Microsoft Studios Analysis Project 1

Stella Kitur - DSF-FT-04

Table of contents

01

Business Problem & Data
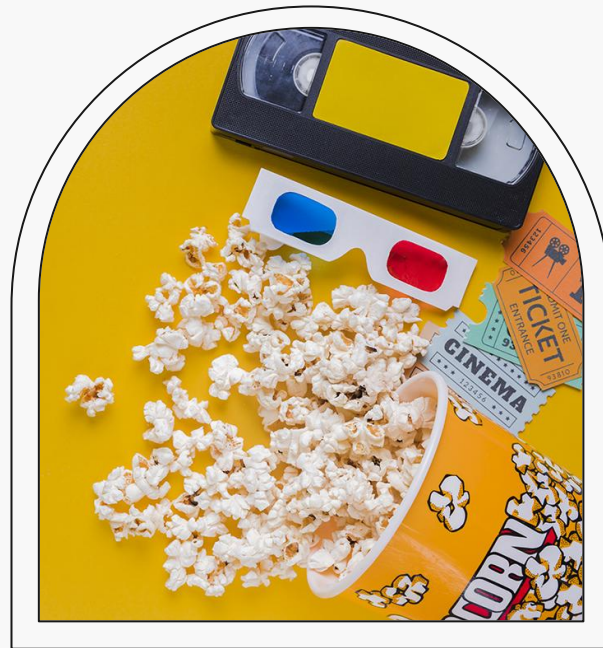
Business Problem & Research Questions

Microsoft wants to establish a successful movie studio that produces original video content that meets the audience's needs and generates substantial profits.

As an analyst, I was tasked to identify the factors that contribute to box office success and provide insights to guide decision-making.

In order to complete this task, these were the 4 main questions addressed.

1. What were the top 5 rated studios?
2. What are the top rated genres?
3. What were the most profitable movies?
4. Who were the top rated directors?

Datasets

There were 3 main datasets that were used in this exploratory data analysis.

These included:
1. Box Office Mojo
2. The Numbers
3. IMDB

Box Office Mojo

	title	studio	domestic_gross	foreign_gross	year
0	Toy Story 3	BV	415000000.0	652000000	2010
1	Alice in Wonderland (2010)	BV	334200000.0	691300000	2010
2	Harry Potter and the Deathly Hallows Part 1	WB	296000000.0	664300000	2010
3	Inception	WB	292600000.0	535700000	2010
4	Shrek Forever After	P/DW	238700000.0	513900000	2010
...	...	...	...	...	...
3382	The Quake	Magn.	6200.0	NaN	2018
3383	Edward II (2018 re-release)	FM	4800.0	NaN	2018
3384	El Pacto	Sony	2500.0	NaN	2018
3385	The Swan	Synergetic	2400.0	NaN	2018
3386	An Actor Prepares	Grav.	1700.0	NaN	2018

3387 rows × 5 columns

`bom_movies_gross`

[6]

domestic_gross	
count	3359.00
mean	28745845.07
std	66982498.24
min	100.00
25%	120000.00
50%	1400000.00
75%	27900000.00
max	936700000.00

When the data was loaded there were 3387 rows and 5 columns.

These are the summary statistics of the dataset

The Numbers

numbers_data

[7]

	id	release_date	movie	production_budget	domestic_gross	worldwide_gross
0	1	Dec 18, 2009	Avatar	$425,000,000	$760,507,625	$2,776,345,279
1	2	May 20, 2011	Pirates of the Caribbean: On Stranger Tides	$410,600,000	$241,063,875	$1,045,663,875
2	3	Jun 7, 2019	Dark Phoenix	$350,000,000	$42,762,350	$149,762,350
3	4	May 1, 2015	Avengers: Age of Ultron	$330,600,000	$459,005,868	$1,403,013,963
4	5	Dec 15, 2017	Star Wars Ep. VIII: The Last Jedi	$317,000,000	$620,181,382	$1,316,721,747
...	...	...	...	...	...	...
5777	78	Dec 31, 2018	Red 11	$7,000	$0	$0
5778	79	Apr 2, 1999	Following	$6,000	$48,482	$240,495
5779	80	Jul 13, 2005	Return to the Land of Wonders	$5,000	$1,338	$1,338
5780	81	Sep 29, 2015	A Plague So Pleasant	$1,400	$0	$0
5781	82	Aug 5, 2005	My Date With Drew	$1,100	$181,041	$181,041

5782 rows × 6 columns

	id
count	5782.00
mean	50.37
std	28.82
min	1.00
25%	25.00
50%	50.00
75%	75.00
max	100.00

When the data was loaded there were 5782 rows and 6 columns.

The values in production_budget all the way to worldwide_gross were later cleaned in order to be a **float data type.**

These are the summary statistics.

The IMDB

	primary_title	original_title	start_year	runtime_minutes	genres	averagerating	numvotes	person_id	primary_name
0	Sunghursh	Sunghursh	2013	175.0	Action,Crime,Drama	7.0	77	nm0712540	Harnam Singh Rawail
1	Sunghursh	Sunghursh	2013	175.0	Action,Crime,Drama	7.0	77	nm0712540	Harnam Singh Rawail
2	Sunghursh	Sunghursh	2013	175.0	Action,Crime,Drama	7.0	77	nm0712540	Harnam Singh Rawail
3	Sunghursh	Sunghursh	2013	175.0	Action,Crime,Drama	7.0	77	nm0712540	Harnam Singh Rawail
4	One Day Before the Rainy Season	Ashad Ka Ek Din	2019	114.0	Biography,Drama	7.2	43	nm0002411	Mani Kaul

The data that was loaded from the IMDB database.

The dataset was quite large, with 181,387 rows and 9 columns
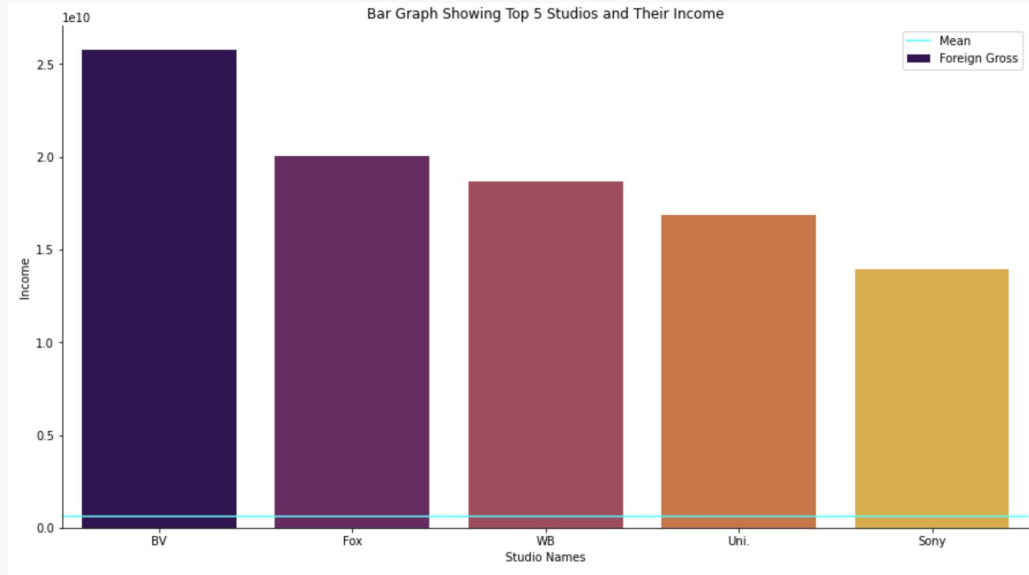
02

Methods & Results

Methods

The following steps were taken in order to prepare the datasets for analysis in order to answer the research questions.

- Removed unnecessary columns & combined data from different sources
- Checked for any missing information & unusual data points
- Filled in the missing information and dealt with the unusual data points
- Select appropriate methods based on the type of data & research questions
- End result: Only the most relevant information included.

The results are presented in the preceding slides.

Top Studios and Income

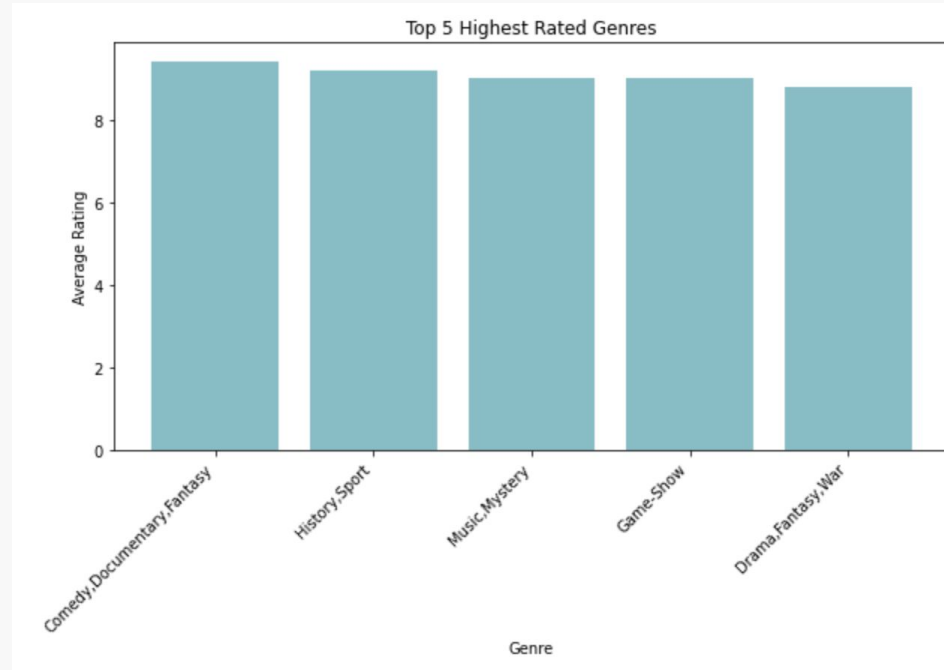Bar Graph Showing Top 5 Studios and Their Income

Analysis

This bar chart illustrates the top studios based on the income.
BV- Buena Vista is the highest studio based on the dataset, while Sony is the lowest compared to the others.
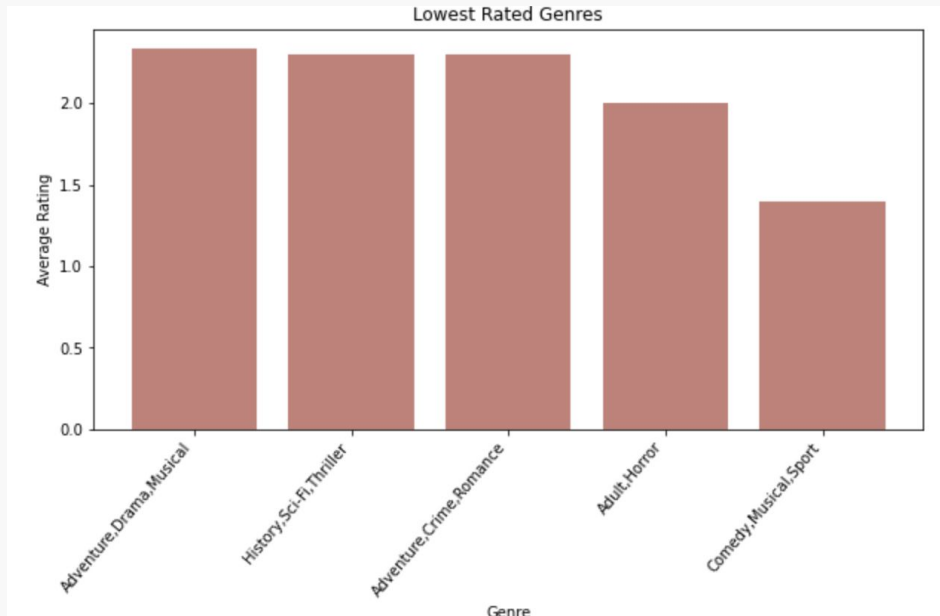
Highest rated Genres

Top 5 Highest Rated Genres

Analysis

This bar chart illustrates the highest rated genres. With the highest being the *comedy, documentary, fantasy* genre.
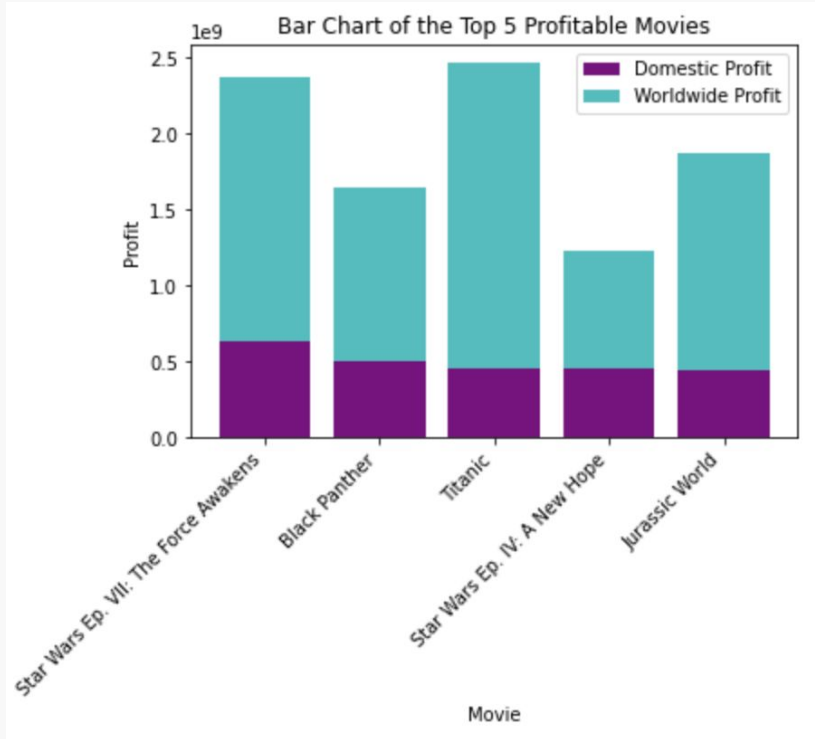
Lowest rated Genres

Analysis

This bar chart illustrates the lowest rated genres. With the lowest being the *comedy, musical, sport* genre that had an average rating of 1.5 .

Profitable Movies

Analysis

The bar chart illustrates that the worldwide profit for all the movies were significantly greater compared to the domestic profits.

Top 5 Directors & Movies

Lindsay Thompson

Fly High: Story of the Disc Dog

Ivana Diniz

Exteriores: Mulheres Brasileiras na Diplomacia

Emre Oran

Calamity Kevin

Loreto Di Cesare

Revolution Food

Masahiro Hayakawa

The Paternal Bond: Barbary Macaques

03

Recommendations

Recommendations

Focus on top rated genres

Prioritizing genres with higher ratings can help ensure content is well-received by an audience, potentially improving customer satisfaction, as Microsoft may want to explore this approach for their content.

Focus on worldwide audience

The highest profits were generated by the worldwide gross.

To increase profits, it's better to target a global audience instead of just focusing on the local audience.

Create with top rated directors

I would recommend that Microsoft considers directing movies with those that have a higher rating due to the quality of work that is produced by them.

Understanding the competition

Based on the analysis carried out,

I would recommend that further research is done to determine why the top studios are at the top.

Thanks!

E-mail: stellaacherotich@gmail.com
Linkedin: www.linkedin.com/in/stella-cherotich/
GitHub: www.github.com/stellacherotich/