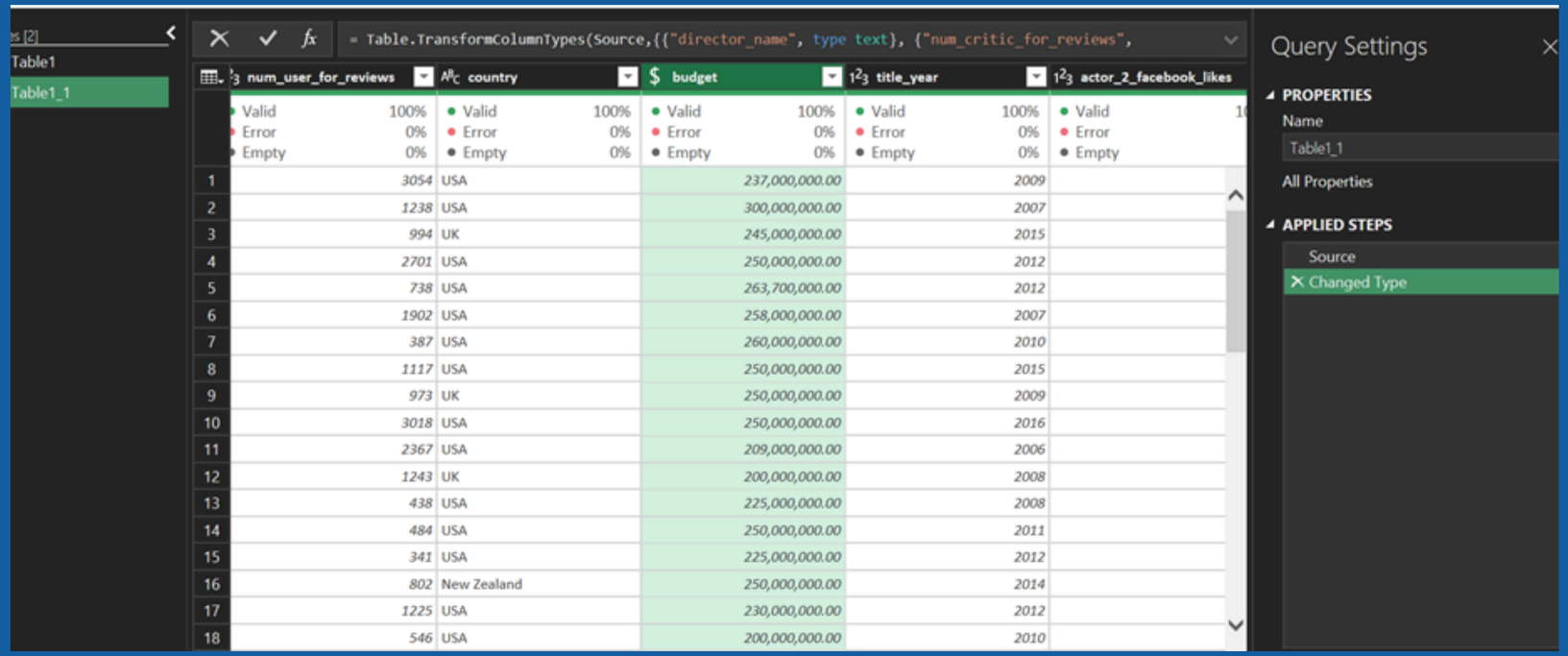


	...get	title_year	actor_2_facebook_likes	imdb_score	movie_facebook_likes
	...d 95% ...r 0% ...pty 5%	Valid 99% Error 0% Empty < 1%	Valid 99% Error 0% Empty < 1%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
1	237000000	2009	936	7.9	33(
2	300000000	2007	5000	7.1	
3	245000000	2015	393	6.8	85(
4	250000000	2012	23000	8.5	164(
5	263700000	2012	632	6.6	24(
6	258000000	2007	11000	6.2	
7	260000000	2010	553	7.8	29(
8	250000000	2015	21000	7.5	118(
9	250000000	2009	11000	7.5	10(
10	250000000	2016	4000	6.9	197(
11	209000000	2006	10000	6.1	
12	200000000	2008	412	6.7	
13	225000000	2008	216	6.6	
14	250000000	2011	11000	6.7	58(
15	225000000	2012	816	6.8	40(
16	250000000	2014	972	7.5	65(
17	230000000	2012	10000	7	56(
18	200000000	2010	882	6.7	17(
19					

Split columns by delimiter



Change data types



DESCRIPTIVE ANALYSIS

In this section we look at the mean, median and range values of Budget, Gross, IDMB Score, Duration and Movie Facebook Likes

	Descriptive Analysis		
Particulars	Average	Median	
Budget	181623232.3	175000000	
Gross	192773425.9	161087183	
IDMB Score	6.753535354	6.7	
Duration	122.7474747	119	
Movie Facebook Likes	36040.40404	24000	

	Range		
Particulars	Maximum	(Minimum)	Range_Value
Budget	300000000	100000000	200000000
Gross	760505847	21379315	739126532
IDMB Score	8.5	4.2	4.3
Duration	206	85	124
Movie Facebook Likes	197000	0	197000

Insights from the descriptive analysis

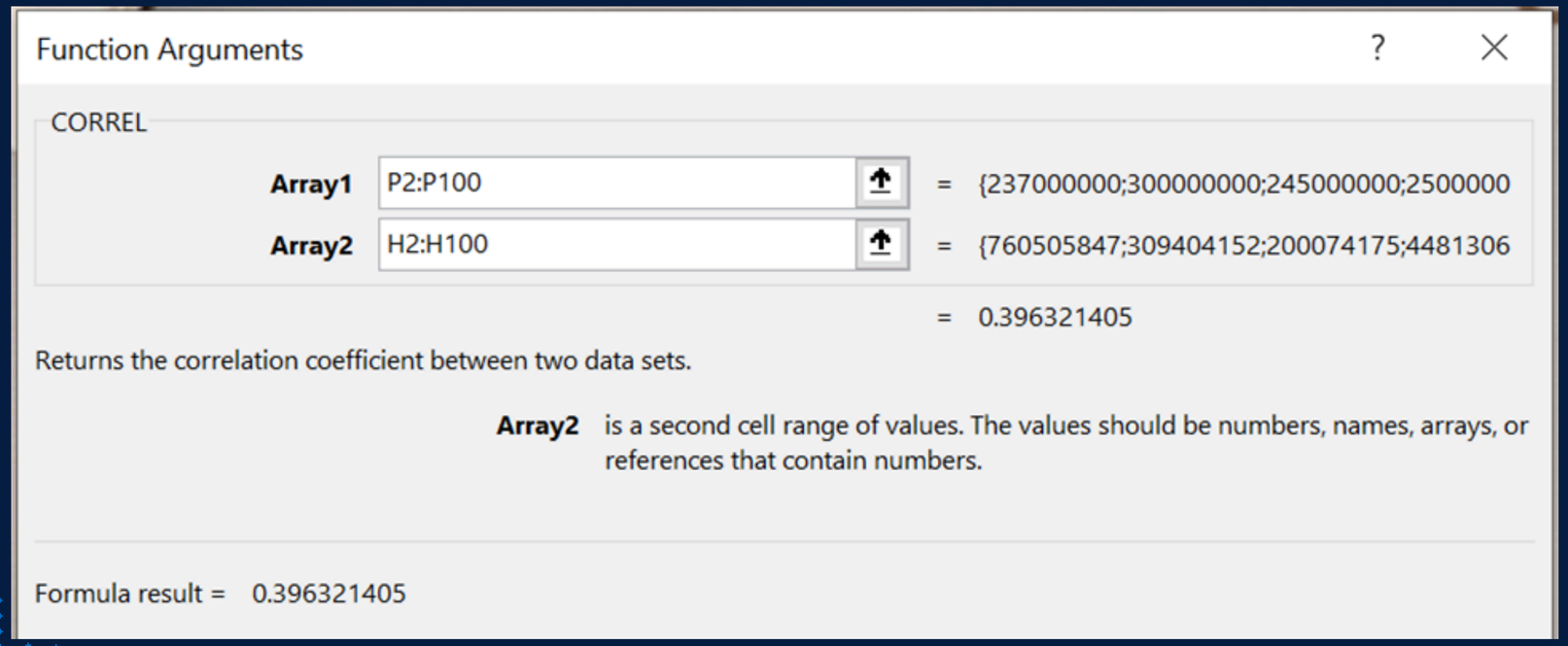
Budget: The average budget is around $181.62 million, with a minimum of $100 million and a maximum of $300 million.

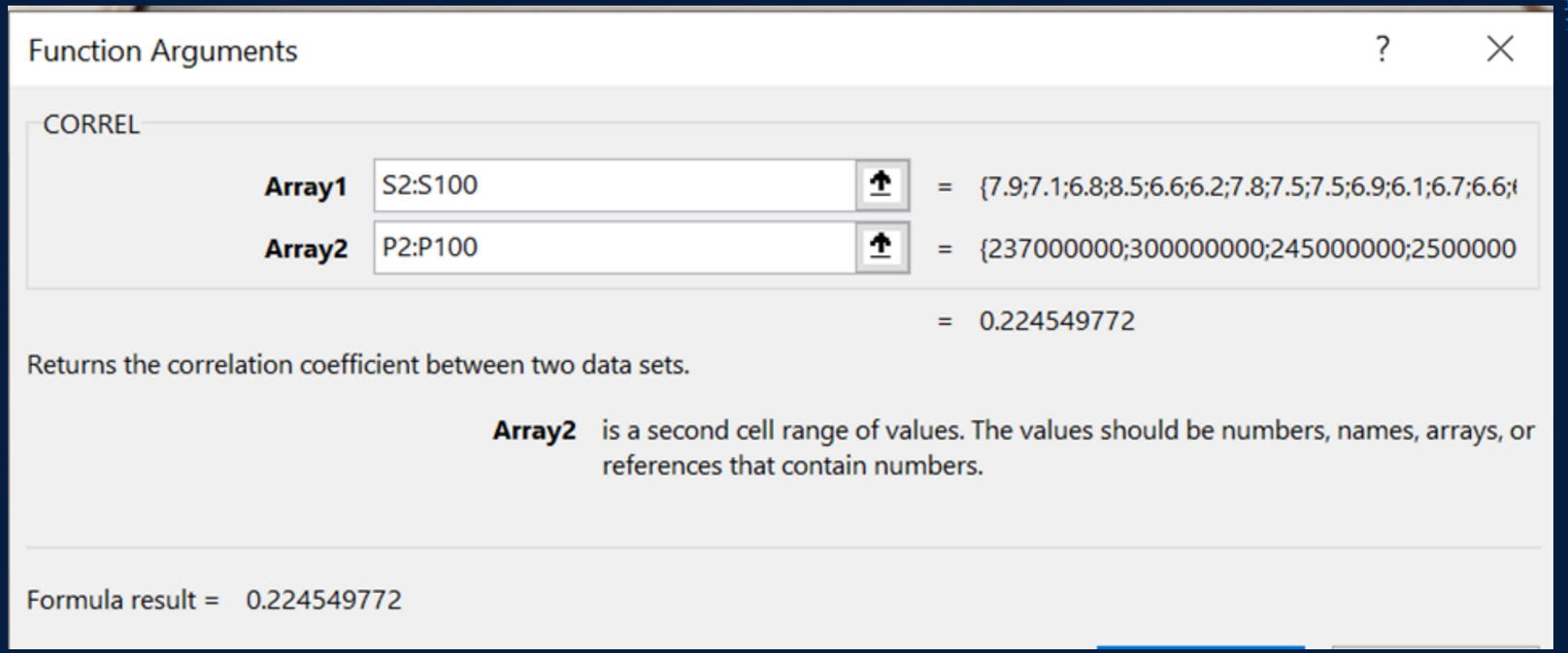
Gross Earnings: The average gross revenue is $192.77 million, with a maximum of $760.51 million and a minimum of $213..79 million.

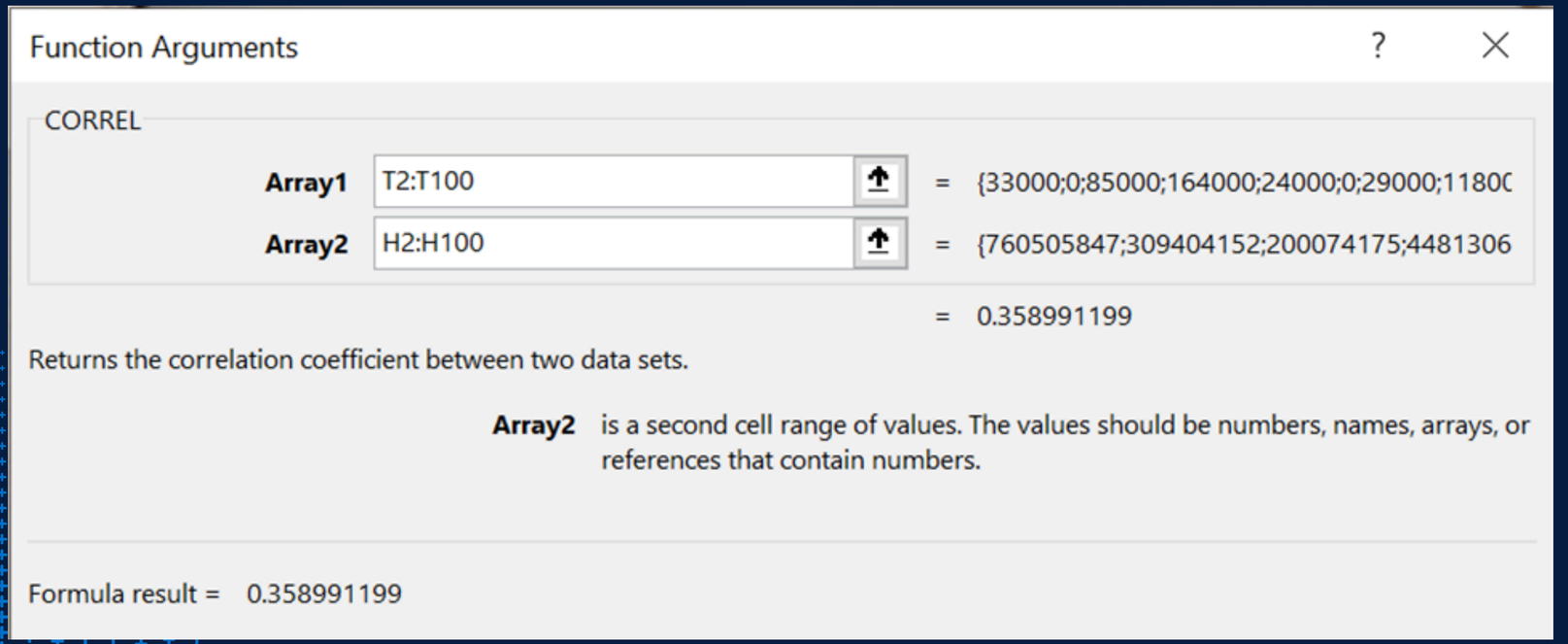
IMDB Score: The average IMDB score is 4.3, ranging from 4.2 to 8.5.

CORRELATION ANALYSIS

Here, we look at the relationships that exist between Budget Vs Gross, IDMB Score Vs Budget and Movie Facebook Likes Vs Gross




Function Arguments
CORREL
Array1
S2:S100
= {7.9;7.1;6.8;8.5;6.6;6.2;7.8;7.5;7.5;6.9;6.1;6.7;6.6;
Array2
P2:P100
= {237000000;300000000;245000000;2500000
= 0.224549772
Returns the correlation coefficient between two data sets.
Array2 is a second cell range of values. The values should be numbers, names, arrays, or references that contain numbers.
Formula result = 0.224549772


Function Arguments
CORREL
Array1
T2:T100
= {33000;0;85000;164000;24000;0;29000;1180
Array2
H2:H100
= {760505847;309404152;200074175;4481306
= 0.358991199
Returns the correlation coefficient between two data sets.
Array2 is a second cell range of values. The values should be numbers, names, arrays, or references that contain numbers.
Formula result = 0.358991199

Insights from the correlation analysis

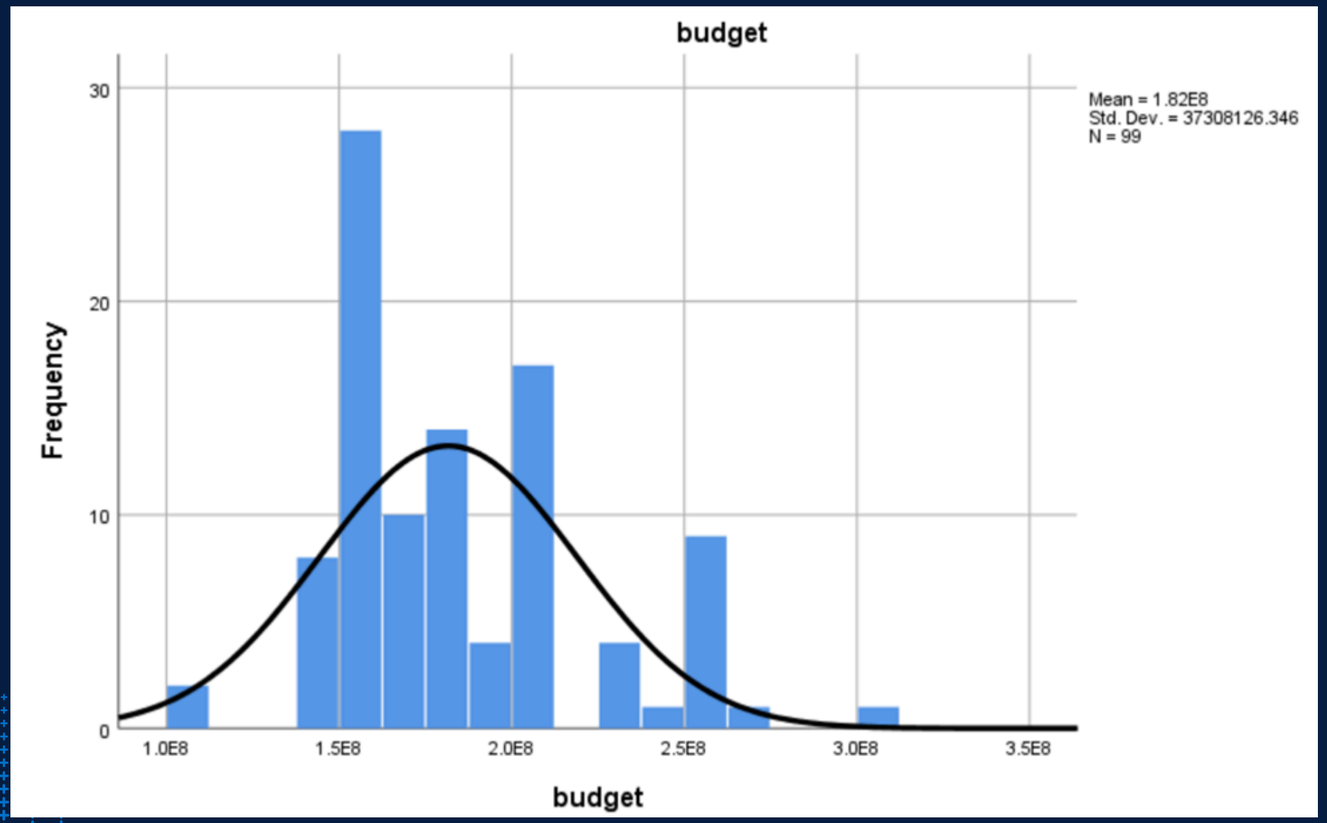
Budget vs Gross: There is a positive correlation (0.40), meaning higher budgets does guarantee higher gross earnings.

IMDB Score vs Gross: Moderate positive correlation (0.22), indicating that higher-rated movies tend to perform better at the box office.

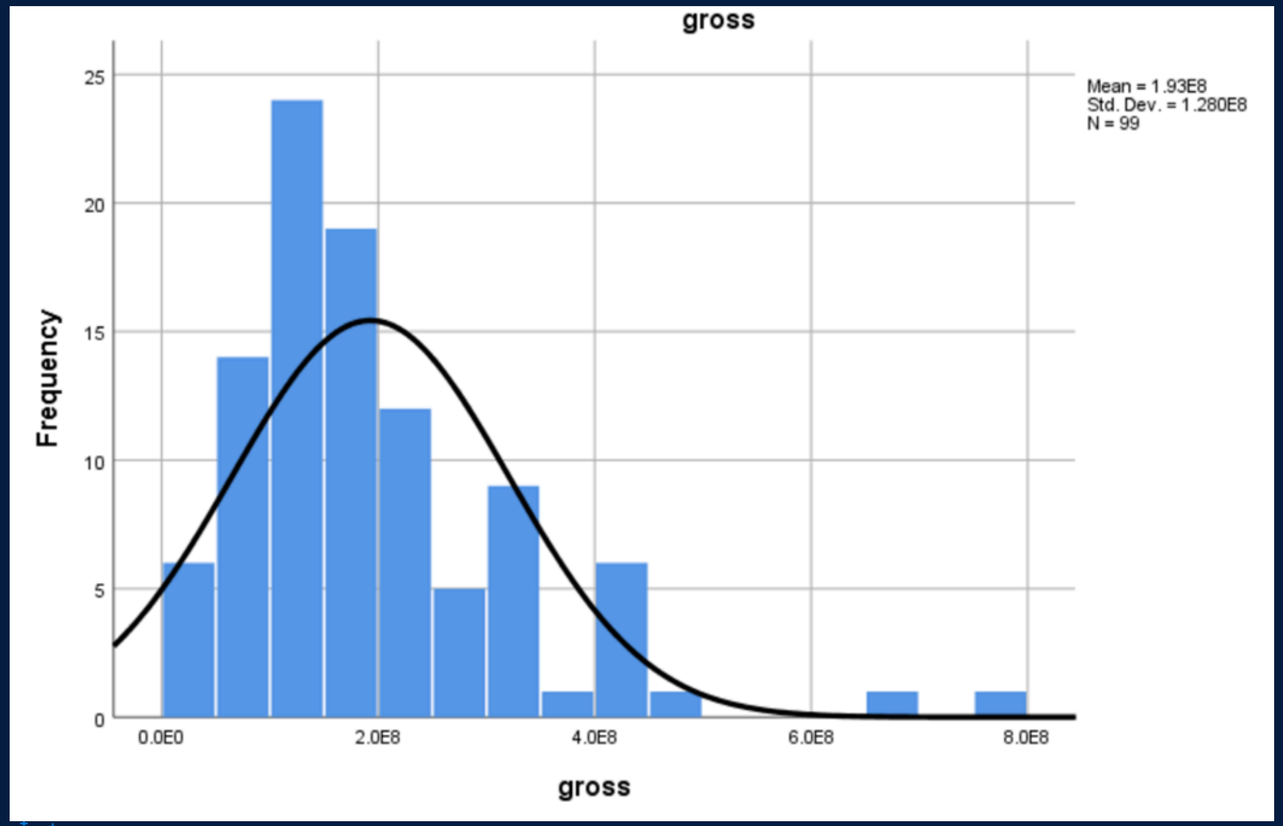
Social Media Impact: There is a strong positive correlation between gross earnings and movie Facebook likes (0.40) suggesting that movies with larger social media presences tend to be associated with higher-box office performances.

VISUALIZATIONS

Histograms of Budget



Histograms of Gross earnings

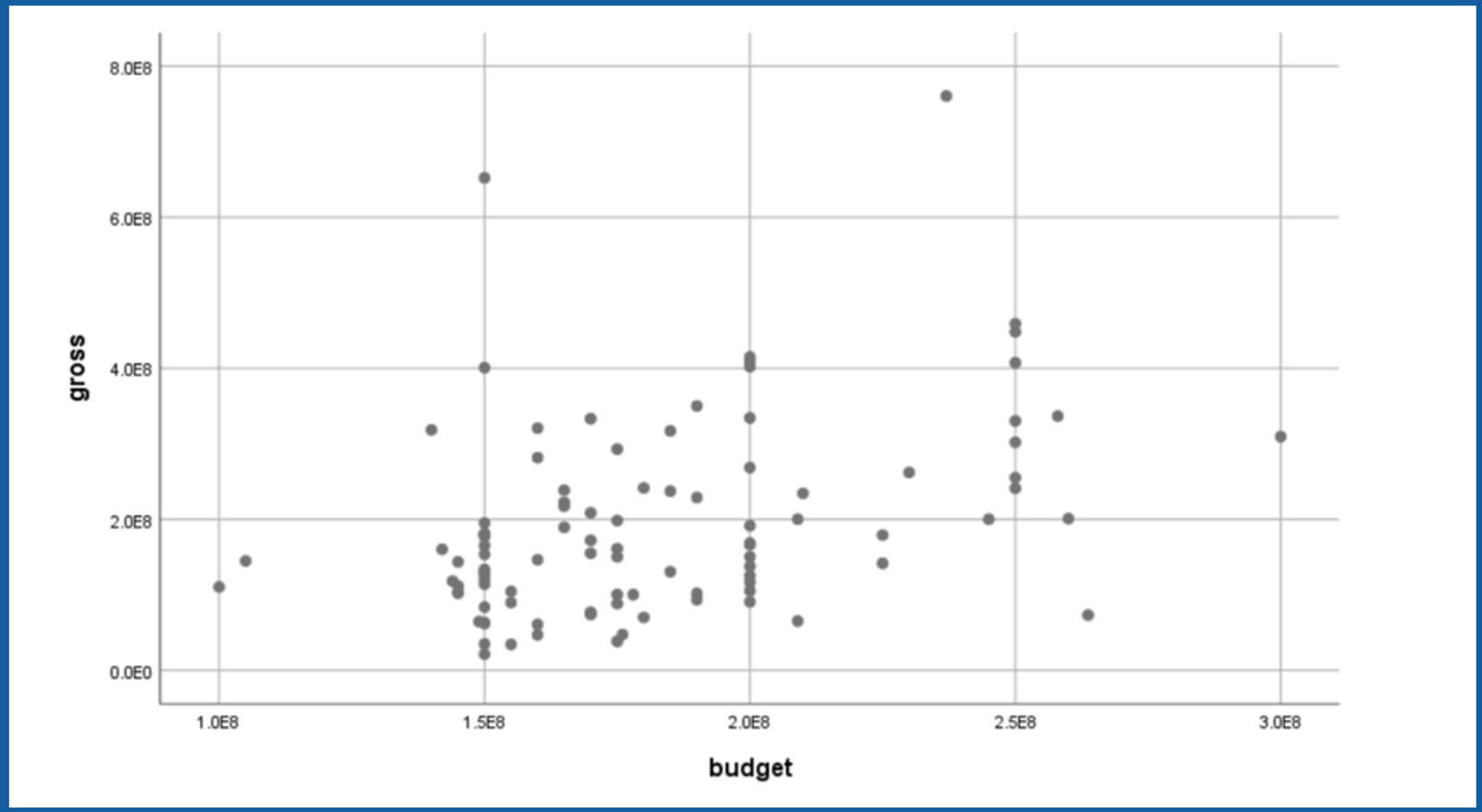


Insights from the histograms

The distribution of budgets shows that all the movies analyzed are mixed of moderate to high-budget blockbusters.

Gross earnings have a wide range, with a few movies earning significantly more than others.

Scatter Plots of Budget Vs. Gross



Insights from scatter plot

The scatter plot shows that while budget and gross earnings do not follow a direct linear trend, some high-budget movies did achieve high gross revenue.

Heat Map

	duration	*budget*	*gross*	*mdb_score*	*facebook_likes*
duration	1				
budget	0.465431	1			
gross	0.293289	0.396321	1		
imdb_score	0.114431	0.22455	0.465188	1	
movie_facebook_likes	0.29984	0.159392	0.358991	0.474635	1

Insights from heat map

The heatmap highlights the relationships among duration of movies, budget, gross earnings, IMDB scores, and social media presence. It confirms that social media metrics for movies correlate strongly with IMDB scores.

RECOMMENDATION

From the various insights uncovered after an indepth exploration of data analysis reveals that in as much higher budgets for movies are good, they do not necessarily guarantee the success of movies at the box-office.

Movie owners need to engage high performing players ranging from directors to actors in order to ensure the success of their movies.

Also there is need for effective PR most especially on social media platforms to engage netizens as more movie lovers tend to be on the various social media platforms. It has been proven that social media does drive movie success.