

# CSC343 Project - Investigation Process and Result

## 1. Factor 1: Experience of the Producing Team

For the first question, we tried to find out how the experience of the producing team affects the movie rating. To make this question more concrete and more approachable, we shift our attention to directors instead of the whole movie-making team, then we assume that a director has more experience of making a high-rating movie if the director has won more movie awards.

In our dataset, we found the average rating of movies that are directed by award-nominated directors (who has been nominated for an award), and we found the average rating of movies that are directed by award-winning directors (who has won any award); then, we compare these two results with the average rating of all movies in our dataset. Our result (Figure 1) shows a positive correlation between the experience of the director with average movie rating.

category	average_movie_rating
All Movies	7.0368347826087
With Nominated Director	7.08041958041958
With Awarded Director	7.10322425409048
(3 rows)	

Figure 1: average movie rating from directors with different award related status

However, the increment in the average rating is not significant. After reviewing our dataset, we find two details that may cause this insignificant difference. One is that our dataset only keeps the movies that have over 600,000 voting on IMDb. This ensures that our rating for the movie can represent most viewers' opinion; however, since popularity are likely to have a positive relationship with rating, the average rating of the movies in our dataset is high. Another detail is that our dataset for award records includes a large variety of awards. Those awards are not equally important in the movie industry, and it is highly possible for a director to be nominated or to win any award.

To solve the first problem, we can include more movies in our dataset that do not have over 600,000 votes. However, which movies should be kept is a decision we made when we were cleaning our dataset; this threshold allows us to find representative data and to keep our dataset relatively small (2876 movies) and manageable. Therefore, we will focus on reducing the effects brought by the award record dataset. We came up with a follow-up question to find out the average movie rating of movies directed by directors that have won over 30 awards. This does

show a larger difference in the average rating of movies that are directed by directors with experienced directors and inexperienced directors (Figure 2).

category	average_movie_rating
All Movies	7.0368347826087
With Nominated Director	7.08041958041958
With Awarded Director	7.10322425409048
With Excellent Awarded Director(won any awards for more than 30 times) (4 rows)	7.54368029739776

Figure 2: result for average movie rating categorized by more detailed award related status

## 2. Factor 2: Financial Support

For the second question, we were investigating how financial factors linked to movie ratings.

Our expectation before digging into the query process was that rating shall be in direct proportion to the budget and revenue. In the process, we divide the categories into four: movies with low budget, low revenue, and high rating (figure 3); movies with low budget, low revenue and low rating (figure 4); movies with high budget, high revenue, and high rating (figure 5); movies with high budget, high revenue, and low rating (figure 6). We set the definition of low budget, low revenue to be budget < 10000000 and USA Revenue < 10000000 and Global revenue < 20000000 (in US dollars), and definition of high budget, high revenue to be budget > 90000000 and USA Revenue > 150000000 and revenue > 300000000, based on the proportion of the number of movies being filtered out and the movies that pass the financial threshold (about 3% of movies passed the financial threshold of low budget and revenue, and same as high budget and revenue). The definition of high rating is larger than 8.0 and the definition of low rating is less than 6.5, also based on the proportion of the number of movies that pass the rating threshold in our database.

mid	low_budget_low_revenue_high_rating_movies	myear	avg_rating
tt0017136	Metropolis	1927	8.1
tt0021749	City Lights	1931	8.3
tt0027977	Modern Times	1936	8.4
tt0033467	Citizen Kane	1941	8.1
tt0034583	Casablanca	1942	8.4
tt0040522	Ladri di biciclette	1948	8.1
tt0042192	All About Eve	1950	8.1
tt0042876	Rashômon	1950	8.1
tt0043014	Sunset Blvd.	1950	8.2
tt0045152	Singin' in the Rain	1952	8.2
tt0046912	Dial M for Murder	1954	8.1
tt0047478	Shichinin no samurai	1954	8.4
tt0052357	Vertigo	1958	8.2
tt0056592	To Kill a Mockingbird	1962	8.2
tt0057012	Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb	1964	8.3
tt0071853	Monty Python and the Holy Grail	1975	8.2
tt0079944	Stalker	1979	8.1
tt0087544	Kaze no tani no Naushika	1984	8.1
tt0088258	This Is Spinal Tap	1984	8.1
tt0094625	Akira	1988	8.1
tt0095327	Hotaru no haka	1988	8.4
tt0105236	Reservoir Dogs	1992	8.2
tt0112471	Before Sunrise	1995	8.1
tt0113247	La haine	1995	8.1
tt0113568	Kôkaku Kidôtai	1995	8.1
tt0118849	Bachha-Ye aseman	1997	8.2
tt0120735	Lock, Stock and Two Smoking Barrels	1998	8.2
tt0169102	Lagaan: Once Upon a Time in India	2001	8.2
tt0180093	Requiem for a Dream	2000	8.2
tt0246578	Donnie Darko	2001	8.1
tt0353969	Salinui chueok	2003	8.1
tt0364569	Oldeuboi	2003	8.3
tt0861739	Tropa de Elite	2007	8.2
tt1255953	Incendies	2010	8.2
tt1562872	Zindagi Na Milegi Dobara	2011	8.3
tt2082197	Barfi!	2012	8.1
tt2106476	Jagten	2012	8.2
tt2265171	Serbuan maut 2: Berandal	2014	8.1

(38 rows)

Figure 3: movies with low budget, low revenue, and high rating

mid	low_budget_low_revenue_low_rating_movies	myear	avg_rating
tt0368226	The Room	2003	4.4
tt0873886	Red State	2011	6.3
tt1034314	Iron Sky	2012	6.1
tt1242432	I Spit on Your Grave	2010	6.4
tt1278340	DV[[d snV[[	2009	6.4
tt1467304	The Human Centipede (First Sequence)	2009	4.5
tt1470827	Monsters	2010	6.4
tt1602613	Only God Forgives	2013	5.8
tt1974419	The Neon Demon	2016	6.2
tt2910814	The Signal	2014	6.1
tt3605418	Knock Knock	2015	4.8
(11 rows)			

Figure 4: movies with low budget, low revenue and low rating

mid	high_budget_high_revenue_high_rating_movies	myear	avg_rating
tt0103064	Terminator 2: Judgment Day	1991	8.4
tt0120737	The Lord of the Rings: The Fellowship of the Ring	2001	8.7
tt0167260	The Lord of the Rings: The Return of the King	2003	8.8
tt0167261	The Lord of the Rings: The Two Towers	2002	8.6
tt0172495	Gladiator	2000	8.5
tt0198781	Monsters, Inc.	2001	8.1
tt0266543	Finding Nemo	2003	8.1
tt0325980	Pirates of the Caribbean: The Curse of the Black Pearl	2003	8.1
tt0372784	Batman Begins	2005	8.2
tt0435761	Toy Story 3	2010	8.2
tt0440963	The Bourne Ultimatum	2007	8.1
tt0468569	The Dark Knight	2008	8.9
tt0816692	Interstellar	2014	8.6
tt0848228	The Avengers	2012	8.1
tt0892769	How to Train Your Dragon	2010	8.2
tt0910970	WALL-E	2008	8.3
tt1049413	Up	2009	8.2
tt1201607	Harry Potter and the Deathly Hallows: Part 2	2011	8.2
tt1345836	The Dark Knight Rises	2012	8.5
tt1375666	Inception	2010	8.7
tt1853728	Django Unchained	2012	8.4
tt1877832	X-Men: Days of Future Past	2014	8.1
tt2015381	Guardians of the Galaxy	2014	8.1
tt2096673	Inside Out	2015	8.1
tt2380307	Coco	2017	8.4
tt2948356	Zootopia	2016	8.1
tt3315342	Logan	2017	8.2
tt4154756	Avengers: Infinity War	2018	8.5
tt4154796	Avengers: Endgame	2019	8.4
tt8579674	1917	2019	8.3
(30 rows)			

Figure 5: movies with high budget, high revenue, and high rating

mid	high_budget_high_revenue_low_rating_movies	myear	avg_rating
tt0112462	Batman Forever	1995	5.5
tt0120667	Fantastic Four	2005	5.8
tt0120755	Mission: Impossible II	2000	6.1
tt0120912	Men in Black II	2002	6.2
tt0133152	Planet of the Apes	2001	5.7
tt0163025	Jurassic Park III	2001	6
tt0170016	How the Grinch Stole Christmas	2000	6.4
tt0177971	The Perfect Storm	2000	6.4
tt0181852	Terminator 3: Rise of the Machines	2003	6.4
tt0213149	Pearl Harbor	2001	6.3
tt0246460	Die Another Day	2002	6.1
tt0348150	Superman Returns	2006	6.2
tt0367882	Indiana Jones and the Kingdom of the Crystal Skull	2008	6.2
tt0413267	Shrek the Third	2007	6.1
tt0413300	Spider-Man 3	2007	6.3
tt0800320	Clash of the Titans	2010	5.9
tt0892791	Shrek Forever After	2010	6.4
tt1046173	G.I. Joe: The Rise of Cobra	2009	5.9
tt1055369	Transformers: Revenge of the Fallen	2009	6.2
tt1078912	Night at the Museum: Battle of the Smithsonian	2009	6.1
tt1190080	2012	2009	5.9
tt1216475	Cars 2	2011	6.3
tt1291150	Teenage Mutant Ninja Turtles	2014	6
tt1324999	The Twilight Saga: Breaking Dawn - Part 1	2011	5.5
tt1386697	Suicide Squad	2016	6.2
tt1623205	Oz the Great and Powerful	2013	6.4
tt1673434	The Twilight Saga: Breaking Dawn - Part 2	2012	6.1
tt1735898	Snow White and the Huntsman	2012	6.3
tt2109248	Transformers: Age of Extinction	2014	5.9
tt2126355	San Andreas	2015	6.2
tt4881806	Jurassic World: Fallen Kingdom	2018	6.3
(31 rows)			

Figure 6: movies with high budget, high revenue, and low rating

category	count
high_budget_high_revenue_high_rating	30
high_budget_high_revenue_low_rating	31
low_budget_low_revenue_high_rating	38
low_budget_low_revenue_low_rating	11

(4 rows)

Figure 7: combined result

Surprisingly, in our combined result (Figure 7), movies with low budget, low revenue, and high rating have a lot more counts than movies with low budget, low revenue and low rating; while counts of movies with high budget, high revenue, and high rating and movies with high budget, high revenue, and low rating are very similar. Movies with low budget and low revenue seem to have a higher chance of gaining a high rating. We analyzed the reason to be that the release year of many of them were very early (Figure 3). The amount of budget and revenue of these movies might seem low nowadays due to inflation, but the money was valued more in the old times. So in follow-up, we decided for movies with low budget and low revenue, selecting only those whose release year was later than 2000. Below is the new counting result for follow-up, and we concluded that financial factors are not directly linked to movie rating for our database.

category	count
high_budget_high_revenue_high_rating	30
high_budget_high_revenue_low_rating	31
low_budget_low_revenue_high_rating_after2000	10
low_budget_low_revenue_low_rating_after2000	11

(4 rows)

Figure 8: combined result after adding releasing year condition

### 3. Factor 3: Movie Genre

For the third question, we were investigating how movie genres affect movie rating. For each genre, we selected all the movies whose attribute contains the keyword. Then for each genre's movie lists, we joined it with the rating data for each age group, and then combined them all. We excluded 'War' and 'Western' genres because their movie counts are both less than 50. A movie can be included in several genres, so the rating of the movie is counted into several genres' average movie rating. Below is the resulting table, rows are ordered by all\_age\_rating attribute, which is the average rating for all ages.

category	count	all_age_rating	under18	from18to30	from30to45	above45
Biography	197	7.48883248730965	7.44517766497462	7.51928934010152	7.35583756345178	7.30913705583756
History	81	7.48148148148148	7.41975308641975	7.53333333333333	7.35061728395062	7.3641975308642
Music	83	7.3385421686747	7.30602409638554	7.32168674698795	7.1144578313253	7.01927710843374
Animation	179	7.33631284916201	7.16927374301676	7.29888268156425	7.16983240223464	7.09608938547486
Drama	1396	7.3262893982808	7.25902578796562	7.35730659025788	7.19018624641834	7.1086676217765
Sport	54	7.21851851851852	7.09814814814815	7.19259259259259	7.03333333333333	6.87037037037037
Crime	546	7.13608058608058	7.05091575091576	7.16813186813186	7.01080586080587	6.91831501831502
Romance	382	7.0369109947644	6.92670157068063	7.00549738219896	6.82722513089005	6.77670157068062
Family	165	7.00909090909091	6.85151515151515	6.92424242424242	6.79878787878788	6.79696969696969
Mystery	289	7.00899653979239	6.89307958477508	7.03425605536332	6.88996539792388	6.82837370242214
Thriller	545	6.94	6.84275229357799	6.94165137614679	6.81064220183487	6.78018348623853
Adventure	757	6.91809775429326	6.78665785997358	6.88071334214002	6.74914134742404	6.74464993394981
Comedy	960	6.89291666666667	6.8090625	6.86020833333334	6.7040625	6.6221875
Fantasy	301	6.8302325813954	6.65647840531561	6.76744186046512	6.64352159468439	6.65681063122924
Sci-Fi	322	6.80621118012423	6.65745341614906	6.78944099378881	6.68012422360249	6.66645962732919
Action	965	6.80497409326425	6.67440414507772	6.78165803108809	6.63564766839378	6.62113989637308
Horror	293	6.62662116040956	6.48805460750853	6.58293515358362	6.48941979522184	6.47064846416382

Figure 9: combined result of different age groups' ratings towards genres

We can get that biography and history are the top two genres that are rated the highest according to average rating for all ages, while action and horror are rated lowest. In our follow-up, we found movie genre that each age group rated highest and rated lowest:

agerange	rated_highest	h_rating	rated_lowest	l_rating
above45	History	7.3641975308642	Horror	6.47064846416382
from18to30	History	7.53333333333333	Horror	6.58293515358362
from30to45	Biography	7.35583756345178	Horror	6.48941979522184
under18	Biography	7.44517766497462	Horror	6.48805460750853

Figure 10: combined result of movie genre that each age group rated highest and rated lowest

We can get that biography and history are still two genres that are rated the highest over different age groups, and horror is the lowest over all age groups.

In conclusion, for the study domain of movie rating based on our dataset, we conclude that there is a positive relation between the experience of the director and average movie rating; Financial factors (budget and revenue) are not directly linked to average movie rating, and movie genre can affect movie rating.