**Disney Movies Box Office Performance (until 2021)**

MGMT 58600 Fall 2023

MSBAIM 36709953

Chun-Yi Chiang

**Project Objective**

The objective of this project is to perform data analysis on Disney movies based on a given dataset using Python. This analysis involves generating descriptive statistics and conducting linear regression analysis. The tools utilized for this project include seaborn, matplotlib, and pandas. The primary aim is to extract valuable business insights from the analysis results, with a specific focus on understanding the linear relationships between box office performance and three key factors: budget, running time, and IMDb score.

**Project Data**

The dataset used for this project was obtained from Kaggle, specifically the 'Walt Disney Movies' dataset[[1]](#footnote-1). To ensure the data's quality and relevance, the CSV file was preprocessed before importing it into Python. The cleaned dataset includes the following columns:

* title: Movie Name
* Running time : Duration of the movie (in minutes)
* Budget: Budget of the movie (in dollars)
* Box Office: Box office of the movie (in dollars)
* Release date: Released date (datetime)
* imbd: imbd ratings

Subsequently, the data was further refined within the Python environment. Extreme values in both the budget (quartile = 0.1) and box office (quartile = 0.01) were removed. Additionally, any null values were dropped to prepare the data for analysis.

**Analysis**

From current dataset, Disney has released 431 movies in 80 years, making the average release movie amount per year to be 5.3875. The year with most movie releases is 2020 with 13 movies; The years with least movie releases is 1943, 1944, 1947, 1949, 1951, 1952, and 1987 with 1 movie.

We can see from below trend (time series plot), Disney has less released movies until 1990s, and starting from 1990s until now Disney maintain higher movie release amount.

*Amount of Released Movie Release by Year*

A graph showing the time line

Description automatically generated

The average running time for Disney movies is 99.03 with standard deviation of 18.11 minutes. The mximum is 168.0, and minimun of 61.0 minutes. The descriptive statistics of running time match the statement in Disney official website for their most well-known genre, animation, that generally has a 90-minute running time[[2]](#footnote-2).

A graph of a running time distribution

Description automatically generated

For the budget, the average is 63.91 million USD, with standard deviation of 70.54 million USD. the mximum is 410.60 by Pirates of the Caribbean: On Stranger Tides, and minimun of 1.35 million USD.

A graph of a budget distribution

Description automatically generatedA graph of a number of blue bars

Description automatically generated with medium confidence

For the IMDb (scale from 1 to 10), the average is 6.59, with standard deviation of 1.01,the maximum is 8.5 with The Lion King, and minimun of 2.4 with Hannah Montana and Miley Cyrus: Best of Both Worlds Concert.

A graph of a box office distribution

Description automatically generated

For the box office, the average is 21.29 million USD, with standard deviation of 27.80 million USD. When computing with cleaned dataset (excluding outliers), the mximum box office is 145 million USD for Frozen II, and minimun is 1 million USD for The Sword and the Rose. Most of the movies generate under 20 million USD, and very few generate more than 100 million USD.

However, if we don’t exclude 1% extreme value, the highest box office is The Lion King with 1657 million USD, which show how the ‘big hit’ in show business can impact its business. It is also clear how the distribution of box office, which can also be interpreted as revenue, in the bar chart is not showing normal distribution, indicating the revenue might be earned mainly through several hit movies out of all productions.

A graph of a box office

Description automatically generated

From previous descriptive statistic, we learn that The Lion King has the highest in both box office and IMDb rating. Does it simply imply with higher IMDb rate (movie reputation) lead to higher box office? When running correlation coefficients between each pair of variables, it shows highest correlation coefficient between box office and budget, follow by Box Office with IMDb.

A screenshot of a graph

Description automatically generated

To further analysis, perform with hypothesis test for linear regression between box office and budget, running time, and IMDb respectively. Setting significant level as 0.01 (accepting the relationship when p value < 0.01), we received below results:

|  |  |
| --- | --- |
| Box office vs | P Value |
| Running Time | 0.0001796119255926207 |
| IMDb | 1.4258771618660103e-09 |
| Budget | 1.6575245090313437e-38 |

All variables show significant relationship with box office. Among all, budget has the lowest p value, follow by IMDb and Budget, echoing the result of correlation coefficient. The result can also be showing in scatter plots as below.

|  |  |
| --- | --- |
| A graph of blue dots  Description automatically generated |  |
|  |  |

In the correlation coefficient chart, we can also see the high value between budget and running time, implying possible multicollinearity. Testing with same hypothesis test we can get the result that p value is 1.674610805780777e-08, hence we reject the hypothesis that β = 0. In another words, there is significant relationship between budget and running time.

**Summary**

The box office is showing linear relationship with budget, running time and reputation of the movie (IMDb). The most significant variables that influence box office is budget. While budget also has significant relationship with running time. Thus, if perform multiple regression analysis will require to check for multicollinearity.

The simple conclusion from this statistic suggests producing hit movies as much as possible. To achieve the goal, increases the budget might help with the chance for production to become a hit.

1. *Walt Disney Movies. (n.d.). www.kaggle.com. Retrieved October 13, 2023, from https://www.kaggle.com/datasets/dikshabhati2002/walt-disney-movies/* [↑](#footnote-ref-1)
2. Walt Disney Animation Studios. (n.d.). Filmmaking Process. Filmmaking Process. https://disneyanimation.com/process/ [↑](#footnote-ref-2)