

# Exposition



---

James Benedito

# Introduction

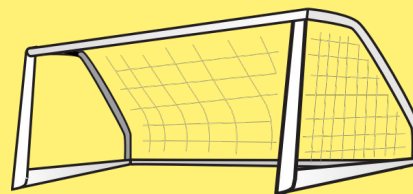
## Problem:

- Microsoft wants to start a new movie studio, but doesn't have experience making movies.



## Goal:

- Based on data analysis, give suggestions on types of movies Microsoft should make.



*“Our mission is to empower every person and every organization on the planet to achieve more.” -Microsoft*

---

# Variables to Consider

- Look at movies with...
  - Highest worldwide gross
  - Highest ratings
- To decide the...
  - Genre of film to make
  - Optimal runtime
  - Optimal budget



# Methods (Process Steps)

Open up datasets in a Jupyter notebook

1

Clean up/filter data using Pandas

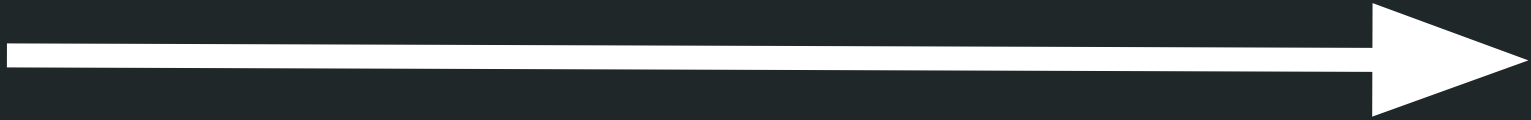
2

Do statistical analysis on data and create visualizations using Matplotlib

3

Generate concrete suggestions based on findings

4



# Data

```
graph LR; Data[Data] --- imdb[im.db]; Data --- tn[tn.movie_budgets.csv];
```

The diagram illustrates the data sources for a central entity labeled 'Data'. Two horizontal lines connect the 'Data' box to two separate source boxes on the right. The top box is labeled 'im.db' and contains a bullet point 'Data from IMDB'. The bottom box is labeled 'tn.movie\_budgets.csv' and contains a bullet point 'Data from The Numbers'.

## **im.db**

- Data from IMDB

## **tn.movie\_budgets.csv**

- Data from The Numbers

# Data

im.db

- Data from IMDB



Parasite Trailer #1 (2019) | Movieclips Indie

2:27

8.5/10  
IMDb

99%  
Rotten Tomatoes

96%  
Metacritic

# Data

**tn.movie\_budgets.csv**

- Data from The Numbers



The Numbers

<https://www.the-numbers.com>



## The Numbers - Where Data and the Movie Business Meet ✓

The film earned a disappointing \$30.11 million in its opening weekend and currently sits with a six-day total of \$35.6 million from North American showings.

### Daily Box Office Chart for ... ✓

Register with The Numbers for free to see additional data and ...

### Weekend Box Office Chart ✓

65 - Cocaine Bear (2023) - Jesus Revolution (2023) - ...

### Movie Release Schedule ✓

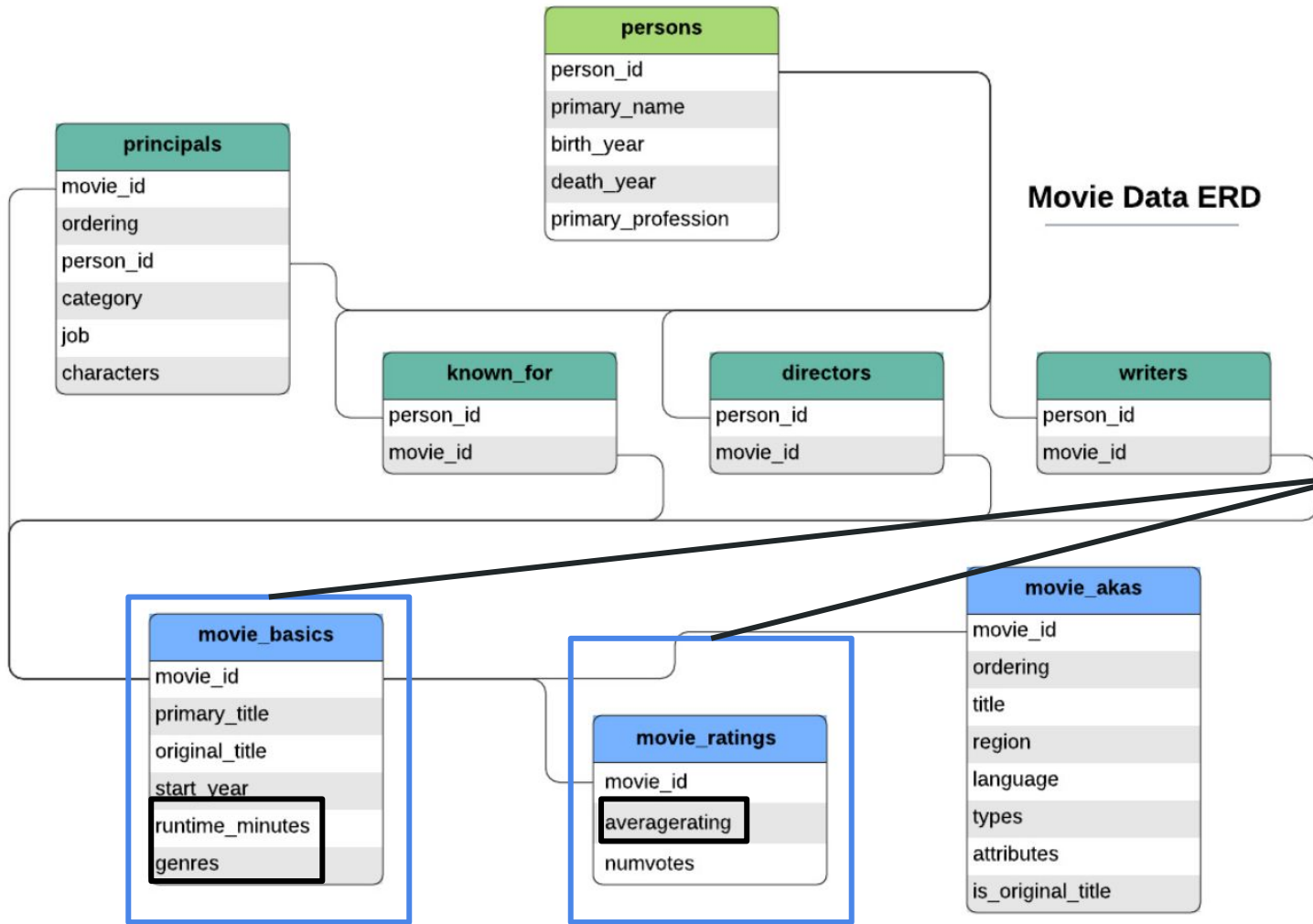
Release Schedule 2023 - Release Schedule 2022 - AIR (2023) - ...

### Box Office Records ✓

Overview of the various movie box office records available at The ...

[More results from the-numbers.com »](#)





**im.db**

-Used to figure out **genre** and **runtime** from **ratings**

Looked at the **movie\_basics** and **movie\_ratings** tables

### tn.movie\_budgets.csv

-Used to figure out an optimal **budget** based on **worldwide gross**

	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

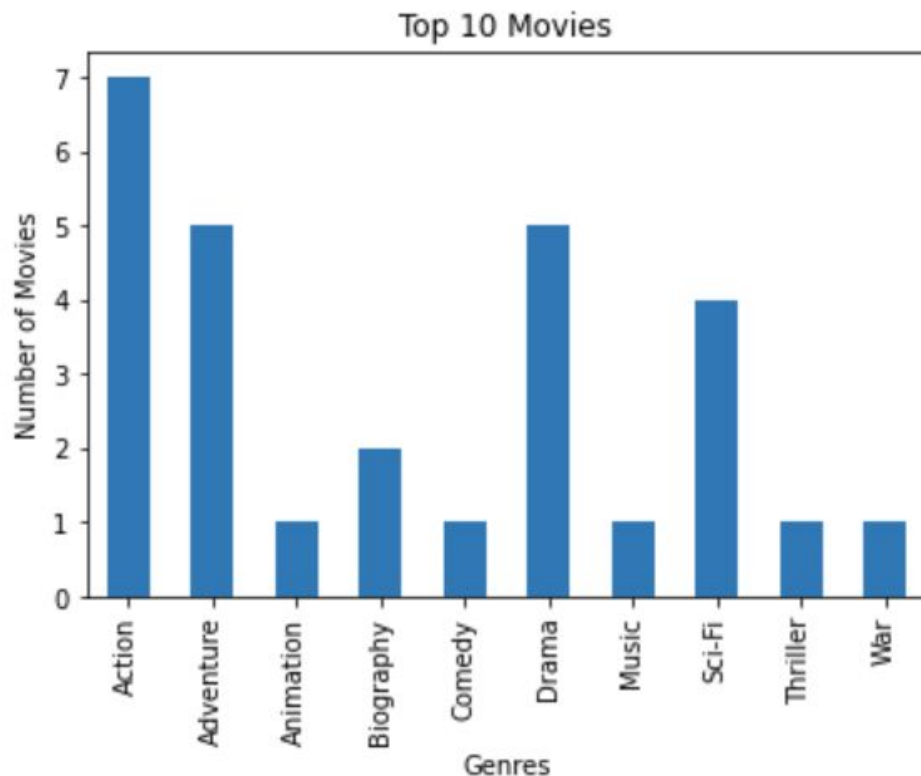
# Results & Business Recommendations

# Results: Genre

Genres	Average_Rating	Number_of_Votes
Action,Drama,War	9.3	100568
Action,Adventure,Sci-Fi	8.8	441135
Action,Adventure,Sci-Fi	8.8	1841066
Adventure,Drama,Sci-Fi	8.6	1299334
Biography,Comedy,Drama	8.5	677343
Drama,Music	8.5	616916
Action,Adventure,Animation	8.5	210869
Action,Biography,Drama	8.5	123638
Action,Adventure,Sci-Fi	8.5	670926
Action,Thriller	8.4	1387769

- Filtered by number of votes (> 100,000)
- Ordered by average rating

# Recommendation #1: Genre



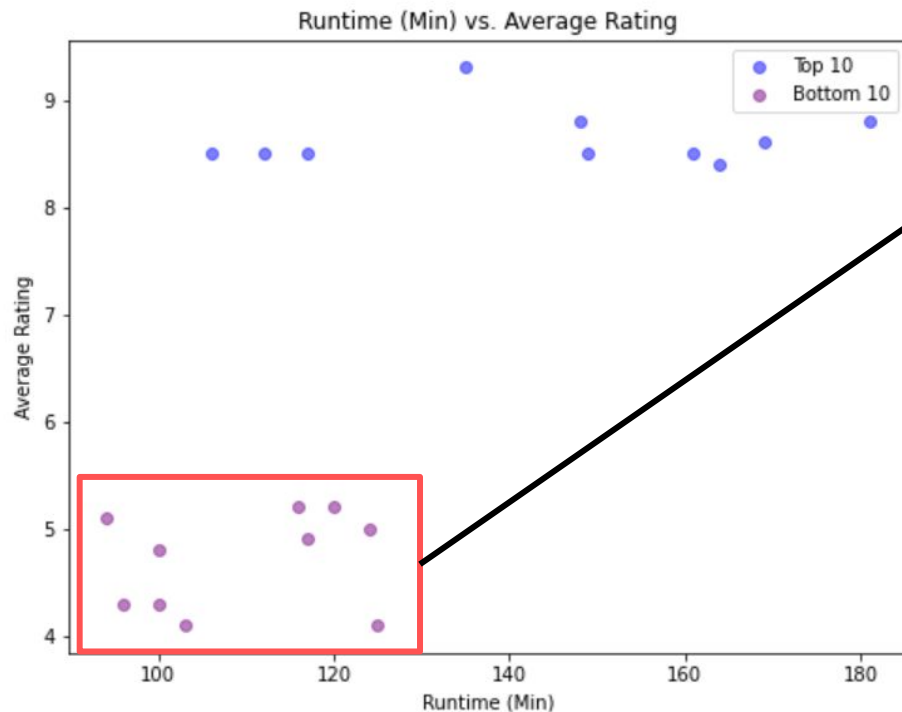
Recommendation: Film should have elements of action, adventure, and drama.

# Results: Optimal Runtime



Data shows a weak correlation between runtime and average rating ( $r = 0.33$ ).

## Recommendation #2: Optimal Runtime



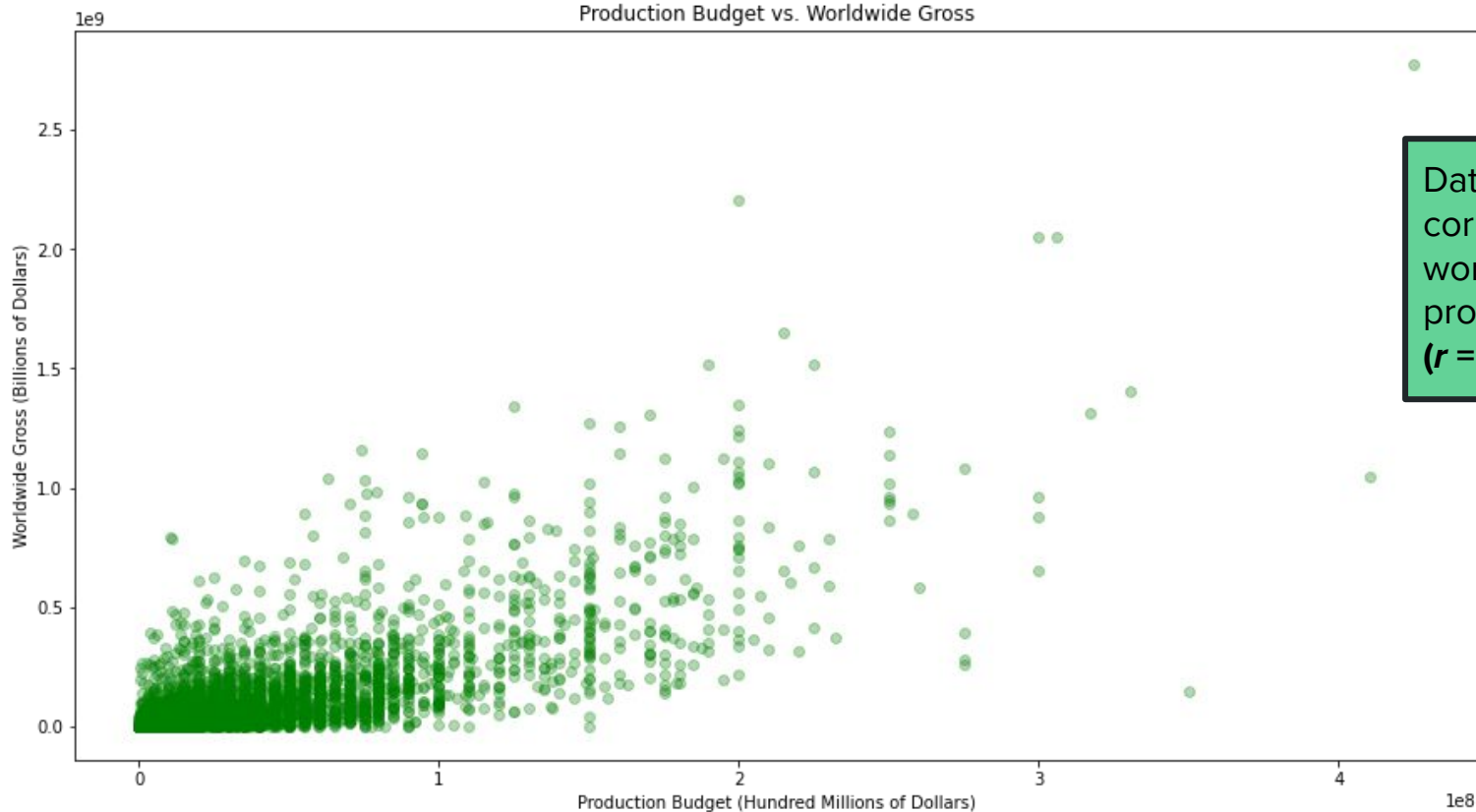
All Bottom 10 movies had a runtime below 130 minutes.

Recommendation: Film runtime should be a minimum of 130 minutes.

Mean for Top 10-rated movies: 144.2

Mean for Bottom 10-rated movies: 109.5

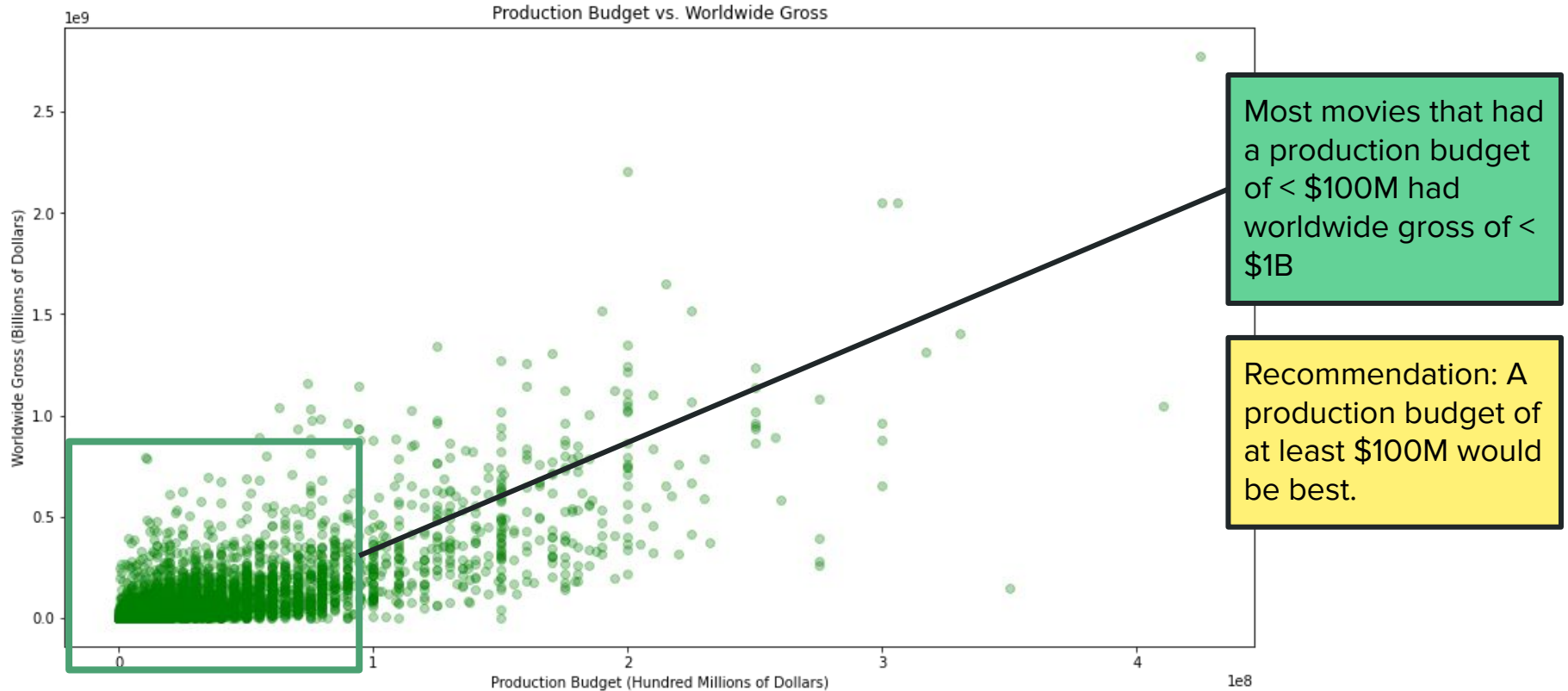
# Results: Optimal Budget



Data shows a strong correlation between worldwide gross and production budget ( $r = 0.75$ ).



# Recommendation #3: Optimal Budget



```
graph TD; A((Limitations and Future Improvement Ideas)) --- B[More datasets should be analyzed]; A --- C[Top 10-rated films table used for genre had arbitrary filter]; A --- D[Production budget dictated by other variables (i.e. genre)]; A --- E[Low, positive correlation between runtime and average ratings];
```

Limitations and  
Future  
Improvement  
Ideas

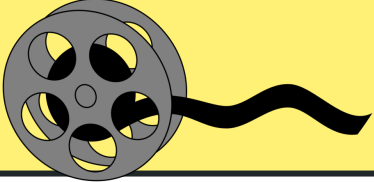


More datasets  
should be analyzed

Top 10-rated films  
table used for genre  
had arbitrary filter

Low, positive correlation  
between runtime and  
average ratings

Production budget  
dictated by other  
variables (i.e. genre)

# Summary

Genres	Runtime	Production Budget
<p data-bbox="363 412 546 536">Action, Adventure, Drama</p> 	<p data-bbox="761 412 1170 492">Minimum of 130 minutes (2 hours, 10 minutes)</p> 	<p data-bbox="1315 412 1634 445">Minimum of \$100M</p> 

# Contact

James Benedito

[jbenedito99@gmail.com](mailto:jbenedito99@gmail.com)

(702) 523-2734

LinkedIn: [www.linkedin.com/in/james-benedito-7b326b217](https://www.linkedin.com/in/james-benedito-7b326b217)

---



Questions?

*That's all Folks!*