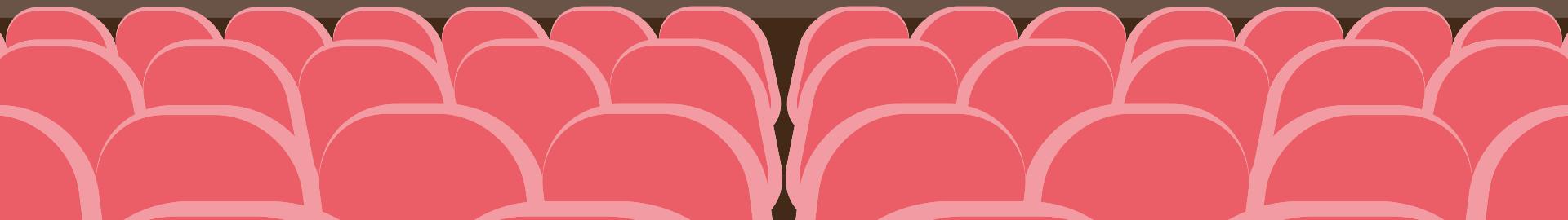


MOVIES & REVENUE

A CASE STUDY & REGRESSION OF MOVIE PROFITABILITY



By performing a regression, I found that movie companies should focus on their hiring decisions in order to maximize revenue. My regression also concluded in an equation to predict revenue based on a variety of factors with an r^2 value of ~0.25. I identified trends in genre, runtime, release dates, and revenue through our case study.

!

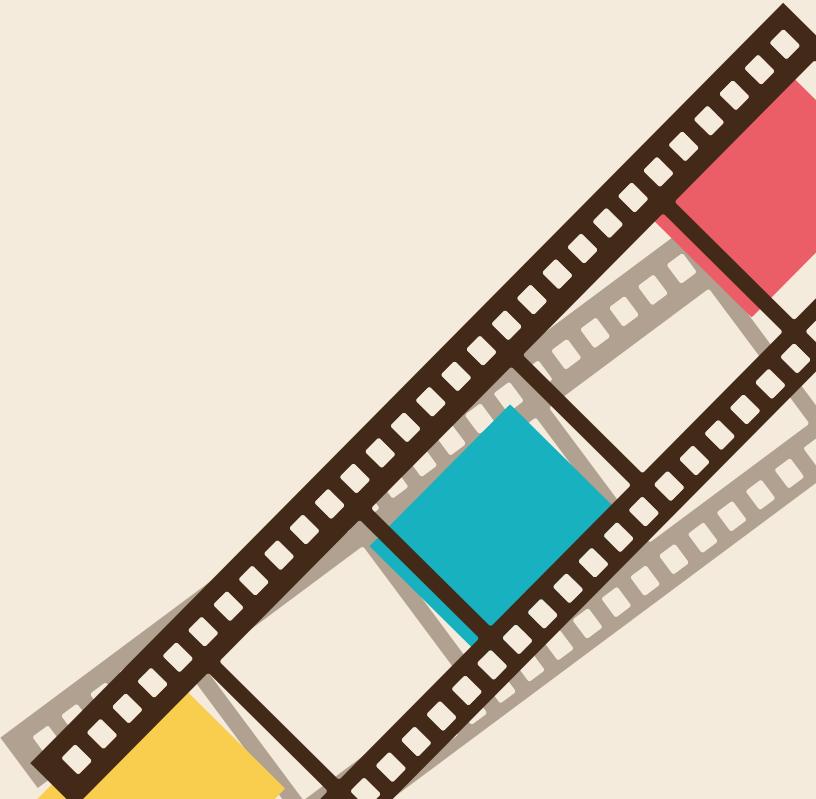
PART 1: CASE STUDY

- Which genre maximizes revenue?
- Does release date affect revenue?
- What types of movies perform better?
- Is there a relationship between runtime and revenue?

!

PART 2: REGRESSION

- Do certain factors have a stronger correlation with revenue?
- Can we create an overall predictive equation for revenue? If so, how accurate is it?
- How strong is the correlation between revenue and all variables?



THE DATASET

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1	id	imdb_id	popularity	budget	revenue	original_title	cast	homepage	director	tagline	keywords	overview	runtime	genres	production_c	release_date	vote_count	vote_average	release_year	budget_adj	revenue_adj	
2	135397	tt0369610	32.985763	150000000	1513528810	Jurassic Wor	Chris Pratt B	http://www. Colin Trevorr	The park is o	monster dnz	Twenty-two		124	Action Aeve	Universal Stu	6/9/15	5562	6.5	2015	137999939	1392445893	
3	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fu	Tom Hardy C	http://www. George Mille	What a Love	future chase	An apocalyp		120	Action Aeve	Village Road	5/13/15	6185	7.1	2015	137999939	348161293	
4	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Wo	http://www. Robert Schw	One Choice	b based on nov	Beatrice Pric		119	Adventure S	Summit Ente	3/18/15	2480	6.3	2015	101199956	271619025	
5	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: T	Harrison For	http://www. J.J. Abrams	Every genera	android spa	Thirty years		136	Action Aeve	Lucasfilm Tr	12/15/15	5292	7.5	2015	183999919	1902723130	
6	168259	tt2820829	9.335014	190000000	1506249360	Furious 7	Vin Diesel P	http://www. James Wan	Vengeance	car race spe	Deckard Sha		137	Action Crim	Universal Pic	4/1/15	2947	7.3	2015	174799923	1385748801	
7	281957	tt1663202	9.1107	135000000	532950503	The Revenan	Leonardo DiC	http://www. Alejandro Gc	(n.	One who father-son	r e In the 1820s,		156	Western Drz	Regency Ente	12/25/15	3929	7.2	2015	124199945	490314247	
8	87101	tt1340138	8.654359	155000000	440603537	Terminator	C Arnold Schw	http://www. Alan Taylor	Reset the fu	saving the w	The year is 2		125	Science Ficti	Paramount P	6/23/15	2598	5.8	2015	142599937	405355076	
9	286217	tt3659388	7.6674	108000000	595380321	The Martian	Matt Damon	http://www. Ridley Scott	Bring Him	Hc based on nov	During a mai		141	Drama Adv	Twentieth Ce	9/30/15	4572	7.6	2015	99359956.3	547749654	
10	211672	tt2293640	7.404165	74000000	1156730962	Minions	Sandra Bullo	http://www. Kyle Balda P	Before Gru,	assistant aff	Minions Stua		91	Family Anir	Universal Pic	6/17/15	2893	6.5	2015	68079970	1064192017	
11	150540	tt2096673	6.326804	175000000	853708609	Inside Out	Amy Poehler	http://movie	Pete Docter	Meet the litt	dream carto	Growing up		94	Comedy Ani	Walt Disney	6/9/15	3935	8	2015	160999929	785411575
12	206647	tt2379713	6.200282	245000000	880674609	Spectre	Daniel Craig http://www. Sam Mendes	A Plan No Or	spy based on A	cryptic me			148	Action Adv	Columbia Pic	10/26/15	3254	6.2	2015	225399901	810220284	
13	76757	tt1617661	6.189369	176000003	183987723	Jupiter Ascer	Mila Kunis C	http://www. Lana Wachow	Expand your	jupiter spac	In a universe		124	Science Ficti	Village Road	2/4/15	1937	5.2	2015	161919932	169268631	
14	246460	tt0470752	6.118847	15000000	36869414	Ex Machina	Dominhall Gl	http://exmra	Alex Garland	There is not	dancing arti	Caleb, a 26 y		108	Drama Scier	DNA Films L	1/21/15	2854	7.6	2015	13799993.9	33919846
15	257344	tt2120120	5.984995	88000000	243637091	Pixels	Adam Sandler	http://www. Chris Columb	Game On.	video game Video game			105	Action Com	Columbia Pic	7/16/15	1575	5.8	2015	80959964.4	224146025	
16	99861	tt2395427	5.944927	280000000	1405035767	Avengers: Ag	Robert Dow	http://marve	Joss Whedor	A New Age	marvel comi	When Tony S		141	Action Adv	Marvel Studi	4/22/15	4304	7.4	2015	257599887	1292632337
17	273248	tt3460252	5.8984	44000000	155760117	The Hateful	I Samuel L. Ja	http://theha	Quentin Tara	No one come	bounty hunte	Bounty hunte		167	Crime Dram	Double Featu	12/25/15	2389	7.4	2015	40479982.2	143299245
18	260346	tt2446042	5.749758	48000000	325771424	Taken 3	Liam Neesor	http://www. Olivier Megalit	Ends Here	revenge mu	Ex-governme		109	Crime Actio	Twentieth Ce	1/1/15	1578	6.1	2015	44159980.6	299709578	
19	102899	tt0478970	5.573184	130000000	518602163	Ant-Man	Paul Rudd Iv	http://marve	Peyton Reed	Heroes Don't	marvel comi	Armed with		115	Science Ficti	Marvel Studi	7/14/15	3779	7	2015	119599947	477113780
20	150689	tt1661199	5.556818	95000000	542351353	Cinderella	Lily James Cat	Blanchett	Kenneth Brai	Midnight is j	cinderella m	When her fa		112	Romance Fa	Walt Disney	3/12/15	1495	6.8	2015	87399961.5	498963025
21	131634	tt1951266	5.476958	160000000	650523427	The Hunger	Jennifer Law	http://www. Francis Lawr	The fire will	revolution st	With the nat		136	War Advent	Studio Babel	11/18/15	2380	6.5	2015	147199953	598481290	
22	158852	tt1964418	5.462138	190000000	209035668	Tomorrow	lai Britt Roberts	http://movie	Brad Bird	Imagine a w	inventor apc	Bound by a s		130	Action Fami	Walt Disney	5/19/15	1899	6.2	2015	174799923	192312730
23	307081	tt1798684	5.337064	30000000	91709827	Southpaw	Jake Gyllenhaal Rachel M	Antoine Fuq	Believe in Hc	sport	Billy "The Gr		123	Action Dram	Escape Artist	6/15/15	1386	7.3	2015	27599987.9	84373003.7	
24	254128	tt2126355	4.907832	110000000	470490832	San Andreas	Dwayne Johr	http://www. Brad Peyton	A rescue pilo	california ea	In the afterr		114	Action Dram	New Line Cir	5/27/15	2060	6.1	2015	101199956	432851375	
25	216015	tt2324441	4.710402	40000000	569651467	Fifty Shades	Dakota Johns	https://www Sam	Taylor-J	Are you curic	based on nov	When colleg		125	Drama Rom	Focus Featur	2/11/15	1865	5.3	2015	36799983.8	524079119
26	318846	tt1596363	4.648046	28000000	133346506	The Big Shor	Christian Bal	http://www. Adam McKa	This is a true	bank fraud	Th men wh		130	Comedy Dra	Paramount P	12/11/15	1545	7.3	2015	25759988.7	122678732	
27	177677	tt2381249	4.566713	150000000	682330139	Mission: Imp	Tom Cruise .	http://www. Christopher	I Desperate	Ti spy sequel i	Ethan and te		131	Action	Paramount P	7/23/15	2349	7.1	2015	137999939	627743452	
28	214756	tt2637276	4.564549	68000000	215863560	Ted 2	Mark Wahlberg Seth MacFar	Ted is Comin	sperm bank	Newlywed c		115	Comedy	Universal Pic	6/25/15	1666	6.3	2015	62559972.5	198594430		
29	207703	tt2802144	4.503789	81000000	403802136	Kingsman: TI	Taron Egert	http://www. Matthev	Wa Manners	ma spy great br	The story of		130	Crime Come	Twentieth Ce	1/24/15	3833	7.6	2015	74519967.2	371497802	
30	314365	tt1895587	4.062293	20000000	88346473	Spotlight	Mark Ruffalc	http://www. Tom McCart	Break the sti	child abuse	The true stor		128	Drama Thril	Participant N	11/6/15	1559	7.8	2015	18399991.9	81278719.4	
31	294254	tt0406784	3.968891	61000000	311256926	Maze Runne	Dylan O'Brie	http://mazi	Wes Ball	The Maze W	based on nov	Thomas and		132	Action Scien	Gotham Gro	9/9/15	1849	6.4	2015	56119975.3	286356246
32	280996	tt3168230	3.927333	0	29355203	Mr. Holmes	Ian McKellen	http://www. Bill Condon	The man bef	london dete	The story is s		103	Mystery Dra	BBC Films S	6/19/15	425	6.4	2015	0	27006774.9	
33	198184	tt1823672	3.899557	49000000	102069268	Chappie	Sharlto Cople	http://www. Neill Blomka	I am consci	artificial inte	Every child c		120	Crime Actio	Columbia Pic	3/4/15	1990	6.6	2015	45079980.2	93903685.2	

tmdb_movies_data

+

-- WORKSHEET --

VARIABLES

I selected 5 variables and observed their relationships with the adjusted revenue.

Genre

Category based on similarities in the narrative element or emotional responses



Runtime

Length of the movie (minutes)



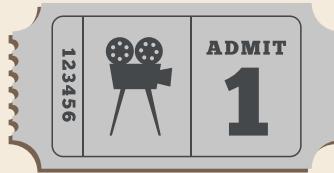
Production Company

Act as the intersection between the movie itself and the film industry



Cast & Director

The stars who act in the movies, and the ones who supervise



Release Date

The time period when the movie is released



Adjusted Revenue

Revenue adjusted for inflation using the base year of 2010

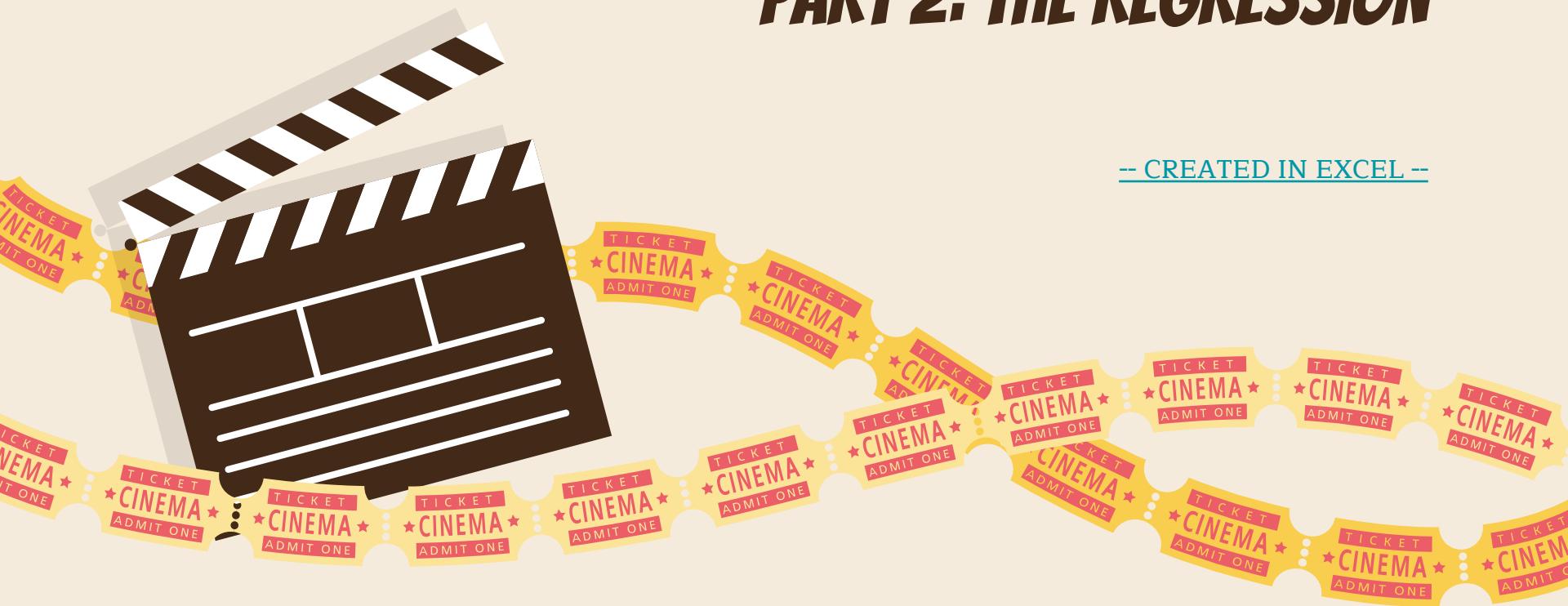
-- VISUALIZATIONS CREATED IN TABLEAU --

PART 1: THE CASE STUDY



PART 2: THE REGRESSION

-- CREATED IN EXCEL --



DATA PREPARATION

By tagging the data and using various Excel functions, I cleaned the dataset and prepared it for regression.

A	B	C	D	E	F	G
Director Names	Director Tags	Series				
Clyde Geronimi	1					
George Lucas	2					
Lee Unkrich	3					
David Yates	4					
Bob Peterson	5					
Pierre Coffin	6					
Robert Stromberg	7					
Colin Trevorrow	8					
Dan Scanlon	9					
Chris Miller	10					
J.J. Abrams	11					
James Cameron	12					
Joe Russo	13					
Jennifer Yuh	14					
Andrew Stanton	15					
Don Hall	16					
Shane Black	17					
George Seaton	18					
Joss Whedon	19					
Peter Jackson	20					
Eric Darnell	21					
Lilly Wachowski	22					

O2 ▾ X ✓ fx =VLOOKUP(L2,M:N,2,FALSE)

L	M	N	O
1 director	Director Names	Director Tag: Director	
2 Colin Trevorrow	Clyde Geronimi	1	8
3 George Miller	George Lucas	2	179
4 Robert Schwentke	Lee Unkrich	3	216
5 J.J. Abrams	David Yates	4	11
6 James Wan	Bob Peterson	5	114
7 Alejandro González Iñárritu	Pierre Coffin	6	#N/A
8 Alan Taylor	Robert Stromberg	7	37
9 Ridley Scott	Colin Trevorrow	8	167
10 Kyle Balda	Dan Scanlon	9	136
11 Pete Docter	Chris Miller	10	42
12 Sam Mendes	J.J. Abrams	11	43
13 Lana Wachowski	James Cameron	12	396
14 Alex Garland	Joe Russo	13	810
15 Chris Columbus	Jennifer Yuh	14	52
16 Joss Whedon	Andrew Stanton	15	19
17 Quentin Tarantino	Don Hall	16	212
18 Olivier Megaton	Shane Black	17	142
19 Peyton Reed	George Seaton	18	173

1. Tag qualitative data

2. Match with original data
using =VLOOKUP

3. Create a worksheet with all
tagged data

A	B	C	D	E	
1	Title	ID	Genre 1	Genre 2	Genre 3
2 Jurassic World	135397	1	2	7	
3 Mad Max: Fury Road	76341	1	2	7	
4 Insurgent	262500	2	7	5	
5 Star Wars: The Force Awakens	140607	1	2	7	
6 Furious 7	168259	1	10	5	
7 The Revenant	281957	17	3	2	
8 Terminator Genisys	87101	7	1	5	
9 The Martian	286217	3	2	7	
10 Minions	211672	6	11	2	
11 Inside Out	150540	4	11	6	
12 Spectre	206647	1	2	10	
13 Jupiter Ascending	76757	7	8	1	
14 Ex Machina	264660	3	7	0	
15 Pixels	257344	1	4	7	
16 Avengers: Age of Ultron	99861	1	2	7	
17 The Hateful Eight	273248	10	3	12	
18 Taken 3	260346	10	1	5	

INDIVIDUAL REGRESSION

Through the regressions I performed, I found that the cast and director for a movie had the strongest correlations with revenue.

Cast: R²= 0.208533852719812, p-value

Director: $R^2 = 0.121383564498438$

Production Co: R²= 0.0767575668226115

Runtime: $R^2 = 0.0308621275345542$

Genre: R²= 0.013040942566452

Release Date: R²= 0.00256862976872277

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	Regression Statistics								
4	Multiple R	0.45665507							
5	R Square	0.20853385							
6	Adjusted R S	0.20816946							
7	Standard Err	128701015							
8	Observations	10866							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	5	4.7396E+19	9.4791E+18	572.274038	0			
13	Residual	10860	1.7988E+20	1.6564E+16					
14	Total	10865	2.2728E+20						
15									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	159387520	2368044.97	67.3076404	0	154745719	164029320	154745719	164029320
18	Cast Member	-11537.682	1245.1853	-9.2658352	2.3005E-20	-13978.472	-9096.8914	-13978.472	-9096.8914
19	Cast Member	-8499.78	989.275074	-8.5919278	9.7225E-18	-10438.94	-6560.6204	-10438.94	-6560.6204
20	Cast Member	-8377.9833	881.781241	-9.5012038	2.5106E-21	-10106.435	-6649.5312	-10106.435	-6649.5312
21	Cast Member	-9285.1835	790.977579	-11.738871	1.2505E-31	-10835.644	-7734.7232	-10835.644	-7734.7232
22	Cast Member	-10198.527	706.535782	-14.434552	8.4947E-47	-11583.466	-8813.5881	-11583.466	-8813.5881

OVERALL REGRESSION

The overall regression equation was reasonably accurate, with an r^2 value of 0.25, in capturing the relationship between the variables we chose and the revenue.



SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.491894							
R Square	0.2419597							
Adjusted R Square	0.2408418							
Standard Error	126017830							
Observations	10866							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	16	5.499E+19	3.437E+18	216.43155	0			
Residual	10849	1.723E+20	1.588E+16					
Total	10865	2.273E+20						
Coefficients Standard Error t Stat P-value Lower 95% Upper 95% Lower 95.0% Upper 95.0%								
Intercept	113278685	5869463.5	19.2996659	1.252E-81	101773464	124783905	101773464	124783905
Genre 1	-365012.5	263032.61	-1.387708	0.1652545	-880604.4	150579.5	-880604.4	150579.5
Genre 2	54457.388	263549.76	0.20663039	0.8363024	-462148.3	571063.06	-462148.3	571063.06
Genre 3	828981.76	281531.44	2.94454422	0.0032412	277128.72	1380834.8	277128.72	1380834.8
Genre 4	908749.87	367086.47	2.4755744	0.0133174	189193.33	1628306.4	189193.33	1628306.4
Release Month	809720.66	352780.25	2.29525504	0.0217375	118206.92	1501234.4	118206.92	1501234.4
Runtime	438724.77	39680.353	11.0564736	2.887E-28	360944.03	516505.51	360944.03	516505.51
Production Company 1	-2020.876	1677.6103	-1.204616	0.2283778	-5309.299	1267.5464	-5309.299	1267.5464
Production Company 2	-6950.763	1626.107	-4.274481	1.932E-05	-10138.23	-3763.297	-10138.23	-3763.297
Production Company 3	-7852.795	2456.718	-3.1964575	0.0013953	-12668.41	-3037.179	-12668.41	-3037.179
Production Company 4	-19669.51	4596.3778	-4.2793497	1.89E-05	-28679.25	-10659.77	-28679.25	-10659.77
Director	-11806.25	1086.4884	-10.86643	2.304E-27	-13935.97	-9676.534	-13935.97	-9676.534
Cast Member 1	-6697.149	1270.1605	-5.2726798	1.37E-07	-9186.896	-4207.403	-9186.896	-4207.403
Cast Member 2	-5990.488	983.48257	-6.0910976	1.159E-09	-7918.294	-4062.683	-7918.294	-4062.683
Cast Member 3	-6908.945	869.59148	-7.945047	2.133E-15	-8613.503	-5204.387	-8613.503	-5204.387
Cast Member 4	-7862.948	781.66541	-10.059225	1.063E-23	-9395.155	-6330.741	-9395.155	-6330.741
Cast Member 5	-8422.778	702.56419	-11.988624	6.577E-33	-9799.932	-7045.624	-9799.932	-7045.624

$$\text{Revenue} = 113278685$$

- 365012.46(Genre 1)
- + 54457.39(Genre 2)
- + 828981.76(Genre 3)
- + 908749.87(Genre 4)
- + 809720.66(Release Month)
- + 438724.771(Runtime)
- 2020.8763(Production Company 1)
- 6950.76(Production Company 2)
- 7852.79(Production Company 3)
- 19669.51(Production Company 4)
 - 11806.25(Director)
 - 6697.15(Cast Member 1)
 - 5590.49(Cast Member 2)
 - 6908.95(Cast Member 3)
 - 7862.95(Cast Member 4)
 - 8422.78(Cast Member 5)