USC University of Southern California

The Effect of Movie Lead's Gender on Movie's Revenue

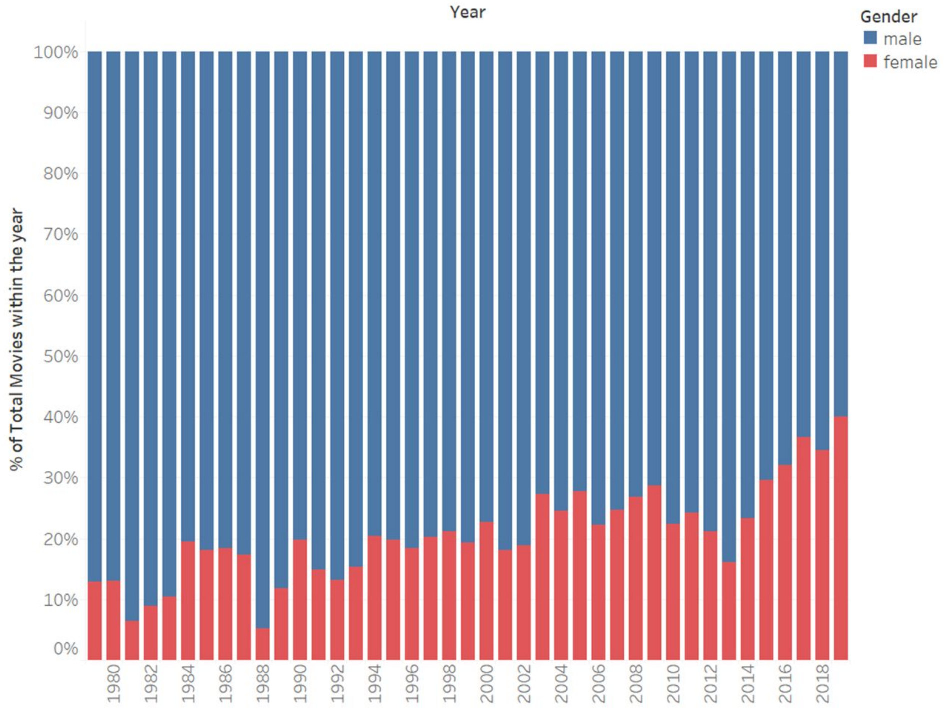
Xiaoxuan (Hayley) Gu
Chaiyapuk (KK) Titinanapun
Xinwen (Gabriel) Wu
Anyi (Aprill) Zhao
Jiaying (Grace) Zheng
Yue (Lydia) Zou

Nov 29, 2022

Background

Background

More films have been produced with a female leading role.



% of Total Count_Released_date for each Year. Color shows details about Gender. The data is filtered on Gross Adj, which includes values greater than or equal to 10. Percents are based on each column of each pane of the table.

USC

Hypothesis and Why Using Data?

Hypothesis:

- Genders of movie leads have impact on movie revenue.

Why using Data?

- Many factors affecting revenue
- Common perception V.S. Data justification*

USC

Ideal Experiment

Create a series pairs of two movies with the **same qualities** but **different genders** of the lead as follow:

- Genres
- Scripts
- Directors
- Two movie leads with same popularity and acting skill
- Constructed in a similar way
- Displayed in the same place

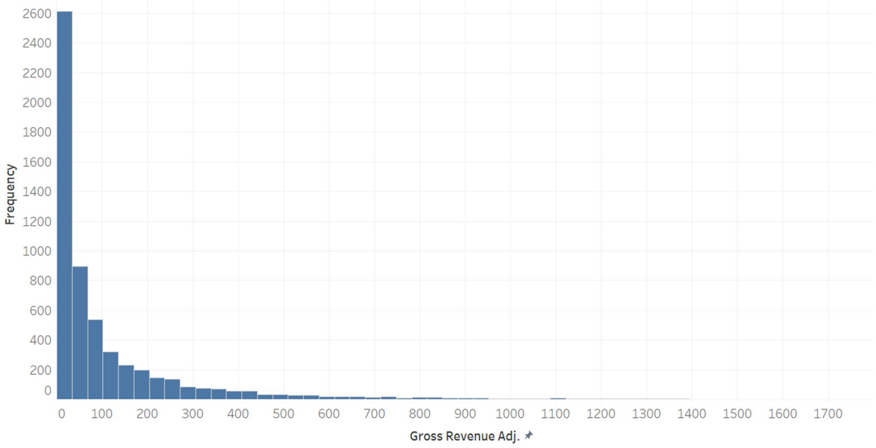
Our Dataset

	Variables	Description
Dependent Variable	Gross	Revenue of a movie (in million dollars); Adjusted by inflation rate; Revenue lower than 10 millions dollars were removed
Independent Variable	Gender	Gender of leading character of the movie
	Budget	Budget of a movie (in million dollars); Adjusted by inflation rate;
	Year	Releasing year of a movie; 1: 1980, 41: 2020
	Scores	IMDB score of a movie; range: 0 - 10
	Rating	Age limits of a movie G and PG; PG-13; R, NC-17 and others
	Genres	Categories of a movie Adventures, action, comedy, drama movies, and others

Revenue Distribution:

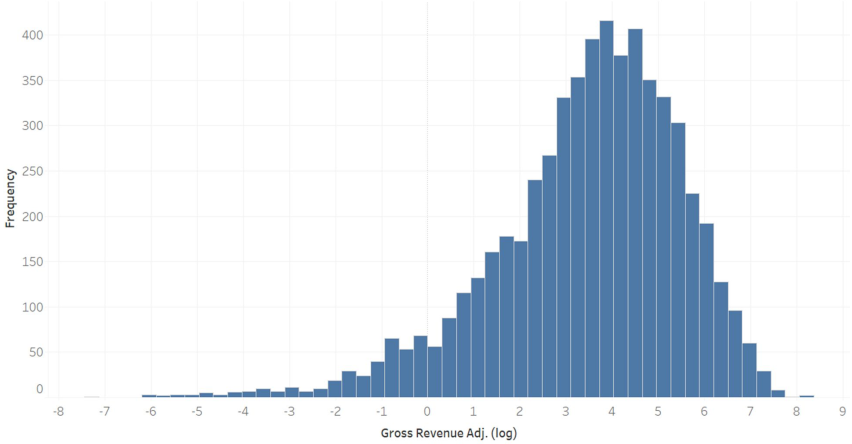
- Highly right-skewed

Distribution of Gross Revenue (linear)



The trend of count of Gross Adj for Gross Adj (bin).

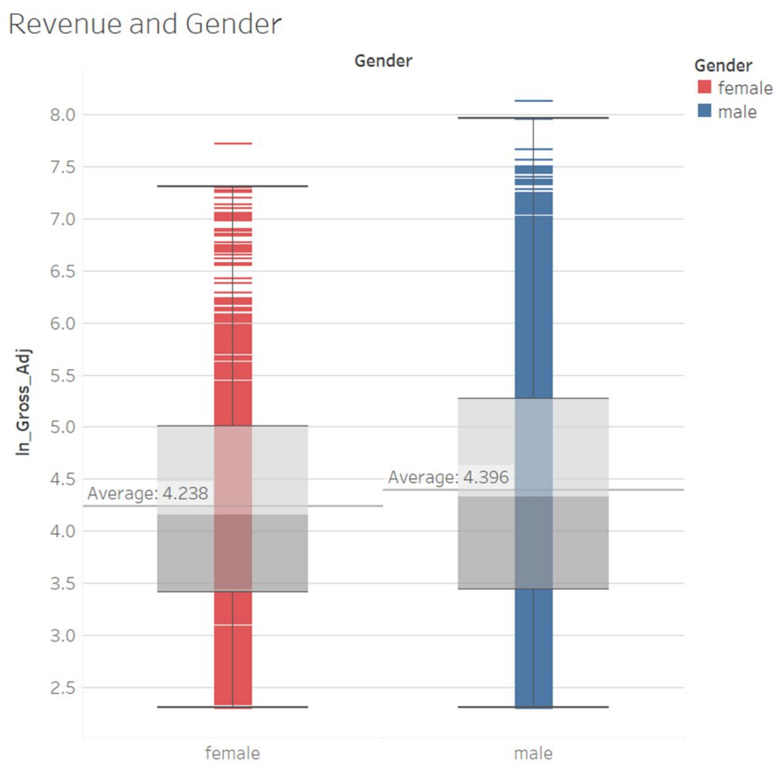
Distribution of Gross Revenue (log)



The trend of count of ln_Gross_Adj for ln_Gross_Adj (bin).

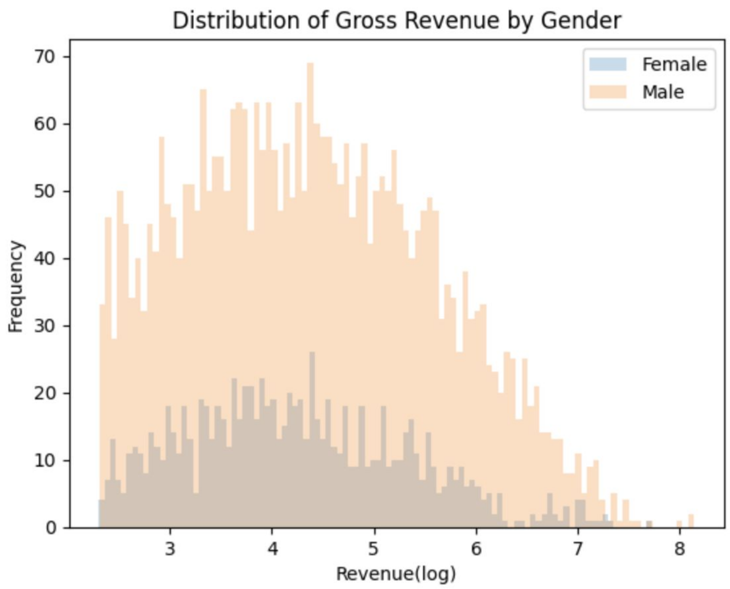
USC

Revenue vs Gender:



Ln_Gross_Adj for each Gender. Color shows details about Gender. The data is filtered on Gross Adj, which includes values greater than or equal to 10.

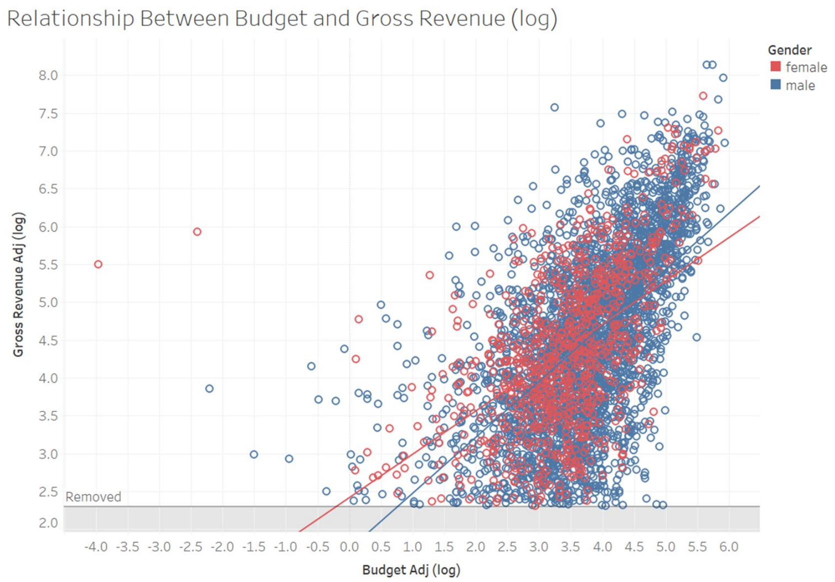
- Males' revenue are higher than Females' in general and on average.
- There are less high revenue samples in female-lead movies



USC

Revenue vs Budget:

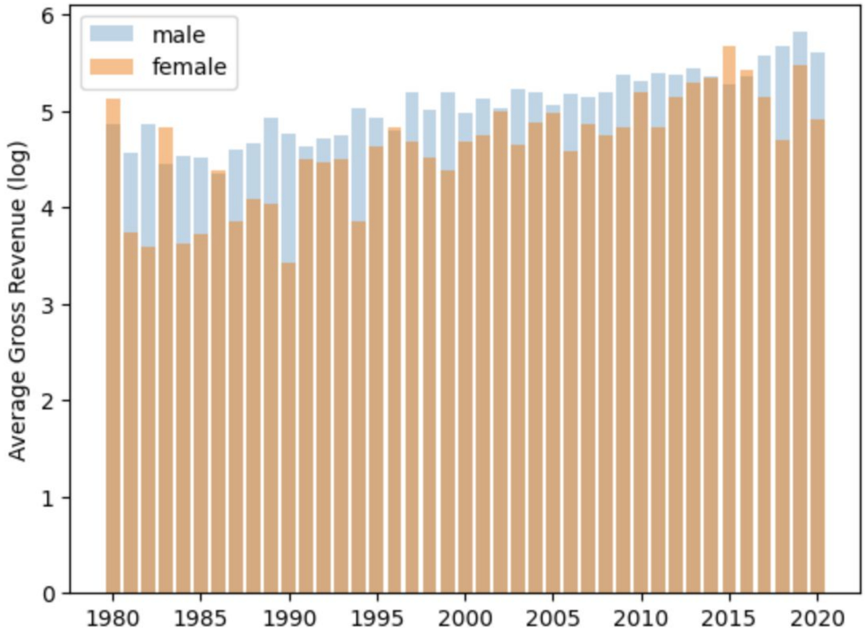
- Positive
- Determine potential



Ln_Budget_Adj vs. ln_Gross_Adj. Color shows details about Gender. Details are shown for Name and Released. The data is filtered on Gross Adj, which includes values greater than or equal to 10.

Revenue vs Year:

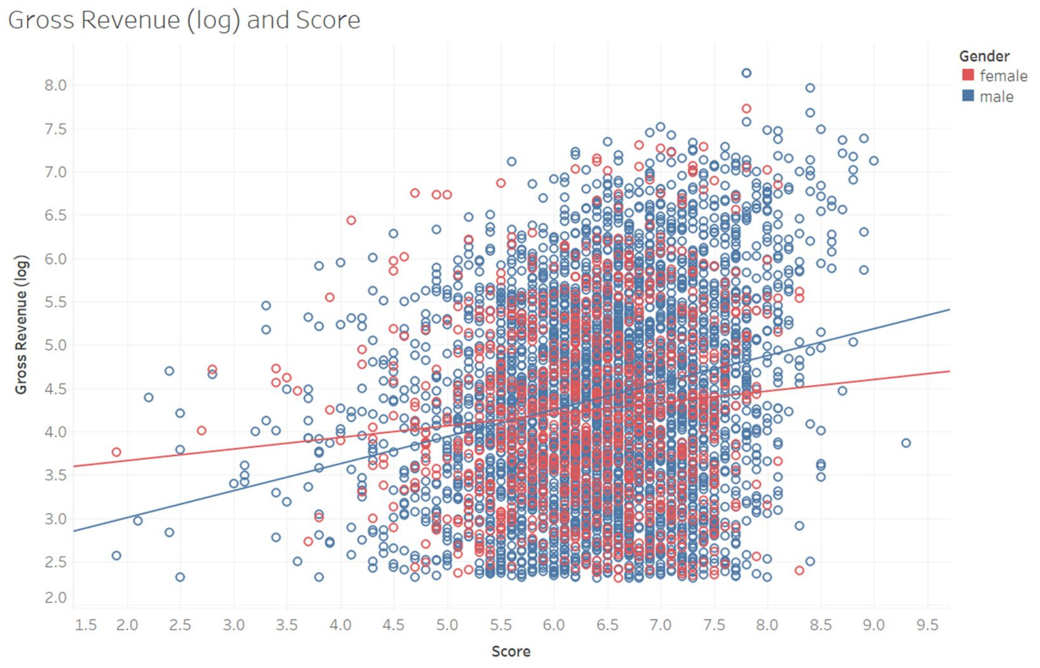
- Inflation adjusted



USC

Revenue vs Score:

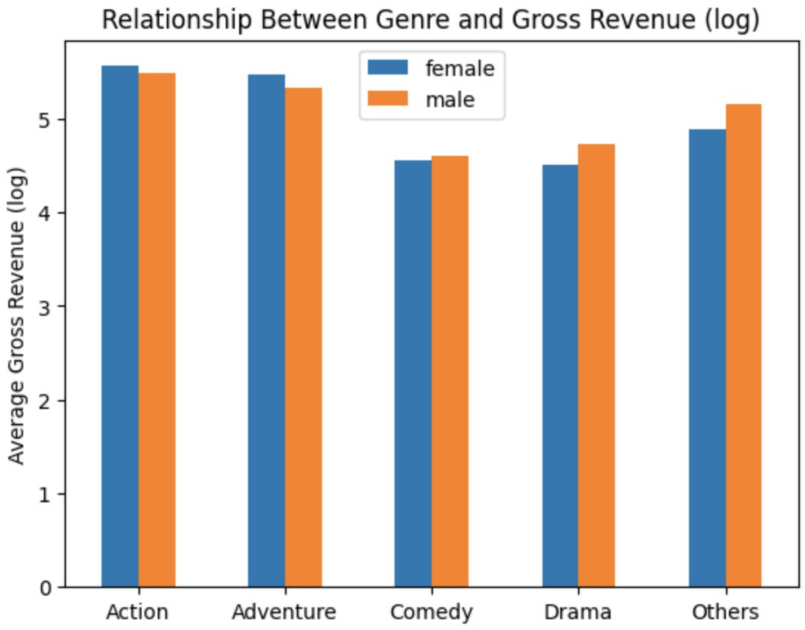
- Quality of a movie



Score vs. ln_Gross_Adj. Color shows details about Gender. The data is filtered on Gross Adj, which includes values greater than or equal to 10.

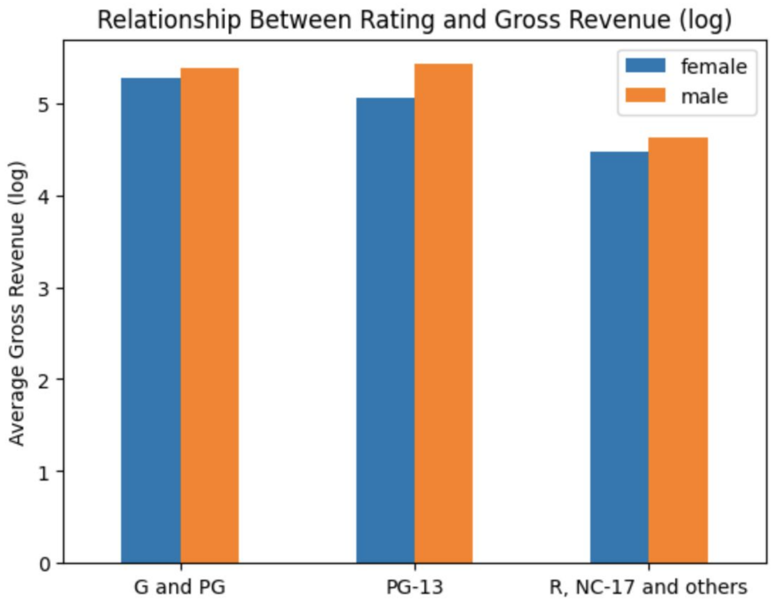
Revenue vs Genre:

- Grouped categories



Revenue vs Rating:

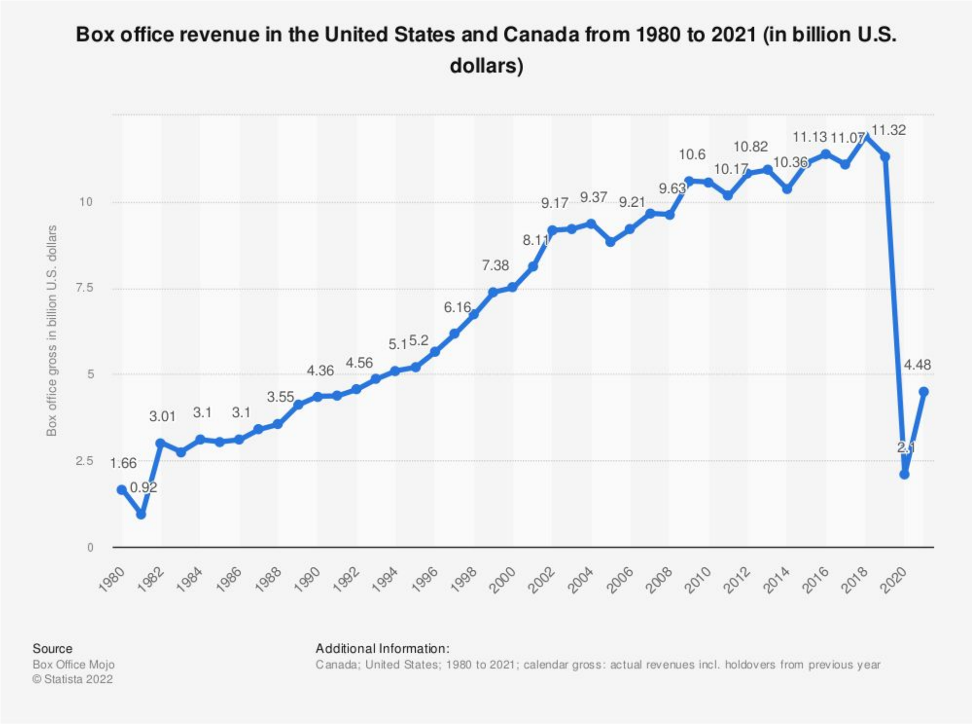
- Limits targeted audience



Different average revenues for different categories

Model and Interpretation

Control Variables



- Budget adj: budget of a movie determines its potential (adjusted by inflation)
- Year: movies industry has been growing in the last 40 years
- Scores: measures quality of a movie
  Rating: limits targeted audience of a movie
- Genres: adventures, action, comedy, drama movies have the highest box office revenue in the North America market

OLS Regression Results			
R-squared	0.441	Adj. R-squared	0.440
F-statistic	340.2	Prob (F-statistic)	0.00

	Coefficient	P-Value
Intercept	0.0793	0.536
Rating[PG-13]	-0.1644	0.000
Rating[R, NC and Others]	-0.4581	0.000
Genre[Adventure]	-0.1921	0.003
Genre[Comedy]	-0.1729	0.000
Genre[Drama]	-0.3357	0.000
Genre[Others]	-0.1207	0.004
ln(Budget)	0.5999	0.000
Year	0.0163	0.000
Score	0.3451	0.000

ln(gross) ~ ln(budget) + year + score + rating + genre

- Without gender
- Adj. R-squared: 0.440

OLS Regression Results			
R-squared	0.443	Adj. R-squared	0.441
F-statistic	308.2	Prob (F-statistic)	0.00

	Coefficient	P-Value
Intercept	0.1578	0.225
Gender[Male]	-0.1230	0.001
Rating[PG-13]	-0.1720	0.000
Rating[R, NC and Others]	-0.4600	0.000
Genre[Adventure]	-0.2002	0.002
Genre[Comedy]	-0.1897	0.000
Genre[Drama]	-0.3588	0.000
Genre[Others]	-0.1346	0.001
ln(Budget)	0.6054	0.000
Year	0.0157	0.000
Score	0.3462	0.000

ln(gross) ~ gender + ln(budget) + year + score + rating + genre

- Gender:
  coef = -0.1230 (p-value=0.001)
  $(e^{-0.1230} - 1) * 100 = -11.57\%$
- Controlling other variables, male-lead movies make 11.57% less revenue compared to female-leads movies
- Adj. R-squared changes little
- The relationships between revenue and control variables are the same

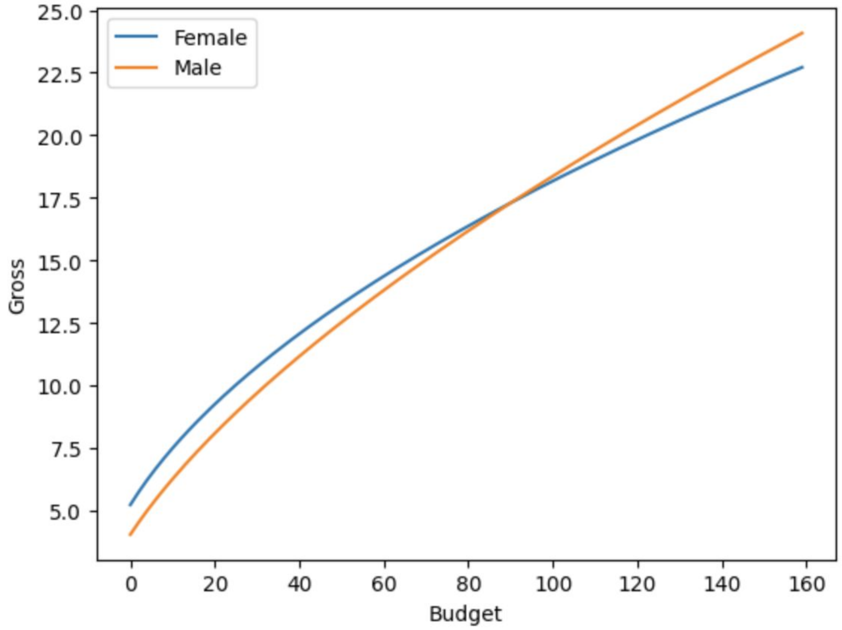
USC

OLS Regression Results			
R-squared	0.444	Adj. R-squared	0.443
F-statistic	281.8	Prob (F-statistic)	0.00

	Coefficient	P-Value
Intercept	0.4576	0.004
Gender[Male]	-0.5145	0.000
Gender[Male]*ln(Budget)	0.1117	0.002
Rating[PG-13]	-0.1725	0.000
Rating[R, NC and Others]	-0.4562	0.000
Genre[Adventure]	-0.1941	0.003
Genre[Comedy]	-0.1837	0.000
Genre[Drama]	-0.3551	0.000
Genre[Others]	-0.1342	0.001
ln(Budget)	0.5196	0.000
Year	0.0154	0.000
Score	0.3456	0.000

ln(gross) ~ gender + gender * ln(budget) + ln(budget) + year + score + rating + genre

- Gender * ln(budget):
  coef = 0.1117 (p-value=0.002)
  = 0.1117% more rev per 1%



OLS Regression Results			
R-squared	0.447	Adj. R-squared	0.446
F-statistic	261.7	Prob (F-statistic)	0.00

	Coefficient	P-Value
Intercept	1.0999	0.000
Gender[Male]	-1.2699	0.000
Gender[Male]*score	0.1813	0.000
Rating[PG-13]	-0.1870	0.000
Rating[R, NC and Others]	-0.4754	0.000
Genre[Adventure]	-0.2072	0.002
Genre[Comedy]	-0.1944	0.000
Genre[Drama]	-0.3593	0.000
Genre[Others]	-0.1430	0.001
ln(Budget)	0.6030	0.000
Year	0.0159	0.000
Score	0.1998	0.000

ln(gross) ~ gender + gender * score + ln(budget) + year + score + rating + genre

- Gender:
  coef = -1.2699 (p-value=0.000)
  $(e^{-1.2699} - 1) * 100 = -71.91\%$
- Gender * score:
  coef = 0.1813 (p-value=0.000)
  $(e^{0.1813} - 1) * 100 = 19.88\%$
- Male will generate more revenue than female when the score is 7 or higher. (percentile 72)

USC

	coef	std err	t	P>|t|
Intercept	0.0297	0.155	0.192	0.848
gender[T.male]	0.0222	0.102	0.219	0.827
rating_adj[T.PG-13]	-0.1706	0.042	-4.050	0.000
rating_adj[T.R, NC-17 and others]	-0.4595	0.041	-11.076	0.000
genre_adj[T.Adventure]	-0.2015	0.065	-3.076	0.002
genre_adj[T.Comedy]	-0.1864	0.040	-4.636	0.000
genre_adj[T.Drama]	-0.3564	0.048	-7.422	0.000
genre_adj[T.Others]	-0.1316	0.042	-3.152	0.002
np.log(budget_adj)	0.6072	0.016	37.109	0.000
year	0.0202	0.003	6.080	0.000
gender[T.male]:year	-0.0056	0.004	-1.524	0.128
score	0.3458	0.016	22.190	0.000

	coef	std err	t	P>|t|
Intercept	0.1404	0.149	0.945	0.345
gender[T.male]	-0.1115	0.084	-1.326	0.185
rating_adj[T.PG-13]	-0.1738	0.042	-4.122	0.000
rating_adj[T.R, NC-17 and others]	-0.4611	0.042	-11.103	0.000
genre_adj[T.Adventure]	-0.1602	0.166	-0.963	0.336
genre_adj[T.Comedy]	-0.1989	0.096	-2.065	0.039
genre_adj[T.Drama]	-0.3951	0.105	-3.760	0.000
genre_adj[T.Others]	-0.0558	0.101	-0.550	0.582
genre_adj[T.Adventure]:gender[T.male]	-0.0471	0.179	-0.263	0.793
genre_adj[T.Comedy]:gender[T.male]	0.0162	0.104	0.157	0.875
genre_adj[T.Drama]:gender[T.male]	0.0565	0.116	0.487	0.626
genre_adj[T.Others]:gender[T.male]	-0.0998	0.109	-0.916	0.360
np.log(budget_adj)	0.6066	0.016	37.067	0.000
year	0.0157	0.001	10.827	0.000
score	0.3465	0.016	22.210	0.000

Other Interaction Terms: Insignificant (P-value >0.05)

ln(gross) ~ gender + gender * year + ln(budget) + year + score + rating + genre

ln(gross) ~ gender + gender * genre + ln(budget) + year + score + rating + genre

Conclusions and Limitations

USC

Limitations

- We didn't control for these things
  - Effects from people
    - Cast List
    - Director
    - Producer
    - Composer
    - Cinematographer
    - Distributor
  - Movie Title
  - Social Factor (Youtube, Tweets, Wikipedia)

- Score and Budget can be a proxy everything else (e.g. Director, Cast, Producer). However, we do not know the separated effect from each of them.
- We do not have a lot of big name female movies (high-budget & high-score) in comparison to male movies.
- Big name male movies:
  - Titanic, Avengers, Pirate of the Caribbean, etc.
- Big name female movies:
  - Star Wars 7-9

Conclusions:

- Gender does **affect** the revenue.
- Controlling all other variables, movies with male leads seem to generate **11.57% less** revenue.
- The slopes for gender with budget and gender with score are **significant**.
    - Within **top 14th percentile budget**, male led movies tend to generate more revenue
    - Within **top 28th percentile score**, male led movies tend to generate more revenue

Thank you

The Effect of Movie Lead's Gender on Movie's Revenue