DISNEY REPORT

Author :- Francis Fernandes

Date :- Oct 29, 2020

TABLE OF CONTENTS

Contents

INTRODUCTION

Disney has released more than 500 films in a wide range of genres. Their movies cater to a broad audience. In this project, we will analyze data to see how Disney has worked with various directors and which of them have been successful in a genre. Disney has grown from the time the first movie was released in 1937 an has been able to make a certain growth and create a niche in the consumer demands.

OBJECTIVE

We will use different ways to get to the conclusion and clean the data and remove missing data. Also we will like to check which genres are growing stronger in popularity and whether Disney has been able to cater to the genres and how it can reduce losses by not going into genres which have not been able to maximize profits.

Based on the data we will try to understand which are the genre's in which Disney has been successful and which are not. Which are the directors who have been able to be a part of Disney's success and who have failed.

Which are the type of movies released by Disney in recent years and have they been successful to read the customers mind.

Is there a relationship between Genre, MPAA-rating and Total Gross?

DATA

We have used 3 datasets in the current project

- Disney_directors – Data consists details of Directors and the movies they have produced
- Disney_movies-total_gross – data consists of movie name, released date, genre, MPAA_rating, Total_gross and inflation_adjusted_gross
- Disney_revenue_1991_2016 – data consists of year, studio_entertainment_N1, Disney_consumer_N2, Disney_interactive_NI3_Rev1, Walt_disney_parks_and_resorts, Disney_Media_networks and total

CHALLENGES

The data is not complete and there are some missing values. The total data consists of 579 records

```
]: # information of rows & columns in "gross" and information stored in it

   gross.info
```

```
]: <bound method DataFrame.info of                          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    $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]>
```

But after cleaning and formatting of data by using .null and drop we have removed the null values and missing data and now we have 513 clean records.

```
# to select missing values in total file  579 rows x 6 columns columns columns
# after .isnull there are total rows 66 rows x 6 columns columns columns selected having null values

gross[gross.isnull().any(axis=1)]

#to drop missing values & final output of 66 rows x 6 columnscolumns
gross_final = gross.dropna()
gross_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 513 entries, 0 to 578
Data columns (total 6 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   title_movie              513 non-null    object
 1   release_date             513 non-null    datetime64[ns]
 2   genre                    513 non-null    object
 3   MPAA_rating              513 non-null    object
 4   total_gross              513 non-null    object
 5   inflation_adjusted_gross  513 non-null    object
dtypes: datetime64[ns](1), object(5)
memory usage: 28.1+ KB
```

DATA-LOAD & CLEANING

We have necessary libraries in python we have imported the same using .csv and formatted the date .parsedates in the necessary data to get the clean formats

```
# import of all required libraries for the project

import pandas as pd
import numpy as np
import altair as alt


# import of all required files for the project


characters = pd.read_csv('data/disney-characters.csv', parse_dates = ['release_date'])
directors = pd.read_csv('data/disney-director.csv')
voice = pd.read_csv('data/disney-voice-actors.csv')
revenue = pd.read_csv('data/disney_revenue_1991-2016.csv')
gross = pd.read_csv('data/disney_movies_total_gross.csv', parse_dates = ['release_date'])
```

The columns have been renamed for the desired output to be self-explanatory. The missing and null values data have been cleaned and columns with missing data have been dropped to maintain consistency in the data.

```
# rename title_movie existing in two dataframes characters and gross
# merge both dataframes voice and gross

gross=gross.rename (columns={'movie_title':'title_movie'})
gross.head


# to select missing values in total file  579 rows x 6 columns columns columns
# after .isnull there are total rows 66 rows × 6 columns columns columns selected having null values

gross[gross.isnull().any(axis=1)]

#to drop missing values & final output of 66 rows × 6 columnscolumns
gross_final = gross.dropna()
gross_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 513 entries, 0 to 578
Data columns (total 6 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   title_movie              513 non-null    object
 1   release_date             513 non-null    datetime64[ns]
 2   genre                    513 non-null    object
 3   MPAA_rating              513 non-null    object
 4   total_gross              513 non-null    object
 5   inflation_adjusted_gross  513 non-null    object
dtypes: datetime64[ns](1), object(5)
memory usage: 28.1+ KB
```

Some of the data having numbers are not in proper format to derive the output therefore we had to change the data from Object to float to sum or calculate min and max and charts.

```
# to remove comma and dollar sign
# to change format type object to float

gross_final1['total_gross'] = gross_final1['total_gross'].str.replace(",", "").str.replace("$", "").str.strip()
gross_final1['inflation_adjusted_gross'] = gross_final1['inflation_adjusted_gross'].str.replace(",", "").str.replace("$", "").str.strip()

gross_final1= gross_final1.assign(new_total_gross=gross_final1["total_gross"].astype("float64"))
gross_final1= gross_final1.assign(new_inflation_adjusted_gross=gross_final1["inflation_adjusted_gross"].astype("float64"))

pd.options.display.float_format = '${:,.2f}'.format

gross_final1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 513 entries, 0 to 578
Data columns (total 9 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   title_movie                  513 non-null    object
 1   release_date                 513 non-null    datetime64[ns]
 2   genre                        513 non-null    object
 3   MPAA_rating                  513 non-null    object
 4   total_gross                  513 non-null    object
 5   inflation_adjusted_gross     513 non-null    object
 6   genre_rating                 513 non-null    object
 7   new_total_gross              513 non-null    float64
 8   new_inflation_adjusted_gross  513 non-null    float64
dtypes: datetime64[ns](1), float64(2), object(6)
memory usage: 40.1+ KB
```

The data has been sorted on genre

```
# to sort the data on genre column

gross_sorted = gross_drop.sort_values(by = 'genre', ascending = False)
gross_sorted
```

	title_movie	release_date	genre	MPAA_rating	genre_rating	new_total_gross	new_inflation_adjusted_gross
387	Open Range	2003-08-15	Western	R	WesternR	$58,331,254.00	$81,547,672.00
401	The Alamo	2004-04-09	Western	PG-13	WesternPG-13	$22,406,362.00	$30,416,359.00
328	Shanghai Noon	2000-05-26	Western	PG-13	WesternPG-13	$56,932,305.00	$89,042,541.00
534	The Lone Ranger	2013-07-02	Western	PG-13	WesternPG-13	$89,302,115.00	$92,597,388.00
202	Tall Tale	1995-03-24	Western	PG	WesternPG	$8,247,627.00	$15,983,331.00
...	...	...	...	...	...	...	...
498	Prince of Persia: Sands of Time	2010-05-28	Action	PG-13	ActionPG-13	$90,759,676.00	$96,971,361.00
410	Ladder 49	2004-10-01	Action	PG-13	ActionPG-13	$74,541,707.00	$101,161,045.00
444	The Guardian	2006-09-29	Action	PG-13	ActionPG-13	$55,011,732.00	$70,801,353.00
364	Bad Company	2002-06-07	Action	PG-13	ActionPG-13	$30,157,016.00	$43,756,218.00
90	Fire Birds	1990-05-24	Action	PG-13	ActionPG-13	$14,760,451.00	$29,485,923.00

513 rows × 7 columns

METHODS & RESULTS

In this project we are going to analyze and check the success of Disney movies. The data consists of 513 movies having complete records with nine features: movie title, director, release date, genre, MPAA rating, total gross, and inflation-adjusted gross.

1. Top 10 Movies of Disney : The data has been sorted by Total Gross and we have considered top 10 movies with maximum total Gross

```
# to sort the data on genre column

gross_sort = gross_drop.sort_values(by = 'new_total_gross', ascending = False)
gross_sort.head(10)
```

	title_movie	release_date	genre	MPAA_rating	genre_rating	new_total_gross	new_inflation_adjusted_gross
564	Star Wars Ep. VII: The Force Awakens	2015-12-18	Adventure	PG-13	AdventurePG-13	$936,662,225.00	$936,662,225.00
524	The Avengers	2012-05-04	Action	PG-13	ActionPG-13	$623,279,547.00	$660,081,224.00
578	Rogue One: A Star Wars Story	2016-12-16	Adventure	PG-13	AdventurePG-13	$529,483,936.00	$529,483,936.00
571	Finding Dory	2016-06-17	Adventure	PG	AdventurePG	$486,295,561.00	$486,295,561.00
558	Avengers: Age of Ultron	2015-05-01	Action	PG-13	ActionPG-13	$459,005,868.00	$459,005,868.00
441	Pirates of the Caribbean: Dead Man'...	2006-07-07	Adventure	PG-13	AdventurePG-13	$423,315,812.00	$544,817,142.00
179	The Lion King	1994-06-15	Adventure	G	AdventureG	$422,780,140.00	$761,640,898.00
499	Toy Story 3	2010-06-18	Adventure	G	AdventureG	$415,004,880.00	$443,408,255.00
532	Iron Man 3	2013-05-03	Action	PG-13	ActionPG-13	$408,992,272.00	$424,084,233.00
569	Captain America: Civil War	2016-05-06	Action	PG-13	ActionPG-13	$408,084,349.00	$408,084,349.00

As we can see star wars & Avengers sequels are the most successful Disney movies in the top ten. Currently Disney seems to be making comic book/ super hero movies (Avengers, Captain America) and Animated movies (Finding Dory, The Lion King and Toy story) and these movies have been able to meet the customer expectations.

2. Data merged

   The data has been merged Directors and gross and have derived the output to check who is the best performing director and which genre he has been successful with :-

```
gross_sort1 = movies_details.sort_values(by = 'new_inflation_adjusted_gross', ascending = False)
gross_sort1.head()

#to check the director and movie with max and min

Lowest_Director_Inflation_gross = (gross_sort1[(gross_sort1.new_inflation_adjusted_gross==gross_sort1.new_inflation_adjusted_gross.min())]
                                   .loc[:,('name','director','genre','new_inflation_adjusted_gross')])
Lowest_Director_Inflation_gross
```

	name	director	genre	new_inflation_adjusted_gross
54	Winnie the Pooh	Stephen J. Anderson	Adventure	$28,375,869.00

```
highest_director_inflation_gross = (gross_sort1[(gross_sort1.new_inflation_adjusted_gross==gross_sort1.new_inflation_adjusted_gross.max())]
                                    .loc[:,('name','director','genre','new_inflation_adjusted_gross')])
highest_director_inflation_gross
```

	name	director	genre	new_inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	David Hand	Musical	$5,228,953,251.00

Based on the above the director with minimum new inflation adjusted gross details are

	name	director	genre	new_inflation_adjusted_gross
54	Winnie the Pooh	Stephen J. Anderson	Adventure	$28,375,869.00

While the best director details are

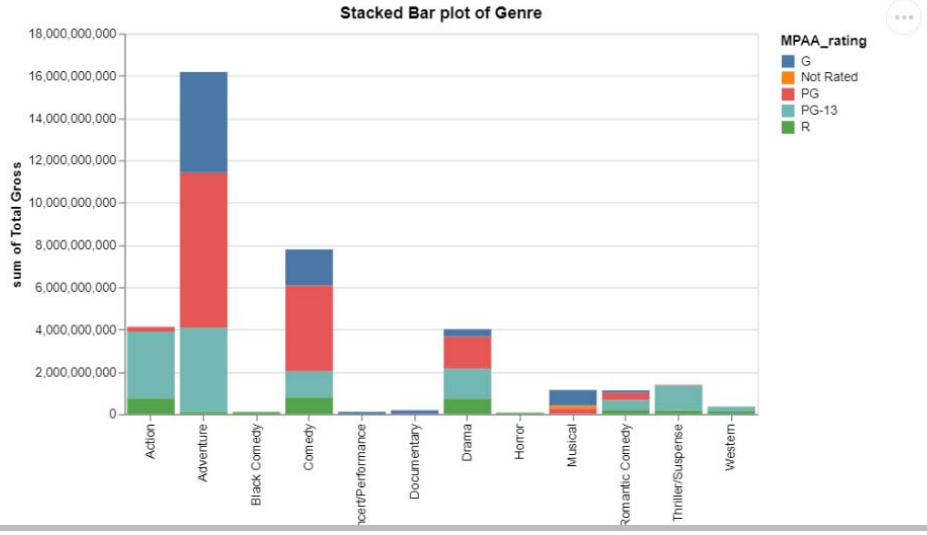
	name	director	genre	new_inflation_adjusted_gross
0	Snow White and the Seven Dwarfs	David Hand	Musical	$5,228,953,251.00

3. Data visualized - 1

- Sacked plot with data using columns genre , MPAA_rating and we have sum of total gross

```
gross_chart= alt.Chart(gross_sorted,width=500, height = 300).mark_bar().encode(
    x = alt.X('genre', title = "Movie category"),
    y = alt.Y('sum(new_total_gross)',title = ("sum of Total Gross")),
    color='MPAA_rating'
).properties(title = "Stacked Bar plot of Genre")
gross_chart
```

221]:



The above plot gives details of all genre stacked against MPAA_rating and sum of today gross.

We can understand that Disney has been successful in movies (Adventure followed by comedy, Drama and action).

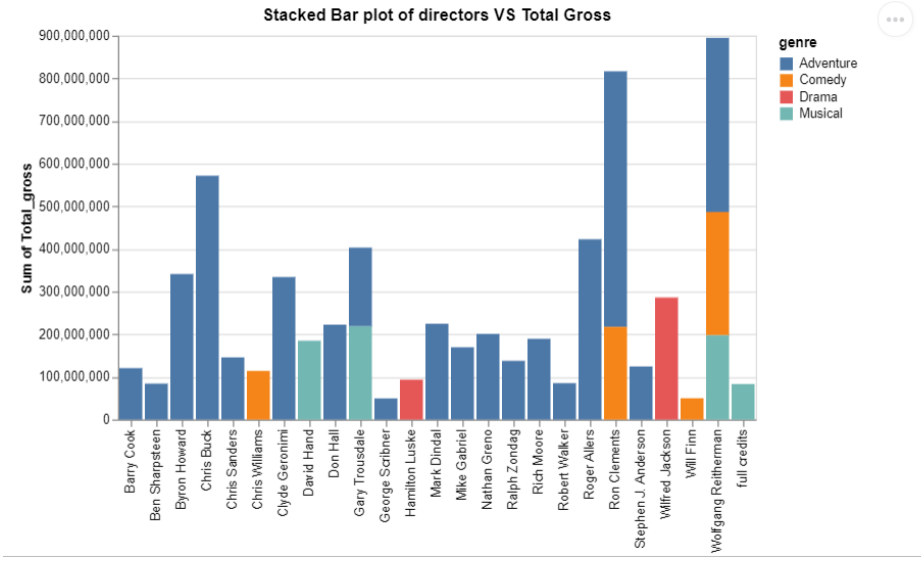
Disney has been successful in the following types of MPAA- Rating – General Audience, PG-13and PG while romance and not rated movies have not performed well.

Disney should produce Adventure and comedy movies.

Disney should avoid movies in the following Black Comedy, Horror, Western, Comedy/ Performance and Documentary because they have not been able to maximize profits and reduce losses.

- Stacked plot (Directors, Total Gross and Genre

```
movie_chart= alt.Chart(Movie_details,width=500, height = 300).mark_bar().encode(
    x = alt.X('director', title = "Directors"),
    y = alt.Y('sum(new_total_gross)',title = ("Sum of Total_gross")),
    color='genre'
).properties(title = "Stacked Bar plot of directors VS Total Gross")
movie_chart
```



The above plot indicates that some directors are performing well in certain types of genre & Disney can produce movies with these directors in the genre of their strength to maximize gross sales and reduce losses.

The least performing directors Will Finn while the most successful director has been “Wolfgang Reitherman” & “Ron Clements”

Wolfgang Reitherman has been successful in generating maximum Total gross in the genre “Adventure and comedy while Ron Clements is in Adventure.

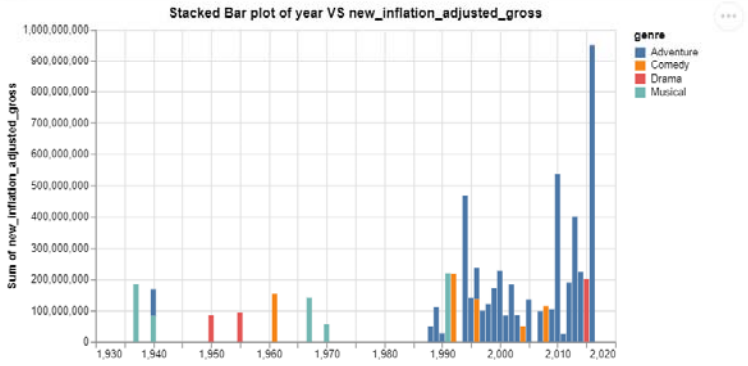
We can notice in the plot Disney has made Adventure movies and most of the directors have successful except for “George Scribner” who has not been able to make even $100,000,000/-

- Plot (year, Total Gross and Genre

```
year_sorted = Movie_details.sort_values(by = 'release_year', ascending = False)
year_sorted


# to sum the data based on years

movie_chart= alt.Chart(year_sorted,width=500, height = 300).mark_bar().encode(
    x = alt.X('release_year', title = "years"),
    y = alt.Y('sum(new_total_gross)',title = ("Sum of new_inflation_adjusted_gross")),
    color='genre'
).properties(title = "Stacked Bar plot of year VS new_inflation_adjusted_gross")
movie_chart
```



Additionally, the analysis of data has shown a progressive increase of Total Gross indicating that Disney has been successful in adventure genre from 2000 and have a high box office demand and hence there has been an increase in gross sales from 2000 and Disney has been successfully catering to the demand.

From the year 2000 Disney has increased producing movies in Adventure and comedy and drama genre and have not produced Musical movies due to less returns.

CONCLUSION

Based on this report we can conclude that each director have their strengths and Disney has overall performed well. Some directors have performed well while others have not. The Directors who have performed well have performed exceptionally well in certain types of genre.

If Disney must increase their profits and have success at the box-office, they need to work with these directors in the genre they are good at.

Disney has been able to meet box office expectations by making sequels to star wars and Avengers and have been able to meet the consumer demands.

We could say that Disney movies with plots that fit into the action and adventure genre, according to our data, tend to do better in terms of Total gross and adjusted gross than any other genres. So we could expect more Marvel, Star Wars, and live-action movies in the upcoming years and they will reduce Horror or drama movies.

REFRENCES

We have used the data from https://data.world/kgarrett/disney-character-success-00-16