

---

# The Movie Industry

---

**Group 2:** Geunwoo Park(G01428707), Hyeyeon Kweon(G01446058),  
Juheon Kim(G01280321), Jungbin Eom (G01447366),  
Soobin Jang (G01399570), WonJune Lee (G01403890)

---

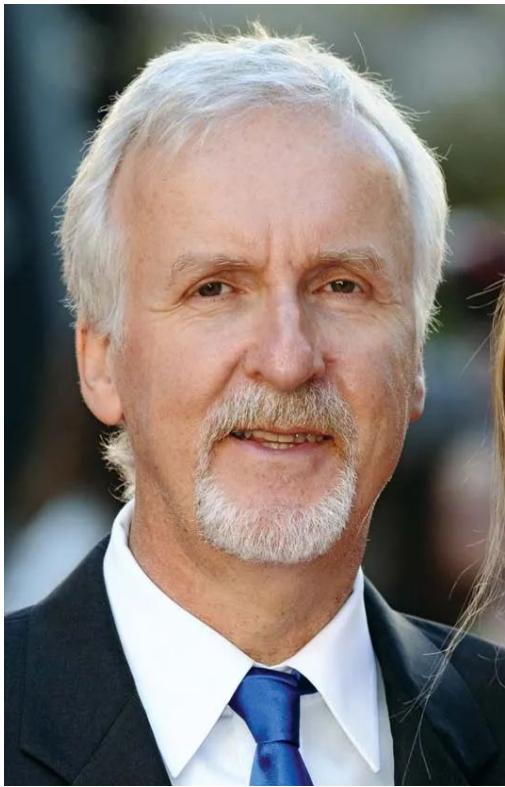




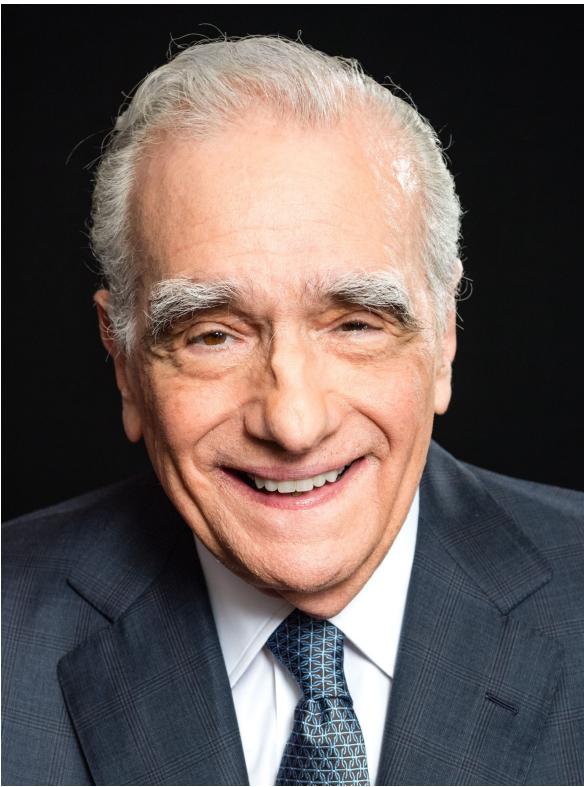


[COVID's Impact on the Film Industry: "The Biggest Shift in the History of Hollywood"](#)





James Cameron



Martin Scorsese

**“Cinemas may not  
survive the impact  
of the pandemic.”**



4D Cinema Events



Drive-in Theaters

# Movie Industry Dataset

## ❑ Movie Industry

- ❑ <https://www.kaggle.com/datasets/danielgrijalvas/movies>

## ❑ Variables:

- ❑ Genre
- ❑ Year
- ❑ Score
- ❑ Votes
- ❑ Budget
- ❑ Gross
- ❑ Director and Star

movies.csv (1.35 MB)

Detail   Compact   Column

### About this file

Three decades of movie data (1986-2016), scraped from IMDb using Python.

name name of the movie	rating rating of the movie (R, PG-13, etc.)	genre main genre of the film	# year release year	released release date	# score IMDb score	# votes count of IMDb user votes	director the director of the film
<b>7512</b> unique values	R PG-13 Other (1859)	48% 28% 24%	Comedy Action Other (3718)	29% 22% 48%		<b>3415</b> unique values	
The Shining	R	Drama	1980	June 13, 1980 (United States)	8.4	927000.0	Stanley Kubrick
The Blue Lagoon	R	Adventure	1980	July 2, 1980 (United States)	5.8	650000.0	Randal Kleiser
Star Wars: Episode V - The Empire Strikes Back	PG	Action	1980	June 20, 1980 (United States)	8.7	1200000.0	Irvin Kershner
Airplane!	PG	Comedy	1980	July 25, 1980 (United States)	7.7	221000.0	Jim Abrahams
Caddyshack	R	Comedy	1980	May 9, 1980 (United States)	7.3	188000.0	Harold Ramis
Friday the 13th	R	Horror	1980	June 28, 1980 (United States)	6.4	123000.0	Sean S. Cunningham
The Blues Brothers	R	Action	1980	June 28, 1980 (United States)	7.9	188000.0	John Landis

kaggle



10 of 15 column

Add Sugges

**“What are the factors that contribute to the success of films in the film industry?”**



# Reading Dataset

```
df = pd.read_csv("movies.csv")
df.head()
```

	name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	company	runtime
0	The Shining	R	Drama	1980	June 13, 1980 (United States)	8.4	927,000.0	Stanley Kubrick	Stephen King	Jack Nicholson	United Kingdom	19,000,000.0	46,998,772.0	Warner Bros.	146.0
1	The Blue Lagoon	R	Adventure	1980	July 2, 1980 (United States)	5.8	65,000.0	Randal Kleiser	Henry De Vere Stacpoole	Brooke Shields	United States	4,500,000.0	58,853,106.0	Columbia Pictures	104.0
2	Star Wars: Episode V - The Empire Strikes Back	PG	Action	1980	June 20, 1980 (United States)	8.7	1,200,000.0	Irvin Kershner	Leigh Brackett	Mark Hamill	United States	18,000,000.0	538,375,067.0	Lucasfilm	124.0
3	Airplane!	PG	Comedy	1980	July 2, 1980 (United States)	7.7	221,000.0	Jim Abrahams	Jim Abrahams	Robert Hays	United States	3,500,000.0	83,453,539.0	Paramount Pictures	88.0
4	Caddyshack	R	Comedy	1980	July 25, 1980 (United States)	7.3	108,000.0	Harold Ramis	Brian Doyle-Murray	Chevy Chase	United States	6,000,000.0	39,846,344.0	Orion Pictures	98.0

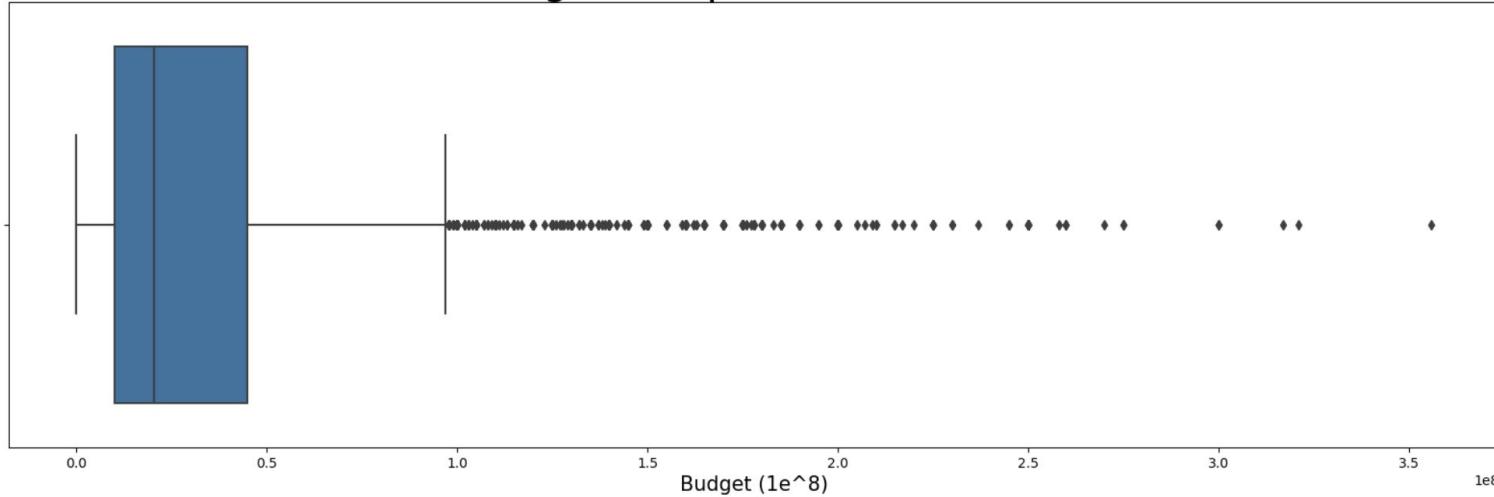
# Sample

```
df.sample(10, random_state=55)
```

		name	rating	genre	year	released	score	votes	director	writer	star	country	budget	gross	compan
3133	Bean	PG-13	Adventure	1997		November 7, 1997 (United States)	6.5	100,000.0	Mel Smith	Rowan Atkinson	Rowan Atkinson	United Kingdom	18,000,000.0	251,212,670.0	Polygram Film Entertainment
6747	Irrational Man	R	Comedy	2015		August 7, 2015 (United States)	6.6	60,000.0	Woody Allen	Woody Allen	Joaquin Phoenix	United States	11,000,000.0	27,391,084.0	Gravie Production
1336	Monkey Shines	R	Drama	1988		July 29, 1988 (United States)	6.2	11,000.0	George A. Romero	Michael Stewart	Jason Beghe	United States	7,000,000.0	5,344,577.0	Orion Picture
6343	The Heat	R	Action	2013		June 28, 2013 (United States)	6.6	165,000.0	Paul Feig	Katie Dippold	Sandra Bullock	United States	43,000,000.0	229,930,771.0	Twentieth Century Fox
6846	Captain America: Civil War	PG-13	Action	2016		May 6, 2016 (United States)	7.8	694,000.0	Anthony Russo	Christopher Markus	Chris Evans	United States	250,000,000.0	1,153,337,496.0	Marvel Studio
5774	All Good Things	R	Crime	2010		December 3, 2010 (Canada)	6.3	54,000.0	Andrew Jarecki	Marcus Hinchey	Ryan Gosling	United States	NaN	1,754,389.0	Groundswell Production
4699	Lord of War	R	Action	2005		September 16, 2005 (United States)	7.6	300,000.0	Andrew Niccol	Andrew Niccol	Nicolas Cage	France	50,000,000.0	72,617,068.0	Entertainment Manufacturing Company
277	The Final Option	R	Action	1982		September 16, 1983 (United States)	6.4	3,000.0	Ian Sharp	George Markstein	Lewis Collins	Switzerland	6,000,000.0	2,666,873.0	Richard Light Horse Production
2395	Mr. Nanny	PG	Action	1993		October 8, 1993 (United States)	3.8	8,900.0	Michael Gottlieb	Edward Ruggoff	Hulk Hogan	United States	10,000,000.0	4,348,572.0	New Line Cinema

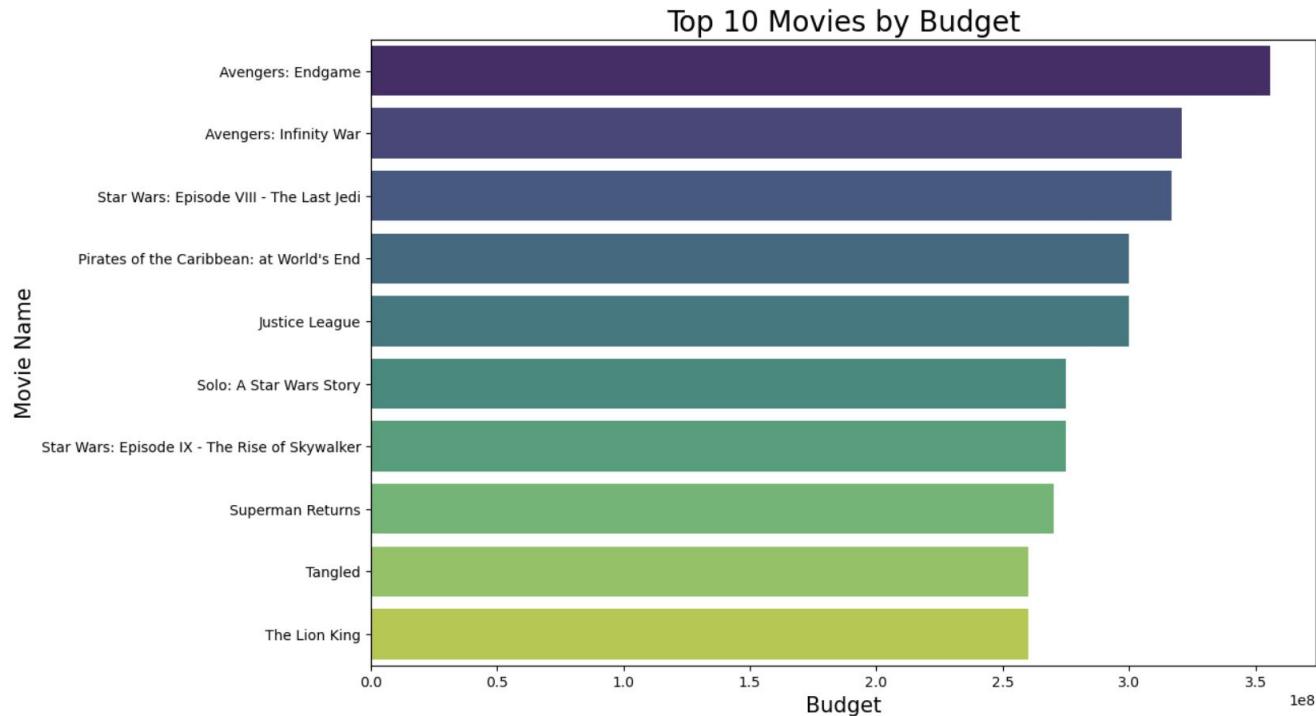
# Budget Outlier

Budget box plot with outliers



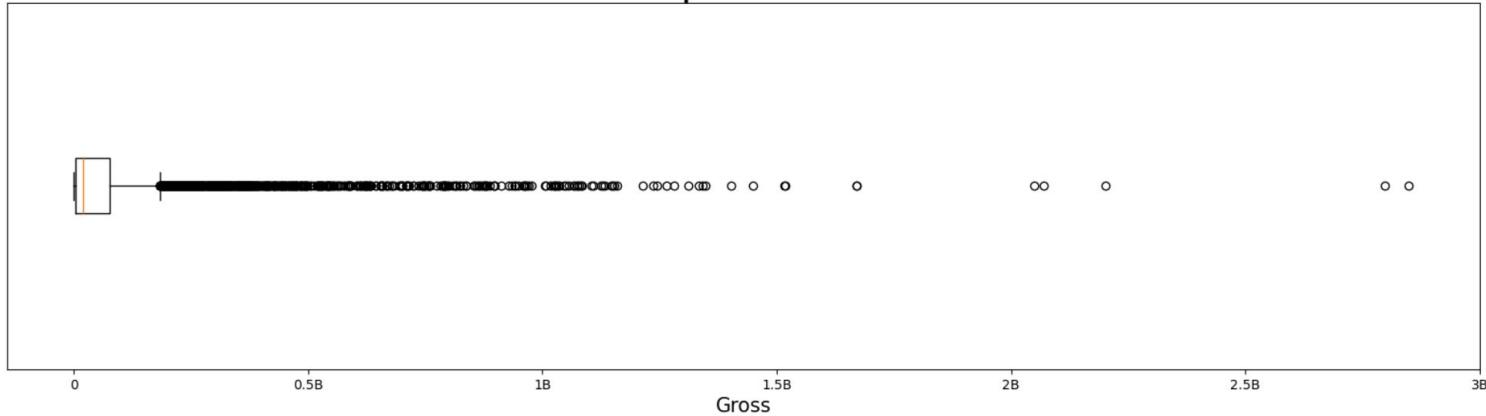
- **Median Budget:** Located near center, typical budget indicator.
- **IQR:** Narrow, majority of budgets are similar.
- **Positive Skewness:** Longer right whisker, higher budget movies.
- **Outliers:** Numerous on right side, very high budgets.

# The outliers are ...



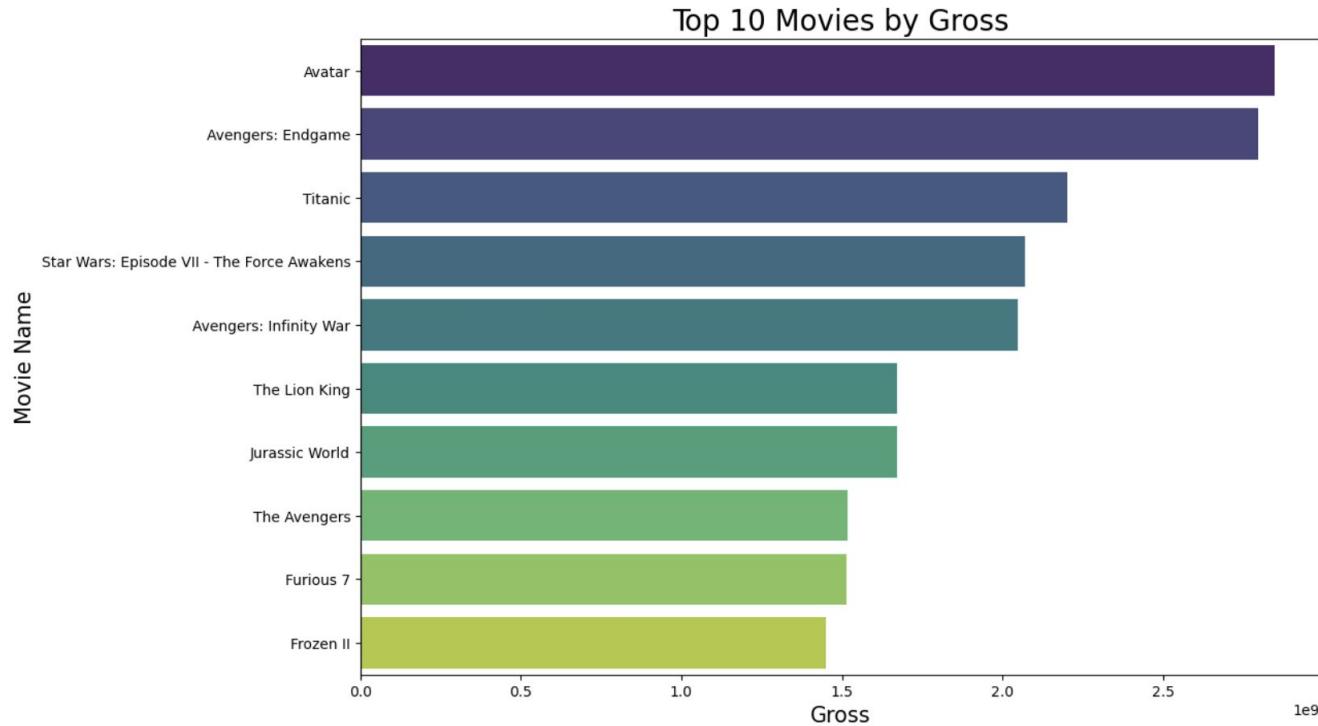
# Gross Outlier

Gross box plot with outliers

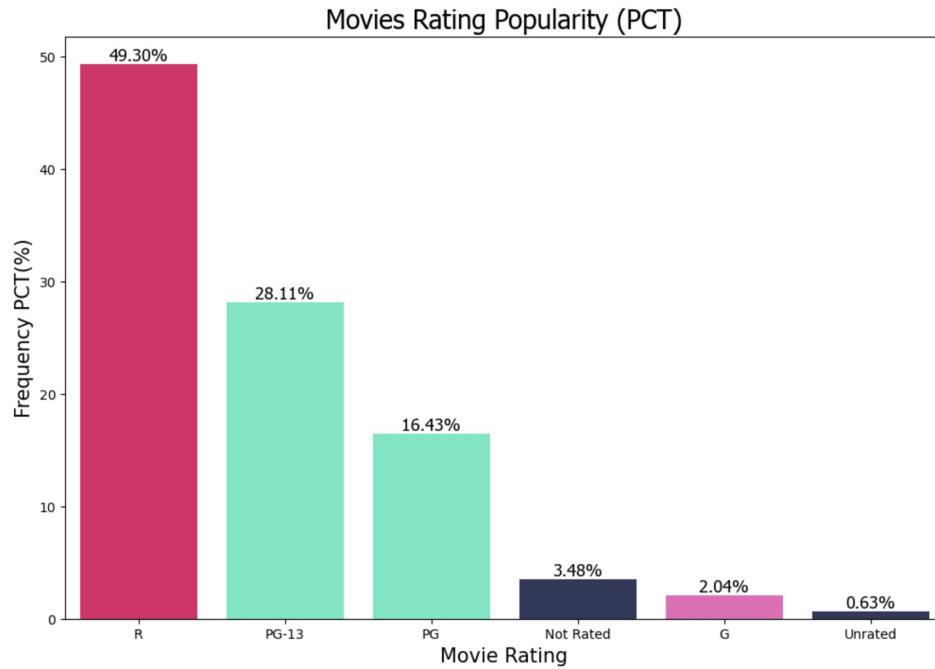


- **Outliers:** Dots beyond whiskers, significantly higher/lower earnings.
- **Median:** Near left side, typical earnings for most movies.
- **Skewness:** Positive, Longer right whisker, higher earnings for some movies.
- **Right Side Outliers:**
  - High gross earnings, Up to and beyond 3 billion dollars.
  - Likely blockbusters or internationally appealing movies.

# The outliers are ...

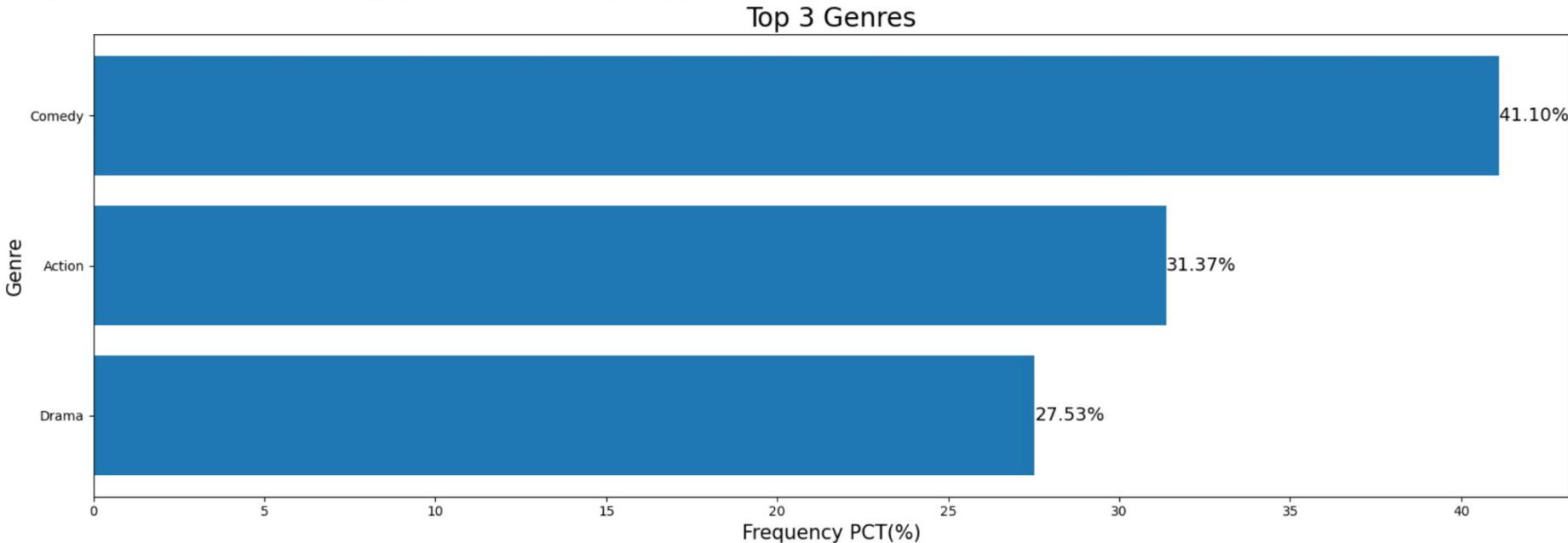


# Movies Rating Popularity (PCT)

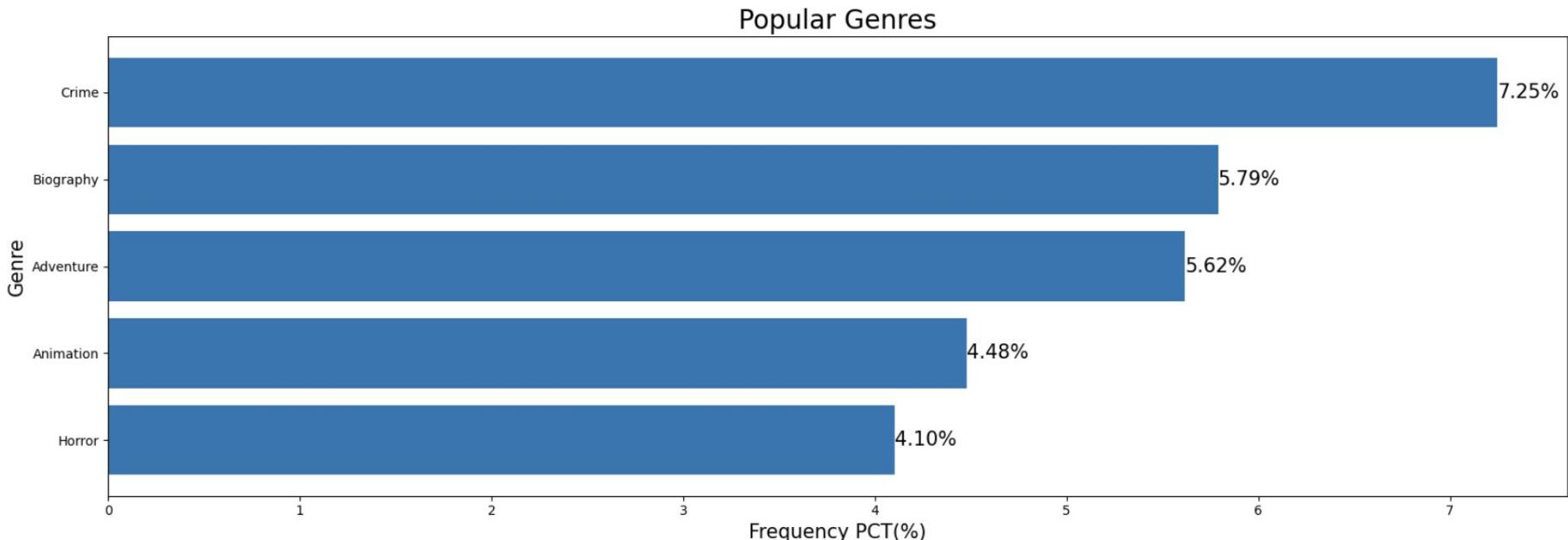


- **Most Common Rating :** 'R', 49.30%, mature audience
- **Second Most Common:** 'PG-13', 28.11%, suitable for teenagers with parental guidance.
- **PG Rating:** 16.43%, family-friendly.
- **Less Common Ratings:**
  - 'Not Rated': 3.48%, various films without official rating.
  - 'G': 2.04%, general audiences.
  - 'Unrated': 0.63%, negligible portion.

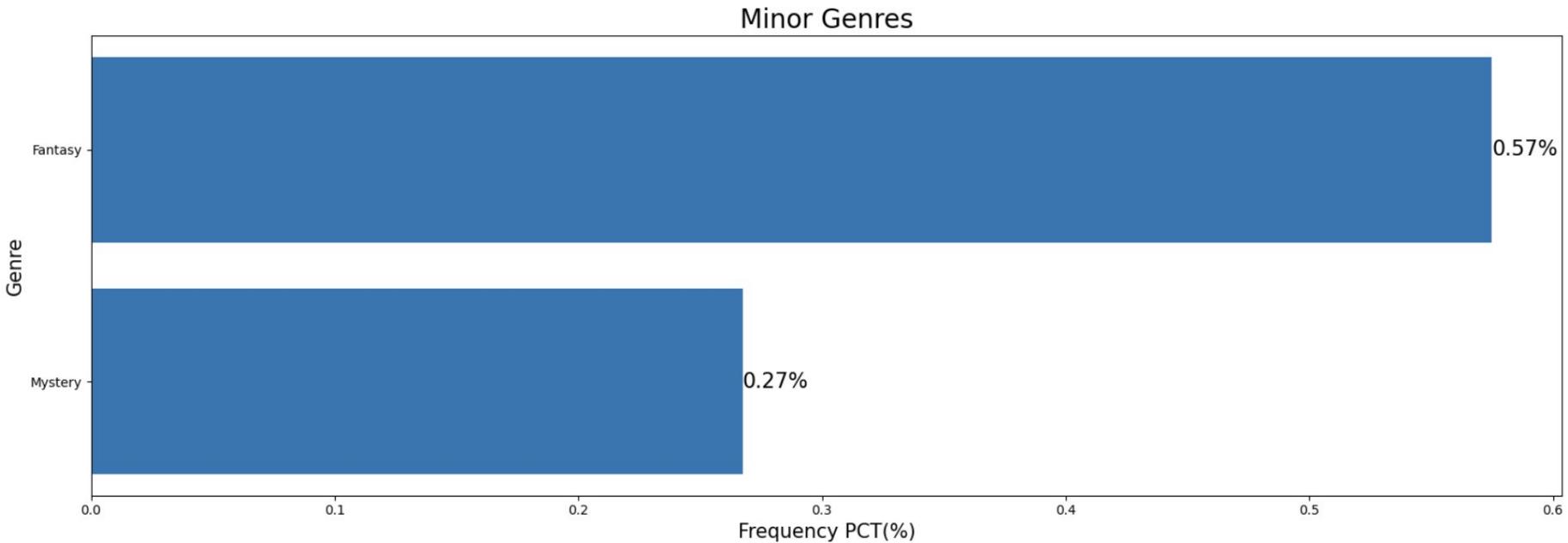
# Top 3 Movie Genres & Popularity (PCT)



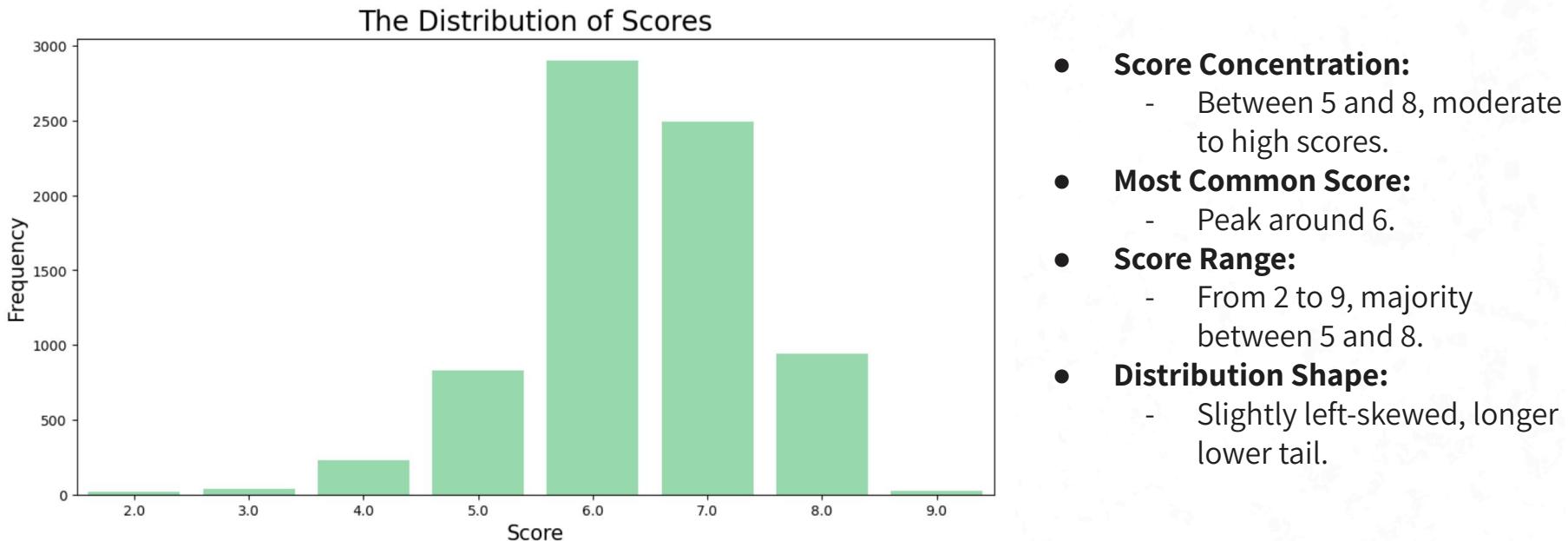
# 5 Popular Movie Genres & Popularity (PCT)



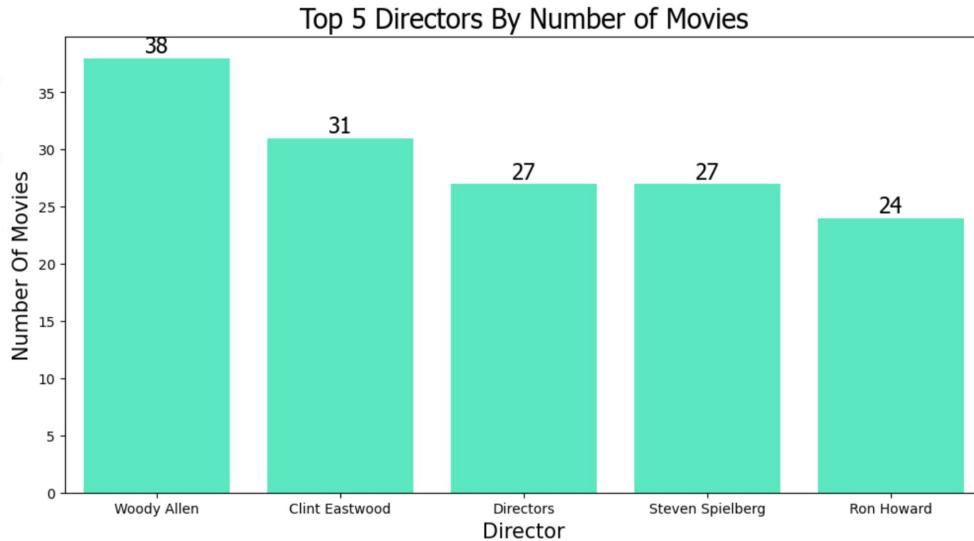
## 2 Minor Movie Genres & Popularity (PCT)



# The Distribution of Scores



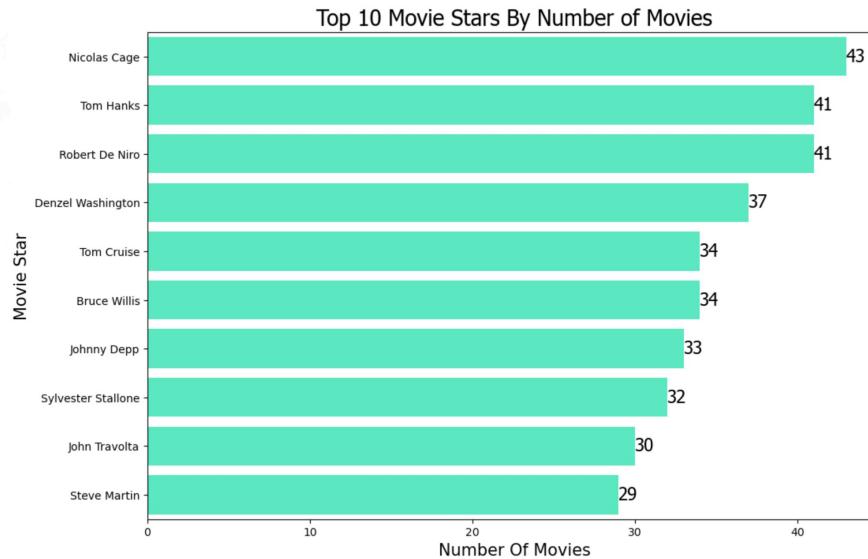
# Top 5 Directors By Number of Movies



- **Woody Allen:**
  - 38 movies, most prolific director.
- **Clint Eastwood:**
  - 31 movies, known for versatility and long-standing career.
- **Steven Spielberg:**
  - 27 movies, significant contributions to art and commerce in cinema.
- **Unnamed Director 1:**
  - 27 movies.
- **Ron Howard:**
  - 24 movies, versatile across various genres.



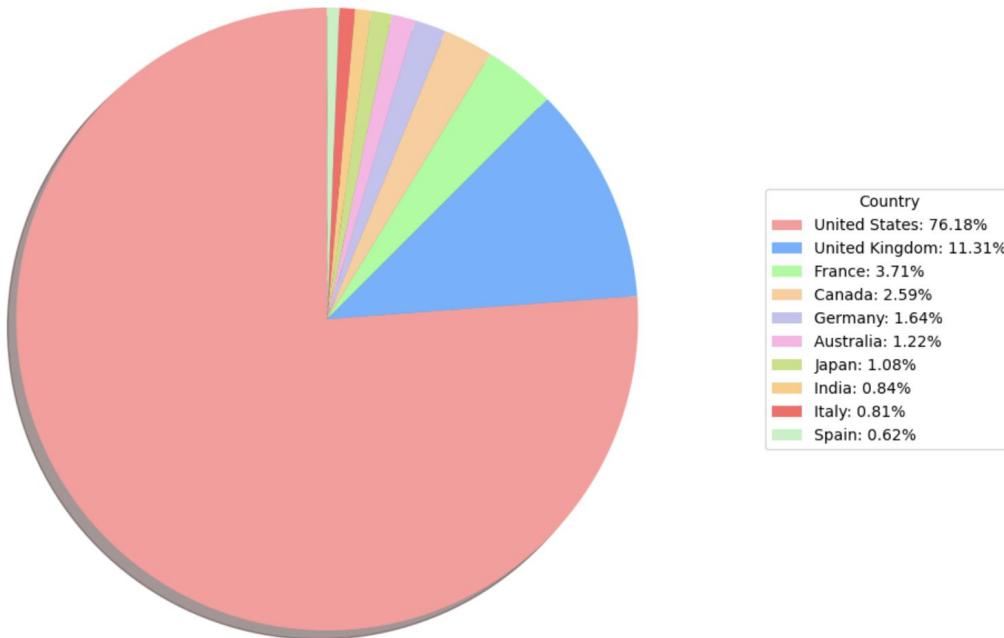
# Top 10 Movie Stars By Number of Movies



- **Nicolas Cage:** 43 movies, most prolific actor.
- **Tom Hanks & Robert De Niro:** 41 movies each, versatile and long-standing careers.
- **Denzel Washington:** 37 movies, known for powerful performances.
- **Tom Cruise & Bruce Willis:** 34 movies each, famous for action roles.
- **Johnny Depp:** 33 movies, known for eclectic roles.
- **Sylvester Stallone:** 32 movies, iconic roles in Rocky and Rambo.
- **John Travolta:** 30 movies, significant in drama and comedy.
- **Steve Martin:** 29 movies, notable for diverse roles in comedy and drama.

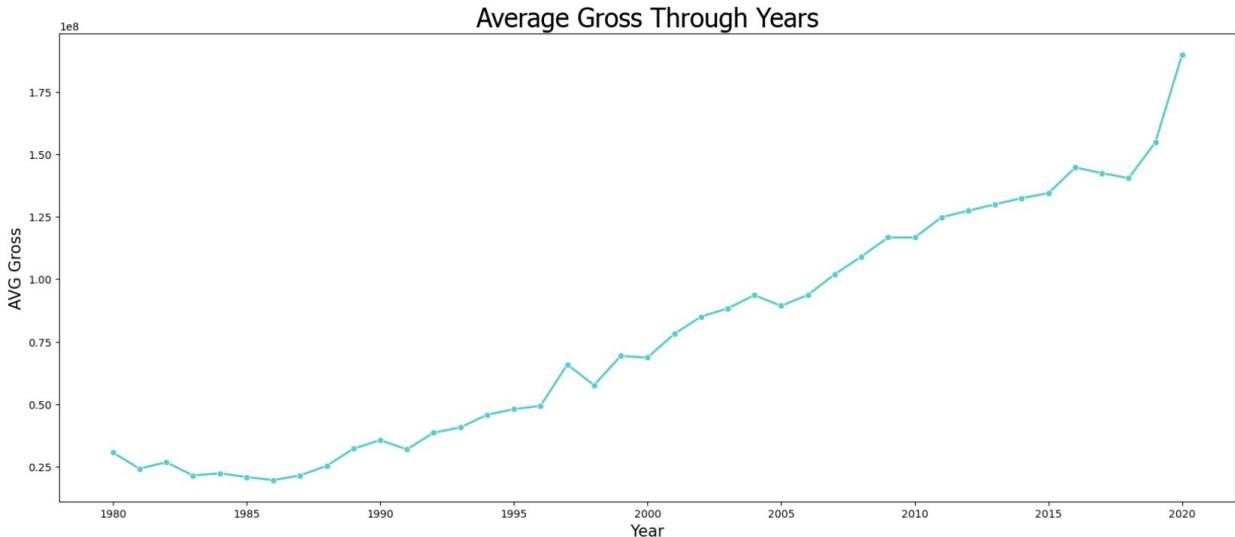
# Top 10 Country By Released Movies

## Top 10 Countries by Released Movies



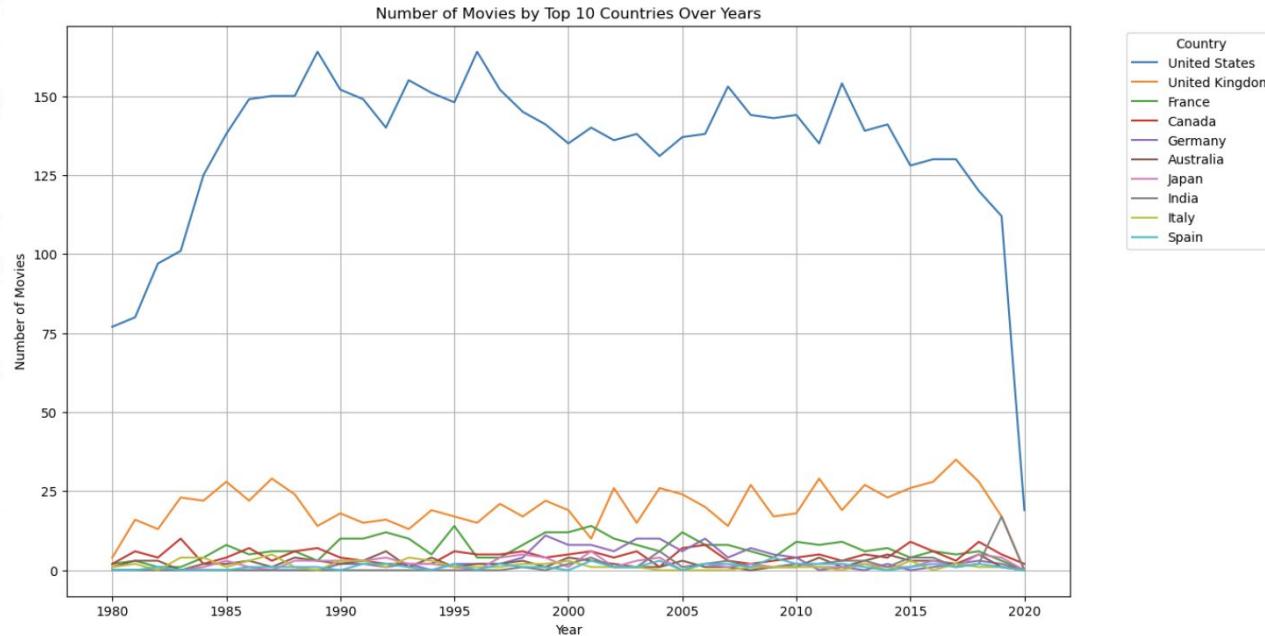
- **Dominance:** The United States leads significantly with 76.18% of movie releases.
- **Major Players:** The United Kingdom (11.31%) and France (3.71%) follow.
- **Diversity:** The remaining countries (Canada, Germany, Australia, Japan, India, Italy, and Spain) each contribute less than 3%, indicating a concentrated distribution of movie production in a few key countries.

# AVG Gross Through Years



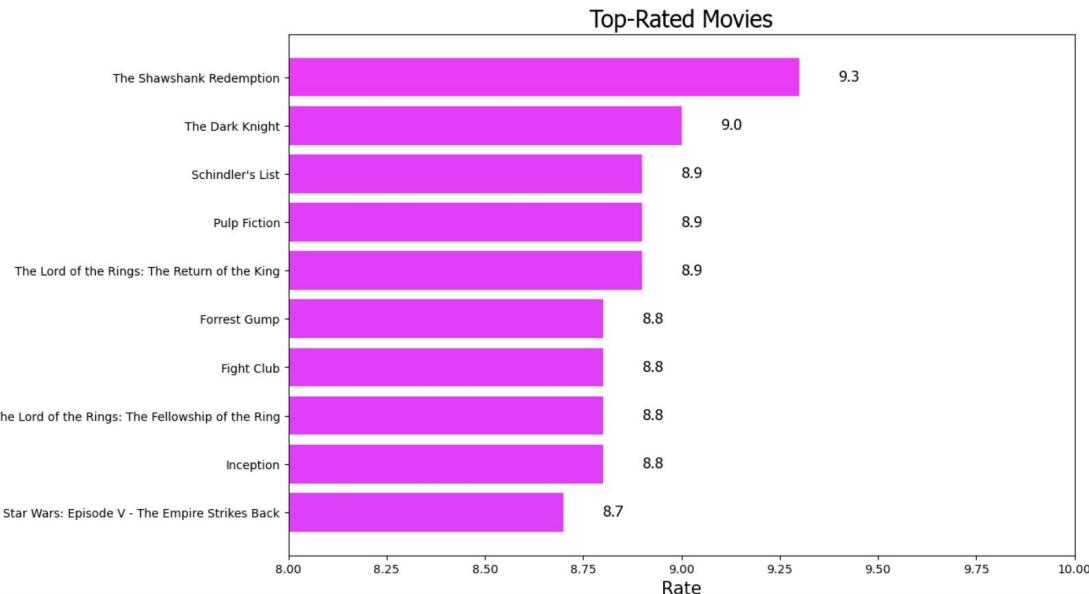
- **Trend:** Average gross revenue has been steadily increasing from 1980 to 2020.
- **Inflection Points:**
  - Mid-1980s: Initial upward shift.
  - Late 1990s: Significant rise, likely due to blockbusters and global markets.
  - 2010s to 2020: Sharp increase, driven by technology and streaming services.

# Number of Movies Through Years



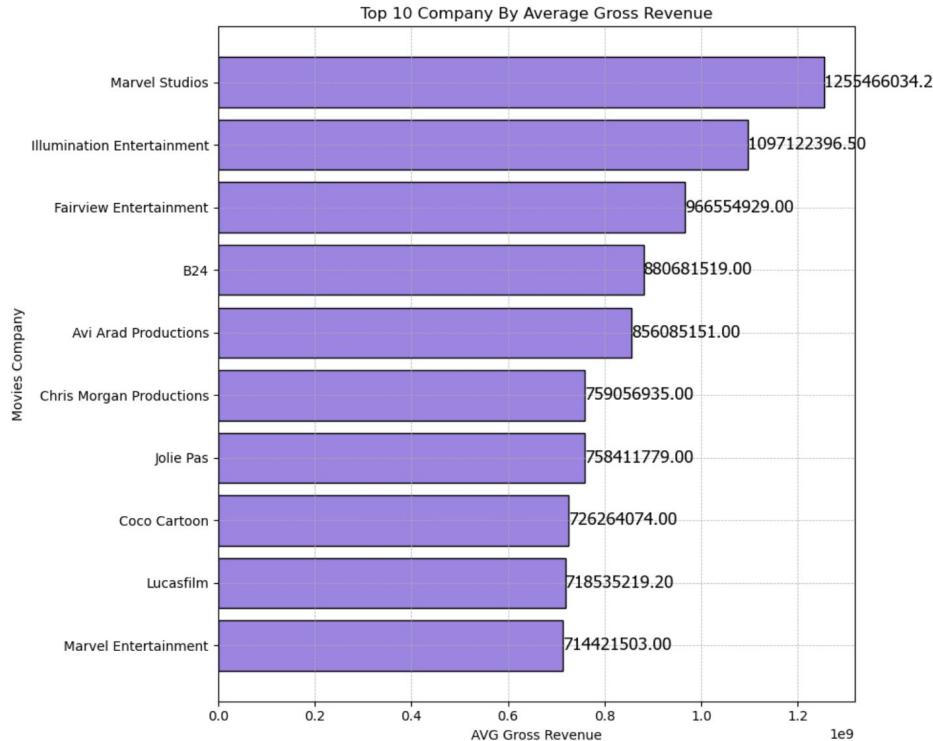
- **Market Dominance:** The U.S. has been the dominant player in the movie industry for decades, which may influence global film trends and market dynamics.
- **Stability:** Other countries have consistent but lower levels of movie production, indicating more stable or niche markets.
- **Impact of External Factors:** The sharp decline around 2020 could be due to external factors like the COVID-19 pandemic affecting movie production globally.

# Top-Rated Movies



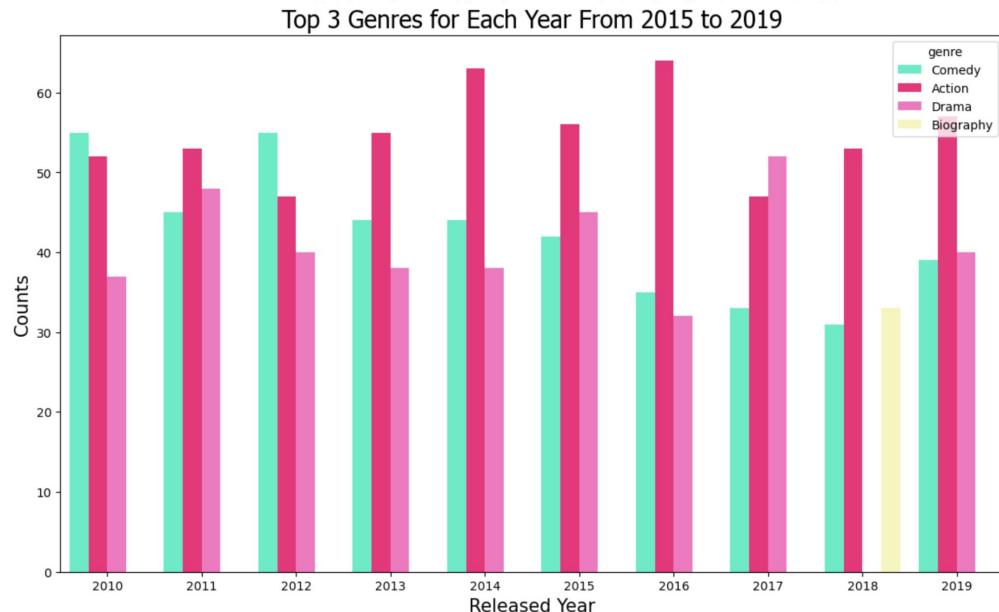
- **The Shawshank Redemption:** Highest score, 9.3.
- **The Dark Knight:** Score of 9.0, highly acclaimed.
- **Schindler's List, Pulp Fiction, The Lord of the Rings: The Return of the King:** Each with a score of 8.9, renowned for excellence.
- **Forrest Gump, Fight Club, The Lord of the Rings: The Fellowship of the Ring, Inception:** Each with a score of 8.8, recognized for storytelling.
- **Star Wars: Episode V - The Empire Strikes Back:** Score of 8.7, lasting appeal.

# Top 10 Company By Average Gross Revenue



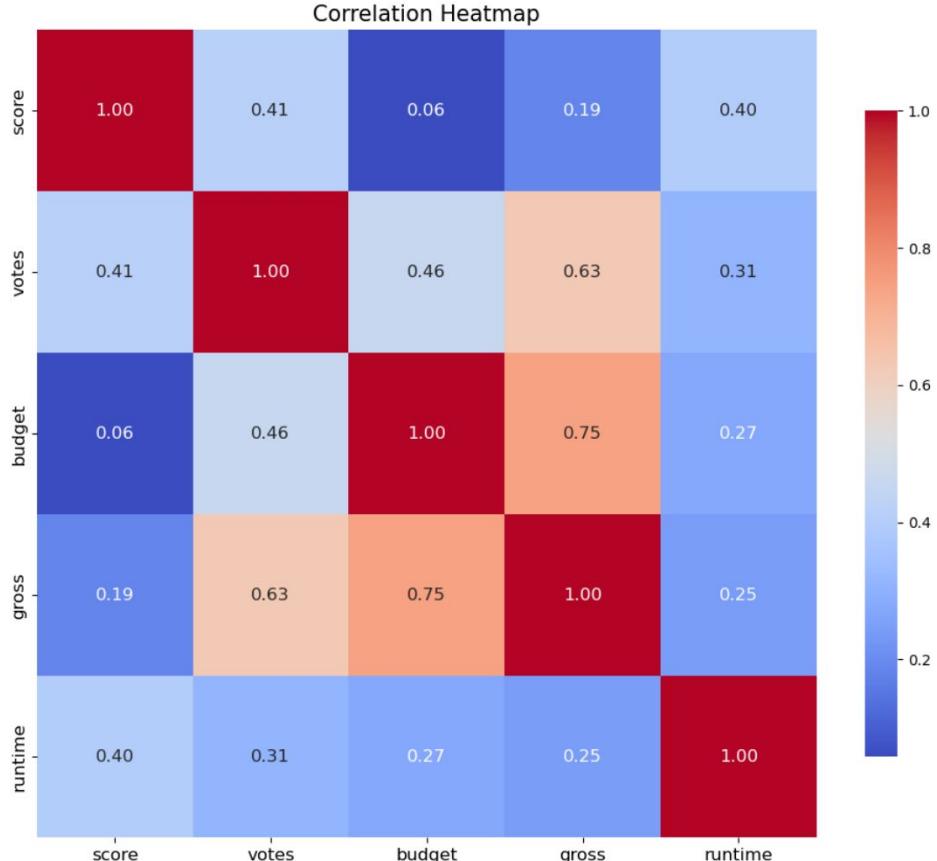
- **Top Company:** Marvel Studios leads with \$1.255 billion average gross revenue.
- **High Performers:** Illumination Entertainment and Fairview Entertainment follow with over \$1 billion each.
- **Overall:** The top 10 companies consistently generate high average gross revenues, indicating strong market positions.

# Top 3 Genres for Each Year From 2015 to 2019



- **Consistent Trends:**
  - **Comedy & Drama:** Appear in top 3 across all years, indicating sustained popularity and significant production volumes.
  - **Action:** Frequently in top 3, reflecting consistent appeal for high-energy films.
- **Notable Yearly Trends:**
  - **2016 & 2017:** Action movies dominate, higher count compared to Comedy and Drama.
  - **2019:** Increased production of Biography films, suggesting a trend in biographical storytelling.

# Correlation Heatmap



Focus on **strong correlation...**

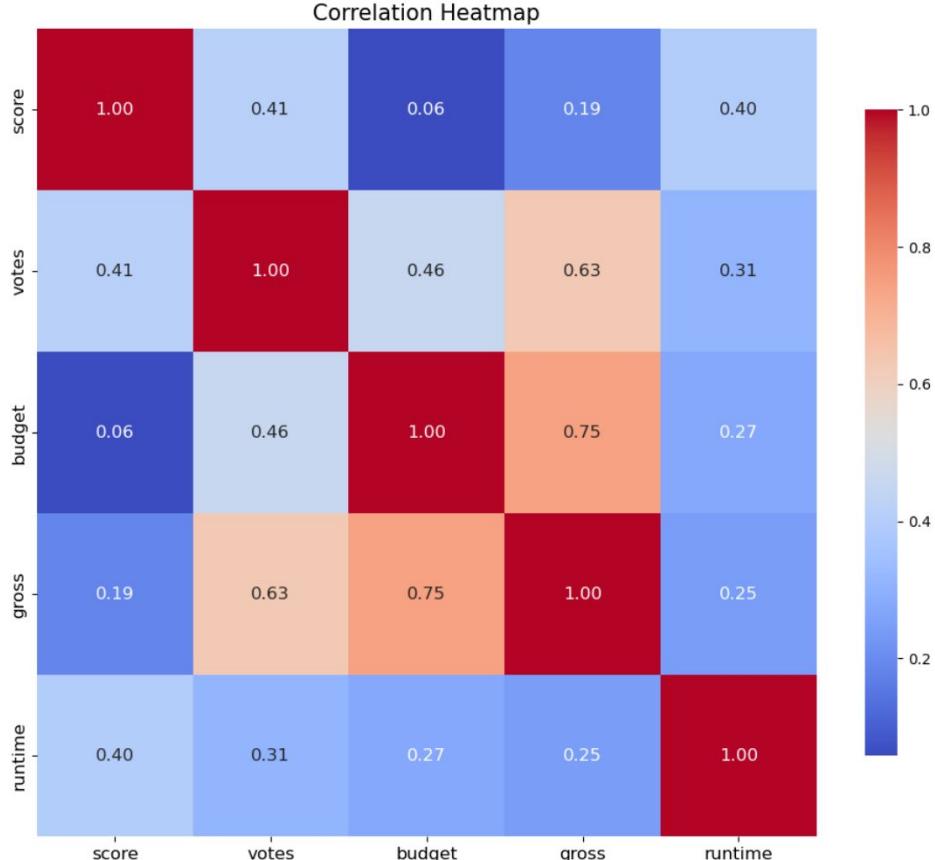
- **Votes & Gross(0.63)**

- Explanation: Higher vote counts correlate strongly with higher gross earnings.
- Implication: Increasing audience engagement can boost a movie's financial success.

- **Budgets & Gross(0.75)**

- Explanation: Larger budgets strongly correlate with higher gross revenue.
- Implication: Investing more in production often leads to higher earnings.

# Correlation Heatmap



Focus on **weak correlation...**

- **Score and Budget(0.06)**
  - Explanation: Almost no relationship between budget and score.
  - Implication: High budgets do not necessarily result in better scores.
- **Score and Gross(0.19)**
  - Explanation: Higher scores slightly correlate with higher gross earnings.
  - Implication: Quality contributes to financial success but is not a strong determinant.

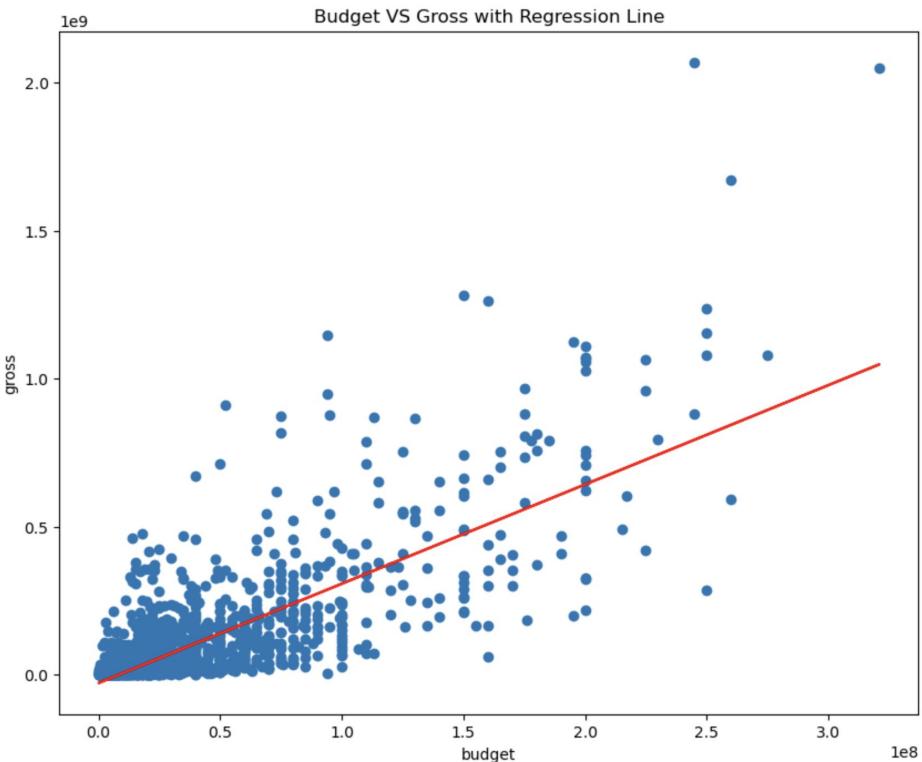
# Linear Regression

---

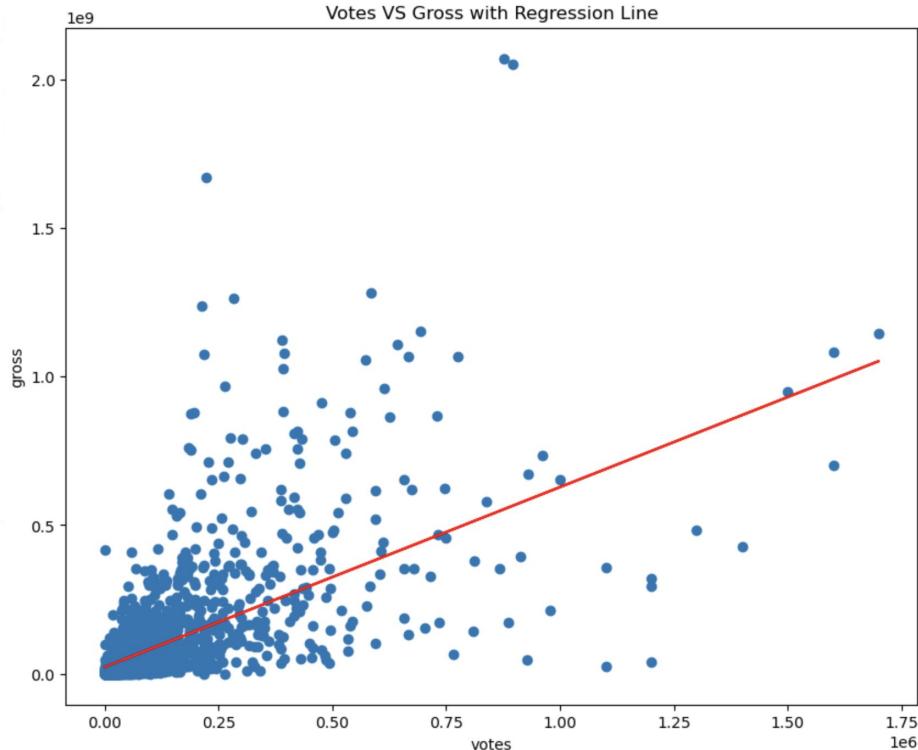
1. Budget vs Gross
  2. Votes vs Gross
  3. Score vs Gross
  4. Runtime vs Gross
-

# Budget VS Gross with Regression Line

- **R-squared value:** 0.6165
- **RMSE:** 106403258.8573
- **Trend:** Significant positive correlation between budget and gross revenue.
- **Distribution:** Most movies have budgets below \$100 million and gross revenues below \$500 million; few outliers.
- **Conclusion:** Higher budgets tend to result in higher gross revenue, indicating a stronger impact of budget on movie success.

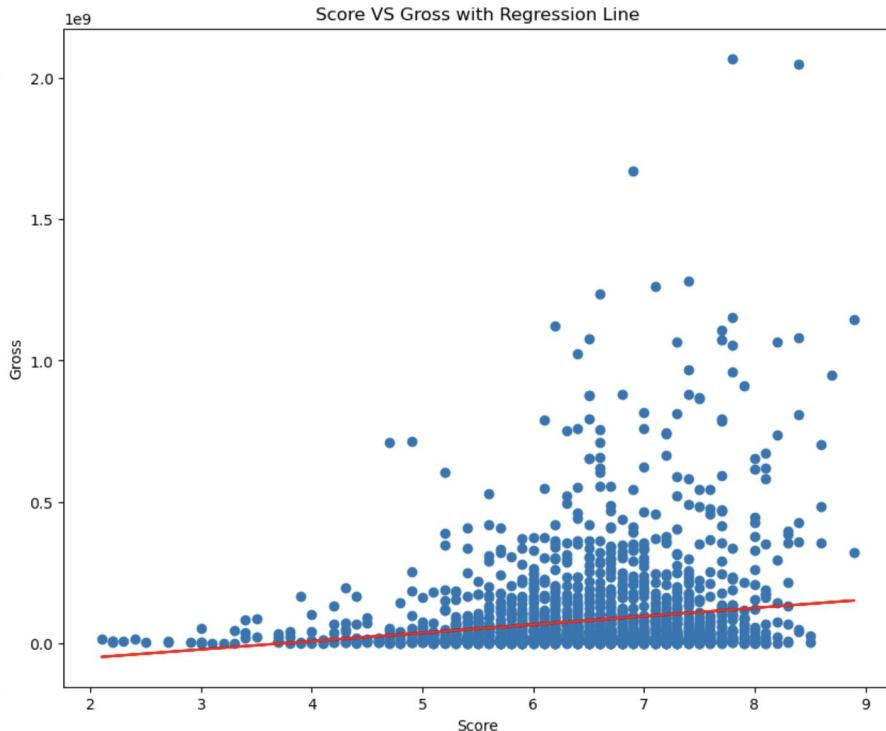


# Votes VS Gross with Regression Line



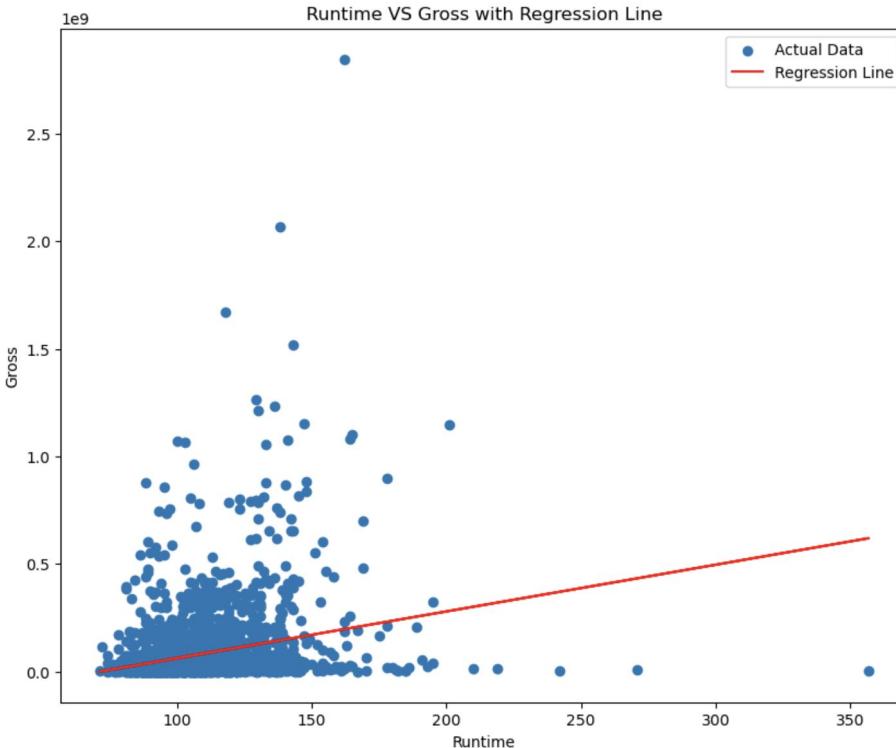
- **R-squared value:** 0.4342
- **RMSE:** 129238650.8175
- **Trend:** Strong positive correlation between votes and gross revenue.
- **Distribution:** Most movies have votes below 250,000 and gross revenues below \$500 million; few outliers.
- **Conclusion:** Higher vote counts tend to result in higher gross revenue, indicating a strong impact of popularity on movie success.

# Score VS Gross with Regression Line



- **R-squared value:** 0.0409
- **RMSE:** 168261955.8808
- **Trend:** Slight increase in revenue with higher scores, but weak correlation.
- **Distribution:** Most movies have scores between 5-8 and gross revenue between \$0-500 million; few outliers.
- **Conclusion:** Score has a weak impact on revenue, suggesting other factors are more influential.

# Runtime VS Gross with Regression Line



- **R-squared value:** 0.0638
- **RMSE:** 166261057.7184
- **Trend:** A slight increase in revenue with longer runtimes, but weak correlation.
- **Distribution:** Most movies are 100-150 minutes with \$0-500 million revenue; few outliers.
- **Conclusion:** Runtime has a weak impact on revenue, suggesting other factors are more influential.

# Multiple Regression

```
df_cleaned = df.dropna(subset=['budget', 'votes', 'score', 'runtime', 'gross'])

X = df_cleaned[['budget', 'votes', 'score', 'runtime']]
y = df_cleaned['gross']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=101)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

r_squared = r2_score(y_test, y_pred)
rmse = mean_squared_error(y_test, y_pred, squared=False)

print(f'R-squared: {r_squared:.4f}')
print(f'Root Mean Squared Error (RMSE): {round(rmse, 4)}')
```

R-squared: 0.6958

Root Mean Squared Error (RMSE): 94767387.124

- **R-squared value:** relatively strong relationship
- **RMSE :** High. Model captures the trend but has significant prediction errors
- **Decision Making :** Informs budget allocation, marketing strategies, and runtime decisions
- **Model Improvement :** Potential to increase accuracy with additional features.

# Conclusions

- **High Budget Impact:** Higher budgets significantly boost movie gross earnings.
- **Audience Engagement:** Increased engagement through votes enhances financial success.
- **Genre Popularity:** Comedy, action, and drama dominate, but niche genres have potential.
- **Target Audiences:** Focus on mature and teenage audiences based on rating distribution.
- **Industry Growth:** Steady growth observed, but vulnerable to global disruptions like COVID-19.

# Conclusions

- **Key Contributors:** Prolific directors and actors significantly influence trends.
- **Investment in Production:** High-quality production and marketing are crucial for success.
- **Adaptable Strategies:** Implement flexible release strategies, including OTT platforms.
- **Innovative Marketing:** Use innovative marketing techniques to reach and engage audiences.
- **Data-Driven Decisions:** Continuous data analysis is essential for informed decision-making.

# Implications and Recommendations

## 01 Investment in Production

high budgets for quality production, marketing, and securing top talent

## 02 Audience Engagement

Active marketing and engagement strategies – utilizing social media, interactive content, and engaging with audiences

# Implications and Recommendations

## 03    Genre Focus

Primary focus on comedy, action, drama, but also exploring and investing in niche genres

## 04    Adaptability to Global Events

develop strategies to mitigate the impact of global disruptions like the COVID-19 pandemic  
– flexible release strategies, including OTT platforms, and innovative marketing techniques

## 05    Data-Driven Decisions

Continue on analyzing data

# Conclusion

- **Strategic Recommendations:**
  - Utilize data-driven decision-making.
  - Adopt flexible distribution strategies, including OTT platforms.
  - Embrace innovative marketing techniques.
- **Future Outlook:**
  - Promising through blending traditional and digital experiences; theaters can attract audiences back through continuous adaptation and innovation.

## COVID-19



---

# THANK YOU

---

**CREDITS:** This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

