final_project_disney_olivier

March 18, 2024

0.1 The Walt Disney Company: An Analysis of Trends, Rating and Revenue

1 Introduction

The Walt Disney Company has been synonymous with family entertainment for nearly a century. By exploring various aspects of the Disney movie dataset, ranging from financial success measured by gross revenues to categorical distinctions such as MPAA ratings and directors, we can uncover insights into the company's production and distribution strategies, as well as audience preferences over the years.

In this analysis, we will try to understand the following questions:

1. **Trend Analysis**: We aim to understand the financial trajectory of Disney movies by examining the change in gross revenues over the years and identifying trends in MPAA ratings, which may reflect evolving content strategies and audience demographics.
2. **Comparative Analysis**: By adjusting gross revenues for inflation, we can identify which Disney movies have truly resonated with audiences in terms of financial success and compare nominal versus real revenue trends.
3. **Categorical Analysis**: We will investigate the prevalence of MPAA ratings among Disney movies and assess the financial performance within each rating category. Additionally, we will explore the influence of directors on the Disney movie landscape.

2 Description of the data

In our analysis, we will use three datasets:

1. **Disney voice Actors**: Contains information about characters, the voice actors who portray them, and the associated movies.
2. **Disney Directors**: Lists the directors along with their directed Disney movies
3. **Disney movies total gross**: which Offers financial data, including the movie title, release date, MPAA rating, total gross, and inflation-adjusted gross revenue, which allows for a fair comparison across different time periods. through the functions we learned in this course, I will merge them to have one complete dataset which i will use to perform the analysis.

In this analysis, I aim to explore several questions: 1. How have Disney movie gross revenues changed over the years? 2. What is the trend in the ratings (MPAA) of Disney movies over time? 3. Which Disney movie has the highest gross revenue when adjusted for inflation? 4. How does the total gross compare to the inflation-adjusted gross over the years? 5. What is the most

common MPAA rating for Disney movies, and how do movies within each rating category perform financially?

Before moving further, let's import the packages that i will use in this analysis

[13]:
```
#import library that we need for this analysis
import altair as alt
import pandas as pd
```

Import all the data we will use in this analysis

[14]:
```
#import data that we need for the analysis
voice_actors = pd.read_csv("data/disney-voice-actors.csv")
```

[15]:
```
#rename the movie column name into movie_title
voice_actors = voice_actors.rename(columns={'movie': 'movie_title'})
voice_actors.head()
```

[15]:

	character	voice-actor	movie_title
0	Abby Mallard	Joan Cusack	Chicken Little
1	Abigail Gabble	Monica Evans	The Aristocats
2	Abis Mal	Jason Alexander	The Return of Jafar
3	Abu	Frank Welker	Aladdin
4	Achilles	None	The Hunchback of Notre Dame

[16]:
```
disney_character = pd.read_csv("data/disney-characters.csv")
```

[17]:
```
disney_director = pd.read_csv("data/disney-director.csv")
```

[18]:
```
#rename the movie column name into movie_title
disney_director = disney_director.rename(columns={'name': 'movie_title'})
disney_director.head()
```

[18]:

	movie_title	director
0	Snow White and the Seven Dwarfs	David Hand
1	Pinocchio	Ben Sharpsteen
2	Fantasia	full credits
3	Dumbo	Ben Sharpsteen
4	Bambi	David Hand

[19]:
```
disney_movies_total_gross = pd.read_csv("data/disney_movies_total_gross.csv")
disney_movies_total_gross.head()
```

[19]:

	movie_title	release_date	genre	MPAA_rating	\
0	Snow White and the Seven Dwarfs	Dec 21, 1937	Musical	G	
1	Pinocchio	Feb 9, 1940	Adventure	G	
2	Fantasia	Nov 13, 1940	Musical	G	
3	Song of the South	Nov 12, 1946	Adventure	G	

```
4                      Cinderella  Feb 15, 1950      Drama           G

    total_gross inflation_adjusted_gross
0  $184,925,485           $5,228,953,251
1   $84,300,000           $2,188,229,052
2   $83,320,000           $2,187,090,808
3   $65,000,000           $1,078,510,579
4   $85,000,000             $920,608,730
```

[20]:
```
#Let's firt merge disney_movie_total_gross with disney_director

disney_movie_gross_director = pd.merge(disney_movies_total_gross,␣
 ↪disney_director, on= 'movie_title', how='left')
disney_movie_gross_director.head()
```

[20]:
```
                       movie_title  release_date      genre MPAA_rating  \
0  Snow White and the Seven Dwarfs  Dec 21, 1937    Musical           G
1                        Pinocchio   Feb 9, 1940  Adventure           G
2                         Fantasia  Nov 13, 1940    Musical           G
3                 Song of the South  Nov 12, 1946  Adventure           G
4                       Cinderella  Feb 15, 1950      Drama           G

    total_gross inflation_adjusted_gross         director
0  $184,925,485           $5,228,953,251       David Hand
1   $84,300,000           $2,188,229,052   Ben Sharpsteen
2   $83,320,000           $2,187,090,808     full credits
3   $65,000,000           $1,078,510,579              NaN
4   $85,000,000             $920,608,730  Wilfred Jackson
```

[21]:
```
disney_movie_gross_director.to_csv('disney_movie_gross_director.csv',␣
 ↪index=False)
```

Now let's merge disney_movie_gross_director with voice_actors to have a final dataset that we will use in our analysis

[22]:
```
#disney_movie_gross_final = pd.merge(disney_movie_gross_director, voice_actors,␣
 ↪on= 'movie_title', how='left')
#disney_movie_gross_final.head()
```

Let's Perfom some data wrangling on the disney_movie_gross_director so that we are able to start our exploring/ answering our questions.

first, I convert **Inflation_adjusted_gross** and **total_gross** from string to numerical type, removing currency formating.

Second, I will clean date structure in the right format so that i am able to use.

[23]:
```
# Convert 'inflation_adjusted_gross' column to string, remove dollar signs and␣
 ↪commas,
# and then convert back to numeric format for further analysis
disney_movie_gross_director['inflation_adjusted_gross'] =␣
 ↪disney_movie_gross_director['inflation_adjusted_gross'].astype(str)
disney_movie_gross_director['inflation_adjusted_gross'] = pd.
 ↪to_numeric(disney_movie_gross_director['inflation_adjusted_gross'].str.
 ↪replace('[\$,]', '', regex=True))

# Total Gross
disney_movie_gross_director['total_gross'] =␣
 ↪disney_movie_gross_director['total_gross'].astype(str)
disney_movie_gross_director['total_gross'] = pd.
 ↪to_numeric(disney_movie_gross_director['total_gross'].str.replace('[\$,]',␣
 ↪'', regex=True))
```

[24]:
```
# Check the data types and look for any remaining missing values
disney_movie_gross_director.info()
disney_movie_gross_director.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 579 entries, 0 to 578
Data columns (total 7 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   movie_title             579 non-null    object
 1   release_date            579 non-null    object
 2   genre                   562 non-null    object
 3   MPAA_rating             523 non-null    object
 4   total_gross             579 non-null    int64
 5   inflation_adjusted_gross  579 non-null    int64
 6   director                49 non-null     object
dtypes: int64(2), object(5)
memory usage: 36.2+ KB
```

[24]:
```
movie_title                 0
release_date                0
genre                      17
MPAA_rating                56
total_gross                 0
inflation_adjusted_gross    0
director                  530
dtype: int64
```

Clearly, The Information about movie director is lacking in this file, Not all the movies have their director assigned.

[25]:
```
#Convert release_date to datetime and extract year
disney_movie_gross_director = disney_movie_gross_director.copy()

# Convert release_date to datetime and extract year
disney_movie_gross_director['release_date'] = pd.
 ↪to_datetime(disney_movie_gross_director['release_date'])
disney_movie_gross_director['release_year'] =␣
 ↪disney_movie_gross_director['release_date'].dt.year
```

[26]:
```
# Group by decades
disney_movie_gross_director.loc[:, 'decade'] =␣
 ↪(disney_movie_gross_director['release_year'] // 10) * 10
```

[27]:
```
# Adjust gross earnings by the count of movies released for each rating

rating_gross_adjusted = disney_movie_gross_director.
 ↪groupby('MPAA_rating')['inflation_adjusted_gross'].sum() /␣
 ↪disney_movie_gross_director.groupby('MPAA_rating').size()
rating_gross_adjusted = rating_gross_adjusted.reset_index()
rating_gross_adjusted.columns = ['MPAA_rating', 'adjusted_gross_per_movie']
```

[28]:
```
disney_movie_gross_director.to_csv('disney_movie_gross_director.csv',␣
 ↪index=False)
```

Now, I will try to explore the questions mentionned above

3 How have Disney movie gross revenues changed over the years?

[29]:
```
# first, let's convert release_date to datetime and extract the year
disney_movie_gross_director['release_date'] = pd.
 ↪to_datetime(disney_movie_gross_director['release_date'], errors='coerce')
disney_movie_gross_director['year'] =␣
 ↪disney_movie_gross_director['release_date'].dt.year

# second, we group by year and sum the total gross
disney_annual_gross = disney_movie_gross_director.
 ↪groupby('year')['total_gross'].sum().reset_index()

# third, we create the chart with bold and clear color for the line
chart = alt.Chart(disney_annual_gross).mark_line(color='blue', strokeWidth=3).
 ↪encode(
    x=alt.X('year:O', axis=alt.Axis(title='Year')),
    y=alt.Y('total_gross:Q', axis=alt.Axis(title='Total Gross Revenue')),
    tooltip=['year', 'total_gross']
).properties(title='Disney Movie Gross Revenues Over the Years')
```

```
chart
```

[29]: alt.Chart(…)

3.1 Comment about the above graph

Here are few observation for the graph above: 1. The graph indicates an overall upward trend in Disney's movie gross revenues over the years. This suggests that Disney movies have been generating more revenue as time progresses 2. There are noticeable fluctuations in the gross revenue from year to year, indicating variability in the performance of Disney movies. 3. There is a significant upward spike in gross revenue in the most recent years. at this point of the analysis, we can't tell why. this increase could be a result of so many factors, that hopefully we can explore more as we delve deeper in the data.

4 What is the trend in the ratings (MPAA) of Disney movies over time?

[30]:
```
#Visualize Number of Movies by Rating Over Time
movies_by_rating = alt.Chart(disney_movie_gross_director).mark_bar().encode(
    x='release_year:O',
    y='count():Q',
    color='MPAA_rating:N',
    tooltip=['release_year', 'count()', 'MPAA_rating']
).properties(
    width=600,
    height=300,
    title='Count of Disney Movies by Rating Over Time')
movies_by_rating
```

[30]: alt.Chart(…)

4.1 Comment about the above graph

The bar graph above shows the count of disney movies released over time, broken down by MPAA rating. from the graph, we can observe several trends:

1. **Dominance of G and PG Ratings**: G and PG-rated movies are the most frequently released films by Disney, which aligns with the company's family-friendly branding. disney movies
2. **Increase in PG-13 Releases**: There has been a noticeable increase in PG-13 releases starting in the early 2000s, indicating a shift to produce content for a slightly older audience.
3. **Peak of Releases**: The highest number of releases in a single year occurred between 1991-mid 2000s, with a mix of G, PG, and PG-13 rated movies.
4. **Decline in G-rated Releases**: There is a visible decline in G-rated releases after the peak, with PG and PG-13 movies becoming more common.

Overall, the graph suggests that disney has diversified its audience reach by producing more films with higher age ratings (PG and PG-13) over time.

5 Which Disney movie has the highest gross revenue when adjusted for inflation?

[31]:
```
# Find the movie with the highest inflation-adjusted gross revenue
highest_grossing_movie = disney_movie_gross_director.loc[
    disney_movie_gross_director['inflation_adjusted_gross'].idxmax()
]
highest_grossing_movie
```

[31]:
```
movie_title                 Snow White and the Seven Dwarfs
release_date                            1937-12-21 00:00:00
genre                                               Musical
MPAA_rating                                               G
total_gross                                       184925485
inflation_adjusted_gross                         5228953251
director                                        David Hand
release_year                                          1937
decade                                                1930
year                                                  1937
Name: 0, dtype: object
```

[32]:
```
# let's start by Sorting  the data to find the top 10 movies with the highest␣
 ↪inflation-adjusted gross revenue
top_10_grossing_movies = disney_movie_gross_director.nlargest(10,␣
 ↪'inflation_adjusted_gross')

# Create a bar chart
chart = alt.Chart(top_10_grossing_movies).mark_bar().encode(
    x=alt.X('movie_title:N', sort='-y', title='Movie Title'),
    y=alt.Y('inflation_adjusted_gross:Q', title='Inflation Adjusted Gross␣
 ↪Revenue'),
    color='movie_title:N',
    tooltip=['movie_title:N', 'inflation_adjusted_gross:Q']
).properties(title='Top 10 Disney Movies by Inflation Adjusted Gross Revenue')

chart
```

[32]:
```
alt.Chart(…)
```

6 How does the total gross compare to the inflation-adjusted gross over the years?

[33]:
```
# First let's convert release_date to datetime and extract the year
disney_movie_gross_director['release_date'] = pd.
 ↪to_datetime(disney_movie_gross_director['release_date'], errors='coerce')
disney_movie_gross_director['year'] =␣
 ↪disney_movie_gross_director['release_date'].dt.year


# second, Group by year and sum the total gross and inflation_adjusted_gross
disney_annual_gross = disney_movie_gross_director.groupby('year').agg({
    'total_gross': 'sum',
    'inflation_adjusted_gross': 'sum'
}).reset_index()

# Create the chart
total_gross_chart = alt.Chart(disney_annual_gross).mark_line(color='blue').
 ↪encode(
    x=alt.X('year:O', axis=alt.Axis(title='Year')),
    y=alt.Y('total_gross:Q', axis=alt.Axis(title='Gross Revenue')),
    tooltip=['year', 'total_gross']
)

adjusted_gross_chart = alt.Chart(disney_annual_gross).mark_line(color='green').
 ↪encode(
    x='year:O',
    y=alt.Y('inflation_adjusted_gross:Q', axis=alt.Axis(title='Inflation␣
 ↪Adjusted Gross Revenue')),
    tooltip=['year', 'inflation_adjusted_gross']
)

# Combine the two charts
combined_chart = alt.layer(total_gross_chart, adjusted_gross_chart).
 ↪resolve_scale(
    y='independent'
).properties(title='Disney Movie Gross vs. Inflation Adjusted Revenues Over the␣
 ↪Years')

combined_chart
```

[33]:
```
alt.LayerChart(…)
```

- There are significant peaks on the graph, with the highest points occurring in the latter years. This might correlate with the release of highly successful movies or series of movies in those years
- The gap between nominal and inflation-adjusted gross revenues appears to be widening over

time, particularly from the 1990s onward. This widening gap may suggest that earlier movies have retained strong value when adjusted for inflation, indicating lasting popularity.

- There is a sharp increase in both nominal and inflation-adjusted revenues in the most recent years shown, which may reflect a period of particularly successful movie releases, possibly driven by successful franchises, new technologies in filmmaking, or marketing strategies.

7 What is the most common MPAA rating for Disney movies, and how do movies within each rating category perform financially?

[34]:
```
# Get the most common MPAA rating
mpaa_rating_counts = disney_movie_gross_director['MPAA_rating'].value_counts()

# Calculate the average financial performance by MPAA rating
average_financials_by_rating = disney_movie_gross_director.
 ↪groupby('MPAA_rating')[['total_gross', 'inflation_adjusted_gross']].mean()

print(mpaa_rating_counts)
print(average_financials_by_rating)
```

```
PG           187
PG-13        145
R            102
G             86
Not Rated      3
Name: MPAA_rating, dtype: int64
              total_gross  inflation_adjusted_gross
MPAA_rating
G            9.209061e+07              2.912610e+08
Not Rated    5.046259e+07              2.998734e+08
PG           7.362521e+07              1.015414e+08
PG-13        8.118074e+07              1.029486e+08
R            2.936536e+07              5.530581e+07
```

The data in the output above suggest the followig:

1. MPAA_rating Distribution - 'PG' (Parental Guidance) is the most common MPAA rating for Disney movies with 187 films, indicating Disney's strong focus on family-friendly content. - 'PG-13' (Parents Strongly Cautioned) is the next most common rating with 145 films, showing that Disney also produces content targeted at a slightly older audience. - 'R' (Restricted) rated films are less common with 102 films, which is interesting given that Disney is typically associated with content suitable for children and families. - 'G' (General Audiences) rated films are fewer with 86, which might suggest a shift over time to content that requires parental guidance, possibly due to changes in content standards or marketing strategies. - There are only 3 movies that are 'Not Rated', indicating that almost all Disney movies go through the MPAA rating process

2.Financial performance by Rating

- 'G' rated films have the highest average inflation-adjusted gross revenue, suggesting that these films have performed very well over time when accounting for inflation. This indicates that

Disney's 'G' rated content has had lasting appeal
- 'Not Rated' films, despite being few, have a high inflation-adjusted gross average similar to 'G' rated films
- 'PG' and 'PG-13' movies have a similar average total gross and inflation-adjusted gross, showing consistent performance in these categories. The 'PG' rating has a slightly lower average than 'PG-13', which might reflect the broader appeal of 'PG-13' movies to both younger and adult audience
- 'R' rated films have the lowest average gross and inflation-adjusted gross.

Overall, with the data at our disposal, it indicates that Disney's strength lies in producing 'PG' and 'G' rated films that cater to a family audience, and these films have historically had a strong financial performance. However, they also have a considerable number of 'PG-13' and 'R' rated films, which suggests a diverse portfolio targeting different age demographics.

To deleve deeper the analysis, let's try to understand the relationship between total_gross, inflation_adjusted_gross and the movies rating. this will help us visualize how Disney's movies perform financially within each MPAA rating category which can aid in understanding Disney's revenue distribution across different audience demographics.

[38]:
```
# first, we will filter out the rows where MPAA_rating is NaN or 'Not Rated'
disney_data_filtered = disney_movie_gross_director.
 ↪dropna(subset=['MPAA_rating'])
disney_data_filtered = disney_data_filtered[disney_data_filtered['MPAA_rating']␣
 ↪!= 'Not Rated']

# Create scatter plots for each MPAA rating category
scatter_plot = alt.Chart(disney_data_filtered).mark_circle(size=60).encode(
    x='total_gross',
    y='inflation_adjusted_gross',
    color='MPAA_rating',
    tooltip=['movie_title', 'total_gross', 'inflation_adjusted_gross',␣
 ↪'MPAA_rating']
).interactive().properties(
    title='Scatter Plot of Disney Movie Revenues by MPAA Rating'
)

scatter_plot
```

[38]:
```
alt.Chart(…)
```

Here are some insights for the graph above:

- There appears to be a positive correlation between total gross and inflation-adjusted gross across all MPAA ratings, which is expected since inflation-adjusted gross is derived from total gross adjusted for the time value of money.
- Most data points are concentrated in the lower left quadrant of the graph, suggesting that a majority of the movies have lower gross revenues. There are, however, a few outliers with significantly higher revenues, particularly in the 'G' and 'PG' categories.
- The graph shows that there are some exceptionally high-grossing movies in the 'G' and 'PG'

categories, which may include timeless classics or blockbuster hits

- There are fewer 'R' rated movies, and their gross revenue does not reach as high as the more family-oriented 'G' and 'PG' movies, which aligns with Disney's reputation as a creator of family-friendly content.
- The dense clustering of points at lower revenue levels across all ratings suggests that while Disney has had blockbuster hits, a significant number of their films achieve moderate financial success

Finally, Let's create a function of top performing Movies by gross that takes arguments:

1. data_path: A string representing the path to a CSV file containing Disney movie data. The CSV file is expected to have a column named 'gross' that contains the gross revenue for each movie.
2. n: An optional integer argument that specifies the number of top-performing movies to return based on their gross revenue. for this examples, we will return top 10 perfoming movies n=10.

[46]:
```
import pandas as pd

def top_performing_movies_by_gross(data_path, n=10):
    """
    Returns the top n performing movies by total_gross.

    Args:
    - data_path (str): The path to the CSV file containing the Disney movie␣
 ↪data.
    - n (int): The number of top-performing movies to return. Default is 10.

    - DataFrame: A DataFrame containing the top n performing movies sorted by␣
 ↪gross revenue.
                 The DataFrame will contain the same columns as the original␣
 ↪CSV file.
    """
    # Read the CSV file
    df = pd.read_csv(data_path)

    # Sort the DataFrame by the 'gross' column in descending order
    sorted_df = df.sort_values(by='total_gross', ascending=False)

    # Return the top n rows
    return sorted_df.head(n)

# Example usage
disney_data_path = "./disney_movie_gross_director.csv"
top_movies = top_performing_movies_by_gross(disney_data_path)
top_movies
```

[46]:
```
                          movie_title release_date      genre MPAA_rating  \
564  Star Wars Ep. VII: The Force Awakens   2015-12-18  Adventure       PG-13
```

```
524                          The Avengers   2012-05-04     Action       PG-13
578          Rogue One: A Star Wars Story   2016-12-16  Adventure       PG-13
571                          Finding Dory   2016-06-17  Adventure          PG
558                Avengers: Age of Ultron   2015-05-01     Action       PG-13
441  Pirates of the Caribbean: Dead Man'…   2006-07-07  Adventure       PG-13
179                          The Lion King   1994-06-15  Adventure           G
499                            Toy Story 3   2010-06-18  Adventure           G
532                             Iron Man 3   2013-05-03     Action       PG-13
569             Captain America: Civil War   2016-05-06     Action       PG-13

     total_gross  inflation_adjusted_gross      director  release_year  decade
564    936662225                 936662225           NaN          2015    2010
524    623279547                 660081224           NaN          2012    2010
578    529483936                 529483936           NaN          2016    2010
571    486295561                 486295561           NaN          2016    2010
558    459005868                 459005868           NaN          2015    2010
441    423315812                 544817142           NaN          2006    2000
179    422780140                 761640898  Roger Allers          1994    1990
499    415004880                 443408255           NaN          2010    2010
532    408992272                 424084233           NaN          2013    2010
569    408084349                 408084349           NaN          2016    2010
```

To Test if the function behaves as expected, we run units test to see if the function returns a pandas dataframe or the correct number of top peforming movies, we use two unit test below:

1. **test_return_type
2. **test_number_of_rows

[45]:
```python
import pandas as pd
import unittest

def top_performing_movies_by_gross(data_path, n=10):
    """
    Returns the top n performing movies by gross revenue.

    Args:
    - data_path (str): The path to the CSV file containing the Disney movie␣
 ↪data.
    - n (int): The number of top-performing movies to return.

    Returns:
    - DataFrame: A DataFrame containing the top n performing movies sorted by␣
 ↪gross revenue.
    """
    # Read the CSV file
    df = pd.read_csv(data_path)

    # Sort the DataFrame by the 'total_gross' column in descending order
```

```
    sorted_df = df.sort_values(by='total_gross', ascending=False)

    # Return the top n rows
    return sorted_df.head(n)

class TestTopPerformingMoviesByGross(unittest.TestCase):
    def test_return_type(self):
        disney_data_path = "./disney_movie_gross_director.csv"
        result = top_performing_movies_by_gross(disney_data_path)
        self.assertIsInstance(result, pd.DataFrame, "The function should return␣
 ↪a DataFrame.")

    def test_number_of_rows(self):
        disney_data_path = "./disney_movie_gross_director.csv"
        n = 5
        result = top_performing_movies_by_gross(disney_data_path, n)
        self.assertEqual(len(result), n, f"The function should return {n} rows.
 ↪")

if __name__ == "__main__":
    unittest.main(argv=[''], exit=False, verbosity=2, testRunner=unittest.
 ↪TextTestRunner())
```

```
..
----------------------------------------------------------------------
Ran 2 tests in 0.013s

OK
```

8 Conclusion

Our exploration of the Disney movie dataset has yielded several insightful observations about the company's film production and distribution strategies, as well as audience preferences over the years. Here are some key takeaways:

1. **Trend Analysis**: Disney's movie gross revenues have shown an overall upward trend, with some fluctuations from year to year. This suggests that Disney movies have been increasingly successful in generating revenue over time. The most recent years have seen a significant spike in gross revenue, which could be attributed to a variety of factors such as successful movie franchises or advancements in filmmaking and marketing strategies.
2. **Comparative Analysis**: When adjusting gross revenues for inflation, "**Snow White and the Seven Dwarfs**" emerges as the Disney movie with the highest gross revenue, highlighting its enduring popularity. The comparison between nominal and inflation-adjusted gross revenues over the years reveals a widening gap, indicating that earlier movies have retained strong value when adjusted for inflation.
3. **Categorical Analysis**: The most common MPAA rating for Disney movies is 'PG,' reflecting the company's focus on family-friendly content. However, there has been an increase

in ‘PG-13’ releases since the early 2000s, indicating a shift towards content for older audiences. Financially, ‘G’ rated films have the highest average inflation-adjusted gross revenue, suggesting that these films have had lasting appeal.

If time permitted, i could have explored other questions such as: if there is a correlation between the release date of a movies (such as a particular month or season) and its financial success or whether there is a pattern in the collaboration of directors and voice actors (I wanted to explore this questions but found out that both director and voice actors data is missing for nearly over 500 data point). an interesting question to explore is also looking at how diverse are the characters in Disney movies and how does it affect gross revenues.

In conclusion, Disney’s film production and distribution strategies have evolved over the years to cater to changing audience preferences, while maintaining a strong focus on family-friendly content.