

# Superhero Films Statistical Analysis

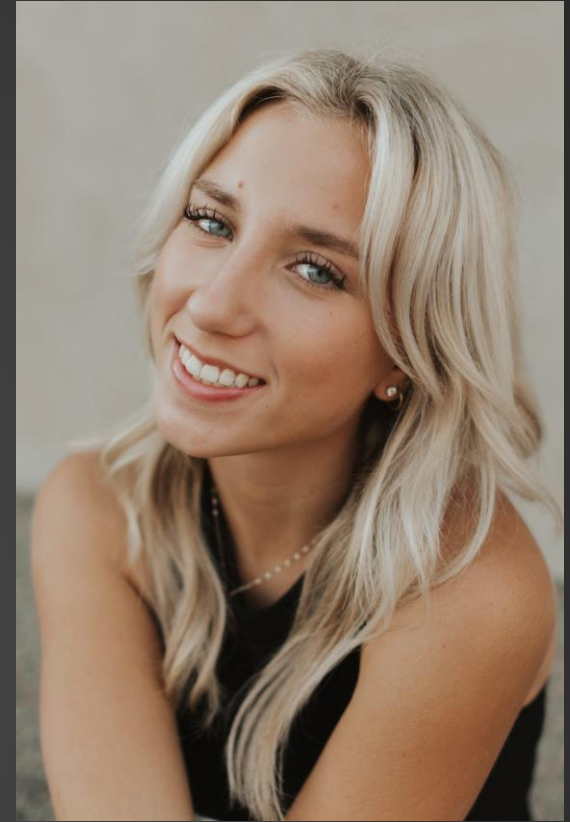




Ava Hatch



Raksha Kumar



Julia DiMartini

# Introduction

# Agenda

Executive  
Summary

Problem  
Statement

Data  
Analysis

Correlation  
Matrix

Residual  
Plots

Final  
Model

Conclusion

# Executive Summary

## Problem

What are the factors that affect the success of a superhero film? The analysis we have completed is helpful to show what key variables play into accomplished superhero films. This could help other filmmakers to have this deeper understanding of how to help their own superhero films succeed.

## Data source

Box office of DC and Marvel superhero movies dataset from Kaggle for our analysis.

## Methods

Multi-variable linear regression, multicollinearity, parsimonious model, pivot table and residual plots

## Measures

Variables, multicollinearity/correlation and the p-values of variables

## Conclusion

The most impactful variable was the Franchise/Distributor (Walt Disney), with a positive relationship. The least impactful variable was Budget, with a positive relationship. Additionally, Minutes, Rotten Tomato Rating, Budget, and Male Protagonist were all very significant variables in predicting the box office success of a film.

# Problem Statement – What are we trying to solve?

## English Problem Statement:

What contributes to the success of a superhero film, and what causes high box office sales?

## Big Analytical Question:

Which variables are significant contributors to the success of box office tickets sold for a superhero film?

## Further Questions:

What factor is the most impactful on box office sales for a superhero film?  
What factor is the least impactful on box office sales for a superhero film?

# Problem Statement – How we are going to solve it?

Dependent variable: Box Office Gross Worldwide

```
graph TD; A[Dependent variable: Box Office Gross Worldwide] --> B[Independent variables: Budget, Marvel Y/N, MPAA Rating, Minutes, Male or Female Protagonist, Year, Franchise/Distributor and Rotten Tomato Ratings]; A --> C[Strategy: Create a multi-variable linear regression model to estimate what factors heavily affect the success of box office ticket sales for a superhero film]; B --> C;
```

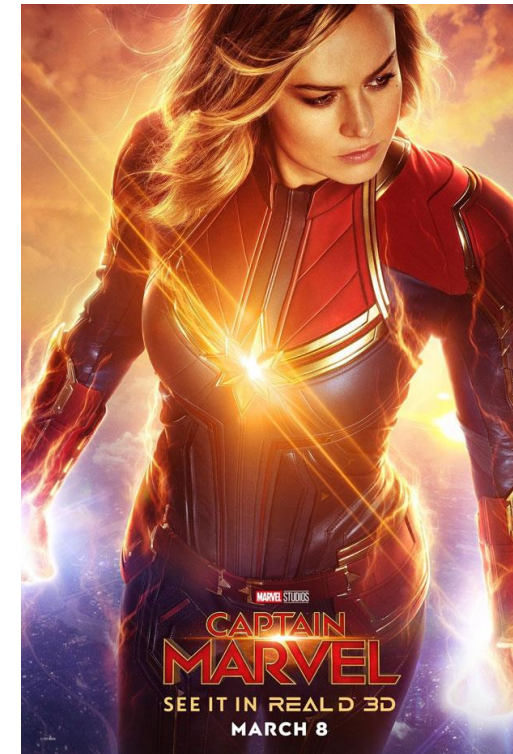
Independent variables: Budget, Marvel Y/N, MPAA Rating, Minutes, Male or Female Protagonist, Year, Franchise/Distributor and Rotten Tomato Ratings

Strategy: Create a multi-variable linear regression model to estimate what factors heavily affect the success of box office ticket sales for a superhero film



# Problem Statement – What are our assumptions?

- The superhero films used in the model were released in the U.S. between the years 1978-2024
- Box Office Gross Worldwide, Budget, MPAA Rating, Marvel Y/N, Minutes, Male or Female Protagonist, Year, Franchise/Distributor and Rotten Tomato Ratings are the variables we used for our analysis.
- We found our dataset on the Kaggle website
  - The data from Kaggle was collected primarily from Box Office Mojo, alongside additional sources such as: Deadline, Variety, The Numbers, Forbes and Wizard Magazine.
- For our alpha, we used a significance level of 0.05
- The Box Office Gross Worldwide result is counted under the following assumptions: the viewers must have watched the film in the theater and the count of box office gross started on the day the film was released.



[This Photo](#) by Unknown Author is licensed under [CC BY-NC-ND](#)

# Data

- **114 rows**
  - 113 Observations
  - 1 row for labels
- **Originally 22 Columns**
  - 22 variables describing Marvel and DC Movies
  - We utilized 9 of the 22 variables
  - 8 independent variables
  - 1 dependent variable



# Data Continued...

- Multi-variable Linear Regression

- **Dependent Variable**: Box Office Gross Worldwide
- **Independent Variables**: Budget, MPAA Rating, Marvel Y/N, Minutes, Male or Female Protagonist, Year, Franchise/Distributor and Rotten Tomato Ratings
- We chose these nine variables out of the dataset, because we believe that they are the most impactful in determining the success of a film's gross total amount of box office ticket sales.

- Hypothesis

- **Alpha** = 0.05
- **H<sub>0</sub>** = The coefficients of the independent variables equal zero meaning that they do not have an impact on the outcome of the box office gross worldwide (null hypothesis).
- **H<sub>a</sub>** = The coefficients of the independent variables do not equal zero meaning that they do have an impact on the outcome of the box office gross worldwide (alternative hypothesis).

# Definitions Provided of Variables

- **Box office gross Worldwide:** The sum total gross earnings for all territories in U.S. dollars.
- **Budget:** The production budget for the movie in U.S. dollars. This does not include any additional expenses relating to the movie beyond what it cost to make, most notably the marketing budget which can equal or even exceed the production budget on major motion pictures.
- **MPAA Rating:** The age rating awarded under the Motion Picture Association film rating system.
- **Marvel Y/N:** If the film was a Marvel movie or not.
- **Minutes:** The length of the film but presented only in minutes.
- **Male/Female-led:** This shows whether the film had a male or female lead.
- **Rotten Tomatoes Critic Score:** The average score from all professional critics on the “Tomatometer” from the popular review aggregator website.
- **Year:** The calendar year the film was released theatrically in the United States of America.
- **Franchise/Distributor:** The name of the film studio that distributed the movie within the United States.

# Methodology: Data Preparation

- Exported dataset from Kaggle into Excel
  - Adjusted columns to fit the right size and changed wording to increase readability
- Adjusted data
  - Chose most important eight variables out of list that will most accurately display why Marvel is one of the most popular movie franchises
  - Removed character family, length, gross to budget, domestic % and U.S release date, inflation adjusted worldwide gross, inflation adjusted budget, phase of production, distributor category converted to 5 categories

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2	Film	Budget	Warner Br	Walt Disne	Twentieth	Sony Pictu	PG dumm	PG-13 dur	Minutes	Marvel du	Rotten To	Male dur	Year	Success du	Box office	gross Worldw
3	Superman	\$ 65,000,000.00	1	0	0	0	1	0	143	0	94	1	1978	1	5	300,478,440.00
4	Superman	\$ 54,000,000.00	1	0	0	0	1	0	127	0	83	1	1981	1	5	236,385,706.00
5	Superman	\$ 39,000,000.00	1	0	0	0	1	0	125	0	29	1	1983	0	5	80,250,623.00
6	Supergirl	\$ 35,000,000.00	0	0	0	0	1	0	124	0	8	0	1984	0	5	14,296,438.00
7	Howard th	\$ 37,000,000.00	0	0	0	0	1	0	110	1	13	1	1986	0	5	37,962,774.00
8	Superman	\$ 17,000,000.00	1	0	0	0	1	0	90	0	10	1	1987	0	5	30,281,020.00
9	The Return	\$ 3,000,000.00	0	0	0	0	0	1	88	0	44	1	1989	0	5	274,928.00
10	Batman	\$ 35,000,000.00	1	0	0	0	0	1	126	0	76	1	1989	1	5	411,569,241.00
11	The Punish	\$ 9,000,000.00	0	0	0	0	0	0	89	1	25	1	1989	1	5	30,000,000.00
12	Batman Re	\$ 80,000,000.00	1	0	0	0	0	1	126	0	81	1	1992	1	5	266,914,688.00
13	Batman M	\$ 5,600,000.00	1	0	0	0	1	0	76	0	82	1	1993	0	5	5,846,204.00
14	Batman Fe	\$ 100,000,000.00	1	0	0	0	0	1	121	0	40	1	1995	1	5	336,567,158.00
15	Batman &	\$ 125,000,000.00	1	0	0	0	0	1	125	0	11	1	1997	0	5	238,253,988.00
16	Steel	\$ 16,000,000.00	1	0	0	0	0	1	97	0	12	1	1997	0	5	1,801,977.00
17	Blade	\$ 45,000,000.00	0	0	0	0	0	0	120	1	58	1	1998	1	5	131,211,377.00
18	X-Men	\$ 75,000,000.00	0	0	1	0	0	1	104	1	82	1	2000	1	5	296,339,528.00
19	Blade II	\$ 54,000,000.00	0	0	0	0	0	0	117	1	57	1	2002	1	5	155,010,032.00
20	Spider-Ma	\$ 139,000,000.00	0	0	0	1	0	1	121	1	90	1	2002	1	5	825,021,036.00
21	Devilvil	\$ 78,000,000.00	0	0	1	0	0	1	103	1	69	1	2003	0	5	179,179,718.00
22	X2	\$ 110,000,000.00	0	0	1	0	0	1	134	1	85	1	2003	1	5	407,711,549.00
23	Hulk	\$ 137,000,000.00	0	0	0	0	0	1	138	1	63	1	2003	0	5	245,285,165.00
24	The Punish	\$ 33,000,000.00	0	0	0	0	0	0	124	1	29	1	2004	0	5	54,700,105.00
25	Spider-Ma	\$ 200,000,000.00	0	0	0	1	0	1	127	1	93	1	2004	1	5	788,976,453.00
26	Catwoman	\$ 100,000,000.00	1	0	0	0	1	1	104	0	8	0	2004	0	5	82,402,275.00
27	Blade: Trin	\$ 65,000,000.00	0	0	0	0	0	0	113	1	24	1	2004	0	5	131,977,904.00
28	Elektra	\$ 43,000,000.00	0	0	1	0	0	1	97	1	11	0	2005	0	5	56,995,646.00
29	Constantin	\$ 100,000,000.00	1	0	0	0	0	0	121	0	46	1	2005	0	5	230,884,728.00
30	Batman Re	\$ 150,000,000.00	0	0	0	0	0	1	140	0	85	1	2005	1	5	375,298,946.00
31	Fantastic F	\$ 100,000,000.00	0	0	1	0	0	1	106	1	28	1	2005	1	5	333,535,934.00
32	X-Men: Th	\$ 210,000,000.00	0	0	1	0	0	1	104	1	57	1	2006	0	5	460,435,291.00
33	Superman	\$ 270,000,000.00	1	0	0	0	0	1	154	0	74	1	2006	0	5	391,081,192.00
34	Ghost Ride	\$ 110,000,000.00	0	0	0	1	0	1	110	1	27	1	2007	0	5	228,738,393.00
35	Spider-Ma	\$ 218,000,000.00	0	0	0	1	0	1	139	1	63	1	2007	1	5	894,985,371.00
36	Fantastic F	\$ 130,000,000.00	0	0	1	0	1	0	92	1	38	1	2007	0	5	301,913,131.00
37	Iron Man	\$ 140,000,000.00	0	0	0	0	0	1	126	1	94	1	2008	1	5	585,796,247.00
38	The Incred	\$ 150,000,000.00	0	0	0	0	0	1	112	1	67	1	2008	0	5	264,770,996.00
39	The Dark	\$ 185,000,000.00	1	0	0	0	1	1	152	0	94	1	2008	1	5	1,007,336,937.00
40	Punisher: I	\$ 35,000,000.00	0	0	0	0	0	0	103	1	29	1	2008	0	5	10,161,491.00
41	Watchmen	\$ 180,000,000.00	1	0	0	0	0	0	162	0	65	1	2009	0	5	185,382,811.00
42	X-Men Ori	\$ 150,000,000.00	0	0	1	0	0	1	107	1	38	1	2009	0	5	373,062,864.00
43	Iron Man 2	\$ 200,000,000.00	0	0	0	0	0	1	124	1	72	1	2010	1	5	623,993,331.00
44	Jonah Hex	\$ 47,000,000.00	1	0	0	0	0	1	81	0	12	1	2010	0	5	10,905,312.00
45	Thor	\$ 150,000,000.00	0	0	0	0	0	1	115	1	77	1	2011	1	5	449,326,618.00
46	X-Men: Fir	\$ 160,000,000.00	0	0	1	0	0	1	131	1	86	1	2011	0	5	352,616,690.00
47	Green Lan	\$ 200,000,000.00	1	0	0	0	0	1	114	0	25	1	2011	0	5	237,201,172.00
48	Captain Am	\$ 140,000,000.00	0	0	0	0	1	1	124	1	80	1	2011	1	5	370,569,774.00
49	Ghost Ride	\$ 57,000,000.00	0	0	0	1	0	1	96	1	19	1	2012	0	5	132,563,930.00
50	The Avenge	\$ 220,000,000.00	0	1	0	0	0	1	143	1	91	1	2012	1	5	1,518,812,988.00
51	The Amazi	\$ 230,000,000.00	0	0	0	1	0	1	136	1	71	1	2012	1	5	757,930,661.00
52	The Dark	\$ 250,000,000.00	1	0	0	0	0	1	164	0	87	1	2012	1	5	1,085,067,637.00
53	Iron Man 3	\$ 200,000,000.00	0	1	0	0	0	1	130	1	79	1	2013	1	5	1,215,577,205.00
54	Man of Ste	\$ 225,000,000.00	1	0	0	0	0	1	143	0	56	1	2013	1	5	670,145,518.00
55	The Wolve	\$ 120,000,000.00	0	0	1	0	0	1	126	1	71	1	2013	1	5	414,828,246.00
56	Thor: The	\$ 170,000,000.00	0	1	0	0	0	1	112	1	67	1	2013	1	5	644,783,140.00
57	Captain Am	\$ 170,000,000.00	0	1	0	1	1	1	136	1	90	1	2014	1	5	734,421,501.00
58	The Amazi	\$ 200,000,000.00	0	0	0	1	0	1	142	1	51	1	2014	1	5	708,982,323.00
59	X-Men: Da	\$ 200,000,000.00	0	0	1	0	0	1	132	1	90	1	2014	1	5	746,045,700.00
60	Guardians	\$ 170,000,000.00	0	1	0	0	0	1	121	1	92	1	2014	1	5	773,350,147.00
61	Avengers	\$ 250,000,000.00	0	1	0	0	0	1	141	1	76	1	2015	1	5	1,405,018,048.00
62	Ant-Man	\$ 130,000,000.00	0	1	0	0	0	1	117	1	83	1	2015	1	5	519,311,965.00
63	Fantastic F	\$ 120,000,000.00	0	0	1	0	0	1	100	1	9	1	2015	0	5	167,882,881.00
64	Deadpool	\$ 58,000,000.00	0	0	1	0	0	0	108	1	85	1	2016	1	5	782,612,155.00
65	Batman v	\$ 250,000,000.00	1	0	0	0	0	1	151	0	29	1	2016	1	5	874,362,803.00
66	Captain Am	\$ 250,000,000.00	0	1	0	0	0	1	147	1	91	1	2016	1	5	1,155,046,416.00
67	X-Men: Ap	\$ 178,000,000.00	0	0	1	0	0	1	144	1	47	1	2016	1	5	543,934,105.00
68	Suicide Squ	\$ 175,000,000.00	1	0	0	0	0	1	123	0	26	1	2016	1	5	749,200,054.00
69	Doctor Str	\$ 165,000,000.00	0	1	0	0	0	1	115	1	89	1	2016	1	5	677,796,076.00
70	The Lego	\$ 80,000,000.00	0	0	0	0	1	0	104	1	89	1	2017	1	5	312,336,671.00
71	Logan	\$ 97,000,000.00	0	0	1	0	0	0	137	1	94	1	2017	1	5	619,179,950.00
72	Guardians	\$ 200,000,000.00	0	1	0	0	0	0	136	1	85	1	2017	1	5	863,756,051.00
73	Wonder W	\$ 149,000,000.00	1	0	0	0	0	1	141	0	93	0	2017	1	5	823,970,682.00
74	Spider-Ma	\$ 175,000,000.00	0	0	0	1	0	1	133	1	92	1	2017	1	5	880,166,924.00
75	Thor: Ragi	\$ 180,000,000.00	0	1	0	0	0	1	130	1	89	1	2017	1	5	855,301,886.00
76	Justice Lea	\$ 300,000,000.00	1	0	0	0	0	1	120	0	39	1	2017	0	5	663,324,295.00
77	Black Pan	\$ 200,000,000.00	0	1	0	0	0	1	134	1	96	1	2018	1	5	1,349,926,081.00
78	Avengers	\$ 325,000,000.00	0	1	0	0	0	1	149	1	85	1	2018	1	5	2,052,415,639.00
79	Deadpool	\$ 110,000,000.00	0	0	0	0	0	1	119	1	84	1	2018	1	5	734,546,611.00
80	Ant-Man a	\$ 162,000,000.00	0	1	0	0	0	1	118	1	87	0	2018	1	5	622,674,139.00
81	Teen Titani	\$ 10,000,000.00	1	0	0	0	1	0	84	0	91	0	2018	1	5	52,390,236.00
82	Venom	\$ 100,000,000.00	0	0	0	1	0	1	112	1	30	1	2018	1	5	856,085,151.00
83	Spider-Ma	\$ 90,000,000.00	0	0	0	0	1	1	117	1	97	1	2018	1	5	384,298,736.00
84	Aquaman	\$ 160,000,000.00	1	0	0	0	0	1	143	0	66	1	2018	1	5	1,151,965,807.00
85	Captain Mi	\$ 152,000,000.00	0	1	0	0	0	0	123	1	79	0	2019	1	5	1,131,416,446.00
86	Shazam!	\$ 100,000,000.00	1	0	0	0	0	1	132	0	90	1	2019	1	5	367,799,011.00
87	Avengers	\$ 356,000,000.00	0	1												

# Methodology: Multicollinearity

	Budget	Warner Bros.	Studios	Motieth	Centu	ures	Entertlumny vari	dummy va	Minutes	l dummy v	Tomatoes Crit	dummy va	Year
Budget	1												
Warner Bros	-0.12883	1											
Walt Disney Studios	0.512486	-0.38694	1										
Twentieth Century Fo	-0.09775	-0.3055	-0.22429	1									
Sony Pictures Enterta	0.05764	-0.26175	-0.19218	-0.15173	1								
PG dummy variable	-0.40484	0.204961	-0.19218	-0.07415	-0.04308	1							
PG-13 dummy variab	0.558991	-0.1607	0.298649	0.003596	0.136985	-0.64349	1						
Minutes	0.658657	0.091362	0.257259	-0.19127	-0.00818	-0.24637	0.248534	1					
Marvel dummy variab	0.186907	-0.96203	0.402211	0.317552	0.27208	-0.24707	0.181428	-0.05655	1				
Rotten Tomatoes Crit	0.341093	-0.08525	0.34094	-0.08162	0.061413	-0.00965	0.037973	0.46452	0.13268	1			
Male dummy variable	0.002455	0.045156	-0.19316	0.038603	0.151726	-0.00343	-0.06164	0.069252	-0.00866	0.14325099	1		
Year	0.543275	-0.14367	0.347348	-0.02769	0.071375	-0.38276	0.336124	0.265011	0.22683	0.26515928	-0.20624	1	

Regression  
without Marvel  
dummy variable

Regression Statistics	
Multiple R	0.793261
<b>R Square</b>	<b>0.62926</b>
Adjusted R	0.588886
Standard E	2.96E+08
Observatio	113

Regression Statistics	
Multiple R	0.793168
<b>R Square</b>	<b>0.62912</b>
Adjusted R	0.588722
Standard E	2.96E+08
Observatio	113

Regression  
without Warner  
Bros. dummy  
variable

# Methodology: Removing High p-values

We removed the following variables because their p-values were greater than alpha of 0.05:

1. PG
2. Warner Bros.
3. PG-13
4. Year
5. Twentieth Century Fox

	Coefficients	standard Error	t Stat	P-value	Lower 95%	Upper 95%	lower 95.0%	Upper 95.0%
Intercept	-9.1E+08	2.13E+08	-4.28852	4.01E-05	-1.3E+09	-4.9E+08	-1.3E+09	-4.9E+08
Budget	1.686913	0.562294	3.000056	0.003372	0.571988	2.801837	0.571988	2.801837
Walt Disne	3.3E+08	90530455	3.649674	0.000411	1.51E+08	5.1E+08	1.51E+08	5.1E+08
Twentieth	98405407	83076346	1.184518	0.238882	-6.6E+07	2.63E+08	-6.6E+07	2.63E+08
Sony Pictui	3.03E+08	93702776	3.236362	0.00162	1.17E+08	4.89E+08	1.17E+08	4.89E+08
Minutes	5961955	2100055	2.838952	0.005435	1797934	10125977	1797934	10125977
Rotten Ton	2823078	1199805	2.352947	0.020488	444085.9	5202070	444085.9	5202070
Male dum	1.91E+08	80611923	2.368745	0.019677	31110519	3.51E+08	31110519	3.51E+08

	Coefficients	standard Error	t Stat	P-value	Lower 95%	Upper 95%	lower 95.0%	Upper 95.0%
Intercept	-4.1E+09	7E+09	-0.59312	0.554429	-1.8E+10	9.73E+09	-1.8E+10	9.73E+09
Budget	1.653864	0.706376	2.341336	0.021179	0.252604	3.055125	0.252604	3.055125
Warner Bro	-1.8E+07	87811095	-0.20406	0.838718	-1.9E+08	1.56E+08	-1.9E+08	1.56E+08
Walt Disne	3.18E+08	1.17E+08	2.705481	0.008007	84787369	5.51E+08	84787369	5.51E+08
Twentieth	86043802	1.03E+08	0.83155	0.407625	-1.2E+08	2.91E+08	-1.2E+08	2.91E+08
Sony Pictui	2.92E+08	1.17E+08	2.49177	0.014339	59517900	5.24E+08	59517900	5.24E+08
PG dummy	-1.4E+07	1.24E+08	-0.11358	0.909793	-2.6E+08	2.32E+08	-2.6E+08	2.32E+08
PG-13 dur	-3.1E+07	1E+08	-0.31068	0.756682	-2.3E+08	1.67E+08	-2.3E+08	1.67E+08
Minutes	6065645	2215148	2.738258	0.007301	1671388	10459902	1671388	10459902
Rotten Ton	2687024	1278453	2.101777	0.03806	150916.5	5223131	150916.5	5223131
Male dum	2E+08	86097414	2.323348	0.022165	29240172	3.71E+08	29240172	3.71E+08
Year	1624097	3460406	0.469337	0.63984	-5240419	8488612	-5240419	8488612

	Coefficients	standard Error	t Stat	P-value	Lower 95%	Upper 95%	lower 95.0%	Upper 95.0%
Intercept	-4.4E+09	6.63E+09	-0.66227	0.509291	-1.8E+10	8.76E+09	-1.8E+10	8.76E+09
Budget	1.645367	0.698996	2.353899	0.020494	0.258911	3.031823	0.258911	3.031823
Warner Bro	-2E+07	85725151	-0.23159	0.817317	-1.9E+08	1.5E+08	-1.9E+08	1.5E+08
Walt Disne	3.17E+08	1.16E+08	2.719469	0.007687	85687590	5.48E+08	85687590	5.48E+08
Twentieth	85828455	1.03E+08	0.833652	0.406425	-1.2E+08	2.9E+08	-1.2E+08	2.9E+08
Sony Pictui	2.9E+08	1.15E+08	2.513891	0.013505	61183633	5.19E+08	61183633	5.19E+08
PG-13 dur	-2.5E+07	82054306	-0.30036	0.764513	-1.9E+08	1.38E+08	-1.9E+08	1.38E+08
Minutes	6113569	2164046	2.825063	0.005687	1821193	10405945	1821193	10405945
Rotten Ton	2667182	1260319	2.116275	0.036755	167344.9	5167019	167344.9	5167019
Male dum	2.01E+08	84815766	2.37479	0.019429	33187931	3.7E+08	33187931	3.7E+08
Year	1739520	3291788	0.528442	0.59834	-4789725	8268765	-4789725	8268765

	Coefficients	standard Error	t Stat	P-value	Lower 95%	Upper 95%	lower 95.0%	Upper 95.0%
Intercept	-4.4E+09	6.6E+09	-0.66602	0.506887	-1.7E+10	8.69E+09	-1.7E+10	8.69E+09
Budget	1.625484	0.69051	2.354034	0.020468	0.256021	2.994947	0.256021	2.994947
Walt Disne	3.33E+08	91584415	3.63756	0.000432	1.52E+08	5.15E+08	1.52E+08	5.15E+08
Twentieth	99451299	84107459	1.182431	0.239757	-6.7E+07	2.66E+08	-6.7E+07	2.66E+08
Sony Pictui	3.05E+08	95978056	3.174202	0.001982	1.14E+08	4.95E+08	1.14E+08	4.95E+08
PG-13 dur	-2.5E+07	81642389	-0.3086	0.758249	-1.9E+08	1.37E+08	-1.9E+08	1.37E+08
Minutes	6102427	2153550	2.83366	0.005538	1831369	10373485	1831369	10373485
Rotten Ton	2638270	1248347	2.113412	0.036981	162469.6	5114071	162469.6	5114071
Male dum	2.02E+08	84413712	2.389943	0.018669	34329275	3.69E+08	34329275	3.69E+08
Year	1737756	3276621	0.53035	0.597011	-4760649	8236162	-4760649	8236162

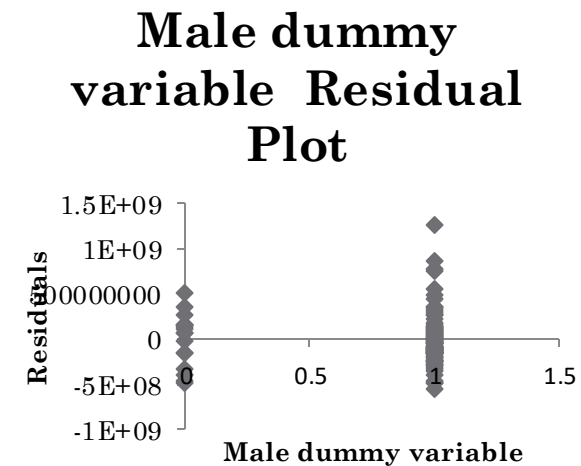
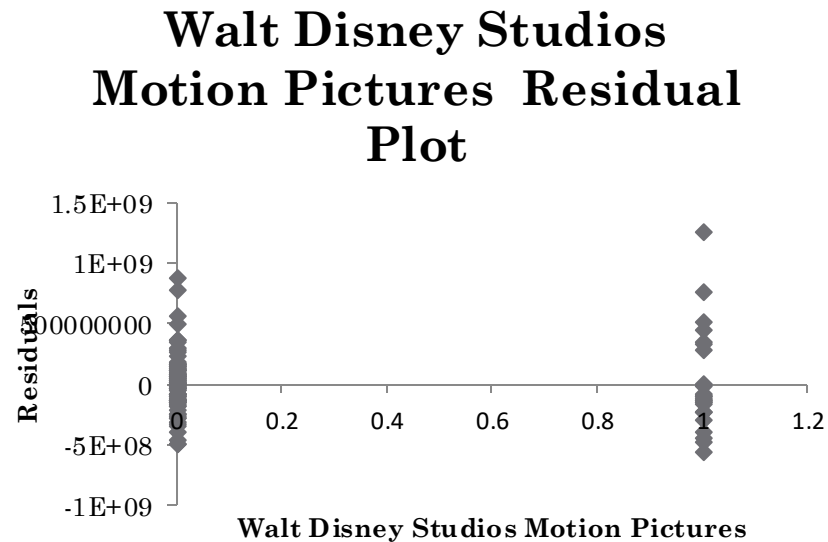
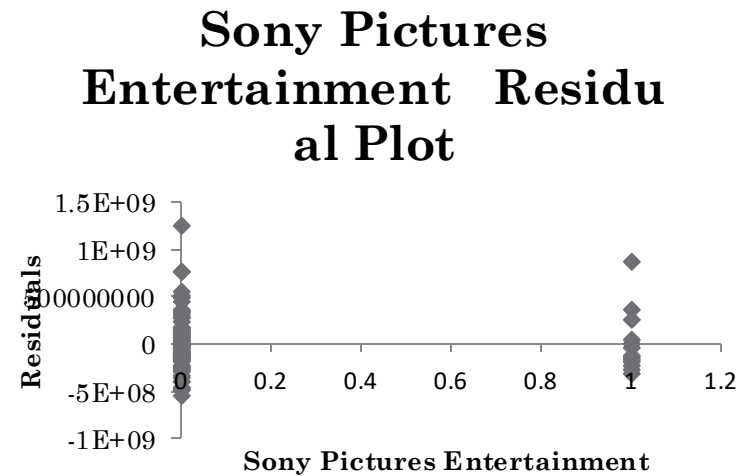
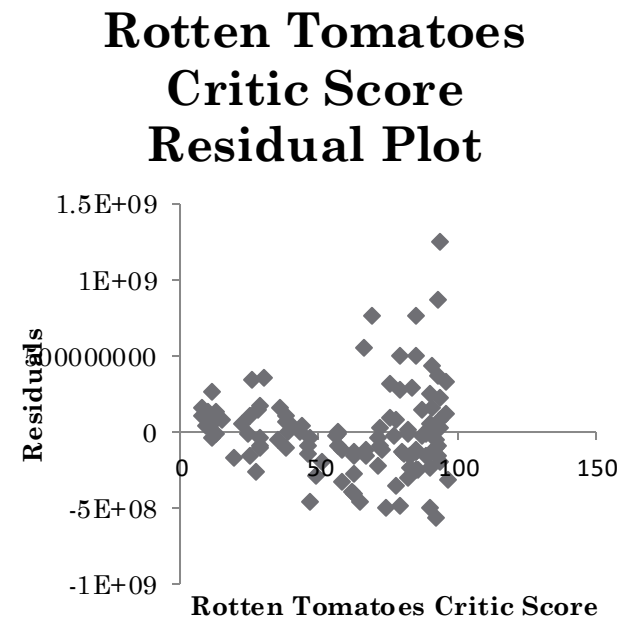
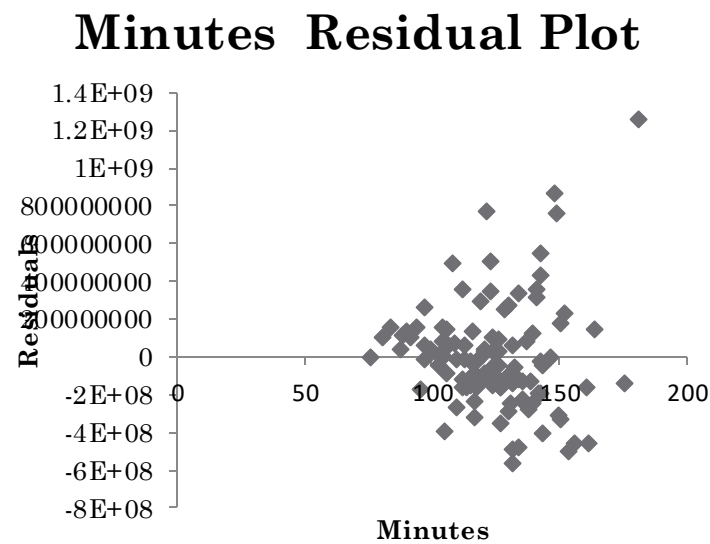
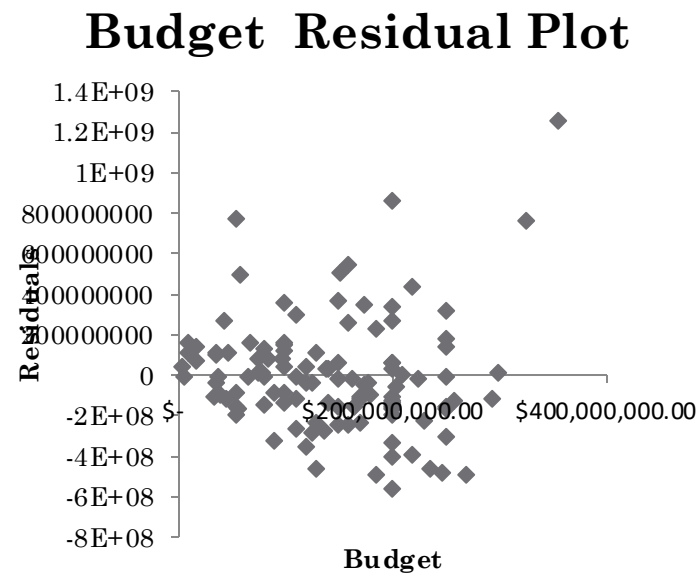
	Coefficients	standard Error	t Stat	P-value	Lower 95%	Upper 95%	lower 95.0%	Upper 95.0%
Intercept	-4.4E+09	6.57E+09	-0.66475	0.507683	-1.7E+10	8.66E+09	-1.7E+10	8.66E+09
Budget	1.539922	0.629644	2.445701	0.016136	0.291313	2.78853	0.291313	2.78853
Walt Disne	3.31E+08	90845684	3.640318	0.000426	1.51E+08	5.11E+08	1.51E+08	5.11E+08
Twentieth	97112629	83400189	1.164417	0.246919	-6.8E+07	2.62E+08	-6.8E+07	2.62E+08
Sony Pictui	3E+08	94258710	3.180428	0.001938	1.13E+08	4.87E+08	1.13E+08	4.87E+08
Minutes	6152208	2138138	2.877367	0.004867	1912200	10392215	1912200	10392215
Rotten Ton	2702085	1225734	2.204463	0.029695	271408.2	5132761	271408.2	5132761
Male dum	2.03E+08	83975225	2.415126	0.017475	36284702	3.69E+08	36284702	3.69E+08
Year	1715665	3261559	0.526026	0.59999	-4752128	8183459	-4752128	8183459



# Final Regression

Regression Statistics								
Multiple R	0.789123016							
R Square	0.622715134							
Adjusted R Square	0.601359387							
Standard Error	291481539.3							
Observations	113							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	6	1.49E+19	2.48E+18	29.15914	2.16E-20			
Residual	106	9.01E+18	8.5E+16					
Total	112	2.39E+19						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-840699887.35	2.04E+08	-4.11423	7.7E-05	-1.2E+09	-435576832.5	-1245822942	-435576832.5
Budget	1.82	0.551471	3.305761	0.001293	0.729687	2.916377818	0.729687003	2.916377818
Walt Disney Studios Motion Pictures	298627087.69	86627894	3.447239	0.000813	1.27E+08	470375312.3	126878863.1	470375312.3
Sony Pictures Entertainment	275368657.00	90869169	3.030386	0.003069	95211692	455525622.4	95211691.61	455525622.4
Minutes	5335277.14	2036182	2.620236	0.010077	1298349	9372205.535	1298348.738	9372205.535
Rotten Tomatoes Critic Score	2976986.39	1195014	2.491172	0.014283	607754.2	5346218.625	607754.1509	5346218.625
Male dummy variable	192004880.02	80760110	2.377472	0.019224	31890107	352119653.3	31890106.75	352119653.3

# Residual Plots





# Final Model

Box Office Gross Worldwide =  $-840699887.35 + 1.82*A + 298627087.69*B + 275368656.9974*C + 5335277.14*D + 2976986.39*E + 192004880.02*F$

A = Budget

B = Walt Disney

C = Sony Pictures Entertainment

D = Minutes

E = Rotten Tomato Rating

F = Male Protagonist

# Conclusion and application

At first, we started with nine variables including the dependent

We took out both MPAA Ratings PG and PG13. Then we took out Warner Bros, Year and 20th Century Fox. This was due to their high p-values

After running our final regression, we found our most significant variable and least significant variable.

# Conclusion Continued

The most significant variable is the Walt Disney Dummy Variable because it showed the highest coefficient. If produced by Walt Disney, the Box Office Gross increases due to the positive relationship

The least significant variable would be Budget, since it had the smallest coefficient. This variable still had a positive coefficient meaning as Budget increases so does the Box Office Gross

# Final Statement

- After completing the statistical regression and evaluating all the variables, we can conclude that Marvel is a very successful movie franchise due to the budget, whether it was produced by Walt Disney or Sony, the number of minutes, Rotten Tomatoes score and if there was a male lead in the movie.



# Work Cited

*Box Office of DC and Marvel Superhero movies*. Kaggle. (n.d.). <https://www.kaggle.com/mdtoomey/box-office-of-dc-and-marvel-superhero-movies>

The dataset: [BAIS Project- dc marvel mvie performance.xls](#)