

# MICROSOFT MOVIE STUDIO PROJECT

## 1. BUSINESS UNDERSTANDING

Microsoft is looking into movie studios as it plans on launching its own. The task at hand is to explore given datasets and understand how much money goes into film creation and what type of films succeed the most. My findings will help Microsoft understand the expected profit that comes from films and how they can make their film successful.

## 2. DATA UNDERSTANDING

### 2.1. DATA COLLECTION

The data was collected from the [box office mojo](#) , [IMDB](#) and [the numbers](#) datasets. In the American and international markets, the major film studios, often known simply as the majors or the Big Five studios, are commonly regarded as the five diversified media conglomerates whose various film production and distribution subsidiaries collectively command approximately 80 to 85% of U.S. box office revenue. The term may also be applied more specifically to the primary motion picture business subsidiary of each respective conglomerate.

### 2.2. DATA DESCRIPTION

Each of the datasets had unique columns , the box office dataset contained information on the domestic and foreign gross earned by movies. The numbers dataset contained the production budget and the world wide gross columns, these columns were to be used to calculate the average budget and return on interest. Finally in the IMDB dataset we used two tables , the movie basics and movie ratings tables. The tables contained information on movie ratings and movie genres which were vital for our data analysis.

## 3. DATA PREPARATION

### 3.1. SELECTING DATA

We focused on specific tables and specific columns which were relevant and impactful on our research. For the box office dataset we used the domestic and foreign gross columns and the production budget and the world wide gross columns for the numbers dataset. Lastly we used the movie basics and movie reviews table from the IMDB database.

### 3.2. DATA CLEANING

This was done to ensure the Validity, Accuracy, Completeness, Consistency and Uniformity of the Data. We replaced all the null values with the most frequent value for the foreign gross column in the box office dataset. For the numbers data set we converted the data type to integers in order to perform arithmetic operations.

## 4. DATA ANALYSIS

### 4.1. EXPLORATORY DATA ANALYSIS

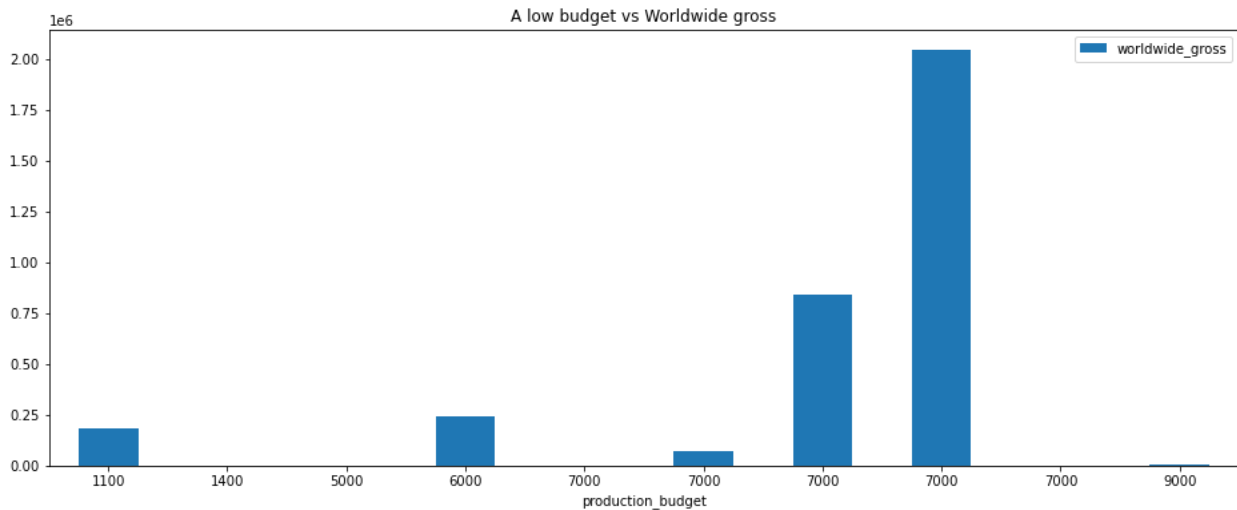
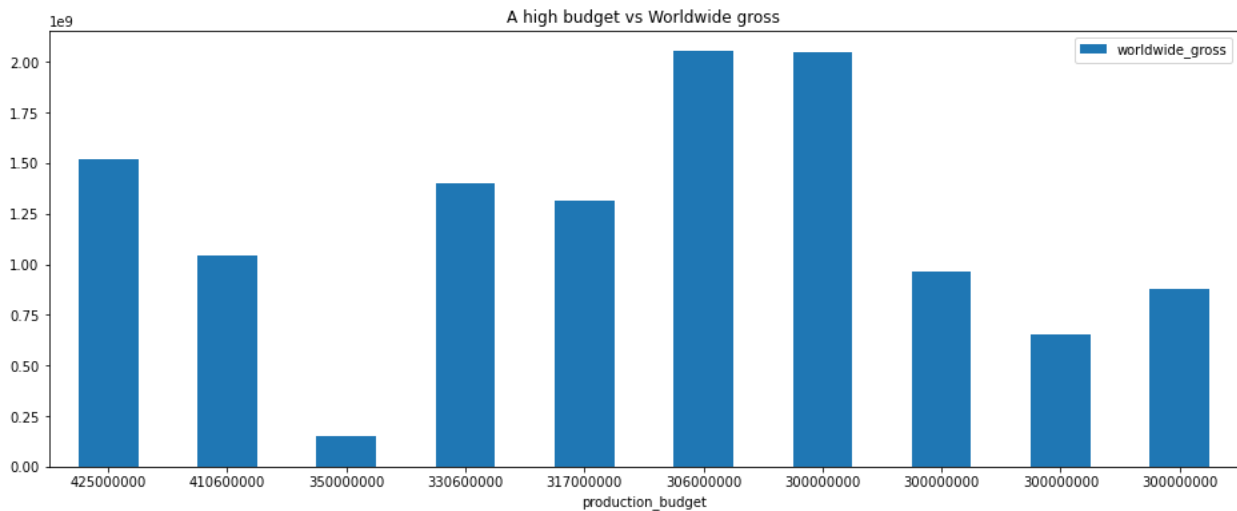
#### UNIVARIATE DATA ANALYSIS

There were few outliers in the numerical dataset however they were dismissed as insignificant as they did not affect our data analysis.

##### Univariate Analysis Recommendation

- a) From the box office dataset we learn that the average domestic gross is 28,745,850 and the average foreign gross is 114,861,500. The average production budget was 31,587,760. From the numbers dataset we observe that a low production budget leads to little or no return, on the other hand, the higher the production budget the higher the worldwide gross.

- b) The movie ratings and movie basics table identifies the most popular movie genres as; Romance, Sci-Fi and Thriller.
- c) The data also demonstrated a high correlation between the worldwide gross and production budget. In addition, as shown in the figures below, not all films that had a high production budget produced high profit margins as well as the inverse.
- d) Average return on interest was: 59,682,180 based on the average difference between the worldwide gross and production budget.



## 4. CONCLUSION

In conclusion, it can be deduced that a high production budget does not necessarily lead to a higher rate of interest. Secondly, even though the worldwide gross increases with the production budget, it can be concluded that the return on interest will not be affected by increasing the production budget.

## 6. RECOMMENDATION

In order to create a successful film I recommend sourcing for technical expertise in production of the three highly rated genres. I would also recommend working with the average production budget in order to maximize the return on interest.