



## Microsoft Movie Studio -A Moringa Data science project

BY PETE NJAGI • [HTTPS://WWW.LINKEDIN.COM/IN/PETEDIGITAL/](https://www.linkedin.com/in/petedigital/)





## Section 1: Overview + Business Understanding

# Before I pitch:

---



## For additional contact me

My name is Pete Njagi

My work email is [Petedigital@hotmail.com](mailto:Petedigital@hotmail.com)

My work phone number is +254 728 752846

I am a Moring Data Science Student and professional software developer

My aim is to: Secure your interests and to potentially raise capital in funding the film



# Overview

---

1. Beginning
  - Overview
2. Business Understanding  
Middle
  - Data Understanding
3. Data Analysis  
End
  - Recommendations
  - Next Steps
  - Thank You



# Business Understanding

---

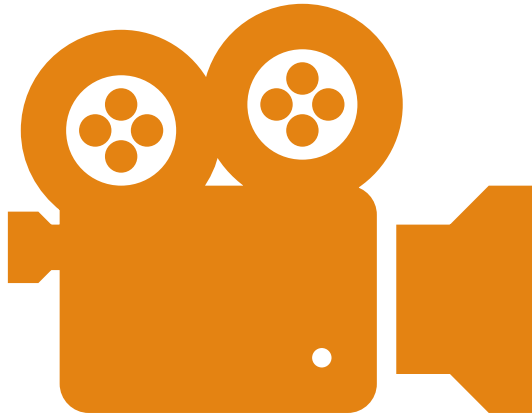
## Introduction

This movie studio pitch is a presentation where I use data on previous films to pitch ideas to a movie studio division to Microsoft for funding and production support.

This pitch is will also be delivered in the form of a slide presentation, highlighting the key elements of the market analysis, and potential audience appeal.

The goal of the pitch is to convince the Microsoft's movie studio of a film project is worth investing in and has the potential for commercial success.

The primary objective is to empower the head of Microsoft's new movie studio with actionable insights derived from a comprehensive analysis of existing trends, enabling informed decision-making on the types of films to produce. This exploration involves understanding key factors contributing to a movie's success, such as genres, release dates, and production budgets.



# Objective

---

My key areas is to pitch the new Microsoft Movie Studio about:

- 1. The type of film they should be making (using data)
- 2. The ideal movie budget
- 3. The ideal director they should hire

# Section 2: Middle.

Data Understanding  
Data Analysis

---



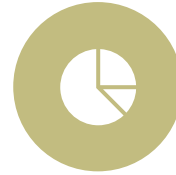
**1. BUSINESS  
UNDERSTANDING**



**2. DATA  
PREPARATION**



**3. DATA  
UNDERSTANDING**



**4. DATA ANALYSIS**



**5. VISUALIZATION**



# Data Understanding

---

In this repository under the file path `` .csv, and .sqlite`` is file data containing the Project dataset. This is the same dataset many filmmakers, and movie studios use to make informed decisions in the film industry.

The features of interest for this analysis will be:

**`**IMDB SQLite3**:`** ``Data from the IMDB Database``

**`**tn.movie_budgets.csv:**`** ``Movie grade (data) including production budgets, and gross revenue``

**`**bom.movie_gross.csv:**`** ``Financial Data on various movies``

**IMDb**





## Data Analysis

---

With a comprehensive understanding of the dataset, the analysis phase will delve deeper into identifying factors that significantly impact a movie's success. Correlation analyses will be conducted to assess relationships between variables, while statistical tests may be employed to validate hypotheses. Machine learning models, if applicable, may be employed to predict box office performance based on selected features. The analysis will aim to uncover patterns, preferences, and trends that can guide decision-making for Microsoft's movie studio.

# Data Understanding: Finding the most popular movies and patterns

### Data Understanding: Finding the most popular movies and patterns

In this section I aim to find what the most popular moves are, in addition I want to find the most highly rated popular movies

So, I shall query the IMDB SQL database to find

**\*\*Highly rated movies\*\***: `Movies with a rating of 8.5 (85% rating and higher)`

**\*\*Popular Movies\*\***: `Movies more than 5000 votes`

# Definition of Blockbuster films that Microsoft Movie Studio's should develop

---

## Blockbusters

Because Microsoft is a large technology company, we should aim to producer high quality films that Microsoft would be proud..

These **are highly popular** films with good ratings

Market Summary > Microsoft Corp

**2.73 trillion** USD

Market capitalization

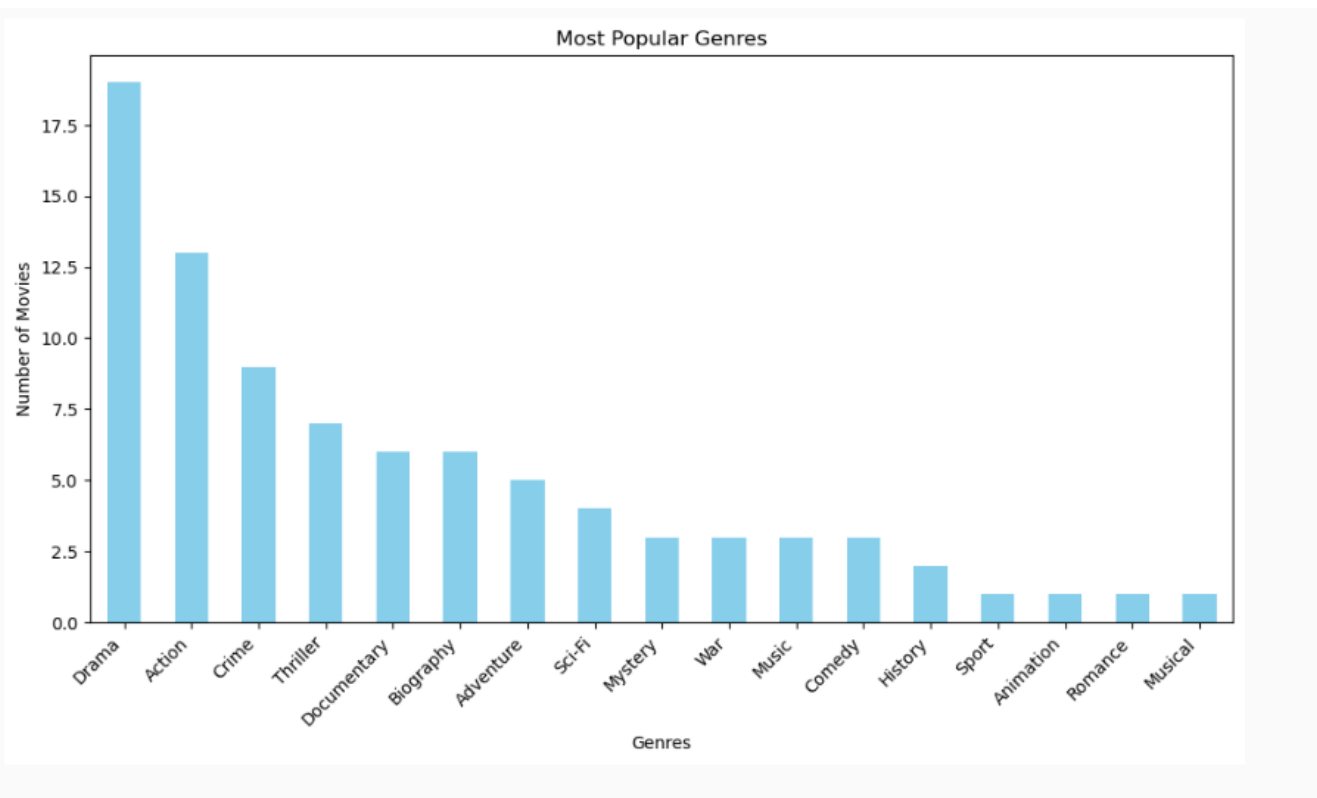
	movie_id	primary_title	original_title	start_year	runtime_minutes	genres	averagerating	numvotes
0	tt1375666	Inception	Inception	2010	148.0	Action,Adventure,Sci-Fi	8.8	1841066
1	tt0816692	Interstellar	Interstellar	2014	169.0	Adventure,Drama,Sci-Fi	8.6	1299334
2	tt1675434	The Intouchables	Intouchables	2011	112.0	Biography,Comedy,Drama	8.5	677343
3	tt4154756	Avengers: Infinity War	Avengers: Infinity War	2018	149.0	Action,Adventure,Sci-Fi	8.5	670926
4	tt2582802	Whiplash	Whiplash	2014	106.0	Drama,Music	8.5	616916
5	tt4154796	Avengers: Endgame	Avengers: Endgame	2019	181.0	Action,Adventure,Sci-Fi	8.8	441135
6	tt4633694	Spider-Man: Into the Spider-Verse	Spider-Man: Into the Spider-Verse	2018	117.0	Action,Adventure,Animation	8.5	210869
7	tt5074352	Dangal	Dangal	2016	161.0	Action,Biography,Drama	8.5	123638
8	tt5813916	The Mountain II	Dag II	2016	135.0	Action,Drama,War	9.3	100568
9	tt1424432	Senna	Senna	2010	106.0	Biography,Documentary,Sport	8.6	55318
10	tt8108198	Andhadhun	Andhadhun	2018	139.0	Crime,Thriller	8.5	43409
11	tt2592910	CM101MMXI Fundamentals	CM101MMXI Fundamentals	2013	139.0	Comedy,Documentary	9.2	41560
12	tt8291224	Uri: The Surgical Strike	Uri: The Surgical Strike	2019	138.0	Action,Drama,War	8.6	30292
13	tt0770802	Samsara	Samsara	2011	102.0	Documentary,Music	8.5	29725
14	tt6316138	Ayla: The Daughter of War	Ayla: The Daughter of War	2017	125.0	Drama,History,War	8.7	26743
15	tt3417422	Drishyam	Drishyam	2013	160.0	Crime,Drama,Thriller	8.8	24326
16	tt8267604	Capernaum	Capharnaüm	2018	126.0	Drama	8.5	20215
17	tt6148156	Vikram Vedha	Vikram Vedha	2017	147.0	Action,Crime,Thriller	8.7	18704
18	tt5354160	Aynabaji	Aynabaji	2016	147.0	Crime,Mystery,Thriller	9.3	18470
19	tt2170667	Wheels	Wheels	2014	115.0	Drama	9.3	17308
20	tt7392212	The Stage	Rangasthalam	2018	170.0	Action,Drama	8.6	15407
21	tt5275892	O.J.: Made in America	O.J.: Made in America	2016	467.0	Biography,Crime,Documentary	8.9	14946
22	tt4987556	Thani Oruvan	Thani Oruvan	2015	160.0	Action,Crime,Thriller	8.5	13747
23	tt4432480	RangiTaranga	RangiTaranga	2015	149.0	Mystery,Thriller	8.5	11203
24	tt7019842	96	96	2018	158.0	Drama,Romance	8.8	10903
25	tt7060344	Ratsasan	Ratsasan	2018	170.0	Action,Crime,Thriller	8.8	10518
26	tt7738784	Peranbu	Peranbu	2018	147.0	Drama	9.4	9629
27	tt6380520	Dhuruvangal Pathinaaru	Dhuruvangal Pathinaaru	2016	105.0	Action,Crime,Mystery	8.5	8560
28	tt5867800	Aruvi	Aruvi	2016	130.0	Drama	8.8	8277
29	tt4640206	Sachin	Sachin	2017	138.0	Biography,Documentary,Drama	8.7	7122
30	tt7465992	Mahanati	Mahanati	2018	177.0	Biography,Drama	8.7	6917
31	tt5963218	Aloko Udupadi	Aloko Udupadi	2017	113.0	Drama,History	9.5	6509
32	tt3327994	Human	Human	2015	190.0	Documentary	8.7	6267
33	tt5959980	Vada Chennai	Vada Chennai	2018	164.0	Action,Crime	8.7	6217
34	tt2077886	The Phantom of the Opera at the Royal Albert Hall	The Phantom of the Opera at the Royal Albert Hall	2011	137.0	Drama,Music,Musical	8.8	5926
35	tt7131622	Once Upon a Time ... in Hollywood	Once Upon a Time ... in Hollywood	2019	159.0	Comedy,Drama	9.7	5600

Here we find the:  
**\*\*Highly rated movies\*\***:  
 `Movies with a rating of 8.5 (85% rating and higher)`

**\*\*Popular Movies\*\***:  
 `Movies more than 5000 votes`

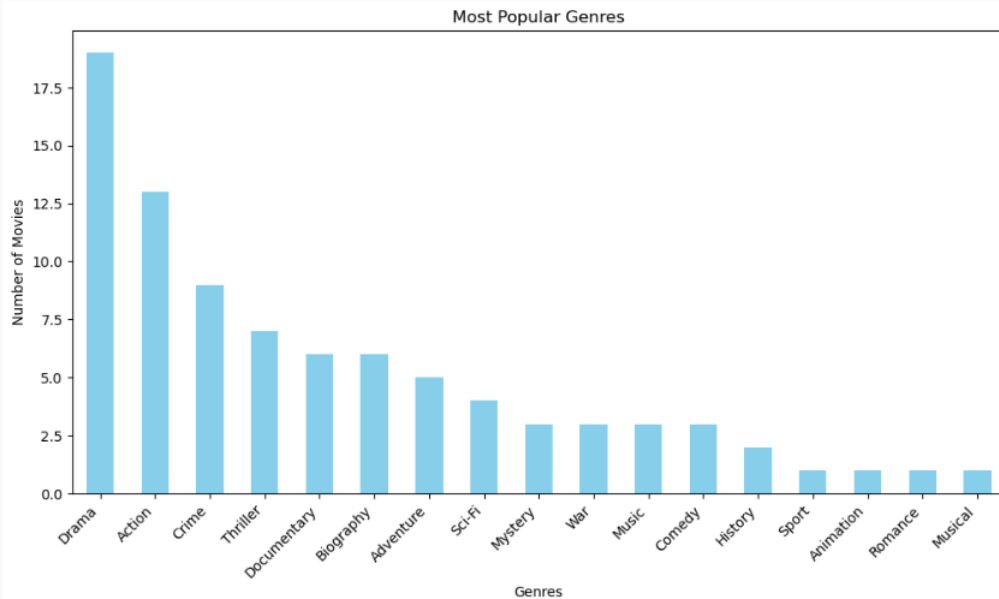
Here I tried to find out the top movies by querying the IMDB database to find blockbusters:`

My definition of a blockbuster in this context is a movie with more than 5000 votes and with a rating higher than 8.5



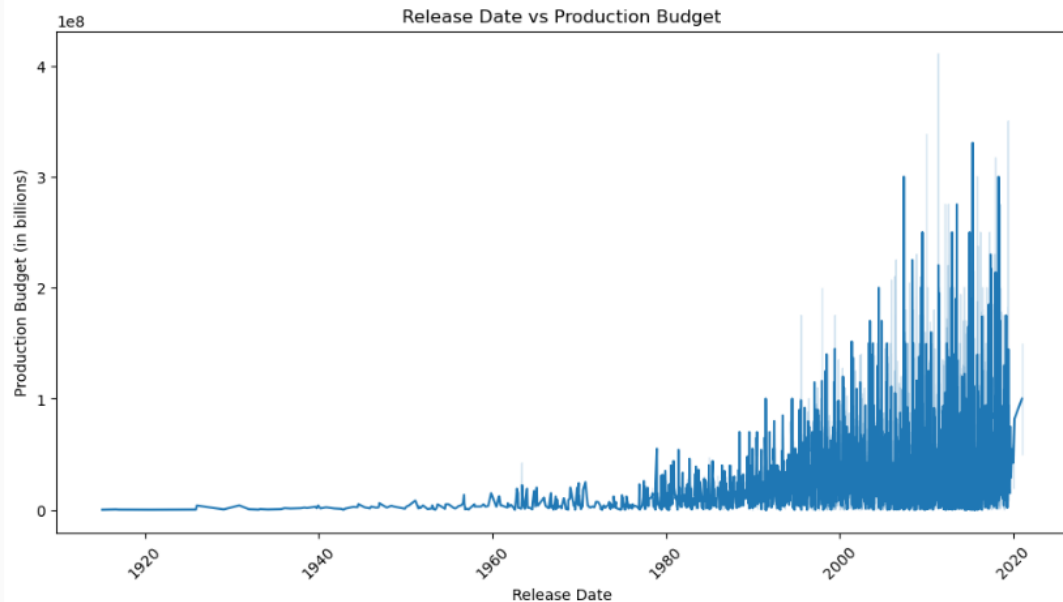
# Drama genre?

---

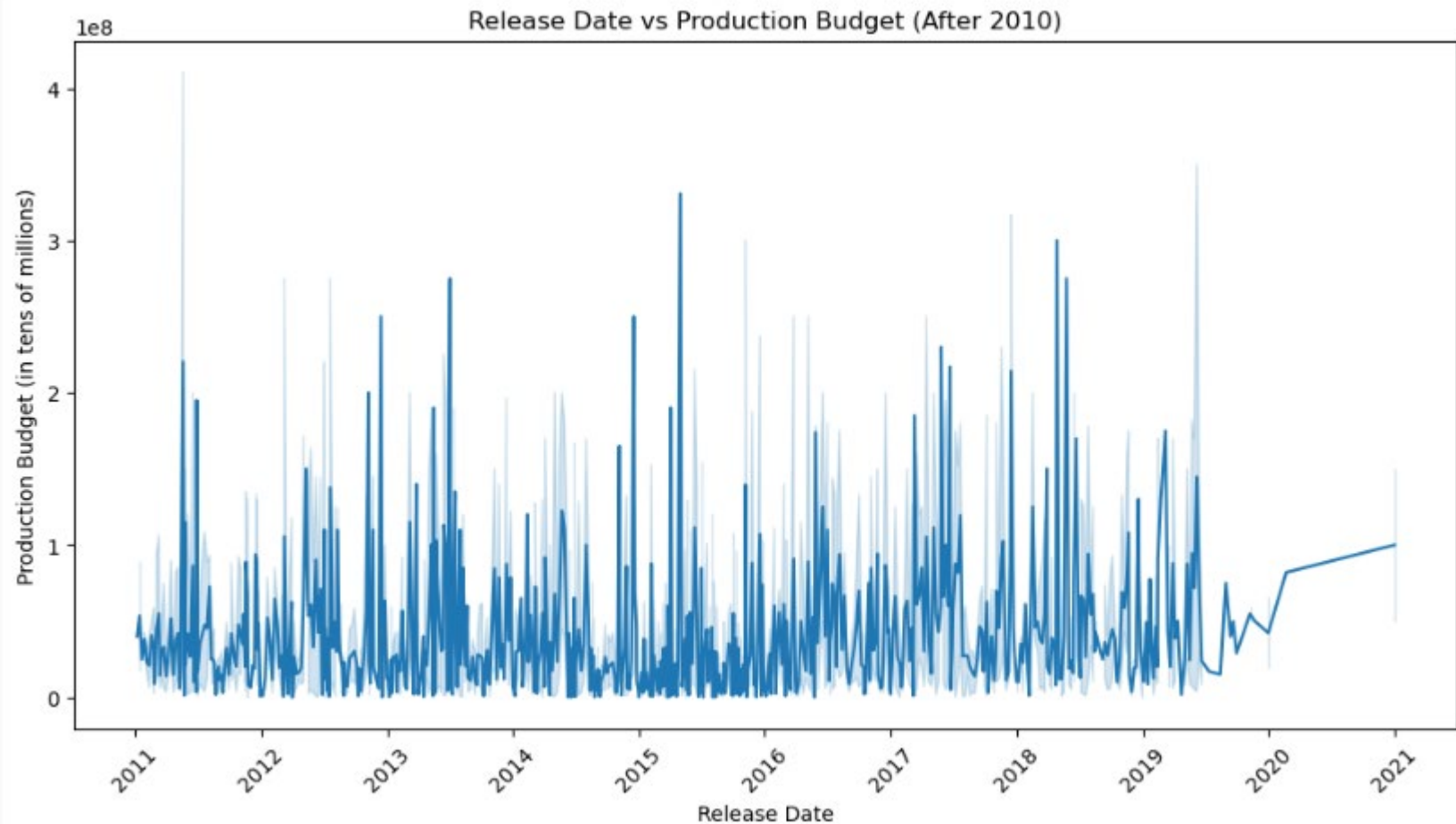


From this bar chart (above), here we can see that a drama is the most popular genre in the film industry at the moment. However, we need to find out if this genre is indeed the best move that the new Microsoft Movie Studio should create

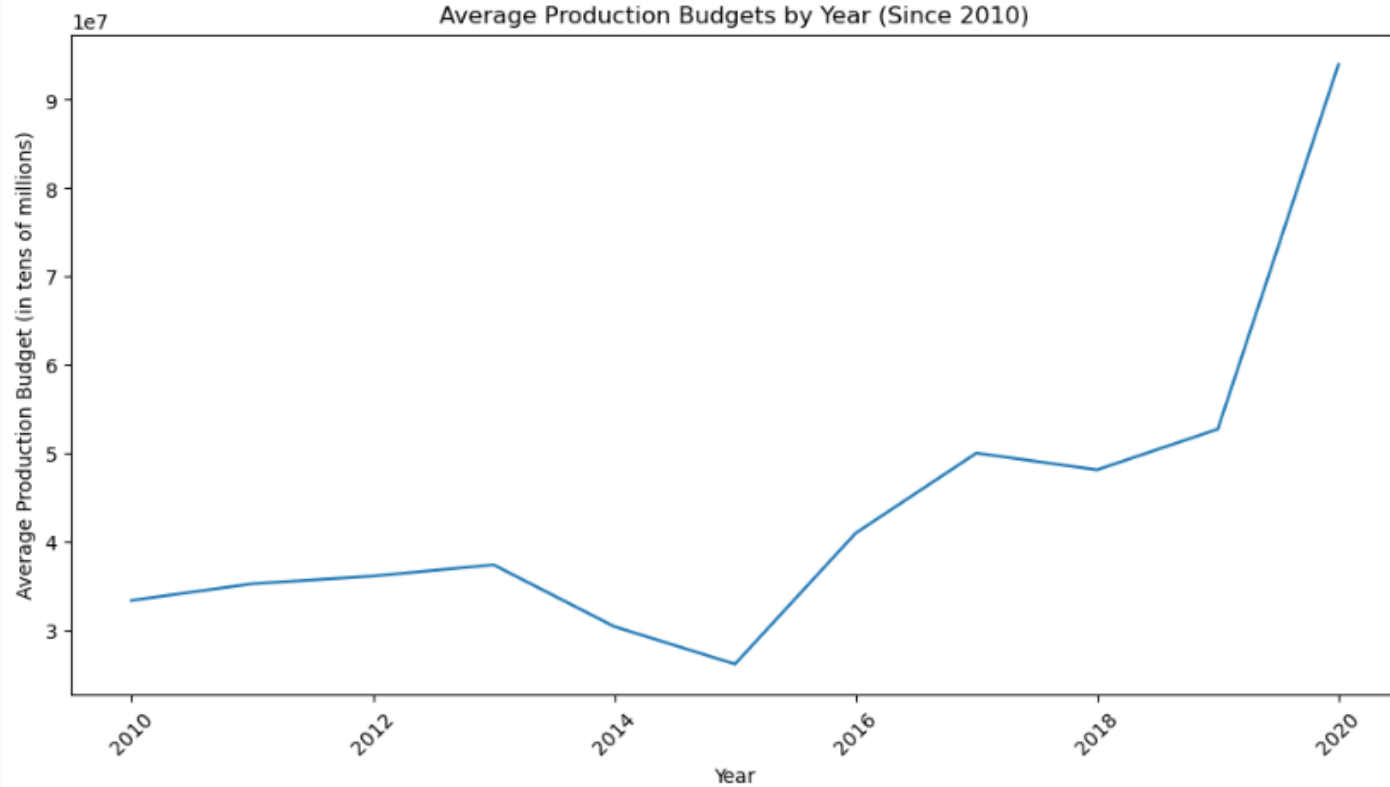
# Historical Information about movie production budgets:



	title	studio	domestic_gross	foreign_gross	year
1872	Star Wars: The Force Awakens	BV	936700000.0	1,131.6	2015
3080	Black Panther	BV	700100000.0	646900000	2018
3079	Avengers: Infinity War	BV	678800000.0	1,369.5	2018
1873	Jurassic World	Uni.	652300000.0	1,019.4	2015
727	Marvel's The Avengers	BV	623400000.0	895500000	2012
2758	Star Wars: The Last Jedi	BV	620200000.0	712400000	2017
3082	Incredibles 2	BV	608600000.0	634200000	2018
2323	Rogue One: A Star Wars Story	BV	532200000.0	523900000	2016
2759	Beauty and the Beast (2017)	BV	504000000.0	759500000	2017
2324	Finding Dory	BV	486300000.0	542300000	2016







Here we can see that movie production budgets are on the raise, and are north of \$USD 90 Million



# Average Domestic Gross

I found that the Average Domestic Gross of a movie: = `USD  
\$28,745,845.07

Using the `bom.movie_gross.csv`

**1. BUSINESS UNDERSTANDING**



```
graph TD; A[1. BUSINESS UNDERSTANDING] --> B[2. DATA PREPARATION]; B --> C[3. DATA UNDERSTANDING]; C --> D[4. DATA ANALYSIS]; D --> E[5. VISUALIZATION];
```

**2. DATA PREPARATION**

**3. DATA UNDERSTANDING**

**4. DATA ANALYSIS**

**5. VISUALIZATION**

**Section 2: End  
Recommendations**

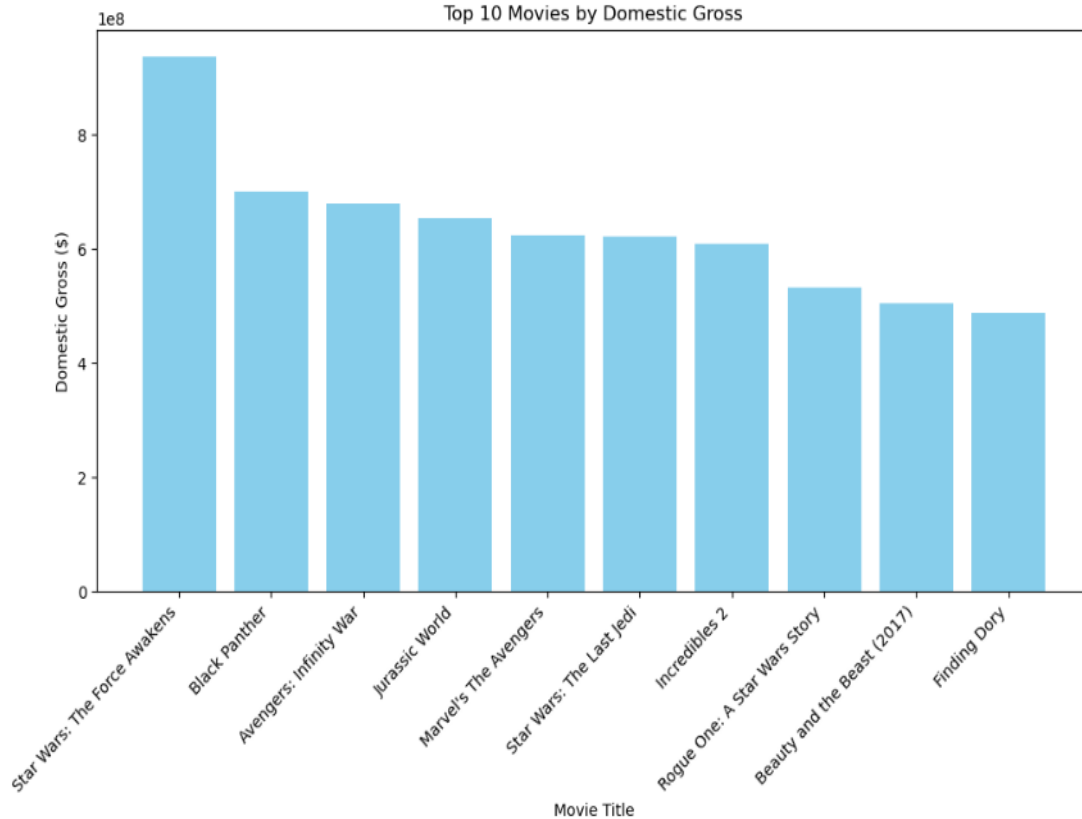


Recommendations for Microsoft Studio's

# Top movies by US domestic Gross

As we can see from this graph:

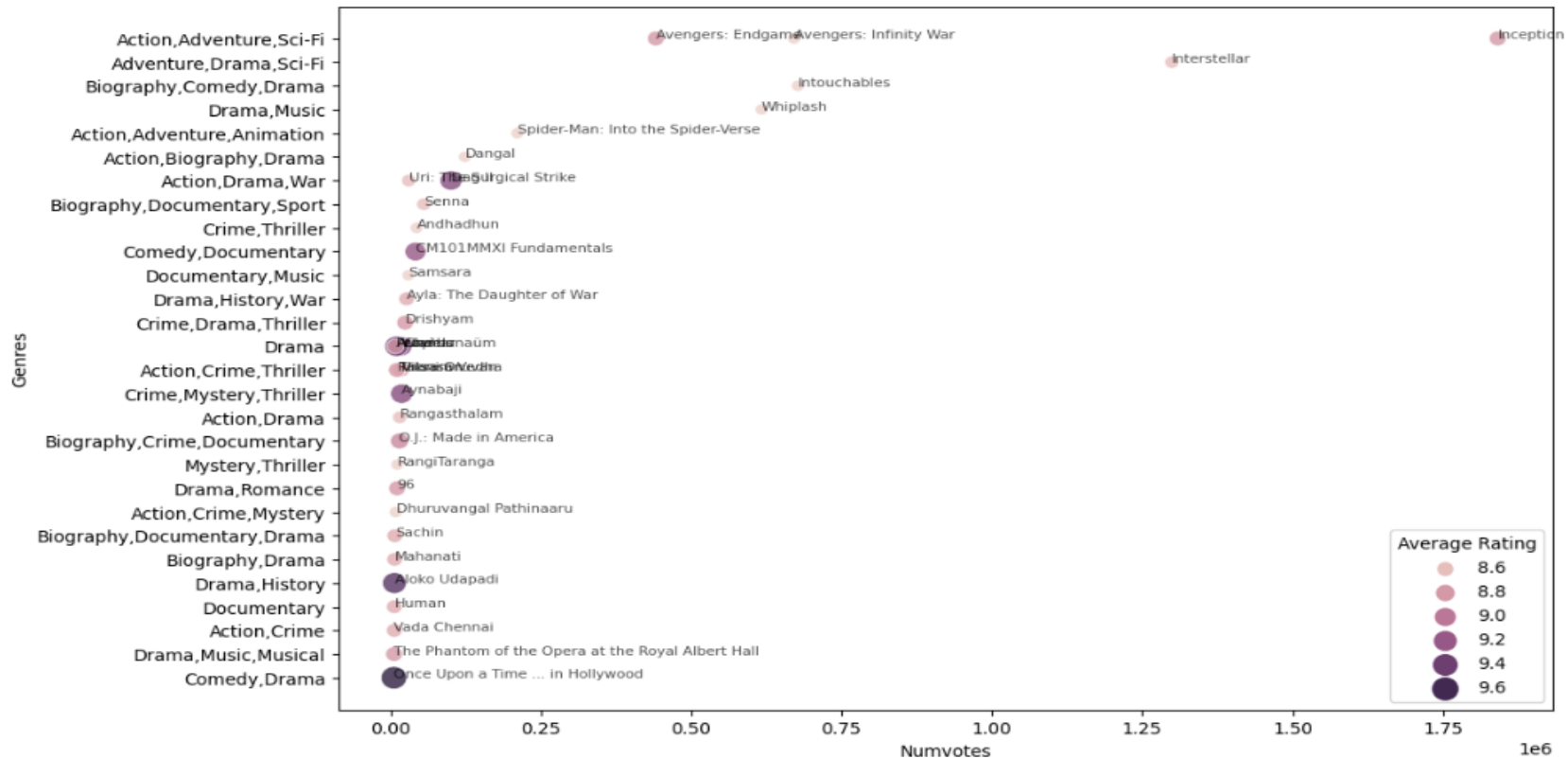
- Blockbusters make Hundreds of millions of dollars
- Movie Sequels also do well, as see with several Star Wars films



---

Although Drama's might be the most popular type of movie, they might not be the most lucrative for Microsoft

Scatter Plot: Genres vs. Numvotes



From this diagram, we see that Action, Sci-Fi genre films can be widely popular (most numvotes), and it is my official recommendation that Microsoft Movie studio's make a Action Sci-Fi film

# Finding the right people: Director to implement the Microsoft Movie Division

movie_id	primary_title	original_title	start_year	runtime_minutes	genres	averagerating	numvotes	person_id	director_name
0	tt1375666	Inception	2010	148.0	Action,Adventure,Sci-Fi	8.8	1841066	nm0634240	Christopher Nolan
1	tt0816692	Interstellar	2014	169.0	Adventure,Drama,Sci-Fi	8.6	1299334	nm0634240	Christopher Nolan
2	tt1675434	The Intouchables	2011	112.0	Biography,Comedy,Drama	8.5	677343	nm0619923	Olivier Nakache
3	tt1675434	The Intouchables	2011	112.0	Biography,Comedy,Drama	8.5	677343	nm0885918	Éric Toledano
4	tt4154756	Avengers: Infinity War	2018	149.0	Action,Adventure,Sci-Fi	8.5	670926	nm0751577	Anthony Russo
5	tt4154756	Avengers: Infinity War	2018	149.0	Action,Adventure,Sci-Fi	8.5	670926	nm0751648	Joe Russo
6	tt2582802	Whiplash	2014	106.0	Drama,Music	8.5	616916	nm3227090	Damien Chazelle
7	tt4154796	Avengers: Endgame	2019	181.0	Action,Adventure,Sci-Fi	8.8	441135	nm0751577	Anthony Russo
8	tt4154796	Avengers: Endgame	2019	181.0	Action,Adventure,Sci-Fi	8.8	441135	nm0751648	Joe Russo
9	tt4633694	Spider-Man: Into the Spider-Verse	2018	117.0	Action,Adventure,Animation	8.5	210869	nm0709056	Peter Ramsey
10	tt4633694	Spider-Man: Into the Spider-Verse	2018	117.0	Action,Adventure,Animation	8.5	210869	nm0745247	Rodney Rothman
11	tt4633694	Spider-Man: Into the Spider-Verse	2018	117.0	Action,Adventure,Animation	8.5	210869	nm2130108	Bob Persichetti
12	tt5074352	Dangal	2016	161.0	Action,Biography,Drama	8.5	123638	nm4318159	Nitesh Tiwari
13	tt5813916	The Mountain II	2016	135.0	Action,Drama,War	9.3	100568	nm3809021	Alper Caglar
14	tt1424432	Senna	2010	106.0	Biography,Documentary,Sport	8.6	55318	nm0438090	Asif Kapadia
15	tt8108198	Andhadhun	2018	139.0	Crime,Thriller	8.5	43409	nm1437189	Sriram Raghavan
16	tt2592910	CM101MMXI Fundamentals	2013	139.0	Comedy,Documentary	9.2	41560	nm2339782	Murat Dündar
17	tt8291224	Uri: The Surgical Strike	2019	138.0	Action,Drama,War	8.6	30292	nm2336554	Aditya Dhar
18	tt0770802	Samsara	2011	102.0	Documentary,Music	8.5	29725	nm0294825	Ron Fricke
19	tt6316138	Ayla: The Daughter of War	2017	125.0	Drama,History,War	8.7	26743	nm8631915	Can Ullkay
20	tt3417422	Drishyam	2013	160.0	Crime,Drama,Thriller	8.8	24326	nm2572228	Jeethu Joseph
21	tt8267604	Capharnaüm	2018	126.0	Drama	8.5	20215	nm1701024	Nadine Labaki
22	tt6148156	Vikram Vedha	2017	147.0	Action,Crime,Thriller	8.7	18704	nm2566836	Gayatri
23	tt6148156	Vikram Vedha	2017	147.0	Action,Crime,Thriller	8.7	18704	nm2575525	Pushkar
24	tt5354160	Aynabaji	2016	147.0	Crime,Mystery,Thriller	9.3	18470	nm2755490	Amitabh Reza Chowdhury
25	tt2170667	Wheels	2014	115.0	Drama	9.3	17308	nm1919905	Donavon Warren
26	tt2170667	Wheels	2014	115.0	Drama	9.3	17308	nm2871077	Tim Gagliardo
27	tt7392212	The Stage	2018	170.0	Action,Drama	8.6	15407	nm1335875	Sukumar
28	tt5275892	O.J.: Made in America	2016	467.0	Biography,Crime,Documentary	8.9	14946	nm1554601	Ezra Edelman
29	tt4987556	Thani Oruvan	2015	160.0	Action,Crime,Thriller	8.5	13747	nm1001345	Mohan Raja
30	tt4432480	RangITaranga	2015	149.0	Mystery,Thriller	8.5	11203	nm4093647	Anup Bhandari
31	tt7019842	96	2018	158.0	Drama,Romance	8.8	10903	nm4242320	C. Prem Kumar
32	tt7060344	Ratsasan	2018	170.0	Action,Crime,Thriller	8.8	10518	nm6442107	Ram Kumar
33	tt7738784	Peranbu	2018	147.0	Drama	9.4	9629	nm3591496	Ram
34	tt7738784	Peranbu	2018	147.0	Drama	9.4	9629	nm6041686	Anjana Krishnakumar
35	tt7738784	Peranbu	2018	147.0	Drama	9.4	9629	nm9988436	Amudhavan Karuppiiah
36	tt6380520	Dhuruvangal Pathinaaru	2016	105.0	Action,Crime,Mystery	8.5	8560	nm8678191	Karthik Naren
37	tt5867800	Aruvi	2016	130.0	Drama	8.8	8277	nm8273413	Arun Prabhu Purushothaman
38	tt4640206	Sachin	2017	138.0	Biography,Documentary,Drama	8.7	7122	nm0259847	James Erskine
39	tt7465992	Mahanati	2018	177.0	Biography,Drama	8.7	6917	nm5645455	Nag Ashwin
40	tt5963218	Aloko Udapadi	2017	113.0	Drama,History	9.5	6509	nm8359417	Chathra Weeraman
41	tt5963218	Aloko Udapadi	2017	113.0	Drama,History	9.5	6509	nm8529623	Bharatha Hettiarachchi
42	tt3327994	Human	2015	190.0	Documentary	8.7	6267	nm0037867	Yann Arthus-Bertrand
43	tt5959980	Vada Chennai	2018	164.0	Action,Crime	8.7	6217	nm4274575	VetriMaaran
44	tt2077886	The Phantom of the Opera at the Royal Albert Hall	2011	137.0	Drama,Music,Musical	8.8	5926	nm0606835	Nick Morris
45	tt2077886	The Phantom of the Opera at the Royal Albert Hall	2011	137.0	Drama,Music,Musical	8.8	5926	nm4711112	Laurence Connor
46	tt7131622	Once Upon a Time ... in Hollywood	2019	159.0	Comedy,Drama	9.7	5600	nm0000233	Quentin Tarantino

# Here are my recommendations for Movie Directors: (based on the data)

	director_name	14	Aditya Dhar	29	Ram Kumar
0	Christopher Nolan	15	Ron Fricke	30	Ram
1	Olivier Nakache	16	Can Ulkay	31	Anjana Krishnakumar
2	Éric Toledano	17	Jeethu Joseph	32	Amudhavan Karuppiah
3	Damien Chazelle	18	Nadine Labaki	33	Karthick Naren
4	Anthony Russo	19	Gayatri	34	Arun Prabhu Purushothaman
5	Joe Russo	20	Pushkar	35	James Erskine
6	Peter Ramsey	21	Amitabh Reza Chowdhury	36	Nag Ashwin
7	Rodney Rothman	22	Donavon Warren	37	Chathra Weeraman
8	Bob Persichetti	23	Tim Gagliardo	38	Bharatha Hettiarachchi
9	Nitesh Tiwari	24	Sukumar	39	Yann Arthus-Bertrand
10	Alper Caglar	25	Ezra Edelman	40	Vetrimaaran
11	Asif Kapadia	26	Mohan Raja	41	Nick Morris
12	Sriram Raghavan	27	Anup Bhandari	42	Laurence Connor
13	Murat Dündar	28	C. Prem Kumar	43	Quentin Tarantino





# Objectives vs Recommendations

---

- |  |   |
|--|---|
| 1. The type of film they should be making (using data) | 1. Action Sc-Fi Films (ideally with many sequels, like Star wars) |
| 2. The ideal movie budget                              | 2. `USD \$28,745,845 or more                                      |
| 3. The ideal director they should hire                 | 3. (See previous slide for film directors' recommendations)       |

# Thanks you for listening to my feature film pitch proposal for Microsoft Movie Studio's

---



This is both student project and ideal movie awareness pitch for Microsoft

My name is Pete Njagi

My work email is [Petedigital@hotmail.com](mailto:Petedigital@hotmail.com)

My work phone number is +254 728 752846

This was a Moringa Data Science student Project

My aim was to: Use Data Science to inform you of the best films that Microsoft Movie Studio's should develop.

---

# Thanks

