MOVIE STUDIO ANALYSIS

Identifying Key Factors For Success in the Film Industry

Overview

- The movie studio industry faces increasing pressure to maximize efficiency, minimize costs, and improve profitability while remaining competitive in an ever-changing market.

- With the growing complexity of film production, distribution, and marketing, studios must leverage data-driven insights to make informed decisions and stay ahead of industry trends.

- The project involves researching current box office trends to identify film genres, themes, and styles that are resonating most with audiences.

- This data-driven analysis will highlight popular film characteristics and emerging viewer preferences, offering valuable insights on what types of movies are achieving commercial success.

Business Problem Statement

Identify Inefficiencies → Optimize production pipelines → Propose actionable recommendations to enhance overall business performance

- The problem revolves around understanding and improving the operational performance of movie studios using data analytics.

- The goal is to identify inefficiencies, optimize production pipelines, and propose actionable recommendations to enhance overall business performance.

- We were charged with exploring what types of films are currently doing the best at the box office.

- We must then translate those findings into actionable insights that the head of your company's new movie studio can use to help decide what type of films to create.

OBJECTIVES
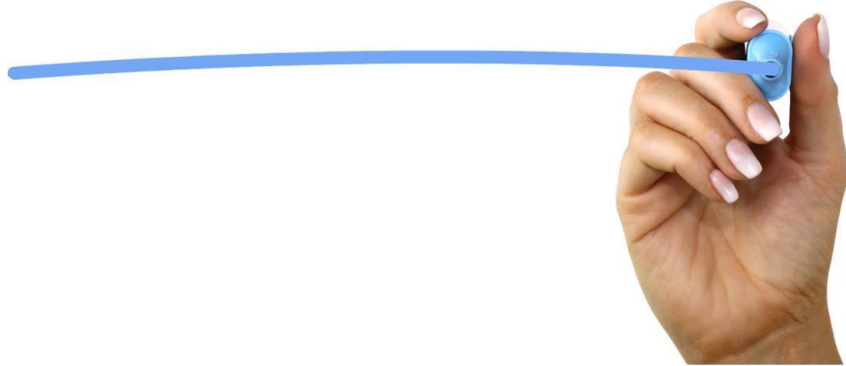
- Analyze the trend in movie sales over time to identify shifts in audience demand.

- Identify key market segments to target for maximum engagement and revenue generation.

- Assess the popularity of different film genres to guide genre selection in alignment with audience preferences.

- Evaluate the impact of movie runtime on popularity to determine optimal lengths for future productions.

DATA UNDERSTANDING

Datasets

```
# getting to understand what is contained in the dataframe
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 2007 entries, 0 to 3353
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   title           2007 non-null   object
 1   studio          2007 non-null   object
 2   domestic_gross  2007 non-null   float64
 3   foreign_gross   2007 non-null   float64
 4   year            2007 non-null   int64
dtypes: float64(2), int64(1), object(2)
memory usage: 94.1+ KB
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 249717 entries, 0 to 314172
Data columns (total 5 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   title            249717 non-null  object
 1   runtime_minutes  249717 non-null  float64
 2   genres           249717 non-null  object
 3   rating           249717 non-null  float64
 4   numvotes         249717 non-null  float64
dtypes: float64(3), object(2)
memory usage: 11.4+ MB
```

bom.movie_gross

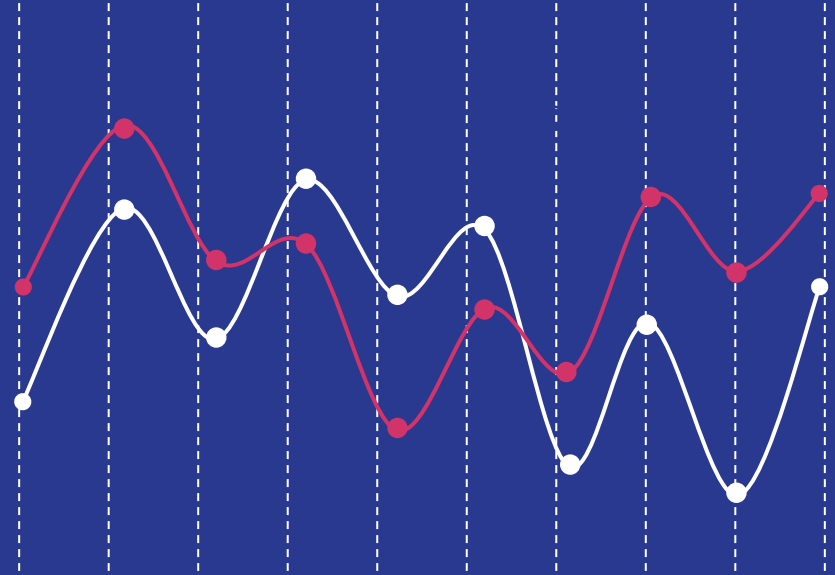**im.db.zip**

Description of dataset

The *bom.movie_gross* has:

❖ 2,007 rows (entries)

❖ 5 data column
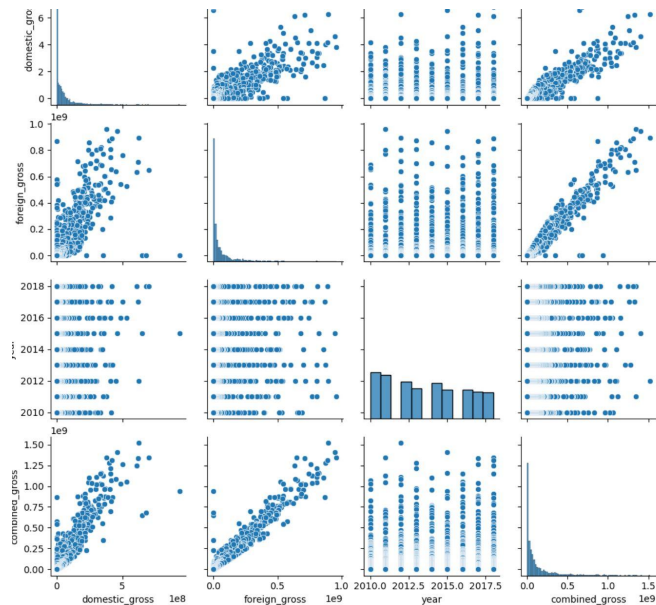
2 objects
2 floats
1 Integer

The *im.db.zip* contains :

❖ 249717 rows (entries)
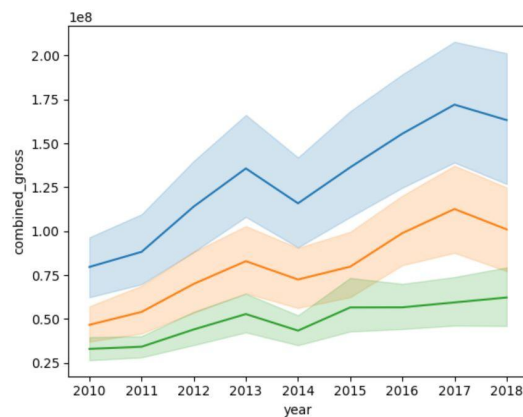
❖ 5 data columns

3 floats
2 objects

Source of data: https://www.imdb.com/
https://www.themoviedb.org/

DATA ANALYSIS

Line graph showing combine gross against year

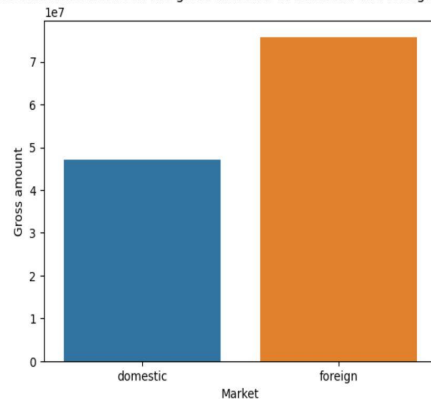The difference between the gross amounts in domestic and foreign markets

Pair plot showing correlation between foreign gross and domestic gross, foreign gross and combined_gross and domestic gross and combined gross

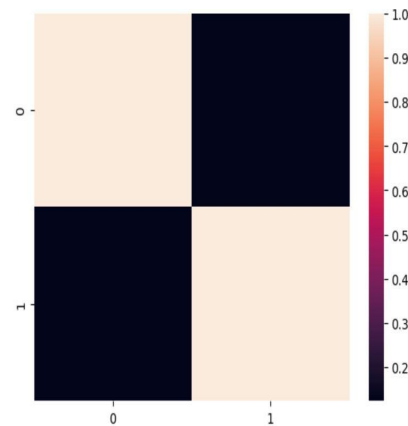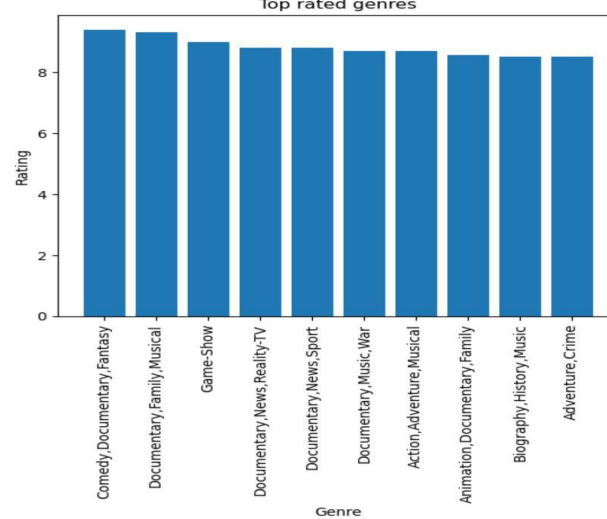*Bar chart showing difference in the amounts gained in the foreign market and in the domestic market.*

Pair Plot showing correlation between the
numerical columns of the second data set

Bar graph of top rated
genres against rating

Heatmap on runtime
minutes and rating

Deductions

Data set 1

- There is a strong positive correlation between foreign gross and domestic gross, foreign gross and combined_gross and domestic gross and combined gross

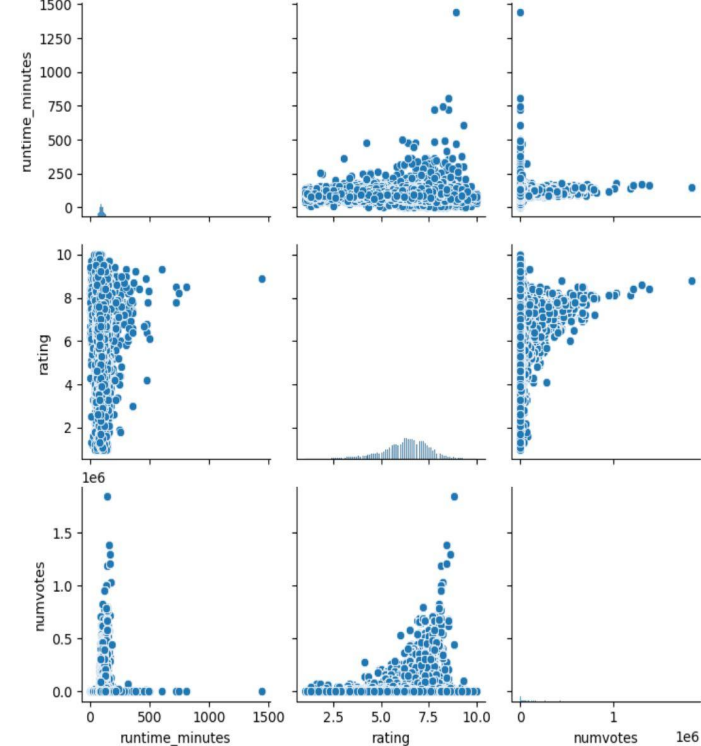- There has been an increase in the sales of the movies over time indicating that the industry's demand has been on the rise.

- There is a difference in the amounts gained in the foreign market and in the domestic market. The foreign market has more returns than the domestic market. This could be attributed to the size of the two markets.

Data set 2

- There appears to be no/minimal correlation in the factors in the second data set. We could further confirm that the runtime of movies does not affect its popularity.

- The heatmap confirms that the runtime does not affect the popularity . The genre of the movie is however anticipated to affect the popularity of the movie.

- The Comedy, Documentary, Fantasy genre is the most popular

STATISTICAL ANALYSIS

Hypothesis testing

There is a claim that there is a difference in the gross amount earned in the domestic and foreign market. To ascertain this we carried out a two sample independent t test.

Null Hypothesis (H0): There is no difference in the gross amount earned in the domestic market compared to that earned in the foreign market.
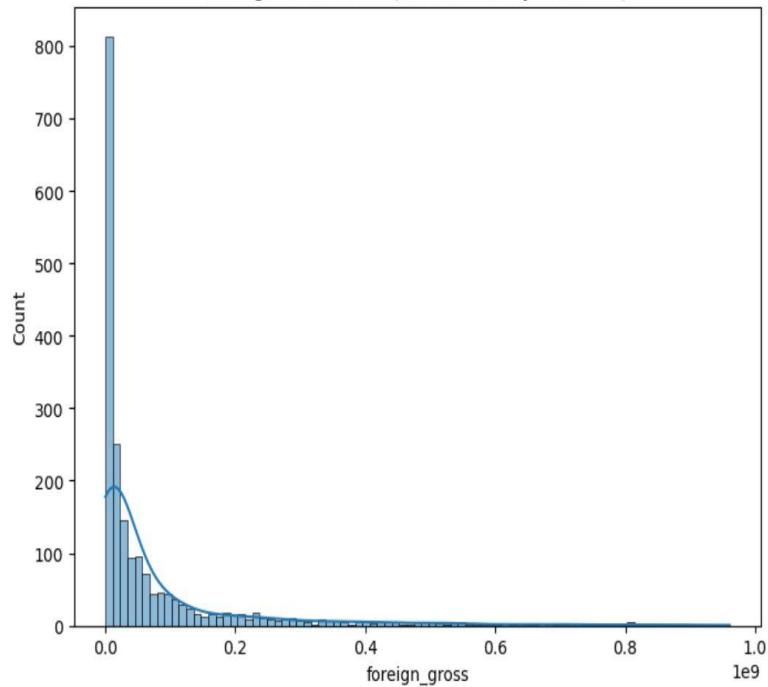
$\mu1 = \mu2$

Alternative Hypothesis (H1): There is a difference in the gross amount earned in the domestic market compared to that earned in the foreign market.

$\mu1 \neq \mu2$

The hypothesis test confirms that the earlier difference between the gross amount earned in the foreign market and domestic market is significant.
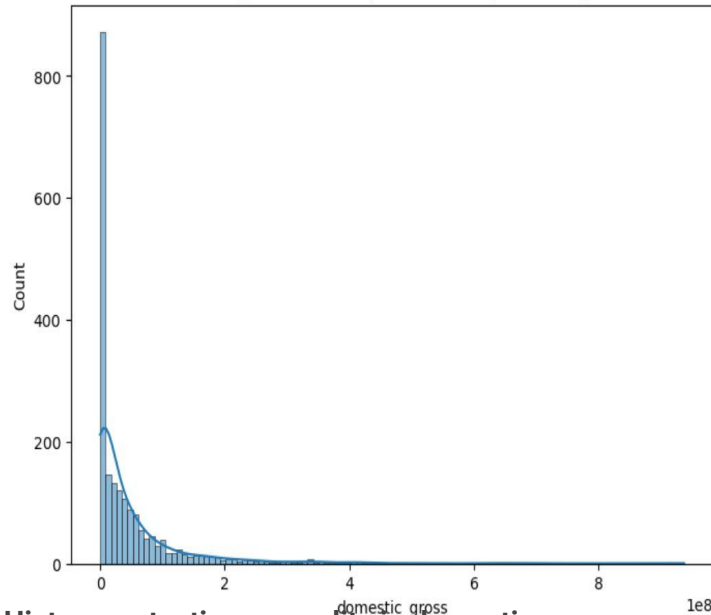
Test for Normality

Histogram testing normality in foreign gross

Histogram testing normality in domestic gross

Regression Model

OLS Regression Results

Dep. Variable:	foreign_gross	**R-squared:**	0.590
Model:	OLS	**Adj. R-squared:**	0.589
Method:	Least Squares	**F-statistic:**	2882.
Date:	Fri, 15 Nov 2024	**Prob (F-statistic):**	0.00
Time:	05:20:54	**Log-Likelihood:**	-39573.
No. Observations:	2007	**AIC:**	7.915e+04
Df Residuals:	2005	**BIC:**	7.916e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	1.467e+07	2.28e+06	6.431	0.000	1.02e+07	1.91e+07
domestic_gross	1.3000	0.024	53.682	0.000	1.252	1.347

Omnibus:	657.742	**Durbin-Watson:**	1.311
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	124753.923
Skew:	-0.242	**Prob(JB):**	0.00
Kurtosis:	41.621	**Cond. No.**	1.09e+08

From the regression model we make the following notes:

➢ Standard Errors assume that the covariance matrix of the errors is correctly specified.

➢ The condition number is large, 1.09e+08. This might indicate that there are strong multicollinearity or other numerical problems.

➢ The model has a coefficient of determination of 59% indicating that the 59% of the variation in the foreign gross is explained by variation in the domestic gross.

Linear Regression - Predicting foreign income
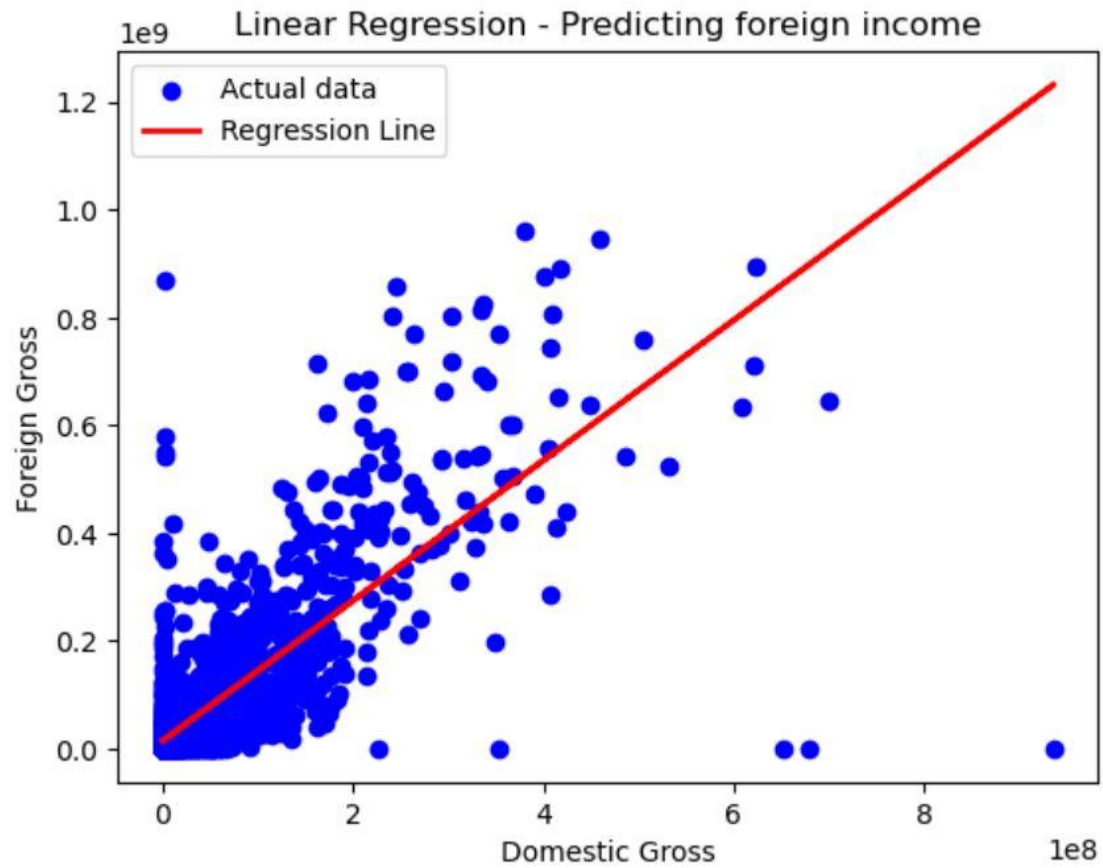
Recommendations

❏ As shown in the domestic gross against year line graph , there has been an increase in the sales of the movies over time indicating that the industry's demand has been on the rise, therefore this venture is viable for the company and could definitely be explored.

❏ The company should focus on maximizing on the foreign market sales.This is because from the market against gross amount bar plot, there is a difference in the amounts gained in the foreign market and in the domestic market. The foreign market has more returns than the domestic market.

❏ The company should consider benchmarking at studios with HC as they are at the top. This would help in adoption of practices that these top studios do to achieve the numbers that they have over the years.

❏ The company's studio should not focus on the runtime but other factors of quality that will make the movies more popular. The heatmap confirms that the runtime does not affect the popularity and thus the company's studio should not focus on the runtime but other factors of quality that will make the movies more popular eg the genre of the movie.

❏ The Comedy, Documentary, Fantasy genre is the most popular and therefore the studio should therefore venture into producing that genre.

Conclusion

The research done and the results found indicate that the company's plan of starting a movie studio is a viable venture and should therefore be explored.

To ensure that the venture is profitable the recommendations mentioned in the report should be taken into consideration.

The industry is ever changing and constant research and always keeping up to date with changes in consumer preferences and the financial performance of the industry.

Presented & researched by

Rawlings Mayabi

Hellen Mwaniki

Faith Karaya

Reagan Saf

Samuel Kamau

Cyrus Wambugu

Colletta Kiilu