TABLE OF CONTENTS

DETERMINING THE MOST APPROPRIATE FILMS FOR A MOVIE STUDIO.

1. BUSINESS UNDERSTANDING

A movie studio is an entertainment company with private facilities to produce films which are handled by a production team. A film on the other hand is an event or story recorded as set of images in motion and shown in a cinema hall or on television. The development of a movie studio is normally in line with a concept based on an event or idea (Vogel, 2001). This evidently shows that different types of films play a critical role in ensuring a movie studio thrives in the technical sector. Caves (2001), argues that for a movie studio to make profits quality devices have to be used to ensure production of quality films. A financial basis is therefore crucial when developing a movie studio since a producer is required to have an idea of how production costs will be recovered. An evaluation on the performance of existing production companies to identify consumer preferences generates insights that promote the success of a new movie studio (Dahan and Hauser, 2002). This shows that the film industry is a social component established by how regularly consumers appreciate a certain film to an extend of referring others to watch it and sparing time aside from their normal duties to concentrate on the concept of a film. To discover the most effective film for a production company our focus will be on income generated by existing movie studios, customer preferences on various films and the types of films available.

1.2 PROBLEM STATEMENT

This research involves the use of variables to determine which types of films are appropriate for developing a movie studio. To investigate on the best films for a production company the research aim towards working with various variables which include; types of films, consumer preferences and income generated by existing movie studios. The main objective is to determine how these variables contribute to huge incomes in the production sector.

2. DATA UNDERSTANDING

2.1 DATA COLLECTION

To determine which films are appropriate for a movie studio, secondary data from various datasets was used; A CSV file from bom.movie_gross.csv.gz and a SQLite database where we retrieved movie basics and movie ratings tables. The required library for our analysis and data from various databases was imported into our Jupyter notebook working space. This was done to retrieve our columns of interest from the variety of datasets available and to understand what our data of interest comprises of. The data was checked for inconsistencies and missing data.

2.2 DATA PREPARATION

The two tables of interest to our study from the SQLite database were joined to retrieve our columns of interest. To prepare our data for analysis, we checked for the missing values available in our dataset and rectified them accordingly. The data had continuous variables which had missing values which was rectified by replacing with the median value. A median

has minimal effect on the overall distribution of a dataset. The categorical variables that had missing values were rectified by filling the values with a text to show the values were missing. The data was then checked to ensure the missing values were eliminated.

2.2.1 DATA DESCRIPTION
The variables of interest to our study were selected to determine how they contribute to the success of a production sector. The types of films available was measured using genres, customer preferences was measured using number of votes and income generated by existing movie studios were measured using domestic gross.

3. DATA ANALYSIS

3.1 UNIVARIATE DATA ANALYSIS
We aim towards conducting various analysis to investigate on the most appropriate films for a movie studio. Univariate Data Analysis was performed to describe each variable where investigation was done on the central tendency and range of values.

**3.1.1 Univariate Analysis on consumer preferences**
Figure 3.1 Box plot on consumer preferences
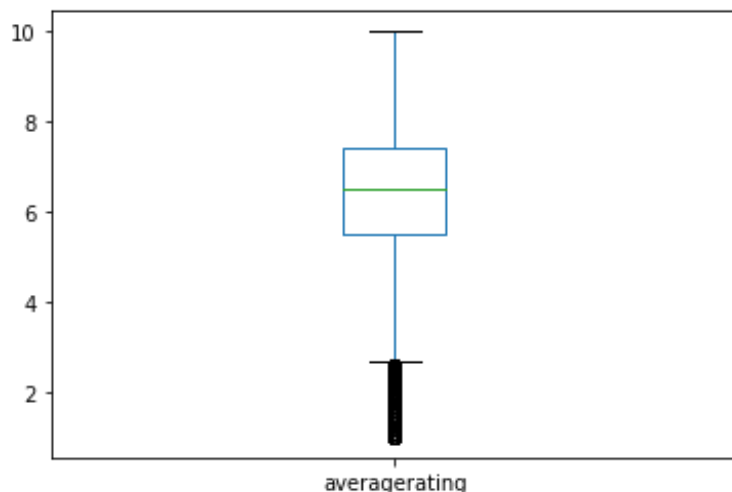
The box plot on figure 3.1 shows that there are different views between consumers on the films they prefer. This proves that the different opinion between consumers is worth to be considered in our analysis. The box also displays a symmetric hence normal distribution since the median lies in the middle of the whiskers which are almost the same on both sides of the boxplot.

3.1.2 Univariate analysis on income generated
Figure 3.2 Box plot on Income generated by existing movie studios
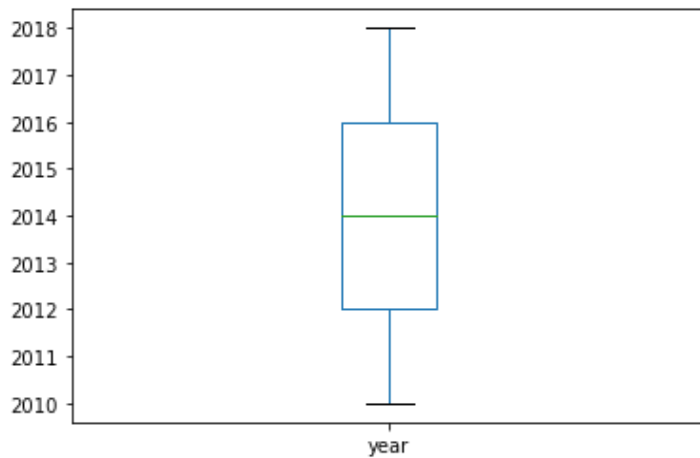
The boxplot on figure 3.2 shows that there are different amounts earned by existing movie studios per year which mainly depends on the films produced. The whiskers are almost the same size on both sides of the box hence the distribution is normal. This is evidently supported by the median which lies in the middle of the boxplot.

3.1.3 Summary Statistics
Table 3.1 Summary Statistics

	Consumer Preferences	Income generated
Mean	3.52	2.87
Median	4.90	1.40
Standard deviation	3.03	6.69

According to table 3.1 consumer preferences have a mean of 3.52 and a standard deviation of 3.03. This indicates that the data is reliable for our analysis since it is not highly spread out as indicated by the standard deviation value. Income generated has a mean of 2.87 and a standard deviation of 6.69 which indicates that the data is highly spread out in reference to the standard deviation value.

3.2 BIVARIATE DATA ANALYSIS
Bivariate analysis is an analysis conducted to determine the relationship that exists between two variables. This analysis implemented the use of scatter plots to determine the relationship that exists between our variables of interest.

3.2.1 Consumer preferences and types of films
According to Figure 3.3, many consumers prefer films based on drama, documentary and comedy. The least preferred films are those based on action, thriller and animation. A movie studio that therefore wishes to start up should most likely major on producing drama, documentary and comedy films.

Fig 3.3 Consumer preferences on various types of films
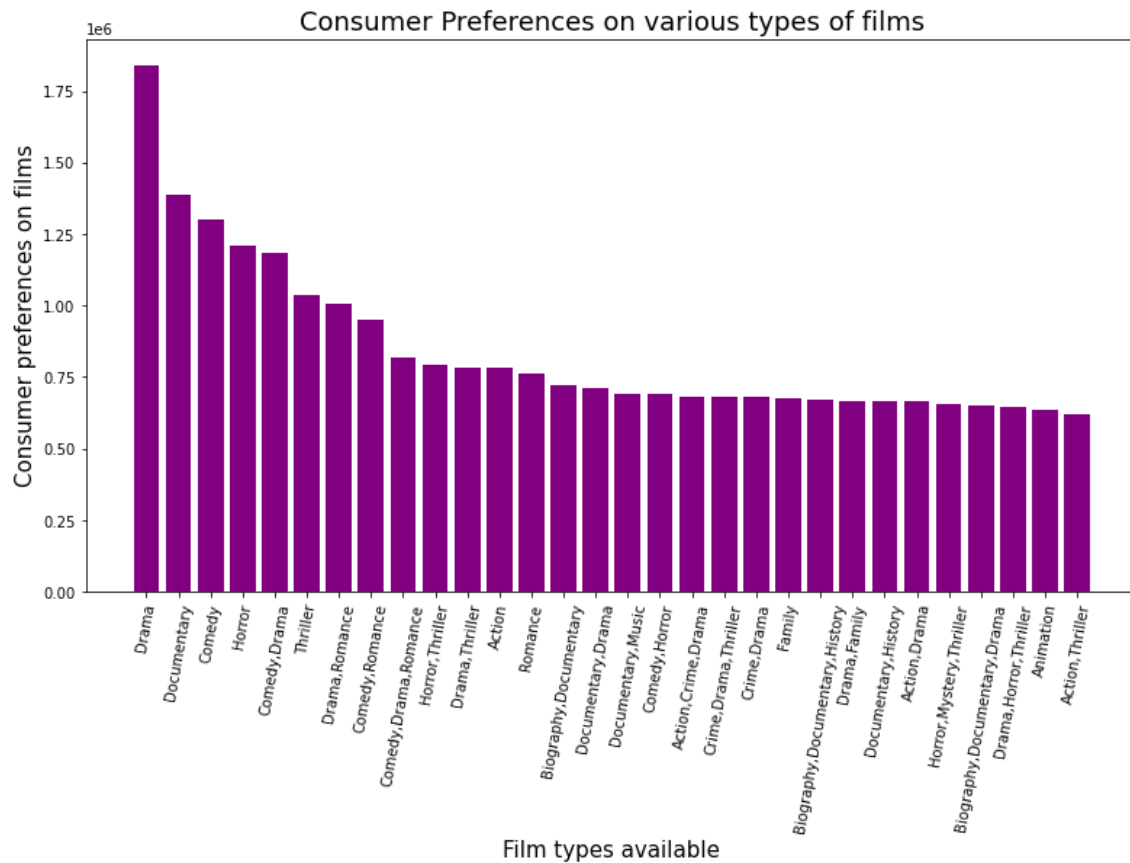
3.2.2 Incomes generated by existing movie studios on various films

Figure 3.4, evidently shows that existing movie studios are generating huge incomes from drama, documentary and comedy. This is an indication that a movie studio that majors on these films is likely to make huge profits. Films based on action, thriller and animation are associated with a low income considering the fact that they accumulate fewer views. The plot also explains the association that consumers have towards the production sector. The higher the views for a particular film, the higher the income generated by movie studios while the lower the views the less the income generated.

Fig 3.4 Income generated from various films by existing movie studios

Income generated from films by existing movie studios
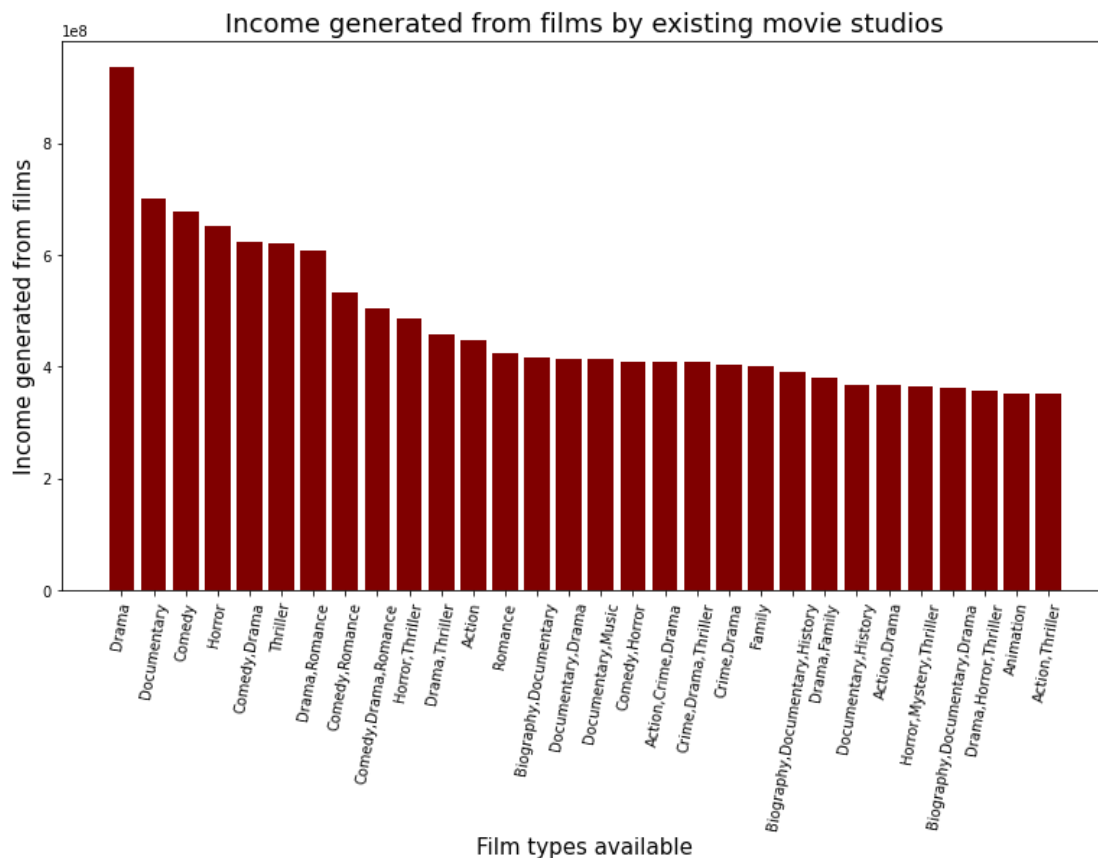
3.2.3 Correlation between types of films and income generated

Fig 3.5 shows that there exists a perfect negative correlation between films and income generated. This proves that the more a film is watched the higher the income generated by movie studios while the less a film is watched the lower the income generated by movie studios. Movie studios have a responsibility to be keen on consumer preferences when producing films in order to avoid generating losses.

Fig 3.5 Correlation between income generated and types of films
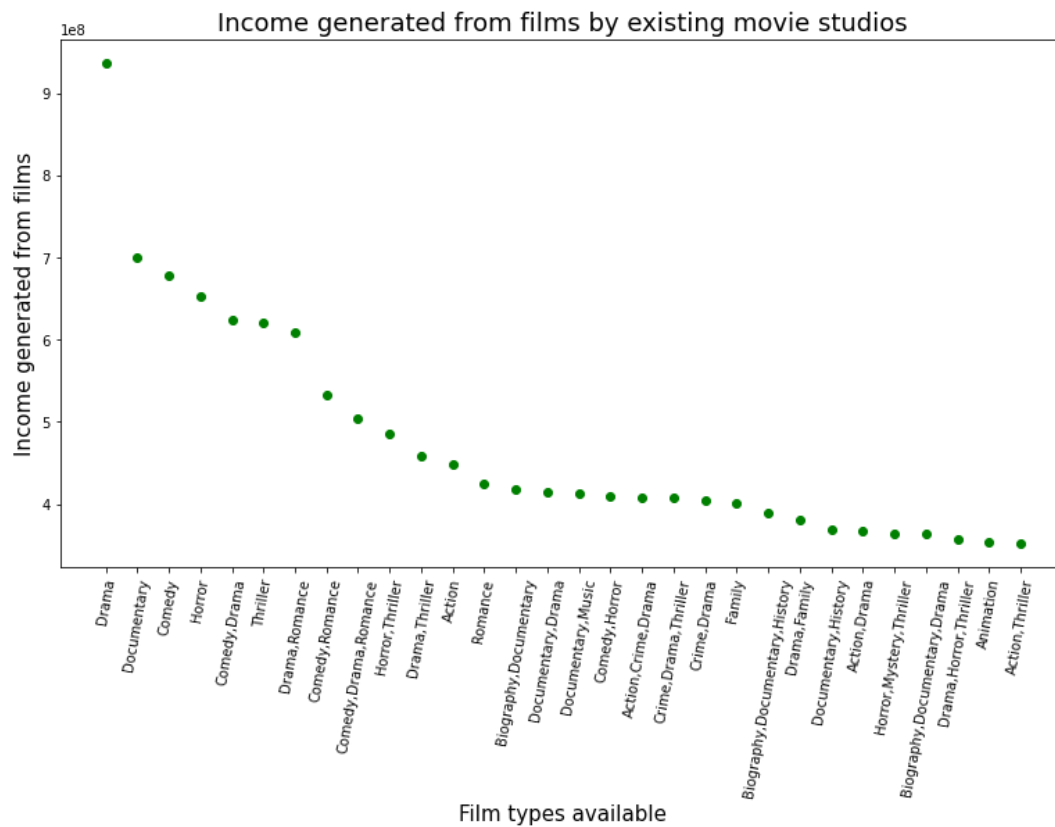
3.2.4 Consumer Preferences on various films

Fig 3.6 displays a positive negative correlation between consumer preferences and films. This shows that the higher the interest in a certain film by a consumer, the higher the chances of watching the film while the lower the interest in a certain film, the lower the chances of watching it. This shows that consumers play a huge role in promoting the production sector since they are the major determinants of what is supposed to be aired.

Fig 3.6 Correlation between consumer preferences and film types
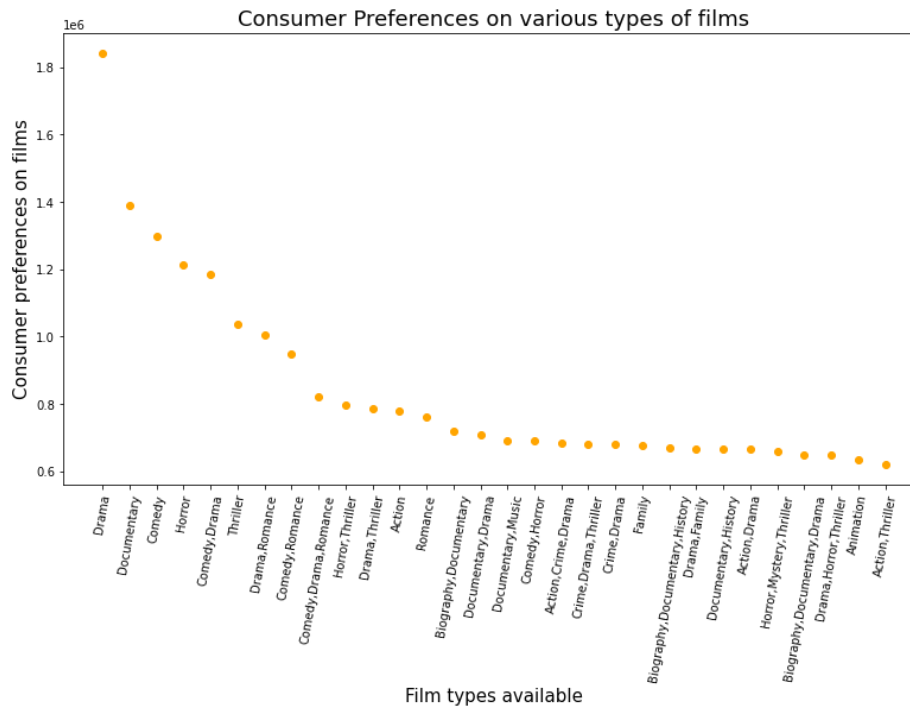
4. EMPIRICAL FINDINGS

This study aimed at conducting an analysis to determine the appropriate films a production company can use to start up a movie studio. The study adopted the use of secondary data to conduct exploratory data analysis to investigate on the relationship between our variables of interest. The analysis shows that the data used was normally distributed and a positive negative relationship exists between types of films, income generated from films and consumer preferences. Films produced by existing companies based on drama, documentary and comedy are successful. A successful film is one that generates high revenues for a production company (McKenzie, 2010). Films based on action, thriller and animation generate low incomes for existing movie studios. The recipients of films produced play a huge role since a film that accumulates less views translates to losses in the production sector. The regular a film is viewed, the more the demand and the higher the feedback for a movie studio in terms of huge revenues. The success of a movie studio requires investment in quality devices to ensure quality production. A film is mainly demanded due to the quality and content displayed therefore to start up a movie studio a stable financial basis is important. This is in line with the findings of Yang (2010) which states that a low quality film is associated with fewer views that can be dealt with by checking on consumer preferences.

4.1 CONCLUSION

This research shows that there exists a relationship between film types, consumer preferences and income generated by movie studios. Films based on drama, documentary and comedy are the best option for film production since existing movie studios are generating huge incomes

through accumulated views by consumers. The higher the content from these films is viewed by consumers, the higher the income earned by production companies.

5. RECOMMENDATIONS
A movie studio should implement film production based on drama, documentary and comedy in order to generate high revenues since they are likely to gain views on their production. Consumers play a huge role in ensuring the success of movie studios hence regular surveys should be conducted to investigate on their interests in film content. The use of technology advanced devices when conducting film production to ensure the content displayed is of high quality.

5.1 RECOMMENDATIONS FOR FURTHER STUDY
There is a recommendation that further study should be conducted to figure out what influences consumers to opt for certain films. This will be beneficial to film producers since they will be able to know the content they are required to input in films. The study can focus on variables not tackled in our study.

Belinda. M. Nyamai,
belinda.nyamai@gmail.com
Data Science student, Moringa School.

REFERENCES

Caves, Richard E. (2001). Creative Industries: Contracts between Art and Commerce. Harvard University Press, Cambridge: MA

Dahan, Ely and John R. Hauser (2002). The Virtual Customer. Journal of Product Innovation Management, 19(5), 332-353.

McKenzie, J. (2010) How do theoretical box office revenues affect DVD retail sales? Australian empirical evidence. *Journal of Cultural Economics,* 34(3):159-179.

Vogel, Harold L. (2001). Entertainment Industry Economics. Cambridge University Press, Cambridge: New York.

Yang, J. (2010) Timing of effort and reward: Three-sided moral hazard in a continuous-time model. *Management Science 56(9)*: 1568-1583.