final_project_disney

March 18, 2024

1 Data Analysis for Disney Movies

1.1 Author : Roopam Arora

1.2 Introduction:

This report analyzes a few datasets related to Disney movies and answers a question related to the movies by using some Python concepts, data structures and modules like Pandas, NumPy and altair for visualization.

1.3 Question(s) of interests

The question considered in this report is “Which adventure movie director had a blockbuster movie with gross earning more than $400M in the 21st Century(the year 2000’s)?” I am interested in knowing the answer to this question to understand who is the skilled Disney director to work with and vote for the Academy awards and also vote for the best song. I am considering the total gross earnings for the movies as against the inflation adjusted gross.

1.4 Dataset description

The below descripitions were taken directly from the website where the datasets were obtained.

“Disney characters, box office success & annual gross income”

The Disney movie dataset is composed of 5 tables - disney_movies_total_gross.csv, disney_revenue_1991-2016.csv, disney-characters.csv, disney-director.csv and disney-voice-actors.csv . Each table is stored in a .csv file and contains different information about Disney movies pieces including gross income, characters, songs, revenue, directors and voice actors. I will be using the disney_movies_total_gross, disney-characters and disney-director tables formally described below:

disney_movies_total_gross.csv This file contains information on movie title, including a release date, genre, MPAA rating, total gross income and inflation adjusted gross income. disney-characters.csv This file includes information on movie title, including a release date, hero, villain and song for every movie. disney-director.csv This file includes information on movie title and the director for each movie.

1.5 Methods & Results:

In this analysis, I plan to import the relevant modules like pandas, altair and numpy. Then read the 3 datasets - disney_movies_total_gross.csv, disney-director.csv and disney-characters.csv. Then I would do some wrangling and cleaning of the datasets to convert them into a format where I can merge them and extract the answer to my question. I am going to use a function for wrangling and formatting and also some test functions to test my function. I am goignt o format my function using flake8 and black formatting. Finally, I will derive the answer to my question. Let's begin.

1.5.1 Let us import the modules, read our tables and do some basic visualizations:

[1]:

```
#import the modules to be used
import pandas as pd
import altair as alt
import numpy as np
```

[2]:

```
#Reading the first dataset - disney_movies_total_gross:
disney_total_gross_df = pd.read_csv('data/disney_movies_total_gross.csv')
disney_total_gross_df
```

[2]:

```
                         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
..                                 …         …        …         …
574         The Light Between Oceans   Sep 2, 2016      Drama       PG-13
575                   Queen of Katwe  Sep 23, 2016      Drama          PG
576                   Doctor Strange   Nov 4, 2016  Adventure       PG-13
577                            Moana  Nov 23, 2016  Adventure          PG
578     Rogue One: A Star Wars Story  Dec 16, 2016  Adventure       PG-13

      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
..             …                      …
574   $12,545,979              $12,545,979
575    $8,874,389               $8,874,389
576  $232,532,923             $232,532,923
577  $246,082,029             $246,082,029
578  $529,483,936             $529,483,936

[579 rows x 6 columns]
```

[3]:

```
#Reading the second dataset - disney_director:
disney_director_df = pd.read_csv('data/disney-director.csv')
#renaming the column to make merging easier
```

```
disney_director_df_new = disney_director_df.rename(columns={"name":␣
 ↪"movie_title"})
disney_director_df_new
```

[3]:

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

[4]:

```
#Reading the third datset - disney_characters:

disney_characters_df = pd.read_csv('data/disney-characters.csv')
```

1.6 Summarizing the datasets

[5]:

```
disney_total_gross_df.describe()
```

[5]:

```
            movie_title  release_date   genre MPAA_rating total_gross  \
count               579           579     562         523         579
unique              573           553      12           5         576
top     The Jungle Book  Dec 25, 1997  Comedy          PG          $0
freq                  3             3     182         187           4

       inflation_adjusted_gross
count                       579
unique                      576
top                          $0
freq                          4
```

[6]:

```
disney_director_df_new.info()
```

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

[7]:

```
disney_characters_df.shape
```

[7]:

```
(56, 5)
```

1.7 Data Wrangling and clean up

[8]:

```
#Wrangling the disney_movies_total_gross dataset using the wrangle_disney_data␣
 ↪function

from wrangle_data import wrangle_disney_data

ch = "[$,]"
col1 = 'total_gross'
disney_gross_wrangled_df = wrangle_disney_data(disney_total_gross_df,col1,ch,␣
 ↪col2 ='release_date')
disney_gross_wrangled_df
```

[8]:

```
                         movie_title release_date      genre MPAA_rating  \
0    Snow White and the Seven Dwarfs   1937-12-21    Musical           G
1                          Pinocchio   1940-02-09  Adventure           G
2                           Fantasia   1940-11-13    Musical           G
3                  Song of the South   1946-11-12  Adventure           G
4                         Cinderella   1950-02-15      Drama           G
..                                …          …        …          …
574         The Light Between Oceans   2016-09-02      Drama       PG-13
575                   Queen of Katwe   2016-09-23      Drama          PG
576                   Doctor Strange   2016-11-04  Adventure       PG-13
577                            Moana   2016-11-23  Adventure          PG
578     Rogue One: A Star Wars Story   2016-12-16  Adventure       PG-13

    total_gross inflation_adjusted_gross
0     184925485           $5,228,953,251
1      84300000           $2,188,229,052
2      83320000           $2,187,090,808
3      65000000           $1,078,510,579
4      85000000             $920,608,730
..          …                      …
574    12545979              $12,545,979
575     8874389               $8,874,389
576   232532923             $232,532,923
577   246082029             $246,082,029
578   529483936             $529,483,936

[579 rows x 6 columns]
```

[9]:

```python
#Wrangling the disney-characters dataset using the wrangle_disney_data function

from wrangle_data import wrangle_disney_data

ch = "[\n,]"
column1 = 'movie_title'
disney_characters_wrangled_df =␣
 ↪wrangle_disney_data(disney_characters_df,column1,ch, col2 ='release_date')
disney_characters_wrangled_df
```

[9]:

```
                               movie_title release_date  \
0          Snow White and the Seven Dwarfs   1937-12-21
1                                Pinocchio   1940-02-07
2                                 Fantasia   1940-11-13
3                                    Dumbo   1941-10-23
4                                    Bambi   1942-08-13
5                           Saludos Amigos   1943-02-06
6                     The Three Caballeros   1945-02-03
7                          Make Mine Music   1946-04-20
8                       Fun and Fancy Free   1947-09-27
9                              Melody Time   1948-05-27
10  The Adventures of Ichabod and Mr. Toad   1949-10-05
11                              Cinderella   1950-02-15
12                     Alice in Wonderland   1951-07-28
13                               Peter Pan   1953-02-05
14                      Lady and the Tramp   1955-06-22
15                         Sleeping Beauty   1959-01-29
16          One Hundred and One Dalmatians   1961-01-25
17                  The Sword in the Stone   1963-12-25
18                         The Jungle Book   1967-10-18
19                          The Aristocats   1970-12-24
20                              Robin Hood   1973-11-08
21  The Many Adventures of Winnie the Pooh   1977-03-11
22                            The Rescuers   1977-06-22
23                   The Fox and the Hound   1981-07-10
24                      The Black Cauldron   1985-07-24
25               The Great Mouse Detective   1986-07-02
26                        Oliver & Company   1988-11-18
27                      The Little Mermaid   1989-11-17
28                 The Rescuers Down Under   1990-11-16
29                    Beauty and the Beast   1991-11-22
30                                 Aladdin   1992-11-25
31                           The Lion King   1994-06-24
32                              Pocahontas   1995-06-23
33             The Hunchback of Notre Dame   1996-06-21
34                                Hercules   1997-06-27
35                                   Mulan   1998-06-19
```

```
36                                  Tarzan   1999-06-18
37                           Fantasia 2000   1999-12-17
38                                Dinosaur   2000-05-19
39                The Emperor's New Groove   2000-12-15
40               Atlantis: The Lost Empire   2001-06-15
41                           Lilo & Stitch   2002-06-21
42                         Treasure Planet   2002-11-27
43                            Brother Bear   2003-11-01
44                       Home on the Range   2004-04-02
45                          Chicken Little   2005-11-04
46                      Meet the Robinsons   2007-03-30
47                                    Bolt   2008-11-21
48               The Princess and the Frog   2009-12-11
49                                 Tangled   2010-11-24
50                         Winnie the Pooh   2011-07-15
51                          Wreck-It Ralph   2012-11-02
52                                  Frozen   2013-11-27
53                              Big Hero 6   2014-11-07
54                                Zootopia   2016-03-04
55                                   Moana   2016-11-23

                          hero                               villian  \
0                   Snow White                            Evil Queen
1                    Pinocchio                             Stromboli
2                          NaN                             Chernabog
3                        Dumbo                            Ringmaster
4                        Bambi                                Hunter
5                  Donald Duck                                   NaN
6                  Donald Duck                                   NaN
7                          NaN                                   NaN
8                 Mickey Mouse                      Willie the Giant
9                          NaN                                   NaN
10  Mr. Toad and Ichabod Crane  Mr. Winkie and The Headless Horseman
11                  Cinderella                         Lady Tremaine
12                       Alice                       Queen of Hearts
13                   Peter Pan                          Captain Hook
14               Lady and Tramp                             Si and Am
15                      Aurora                            Maleficent
16                       Pongo                        Cruella de Vil
17                      Arthur                             Madam Mim
18                      Mowgli                    Kaa and Shere Khan
19          Thomas and Duchess                       Edgar Balthazar
20                  Robin Hood                           Prince John
21             Winnie the Pooh                                   NaN
22     Bernard and Miss Bianca                         Madame Medusa
23              Tod and Copper                            Amos Slade
24                       Taran                           Horned King
```

```
25                       Basil                     Professor Ratigan
26                      Oliver                                 Sykes
27                       Ariel                                Ursula
28     Bernard and Miss Bianca                   Percival C. McLeach
29                       Belle                                Gaston
30                     Aladdin                                 Jafar
31                       Simba                                  Scar
32                  Pocahontas                    Governor Ratcliffe
33                   Quasimodo                         Claude Frollo
34                    Hercules                                 Hades
35                       Mulan                               Shan Yu
36                      Tarzan                               Clayton
37                         NaN                                   NaN
38                      Aladar                                  Kron
39                       Kuzco                                  Yzma
40                 Milo Thatch                      Commander Rourke
41             Lilo and Stitch                                   NaN
42                 Jim Hawkins                           John Silver
43                       Kenai                                Denahi
44                      Maggie                          Alameda Slim
45                   Ace Cluck                             Foxy Loxy
46                       Lewis                                 Doris
47                        Bolt                            Dr. Calico
48                       Tiana                          Dr. Facilier
49                    Rapunzel                         Mother Gothel
50             Winnie the Pooh                                   NaN
51                       Ralph                                 Turbo
52                        Elsa                           Prince Hans
53                 Hiro Hamada                   Professor Callaghan
54                  Judy Hopps                                   NaN
55                       Moana                                   NaN

                                  song
0         Some Day My Prince Will Come
1            When You Wish upon a Star
2                                  NaN
3                            Baby Mine
4                       Love Is a Song
5                     Saludos Amigos\n
6                                  NaN
7                                  NaN
8                                  NaN
9                          Little Toot
10                    The Merrily Song
11                 Bibbidi-Bobbidi-Boo
12                 The Unbirthday Song
13                        You Can Fly!
```

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

1.8 Filtering the first dataset for our analysis

```
[10]: #Finding the movies from first dataset(data/disney_movies_total_gross.csv)␣
       ↪where gross earnings is more than $400M

      disney_movies_400M_df = disney_total_gross_df.
       ↪loc[disney_gross_wrangled_df['total_gross'].astype(int) > 400000000]
      disney_movies_400M_df
```

```
[10]:                                movie_title release_date      genre MPAA_rating  \
      179                             The Lion King   1994-06-15  Adventure           G
      441  Pirates of the Caribbean: Dead Man’…   2006-07-07  Adventure       PG-13
      499                               Toy Story 3   2010-06-18  Adventure           G
      524                              The Avengers   2012-05-04     Action       PG-13
      532                                Iron Man 3   2013-05-03     Action       PG-13
      539                                    Frozen   2013-11-22  Adventure          PG
      558                   Avengers: Age of Ultron   2015-05-01     Action       PG-13
      564  Star Wars Ep. VII: The Force Awakens   2015-12-18  Adventure       PG-13
      569                Captain America: Civil War   2016-05-06     Action       PG-13
      571                              Finding Dory   2016-06-17  Adventure          PG
      578             Rogue One: A Star Wars Story   2016-12-16  Adventure       PG-13

          total_gross inflation_adjusted_gross
      179   422780140             $761,640,898
      441   423315812             $544,817,142
      499   415004880             $443,408,255
      524   623279547             $660,081,224
      532   408992272             $424,084,233
      539   400738009             $414,997,174
      558   459005868             $459,005,868
      564   936662225             $936,662,225
      569   408084349             $408,084,349
      571   486295561             $486,295,561
      578   529483936             $529,483,936
```

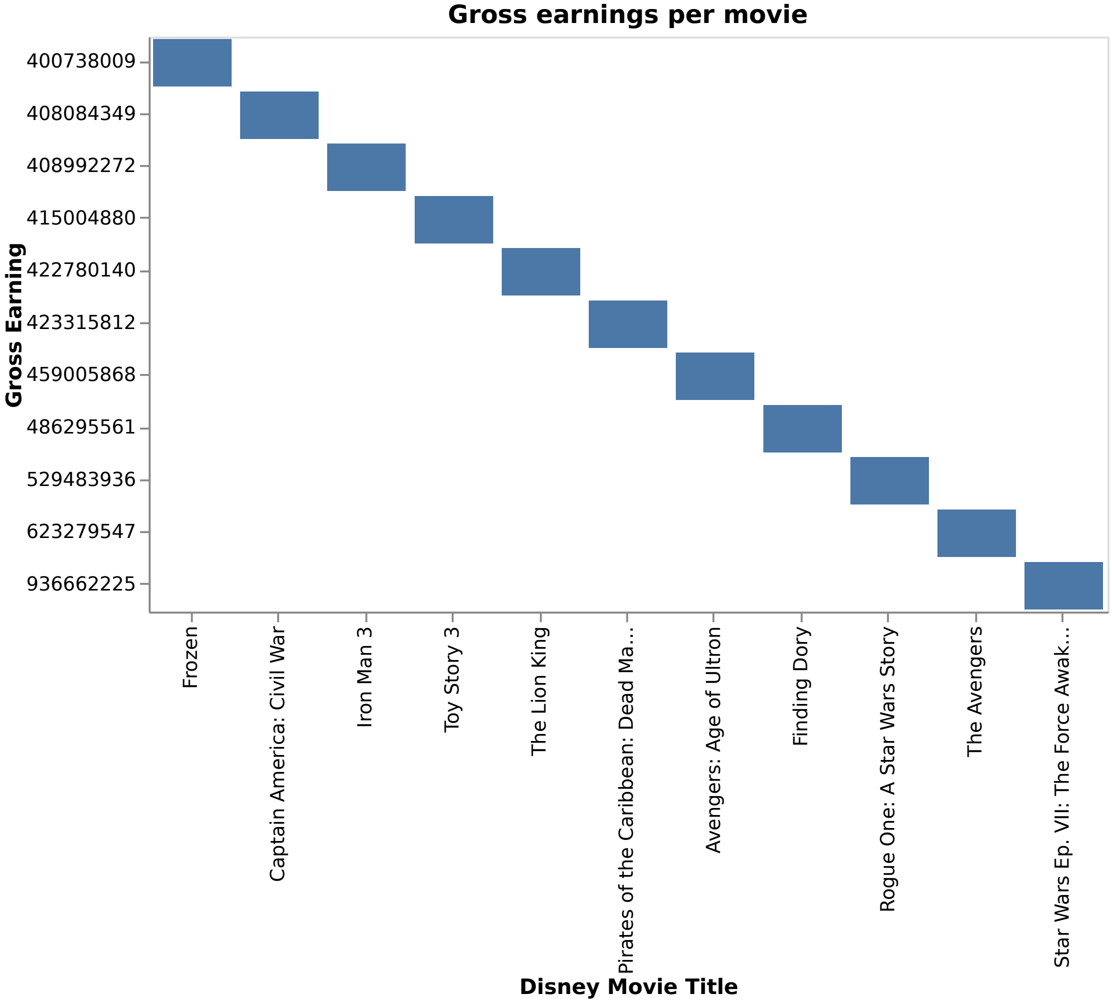
1.9 Visualization with Altair

```
[11]: #Visualizing the first dataset:
      alt.renderers.enable('altair_saver', fmts=["html", "svg"])

      disney_total_gross_plot = alt.Chart(disney_movies_400M_df, width=500,␣
       ↪height=300).mark_bar().encode (x=alt.X('movie_title', sort="y",␣
       ↪title='Disney Movie Title'),
                                                                              ␣
       ↪y=alt.Y('total_gross', title= 'Gross Earning')).properties(title="Gross␣
       ↪earnings per movie")
```

```
disney_total_gross_plot
```

[11]:

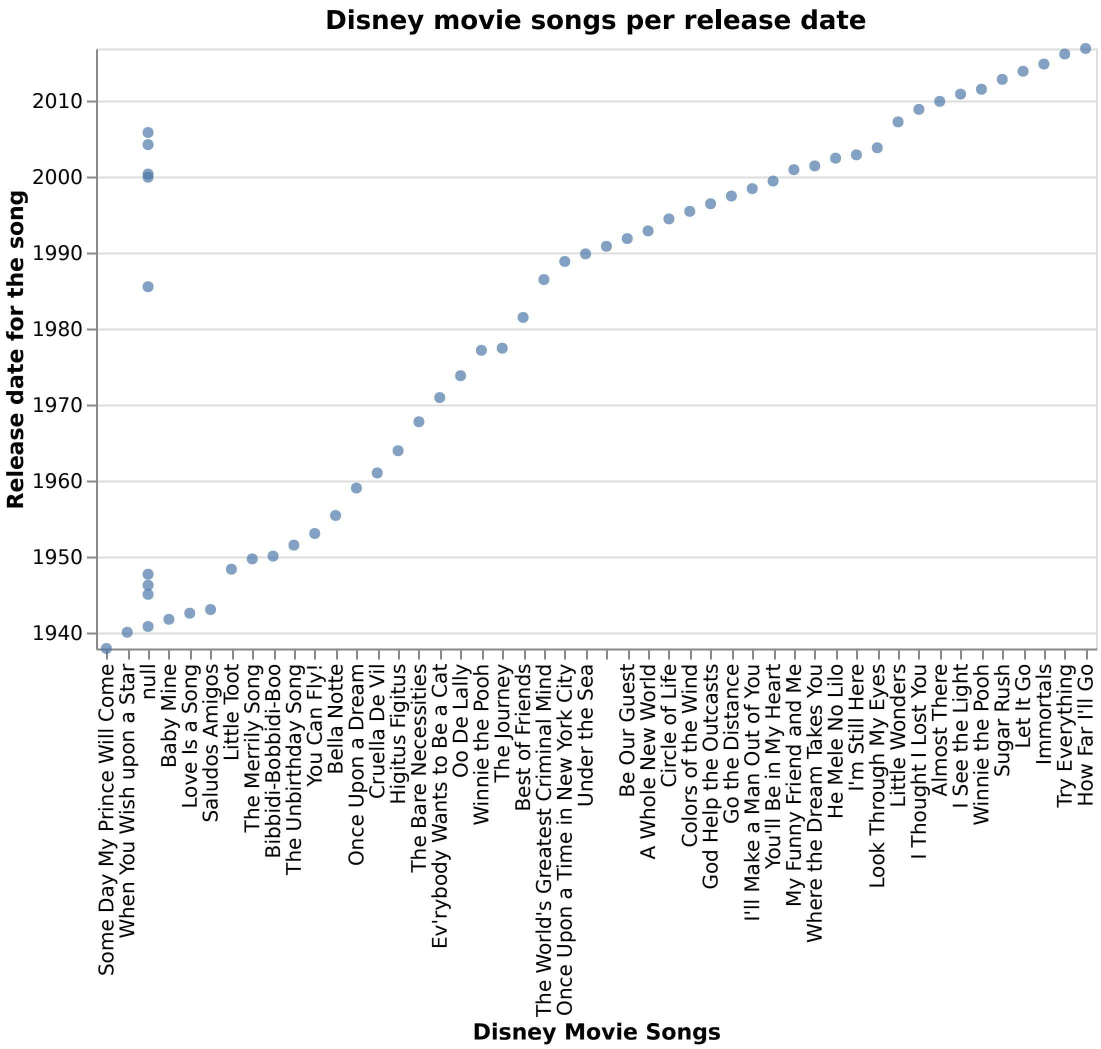


[12]:

```
#Visualizing the first dataset:
disney_characters_plot = alt.Chart(disney_characters_wrangled_df, width=500,␣
 ↪height=300).mark_circle().encode (x=alt.X('song', sort="y", title='Disney␣
 ↪Movie Songs'),
                                                                           ␣
 ↪y=alt.Y('release_date', title= 'Release date for the song')).
 ↪properties(title="Disney movie songs per release date")

disney_characters_plot
```

[12]:



1.10 Code Improvement - black formatting

[13]:
```
!flake8 wrangle_data.py

#Using black formatting to improve our function code
!black wrangle_data.py
```

```
wrangle_data.py:4:80: E501 line too long (81 > 79 characters)
wrangle_data.py:25:80: E501 line too long (86 > 79 characters)
All done!
1 file left unchanged.
```

1.11 Testing the wrangle_data function

```
[14]: !pytest test_wrangle_data.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


test_wrangle_data.py ..

[100%]

============================== 2 passed in 0.81s
===============================
```

```
[15]: #Summarizing the data - combining the 3 datasets into one consolidated␣
       ↪dataframe to make our analysis easier

      merged_gross_character_df = pd.merge(disney_movies_400M_df,␣
       ↪disney_director_df_new, how="left", on=["movie_title"])
      merged_gross_character_df = pd.
       ↪merge(merged_gross_character_df,disney_characters_df, how="left",␣
       ↪on=["movie_title"] )
      merged_gross_character_df = merged_gross_character_df.
       ↪rename(columns={'release_date_x':'release_date','release_date_y':
       ↪'release_date_song'})
      merged_gross_character_df
```

```
[15]:                            movie_title release_date      genre MPAA_rating  \
      0                         The Lion King   1994-06-15  Adventure           G
      1   Pirates of the Caribbean: Dead Man’…   2006-07-07  Adventure       PG-13
      2                           Toy Story 3   2010-06-18  Adventure           G
      3                          The Avengers   2012-05-04     Action       PG-13
      4                            Iron Man 3   2013-05-03     Action       PG-13
      5                                Frozen   2013-11-22  Adventure          PG
      6               Avengers: Age of Ultron   2015-05-01     Action       PG-13
      7   Star Wars Ep. VII: The Force Awakens   2015-12-18  Adventure       PG-13
      8            Captain America: Civil War   2016-05-06     Action       PG-13
      9                          Finding Dory   2016-06-17  Adventure          PG
      10          Rogue One: A Star Wars Story   2016-12-16  Adventure       PG-13

         total_gross inflation_adjusted_gross      director release_date_song  \
      0    422780140             $761,640,898  Roger Allers        1994-06-24
```

```
1    423315812             $544,817,142           NaN               NaT
2    415004880             $443,408,255           NaN               NaT
3    623279547             $660,081,224           NaN               NaT
4    408992272             $424,084,233           NaN               NaT
5    400738009             $414,997,174    Chris Buck        2013-11-27
6    459005868             $459,005,868           NaN               NaT
7    936662225             $936,662,225           NaN               NaT
8    408084349             $408,084,349           NaN               NaT
9    486295561             $486,295,561           NaN               NaT
10   529483936             $529,483,936           NaN               NaT

     hero      villian            song
0   Simba         Scar  Circle of Life
1     NaN          NaN             NaN
2     NaN          NaN             NaN
3     NaN          NaN             NaN
4     NaN          NaN             NaN
5    Elsa  Prince Hans     Let It Go\n
6     NaN          NaN             NaN
7     NaN          NaN             NaN
8     NaN          NaN             NaN
9     NaN          NaN             NaN
10    NaN          NaN             NaN
```

[16]:

```
#Dropping the rows with Nan values

df_dropped = merged_gross_character_df.dropna()
df_dropped
```

[16]:

```
     movie_title release_date      genre MPAA_rating total_gross  \
0  The Lion King   1994-06-15  Adventure           G   422780140
5         Frozen   2013-11-22  Adventure          PG   400738009

  inflation_adjusted_gross      director release_date_song   hero  \
0             $761,640,898  Roger Allers        1994-06-24  Simba
5             $414,997,174    Chris Buck        2013-11-27   Elsa

       villian            song
0         Scar  Circle of Life
5  Prince Hans     Let It Go\n
```

[17]:

```
# Filtering the movie which was made in the 2000's
df_dropped_final = df_dropped.loc[df_dropped['release_date'].dt.year > 2000 ]
print("Highest Earning Adventure Movie Director")
df_dropped_final["director"]
```

```
Highest Earning Adventure Movie Director
```

```
[17]: 5    Chris Buck
      Name: director, dtype: object
```

```
[18]: print("Award Winning Song")
      df_dropped_final["song"].replace("[\n,]","", regex=True)
```

```
Award Winning Song
```

```
[18]: 5    Let It Go
      Name: song, dtype: object
```

1.12 Discussions

In this analysis, I found the director who directed the highest earning movie of the 2000's and most importantly the Academy Award winning song from the movie(Frozen). The song is “Let it go” which is definitely a very popular song. This answers my original question which was a part of this analysis. The other questions that I want to answer are the genre and the MPAA rating of this movie.

2 References

Not all the work in this notebook is original. Some parts were borrowed from online resources. I take no credit for parts that are not mine. They were soley used for illustration purposes

- Data Source All the tables(datset) were derived from this website
- Question Of Interest I got my interest in the question after going through this website