MICROSOFT-MOVIE ANALYSIS PROJECT

BENTA W. IRUNGU
5TH FEBRUARY,2023

PROJECT'S SUMMARY

The project used exploratory data analysis on the movie data with an aim of providing recommendation to the microsoft new production studio on what movies to produce as a new entrant. The analyais was based on the concepts below;

- Production cost: This is in reference to the production budget of different movies
- Studio Preference: This is in terms of the most preffered studios based on their perfomances which can be used for benchmarking
- Income generation: the purpose of anybusiness is to generateincome from production, thus I included this aspect in the analysis
- Mostly produced: I analyzed the genres which are mostly produced to help crosscheck whether the same is relevant when determining what movies to produce

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

Microsoft company needs to know which movies to produce. In order to make the recommendation, an analysis and interpretation of the data in the market is necessary. This analysis used the below objectives;

a. To determine the most profitable studios which can be used for benchmarking
b. To determine what movie genres are mostly produced
c. To determine the cost of producing the most profitable movies

Data

1. imdb.title.basics -This is a csv file used in the analysis it has 146144 observations and 5 columns . See the image to get an idea of how the data set looks like.

```
title_basics.head()
```

tconst	primary_title	original_title	start_year	runtime_minutes	genres
tt0063540	Sunghursh	Sunghursh	2013	175.0	Action,Crime,Drama
tt0066787	One Day Before the Rainy Season	Ashad Ka Ek Din	2019	114.0	Biography,Drama
tt0069049	The Other Side of the Wind	The Other Side of the Wind	2018	122.0	Drama
tt0069204	Sabse Bada Sukh	Sabse Bada Sukh	2018	NaN	Comedy,Drama
tt0100275	The Wandering Soap Opera	La Telenovela Errante	2017	80.0	Comedy,Drama,Fantasy

Data

2. imdb.title.ratings- this is also a csv file which contains 73856 observations and 2 columns. See the image to get an idea of how the data set looks like.

```
title_ratings.head()
```

tconst	averagerating	numvotes
tt10356526	8.3	31
tt10384606	8.9	559
tt1042974	6.4	20
tt1043726	4.2	50352
tt1060240	6.5	21

Data

3. bom.movie_gross – This is a csv data set with 3387 entries and 4 columns. See the image to get an idea of how the data set looks like.

```
movie_gross.head() #opens the first 5 observations
```

title	studio	domestic_gross	foreign_gross	year
Toy Story 3	BV	415000000.0	652000000	2010
Alice in Wonderland (2010)	BV	334200000.0	691300000	2010
Harry Potter and the Deathly Hallows Part 1	WB	296000000.0	664300000	2010
Inception	WB	292600000.0	535700000	2010
Shrek Forever After	P/DW	238700000.0	513900000	2010

Data

4. tn.movie_budgets- This is a csv data set with 5782 observations and 5 columns. See the image to get an idea of how the data set looks like.

```
movie_budgets.head()
```

id	release_date	movie	production_budget	domestic_gross	worldwide_gross
1	Dec 18, 2009	Avatar	$425,000,000	$760,507,625	$2,776,345,279
2	May 20, 2011	Pirates of the Caribbean: On Stranger Tides	$410,600,000	$241,063,875	$1,045,663,875
3	Jun 7, 2019	Dark Phoenix	$350,000,000	$42,762,350	$149,762,350
4	May 1, 2015	Avengers: Age of Ultron	$330,600,000	$459,005,868	$1,403,013,963
5	Dec 15, 2017	Star Wars Ep. VIII: The Last Jedi	$317,000,000	$620,181,382	$1,316,721,747

DATA SET 'movie_gross' PREPROCESSING

Methods

This analysis used the specified data with an aim of answering the specified objectives of the analysis.

Data preparation

Data used in this analysis was prepared for analysis through;

- Checking the missing values using the **.isna()** function.
- Replacing missing data using a **forward fill**
- Replacing missing data with the median using the **.fillna()** function
- Removing the strings objects such as dollar signs and commas from numerical data to allow for statistical computations
- Used the **.astype()** function to convert objects to either integers or float for analysis

Methods

Data analysis

In order to attain results data was analyzed as follows;

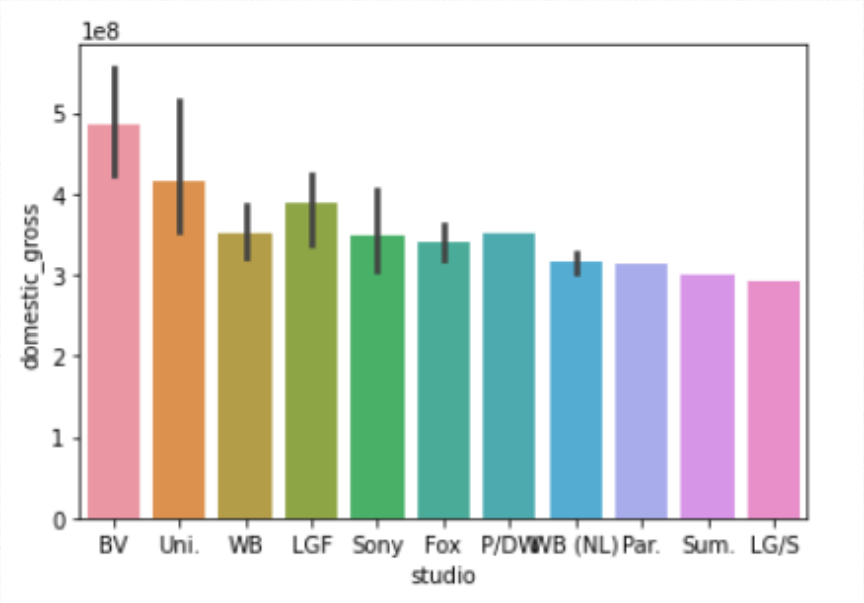
- used the movie_gross data set to determine the most profitable studios which will recommend studios for benchmarking and highlight the industry's top competitors
- Used the movie_budgets data to determine the most profitable movies so as to have an idea of the preference of the target market.
- Used the movie_budgets data to determine the movies that are cost friendly to produce
- To determine the most produced movie genres using the title basics data so as to help make a decision on what movie to produce

With the above analysis, it will be possible to identify what movies to produce, how to produce it and what budget range to use to attain profit. It will also help get a clear understanding of the competitors ability and also have an idea of the most preferred movies by the consumers.

RESULTS

1.**Determining The Most Profitable Studios Results**

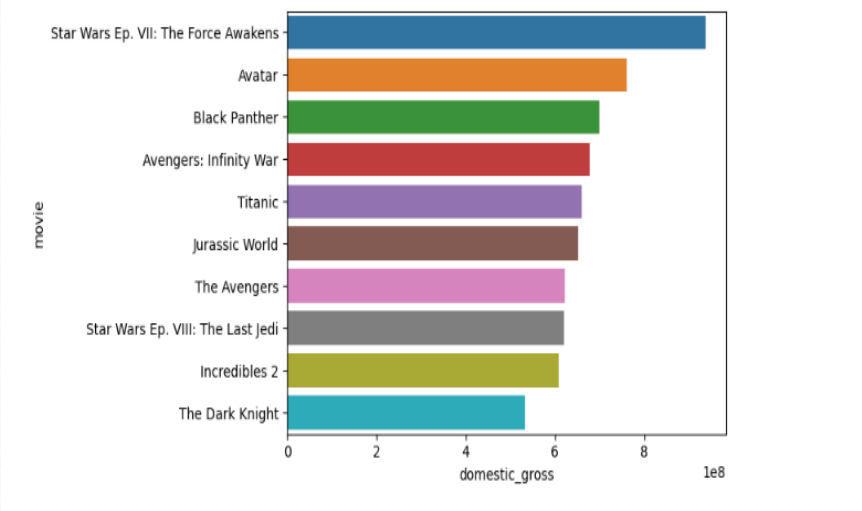
This analysis used movie gross data to analyze the most profitable studios on the basis of the domestic gross. 837 studios out of 3387 had a gross domestic higher than the average domestic gross. The top ten were visualized using a bar plot as illustrated



RESULTS

2. Determining The Most Profitable Movies

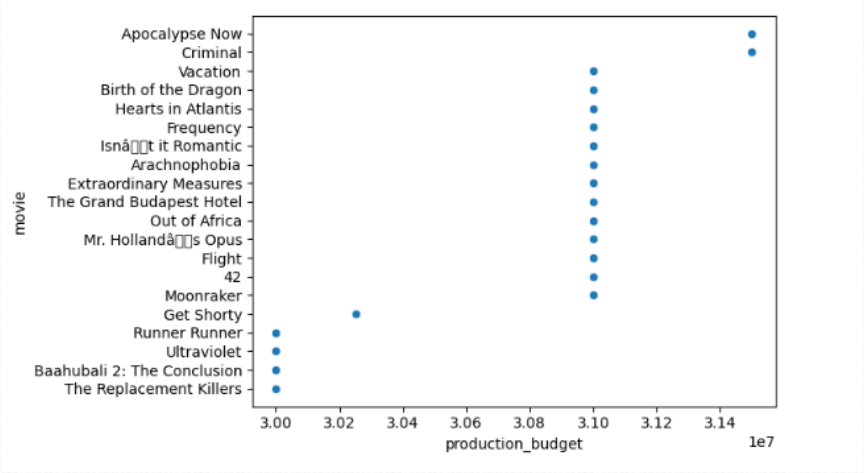
This analysis used movie budget data to analyze the most profitable movies on the basis of the domestic gross. 837 movies out of 3387 had a gross domestic higher than the average domestic gross. The top ten were visualized using a bar plot as illustrated



RESULTS

3. Determine The Movies That Are Cost Friendly To Produce

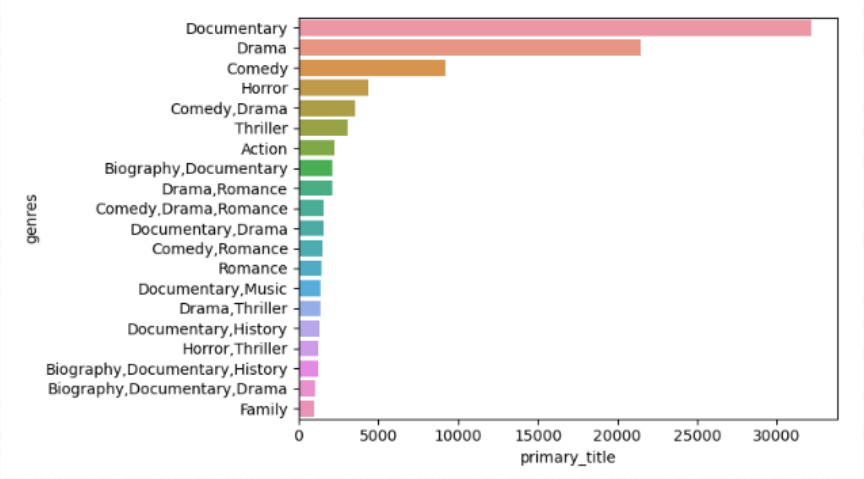
This analysis used movie budget data to analyze movies that are cost friendly(where cost effective movies were those whose production cost was less than the mean of the total production cost. The results were arranged and visualized using a scatter plot. The results indicate that 3969 movies budget (69%) was less than the total mean.



RESULTS

4. Determine Most Produced Movie Genres

This analysis used movie basics data to analyze movies genres that are mostly produced. The results were arranged and visualized using a barplot showing the top ten mostly produced genres as illustrated



Conclusions

From this analysis we can make the following conclusions

- Studios BV, Uni, WB, LGF, Fox studio, P/DWWB, NL, Par, Sum and LG/S has produced movies with the highest domestic gross making them best suitable for benchmarking
- The top ten movies with high domestic gross represent most profitable movies. the Microsoft company can choose to produce movies similar to these ones in order to attain more profits from the same
- There are 2677 movies whose production costs are less than the domestic gross which translates into profits. Microsoft can chose to produce movies which require lesser production cost than the top perfuming movies and still make profits. this option will optimize on resources and enable production of multiple movies using reasonable production budgets

From the analysis of the genres mostly produced, Microsoft should therefore consider producing movie of the top ten genres mostly produced

Recommendations

From this analysis we can make the following recommendations;

- Microsoft new production studio should endeavor to learn how the most profitable studios produce their movies in order to attain profitability
- The top movies has the highest income generation as per this analysis. Microsoft should consider producing movies which with the concept of the top movies.
- The first 3 top movies are Star wars :force awakens whose genre is science fiction/action, Avator movie whose genre is science fiction and Black panther movie whose genre is action/adventure. Microsoft should produce science fiction or action movies in order to generate more income.
- There are 2677 movies whose production costs are less than the domestic gross which translates into profits. Microsoft can chose to produce movies which require lesser production cost than the top perfuming movies and still make profits. this option will optimize on resources and enable production of multiple movies using reasonable production budgets

Future Considerations

The following aspects should be considered in the future analysis

- Analyze the impact of the movie release date on income generation and profitability
- Analyze the impact of the movie directors on the income generation and profitability

Thank You!

Email: bentairungu@gmail.com
GitHub: b-irungu
LinkedIn: https://www.linkedin.com/in/benta-irungu-a6358513a