Pixar Films Data Analysis Project Report

Business Task
The business task is to provide recommendations for Pixar, on how they should produce and release films in the modern era, to ensure survival and profitability in the face of streaming services pulling customers away from traditional cinemas. The business task is also to advise Pixar on how to act in the event of another global emergency, similar to Covid-19, to ensure its survival.

Data Source
The data has been sourced from Maven Analytics' website (https://mavenanalytics.io/data-playground).

Data Organisation
The data has been supplied in 7 CSV files. Relevant data has been combined into one Excel spreadsheet for ease of access and so that analysis & visualisation can take place all in one place.

Bias / Credibility of the Data
The data has been supplied by a trustworthy data agency, Maven. Maven have collated the data themselves from public sources of information online. The data is published by Pixar, (or distributor Disney) after each film is released, so the data originated from the company which the data is about, so can be considered reliable. As an additional layer of safety, in order to ensure that there have been no mistakes during data collection, I have also spot checked the data as part of the data verification process; the data has been found to match the original source of the data. I can therefore conclude that the data is reliable and original enough to use for this project. The data can also be cited, as described above.

The data can be considered comprehensive and current, because it covers a complete range of time, i.e. from 1995 when the first Pixar film was released, up to Pixar's most recent film release, in 2024, and no films have been omitted.

Licence
Maven have licensed the data for use by individuals for data analysis practice purposes. See their website (noted above) for full details.

Privacy
All the data is in the public domain already, and was public knowledge before it was collated by Maven, and presented as a dataset. So there are no privacy concerns for this project.

Security
As the data is not confidential, there are no security concerns. For the record, the data will be stored locally on my PC as well as a backup being stored on Google Drive.

Integrity
There are no reasons to doubt the integrity of the data: as we know, the data has been spot checked against the original source, and has been found to be true. The data is public knowledge and there is no reason why the data would lack integrity. The only exception is the Cinema Rating

field, within the public_response.csv dataset, this contains too many null values to be used reliably so this data has been dropped from the analysis.

Timeframe
I will be looking at data from Pixar's first film release in 1995 (Toy Story), to its most recent release* in 2024 (Inside Out 2).
*At the time of writing. No data after 2024 will be considered.

Process

Tools
I have used Microsoft Excel for this project, as that provided all the tools required to clean, analyse, and visualise the data. Excel can easily handle this relatively small amount of data.

Cleaning
The cleaning process has involved the following (but was not limited to):

- Convert CSV files to XLSX so that formatting, formulas and visualisations can be saved.
- Combine relevant data from all CSVs into one spreadsheet.
- Removed Cinema Ratings, as there was too much missing information for that to be reliable.
- Hide the number of ratings from the consumer review sites, as this was found to be not helpful to this analysis.
- Rename and reformat headings.
- Freeze panes: top row and leftmost column for easier viewing.
- Format monetary cells as dollar values, including commas after every third digit, for easier reading.
- All of the above is for the purpose of making the data more user-friendly and easier to analyse all in one place.
- Box Office figures validation: US & Canada and Rest of the World figures, have been summed and are confirmed to match that of the global box office figures, so we can be sure that these figures have already been accurately calculated at source.
- The film "Luca" was missing data for its budget. So I have researched this information myself, and added this data to the table.
- Convert the CSV "academy.csv" (which contains data of which Academy Awards each film won, or was nominated for) into two columns – 1, denominating the number of Academy Awards won by each film; and 2 – denominating the number of Academy Awards each film was nominated for. Added the "won special achievement" into the "won" column. Discarded "not eligible" and "award not yet created" values, as these don't add any real value to this analysis.
- Add the director and/or co-director(s) of each film (from the file "pixar_people.csv" dataset), to a column in the main analysis spreadsheet, providing this data as an array in alphabetical order. Directors and co-directors are treated as equal for the purpose of this column. The data relating to people for the film "Brave" is believed to be incorrect, as this shows Mark Andrews to have done every job role. To fix this, I have researched this, and added Brenda Chapman as director, alongside Mark Andrews. For the purpose of this analysis we are not interested in other roles so these can be ignored.
- All other columns have been checked for outliers, and null or erroneous values. While outliers do exist, I consider the data in the columns included in the final spreadsheet to be plausible.

Analyse

Organisation

I have organised the data into five main categories:

- General (Release Date, Release Day, Release Month, Time Since Last Release, Duration, Rating)
- Finance – Raw Data (Budget, Box Office (US & Canada), Box Office (Rest of World), Box Office (Worldwide)).
- Finance – Analysed (Budget : Box Office Ratio, Box Office : Budget Ratio, US & Canada Sales as a Percentage of Worldwide Sales, Rest of World Sales as a Percentage of Worldwide Sales).
- Audience Reception (Rotten Tomatoes Score, MetaCritic Score, IMDB Score).
- Academy Awards (number of awards won, number of awards nominated for).
- Directors (directors and co-directors combined into an alphabetical array)

New Field Calculations

I have calculated the following for each film, as part of the analysis.

- Profit made (worldwide), (calculated by subtracting Budget from Global Box Office Sales.) (Also added conditional formatting to highlight the loss-making films.)
- Ratios of Budget to Global Box Office Sales, and vice versa. (calculated by dividing the budget by Global Box Office, and vice versa. Rounded to two decimal places.)
- US & Canada sales as a proportion of worldwide sales. (calculated by dividing US & Canada Box Office Sales by Global Box Office Sales, and formatting as a percentage, with up to two decimal places.)
- Same as above but for Non-US & Canada (Rest of World) Sales.
- Time since last release (in days), (calculated using the DATEDIF function. For the first Pixar film (Toy Story), I have marked this field as N/A as there is no previous film to compare this to.)
- Release Day, (calculated using the TEXT function within Excel to convert the Release Date column to day of the week.) I have also applied conditional formatting to this column to highlight different days of the week, from lightest to darkest, representing earliest to latest in the week.
- Release month, (calculated using the TEXT function within Excel to extract the Month from the Release Date.) I have also applied conditional formatting to this column to highlight different Months, from lightest to darkest, representing earliest to latest in the year.
- (The conditional formatting applied to Release Day and Release Month makes popular day / month combinations stand out more clearly, (explained later), so this also serves as a useful visualisation in itself, which also shows trends over time.)

Aggregations

I have calculated the following aggregations for each field where applicable:

- Total
- Average (mean)
- Minimum
- Maximum
- Name of film for minimum (using Vlookup) *
- Name of film for maximum (using Vlookup) *
- Median (for Time Since Last Release, only)

* where several films are tied for maximum or minimum aggregation, the first film on the list will take precedence.

Pivot Tables in place of aggregations
I have also created small pivot tables for the Release Day, Release Month, and Rating columns, as these are not applicable to the above aggregations. I have marked these with a bold border for easier viewing.

Directors & Awards Analysis – separate table.
Separately to the main spreadsheet, I have created two (hidden) sheets one to show the directors with the most awards, and one to show the directors with the most awards per film. Each sheet contains a table and subsequent pivot table. These tables are based on Director, rather than film, which is why these are separate to the main spreadsheet. For this analysis, I have treated directors and co-directors as equal. The data available is specifically for Academy Awards, no other awards have been considered. I have created the following fields:

- Total number of awards won;
- Total number of awards nominated for;
- Number of films directed by this director;
- Average number of awards won per film;
- Average number of award nominations per film.

I have therefore been able to create two charts to show:

- Directors with most award wins (also showing number of nominations achieved.)
- Directors with most award wins per film. (Also showing the number of nominations achieved per film.)

Summary of Analysis Findings
<u>Budgets and Box Office Sales</u>
Budgets
Film budgets increased dramatically between 1995 and 2010, but relatively steadily between each film, from a low of $30 Million for Toy Story in 1995, to a peak of $200 Million for Toy Story 3 in 2010. Post 2010, budgets varied slightly, hitting a low of $150 Million for Soul in 2020. While budgets never exceeded the $200 Million record set by Toy Story 3, the record has been frequently matched since 2010. The average budget is $161.6 Million.

Box Office Sales
Global Box Office Sales have varied drastically over the years. From a low of $21.8 Million for Turning Red, to a high of $1.698 Billion for Inside Out 2, with an average of $608.6 Million.

Notably Toy Story in 1995 produced a global revenue of $394.4 Million; at the time, this was far higher than expected, giving Pixar its kickstart as a highly regarded animation studio. Despite this, revenues of this level would now be considered low. Toy Story was the world's first animated feature length film, and many expected the film to flop, which is why this success was so notable. Toy Story also represented a huge Box Office : Budget ratio of 13.15 : 1, by far the highest of all Pixar films, due to Toy Story's shoe string budget of just $30 Million.

Post Toy Story, despite huge variation from film to film, average revenues slowly increased as can be seen in the chart, until the Covid-19 pandemic hit in early 2020. The pandemic caused

revenues to drop dramatically, resulting in Pixar's first loss making Films, Onward, Soul, Luca and Turning Red. Following this, *Lightyear*, (2022) only made a small profit of $26.4 Million. Post 2022, Pixar's films' revenue have increased dramatically again, resulting in a return to profitability. In fact Pixar's most recent release, *Inside Out 2* (2024) was the highest grossing film in Pixar's history, with a global revenue of some $1.698 Billion.

Unsurprisingly, I found there to be a trend of positive correlation between Budget and Box Office figures, however not as strong correlation as might be expected, as can be seen from the scatter chart.

Box Office : Budget Ratio

Box Office : Budget ratio has varied drastically over the years, but averages 4.18 : 1. The worst figures were on the loss making films of the Covid-19 Pandemic, specifically Turning Red, with a ratio of just 0.12 : 1. The highest Box Office : Budget ratio was on Toy Story, with 13.15 : 1, as previously noted. This was mainly because it was made on such a shoe string budget, yet still managed to achieve huge success. Even though this ratio has never been reproduced post Toy Story, profits have still (mostly) been higher than that of Toy Story, due to even higher Box Office figures.

Profit

Profit loosely follows the trend of Global Box Office Figures, usually because of the relatively large Box Office : Budget ratio usually enjoyed by Pixar films. In simple terms, even a "large" budget of $200 Million doesn't make a huge dent into the gigantic Box Office figures. That said, losses during the Covid-19 pandemic amounted to a sizable $338 Million; but these losses have since been recouped by later films, not least Inside Out 2. The average profit per film is $447.02 Million. The lowest was a loss of $153.18 Million (Turning Red, 2022), and the highest was a profit of $1.49 Billion (Inside Out 2, 2024)

Covid-19 Pandemic

During the pandemic some newly released Pixar films were also made available on the streaming service, *Disney+*. This will have undoubtedly had an impact on Box Office figures for these films. Many audience members would have watched the latest Pixar film at home for a small monthly fee that covers a whole family to view the film, as well as a host of other content. Rather than traditionally, a cinema ticket would need to be purchased for each person individually. This also came at a time when the number of people signing up to streaming services rocketed anyway, as many other popular activities had been made impossible due to lockdowns, resulting in many people having the extra disposable income required to sign up for Disney+. So it's likely that during the pandemic, many people would have watched the latest Pixar film without paying anything more than what they would normally do. Disney+ revenue figures are not part of this analysis, we are only considering cinema box office figures, when we reference revenues here.

Post Covid-19 Pandemic

Since *Elemental* (2023), Pixar films have continued to be made available on Disney+, but only around 3 months after the cinema release, which appears to have resulted in a return to high revenues at cinema Box Offices.

Soul, Turning Red, and Luca were subsequently re-released in cinemas in 2024, however that data is not part of this analysis. This only covers the original release of each film.

Sales from US & Canada

The proportion of sales from the US & Canada market, compared to the rest of the world, varies quite a lot from film to film. Ignoring the Covid-19 pandemic, the figure ranges from around 26% (Coco) to 57% (Toy Story).

The pandemic saw this figure drop to as low as 0.78%, for the 2020 release of *Soul.* Since the pandemic, figures have returned to normal so far. On average 37.72% of Box Office revenues come from the US and Canada. This equates to $6.93 Billion over the years. Meaning a further $10.11 Billion of revenue has come from the rest of the world. This shows that both the US & Canada market, and the rest of the world market, are both of vital importance to Pixar.

<u>General</u>

Release Day and Month (US only)*

The most common month for Pixar film releases is June, (16/28 films (57%)). This is followed by November, (7/28 films (25%)). The remaining 5 films were unusually released in March, May or December.

23 of the 28 Pixar films (82%) were released on a Friday. 5 out of 28 (18%) were released on a Wednesday. No Pixar film has ever been released on any other day of the week.

All June releases were also Friday releases.

All Wednesday releases were also November releases, and were also the day before Thanksgiving in the US.

The most popular day / month combination is Friday / June, accounting for 16 of the 28 films (57%). Although it appears to have taken Pixar until Cars (2006) to discover this popular combination.

The second most popular day / month combination is Wednesday / November (day before Thanksgiving in the US), accounting for 5 of the 28 films (18%). This combination was used from the early years, including on Pixar's first three films, and was only repeated on The Good Dinosaur (2015), and Coco (2017).

Other combinations were experimented with between 2001 – 2004, (until Pixar discovered the winning formula, Fri/Jun), and again between 2020 - 2022 when the Covid-19 pandemic caused release-scheduling chaos.

This formula of Friday / June releases appears to work very well for Pixar, as they have stuck with this combination since they discovered it, only deviating from it during the exceptional times of the Covid-19 pandemic, and just 3 other anomalies. The success of Fri/Jun is presumably because the release of a new Pixar film would coincide with parents taking their children to the cinema on a Friday night in the summer, as a treat at the end of the school / working week, when spirits are high, thanks to (hopefully) good weather.

*The data does not cover non-US release dates.

Release Frequency

Pixar films have, on average, become more and more frequently released since Pixar's first release in 1995. The difference in release date between Pixar's first two films, Toy Story, and A Bug's Life was 1099 days, which was by far the longest amount of time audiences had to wait for Pixar to release a new film. This is likely down to Pixar's reliance on less advanced technology and techniques to create films at the time, compared to the modern day, meaning animation took even longer to produce than it does today. Both the technology, and the output, at the time, were cutting edge, and this was not a quick way to make a film. Wait times have varied over the years, see the chart for full details, but the second longest wait time of 728 days, was between Monster's University (2013) and Inside Out (2015). The shortest wait time was between Turning Red (2022) and Lightyear (2022), with just 98 days in between releases. The average (mean) wait time between releases is 386 days. The median wait time between releases is 364 days, these median instances are examples of Pixar's preference of particular day / month combinations, year after year.

Film Durations

Film durations have varied over the years. The shortest Pixar film being Toy Story (1995), at 81 minutes. This is likely because the CG animation process at the time was even more time consuming and labour intensive than it is today; likely due to using less advanced technology and techniques, as well as a relative shoe-string budget. The longest Pixar film was The Incredibles 2 (2018), at 118 minutes. The average Pixar film lasts 100.39 minutes. With the exception of Toy Story, films range from 92 – 118 minutes. Unsurprisingly, films never exceed the 2 hour mark, likely because this would be too long for a child to stay focused on a film, the films are largely aimed at children. Post Toy Story, there appears to be no notable trend in film durations. Duration also appears to have no relation to Box Office Figures.

Ratings

US ratings are fairly evenly split between G and PG, with 13 and 15 films in each rating, respectively. Unsurprisingly, ratings do not exceed PG, as the films are mostly aimed at children.

<u>Audience Ratings</u>

Audience feedback on film review sites *Rotten Tomatoes, MetaCritic* and *IMDB,* has formed part of this analysis: Unsurprisingly I found there to be positive correlation between positive user reviews on all three of these sites, and higher Global Box Office revenues. Having said that, the different review sites do not agree on which film is the best, though two do agree on the worst. According to user-added scores on these sites, the best films are said to be Toy Story (Rotten Tomatoes), Ratatouille (Metacritic), or WALL-E (IMDB). The worst film according to audience reviews on these sites is said to be Cars 2 (Rotten Tomatoes and Metacritic) or Lightyear (IMDB).

<u>Academy Awards</u>

Awards vs Audience Ratings

According to the data, positive audience online reviews tend to correlate with the film winning awards. On the face, this does not come as a surprise, however it is not unusual to hear of film releases where critics / judges do not agree with mass audiences, but that is not what the data shows, although there are outliers. This is, again, looking at all three public review sites mentioned earlier.

Award winning films

No Pixar film has won more than 2 awards, though 5 films have achieved the accolade of two award wins. (The Incredibles, Up, Toy Story 3, Coco and Soul). In total, Pixar films have won 18 Academy Awards. On average, each film wins 0.64 Academy Awards.

Award nominated films
WALL-E was nominated for 5 awards, which is the most awards any Pixar film was nominated for. In total, Pixar films have been nominated for 39 awards over the years, which is an average of 1.39 nominations per film.

(The Nominations field in this dataset does not include awards that were also won, after being nominated.)

Budget vs Awards Won
Possibly surprisingly, higher budgets do not result in more award wins. In fact the opposite happens to be true, although I expect this to be a coincidence as this dataset is relatively small, and there are many outliers. Although this lack of positive correlation may highlight the fact that budget may be used to carry out a particular animation technique, to make a film possible at all, rather than to increase the award-worthiness of the film. This unexpected trend may also be explainable by the possibility that judges take budget into account when granting awards.

Number of Academy Awards vs Global Revenue
Unsurprisingly there is a relationship of positive correlation between these two metrics. Though there are many outliers. Either judges and audiences tend to agree, or audiences are influenced by awards won by a film.

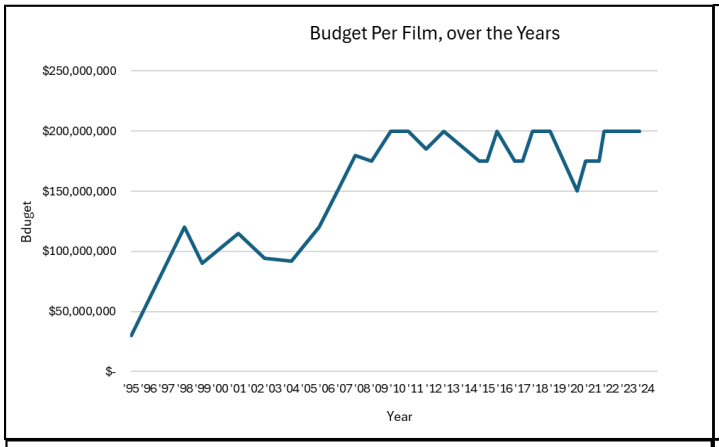
Directors with the most awards
Looking at the charts is the best way to illustrate this. Directors Lee Unkrich and Pete Docter have won the most awards (6 each) from Pixar films. (This includes awards not related to directing, such as *Original Score*. Nevertheless, I have chosen to categorize this metric by Director as the Director oversees the overall creative vision for the film). I have also created another chart to show Directors with the Most Award Wins per Film. Bob Anderson, Kemp Powers and Adrian Molina top this chart, with an average of 2 award wins per film. Both charts also show the number of nominations their films received.

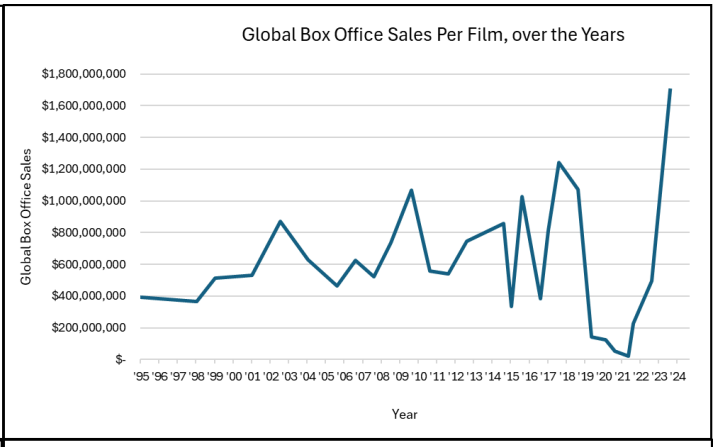
Academy Awards for Pixar films are the only awards included in this analysis.

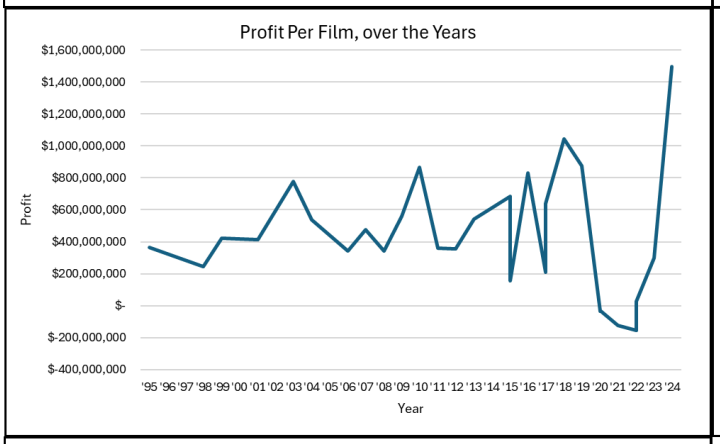
Share

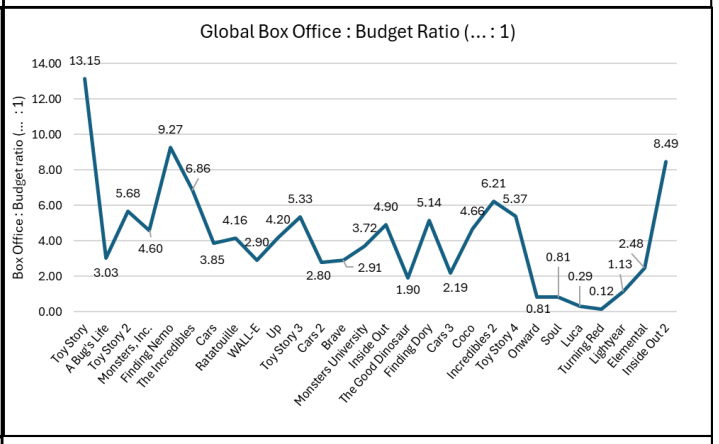
Data Visualisations
I have created a collection of charts and visualisations which support the analysis I have written above, and tell additional stories too. I have included the charts below.

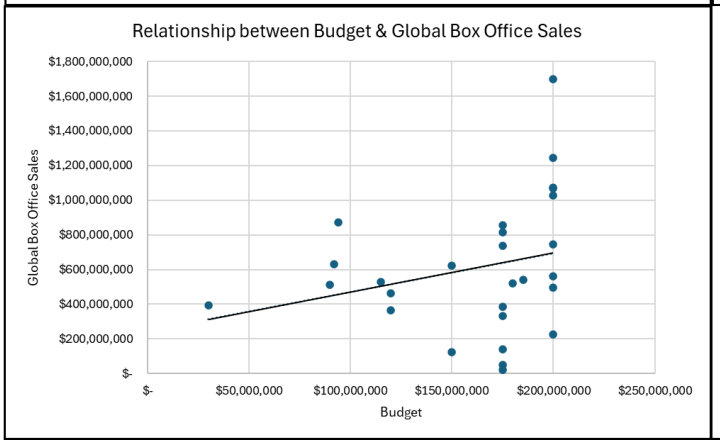
Budgets and Box Office Sales Charts

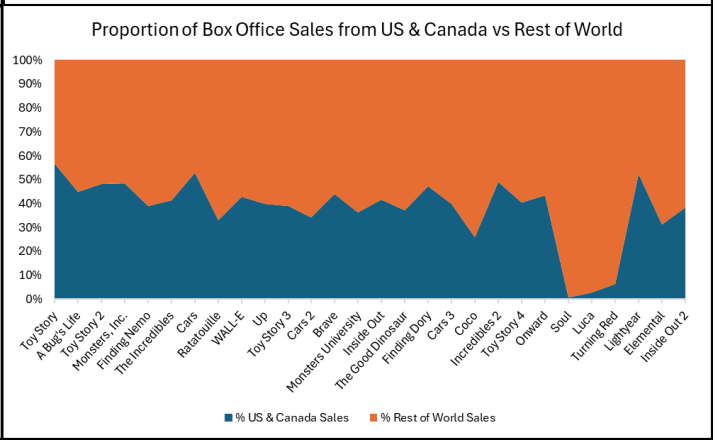




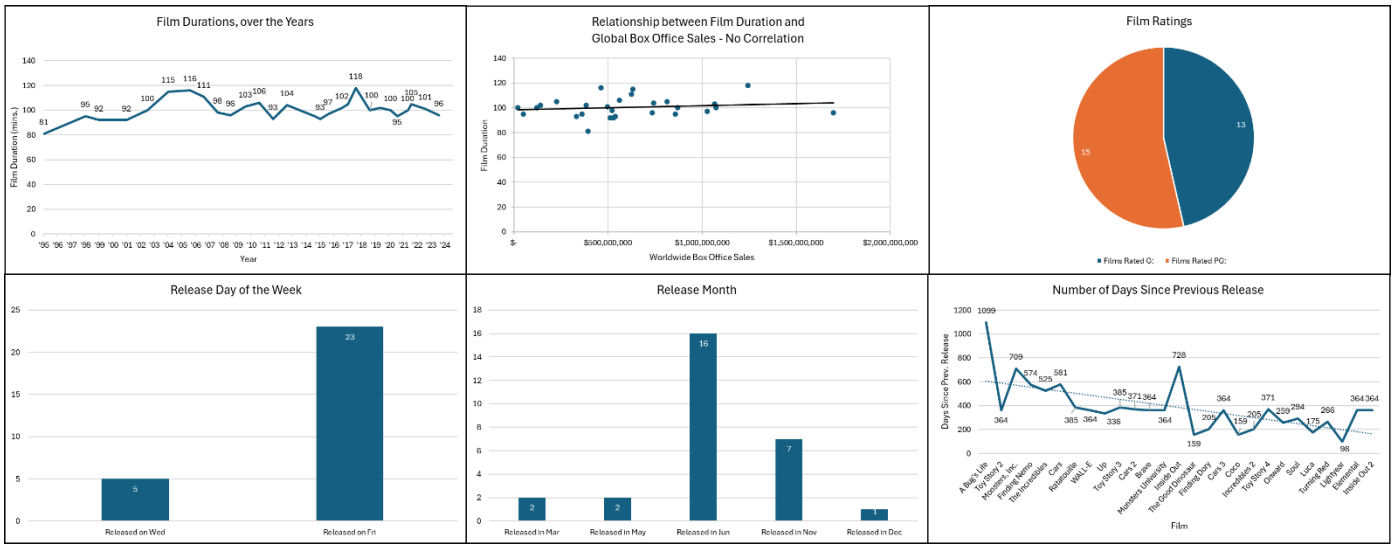






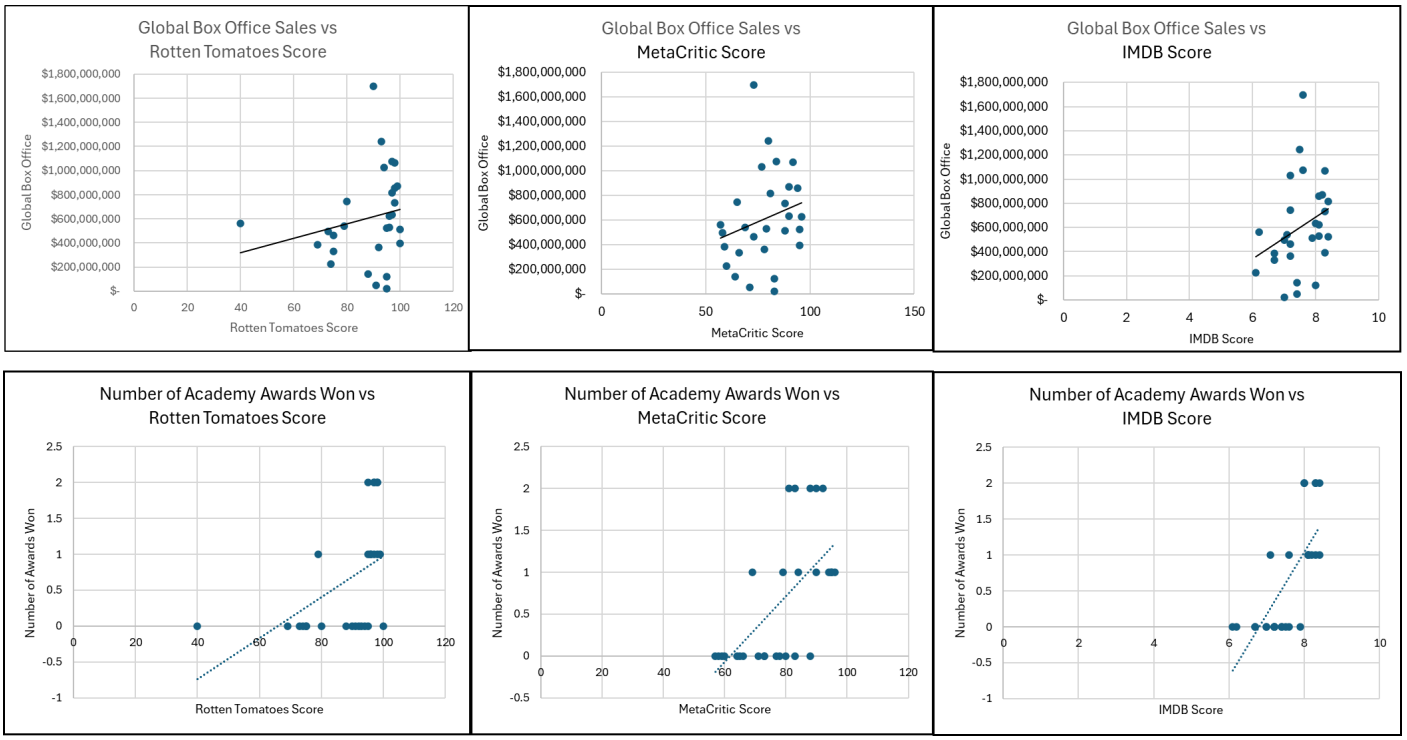


General Charts

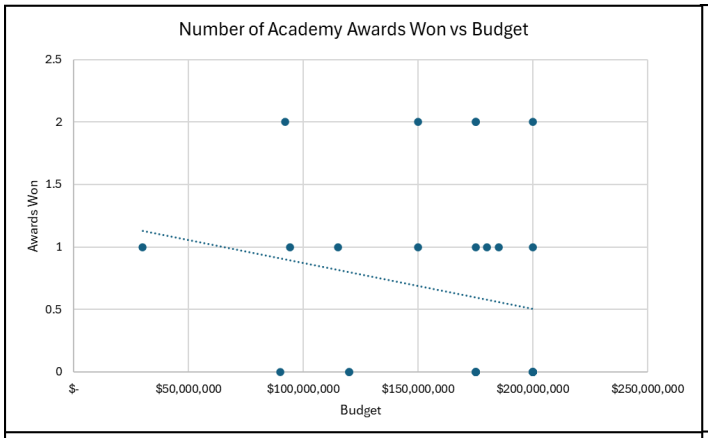


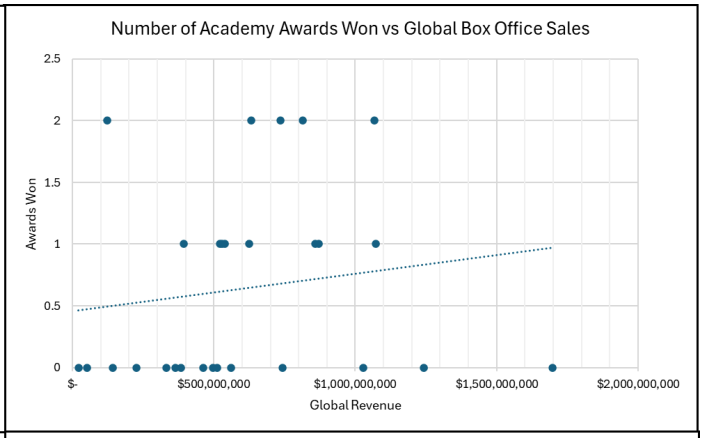
Film	Release Day	Release Month
Toy Story	Wed	Nov
A Bug's Life	Wed	Nov
Toy Story 2	Wed	Nov
Monsters, Inc.	Fri	Nov
Finding Nemo	Fri	May
The Incredibles	Fri	Nov
Cars	Fri	Jun
Ratatouille	Fri	Jun
WALL-E	Fri	Jun
Up	Fri	May
Toy Story 3	Fri	Jun
Cars 2	Fri	Jun
Brave	Fri	Jun
Monsters University	Fri	Jun
Inside Out	Fri	Jun
The Good Dinosaur	Wed	Nov
Finding Dory	Fri	Jun
Cars 3	Fri	Jun
Coco	Wed	Nov
Incredibles 2	Fri	Jun
Toy Story 4	Fri	Jun
Onward	Fri	Mar
Soul	Fri	Dec
Luca	Fri	Jun
Turning Red	Fri	Mar
Lightyear	Fri	Jun
Elemental	Fri	Jun
Inside Out 2	Fri	Jun

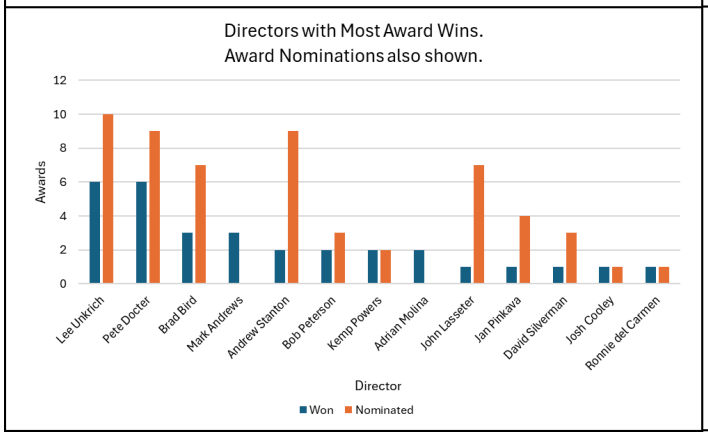
Audience Reception Charts

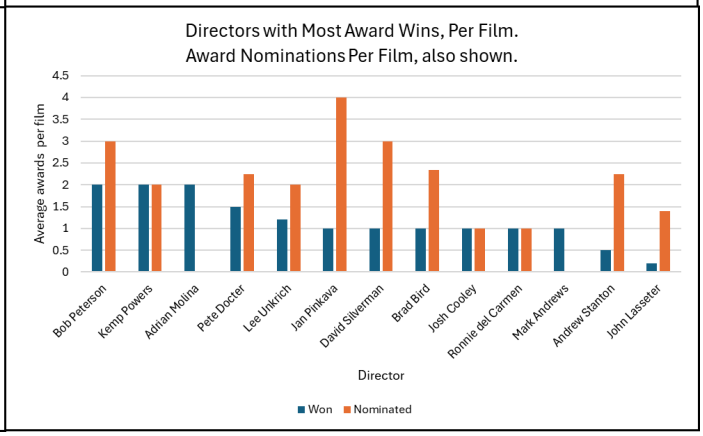


Awards and Directors Charts









Act

Recommendations

I recommend Pixar continue to make films with budgets of $150 - 200 Million. The data has shown that while this level of budget doesn't provide maximum Box Office : Budget ratio, it does tend to provide maximum revenue and profit.

Box Office : Budget ratio can be maximised primarily by making a film on a lower budget, however the data shows that this would produce a smaller profit than allocating a budget of $150 - 200 Million, so is not recommended.

Some films will perform better than others, and Pixar should not let a less successful film influence the budget of a future film. Revenues for the film that follows Inside Out 2 (revenue $1.698 Billion), are unlikely to be as high as Inside Out 2, but I recommend Pixar keep budgets consistent, and keep sight of the bigger picture.

The Covid-19 pandemic represented exceptional circumstances, and the first period in Pixar's history of loss-making films. The effects of the pandemic were unavoidable, and while the world is now better prepared for another pandemic, it is still possible that a major international event such as another pandemic or a war could take place, if this happens then there may be little that Pixar can do to minimise Box Office losses. This analysis has led me to recommend that weathering the storm would be the best approach should an event of similar magnitude happen again, as future film profits can easily recoup the losses of past loss-making films.

I recommend Pixar release films on Disney+ 3 months after cinema release. The data has shown that this practise means audiences will still visit cinemas, and contribute to Box Office revenue, before they stream the film, in place of buying a DVD / Blu Ray. The data shows that this formula ensures Box Office revenues are retained. The data shows that releasing a film on Disney+ any sooner will result in a loss of cinema Box Office revenue, so this is not recommended.

The data shows the US & Canada market, and the Rest of the World market, are both of vital importance to Pixar. Pixar cannot afford to lose either market, so Pixar should continue to make films that appeal to both markets. US-made films are widely popular, in all corners of the globe. The US-centric culture of Pixar films does not deter foreign audiences from watching, as foreign audiences are used to this, and this appears to be a tiny price to pay for good films.

This analysis has identified the release day / month of Friday / June as contributing to the film's financial success. I recommend this release day / month combination continues, with the exception of when Pixar want to release more than 1 film a year. Failing Friday / June, Pixar should consider the Wednesday / November (Thanksgiving) combination for US releases, as this has also been successful in the past. Beyond this, any other combination should be used carefully and sparingly, as results have proven unpredictable.

Pixar's formula for ratings (G/PG) and durations (1.5 - 2 hours) has been set in stone since A Bug's Life, and I don't see any reason not to stick with this formula. It works.

Pixar should be hyper aware of the importance of choosing the right director, when crewing up for their next film. Analysis and visualisations earlier in this report show the most successful directors, in terms of awards won. Award wins mean more revenue; and winning awards does not tend to require a higher budget. So I recommend Pixar choose a Director(s) with a track record of winning awards, in order to generate the most revenue, and therefore profit.

During this analysis, by observing each metric on each film, it is possible to see what does and doesn't result in success. Pixar have been known to produce extremely successful films, and this analysis has identified how to consolidate and maximise their success. As new trends and metrics emerge, Pixar should experiment until they find the sweet spot for each metric which results in the highest profits.