QNT 2020 — Foundations of Analytics — Spring 2026

Predicting Worldwide Gross Income of Feature Films

Group 2 | Alon Shabshin, Vincent Viktor Neyman, Adam Abdelaziz | April 27, 2026

---

1. Introduction and Motivation

The linear progression model we worked on can help predict the gross income for a film being made today given variables that are available for executive producers. Movie releases nowadays have massive budgets, and some make a great gross income, some flop and cause big losses. Many hypothesize what variables matter. In this model we chose to investigate the films budget, average grossing of the director, runtime, Rotten Tomato rating, and whether it was adapted from an existing intellectual property or not. We investigated 25 feature films released in the last few years and performed a linear regression check. Our model helps decide which observable factors most reliably predict worldwide gross.

2. Data and Variable Definitions

Dependent Variable

> **Y — Worldwide Gross ($M):** Worldwide gross is the money made from box office sales in domestic and international theatres. This variable is measured in millions of dollars and is the clearest way to measure a film's commercial success. This is the bottom line of film production. When studios decide to greenlight a project, it is because they believe the project will prove financially successful. The source for this information is Box Office Mojo.

Independent Variables

- **X1 — Production Budget ($M):** We hypothesize that the more budget a film has, the bigger the financial success and ratings would be. The source for this information is from The Numbers and is reported in millions of dollars.
- **X2 — Director Career Average ($M):** This variable also supports the hypothesis that more established directors will make a more successful movie or at least attract a bigger audience. This is a calculation of the directors' prior film worldwide gross. The source is Box Office Mojo as well.
- **X3 — Runtime (minutes):** This variable was chosen as movies recently became very long, and it is there to see the implication of runtime on the overall financial success of the film. This is a combination of

impact. 1. Longer runtimes can deter people from committing to long movies. 2. Longer runtimes can allow for less showing of the movie in theatres. The information was taken from IMDB.

- **X4 — Rotten Tomatoes Score:** The Rotten Tomato score is an aftermath, but we wanted to see the correlation between the films overall rating, and the worldwide gross. The Tomatoemeter. The hypothesis is that critical reception drives word-of-mouth and awards positioning. The source is rottentomatoes.com.
- **X5 — Existing IP (1/0):** This variable supports the hypothesis that movies based on existing intellectual property sequel/remake/adaptation, will bring in a larger audience, as the audience is already familiar with the story and are interested in it. 1= sequel/remake/adaptation, 0 = original screenplay.

3. Correlation Analysis + Models and Interpretation of Results

The correlation data shows that the best predictor of success is the Director's Track Record ($R \sim 0.72$), followed closely by the budget. However, there is an overlap between these two. We assume that either the studios entrust bigger budgets to more reputable directors, or the directors with a lower average require less budget to hire. Because of this, Budget was actually dropped during our backward elimination process. It became redundant once the director's average gross was in the model, as the director's name "absorbed" the financial scale of the project. We also found that while Existing IP shows a solid moderate correlation, Critical Ratings (RT Score) were almost entirely irrelevant to the bottom line ($p = 0.823$).

Our final model is statistically significant ($F < 0.0001$) and explains 63.7% of the variance in box office success (Adjusted $R^2 = 0.637$). All three remaining variables are significant with p-values below 0.05:

DirectorAVG (+2.28): For every additional $1M in the director's career average, the film earns roughly $2.28M more. This confirms the director's previous success is a more efficient predictor than the film budget itself.

Existing IP (+487.39): This comes to show that a movie simply being a sequel or established brand is worth a massive $487M boost in worldwide gross, regardless of other factors.

Runtime (-8.00): Runtime shows a negative relationship where shorter films tend to be more successful financially. For every extra minute of runtime, our model predicts a decrease of $8M. We assume it is because longer movies mean fewer daily screenings per screen, which imposes a physical ceiling on how much revenue a theater can generate in a day, and also, people might have personal preferences to attend shorter screenings.

By removing the insufficient variables, we are left with the following final model:

$$\hat{Y} = 972.48 + 2.28 \cdot (DirectorAVG) - 8.00 \cdot (Runtime) + 487.39 \cdot (ExIP)$$

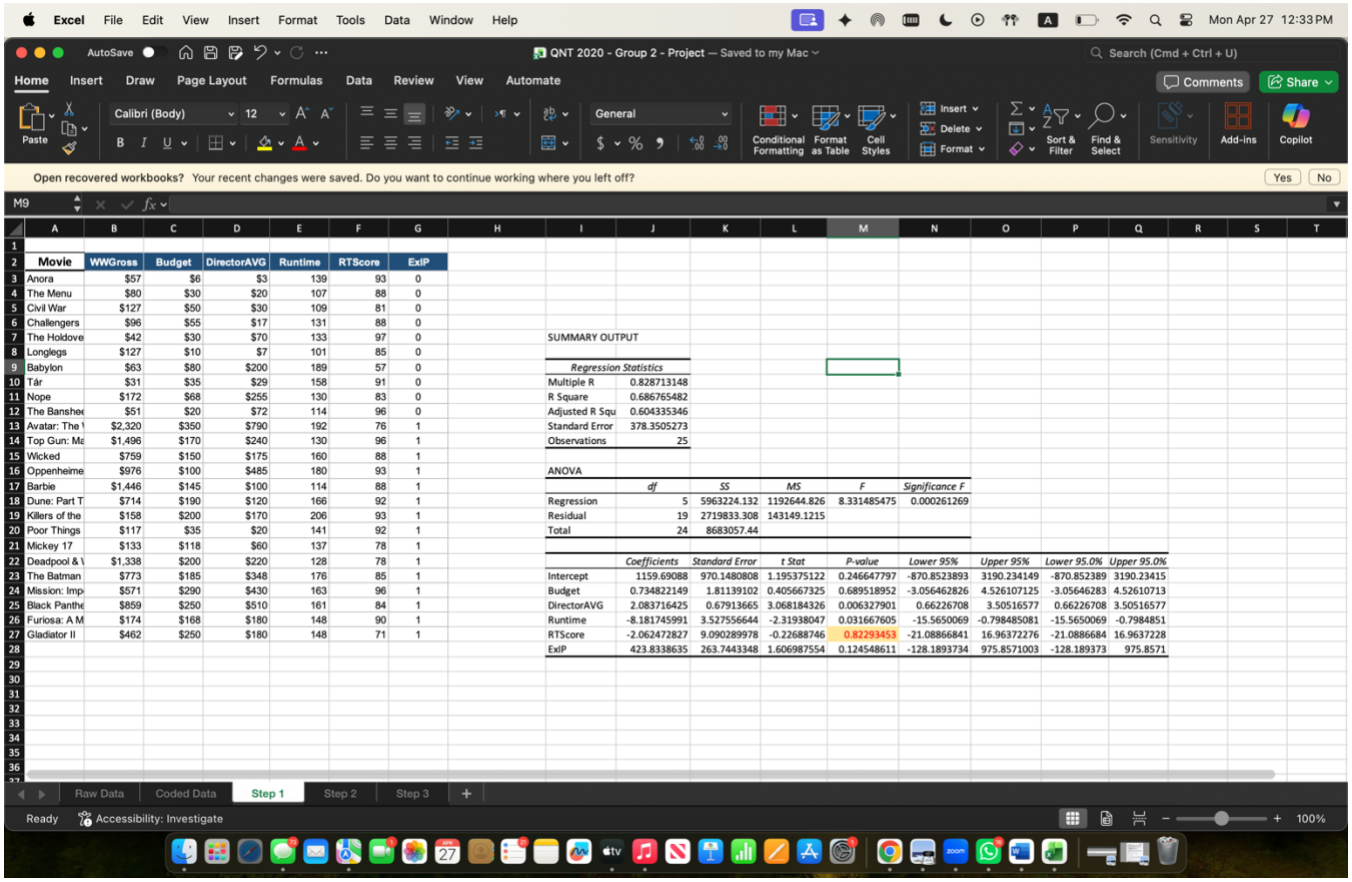
5. Conclusion and Managerial Importance

From our model we can learn the following managerial takeaways:

- **1.** Reputable directors are more likely to increase the financial success of the film. Capital allocated to securing proven directors is more efficient than capital allocated to scaling the entire production scope.
- **2.** Shorter runtimes could help increase the box office gross. The −$8M-per-minute coefficient implies a real economic case for keeping films tight. A 90-minute cut versus a 150-minute cut implies roughly $480M of latent gross.
- **3.** Our model predicts that creating an original movie could start the production off approximately $487M behind a movie based on existing intellectual property. This isn't an argument against original filmmaking, Anora and The Holdovers earned awards prestige, but it explains why studio slates have tilted so heavily toward sequels and remakes.

Appendix — Excel Output

A.1 Initial Model — All 5 Independent Variables



Movie	WWGross	Budget	DirectorAVG	Runtime	RTScore	ExIP
Anora	$57	$6	$3	139	93	0
The Menu	$80	$30	$20	107	88	0
Civil War	$127	$50	$30	109	81	0
Challengers	$96	$55	$17	131	88	0
The Holdove	$42	$30	$70	133	97	0
Longlegs	$127	$10	$7	101	85	0
Babylon	$63	$80	$200	189	57	0
Tár	$31	$35	$29	158	91	0
Nope	$172	$68	$255	130	83	0
The Banshee	$51	$20	$72	114	96	0
Avatar: The	$2,320	$350	$790	192	76	1
Top Gun: Ma	$1,496	$170	$240	130	96	1
Wicked	$759	$150	$175	160	88	1
Oppenheime	$976	$100	$485	180	93	1
Barbie	$1,446	$145	$100	114	88	1
Dune: Part T	$714	$190	$120	166	92	1
Killers of the	$158	$200	$170	206	93	1
Poor Things	$117	$35	$20	141	92	1
Mickey 17	$133	$118	$60	137	78	1
Deadpool & \	$1,338	$200	$220	128	78	1
The Batman	$773	$185	$348	176	85	1
Mission: Imp	$571	$290	$430	163	96	1
Black Panthe	$859	$250	$510	161	84	1
Furiosa: A M	$174	$168	$180	148	90	1
Gladiator II	$462	$250	$180	148	71	1

SUMMARY OUTPUT

Regression Statistics

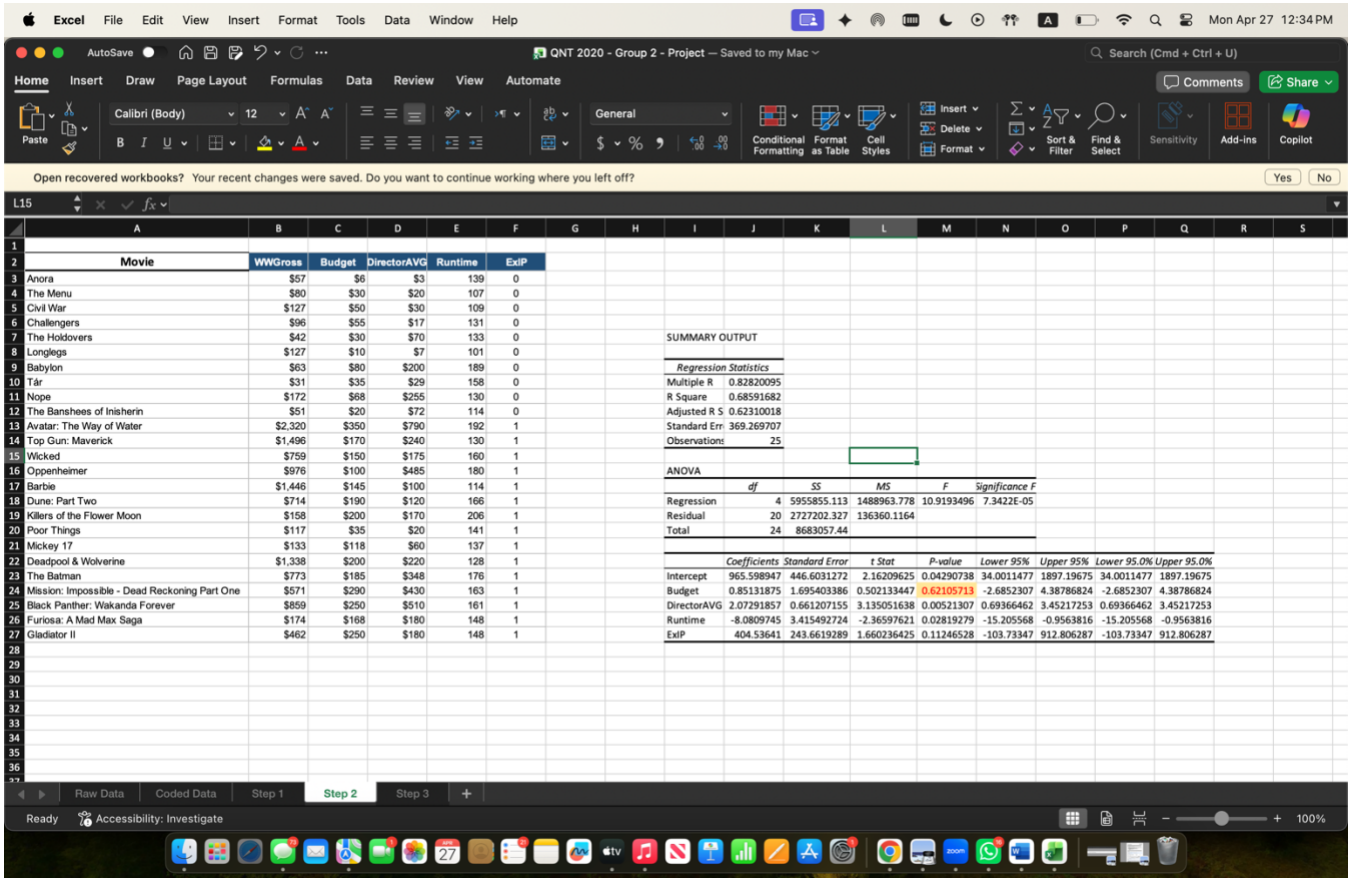
Multiple R	0.828713148
R Square	0.686765482
Adjusted R Squ	0.604335346
Standard Error	378.3505273
Observations	25

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	5963224.132	1192644.826	8.331485475	0.000261269
Residual	19	2719833.308	143149.1215		
Total	24	8683057.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1159.69088	970.1480808	1.195375122	0.246647797	-870.8523893	3190.234149	-870.852389	3190.23415
Budget	0.734822149	1.81139102	0.405667325	0.689518952	-3.056462826	4.526107125	-3.05646283	4.52610713
DirectorAVG	2.083716425	0.67913665	3.068184326	0.006327901	0.66226708	3.50516577	0.66226708	3.50516577
Runtime	-8.181745991	3.527556644	-2.31938047	0.031667605	-15.5650069	-0.798485081	-15.5650069	-0.7984851
RTScore	-2.062472827	9.090289978	-0.22688746	0.82293453	-21.08866841	16.96372276	-21.0886684	16.9637228
ExIP	423.8338635	263.7443348	1.606987554	0.124548611	-128.1893734	975.8571003	-128.189373	975.8571

A.2 Step 1 — After Removing RTScore



Movie	WWGross	Budget	DirectorAVG	Runtime	ExIP
Anora	$57	$6	$3	139	0
The Menu	$80	$30	$20	107	0
Civil War	$127	$50	$30	109	0
Challengers	$96	$55	$17	131	0
The Holdovers	$42	$30	$70	133	0
Longlegs	$127	$10	$7	101	0
Babylon	$63	$80	$200	189	0
Tár	$31	$35	$29	158	0
Nope	$172	$68	$255	130	0
The Banshees of Inisherin	$51	$20	$72	114	0
Avatar: The Way of Water	$2,320	$350	$790	192	1
Top Gun: Maverick	$1,496	$170	$240	130	1
Wicked	$759	$150	$175	160	1
Oppenheimer	$976	$100	$485	180	1
Barbie	$1,446	$145	$100	114	1
Dune: Part Two	$714	$190	$120	166	1
Killers of the Flower Moon	$158	$200	$170	206	1
Poor Things	$117	$35	$20	141	1
Mickey 17	$133	$118	$60	137	1
Deadpool & Wolverine	$1,338	$200	$220	128	1
The Batman	$773	$185	$348	176	1
Mission: Impossible - Dead Reckoning Part One	$571	$290	$430	163	1
Black Panther: Wakanda Forever	$859	$250	$510	161	1
Furiosa: A Mad Max Saga	$174	$168	$180	148	1
Gladiator II	$462	$250	$180	148	1

SUMMARY OUTPUT

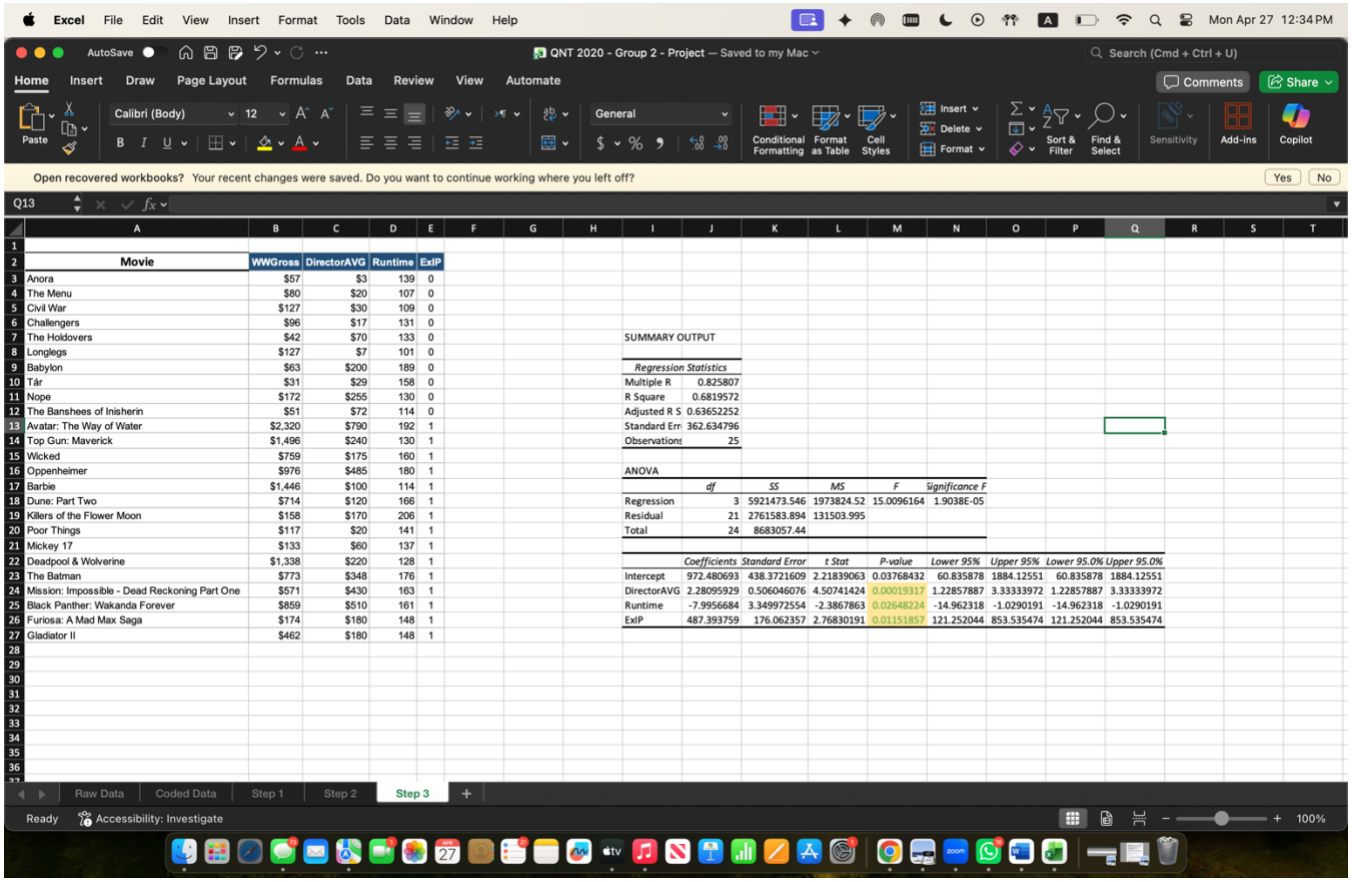
Regression Statistics	
Multiple R	0.82820095
R Square	0.68591682
Adjusted R S	0.62310018
Standard Err	369.269707
Observations	25

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	5955855.113	1488963.778	10.9193496	7.3422E-05
Residual	20	2727202.327	136360.1164		
Total	24	8683057.44			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	965.598947	446.6031272	2.16209625	0.04290738	34.0011477	1897.19675	34.0011477	1897.19675
Budget	0.85131875	1.695403386	0.502133447	0.62105713	-2.6852307	4.38786824	-2.6852307	4.38786824
DirectorAVG	2.07291857	0.661207155	3.135051638	0.00521307	0.69366462	3.45217253	0.69366462	3.45217253
Runtime	-8.0809745	3.415492724	-2.36597621	0.02819279	-15.205568	-0.9563816	-15.205568	-0.9563816
ExIP	404.53641	243.6619289	1.660236425	0.11246528	-103.73347	912.806287	-103.73347	912.806287

A.3 Step 2 — After Removing RTScore and Budget



Movie	WWGross	DirectorAVG	Runtime	ExIP
Anora	$57	$3	139	0
The Menu	$80	$20	107	0
Civil War	$127	$30	109	0
Challengers	$96	$17	131	0
The Holdovers	$42	$70	133	0
Longlegs	$127	$7	101	0
Babylon	$63	$200	189	0
Tár	$31	$29	158	0
Nope	$172	$255	130	0
The Banshees of Inisherin	$51	$72	114	0
Avatar: The Way of Water	$2,320	$790	192	1
Top Gun: Maverick	$1,496	$240	130	1
Wicked	$759	$175	160	1
Oppenheimer	$976	$485	180	1
Barbie	$1,446	$100	114	1
Dune: Part Two	$714	$120	166	1
Killers of the Flower Moon	$158	$170	206	1
Poor Things	$117	$20	141	1
Mickey 17	$133	$60	137	1
Deadpool & Wolverine	$1,338	$220	128	1
The Batman	$773	$348	176	1
Mission: Impossible - Dead Reckoning Part One	$571	$430	163	1
Black Panther: Wakanda Forever	$859	$510	161	1
Furiosa: A Mad Max Saga	$174	$180	148	1
Gladiator II	$462	$180	148	1

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.825807
R Square	0.6819572
Adjusted R S	0.63652252
Standard Err	362.634796
Observations	25

ANOVA

	df	SS	MS	F	Significance F
Regression	3	5921473.546	1973824.52	15.0096164	1.9038E-05
Residual	21	2761583.894	131503.995		
Total	24	8683057.44			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	972.480693	438.3721609	2.21839063	0.03768432	60.835878	1884.12551	60.835878	1884.12551
DirectorAVG	2.28095929	0.506046076	4.50741424	0.00019317	1.22857887	3.33333972	1.22857887	3.33333972
Runtime	-7.9956684	3.349972554	-2.3867863	0.02648224	-14.962318	-1.0290191	-14.962318	-1.0290191
ExIP	487.393759	176.062357	2.76830191	0.01151857	121.252044	853.535474	121.252044	853.535474

A.6 Data Source

Variable	Definition	Primary Source	URL Pattern
Y	Worldwide gross (USD millions, lifetime theatrical)	Box Office Mojo	https://www.boxofficemojo.com
X1	Production budget (USD millions, reported)	The Numbers / IMDb	https://www.the-numbers.com
X2	Director's prior career-average worldwide gross across all theatrical features (excludes the film itself)	Box Office Mojo	https://www.boxofficemojo.com
X3	Theatrical runtime in minutes (single deterministic value per film)	IMDb / Box Office Mojo movie page	Listed under 'Runtime' on each title page
X4	Rotten Tomatoes Tomatometer (critic %) at end of theatrical run	Rotten Tomatoes	https://www.rottentomatoes.com
X5	Existing IP (1 = sequel/remake/adaptation; 0 = original screenplay)	Editorial classification	Defined per project specification