MICROSOFT MOVIE ANALYSIS PROJECT

Link to GitHub Repository: https://github.com/Bonnie10/Microsoft-Project

1. BUSINESS UNDERSTANDING

1.1 UNDERSTANDING THE PROBLEM

The role assumed here is for a data scientist outsourced by Microsoft on the dos and don'ts while getting into the movie production sector. Thus, the project aims at providing directives on the possible lookout factors before entering the industry. The provision of a path on the vital element is imminent, from which deductions are made based on the evaluation of the factors. It answers some of the future questions regarding this entry on what type of movies the company should produce based on customer consumption. Also, the project evaluates the sales in markets for a transparent target market, either domestic or foreign, based on the movies' gross sales in both markets. Considering that profit making is a sole purpose in this organization, just like any other, a cost-benefit analysis concerning the production budget, foreign and worldwide gross sales conflict to give way forward on whether the firm should enter the market.

1.2 PROBLEM STATEMENT

The problem of this project is to find out whether Microsoft should enter the movie production industry or not. If yes, what is the relationship between some lookout factors, and which top movies should venture into the initial production to break even and compete with other companies in the industry?

2. DATA UNDERSTANDING

2.1 DATA COLLECTION

The firm provides the data used for this project. Different data types with various specifications are available for choosing the most relevant data type for making deductions. Looking into the data descriptions, the information contained, the shape, and the content of the initial and last columns of all datasets provided is critical in determining the relevant data for drawing the project's conclusions.

3. DATA SELECTING

3.1 SELECTING DATA

This project uses several columns of the various provided dataset. The budget, domestic and foreign gross columns are essential in the determination of whether the firm should enter the market or not. That is, whether the business is lucrative enough domestically to only rely on domestic supply and the extent of the foreign market to know the supply scope for profit maximization, the firms' sole aim. The popularity and studio-type columns are vital in knowing which movies are popular and the most lucrative studios to set up. The vote count column is crucial to draw an inclusive conclusion by mapping it against the popularity concept to deduce their relationship concerning what movies to produce.

3.2 DATA CLEANING

For validity and accuracy, the data used in this project is cleaned up by checking for null values in the relevant, usable columns. Although most of the data used in the project were initially clean, the few missing values are dealt with by replacing the missing values with the mean. Otherwise, one column is dropped for missing more than half the values, which results in using available data in deductions making.

4. DATA ANALYSIS

4.1 EXPLORATORY DATA ANALYSIS

4.1.1 UNIVARIATE DATA ANALYSIS

a. Numerical Data

There were a lot of missing values, 1350, in the foreign gross column for bommovie data. For the domestic gross that is used in this section after dropping the foreign gross column, just a few values are missing, 28. For this case, the replacement option works out of the normality in the dataset distribution. There were no missing values in the rest of the used data. There was no need for duplicate clean-up as there were non, and the numerical data used for this is not sensitive. Thus, there is room for duplicates.

b. Categorical Data

This project's obsession lies in the market outbursts; thus, the top figures matter most. For instance, the statistics in domestic gross are central to the choice of best performing studio. In such a case, there is no missing categorical data for the required data set. Therefore, no issues with missing categorical data in this case. Duplicates are allowed in this case out of the possibility of a studio producing different titled movies in different years. The case is no different with all other categorical data used in this project.

c. Summary statistics.

i. In some instances, 2 out of the 5 used, the budget of producing a movie exceeds the domestic gross income. For instance, pirates' budget = 410,600,000 vs domestic gross = 241,063,875. In other instances, the profits earned are merely above the budget. For instance, avengers' budget = 330,600,000 vs domestic gross = 459,005,868.

ii. At no instance does the budget surpass the worldwide gross income. At some point, the worldwide profit is supernormal. For instance, Avatar's budget is $ 425,000,000 vs worldwide gross of $ 2,776,345,279.

iii. The BV studio ranks highly with regard to domestic gross, 1.841903e+10.

iv. As per the tmd dataset, the box, Home, The Gift, Eden and Lucky are the highest voted movies with 39.5, 39.4, 37.1, 34.6, 34.4 votes respectively.

d. Univariate analysis recommendation

There is a clear arrangement of all variables being tested in ascending orders. This distinction makes it easy for the firm to choose everything it intends to invest on.

4.1.2 BIVARIATE ANALYSIS

a. Numeric

There is a low positive correlation between popularity of a movie and vote count.

b. Categorical

The relationship between popularity and vote counts is weakly positive. A movie being popular is not a guarantee that people will vote it most.

5. CONCLUSION

The cost-benefit analysis compares the amount spent producing a movie and its domestic and foreign grosses. For which studio to set up, the project uses the most profitable studios in the domestic market out of the missing nature of the foreign gross values. The case applies to the movie's popularity, unlike the number of votes that is lowly positively correlated with popularity.

6. RECOMMENDATIONS

Basing the arguments from the deductions above, Microsoft should enter the movie production industry to boost its profits. However, it should look for a market for its products away from its country since the gain in domestic sales is insufficient and may sometimes drive the organization into losses. Thus, expansion to the global mindset is key to entering the movie production sector.

Also, among the many studios, it should invest in, the BV studio is the most sufficient to boost the company’s profit. Despite the unpromising nature of the domestic market, as realized above,

this studio gives the highest payoffs at the domestic level. Such high domestic incomes would alternatively promise more at the worldwide level where the movies are selling highly.

Again, there are movies that Microsoft must prioritize to achieve highly desired sales. These movies are the top movies, especially in terms of popularity. They are; Blade Runner 2049, Thor: Ragnarok, Avengers: Infinity War, John Wick, and Big Hero 6. Here, there is the need to overlook the voting concept since it is just slightly correlated with popularity.