

# Data Analysis Presentation

...

Jake McClain - INFO450 - 11/17/22

# Preamble

Good Movie Co. is an upstart film company that after a run of three unsuccessful movies, has one last chance for profit before filing for bankruptcy. After their own attempts have failed they want to take a more analytical approach for the creation of a successful movie.



# Problem Statement

I have been hired by this company to analyze the current data in the movie industry to come up with a formula that determines a successful movie. Ideally, Good Movie Co. can follow this formula for a heightened chance for saving their company. The two datasets that I used for this analysis include one that tracks the gross income of movies, and then the IMDB rating of movies. These are the two key metrics that I have determined on what makes for a successful movie, both revenue and high critic reviews.

# The Datasets

## Gross Movies:

	Title	World Sales (in \$)
0	Star Wars: Episode VII - The Force Awakens	2069521700
1	Avengers: Endgame	2797501328
2	Avatar	2847246203
3	Black Panther	1347597973
4	Avengers: Infinity War	2048359754
..	...	...
913	The Notebook	117815157
914	Jimmy Neutron: Boy Genius	102992536
915	Eat Pray Love	204594016
916	The Texas Chainsaw Massacre	107363905
917	Zookeeper	169852759

## IMDB Movies:

	Title	IMDB_Rating	Runtime	Genre	Director
0	The Shawshank Redemption	9.3	142 min	Drama	Frank Darabont
1	The Godfather	9.2	175 min	Crime, Drama	Francis Ford Coppola
2	The Dark Knight	9.0	152 min	Action, Crime, Drama	Christopher Nolan
3	The Godfather: Part II	9.0	202 min	Crime, Drama	Francis Ford Coppola
4	12 Angry Men	9.0	96 min	Crime, Drama	Sidney Lumet
..	...	...	...	...	...
995	Breakfast at Tiffany's	7.6	115 min	Comedy, Drama, Romance	Blake Edwards
996	Giant	7.6	201 min	Drama, Western	George Stevens
997	From Here to Eternity	7.6	118 min	Drama, Romance, War	Fred Zinnemann
998	Lifeboat	7.6	97 min	Drama, War	Alfred Hitchcock
999	The 39 Steps	7.6	86 min	Crime, Mystery, Thriller	Alfred Hitchcock

After inserting the two data sets, I have narrowed down variables that determine what create high sales and reviews, which has been identified as the factors that determine a successful movie.

# Data Merge

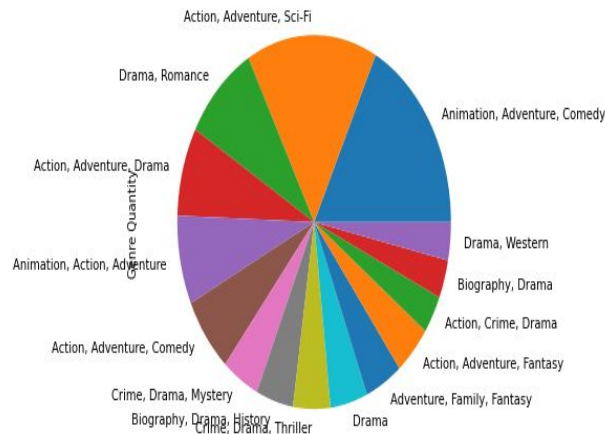
	Title	World Sales (in \$)	IMDB_Rating	Runtime	Genre	Director
0	Star Wars: Episode VII - The Force Awakens	2069521700	7.9	138 min	Action, Adventure, Sci-Fi	J.J. Abrams
1	Avengers: Endgame	2797501328	8.4	181 min	Action, Adventure, Drama	Anthony Russo
2	Avatar	2847246203	7.8	162 min	Action, Adventure, Fantasy	James Cameron
3	Avengers: Infinity War	2048359754	8.4	149 min	Action, Adventure, Sci-Fi	Anthony Russo
4	Titanic	2201647264	7.8	194 min	Drama, Romance	James Cameron
..	...	...	...	...	...	...
180	Die Hard	141603197	8.2	132 min	Action, Thriller	John McTiernan
181	Apocalypse Now	92158064	8.4	147 min	Drama, Mystery, War	Francis Ford Coppola
182	Brokeback Mountain	178062759	7.7	134 min	Drama, Romance	Ang Lee
183	Alien	106285522	8.4	117 min	Horror, Sci-Fi	Ridley Scott
184	The Notebook	117815157	7.8	123 min	Drama, Romance	Nick Cassavetes

I decided to merge these datasets on an inner join, this was that most obvious choice for my analysis because it allows me to track movies with both gross earnings and the imdb rating.

# Closer Look

After inspecting the data we can see that categories are largely dominated by action/adventure movies with different subgenres such as animation, comedy, and sci-fi. Other leading genres include crime dramas and romance films .

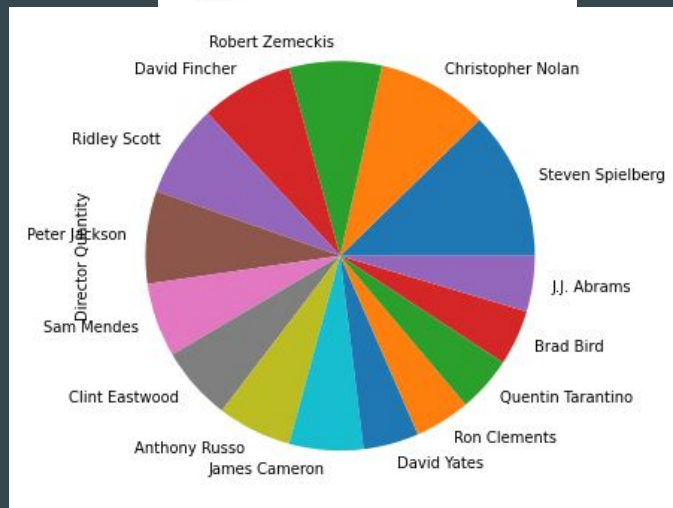
```
Genre
Animation, Adventure, Comedy    16
Action, Adventure, Sci-Fi       14
Drama, Romance                  8
Action, Adventure, Drama        7
Animation, Action, Adventure     7
Action, Adventure, Comedy       6
Crime, Drama, Mystery           4
Biography, Drama, History       4
Crime, Drama, Thriller          4
Drama                           4
Adventure, Family, Fantasy      4
Action, Adventure, Fantasy      4
Action, Crime, Drama            3
Biography, Drama               3
Drama, Western                 3
dtype: int64
```



# Closer Look

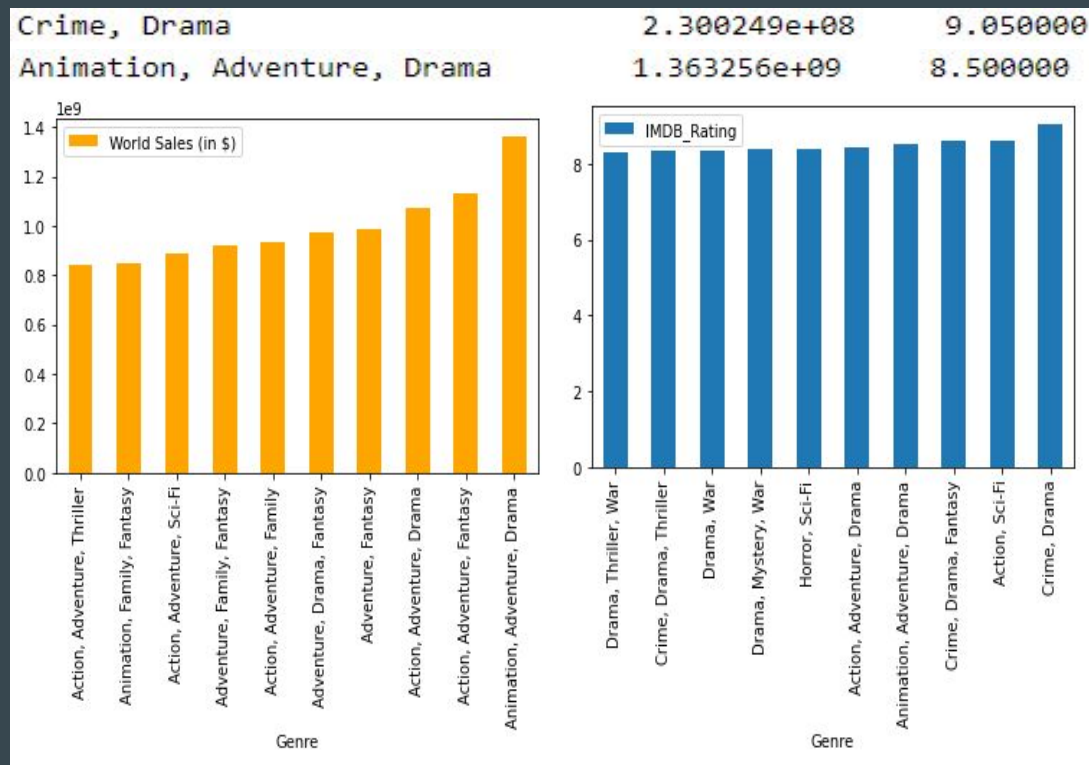
I was also important to obtain the information on the top directors from the merge. Being the leading force on movie production, they are an important factor on driving a successful movie. Spielberg, Nolan, Zemeckis, and Fincher have the highest inclusion from the data merge.

Director	
Steven Spielberg	8
Christopher Nolan	6
Robert Zemeckis	5
David Fincher	5
Ridley Scott	5
Peter Jackson	5
Sam Mendes	4
Clint Eastwood	4
Anthony Russo	4
James Cameron	4
David Yates	3
Ron Clements	3
Quentin Tarantino	3
Brad Bird	3
J.J. Abrams	3
dtype: int64	



# Data Grouping 1

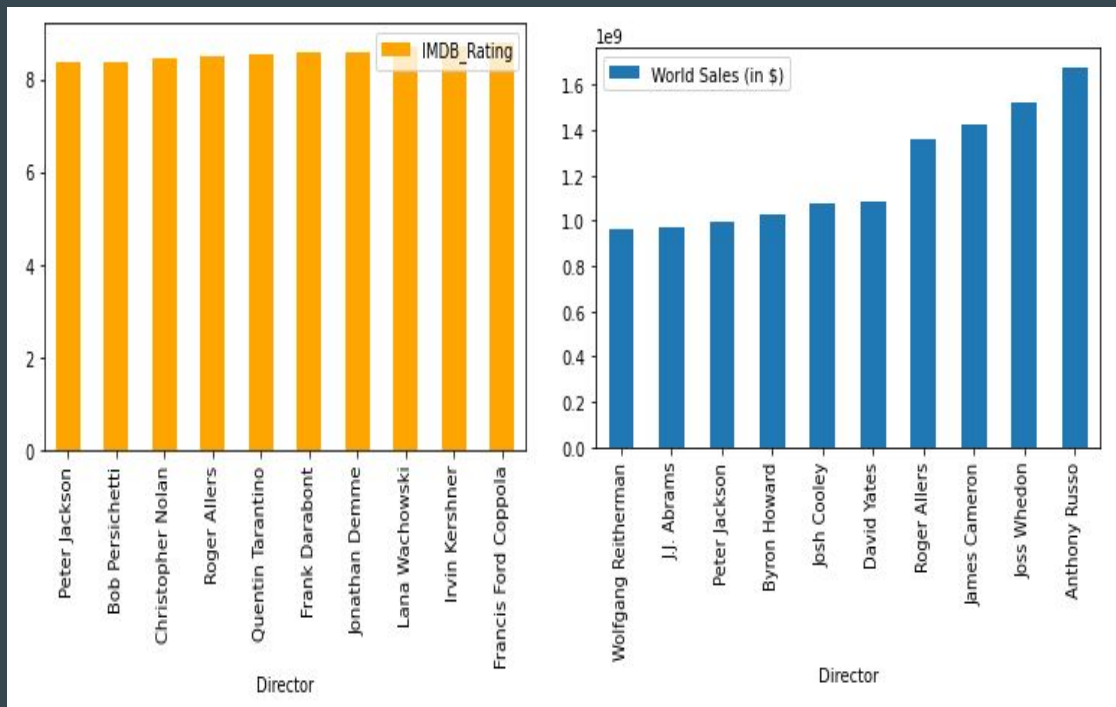
Just because a genre is popular, it doesn't determine that it's successful by our definition. This is why I grouped the genres together and found the average for both sales and rating.





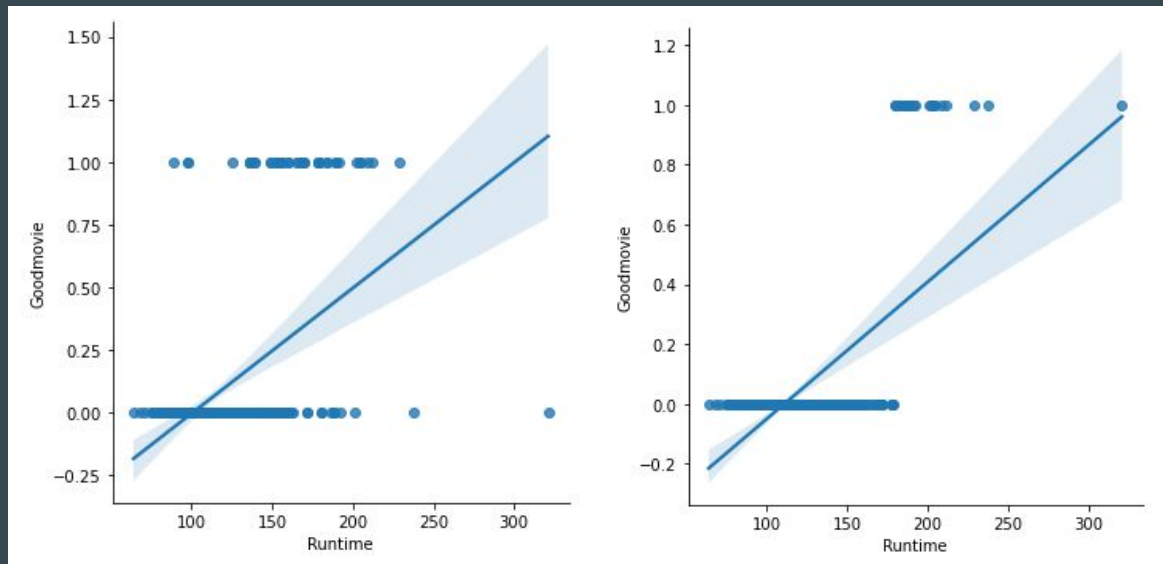
# Data Grouping 2

Doing the same grouping for the directors were also necessary to gather a further understanding of what makes a successful movie.



# The Runtime Factor

I used the predictive portion of my analysis to determine what a successful movie. If  $y=0$  the movie was unsuccessful and if  $y=1$  the movie was determined to be a success. The average successful movie was determined to be almost 3 hours in length!



# Drawing Conclusions

After my analysis I can determine that the best possible course of action for Good Movie Co. would be to develop a lengthy, adventurous, movie that doesn't shy away from more serious themes. Directors that could be hired to lead this process would be someone like Roger Allers or Peter Jackson. Creating an animated versus live action movie is negligible when it comes to the determination of a successful movie as it largely gains overall sales with a reduction in score. The trends of the movie industry should be closely watched afterwards in order to keep Good Movie Co. afloat, like any other artistic industry.