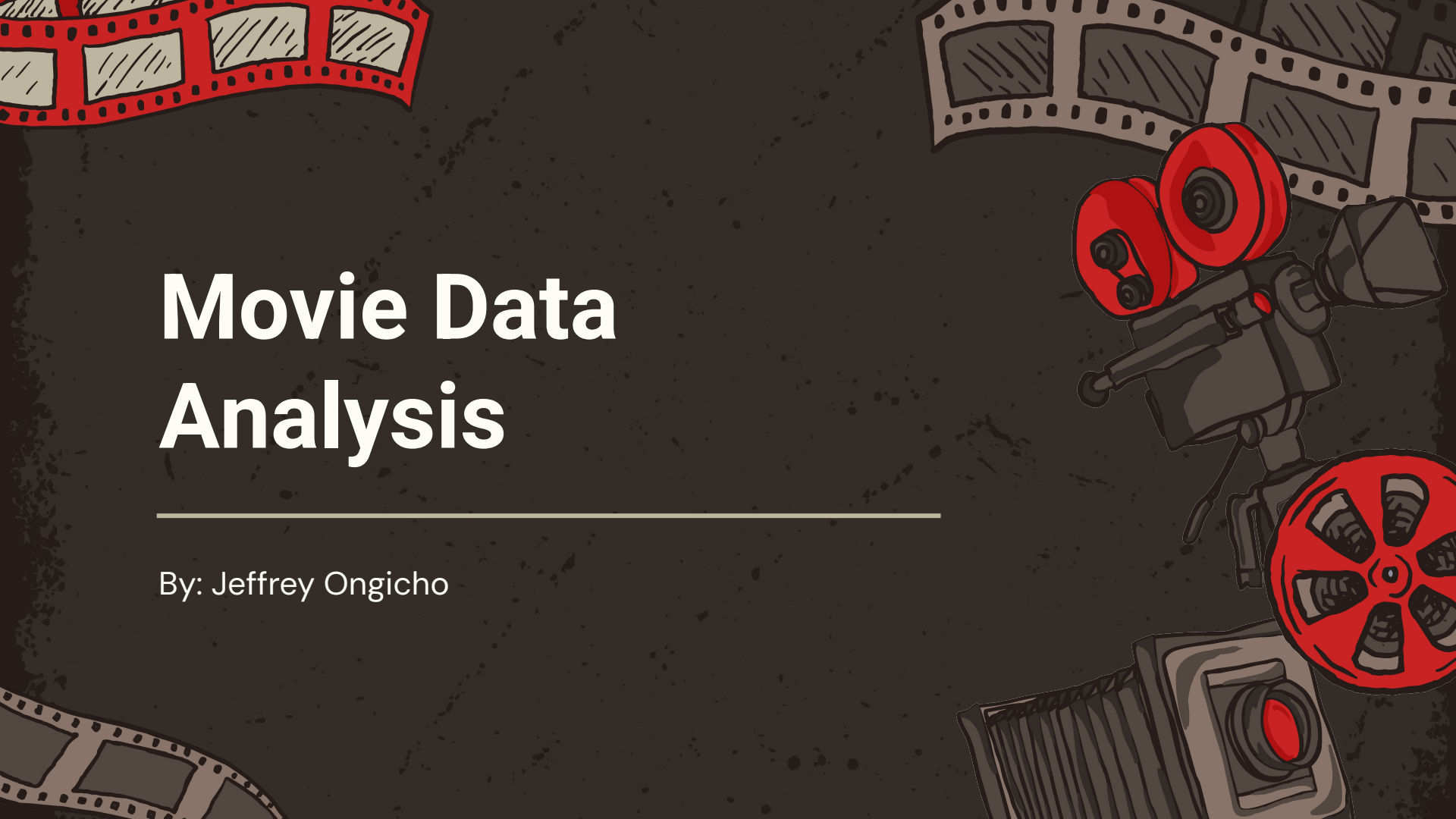


# Movie Data Analysis

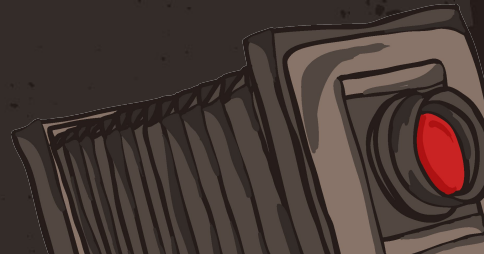
---

By: Jeffrey Ongicho



# Overview

This project is an exploratory data analysis on film data obtained from various sources to help our company as they venture into the filmmaking industry. The task is to provide insights from the analysis to help the company make decisions on the kind of films to make to maximize profit and audience engagement. The analysis will focus on identifying trends in genres and box office performance.



# Business Understanding

The company seeks to enter the movie industry after seeing other big companies make original video content. They decide to create a studio, however they have no idea on the type of movies to create. The task is to determine the type of films currently doing well at the box office and help the company determine the kind of movies to make.

Some of the key business questions to be answered include:

- \* What genre of films does the best at the box office?
- \* Is there a correlation between the production budget and the performance of a movie at the box office?
- \* What is the correlation between movie runtime and the Return of Investment?
- \* What is the average production budget for films over the years?



# Data Understanding

The data used in this analysis contains data collected from various popular movie sites such as Box Office Mojo, IMDb, rotten tomato reviews. It contains detailed information on movie titles, actors, directors, box office earnings, and movie ratings. In this analysis, the key variables were the genres, the box office earnings and the movie runtime.





# Data Preparation and Cleaning

Data preparation involved a lot of steps because multiple datasets were involved. This included checking for common columns in the datasets and combining them to provide more information towards the analysis. Other steps included:

- Checking the shape of the data-frames
- Confirming the column data types .
- Formatting column values – Changing column names for easier readability.
- Handling missing values by imputation as well as dropping rows.



# Recommendations

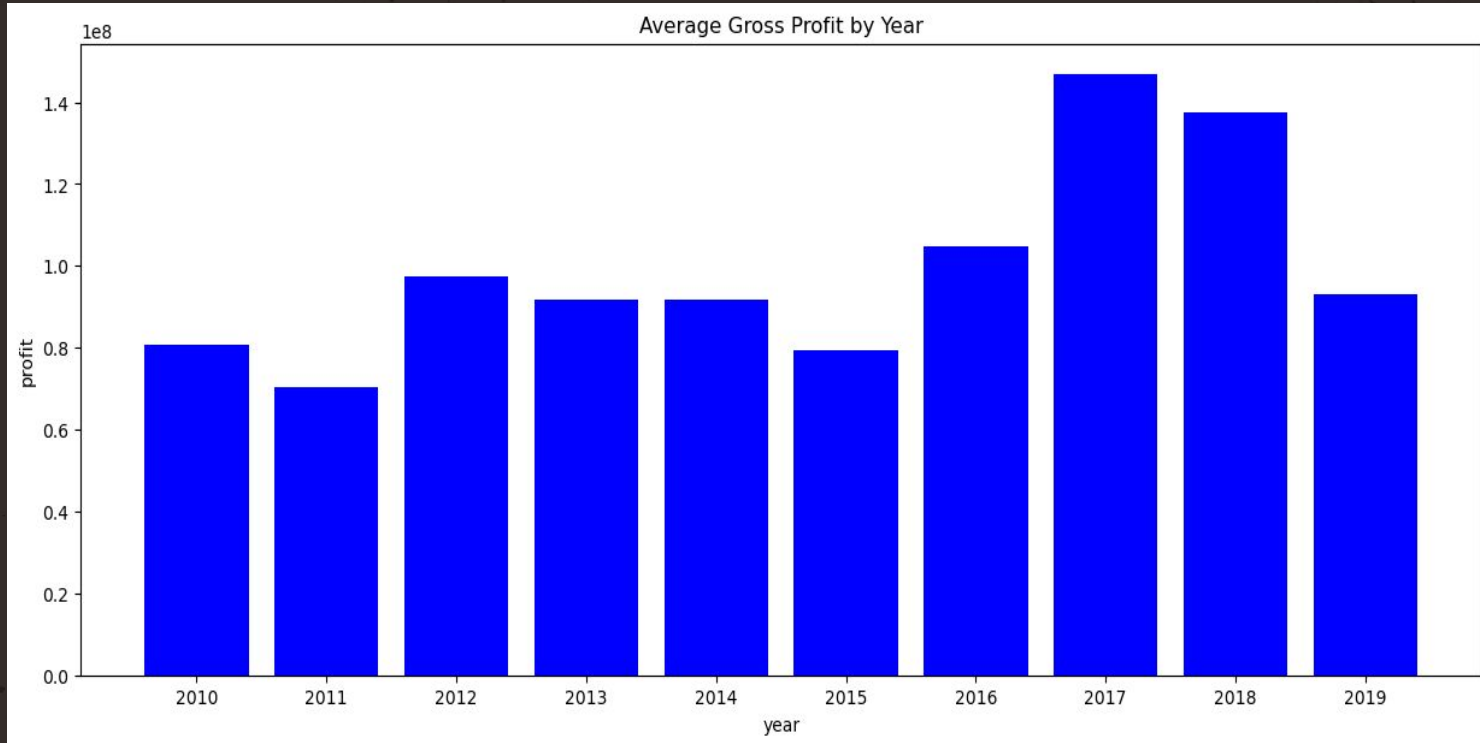
The following are recommendations drawn from the analysis done:

1. The company should have a sizeable production budget when creating the films. There was a strong positive relationship between production budget and the global gross profit of a film. More money towards production would lead to higher production quality in terms of graphics and cinematography drawing in more viewers.
2. The company should make movies of a longer runtime. Whilst there some isolated scenarios where short films still earned a hefty global gross profit, I found a general trend of profit increasing for longer runtime.
3. The company should make movies that are animations as that was the genre that had the highest average global gross profit of all genres. Action, adventure, musicals and scientific fiction should also be considered as they also had good performance at the box office.



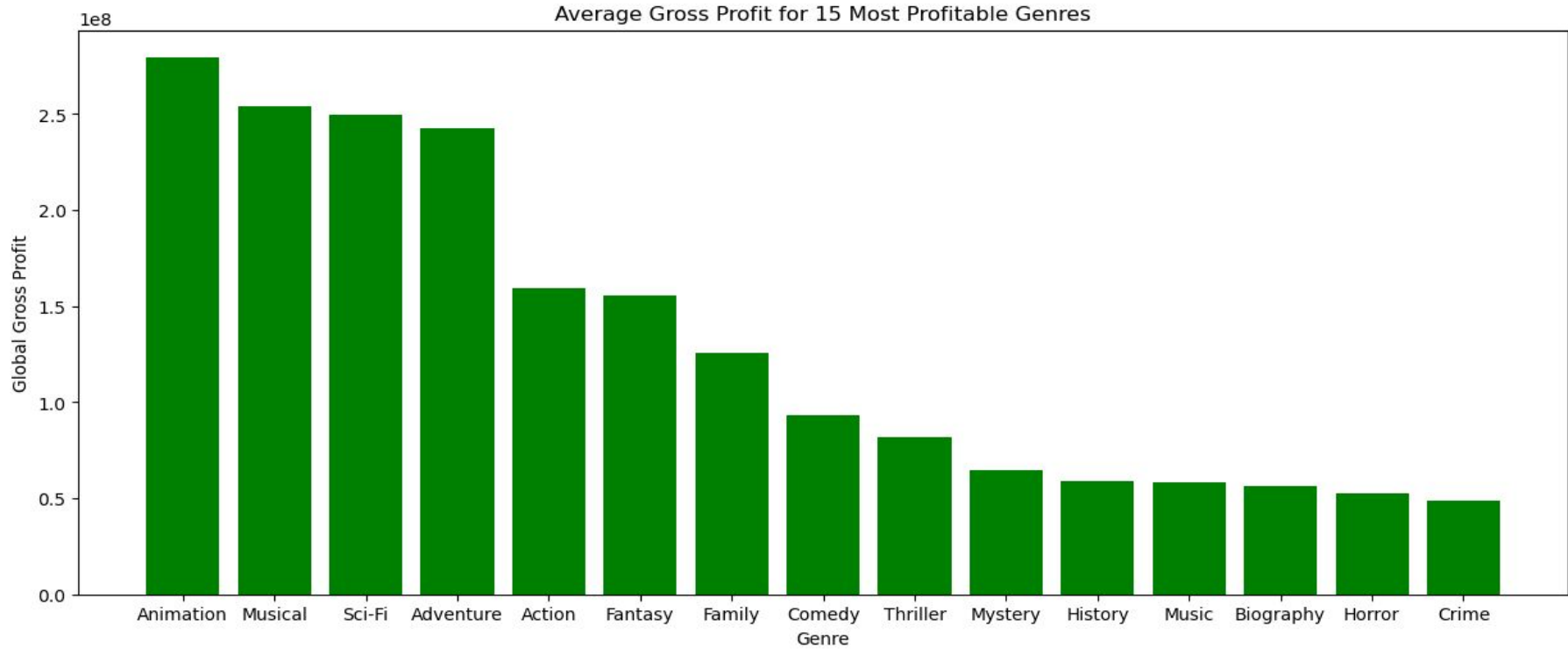
# Annual Gross Profit

This visualization shows the annual gross profit:



# Top 15 genres in Average Gross Profit

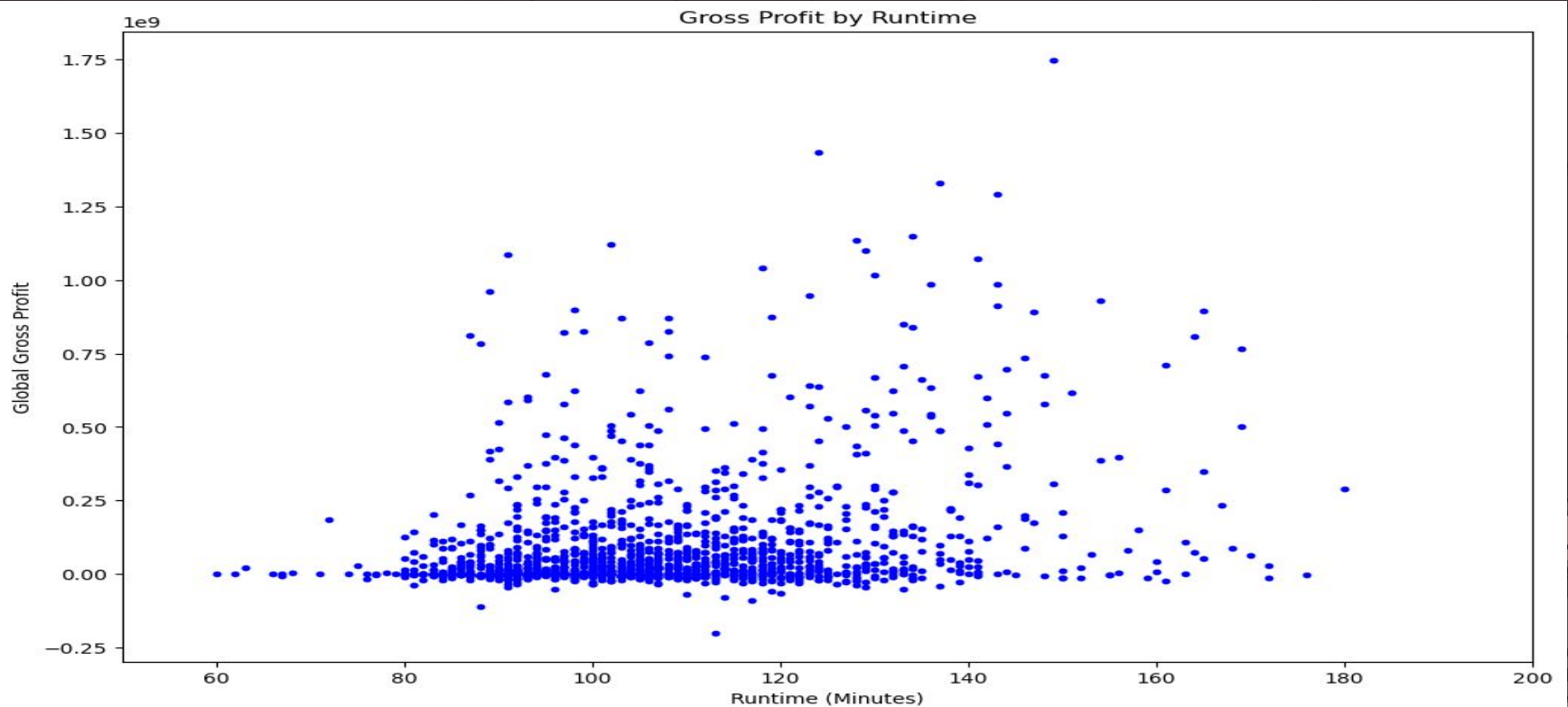
This visualization shows the top 15 genres in terms of average gross profit:





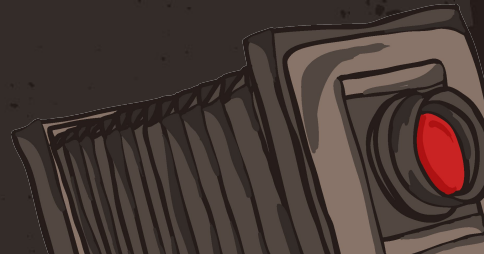
# Runtime against gross profit

This visualization shows runtime against gross profit:



# Conclusion

In conclusion, the comprehensive analysis done has provided useful insights for my company as we venture into the filmmaking industry. It is important that these recommendations be included together with other factors to make the best decision for the company. These insights will act as a fundamental framework, enabling us to understand the kinds of movies that appeal to audiences the most and are profitable.





Cinema is a matter  
of what's in the  
frame and what's  
out.

Martin Scorsese

BrainyQuote®





THANK YOU!

- Here are my Contact Details:
- Email: [jeffr9471@gmail.com](mailto:jeffr9471@gmail.com)
- GitHub: <https://github.com/Jeffrey-sinos>
- LinkedIn: <https://www.linkedin.com/in/jeff-rogers64925124a/>

