

# Microsoft Movie Studios

---

by Mario Mocombe

# Summary

Microsoft is creating a new movie studio and is inquiring about what types of films are currently doing the best at the box office. The datasets used in answering the business problem are from various movie websites and methods used were exploratory data analysis, data cleaning, and data manipulation.

# Summary

- Adventure, Action, Comedy are the top-grossing film genres
- May is the best month of the year to release a movie
- Strong correlation between high budget and higher profits
- Christopher Nolan should be hired as director

# Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

# Business Problem

Microsoft sees all the big companies creating original video content and want to get in on the fun. They have decided to create a new movie studio, but don't know anything about creating movies. We are here today to discuss what types of films are currently doing the best at the box office and hopefully Microsoft Studios can use these findings to make informed business decisions.

# Data

IMDB Database: Contains information on genre, online user votes, average user rating, identity and roles of people involved in the films. The most relevant tables are movie\_basics and movie\_ratings.

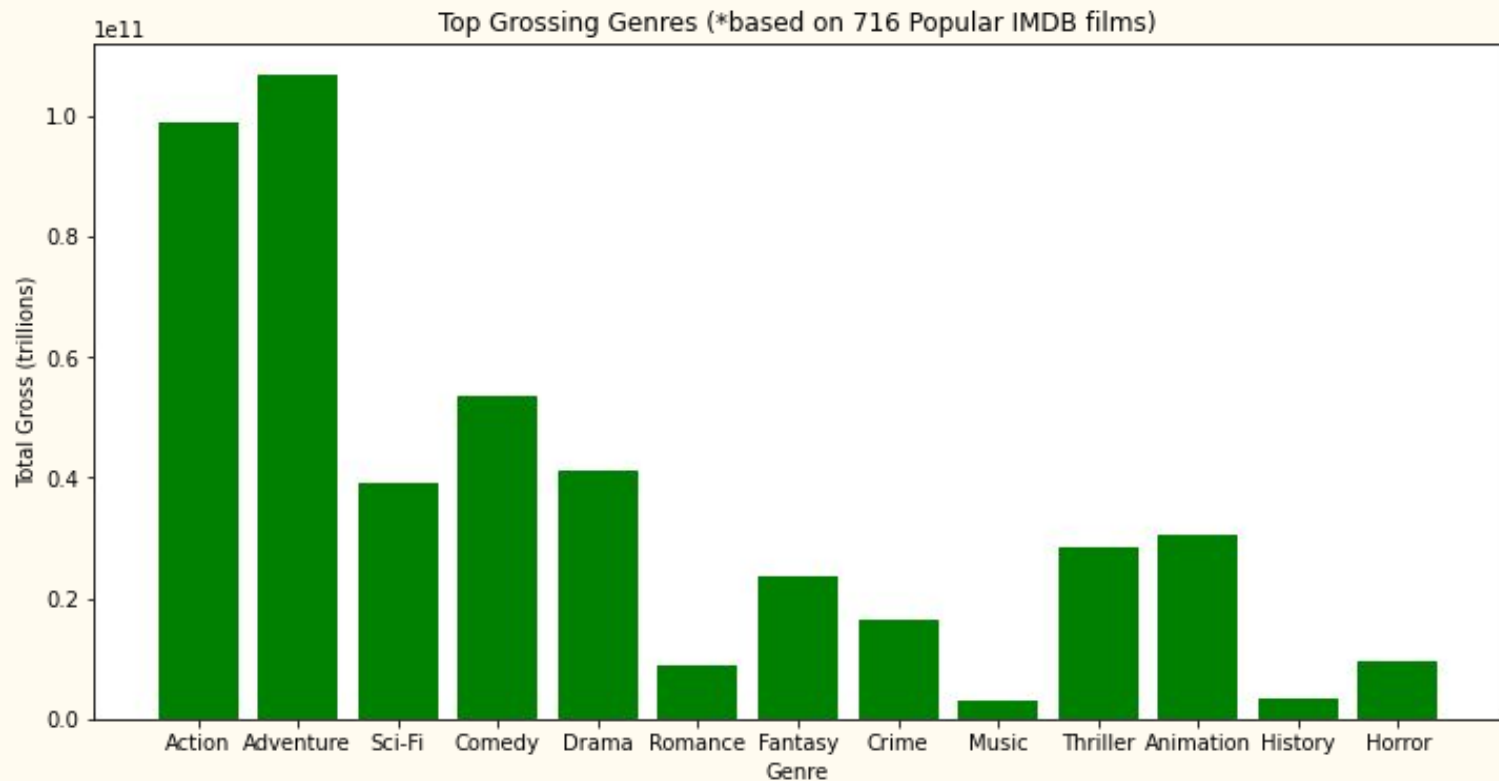
Box Office Mojo Database: Contains domestic and foreign gross information.

The Numbers Database: Contains box office data, release dates, and budget data.

# Methods

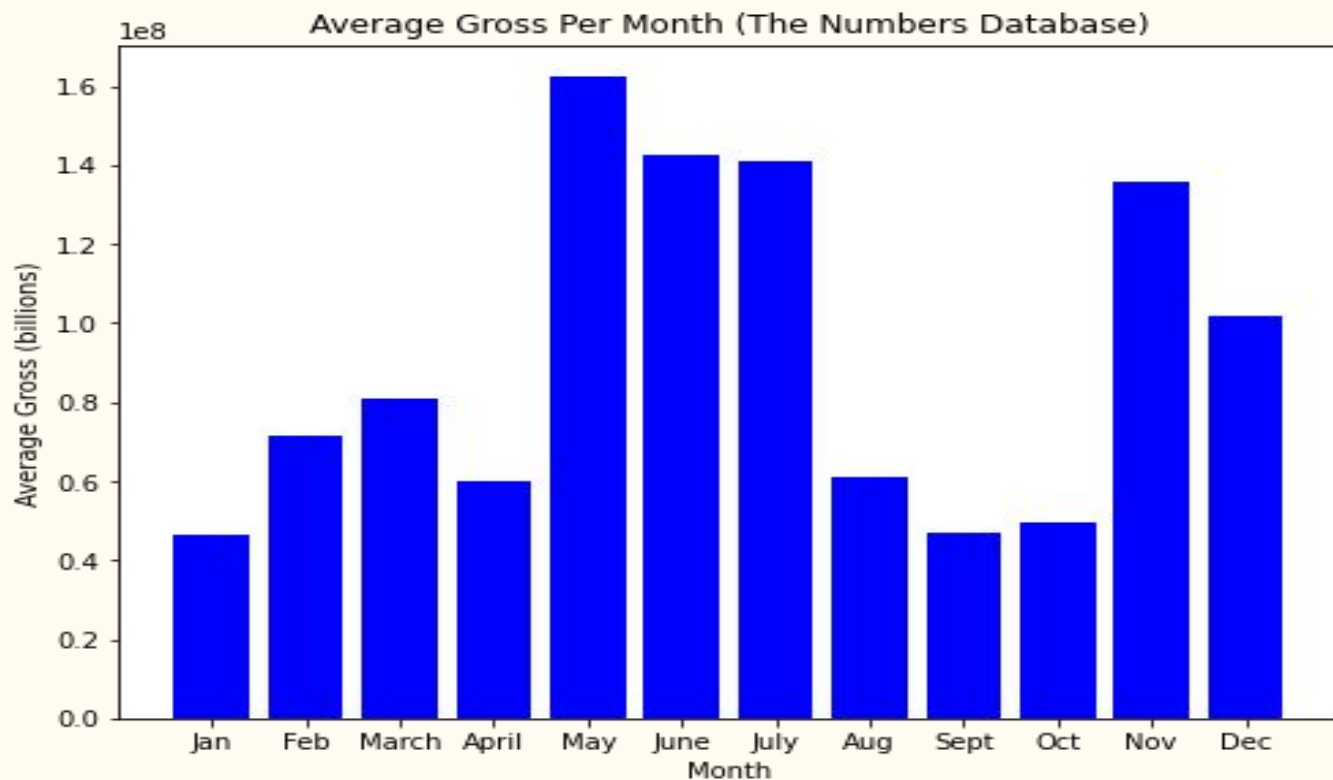
- Data cleaning: Identifying and handling missing values
- Data aggregation: Joining dataframes and using summary statistics
- Exploratory Data Analysis: Looking at the data in-depth, finding useful patterns and spotting anomalies while using visual aids to understand the data.

# Results

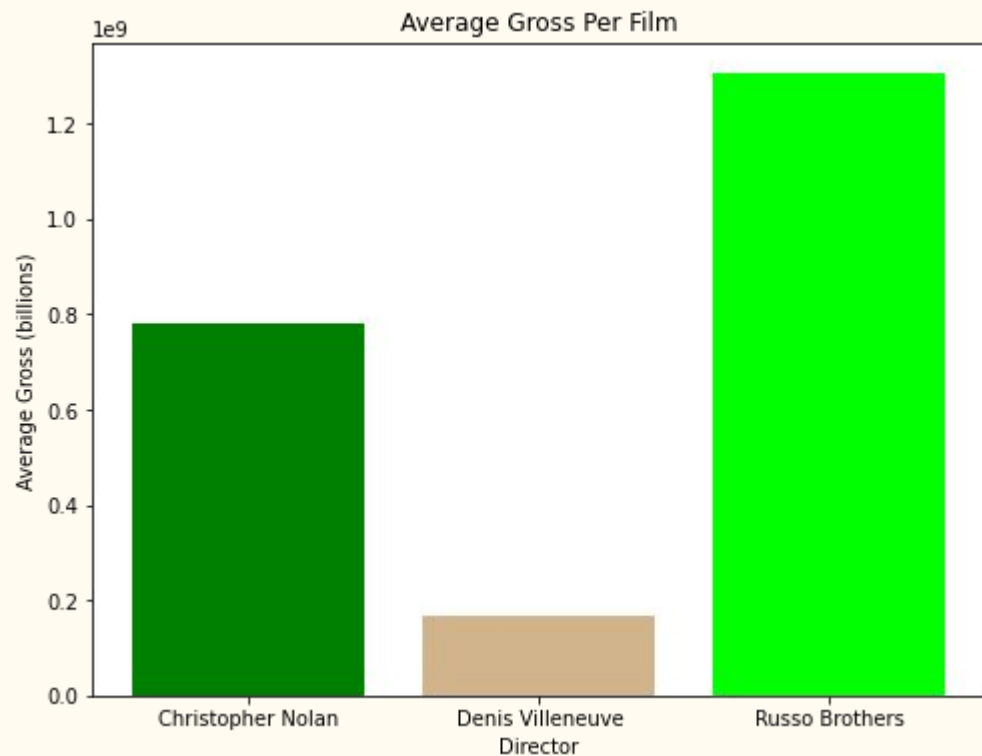




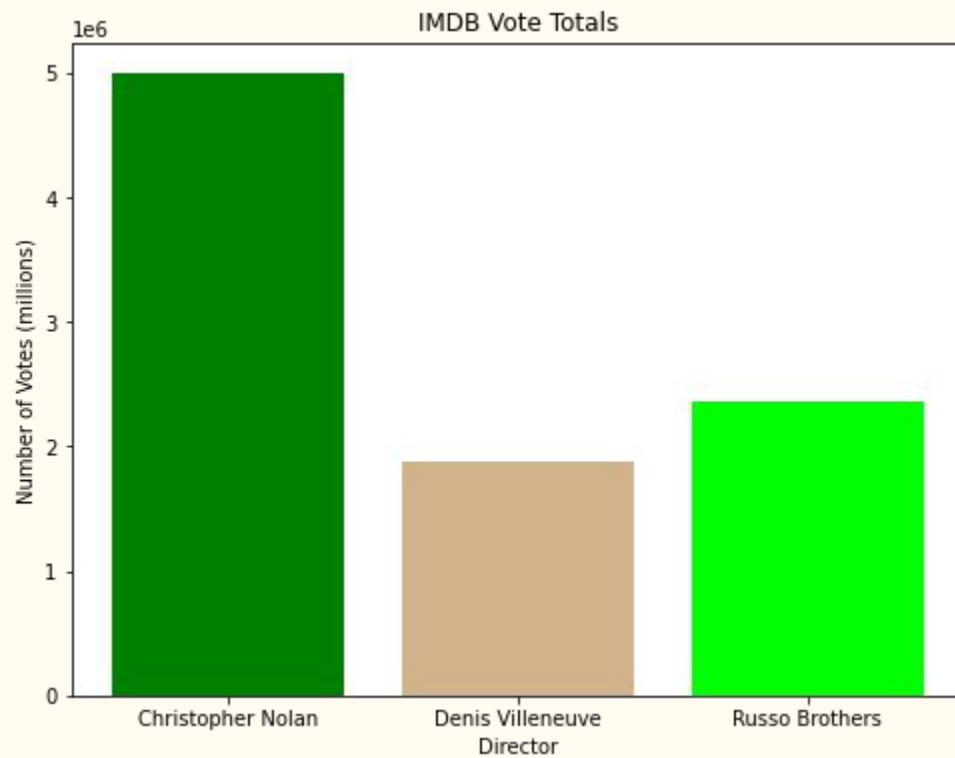
# Results



# Results



# Results



# Conclusions

- Collaborate with Christopher Nolan and release a big budget action-adventure film
- Aim for a May release so that the film can benefit from the summer box office boom.

## Limitations

- Release window may be overcrowded by other films.
- Analysis does not make predictions farther out into the future.
- Older films will probably have less online engagement on IMDB
- It would be helpful to have the following data in the future: movie streaming numbers, digital purchase and rental sales, VHS and DVD sales, and data during the pandemic and the post-pandemic recovery period.

# Thank You!

**Email:** [mario.mocombe@gmail.com](mailto:mario.mocombe@gmail.com)

**GitHub:** [@mmocombe](https://github.com/mmocombe)

**LinkedIn:** [linkedin.com/in/mario-mocombe-80117211/](https://www.linkedin.com/in/mario-mocombe-80117211/)