final_project

November 6, 2023

1 Exploratory data analysis of Disney datasets

1.1 Author: Gloria Mo

1.2 Introduction

1.2.1 Question of Interest

I will be investigating the relationship between a Disney movie and its success with having a theme song. I am interested to see if Disney movies with a theme song perform better than Disney movies without a theme song in terms of inflation adjusted gross. I expect movies with a theme song to have higher inflation adjusted gross than Disney movies without a theme song. This inference is because the general audience of Disney movies are usually younger in age, thus, I believe having a theme song would be more entertaining and can also increase the popularity of the movie if the song is successful. To add on, I believe a big part that makes a Disney movie "Disney" is if the movie can feel fantastical and magical for the viewers and I believe a musical production contributes highly to making Disney movies successful.

Question: Do Disney movies that feature a theme song generate higher inflation adjusted gross than Disney movies without a theme song?

1.2.2 Dataset Description

The dataset used for Disney movies is taken from this website:

The Disney dataset is composed of 5 tables: disney_movies_total_gross.csv, disney_revenue_1991-2006.csv, disney-characters.csv, disney-director.csv, disney-voice-actors.csv. Each table is stored in a `.csv` file and contains different information about Disney movies including the total gross, total revenue, characters, directors, and voice actors.

The datasets I used in this analysis are: * **disney-characters.csv** * This file contains information on Disney movies such as the movie title, the release date, the villian and heros involved in the storyline, and the song title if the movie has one. * **disney_movies_total_gross.csv** * This file contains information on Disney movies such as release date, genre, MPAA rating, total gross, and inflation adjusted gross.

1.3 Methods and Results

I will only be using the disney characters and total gross tables as all the information I require to answer my question can be answered with this data.

First we will import in all our data and visualize how the tables look:

[1]:
```
# import all required libraries
import altair as alt
import pandas as pd
import numpy as np

# import all the required files
characters = pd.read_csv('data/disney-characters.csv')
gross = pd.read_csv("data/disney_movies_total_gross.csv")
```

Let's take a look at the tables:

[2]:
```
characters.head()
```

[2]:

	movie_title	release_date	hero	villian	song
0	\nSnow White and the Seven Dwarfs	December 21, 1937	Snow White	Evil Queen	Some Day My Prince Will Come
1	\nPinocchio	February 7, 1940	Pinocchio	Stromboli	When You Wish upon a Star
2	\nFantasia	November 13, 1940	NaN	Chernabog	NaN
3	Dumbo	October 23, 1941	Dumbo	Ringmaster	Baby Mine
4	\nBambi	August 13, 1942	Bambi	Hunter	Love Is a Song

[3]:
```
gross.head()
```

[3]:

	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

Since I am interested in the variable inflation_adjusted_gross, I would like to eventually sort this value in descending order to compare different relationships. First, I will remove the '$' in the column so that inflation adjusted gross can be properly sorted.

[4]:
```
# Remove the $ so that inflation adjusted gross can be sorted properly
gross["inflation_adjusted_gross"] = gross["inflation_adjusted_gross"].
 ↪replace("[$,]", "", regex=True).astype(int).round(8)
gross = gross.sort_values(by=['inflation_adjusted_gross'], kind="mergesort",␣
 ↪ascending=False)
gross_clean = gross.dropna(subset=['inflation_adjusted_gross'])

gross_clean.head()
```

[4]:

	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	5228953251
1	Pinocchio	Feb 9, 1940	Adventure	G	$84,300,000	2188229052
2	Fantasia	Nov 13, 1940	Musical	G	$83,320,000	2187090808
8	101 Dalmatians	Jan 25, 1961	Comedy	G	$153,000,000	1362870985
6	Lady and the Tramp	Jun 22, 1955	Drama	G	$93,600,000	1236035515

To begin, I want to have a dataframe that shows only movies **with** a theme song. In the 'characters' dataframe with 56 disney movies, only 45 movies have theme songs.

[5]:
```
# Clean up blank spaces in movie_title column
characters_clean = characters.assign(movie_title = characters['movie_title'].
 ↪str.strip())
characters_clean = characters_clean.assign(song = characters_clean['song'].str.
 ↪strip())
characters_clean = characters_clean.replace(r'^\s*$', np.nan, regex=True)

# Drop all movies with no theme song
characters_song = characters_clean.dropna(subset=['song']).
 ↪reset_index(drop=True)
characters_song
```

[5]:

	movie_title	release_date
0	Snow White and the Seven Dwarfs	December 21, 1937
1	Pinocchio	February 7, 1940
2	Dumbo	October 23, 1941
3	Bambi	August 13, 1942
4	Saludos Amigos	February 6, 1943
5	Melody Time	May 27, 1948
6	The Adventures of Ichabod and Mr. Toad	October 5, 1949
7	Cinderella	February 15, 1950

```
8                      Alice in Wonderland      July 28, 1951
9                                Peter Pan   February 5, 1953
10                     Lady and the Tramp      June 22, 1955
11                        Sleeping Beauty   January 29, 1959
12         One Hundred and One Dalmatians   January 25, 1961
13                 The Sword in the Stone  December 25, 1963
14                       The Jungle Book   October 18, 1967
15                        The Aristocats  December 24, 1970
16                            Robin Hood   November 8, 1973
17  The Many Adventures of Winnie the Pooh     March 11, 1977
18                             The Rescuers      June 22, 1977
19                   The Fox and the Hound      July 10, 1981
20               The Great Mouse Detective       July 2, 1986
21                        Oliver & Company  November 18, 1988
22                     The Little Mermaid  November 17, 1989
23                   Beauty and the Beast  November 22, 1991
24                                Aladdin  November 25, 1992
25                            The Lion King      June 24, 1994
26                               Pocahontas      June 23, 1995
27            The Hunchback of Notre Dame      June 21, 1996
28                                Hercules      June 27, 1997
29                                   Mulan      June 19, 1998
30                                  Tarzan      June 18, 1999
31                The Emperor's New Groove  December 15, 2000
32               Atlantis: The Lost Empire      June 15, 2001
33                            Lilo & Stitch      June 21, 2002
34                         Treasure Planet  November 27, 2002
35                            Brother Bear   November 1, 2003
36                     Meet the Robinsons     March 30, 2007
37                                   Bolt  November 21, 2008
38               The Princess and the Frog  December 11, 2009
39                                Tangled  November 24, 2010
40                         Winnie the Pooh      July 15, 2011
41                          Wreck-It Ralph   November 2, 2012
42                                  Frozen  November 27, 2013
43                              Big Hero 6   November 7, 2014
44                                Zootopia      March 4, 2016
45                                   Moana  November 23, 2016

                          hero                                villian  \
0                  Snow White                             Evil Queen
1                   Pinocchio                              Stromboli
2                       Dumbo                             Ringmaster
3                       Bambi                                 Hunter
4                  Donald Duck                                    NaN
5                          NaN                                    NaN
6   Mr. Toad and Ichabod Crane  Mr. Winkie and The Headless Horseman
```

7	Cinderella	Lady Tremaine
8	Alice	Queen of Hearts
9	Peter Pan	Captain Hook
10	Lady and Tramp	Si and Am
11	Aurora	Maleficent
12	Pongo	Cruella de Vil
13	Arthur	Madam Mim
14	Mowgli	Kaa and Shere Khan
15	Thomas and Duchess	Edgar Balthazar
16	Robin Hood	Prince John
17	Winnie the Pooh	NaN
18	Bernard and Miss Bianca	Madame Medusa
19	Tod and Copper	Amos Slade
20	Basil	Professor Ratigan
21	Oliver	Sykes
22	Ariel	Ursula
23	Belle	Gaston
24	Aladdin	Jafar
25	Simba	Scar
26	Pocahontas	Governor Ratcliffe
27	Quasimodo	Claude Frollo
28	Hercules	Hades
29	Mulan	Shan Yu
30	Tarzan	Clayton
31	Kuzco	Yzma
32	Milo Thatch	Commander Rourke
33	Lilo and Stitch	NaN
34	Jim Hawkins	John Silver
35	Kenai	Denahi
36	Lewis	Doris
37	Bolt	Dr. Calico
38	Tiana	Dr. Facilier
39	Rapunzel	Mother Gothel
40	Winnie the Pooh	NaN
41	Ralph	Turbo
42	Elsa	Prince Hans
43	Hiro Hamada	Professor Callaghan
44	Judy Hopps	NaN
45	Moana	NaN

	song
0	Some Day My Prince Will Come
1	When You Wish upon a Star
2	Baby Mine
3	Love Is a Song
4	Saludos Amigos
5	Little Toot

```
6                     The Merrily Song
7                   Bibbidi-Bobbidi-Boo
8                   The Unbirthday Song
9                          You Can Fly!
10                          Bella Notte
11                    Once Upon a Dream
12                       Cruella De Vil
13                      Higitus Figitus
14                 The Bare Necessities
15         Ev'rybody Wants to Be a Cat
16                           Oo De Lally
17                      Winnie the Pooh
18                          The Journey
19                      Best of Friends
20  The World's Greatest Criminal Mind
21   Once Upon a Time in New York City
22                        Under the Sea
23                         Be Our Guest
24                   A Whole New World
25                      Circle of Life
26                  Colors of the Wind
27                God Help the Outcasts
28                      Go the Distance
29          I'll Make a Man Out of You
30               You'll Be in My Heart
31              My Funny Friend and Me
32           Where the Dream Takes You
33                      He Mele No Lilo
34                       I'm Still Here
35                 Look Through My Eyes
36                       Little Wonders
37                 I Thought I Lost You
38                          Almost There
39                      I See the Light
40                      Winnie the Pooh
41                           Sugar Rush
42                            Let It Go
43                            Immortals
44                       Try Everything
45                      How Far I'll Go
```

Next I want a dataframe showcasing only movies **without** a theme song. There are 10 movies with no theme songs in the 'characters' table.

[8]:
```
# drop all movies with a theme song
characters_no_song = characters_clean[characters_clean['song'].isnull()]
characters_no_song = characters_no_song.reset_index(drop=True)
```

```
characters_no_song
```

[8]:

	movie_title	release_date	hero	villian	song
0	Fantasia	November 13, 1940	NaN	Chernabog	NaN
1	The Three Caballeros	February 3, 1945	Donald Duck	NaN	NaN
2	Make Mine Music	April 20, 1946	NaN	NaN	NaN
3	Fun and Fancy Free	September 27, 1947	Mickey Mouse	Willie the Giant	NaN
4	The Black Cauldron	July 24, 1985	Taran	Horned King	NaN
5	The Rescuers Down Under	November 16, 1990	Bernard and Miss Bianca	Percival C. McLeach	NaN
6	Fantasia 2000	December 17, 1999	NaN	NaN	NaN
7	Dinosaur	May 19, 2000	Aladar	Kron	NaN
8	Home on the Range	April 2, 2004	Maggie	Alameda Slim	NaN
9	Chicken Little	November 4, 2005	Ace Cluck	Foxy Loxy	NaN

With the two tables of movies with theme songs and without theme songs, I am able to compare the relationship with inflation adjusted gross between the two tables.

As a first visualization, I want to look at how movies with theme songs perform in terms of inflation adjusted gross. To do this, I will use the movies with songs table and merge it with the gross table to find movies that are included in both dataframes. I will use the 'movie_title' column as the common key column to merge the two dataframes on.

[9]:
```
# Merge movies with songs with gross df with all columns showing
characters_song_gross_merged = characters_song.merge(gross_clean,␣
↪left_on='movie_title', right_on='movie_title', how='left')
characters_song_gross_merged.head()
```

[9]:

	movie_title	release_date_x	hero	villian	song	release_date_y	genre
0	Snow White and the Seven Dwarfs	December 21, 1937	Snow White	Evil Queen	Some Day My Prince Will Come	Dec 21, 1937	Musical
1	Pinocchio	February 7, 1940	Pinocchio				
2	Dumbo	October 23, 1941	Dumbo				
3	Bambi	August 13, 1942	Bambi				
4	Saludos Amigos	February 6, 1943	Donald Duck				

```
1   Stromboli     When You Wish upon a Star    Feb 9, 1940  Adventure
2  Ringmaster                    Baby Mine            NaN         NaN
3      Hunter               Love Is a Song            NaN         NaN
4         NaN               Saludos Amigos            NaN         NaN

  MPAA_rating   total_gross  inflation_adjusted_gross
0           G  $184,925,485              5.228953e+09
1           G   $84,300,000              2.188229e+09
2         NaN           NaN                       NaN
3         NaN           NaN                       NaN
4         NaN           NaN                       NaN
```

[10]:
```
# Sort in descending order of inflation adjusted gross and display the top 20
top_20_song_gross_merged = characters_song_gross_merged.
 ↪sort_values(by=['inflation_adjusted_gross'], ascending=False)[:20]
top_20_song_gross_merged = top_20_song_gross_merged.reset_index(drop=True)

# Drop row 8 as The Jungle Book has a duplicate entry as one was released in␣
 ↪2016
top_20_song_gross_merged = top_20_song_gross_merged.drop([8,8]).
 ↪reset_index(drop=True)
top_20_song_gross_merged
```

[10]:
```
                        movie_title     release_date_x                hero  \
0   Snow White and the Seven Dwarfs  December 21, 1937          Snow White
1                          Pinocchio   February 7, 1940           Pinocchio
2                Lady and the Tramp      June 22, 1955      Lady and Tramp
3                         Cinderella  February 15, 1950          Cinderella
4                    The Jungle Book   October 18, 1967              Mowgli
5                      The Lion King      June 24, 1994               Simba
6                            Aladdin  November 25, 1992             Aladdin
7                             Frozen  November 27, 2013                Elsa
8               Beauty and the Beast  November 22, 1991              Belle
9               Alice in Wonderland      July 28, 1951              Alice
10                         Zootopia      March 4, 2016          Judy Hopps
11                           Tarzan      June 18, 1999              Tarzan
12                       Pocahontas      June 23, 1995          Pocahontas
13                    The Aristocats  December 24, 1970  Thomas and Duchess
14                           Moana  November 23, 2016               Moana
15                       Big Hero 6   November 7, 2014         Hiro Hamada
16                The Little Mermaid  November 17, 1989              Ariel
17                           Mulan      June 19, 1998              Mulan
18                         Tangled  November 24, 2010            Rapunzel

                villian                          song release_date_y  \
0           Evil Queen  Some Day My Prince Will Come   Dec 21, 1937
1            Stromboli     When You Wish upon a Star    Feb 9, 1940
```

2	Si and Am	Bella Notte	Jun 22, 1955
3	Lady Tremaine	Bibbidi-Bobbidi-Boo	Feb 15, 1950
4	Kaa and Shere Khan	The Bare Necessities	Oct 18, 1967
5	Scar	Circle of Life	Jun 15, 1994
6	Jafar	A Whole New World	Nov 11, 1992
7	Prince Hans	Let It Go	Nov 22, 2013
8	Gaston	Be Our Guest	Nov 13, 1991
9	Queen of Hearts	The Unbirthday Song	Mar 5, 2010
10	NaN	Try Everything	Mar 4, 2016
11	Clayton	You'll Be in My Heart	Jun 16, 1999
12	Governor Ratcliffe	Colors of the Wind	Jun 10, 1995
13	Edgar Balthazar	Ev'rybody Wants to Be a Cat	Apr 24, 1970
14	NaN	How Far I'll Go	Nov 23, 2016
15	Professor Callaghan	Immortals	Nov 7, 2014
16	Ursula	Under the Sea	Nov 15, 1989
17	Shan Yu	I'll Make a Man Out of You	Jun 19, 1998
18	Mother Gothel	I See the Light	Nov 24, 2010

	genre	MPAA_rating	total_gross	inflation_adjusted_gross
0	Musical	G	$184,925,485	5.228953e+09
1	Adventure	G	$84,300,000	2.188229e+09
2	Drama	G	$93,600,000	1.236036e+09
3	Drama	G	$85,000,000	9.206087e+08
4	Musical	Not Rated	$141,843,000	7.896123e+08
5	Adventure	G	$422,780,140	7.616409e+08
6	Comedy	G	$217,350,219	4.419692e+08
7	Adventure	PG	$400,738,009	4.149972e+08
8	Musical	G	$218,951,625	3.630177e+08
9	Adventure	PG	$334,191,110	3.570635e+08
10	Adventure	PG	$341,268,248	3.412682e+08
11	Adventure	G	$171,091,819	2.839003e+08
12	Adventure	G	$141,579,773	2.743710e+08
13	Musical	G	$55,675,257	2.551615e+08
14	Adventure	PG	$246,082,029	2.460820e+08
15	Adventure	PG	$222,527,828	2.292492e+08
16	Adventure	G	$111,543,479	2.237260e+08
17	Adventure	G	$120,620,254	2.168078e+08
18	Adventure	PG	$200,821,936	2.143885e+08

After sorting the table in descending order of inflation adjusted gross and selecting only the top 20 movies to display, we can generate a bar plot to visualize which 20 movies with theme songs have the highest inflation adjusted gross.

[11]:
```
# Use altair to generate a bar plot
top_20_song_gross_merged_plot = (
    alt.Chart(top_20_song_gross_merged, width=500, height=300)
    .mark_bar()
```

```
    .encode(
        x=alt.X("movie_title:O", title="Movie Title", sort='y'),
        y=alt.Y("inflation_adjusted_gross:Q", title="Inflation Adjusted Gross"),
    )
    .properties(title="Top 20 movies with Theme Songs in Inflation Adjusted␣
 ↪Gross")
)
top_20_song_gross_merged_plot
```

[11]:
```
alt.Chart(…)
```

From the plot, we can see the movies with theme songs and how they perform in terms of inflation adjusted gross. We can see that Snow White and the Seven Dwarfs takes the top spot with over 5 billion dollars generated in inflation adjusted gross.

Next we will do perform the same steps to see which movies with no theme song have the highest inflation adjusted gross.

[13]:
```
# Merge characters_no_song with gross df
characters_no_song_merged = characters_no_song.merge(gross_clean,␣
 ↪left_on='movie_title', right_on='movie_title', how='left')

# Sort in descending order of inflation adjusted gross and display the top 20
characters_no_song_merged = characters_no_song_merged.
 ↪dropna(subset=['inflation_adjusted_gross'])
top_20_characters_no_song_merged = characters_no_song_merged.
 ↪sort_values(by=['inflation_adjusted_gross'], ascending=False).
 ↪reset_index(drop=True)
top_20_characters_no_song_merged
```

[13]:

	movie_title	release_date_x	hero	villian	song	release_date_y	genre	MPAA_rating	total_gross	inflation_adjusted_gross
0	Fantasia	November 13, 1940	NaN	Chernabog	NaN	Nov 13, 1940	Musical	G		
1	Dinosaur	May 19, 2000	Aladar	Kron	NaN	May 19, 2000	Adventure	PG		
2	Chicken Little	November 4, 2005	Ace Cluck	Foxy Loxy	NaN	Nov 4, 2005	Adventure	G		
3	Home on the Range	April 2, 2004	Maggie	Alameda Slim	NaN	Apr 2, 2004	Comedy	PG		
4	The Rescuers Down Under	November 16, 1990	Bernard and Miss Bianca	Percival C. McLeach	NaN	Nov 16, 1990	Adventure	G		
5	The Black Cauldron	July 24, 1985	Taran	Horned King	NaN	Jul 24, 1985	Adventure	NaN		

```
0   $83,320,000              2.187091e+09
1  $137,748,063              2.154390e+08
2  $135,386,665              1.779547e+08
3   $50,026,353              6.791017e+07
4   $27,931,461              5.579673e+07
5   $21,288,692              5.055314e+07
```

[14]:
```
# Use altair to generate a bar plot
top_20_characters_no_song_merged_plot = (
    alt.Chart(top_20_characters_no_song_merged, width=500, height=300)
    .mark_bar()
    .encode(
        x=alt.X("movie_title:O", title="Movie Title", sort='y'),
        y=alt.Y("inflation_adjusted_gross:Q", title="Inflation Adjusted Gross"),
    )
    .properties(title="Top Movies Without a Theme Song in Inflation Adjusted␣
 ↪Gross")
)
top_20_characters_no_song_merged_plot
```

[14]:
```
alt.Chart(…)
```

We can see from this graph that Fantasia is the movie with no theme song that has the highest inflation adjusted gross at over 2 billion dollars.

We can also immediately see by the number of columns shown in the Top movies with no theme song in inflation adjusted gross, there are only 6 movies displayed compared to the previous graph which hold a lot more movies. This could suggest two things– one is that there are more Disney movies produced with theme songs or that the disney_movies_total_gross.csv does not include all of the songs included in the disney-characters.csv, therefore, they were not merged into the characters_no_song_merged dataframe.

Another observation is that the top movie, Fantasia, has an inflation adjusted gross a little bit under 2.2 billion dollars which is almost 3 billion dollars less than what Snow White and the Seven Dwarfs generated. This shows that the top movie with a theme song performed significantly better than the top movie without a theme song in inflation adjusted gross.

Next I will merge the characters dataframe and gross dataframe into one full dataframe that include both movies with and without theme songs. I will also create a new column 'has_theme_song' that shows the boolean value of if the movie has a theme song.

[15]:
```
# Graph of all movies and inflated adjusted gross

# Merge characters df with gross df with all columns showing
merged_df = characters_clean.merge(gross_clean, left_on='movie_title',␣
 ↪right_on='movie_title', how='outer')

# Create new column 'has_theme_song' which shows the boolean value of if the␣
 ↪movie has a theme song
```

```
merged_df['has_theme_song'] = merged_df['song'].notna()
merged_df.head()
```

[15]:

	movie_title	release_date_x	hero	villian	song	release_date_y	genre	MPAA_rating	total_gross	inflation_adjusted_gross	has_theme_song
0	Snow White and the Seven Dwarfs	December 21, 1937	Snow White	Evil Queen	Some Day My Prince Will Come	Dec 21, 1937	Musical	G	$184,925,485	5.228953e+09	True
1	Pinocchio	February 7, 1940	Pinocchio	Stromboli	When You Wish upon a Star	Feb 9, 1940	Adventure	G	$84,300,000	2.188229e+09	True
2	Fantasia	November 13, 1940	NaN	Chernabog	NaN	Nov 13, 1940	Musical	G	$83,320,000	2.187091e+09	False
3	Dumbo	October 23, 1941	Dumbo	Ringmaster	Baby Mine	NaN	NaN	NaN	NaN	NaN	True
4	Bambi	August 13, 1942	Bambi	Hunter	Love Is a Song	NaN	NaN	NaN	NaN	NaN	True

Now that we have a dataframe with all movies merged with the inflation adjusted gross, we can group by our newly created column of whether or not a movie has a theme song and calculate the average inflation adjusted gross. I will import and use a script I created with a custom function to group by a grouping column and applies a specific aggregating function with the default being finding the mean.

[16]:
```
# import script
import sample_script as ps

# run function on data to find the mean of inflation adjusted gross for movies␣
↪with and without a theme song
theme_song_groups = ps.group_by_aggregate(merged_df, 'has_theme_song',␣
↪'inflation_adjusted_gross')
theme_song_groups
```

[16]:

has_theme_song	mean
False	9.457005e+07
True	4.362150e+08

Grouping the movies by song and no song and aggregating the mean value for each group answers our original question of whether movies with songs or movies without songs perform better in terms of inflation adjusted gross. Disney movies that include songs generated on average almost 450 million dollars while movies that don't include songs generated on average almost 100 million dollars. The gap is quite large and from these values we can infer that movies with theme songs

tend to perform better in terms of inflation adjusted gross.

We can also generate a bar plot to visualize the gap between the two groups.

[18]:
```
# Use altair to generate a bar plot
theme_song_groups_plot = (
    alt.Chart(theme_song_groups, width=500, height=300)
    .mark_bar()
    .encode(
        x=alt.X("has_theme_song:N", title="Has Theme Song"),
        y=alt.Y("mean:Q", title="Average Inflation Adjusted Gross"),
    )
    .properties(title="Average Inflation Adjusted Gross for Disney Movies With␣
 ↪and Without a Theme Song")
)
theme_song_groups_plot
```

[18]:
```
alt.Chart(…)
```

1.4 Discussions

In this analysis, I analyzed the Disney dataset and tried to determine whether movies with a theme song or movies without a theme song performed better in inflation adjusted gross. To start off, I explored how movies with theme songs performed separately from how movies without theme songs performed. I found out that Snow White and the Seven Dwarfs is the movie with a theme song that has the highest inflation adjusted gross at approximately \$5 billion. I also found out that Fantasia is the movie without a theme song that has the highest inflation adjusted gross at approximately \$2 billion. Both the top movies seemed to perform significantly better than the rest of the movies, respectively. For example, Snow White generated around \$3 billion more in inflation adjusted gross compared to the second highest movie with a theme song, Pinocchio. To add on, Fantasia generated around \$2 billion more in inflation adjusted gross than the second highest movie without a theme song, Dinosaur. It is possible these two movies are outliers in their respective tables as they drastically generated more in inflation adjusted gross compared to the rest of the movies. From this visualization we can see that there are many more movies with theme songs compared to movies without theme songs. On top of that, Snow White generated almost \$3 billion more in inflation adjusted gross compared to Fantasia, the top movie without a theme song.

We determined after grouping the two groups of movies with songs and movies without songs, that movies with songs actually do perform better in inflation adjusted gross. Movies with songs generated, on average, \$450 million. Movies without theme songs generated, on average, \$100 million. This is a big discrepancy between the two, implying that on average movies with theme songs do perform better in inflation adjusted gross. This is not surprising to me as I originally believed that movies that feature a theme song would perform better in inflation adjusted gross as nowadays I rarely see any Disney movies that don't feature songs or a musical aspect in the film. To add on, I inferred that the audience of Disney movies tend to be younger in age, thus, featuring music in the film can make the film more entertaining. Furthermore, if movies include catchy songs, this can attract a lot of attention to the movie itself, making the movie more successful and talked about as viewers will associate the songs with the movie.

It is important to note that this implication should not be generalized as this finding may be very specific to the unique datasets used. One reason being is that the Disney datasets only date up to 2016 which means the datasets are quite old and missing many recent films. Another reason could be that the ***disney_movies_total_gross.csv*** does not include all of the movies in the ***disney-characters.csv*** as previously mentioned. This would mean that when merging on the dataframes, some movies that are not common in these dataframes are dropped, thus, many points that may contribute to this finding may not be included.

Another question that could be explored in the future is whethere there is a change in the inflation adjusted gross in movies with songs over the years. One could look at whether movies with theme songs increase in inflation adjusted gross over the years as perhaps directors would feel more inclined to produce more movies with theme songs as they can become more successful than without one. If there is an increase in inflation adjusted gross over the years this may suggest a change in marketing strategy where directors also decide to include more songs and invest more money into making higher quality song productions because movies with theme songs perform better with movies without theme songs.

1.5 References

- Data Source
  - The Disney database used in this work was created by Kelly Garrett
- Data Analysis
  - The functions for analysis were inspired by Elijah Willie
- Data functions
  - For cleaning up blank spaces
  - To remove the $ in columns