ASSESSMENT TASK 3: MOVIE COLLECTIONS ANALYSIS

Peter McNamara (88451999)

Multi-Level Analysis of Movie Collections

In April 2019, Disney released the upcoming schedule of movies they will be releasing (Quinn 2019). The list is made up of many sequels, remakes and crossovers, such as Toy Story 4, Frozen 2, 4 new Avatar movies and of course more Star Wars movies. As part of the previous analysis the team did on movie revenue, it was found that franchise and sequel movies did much better at the box office than original content movies.

However, this analysis did not consider the variability of individual series. For example, in 2016 and 2017, the Jumanji and Independence Day collections were each rebooted with a sequel. Jumanji: Welcome to the Jungle made a 490% return whilst Independence Day: Resurgence made just 29% profit. This report will explore how profitable sequels are, accounting for this variability between collections. It will also seek to understand what factors make a successful series.

Data Set and Enrichment

The data for analysis has come from three movie sources:

- The Movie Database (TMDB) - https://www.themoviedb.org
- Open Movie Database (OMDB) - https://www.omdbapi.com
- The Numbers - https://www.the-numbers.com

The data spans movie history from 1960 to 2019, so inflation needs to be considered to be able to compare monetary values. Using data from the US Bureau of Labour Statistics which supplies the Consumer Price Index (CPI) rate for each year from 1947 onwards, all revenue and budget amounts were converted to 2018 dollars using the formula below

$$\mathbf{2018}\ \boldsymbol{value} = \boldsymbol{Original\ value} \times \frac{\mathbf{2018}\ \boldsymbol{CPI\ rate} - \boldsymbol{Movie\ year\ CPI\ rate}}{\boldsymbol{Movie\ year\ CPI\ rate}}$$

To measure box office success, Return on Investment (ROI) has been used which is stated as:

$$\boldsymbol{ROI} = \frac{\boldsymbol{Revenue} - \boldsymbol{Budget}}{\boldsymbol{Budget}}$$

Within the movie industry, a movie budget is quoted as the cost of making the movie, but does not include marketing, royalties, etc. According to (Velikovsky 2012), 45% of the cost of a movie is hidden, so for analysis in this report, the budget has been inflated to be comparable to the revenue earned as thus:

$$\boldsymbol{Adjusted\ Budget} = \frac{\boldsymbol{Budget}}{\mathbf{0.55}}$$

All figures quoted in this report are in US dollars and have been adjusted above for inflation, and as per the adjusted budget calculation. For analysis of sequels, 348 collections have been used, containing 982 movies.

Sequels and Profitability

From the analysis done previously, collections usually begin with a very successful first movie which then turns into multiple sequels that perform better than original content films but not as well as the first movie. Figure 1 shows the ROI per sequel number within a series. Since most chains are 4 movies or less, series that span 5 or more movies have been combined. The median ROI for the first movie is 190% with subsequent movies falling to about 100% profitability. For original content, the median profitability is -1.5%.

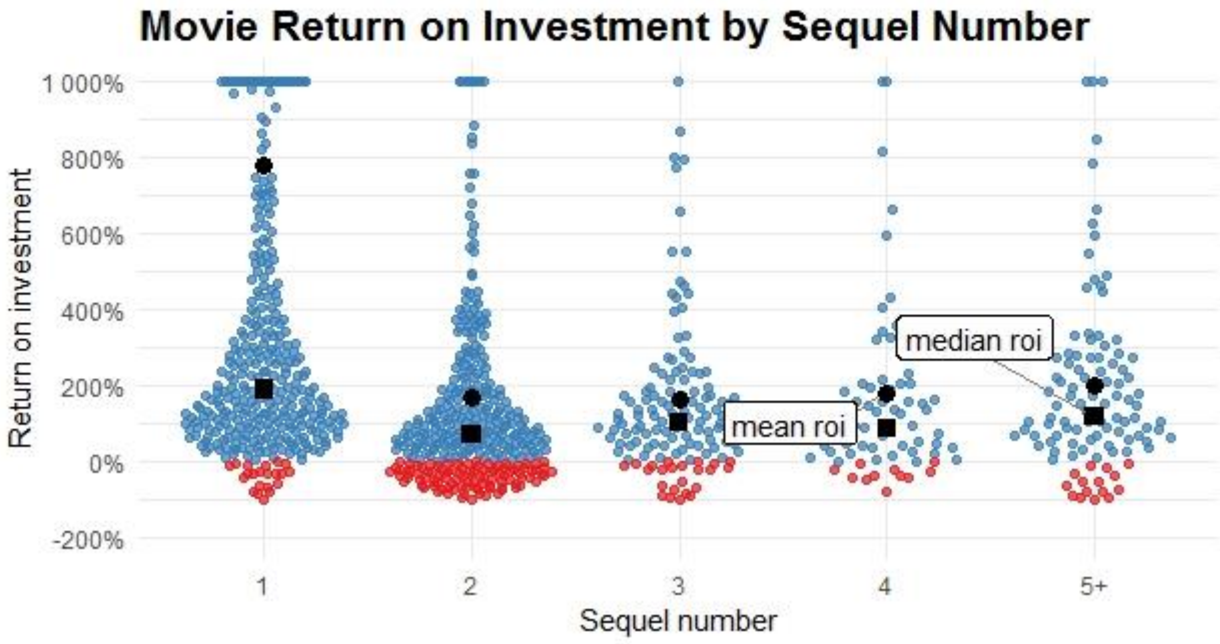

Figure 1. Return on Investment for Movies

However, when grouping movies by their collection and excluding the very profitable first movie, not all series are so successful. Figure 2 shows the ROI for each series by the number of movies in the collection. The ROI is calculated without the original movie that started the chain. Here, ROI for collections that only span 2 movies show a marked drop, with a median ROI of 25% compared to over 100% for longer chains.

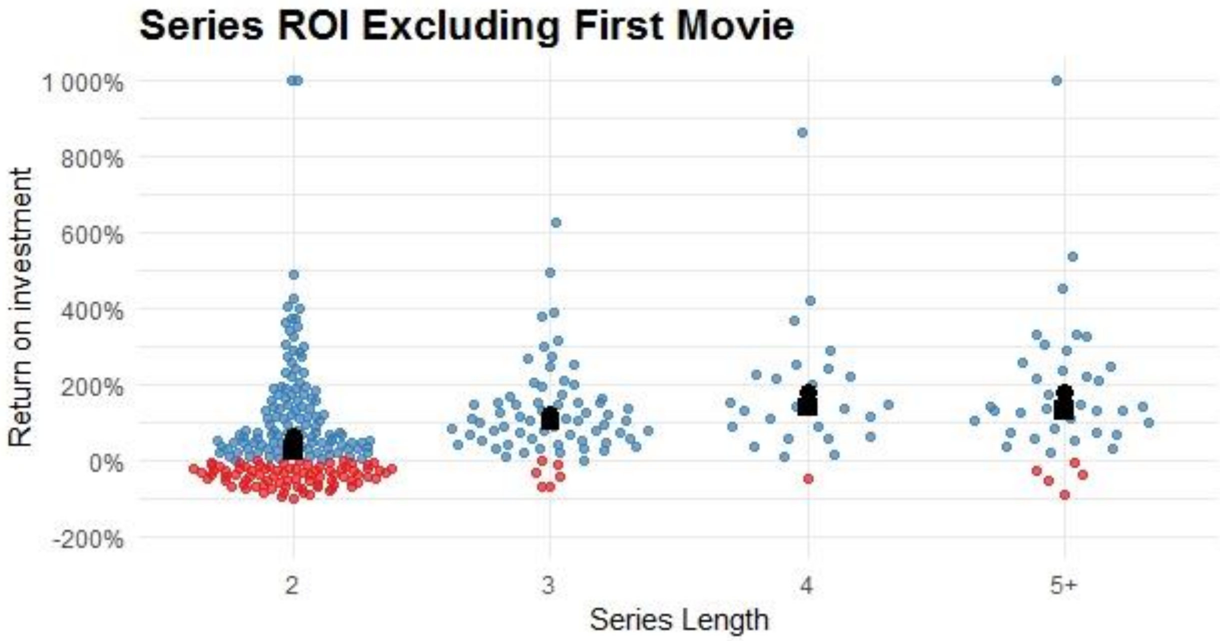

Figure 2. Series ROI Excluding First Movie

Types of Sequels

JT Velikovsky contends that there are 2 main types of sequels (Velikovsky 2012). Firstly, there are those that are repeats of the first movie, generally made because the original was very successful and so warranted a second outing. Figure 3 shows series that had incredible success from a low budget. The points and line in red are the adjusted budget and the blue points and line are the revenue earned. Typically, budgets increase as the series continues and revenue decreases.

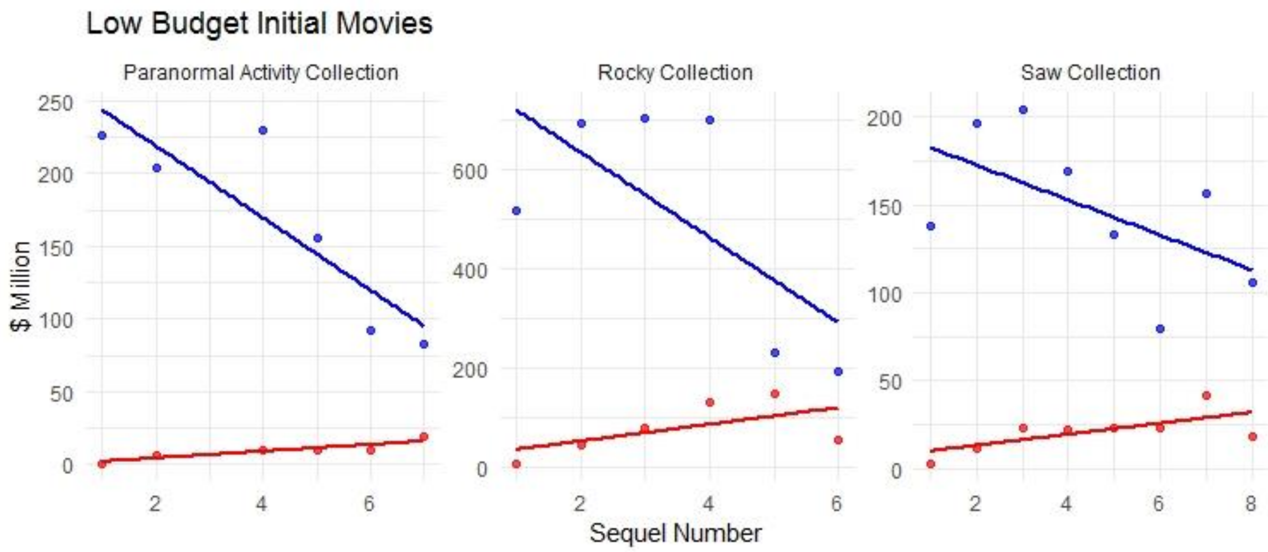

Figure 3. Low Budget Initial Movies

There are also well funded 'blockbuster' movies, shown in figure 4. Here, the general trend is for revenue to increase, at least for a while. The budgets continue to at least match the original film.

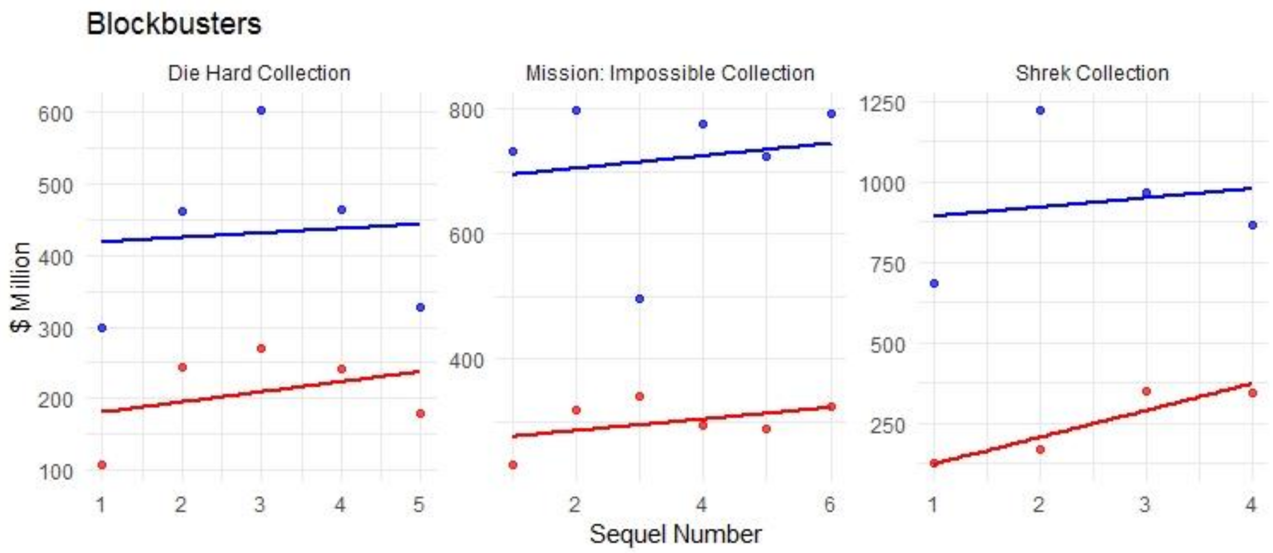

Figure 4. Blockbuster series

Finally, there are sagas or multi-movie stories where each movie progresses an overall story. They have a finite end (although often seem to be extended at the end, such as book 7 in Harry Potter becoming 2 films). They are also generally well-funded throughout and can maintain their audience. My domain expertise has been used to define the collections that are sagas. Appendix A – Sagas shows the full list of collections defined as sagas.

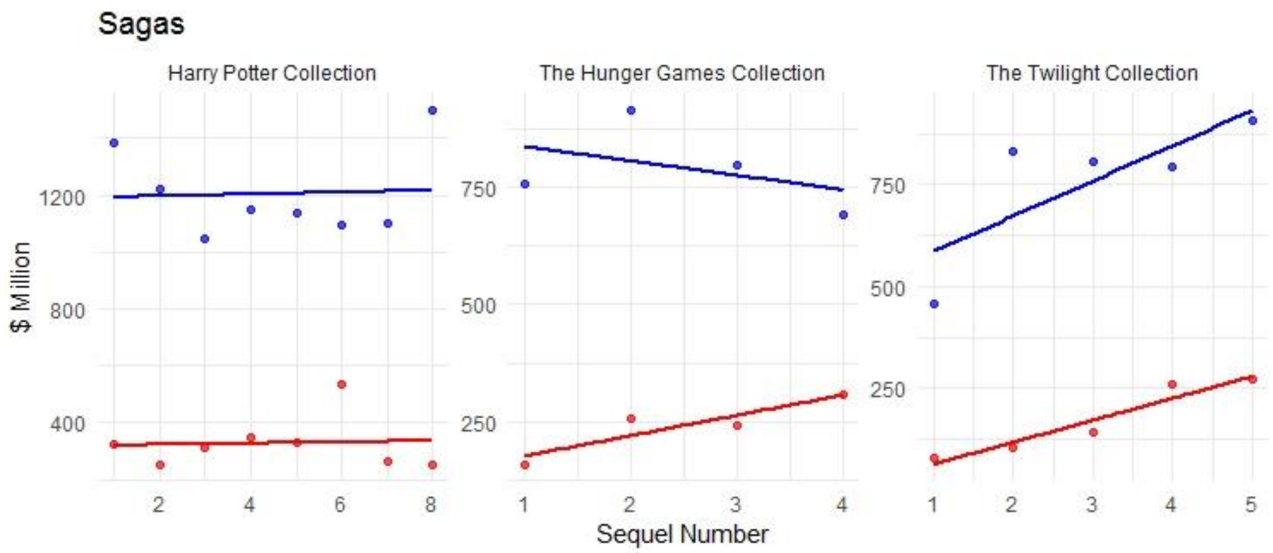

Figure 5. Sagas

There is much variation between collections, with differing slopes and intercepts for each chain. Some series continue to capture the imagination of the cinema going public while others become staid and decline. Paranormal Activity is one of the most successful movies of all time, returning a profit of over $220 million for an adjusted budget of under $1 million. The latest instalment still made $60 million in profit but could not match the stellar first-time performance.

Multi-Level Analysis

Multi-level analysis is a technique that deals with hierarchical data, drawing conclusions about groups and subject-to-subject variability. Here, each movie within a collection share similar traits. They are within the same genre, come from the same source, usually share the same actors and often the same directors. So, there is covariance between the movies within a sequel chain.

Thus, for analyzing collections of movies, a multi-level model will be used. Each collection will be a group (or subject) with each movie being a point in time observation about the series. In this way the variance between collections can be accounted for. To build the model, various combinations were tried and then compared using Analysis of Variance (ANOVA). Appendix B – ANOVA Model Comparison contains more details. The models were compared using likelihood ratio tests and the final model to predict revenue is defined as:

$$\boldsymbol{Revenue \sim Budget + Sequel\ Number + Saga + (1 + Sequel\ Number \mid Collection)}$$

Budget and Revenue are in millions of US dollars. The random effects being catered for is the sequel within the collection and both slope and intercept vary. The results of this model are shown in figure 6 below.

```
Linear mixed model fit by REML ['lmerMod']
Formula: I(revenue.final/1000000) ~ I(budget.adjusted/1000000) +
collection.sequel.number + Sequel.Source +
(collection.sequel.number | tmdb.collection.name)
   Data: movies.coll

REML criterion at convergence: 13447

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.050 -0.335 -0.102  0.207  8.088

Random effects:
 Groups               Name                     Variance Std.Dev. Corr
 tmdb.collection.name (Intercept)              53859    232
                      collection.sequel.number  4620     68      -0.68
 Residual                                      29844    173
Number of obs: 982, groups:  tmdb.collection.name, 348

Fixed effects:
                           Estimate Std. Error t value
(Intercept)                192.8176    19.7577    9.76
I(budget.adjusted/1000000)   1.5325     0.0816   18.78
collection.sequel.number   -54.1094     7.7317   -7.00
Sequel.SourceSaga          361.4601    54.6027    6.62

Correlation of Fixed Effects:
            (Intr) I(./10 cllc..
I(./1000000 -0.361
cllctn.sql. -0.662 -0.169
Sequl.SrcSg -0.037 -0.144 -0.017
```

Figure 6. Model Summary

Fixed Effects

From the model above, a collection's revenue will be defined by the following parameters:

- Starting intercept of $192.82M.
- If it is a saga, the intercept increases by $361.46 Million
- Then for each $1 million of budget the revenue returned will be $1.53 Million.
- Each movie sequel will then lose $54.11 Million. Note that the first movie will also lose this amount, so the starting intercept is really $138.71 Million

Running a similar model for standalone movies, predicting revenue based on budget, the intercept is $11.12 million with $1.10 million added for each $1 million of adjusted budget. For the median original content movie (having an adjusted budget of $63.83 million), the expected revenue is $81.53 million. Table 1 shows the expected revenue for a series with this same budget, based on the fixed effects.

1st Movie	2nd Movie	3rd Movie	4th Movie	5th Movie
236.37	182.26	128.15	74.04	19.93

Table 1 Expected Revenue for Median Budget New Series ($ Million)

By the 4th instalment the expected revenue has dropped to just $74 million. So, whilst sequels typically start higher than non-sequels, they will become unprofitable unless the series is embraced by the cinema going public.

Being part of a saga is also a large boost to a franchise. The starting revenue doubles, adding $361 Million to each movie in the chain. This reinforces the idea that movies that progress a storyline will do better than just repeating a formula.

Random Effects

The multi-level model provides variability for each group (series in this case). The model chosen has random variability for both the intercept and slope of each series. The mean of each is 0 and figures 7 and 8 show the distribution of the intercept and slope (i.e., Sequel number) respectively. Both show the most common values are relatively close to 0 but then spread out with the Intercept having a longer tail positively and Sequel number having a long tail negatively. The standard deviation for both elements is high being 232 for the intercept and 68 for sequels. This indicates the random effect caters for much of the variation between collections.

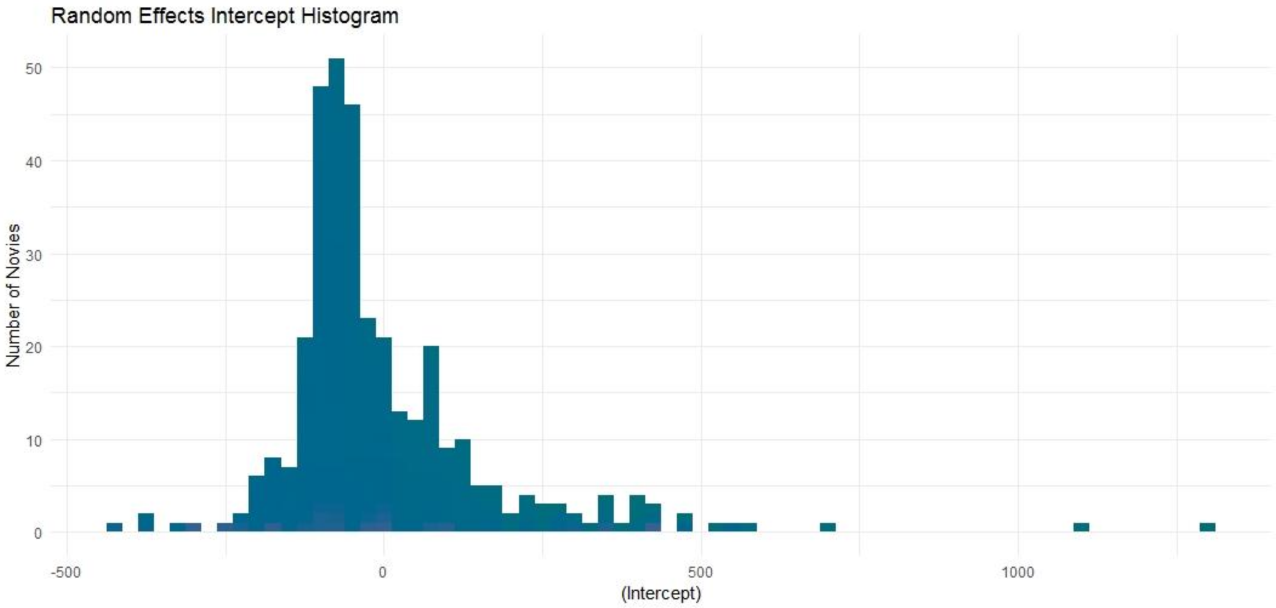

Figure 7. Random Effects Intercept Histogram

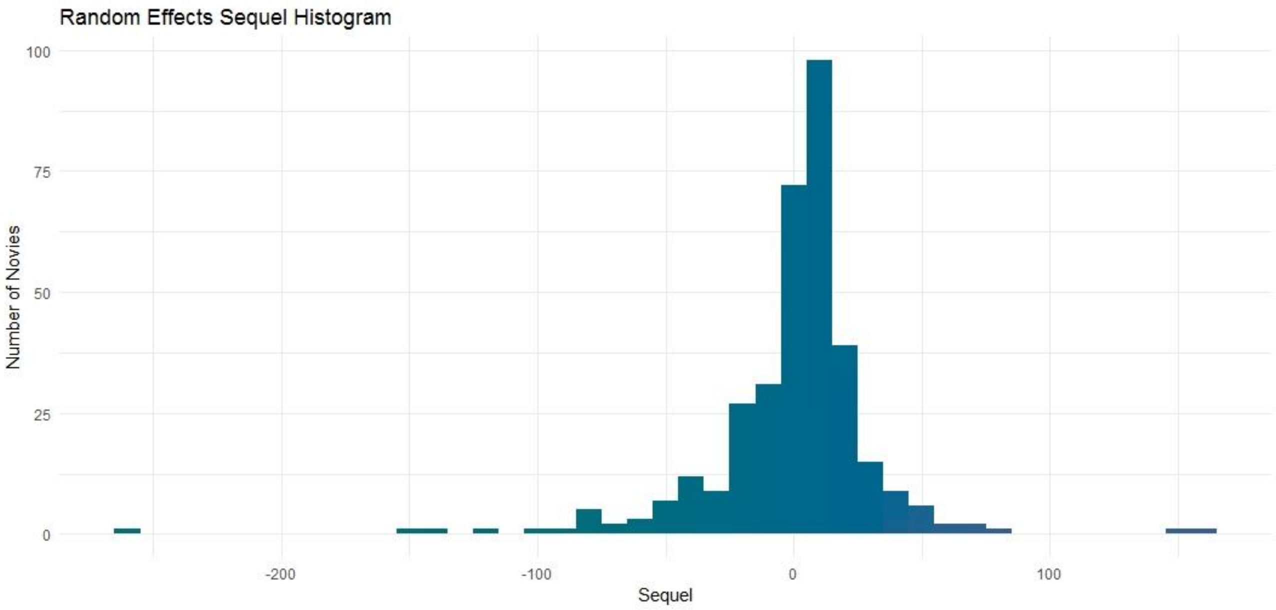

Figure 8. Random Effects Sequel Histogram

There is a strong negative correlation between intercept and sequel number for random effects, being -0.68 as shown in figure 6. This makes sense in that series that start very strongly (and thus have a high intercept) will have a more negative slope for sequels as later movies in the chain cannot match the original. Those that do not start strongly but gain commissions for further sequels tend to do better as the collection continues. It could be thought of as an example of regression to the mean.

Tables 2 and 3 below show the 10 movies with the lowest and highest intercepts. This highlights that usually when the intercept is positive, the sequel slope is negative and vice-versa. The exception is The Avengers Collection with both a high intercept and a high sequel. This demonstrates just how phenomenal the Marvel Cinematic Universe has been and is looked at further in Appendix C – The Marvel Effect. Another series with both positive intercept and sequel is James Bond and is discussed in Appendix D - James Bond: The Never-ending Story.

Collection	Intercept	Sequel
The Girl - Millennium Collection	-420.56	37.93
Divergent Collection	-363.20	4.02
Kill Bill Collection	-362.74	34.89
The Bourne Collection	-325.44	13.77
The Fast and the Furious Collection	-299.92	146.20
The Dark Knight Collection	-252.38	64.56
Blade Runner Collection	-228.41	16.33
The Chronicles of Riddick Collection	-220.79	41.32
G.I. Joe (Live-Action) Collection	-212.44	34.07
The Maze Runner Collection	-204.42	2.178

Table 2 Lowest Random Effects Intercept

Collection	*Intercept*	*Sequel*
Shrek Collection	427.78	-13.10
The Avengers Collection	429.97	162.72
Home Alone Collection	478.30	-97.49
James Bond Collection	482.48	29.55
Finding Nemo Collection	519.70	-70.08
Jurassic Park Collection	551.62	33.43
Indiana Jones Collection	570.97	-44.81
The Jungle Book Collection	691.82	-138.14
The Jaws Collection	1104.79	-257.24
Star Wars Collection	1312.37	-124.71

Table 3 Highest Random Effects Intercept

The table also highlights the evolution of the Fast and the Furious franchise. From a successful average budget movie, the series waned over the following 2 sequels. It then rebooted with the original cast returning to great success, collecting over $1 billion each on the most recent 2 movies. The series has kept fresh by morphing from a streetcar racing genre to heist and then on to a spy film. From the above data, whilst the intercept is negative, the sequel slope of $146 Million shows how successful the franchise has become.

The model can also be used to predict upcoming movies in a collection. The expected revenue for the new offerings in the James Bond, Star Wars, Toy Story and Men in Black franchises are below, assuming the same budget as the previous movie.

Collection	*Sequel*	*Budget (Millions)*	*Revenue Previous Movie ($ Million)*	*Predicted Revenue ($ Million)*	*Revenue Change (%)*
Star Wars Collection	9	590.4	1,364.91	1,162.1	-14.86%
Toy Story Collection	4	418.7	1,228.56	939.4	-23.54%
James Bond Collection	26	577.9	933.09	922.3	-1.16%
Men In Black Collection	4	427.5	682.51	629.4	-7.78%

Table 4 Predicted Revenue for Upcoming Films

Conclusion

The original analysis found that franchises are indeed a ‘safer bet’ than standalone movies. However, upon deeper analysis, it has been found that not all sequels are equal. Nearly 60% of collections analyzed span only 2 movies, with a median profit of just 25% on the 2nd movie in the series, compared to a median profit of 190% on the original film. If the collection can survive beyond this to further sequels, then the profitability is typically 100% or more.

When considering the final movie in 2-movie series, over 40% are unprofitable compared to approximately 50% of original content movies failing to make a profit. Profitability increases significantly for multi-sequel chains with only 14% of subsequent movies being unprofitable when the series length is 3 movies or more.

Using multi-level analysis, the trend discovered is for a sequel to do worse than its predecessor by approximately $55 million. Drilling down to individual series, there is great variability with a few being able to buck the decline and improve as the series goes on. These collections successfully navigate the plot line of being familiar but different enough to stay fresh (Velikovsky 2012). Multi-movie stories that progress rather than repeat are more successful at maintaining audience engagement. The powerhouse franchise for this is Marvel that has been able to weave several collections into a hugely profitable continuum.

Reflection

Upon reflection, this report has led to greater understanding about the profitability of franchises and sequels. Particularly for 2-movie series, there is almost as much uncertainty about making a sequel as creating a standalone movie. This is somewhat surprising given the boost we found that franchise movies typically receive. However, much of this lift was found to be attributable to the original movie, rather than ongoing instalments. Sequels are still ‘safer bets’ in the movie industry. If the upcoming Disney offerings are any indication, they will continue to make up at least 25% of all movies made and will persist in dominating the box office in the coming years.

References

Quinn, K. (2019). "Disney reveals the future of cinema and for grown-ups it's not good." from https://www.smh.com.au/entertainment/movies/disney-reveals-the-future-of-cinema-and-for-grown-ups-it-s-not-good-20190508-p51lbr.html.

Snow, S. (2016). "Why Hollywood Makes More Sequels Every Year (Even Though We Like Them Less)." from https://www.linkedin.com/pulse/why-hollywood-makes-more-sequels-every-year-even-though-shane-snow/.

Velikovsky, J. (2012). "On Feature Films and RoI (return-on-investment)." from https://storyality.wordpress.com/2012/12/17/storyality-24-on-feature-films-and-roi-return-on-investment/.

Winter, B. (2014). "A very basic tutorial for performing linear mixed effects analyses (Tutorial 2)." from http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf.

Appendix A – Sagas

The table bellow shows what has been classified as multi-movie stories, using my own domain knowledge. The view was taken that the series should be planned from the start so, The Jack Ryan series for example has not been included as this was adaptations of books rather than a planned-out series – it was deemed that the series continued until the movies lost the publics' interest, not because the source material was exhausted.

Collection	***Series Length***	***Year of 1st Movie***	***Series ROI***
Star Wars Collection	8	1977	4.48
Divergent Collection	3	2014	0.38
Fifty Shades Collection	3	2015	3.87
Harry Potter Collection	8	2001	2.68
Kill Bill Collection	2	2003	0.67
The Bourne Collection	5	2002	0.74
The Dark Knight Collection	3	2005	1.20
The Girl - Millennium Collection	2	2011	0.06
The Godfather Collection	3	1972	4.86
The Hobbit Collection	3	2012	1.15
The Hunger Games Collection	4	2012	2.28
The Lord of the Rings Collection	3	2001	4.39
The Maze Runner Collection	3	2014	2.35
The Twilight Collection	5	2008	3.43

Table 5 Collection marked as Sagas

Most sagas have been very successful. As seen before with Harry Potter (figure 5), it was very consistent with its revenue in each outing.

Appendix B – ANOVA Model Comparison

To evaluate the best model for describing this system, several different combinations were tried. This included varying just the slope, just the intercept and both the intercept and slope. The models were then compared using ANOVA. The results are shown in figure below.

```
Data: movies.coll
Models:
model.lm1: I(revenue.final/1000000) ~ I(budget.adjusted/1000000)
model.lm: I(revenue.final/1000000) ~ I(budget.adjusted/1000000) + collection.sequel.number +
model.lm:     Sequel.Source
model1: I(revenue.final/1000000) ~ I(budget.adjusted/1000000) + collection.sequel.number +
model1:     Sequel.Source + (0 + collection.sequel.number | tmdb.collection.name)
model2: I(revenue.final/1000000) ~ I(budget.adjusted/1000000) + collection.sequel.number +
model2:     Sequel.Source + (1 | tmdb.collection.name)
model3: I(revenue.final/1000000) ~ I(budget.adjusted/1000000) + collection.sequel.number +
model3:     (collection.sequel.number | tmdb.collection.name)
model.full: I(revenue.final/1000000) ~ I(budget.adjusted/1000000) + collection.sequel.number +

model.full:     Sequel.Source + (collection.sequel.number | tmdb.collection.name)
           Df   AIC   BIC logLik deviance Chisq Chi Df            Pr(>Chisq)
model.lm1   3 13784 13799  -6889    13778
model.lm    5 13700 13725  -6845    13690 87.61      2 < 0.0000000000000002 ***
model1      6 13624 13653  -6806    13612 78.49      1 < 0.0000000000000002 ***
model2      6 13525 13555  -6757    13513 98.64      0 < 0.0000000000000002 ***
model3      7 13522 13556  -6754    13508  5.04      1                0.025 *
model.full  8 13482 13521  -6733    13466 41.82      1         0.0000000001 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 9. Results of ANOVA comparison

The Likelihood Ratio Test is used to compare models, which tests for significant differences between the models as factors are added or removed (Winter 2014). The above ranks the models in order of significance so model.full gives the best results. This model is defined as:

$$\boldsymbol{Revenue \sim Budget + Sequel\ Number + Saga + (1 + Sequel\ Number \mid Collection}$$

Appendix C – The Marvel Effect

The Marvel Cinematic Universe has been a phenomenon, producing 21 movies in the last 10 years (and still more to come). Figure 10 shows the release dates for these movies and the return on investment for each. The movies overlap and characters crossover between movies, almost making one continuous saga. There have been 10 movies released (with another spider-man due in July 2019) in just over 3 years since May 2016. Compare this to the median time between sequels of just under 3 years means the market has been saturated with Marvel.

Only 3 movies have made a profit of less than 100% - the start of the Iron Man, Thor and Captain America collections with the worst being Captain America: The First Avenger which still made a respectable 45% ROI.

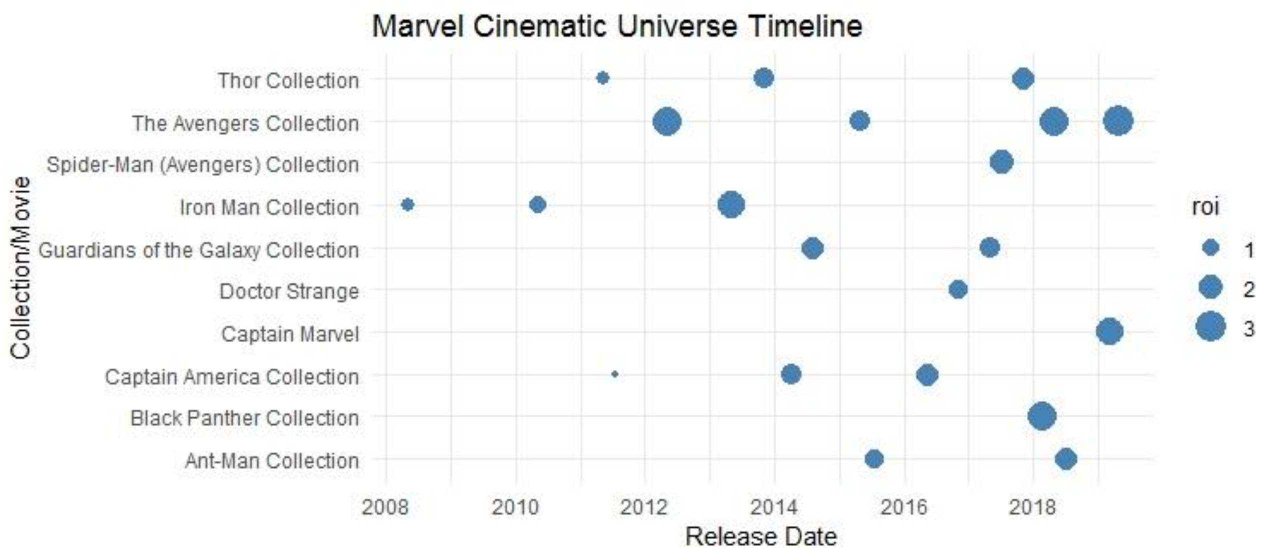

Figure 10. Marvel Cinematic Universe Timeline

Looking at each series individually gives the breakdown below in figure 11. Every series increased revenue whilst keeping budgets contained.

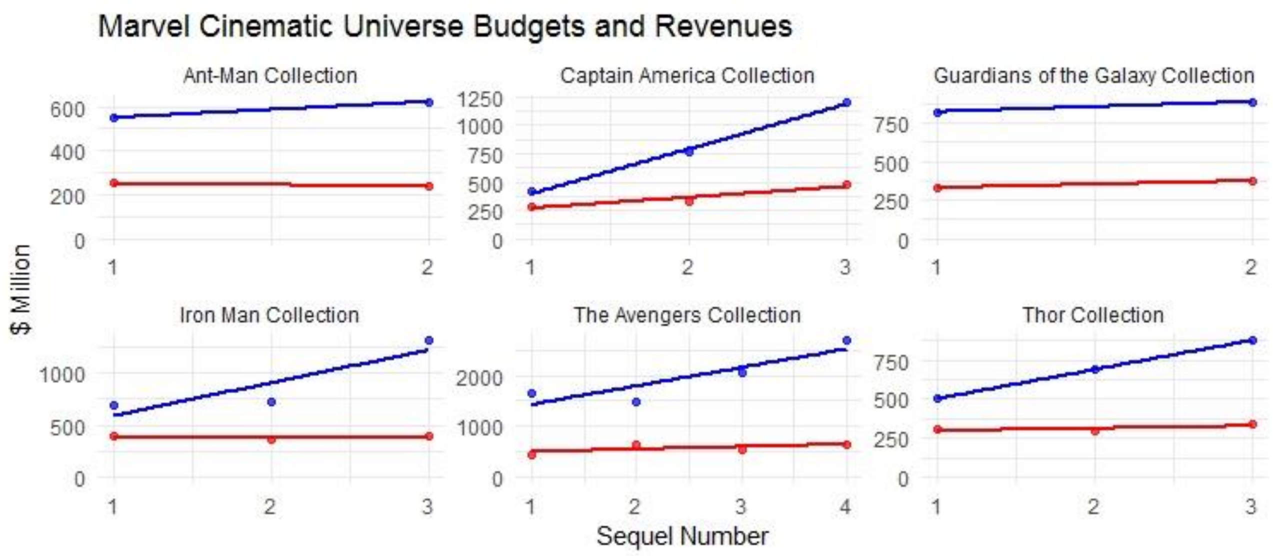

Figure 11. Marvel Cinematic Universe Budgets and Revenues

Table 6 shows the Random Effects intercepts and sequel slopes for the Marvel series. The Avengers has been wildly successful, starting with a high intercept and also having a positive slope for sequels, which is unusual.

Collection	***Intercept***	***Sequel***
Ant-Man Collection	62.93	2.40
Captain America Collection	-25.41	66.92
Guardians of the Galaxy Collection	156.04	-11.83
Iron Man Collection	9.13	80.02
The Avengers Collection	429.97	162.72
Thor Collection	-2.24	47.24

Table 6 Random Effects for Marvel Collections

The movie franchise has culminated with the wildly successful Avengers: Endgame released this year and taking over $2.6 Billion in revenue. This box office dominance may be hard to emulate as the series now needs to almost start again. They do have a formula that has worked so far, though, so would be aware of the path to take..

Appendix D - James Bond: The Never-ending Story

No discussion of sequels would be complete without referencing James Bond. Started over 50 years ago, the franchise will soon be releasing its 25th instalment. The next longest series is Friday the 13th with 12 movies. This longevity could be due to the series treading the line between being familiar enough to engage the audience but changing enough to maintain interest. Within the framework of super-spy versus super-villain, the plots move with the time it is in.

Figure 12 shows the revenue (in blue) and budget (in red) for each of the movies in the chain (note, the source data has included Never Say Never Again as one of the bond movies even though it is not officially part of the collection). The revenue has gone through peaks and troughs, whilst the budget is generally increasing. Slides in revenue were able to be arrested by changing the lead actor, thus rebooting the collection and essentially restarting the sequel count.

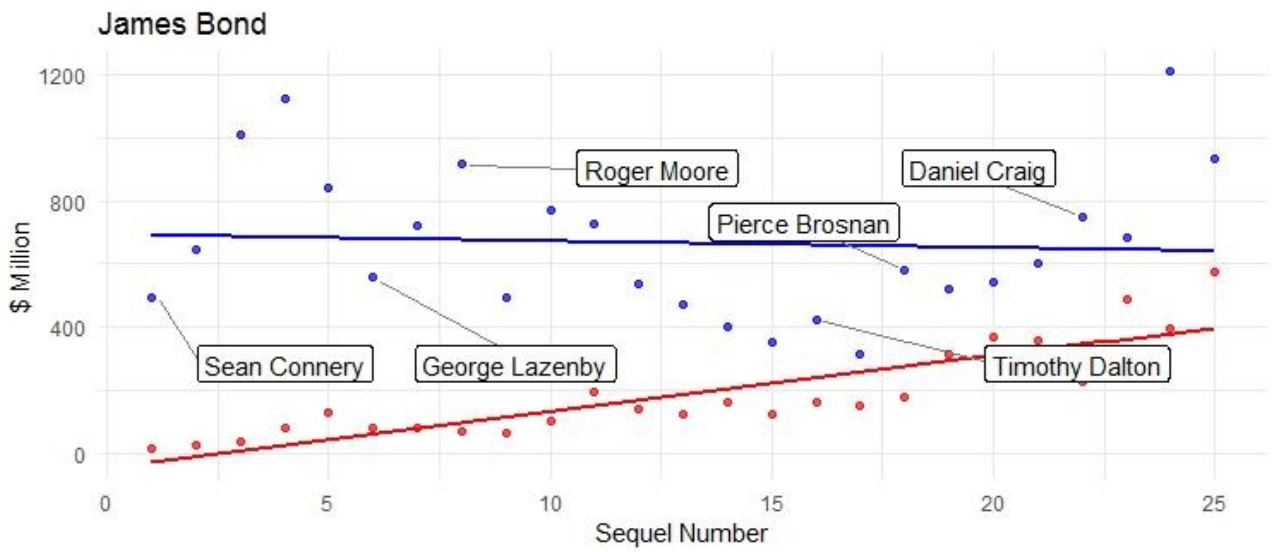

Figure 12 Bond Movies

By all accounts, the movie series has been incredibly successful. The series return on investment stands at 2.55 and inspecting the random effects parameters, the intercept is $482.48 million with a positive sequel slope of $29.55 million, so another indicator of just how lucrative and unique the collection has been. With talk that Daniel Craig is on his last Bond film, the series is set for another refresh.