Kosisochi Chuma-Ibe

Data Warehousing and BI Final Project Report

DESCRIPTIVE DASHBOARDS OF WALT DISNEY BOX OFFICE DATA

OBJECTIVE

To create descriptive dashboards that examine several aspects of the Walt Disney datasets in order to observe any interesting characteristics or trends in the box office data from the years 1937 – 2016. Data source: data.world - https://data.world/

DATASETS INFORMATION

1) Disney total box office gross revenue dataset: This dataset has the following columns -Movie title, release date, genre, MPAA rating, total gross revenue, inflation adjusted gross revenue
2) Disney characters dataset: This dataset has the following columns - Movie title, release date, hero, hero gender, villain, song
3) Disney directors dataset: This dataset has the following columns - Movie title and director

STEPS

Firstly, pre-processing in Tableau Prep builder

1) Replacing null values: I initially decided to replace the null values in some of the columns with "unknown" because I felt that there would be some information value in them, and did not want to lose that. Later on, I ended up excluding them regardless when developing my story

2) Joins: I did my joins on the movie title column because it was the column in all three datasets had in common. This was a risky move but it worked. I did 2 joins – a left join and a right join. As a result of the joins, some columns had mostly nulls, and since my aim was to get as many movie titles as possible out of each join, I deleted the columns which had more/mostly nulls because they were not relevant

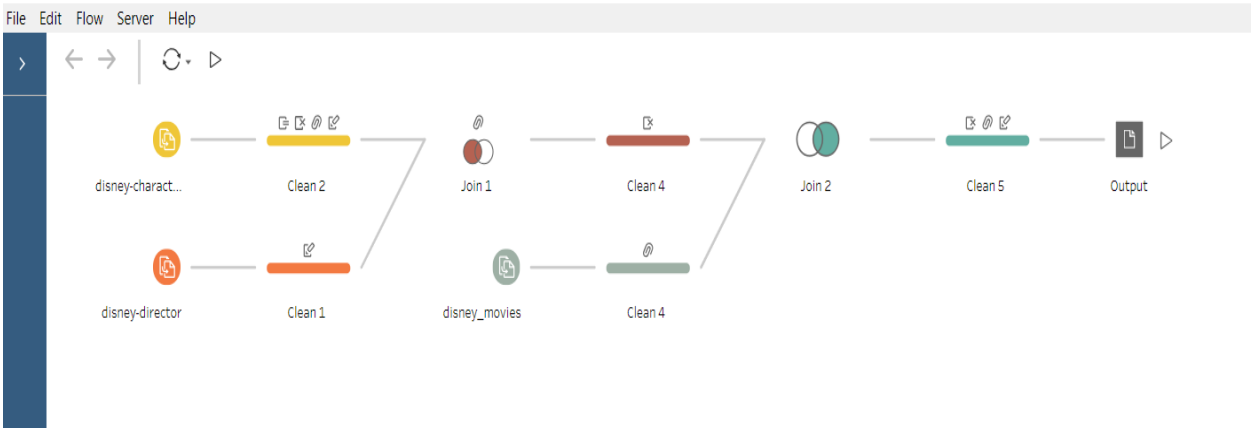
3) Removing irrelevant columns: I removed some of the fields/columns that were irrelevant to my analysis such as the villain and song column

4) Renaming columns: I renamed some of the columns so that there could be a uniform naming format across all the columns, and this also helped me to perform the joins.

5) Trimming extra spaces: The data in some columns came in with extra spaces, which I had to get rid of

6) Formatting similar values that had different spellings/representations: While performing the joins, I observed that some of the values were represented differently. For instance, there were two versions of certain movies, e.g. Lilo and Stitch, Lilo & Stitch. I had to format these values so they would match since they referred to the same movie

Tableau Prep Builder Workflow:



Secondly, I loaded my output file from the Tableau prep Builder workflow into Tableau Desktop, and began creating my visualizations.

CHALLENGES

1) Finding the story: One major challenge I faced was finding or extracting the story from the data. Initially I had intended to explore possibilities such as
   - Who are the most prolific directors? Is there a trend in the highest grossing movies and the directors involved, or are some directors more bankable than others when we consider the box office success of their movies?
   - Is there a trend in the gender of the hero/lead character in the highest grossing movies? Is there a trend in the gender of the lead character in certain genres of movies? Does the gender of the lead character affect the box office success?

   Upon exploring the data, I discovered I did not have enough data to support some of the suspicions I earlier had. In deciding on what constituted a meaningful story, I pondered on the level of detail I could delve into that would provide useful insights - too much detail and it becomes redundant/irrelevant, too little detail, and you miss significant observations. I decided to group my data into seasons and decades, and formed my story around those

   Decades : 30s, 40s, 50s etc.

   Seasons: Winter – December to February, Spring – March to May, Summer – June to August, Fall – September to November

2) Incomplete data/missing values: Another reason I initially struggled while using the data to develop my story was due to the sheer amount of missing values that were present. Initially I tried filling in these values by going online to research on the information, but this would have lead to

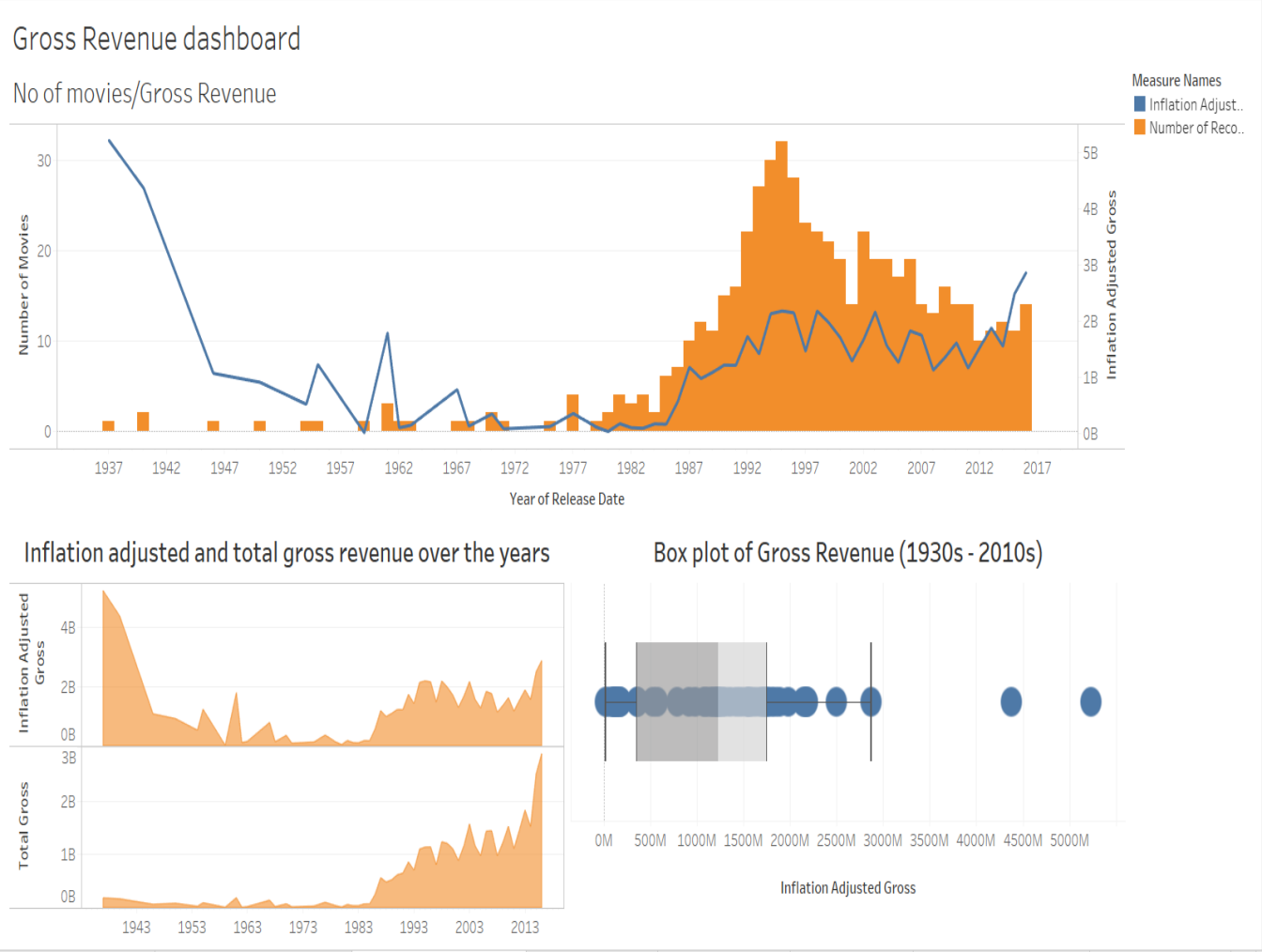
a lot of inconsistencies down the line, and there were too many missing values for me to successfully fill everything in.

3) Mismatched dataset lengths: My datasets had different lengths, but this did not pose a major problem to creating the story I later decided on

4) Learning curve of software: Tableau prep builder provided a simple option for cleaning data prior to creating visualizations in Tableau Desktop, so I decided to use it. However, I had never used it before, so I had to go through the process of learning how to properly use it to format my data into the way that I wanted

RESULTS/OBSERVATIONS

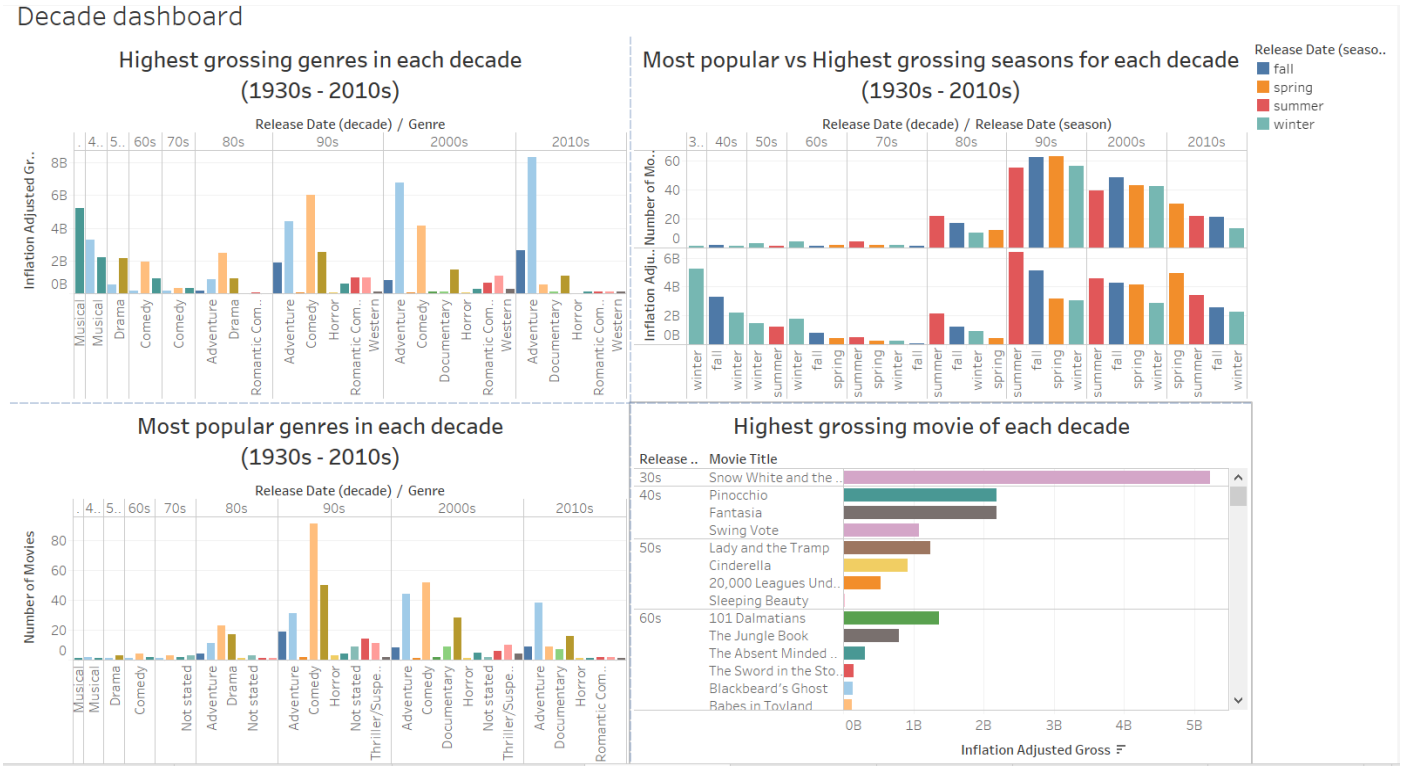
1) GROSS REVENUE DASHBOARD: This dashboard enables us view the yearly comparison between the number of movies released, and the Inflation adjusted gross revenue. It also shows the comparison between the total gross revenue, and the inflation adjusted gross revenue, as well as a box plot of the inflation adjusted gross revenue distribution



From the gross revenue dashboard, I made the following interesting observations:

- There were gap years. There seemed to be some gap years where no movies were released, but this could be due to the amount of time it takes to actually produce and release a movie
- Though movie production has increased, Disney seems to have made more money in the past compared to now. Interestingly, the highest Disney movie earnings ever, came from about three of their earliest movies which did exceptionally well
- Between 1942 – 1945, revenue dropped drastically. I observed that this was around the time World War 2 happened, and many of the Disney animators had become part of the war efforts. Movie production dropped too, and the movies that were previously produced and were released during this time did not do well.
- Mostly, when movie production increases, revenue increases too but there have been lots of ups and downs. This can be traced to the fact that though you can try to predict how well a movie will do in certain demographics based on past data, human nature tends to have a bit of unpredictability so there may be a bit of error, and a movie that was expected to be a box office success may fail to meet that expectation.
- Peak movie production was in the 90s

2) DECADE DASHBOARD: This dashboard explores different characteristics of Disney data set by decade



From the decade dashboard, I made the following interesting observations:

Up to the 60s

- most popular release season & highest grossing seasons: fall and winter

In the 70s and 80s

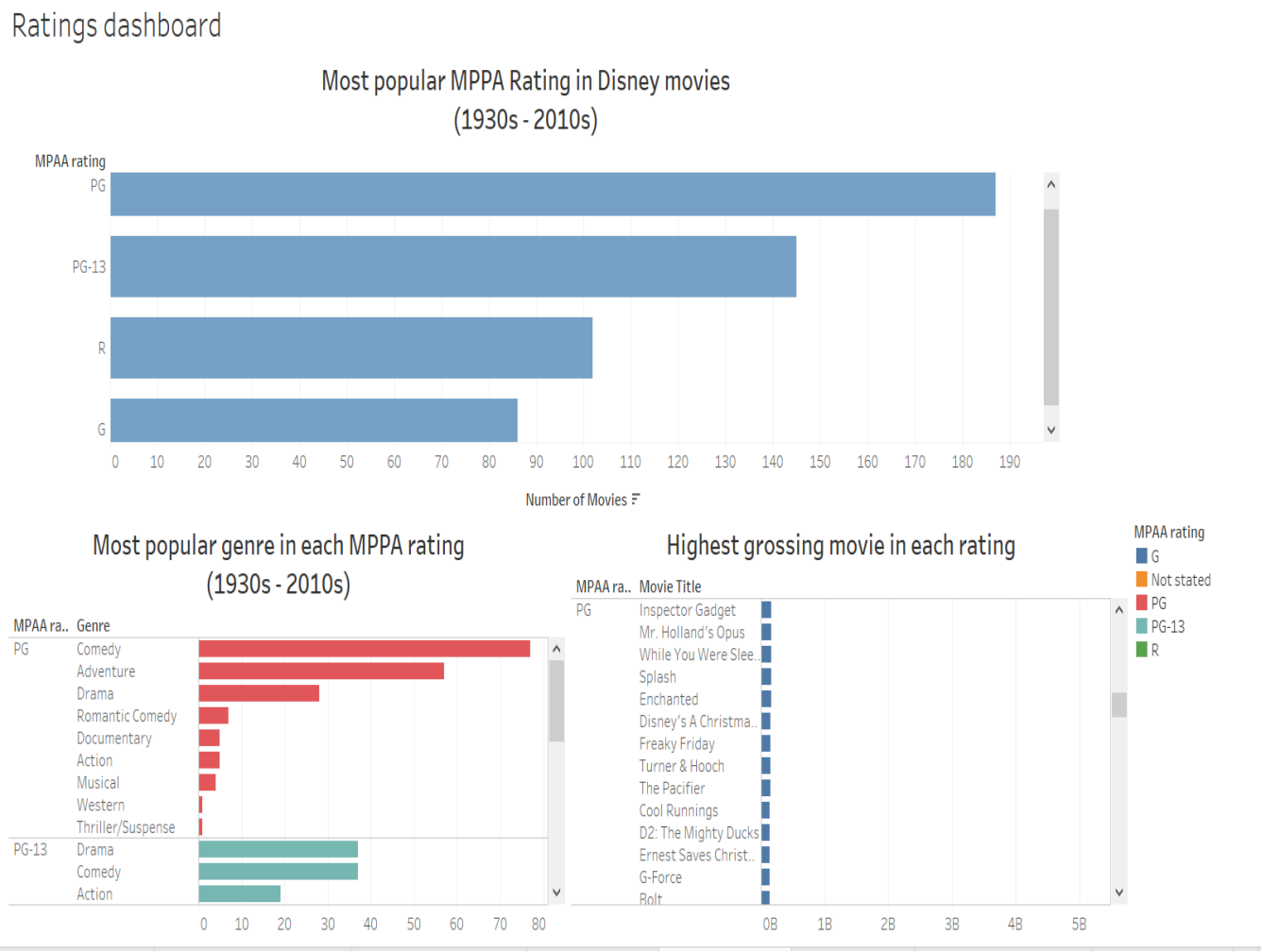
- summer took over as the most popular release season and highest grossing seasons
- spring was the least grossing/lucrative season

Since the 90s & 2000s

- spring and fall became the most popular release times
- summer remained the highest grossing season, till the 2010s when spring took over

In the 2000s, adventure was the highest grossing genre even though comedy was the most popular genre. The highest grossing movie was Pirates of the Caribbean

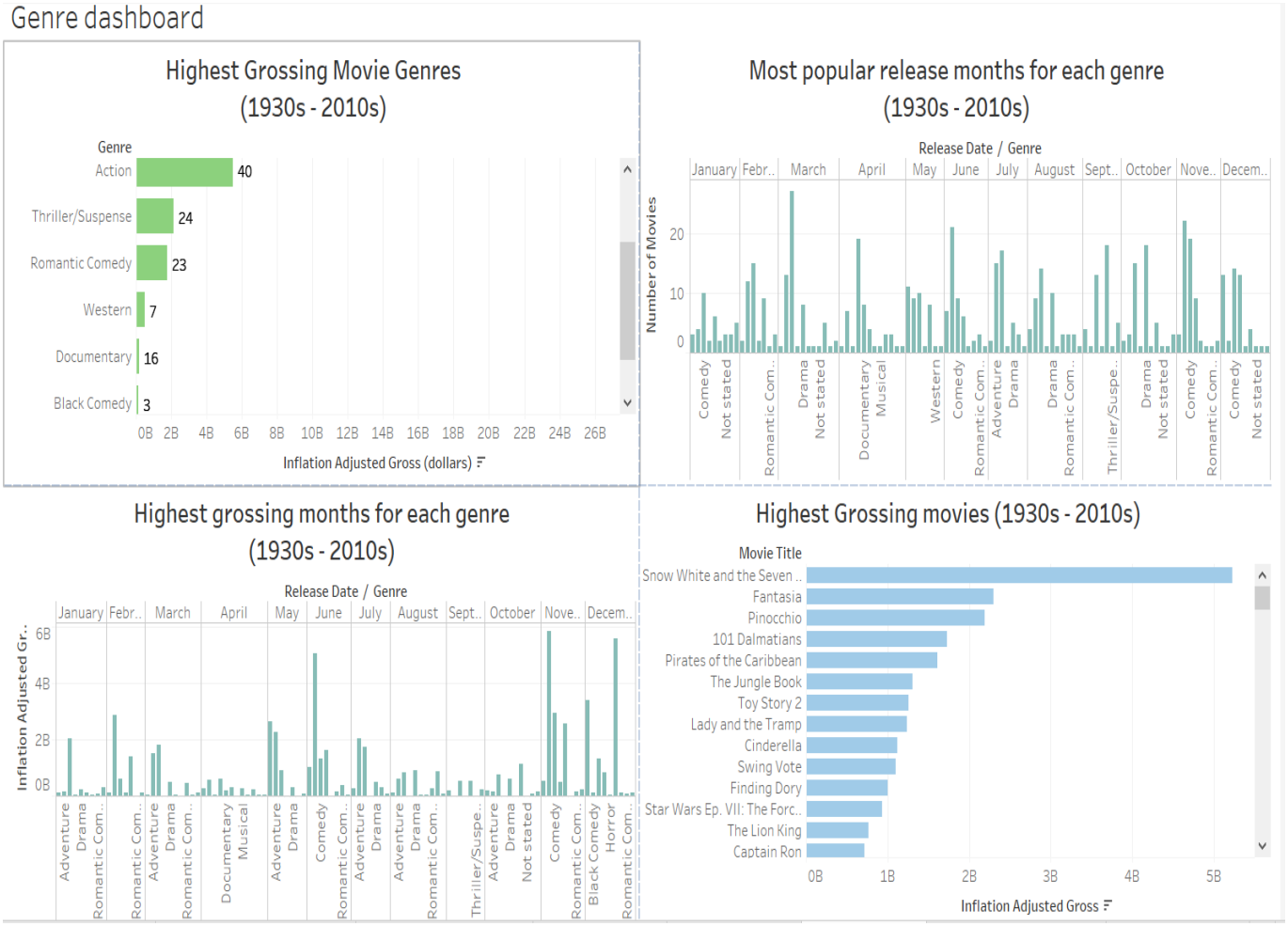
3) RATINGS DASHBOARD: This dashboard explores the MPPA rating dimension of the dataset



From the ratings dashboards, I made the following interesting observations:

- Disney makes PG movies the most overall
- Majority of R rated movies fall under the Drama & Comedy genre
- Most successful R rated movie: Pretty woman with Julia Roberts and Richard Gere

4) GENRE DASHBOARD: This dashboard explores the genre dimension of the dataset

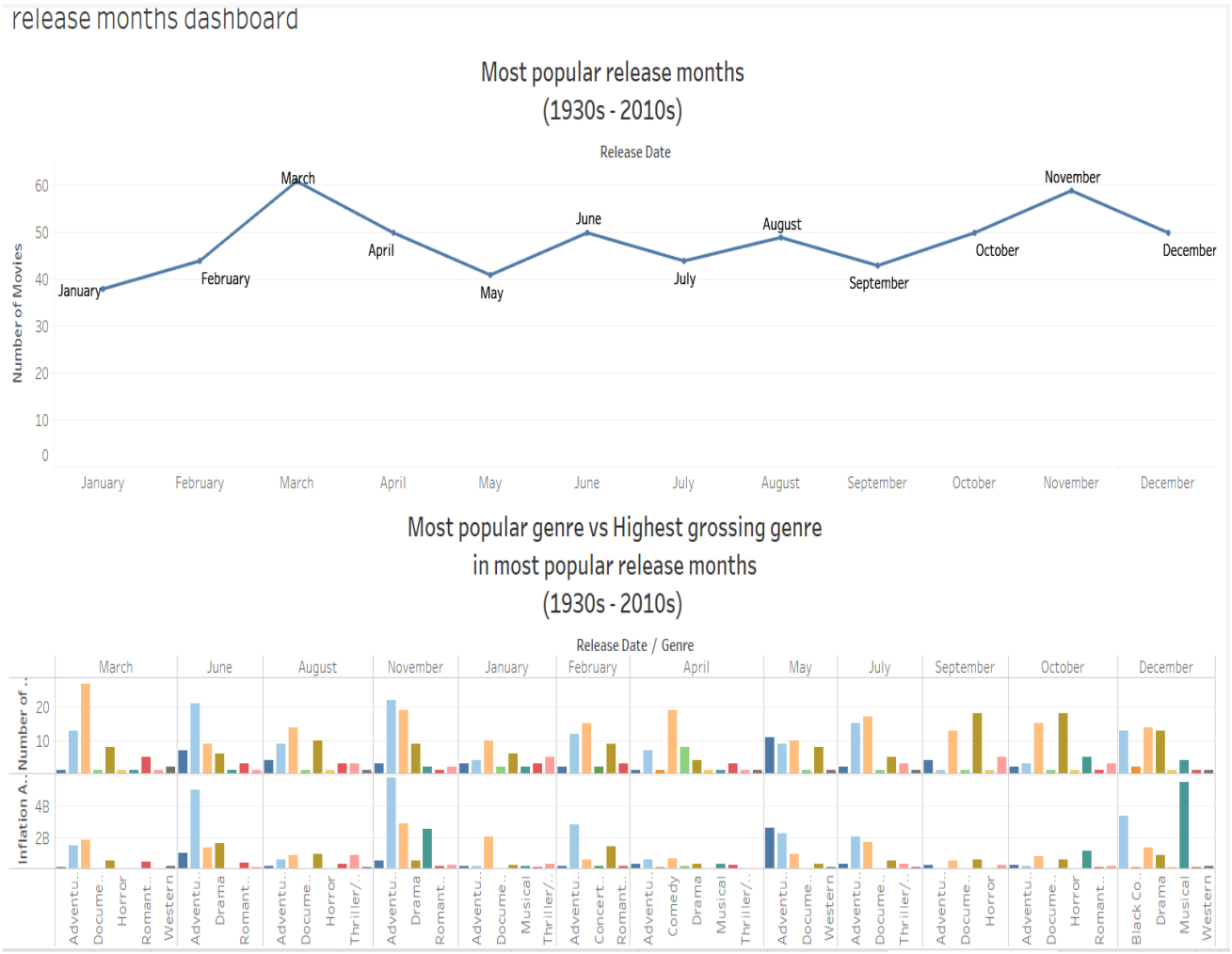


From the genre dashboard, I made the following interesting observations:

- Adventure movies are the highest grossing genre at Disney, but Comedy movies are the most popular, followed by Drama
- Most Drama movies are released in September and October. There could be something about the colder months that makes people desire more intense, emotional content. The highest grossing drama movie was surprisingly the Lady and the Tramp – an animation
- March has the highest number of romantic movie releases. I had expected this to be February, since it was considered the month of love, and seemed like the time you would expect

romantic movie releases to be plentiful. The highest grossing romantic movies so far have been Pretty Woman and The proposal

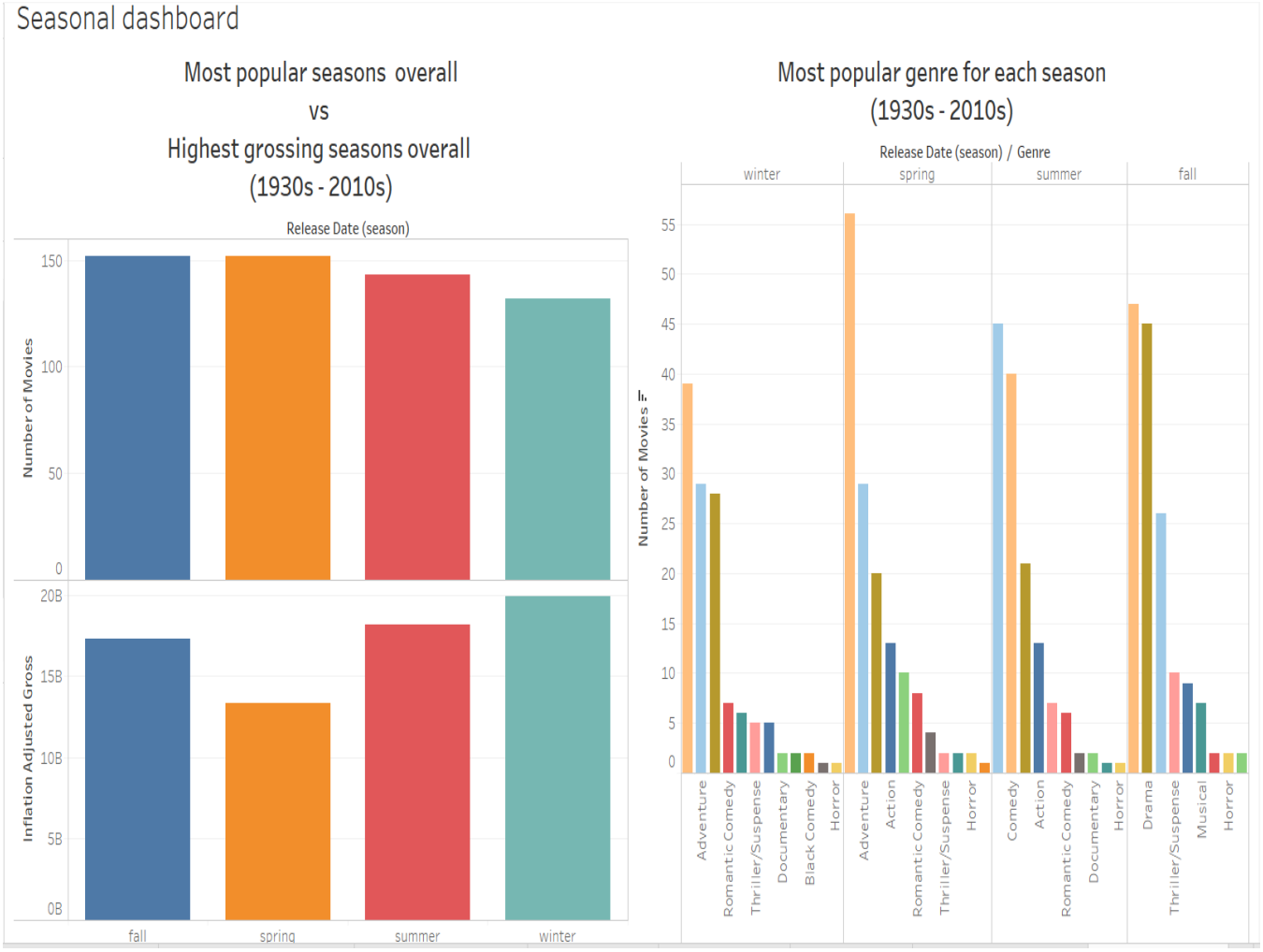
5) RELEASE MONTHS DASHBOARD: This dashboard explores characteristics of the data set from the perspective of release months



From the release months dashboard, I made the following interesting observations:

- Most popular release months overall – March and November

6) SEASONAL DASHBOARD: This dashboard explores the dataset from the perspective of seasons



From the seasonal dashboard, I made the following interesting observations:

- Overall, most popular season to release a movie – fall and Spring
- Highest grossing season overall – Winter. Outlier values may have contributed to this effect
- Most popular genre released in the Winter – comedy
- Different genres are released all through the year, but have certain seasons where they peak or are most popular