

Exploratory Data Analysis of the Disney Datasets - What Makes Disney the Most Money?

Marina Galvao

Introduction

Questions of Interest

In my analysis, I will be investigating what makes a Disney movie successful.

My questions are:

1) On average, what genre of movies tends to have the highest inflation-adjusted gross? 2) Do the movies with the highest inflation-adjusted gross all come from the same genre? 3) Do the movies with the highest inflation-adjusted gross all come from the same director who has an in-depth understanding of the features that make for our most beloved characters? 4) Is the timeline under which a movie is released related to its inflation-adjusted gross?

This is interesting because it helps us understand which elements influence box office performance and it gives us an idea as to what degree these elements impact a movie's success.

Dataset Description

The dataset we are working with is from [Kaggle](#) and includes five tables containing information about Disney. We will be working with the `disney_movies_total_gross` table and the `disney-director` table.

- `disney_movies_total_gross.csv`: Listed here in the order of the columns, the file contains a column for the movie title, the movie release date, the movie genre, the movie MPAA rating, total gross value earned from the movie, and the inflation-adjusted total gross value earned (which we will be using as this measure is comparable across timelines).
- `disney-director.csv`: Listed here in the order of the columns, the file contains a column for the movie title, and a column for the director's name.

Methods and Results

```
In [1]: # First, let's import the libraries needed for the analysis

import pandas as pd
import altair as alt

# Now, let's import the required files

disney_gross = pd.read_csv('data/disney_movies_total_gross.csv')
disney_director = pd.read_csv('data/disney-director.csv')
```

Table 1. Disney Inflation Adjusted Gross File, First 5 Rows

```
In [2]: # Let's have a look at each file, starting with the disney_gross file's first 5 rows

disney_gross.head()
```

```
Out[2]:
```

	movie_title	release_date	genre	MPAA_rating	total_gross	inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	Dec 21, 1937	Musical	G	\$184,925,485	\$5,228,953,251
1	Pinocchio	Feb 9, 1940	Adventure	G	\$84,300,000	\$2,188,229,052
2	Fantasia	Nov 13, 1940	Musical	G	\$83,320,000	\$2,187,090,808
3	Song of the South	Nov 12, 1946	Adventure	G	\$65,000,000	\$1,078,510,579
4	Cinderella	Feb 15, 1950	Drama	G	\$85,000,000	\$920,608,730

Table 2. Summary of Disney Director File, First 5 Rows

```
In [3]: # Now the disney_director file

disney_director.head()
```

```
Out[3]:
```

	name	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

Here, we can observe that in the `disney_gross` file, the movie name is referred to as `"movie_title"`. On the `disney_director` file, however, the movie name is referred to as `"name"`

```
In [4]: # Now let's get some info about each file

# Starting with the disney_gross dataframe

disney_gross.info()
disney_gross.dtypes

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 579 entries, 0 to 578
Data columns (total 6 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   object
5   inflation_adjusted_gross 579 non-null   object
dtypes: object(6)
memory usage: 27.3+ KB
```

```
Out[4]: movie_title            object
release_date           object
genre                  object
MPAA_rating            object
total_gross            object
inflation_adjusted_gross object
dtype: object
```

The result shows us the column types in this dataframe are all objects, and that some columns have null values (genre and MPAA_rating).

```
In [5]: # Now the disney_director dataframe

disney_director.info()
disney_director.dtypes
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 56 entries, 0 to 55
Data columns (total 2 columns):
#   Column    Non-Null Count  Dtype
---  -
0   name      56 non-null     object
1   director  56 non-null     object
dtypes: object(2)
memory usage: 1.0+ KB
```

```
Out[5]: name      object
director object
dtype: object
```

Here we see we also have a dataframe with objects as the column types, but unlike the previous one, we do not see any null values.

To start off our analysis, I want to manipulate the data to help me answer the first question previously presented: 1) On average, what genre of movies tends to have the highest inflation-adjusted gross?

To do so, I need to group the `disney_gross` file by genre, and calculate the mean of the `inflation_adjusted_gross` for each genre. This way I can see which genre is most successful on average.

Since we saw some nan values are present in the genre column of `disney_gross`, we will need to get rid of these in order to answer our question.

Table 3. Summary of Genre and Mean Inflation-Adjusted Gross

```
In [6]: # Dropping nan values from disney_gross genre column as we don't want to include missing values
disney_gross = disney_gross.dropna(subset=['genre'])

# We first need to convert 'inflation_adjusted_gross' to numeric; we remove dollar signs and commas, then convert to float
disney_gross['inflation_adjusted_gross'] = disney_gross['inflation_adjusted_gross'].replace({'\': '', ',': ''}, regex=True).astype(float)

# Now, let's group by genre and calculate the mean inflation_adjusted_gross for each group, sorting from highest to lowest
# We will also reset the index and rename the column from 'inflation_adjusted_gross' to 'mean_inflation_adjusted_gross' for better readability
grouped_by_genre = disney_gross.groupby('genre')['inflation_adjusted_gross'].mean()
grouped_by_genre_df = grouped_by_genre.sort_values(ascending=False).reset_index()
grouped_by_genre_df = grouped_by_genre_df.rename(columns={'inflation_adjusted_gross': 'mean_inflation_adjusted_gross'})
grouped_by_genre_df
```

```
Out[6]:
```

	genre	mean_inflation_adjusted_gross
0	Musical	6.035979e+08
1	Adventure	1.903974e+08
2	Action	1.374734e+08
3	Thriller/Suspense	8.965379e+07
4	Comedy	8.466773e+07
5	Romantic Comedy	7.777708e+07
6	Western	7.381571e+07

	genre	mean_inflation_adjusted_gross
7	Drama	7.189302e+07
8	Concert/Performance	5.741084e+07
9	Black Comedy	5.224349e+07
10	Horror	2.341385e+07
11	Documentary	1.271803e+07

From looking at the table above, we can see the genre 'Musical' results in the highest inflation adjusted gross. Let's create a chart for this so we can have a better look.

```
In [7]: # Let's use Altair here and create a bar graph with Movie Genre in the x-axis and Mean Inflation-Adjusted Gross in $ in the y-axis

mean_inflation_chart = alt.Chart(grouped_by_genre_df).mark_bar().encode(
    x=alt.X(
        'genre:N',
        title='Movie Genre',
        sort='-y'
    ),
    y=alt.Y(
        'mean_inflation_adjusted_gross:Q',
        title='Mean Inflation-Adjusted Gross in $',
        sort='-x'
    )
).properties(
    width=500,
    height=300,
    title='Disney Movie Genre and Mean Inflation-Adjusted Gross'
)

mean_inflation_chart
```

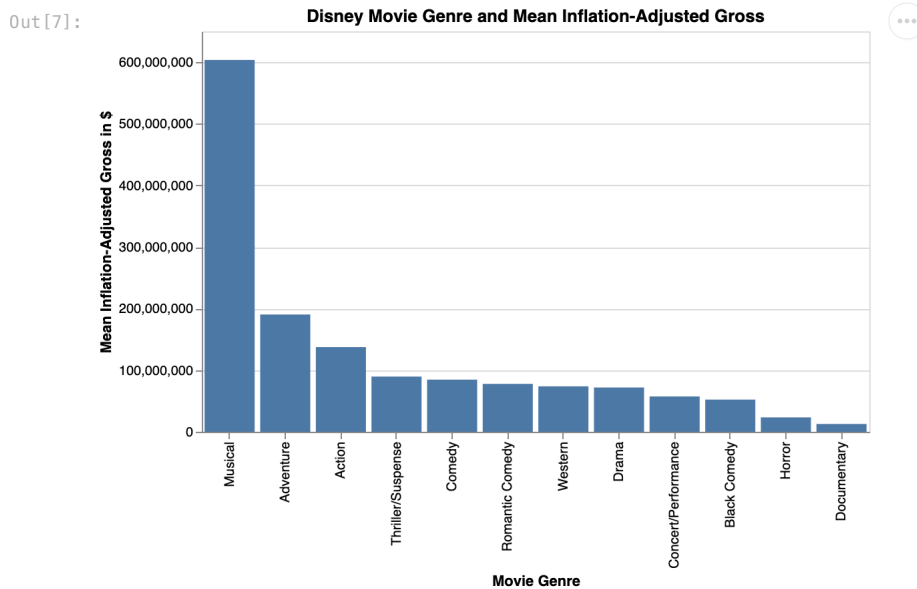


Figure 1. Disney Movie Genre and Mean Inflation-Adjusted Gross

There is an obvious discrepancy between musicals and other genres, with musicals having the highest mean inflation adjusted gross (followed by adventure, action, thriller/suspense, comedy, romantic comedy, western, drama, concert/performance, black comedy, horror, and documentary).

Now let's tackle our second question:

2) Do the movies with the highest inflation-adjusted gross all come from the same genre?

To understand this, we can order the disney_gross data from highest inflation-adjusted gross to lowest, and see if the highest profiting movies have a tendency to be musicals.

Table 4. Top 10 Movies with Highest Inflation-Adjusted Gross

```
In [8]: # Let's do this for the top 10 movies, sorting by inflation-adjusted gross

sorted_by_inflation = disney_gross.sort_values(by='inflation_adjusted_gross', ascending=False).head(10)
sorted_by_inflation
```

Out [8]:

	movie_title	release_date	genre	MPAA_rating	total_gross	inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	Dec 21, 1937	Musical	G	\$184,925,485	5.228953e+09
1	Pinocchio	Feb 9, 1940	Adventure	G	\$84,300,000	2.188229e+09

	movie_title	release_date	genre	MPAA_rating	total_gross	inflation_adjusted_gross
2	Fantasia	Nov 13, 1940	Musical	G	\$83,320,000	2.187091e+09
8	101 Dalmatians	Jan 25, 1961	Comedy	G	\$153,000,000	1.362871e+09
6	Lady and the Tramp	Jun 22, 1955	Drama	G	\$93,600,000	1.236036e+09
3	Song of the South	Nov 12, 1946	Adventure	G	\$65,000,000	1.078511e+09
564	Star Wars Ep. VII: The Force Awakens	Dec 18, 2015	Adventure	PG-13	\$936,662,225	9.366622e+08
4	Cinderella	Feb 15, 1950	Drama	G	\$85,000,000	9.206087e+08
13	The Jungle Book	Oct 18, 1967	Musical	Not Rated	\$141,843,000	7.896123e+08
179	The Lion King	Jun 15, 1994	Adventure	G	\$422,780,140	7.616409e+08

We can see that out of the ten rows, three are musicals (the 1st, 3rd, and 9th row). Since there is not an obvious trend towards musicals being the most profitable here, it may be that some values such as the first movie and third movie are skewing the mean from the data, and that being a musical is not the key to being successful as a movie.

```
In [9]: # Let's graph this data as a bar graph to see what this might show us

sorted_by_inflation_chart = alt.Chart(sorted_by_inflation,
).mark_bar().encode(x=alt.X('movie_title:N',
title='Movie Title', sort='-y'), y=alt.Y('inflation_adjusted_gross:Q',
title='Inflation-Adjusted Gross in $', sort='-x'
)).properties(width=500, height=300, title='Disney Movie and Inflation-Adjusted Gross')

sorted_by_inflation_chart
```

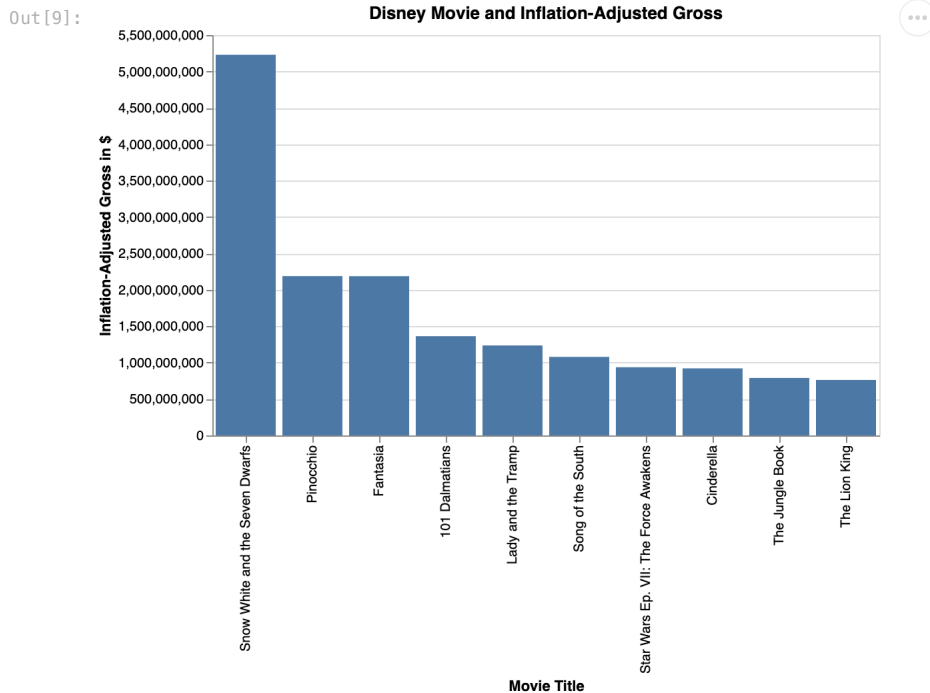


Figure 2. Disney Movie and Inflation-Adjusted Gross

The graph shows Snow White and the Seven Dwarfs with the highest value (5.228953e+09) followed by Pinocchio, Fantasia, 101 Dalmatians, Lady and the Tramp, Song of the South, Star Wars Ep. VII: The Force Awakens, Cinderella, The Jungle Book, and The Lion King. The result shows just how significant the difference is between Snow White and the Seven Dwarfs and other movies.

Table 5. Disney Director Table, Column Renamed

Now let's tackle our third question:

3) Do the movies with the highest inflation-adjusted gross all come from the same director who has an in-depth understanding of the features that make for our most beloved characters?

```
In [10]: # In order to answer this, let's merge the disney_director dataset with the disney_gross dataset using the name of the movies
# In doing so, we can see if the movies with the highest inflation-adjusted gross belong to the same directors

# But first, let's re-name the column 'name' to 'movie_title' in the disney_director dataframe,
# so that we can merge the data without any issues

disney_director = disney_director.rename(columns={'name': 'movie_title'})

disney_director
```

Out[10]:

	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
5	Saludos Amigos	Jack Kinney
6	The Three Caballeros	Norman Ferguson
7	Make Mine Music	Jack Kinney
8	Fun and Fancy Free	Jack Kinney
9	Melody Time	Clyde Geronimi
10	The Adventures of Ichabod and Mr. Toad	Jack Kinney
11	Cinderella	Wilfred Jackson
12	Alice in Wonderland	Clyde Geronimi
13	Peter Pan	Hamilton Luske
14	Lady and the Tramp	Hamilton Luske
15	Sleeping Beauty	Clyde Geronimi
16	101 Dalmatians	Wolfgang Reitherman
17	The Sword in the Stone	Wolfgang Reitherman
18	The Jungle Book	Wolfgang Reitherman
19	The Aristocats	Wolfgang Reitherman
20	Robin Hood	Wolfgang Reitherman
21	The Many Adventures of Winnie the Pooh	Wolfgang Reitherman
22	The Rescuers	Wolfgang Reitherman
23	The Fox and the Hound	Art Stevens
24	The Black Cauldron	Ted Berman
25	The Great Mouse Detective	Ron Clements
26	Oliver & Company	George Scribner
27	The Little Mermaid	Ron Clements
28	The Rescuers Down Under	Mike Gabriel
29	Beauty and the Beast	Gary Trousdale
30	Aladdin	Ron Clements
31	The Lion King	Roger Allers
32	Pocahontas	Mike Gabriel
33	The Hunchback of Notre Dame	Gary Trousdale
34	Hercules	Ron Clements
35	Mulan	Barry Cook
36	Tarzan	Chris Buck
37	Fantasia 2000	full credits
38	Dinosaur	Ralph Zondag
39	The Emperor's New Groove	Mark Dindal
40	Atlantis: The Lost Empire	Gary Trousdale
41	Lilo & Stitch	Chris Sanders
42	Treasure Planet	Ron Clements
43	Brother Bear	Robert Walker
44	Home on the Range	Will Finn
45	Chicken Little	Mark Dindal
46	Meet the Robinsons	Stephen J. Anderson
47	Bolt	Chris Williams
48	The Princess and the Frog	Ron Clements
49	Tangled	Nathan Greno
50	Winnie the Pooh	Stephen J. Anderson
51	Wreck-It Ralph	Rich Moore
52	Frozen	Chris Buck

	movie_title	director
53	Big Hero 6	Don Hall
54	Zootopia	Byron Howard
55	Moana	Ron Clements

Table 6. Data for Inflation-Adjusted Gross and Movie Director Merged

In [11]: *# Now, we can go ahead and merge*

```
combined_df = pd.merge(disney_director, disney_gross, on='movie_title', how='inner').sort_values(by='inflation_adjusted_gross',
combined_df
```

Out [11]:	movie_title	director	release_date	genre	MPAA_rating	total_gross	inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	David Hand	Dec 21, 1937	Musical	G	\$184,925,485	5.228953e+09
1	Pinocchio	Ben Sharpsteen	Feb 9, 1940	Adventure	G	\$84,300,000	2.188229e+09
2	Fantasia	full credits	Nov 13, 1940	Musical	G	\$83,320,000	2.187091e+09
8	101 Dalmatians	Wolfgang Reitherman	Jan 25, 1961	Comedy	G	\$153,000,000	1.362871e+09
6	Lady and the Tramp	Hamilton Luske	Jun 22, 1955	Drama	G	\$93,600,000	1.236036e+09
3	Cinderella	Wilfred Jackson	Feb 15, 1950	Drama	G	\$85,000,000	9.206087e+08
11	The Jungle Book	Wolfgang Reitherman	Oct 18, 1967	Musical	Not Rated	\$141,843,000	7.896123e+08
24	The Lion King	Roger Allers	Jun 15, 1994	Adventure	G	\$422,780,140	7.616409e+08
23	Aladdin	Ron Clements	Nov 11, 1992	Comedy	G	\$217,350,219	4.419692e+08
44	Frozen	Chris Buck	Nov 22, 2013	Adventure	PG	\$400,738,009	4.149972e+08
13	The Jungle Book	Wolfgang Reitherman	Apr 15, 2016	Adventure	PG	\$364,001,123	3.640011e+08
22	Beauty and the Beast	Gary Trousdale	Nov 13, 1991	Musical	G	\$218,951,625	3.630177e+08
5	Alice in Wonderland	Clyde Geronimi	Mar 5, 2010	Adventure	PG	\$334,191,110	3.570635e+08
46	Zootopia	Byron Howard	Mar 4, 2016	Adventure	PG	\$341,268,248	3.412682e+08
29	Tarzan	Chris Buck	Jun 16, 1999	Adventure	G	\$171,091,819	2.839003e+08
25	Pocahontas	Mike Gabriel	Jun 10, 1995	Adventure	G	\$141,579,773	2.743710e+08
9	101 Dalmatians	Wolfgang Reitherman	Nov 27, 1996	Comedy	G	\$136,189,294	2.587289e+08
14	The Aristocats	Wolfgang Reitherman	Apr 24, 1970	Musical	G	\$55,675,257	2.551615e+08
47	Moana	Ron Clements	Nov 23, 2016	Adventure	PG	\$246,082,029	2.460820e+08
45	Big Hero 6	Don Hall	Nov 7, 2014	Adventure	PG	\$222,527,828	2.292492e+08
20	The Little Mermaid	Ron Clements	Nov 15, 1989	Adventure	G	\$111,543,479	2.237260e+08
28	Mulan	Barry Cook	Jun 19, 1998	Adventure	G	\$120,620,254	2.168078e+08
30	Dinosaur	Ralph Zondag	May 19, 2000	Adventure	PG	\$137,748,063	2.154390e+08
41	Tangled	Nathan Greno	Nov 24, 2010	Adventure	PG	\$200,821,936	2.143885e+08
33	Lilo & Stitch	Chris Sanders	Jun 21, 2002	Adventure	PG	\$145,771,527	2.115067e+08
4	Cinderella	Wilfred Jackson	Mar 13, 2015	Drama	PG	\$201,151,353	2.011514e+08
43	Wreck-It Ralph	Rich Moore	Nov 2, 2012	Adventure	PG	\$189,412,677	2.003550e+08
26	The Hunchback of Notre Dame	Gary Trousdale	Jun 21, 1996	Adventure	G	\$100,138,851	1.909888e+08
27	Hercules	Ron Clements	Jun 13, 1997	Adventure	G	\$99,112,101	1.820294e+08
37	Chicken Little	Mark Dindal	Nov 4, 2005	Adventure	G	\$135,386,665	1.779547e+08
15	The Rescuers	Wolfgang Reitherman	Jun 22, 1977	Adventure	NaN	\$48,775,599	1.597439e+08
10	The Sword in the Stone	Wolfgang Reitherman	Dec 25, 1963	Adventure	NaN	\$22,182,353	1.538708e+08
31	The Emperor's New Groove	Mark Dindal	Dec 15, 2000	Adventure	G	\$89,296,573	1.367893e+08
39	Bolt	Chris Williams	Nov 21, 2008	Comedy	PG	\$114,053,759	1.337025e+08
16	The Fox and the Hound	Art Stevens	Jul 10, 1981	Comedy	NaN	\$43,899,231	1.331189e+08
32	Atlantis: The Lost Empire	Gary Trousdale	Jun 8, 2001	Adventure	PG	\$84,052,762	1.251881e+08
38	Meet the Robinsons	Stephen J. Anderson	Mar 30, 2007	Adventure	G	\$97,822,171	1.198606e+08
35	Brother Bear	Robert Walker	Oct 24, 2003	Adventure	G	\$85,336,277	1.192183e+08
40	The Princess and the Frog	Ron Clements	Nov 25, 2009	Adventure	G	\$104,400,899	1.163165e+08
19	Oliver & Company	George Scribner	Nov 18, 1988	Adventure	G	\$49,576,671	1.022545e+08
12	The Jungle Book	Wolfgang Reitherman	Dec 25, 1994	Adventure	PG	\$44,342,956	8.893032e+07
36	Home on the Range	Will Finn	Apr 2, 2004	Comedy	PG	\$50,026,353	6.791017e+07
21	The Rescuers Down Under	Mike Gabriel	Nov 16, 1990	Adventure	G	\$27,931,461	5.579673e+07
34	Treasure Planet	Ron Clements	Nov 27, 2002	Adventure	PG	\$38,120,554	5.518914e+07

	movie_title	director	release_date	genre	MPAA_rating	total_gross	inflation_adjusted_gross
18	The Great Mouse Detective	Ron Clements	Jul 2, 1986	Adventure	NaN	\$23,605,534	5.363737e+07
17	The Black Cauldron	Ted Berman	Jul 24, 1985	Adventure	NaN	\$21,288,692	5.055314e+07
42	Winnie the Pooh	Stephen J. Anderson	Jul 15, 2011	Adventure	G	\$26,692,846	2.837587e+07
7	Sleeping Beauty	Clyde Geronimi	Jan 29, 1959	Drama	NaN	\$9,464,608	2.150583e+07

There doesn't seem to be an obvious pattern with a specific director leading to the most profit. In fact, we only see one director repeated in the top 10 movies (Wolfgang Reitherman).

Let's use a function previously created to look into this further. The function will take in a dataframe, a column to group by, and a column to sum. After running it, it will return a dataframe with the grouped by column and the summed column. In this case, I want to use it to group by the 'director', and to sum the 'inflation_adjusted_gross'. This will help us see which directors contribute most to inflation-adjusted gross.

```
In [12]: # Let's start by importing the script

from script1 import group_and_sum

# Let's run black on our script to make sure it is up to standards

!black script1.py
```

All done! 🌟 🍰 🌟
1 file left unchanged.

Table 7. Data Grouped by Director with Inflation-Adjusted Gross Summed

```
In [13]: # Now let's use the script and ask it to group by the 'director' column and sum the 'inflation_adjusted_gross' column

director_total_gross = group_and_sum (combined_df, 'director', 'inflation_adjusted_gross')
director_total_gross
```

```
Out [13]:
```

	inflation_adjusted_gross
director	
David Hand	5.228953e+09
Wolfgang Reitherman	3.432920e+09
Ben Sharpsteen	2.188229e+09
full credits	2.187091e+09
Ron Clements	1.318950e+09
Hamilton Luske	1.236036e+09
Wilfred Jackson	1.121760e+09
Roger Allers	7.616409e+08
Chris Buck	6.988974e+08
Gary Trousdale	6.791946e+08
Clyde Geronimi	3.785693e+08
Byron Howard	3.412682e+08
Mike Gabriel	3.301677e+08
Mark Dindal	3.147439e+08
Don Hall	2.292492e+08
Barry Cook	2.168078e+08
Ralph Zondag	2.154390e+08
Nathan Greno	2.143885e+08
Chris Sanders	2.115067e+08
Rich Moore	2.003550e+08
Stephen J. Anderson	1.482365e+08
Chris Williams	1.337025e+08
Art Stevens	1.331189e+08
Robert Walker	1.192183e+08
George Scribner	1.022545e+08
Will Finn	6.791017e+07
Ted Berman	5.055314e+07

We see that although David Hand only showed up once on Table 6, he was still the director with the highest inflation-adjusted gross due to his direction of the Snow White movie.

```
In [14]: # Now let's quickly test script2 to make sure it doesn't have any failures
```

```
!pytest script2.py
```

```
===== test session starts =====
platform linux -- Python 3.8.5, pytest-6.2.4, py-1.10.0, pluggy-0.13.1
rootdir: /home/jupyter/prog-python-ds-students/release/final_project
plugins: anyio-3.2.1, dash-1.20.0
collected 2 items
```

```
script2.py ..
```

```
[100%]
```

```
===== 2 passed in 0.72s =====
```

What if the timeline under which the movie is released is related to its success? Maybe over the years, Disney movies have gained or lost popularity. Let's test this out and try to answer our last question:

4) Is the timeline under which a movie is released related to its inflation-adjusted gross?

```
In [15]: # First, let's separate the 'release_date' column so we can have the year be its own column
```

```
dates = (combined_df['release_date'].str.split('-', expand=True)
         .rename(columns = {0:'Date',
                           1:'Year'}))
```

```
# Now let's merge the new dataframe with the separated year column to our 'combined_df' previously created, sorting it by year
```

```
dates_concatenated = pd.concat([combined_df.drop('release_date', axis=1), dates], axis=1).sort_values(by='Year')
dates_concatenated
```

```
# Let's create an Altair scatterplot to see what this tells us
```

```
scatter_plot_years = alt.Chart(dates_concatenated).mark_point().encode(
    x=alt.X(
        'Year:0',
        title='Movie Release Year',
        sort='ascending'
    ),
    y=alt.Y(
        'inflation_adjusted_gross:Q',
        title='Inflation-Adjusted Gross in $',
        sort='-y'
    )
).properties(
    width=500,
    height=500,
    title='Movie Inflation-Adjusted Gross and Release Year'
)
```

```
scatter_plot_years
```

```
Out[15]:
```

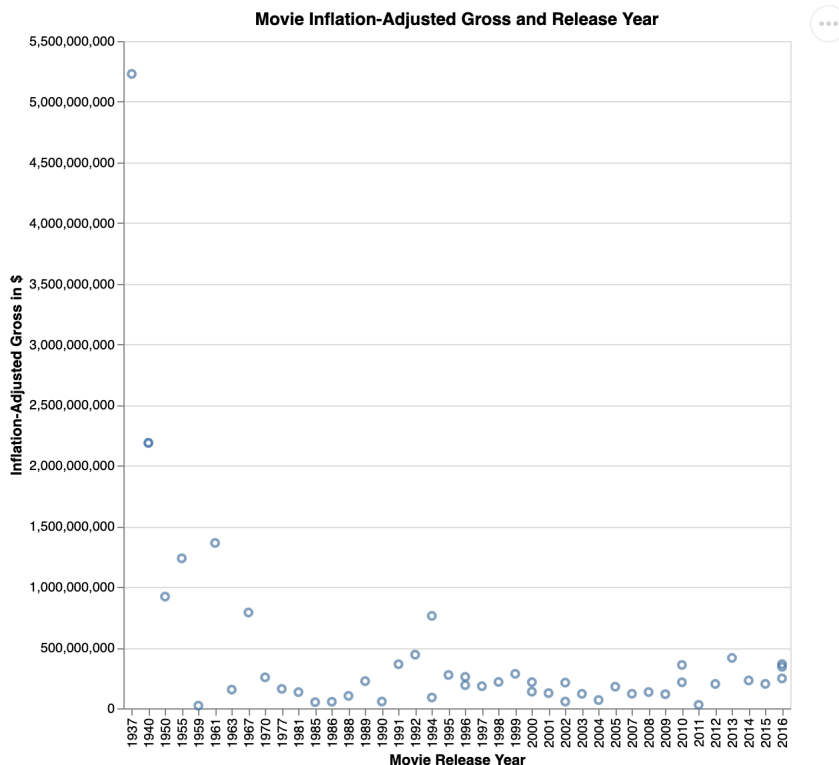


Figure 3. Movie Inflation-Adjusted Gross and Release Year

Figure 3 shows us that there is a somewhat general trend over time, except for a few outliers. The outlier that shows the most difference from the general trend is Snow White and the Seven Dwarves, released in 1937 and with an inflation-adjusted gross of 5.228953e+09. The second outlier is Pinocchio, released in 1940 and with an inflation-adjusted gross of 2.188229e+09. Once again, the data shows that Snow White and the Seven Dwarves has significantly different inflation-adjusted gross values than the rest of the movies. The overall trend shows these outliers decreasing and from 1995 onwards, all of the movies fall under the first line on the plot (between 0 and 500,000,000).

Discussion

Let's tackle each of the questions previously mentioned:

1) On average, what genre of movies tends to have the highest inflation-adjusted gross?

My findings indicate that on average, musicals tend to make the most money. The analysis shows that the Snow White and the Seven Dwarfs musical is the movie with the highest inflation-adjusted gross of all time. It is important to mention that this movie is likely a significant reason as to why musicals are the most profitable genre on average. Snow White and the Seven Dwarfs is seen as a movie that is highly differentiated in terms of inflation-adjusted gross when compared to others (see Figures 2 and 3). The inflation-adjusted gross is 5.228953e+09 which is more than half than the second most-profitable movie (Pinocchio, with an inflation-adjusted gross of 2.188229e+09).

2) Do the movies with the highest inflation-adjusted gross all come from the same genre?

The analysis showed that the movies with the highest inflation-adjusted gross are not all musicals. In fact, only three out of the top ten movies with the highest inflation-adjusted gross are musicals (Table 4).

3) Do the movies with the highest inflation-adjusted gross all come from the same director who has an in-depth understanding of the features that make for our most beloved characters?

The analysis implied that this is not the case. It was observed that the same director did not appear numerous times throughout the movies with highest inflation-adjusted gross (Table 6). However, Table 7 showed that even though David Hand only appeared once on the dataset (directing only one movie), he was still the director with the highest inflation-adjusted gross due to his direction of the Snow White movie.

4) Is the timeline under which a movie is released related to its inflation-adjusted gross?

It appears that there are outlier movies that have high inflation-adjusted gross in comparison to other movies. These took place before 1995 (Figure 3).

What I expected

I expected musicals to be the most popular as this category tends to be the most associated with successful Disney films (such as Frozen or The Lion King). Alongside this, I expected the highest inflation-adjusted gross movies to all be musicals. I expected to see the same director come up under the list of successful films (such as Tim Burton, who has directed many incredibly successful movies). I found it interesting that there are numerous directors and no specific pattern. I expected that the year a movie was released would not have any relation to the inflation-adjusted gross, however, Figure 3 shows inflation-adjusted gross decrease over time, with some outliers displaying very high inflation-adjusted gross in the past.

What my findings mean

My findings show that besides musicals being most popular on average, it is not the only successful genre. This implies other genres possess other factors (perhaps the story line or different animation styles) that are also very appealing to the public. The variety in directors implies Disney does not tend to stick to one director for numerous movies, and that there is diversity in leadership at Disney. The timeline's relationship with inflation-adjusted gross could be caused by the fact that classic movies (such as Snow White and the Seven Dwarfs) remained popular over time, adding to the overall cumulative inflation-adjusted gross. Moreover, other platforms growing over time (i.e. Netflix) could explain decreased interest in Disney with time, as other platforms increase in popularity.

What I would like to learn about

Having performed this analysis, I would like to learn more about Disney's marketing strategies and how these may influence how successful movie releases are. I would also like to better understand how measuring inflation-adjusted gross might have changed over time, as this may account for the changes seen in Figure 3.

References

[Kaggle Walt Disney Dataset](#) - last updated on Kaggle three years ago.