

Final Project

Su Doğa Karaca

December 8, 2022

DATA 22700 Final Project: Telling a Story

The Dataset: Disney Movies 1937-2016 Gross Income (Rahman, 2021)

The following project tells a story about the movies from 1980 to 2016 that were released by Disney productions. Walt Disney Studios is an umbrella studio that embodies multiple film and animation studios. Under the name of Walt Disney Studios, approximately 600 films have been distributed. This dataset that was obtained from Kaggle contains 579 Disney movies that were released from 1937 to 2016. The dataset was built using the actual Walt Disney Box Office history. The attributes included in the dataset are the movie title, release date, genre, Motion Picture Association film rating system (MPAA) rating, total gross, and inflation adjusted gross. Through the data cleaning process, 4 rows that had only the title of the movie and its release date were removed as they were missing most of the attributes discussed in this project. In addition, because this project focuses on the trends starting from 1980s, when Disney started distributing at least one movie a year, the data from 1937 to 1979 was not used in this project. As this project focuses on discussing the time and revenue trends of relatively modern times and as there were only 1-2 movies released every five years up until 1975s, not utilizing the data from 1937 to 1979 did not affect the trends discussed in the following sections.



Walt Disney logo. PNG IMG. Retrieved December 8, 2022, from <https://pngimg.com/image/57582>.

Visualization 1:

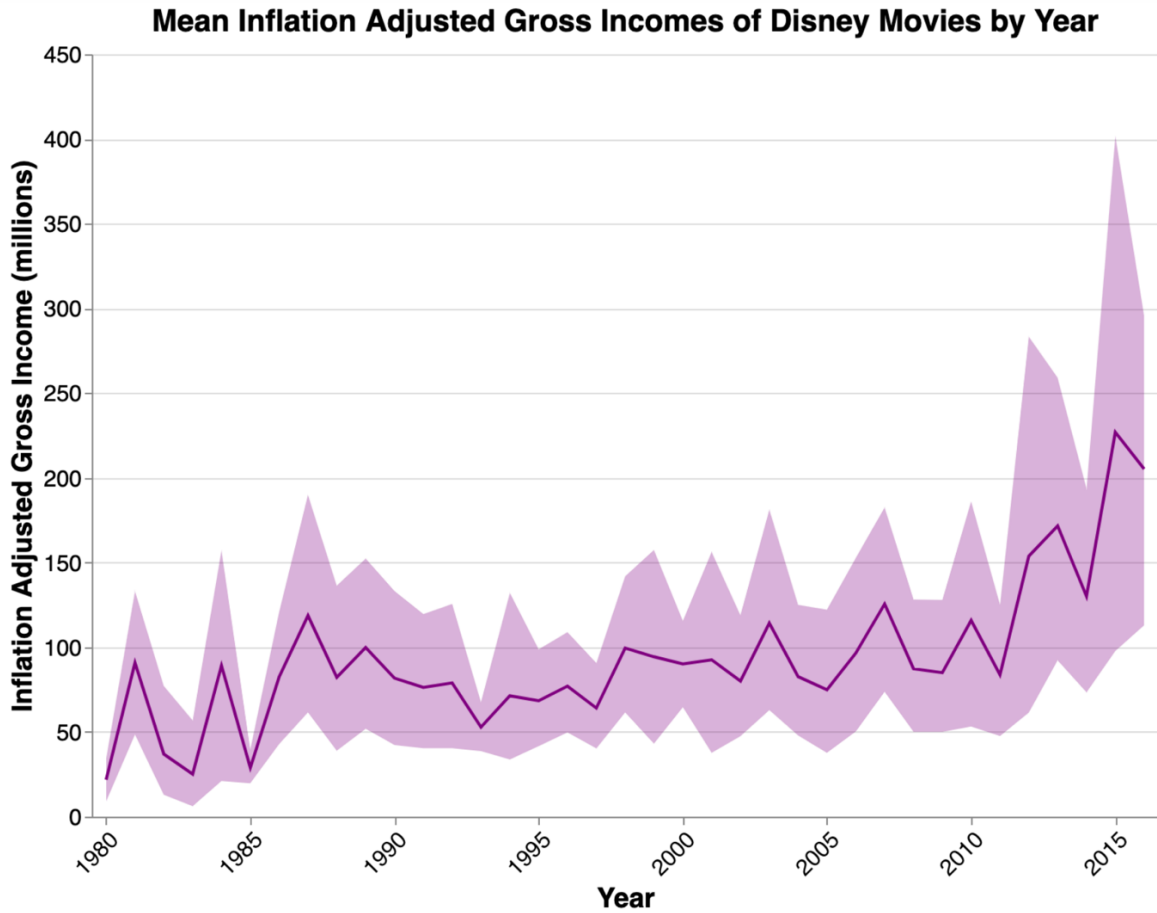


Figure 1. This line chart visualizes the mean inflation adjusted gross incomes of the Disney movies published each year in million dollars between the years 1980 and 2016. Along with the mean inflation adjusted gross income values, the error band for each year is also graphically represented in the visualization, taking the variability of the data into account.

Discussion: It can be seen from the chart that, even though there are sharp fluctuations across the years, the mean inflation adjusted income of each year shows an increasing pattern. It should be noted that the observed variable is the adjusted according to the ticket price inflation. This adjustment is performed by taking estimated ticket sales for the movies of a year and multiplying this number by the accepted mean movie ticket price of the year the calculation is made in. So, this value is resistant to economic differences across years. It is noticeable that the variability pattern around the mean also increases over time. The error bands around the mean value of each year shows how spread the data are around the mean inflation adjusted gross income value. However, this visualization alone does not help to our understanding of why this might be the case. This increasing pattern might be due to having more movies produced each year and the accompanying increase in the gross income range or other confounding variables, such as the acquisition of multiples film studios with varying gross income rates, as Disney is one of the largest media conglomerates acquiring studios like Lucasfilm and Pixar. The following visualizations will discuss these further trends in the data.

Visualization 2:

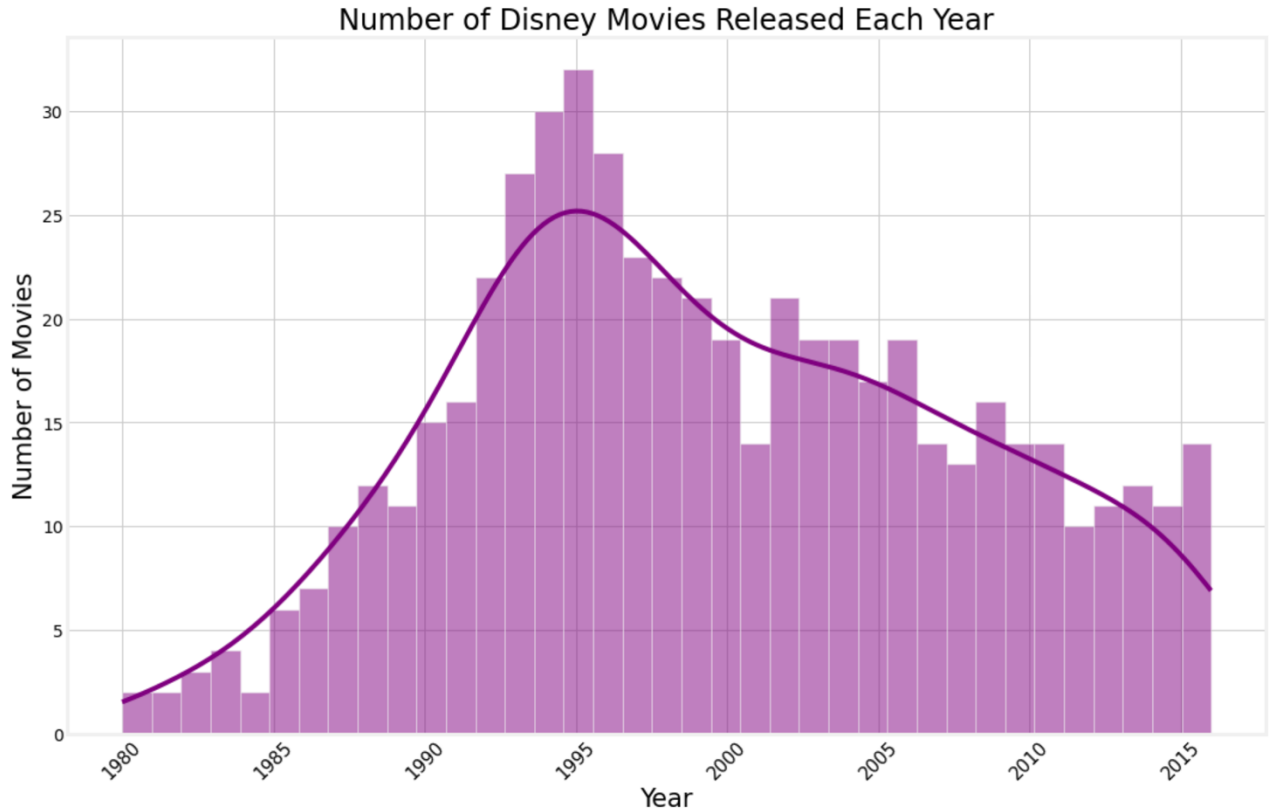


Figure 2. This histogram visualizes the distribution of the Disney movies released each year from 1980 to 2016 with the accompanying kernel density estimate line for a smoothed interpretability.

Discussion: Following the discussion of the Figure 1, this time-series is produced to have a better understanding of the Disney's movie production rate. It is notable that the dataset contains movies that were originally produced under the name Walt Disney. This means that the movies purchased by Disney through the acquisition of other film productions are not included in the data. This means that there are some movies released under the names Buena Vista Pictures Distribution, Inc., Walt Disney Pictures, Walt Disney Animation Studios, Touchstone Pictures, Pixar, Marvel Studios, Lucasfilm, 20th Century Studios, and more. As an example, Toy Story 3 and 4 are not a part of this data because even though they are distributed by Pixar, the production is not made by the Disney Animation Studios. Or only the Star Wars movies that were released after the acquisition of Lucasfilm are a part of the Disney family. It can be seen from Figure 2 that the distribution is unimodal with a peak around the 1995. This period is known as 'The Disney Renaissance.' This period included the years from 1989 to 1999, in which Walt Disney reproduced its high success animations that were produced before 1970s (Pickett, 2014). This reproduction helped Disney to gain authority in the media and entertainment field, making Disney more profitable than its previous decades. However, in accompany of the Figure 1, it can be seen that this success did not contribute to Disney's inflation adjusted gross income. So, the result of the continues increase in the mean inflation adjusted gross income is not the increase in the number of productions and the accompanying increase in the gross income range. This might be because the increase in the production might have resulted in division of the gross income across multiple movies. Therefore, it may be hypothesized that the reason of this increasing trend is the production or acquisition of high revenue movies as the years progress. The following figures will inquire this.

Visualization 3:

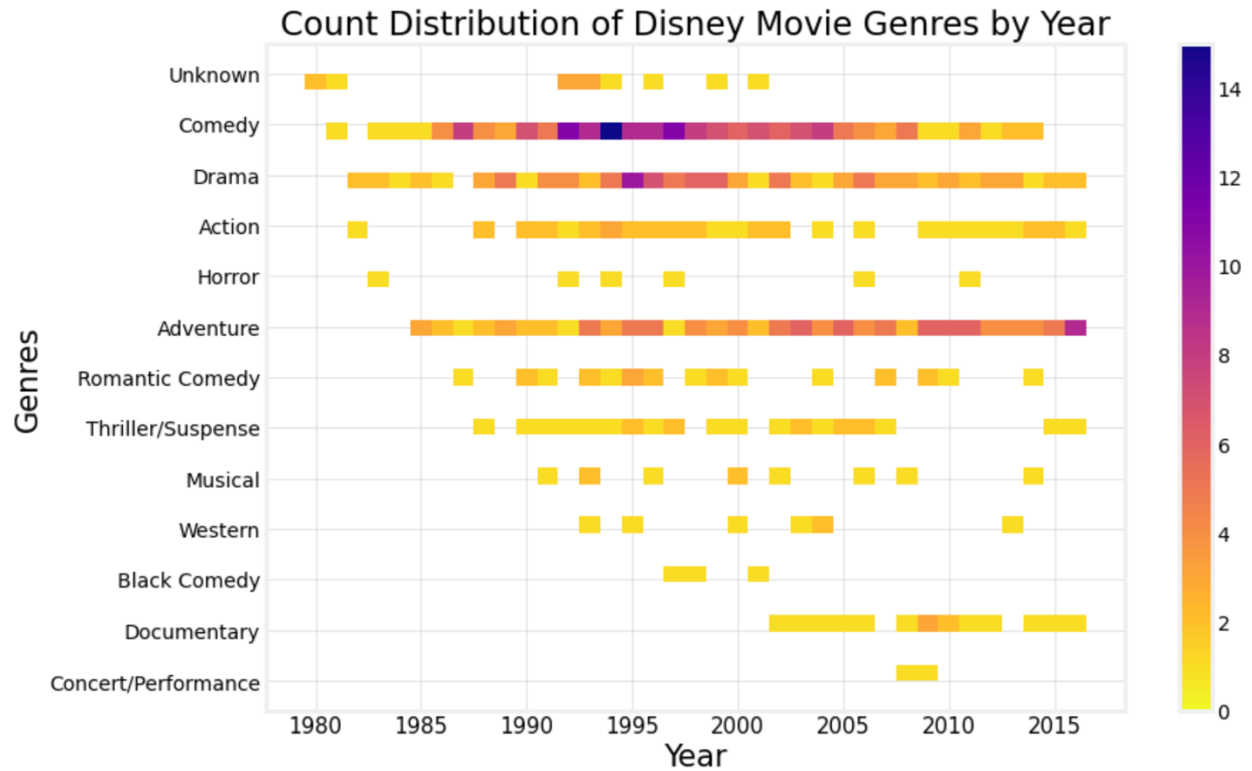


Figure 3. This heatmap density figure visualizes the count distribution of Disney movies released each year from 1980 to 2016 by genres of the movies. The accompanying color map visualizes the number of movies produced each year by genre.

Discussion: From the previous figures, it was seen that the mean gross incomes of the years were not directly associated with the number of movies produced. Therefore, it is worthwhile to check the specific features of the movies produced each year. In Figure 3, the distribution of movies by genres are observed. The movie titles were grouped by genre and then divided by year. From this distribution, it can be seen that not all genres were produced uniformly across this 37-years-long period. The production of drama, comedy, adventure, and action genres were relatively uniform across years. However, it is important to mention that around 1990s Disney focused on producing more comedy and drama movies than any other genres. Yet, the focus of Disney shifted to adventure movies after 2000s. This change in trends might be due to the acquisition of big production companies at given years. For example, in 1988, Disney pictures acquired the Touchstone Pictures (DeKalb, 2020), which is famous for its successful comedy and drama movies such as *Pretty Woman*, *Dead Poets Society*, and *The Prestige*. Moreover, Disney purchased Marvel Entertainment in 2009 (Goldman, 2009) and Lucasfilm in 2012 (BBC, 2012), which are film studios famous for their adventure and action movies such as the *Avengers* series and the *Star Wars* series, respectively. Looking at this information, it might be possible that the changes in the gross income with respect to acquisition of different genre movies might have led Disney to shift their overall genre production. The following visualization will discuss this aspect of the data.

Visualization 4:

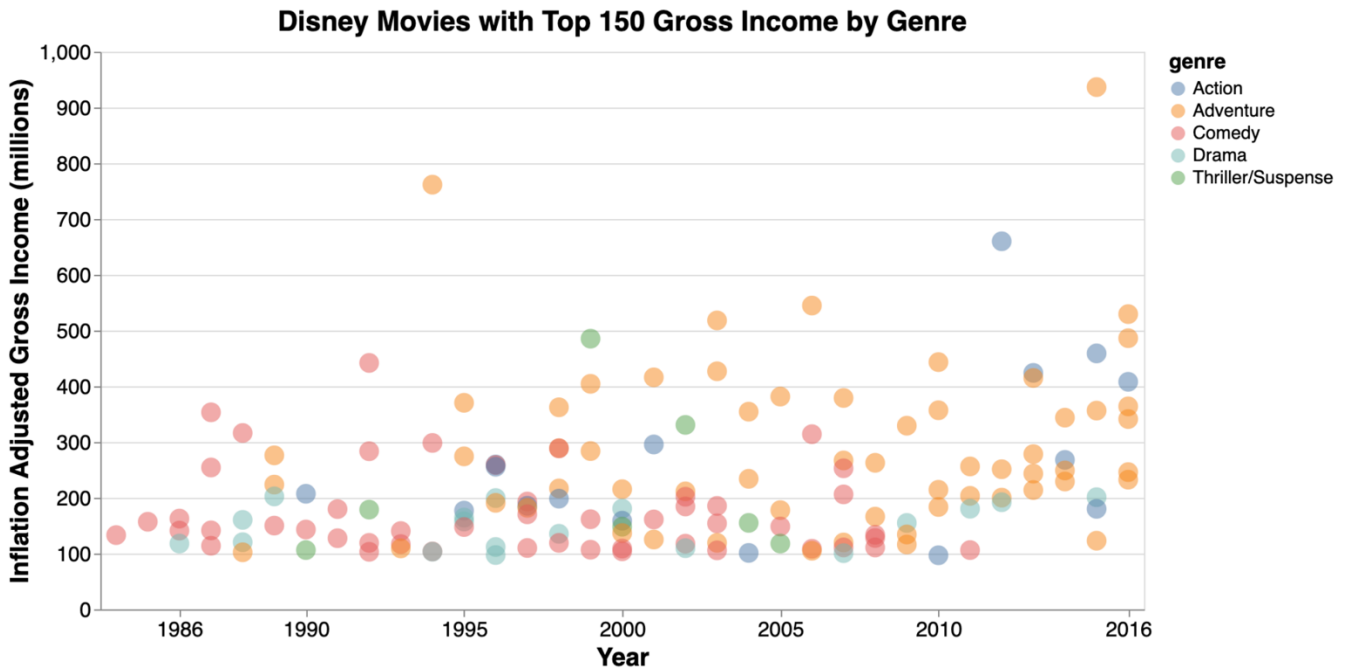


Figure 4. This scatterplot visualizes the top 100 Disney movies with the highest inflation adjusted gross income, released between 1980 and 2016 by the top 5 most stably produced genres of the movies.

Discussion: As discussed with Figure 3, Disney shifted their genre trends from comedy and drama to adventure and action movies, possibly along with the acquisition of large revenue adventure / action movie production companies such as Lucasfilm and Marvel Entertainment. In this scatterplot distribution of the top 100 gross income movies of Disney, the increasing trend of adventure movies can be clearly seen. Both the production of the adventure movies and their inflation adjusted gross income by year seem to increase across this period. It can also be seen that the production of high gross income comedy movies continued until the 2010s, and the production of high gross income drama movies got less dense over time. However, in comparison to adventure movies, such as the Star Wars sequels, these genres' gross income seems to stay stable. This positively increasing gross income trend of adventure movies compared to the other genres might have led Disney to produce more adventure movies together with the acquisition of successful production companies that produce adventure movies.

References

- Pickett, L. (2014, November 17). *Ranking: The disney renaissance from worst to best*. Time. Retrieved December 8, 2022, from <https://time.com/3590521/disney-renaissance/>
- BBC. (2012, October 31). *Disney buys Star Wars Maker lucasfilm from George Lucas*. BBC News. Retrieved December 8, 2022, from <https://www.bbc.com/news/business-20146942>
- DeKalb, C. (2020, April 14). *Disney for adults: The story behind Touchstone Pictures*. Medium. Retrieved December 8, 2022, from <https://medium.com/framerated/disney-for-adults-the-story-behind-touchstone-pictures-9367283b8a5a>
- Goldman, D. (2009, August 31). *Disney to buy Marvel for \$4 billion*. CNNMoney. Retrieved December 8, 2022, from https://money.cnn.com/2009/08/31/news/companies/disney_marvel/
- Rahman, R. (2021, December 7). *Disney Movies 1937-2016 Gross Income*. Retrieved December 8, 2022, from <https://www.kaggle.com/datasets/rashikrahmanpritom/disney-movies-19372016-total-gross>.

Project Feedback for Su Karaca by Jennifer Spinoglio

Visualization 1:

I think the second version provides a better view of the variation from year to year, which I think is interesting and important. As a general reminder, I would rather use values like 100, 200, etc. on the tick marks on the axes and “total gross income (millions),” than have “100,000,000, 200,000,000, etc. on the tick marks, I think it looks clearer and is easier for people to understand. You should also specify which currency this is in for clarity. I would also look more into 1994’s movies so see why this peak is appearing and disappearing. It could be either an error in the data set or a genuine observation. You might also consider adding annotations where in time Disney made major acquisitions/business deals (ex. buying Pixar in 2006) so we can see how that effected their total gross income.

Visualization 2:

I love this way of looking at the distribution of films over time! I think it could also be interesting to interrogate why Disney is making few movies nowadays given that they are making so much more money than they were in the mid-90s, when the number of movies made annually peaks. I would also be sure to correct your y-axis here, you are looking at number of films, which does not have much to do with MPAA ratings. I would also be careful with so many values on the x-axis, maybe consider doing tick marks every 2 or 5 years. Finally, watch out for font size, some of these graphs, when presented at this size, have very small text. Again, I would consider adding annotations of major deals/acquisitions to see if this added or took away films.

Visualization 3:

I really like this kind of visualization! I would be sure to add a title for your figure, as right now I am a bit confused on what you are measuring. You also may want to consider defining some of these terms, or the methodology in which the genres were categorized. For example, were films marked as “unknown” because they fall into two or more of these genres or because they defy it altogether? Who decided what makes “action” different from “adventure”? Did this come from Disney’s categorization or external critics, viewers, etc.?

Visualization Four:

A trendline here would help the viewer visualize the increase in action and adventure movies over time, it’s a bit unclear to me now. In addition, I would again consider changing the scale of the y-axis

(i.e. from 100,000,000 gross income to 100 gross income (millions)) to make it easier to read. I like the colorfulness of this color scheme, although it is a little difficult to tell the orange from the yellow. I would also be careful to mention that these are only the top 100 movies from the data set in the caption.

Overall: One quick note, I would be interested in how this dataset defines a “Disney movie.” Is it all movies released under the Disney name and branding, or all released by Disney, including other names/brands (Touchstone Pictures, etc.)? Does it include Star Wars movies from before Lucasfilm was purchased? I also think it could be cool if you have any suggestions for Disney on what kind of movies are most profitable and what they should do going forward.