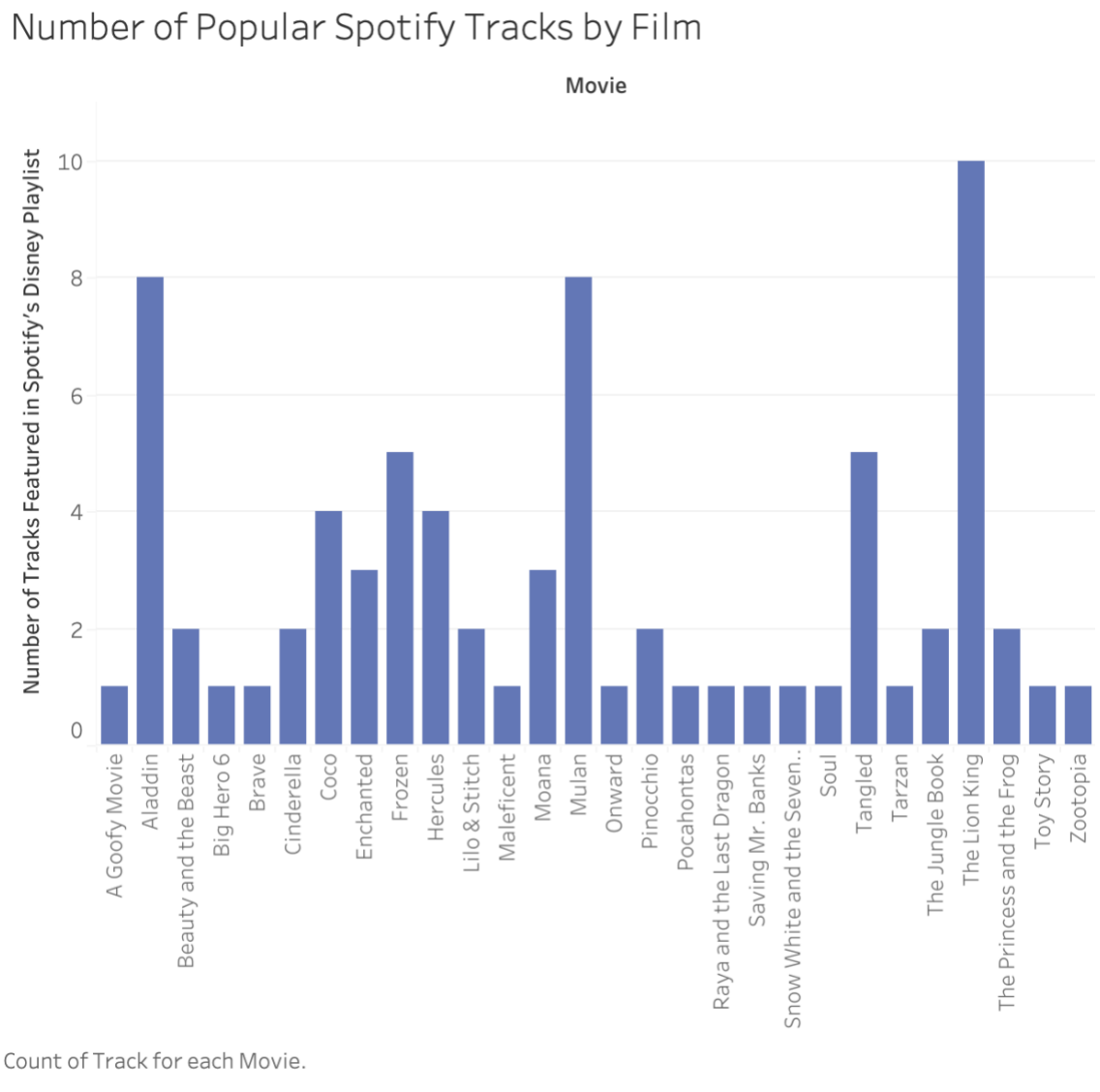
Milestone 5 – Visualizations

1. Popular Tracks by Film

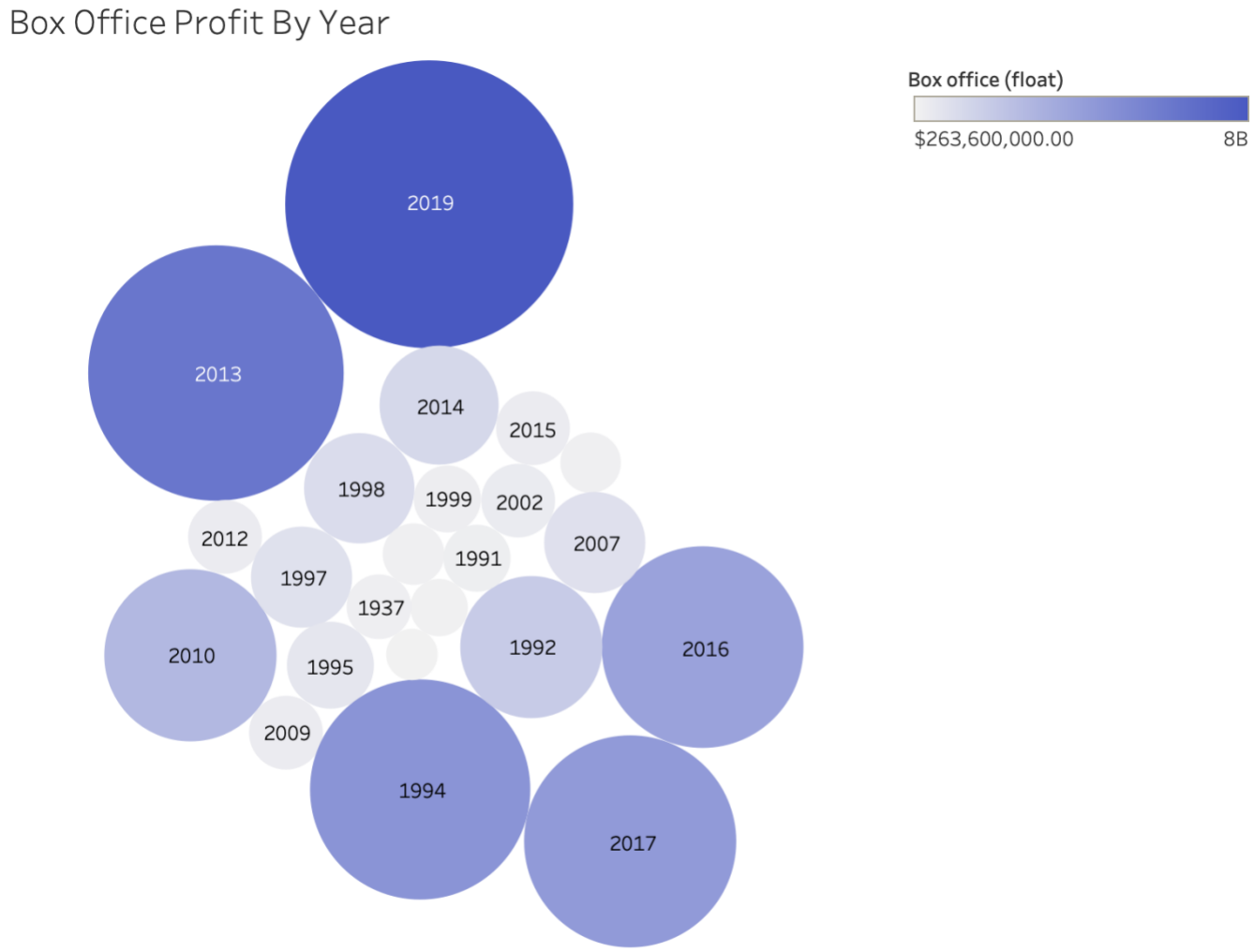
This visualization takes a look at the number of songs represented from each movie in Spotify's featured Disney playlist. This does not represent every Disney song on Spotify, only the 75 most popular ones.



We can see that The Lion King has the most songs featured on this playlist with 10. Aladdin and Mulan each have 8. These three movies have one thing in common: they have had a recent re-release. Since the songs have different sets of artists, they are considered as separate entities in our dataset. That means that there are some songs that may be represented twice.

2. Box Office Profit By Year

Our second visualization looks at the amount of box office profit was made each calendar year. This takes the sum of profits from each movie release in a given year and compares the total against that of other years.

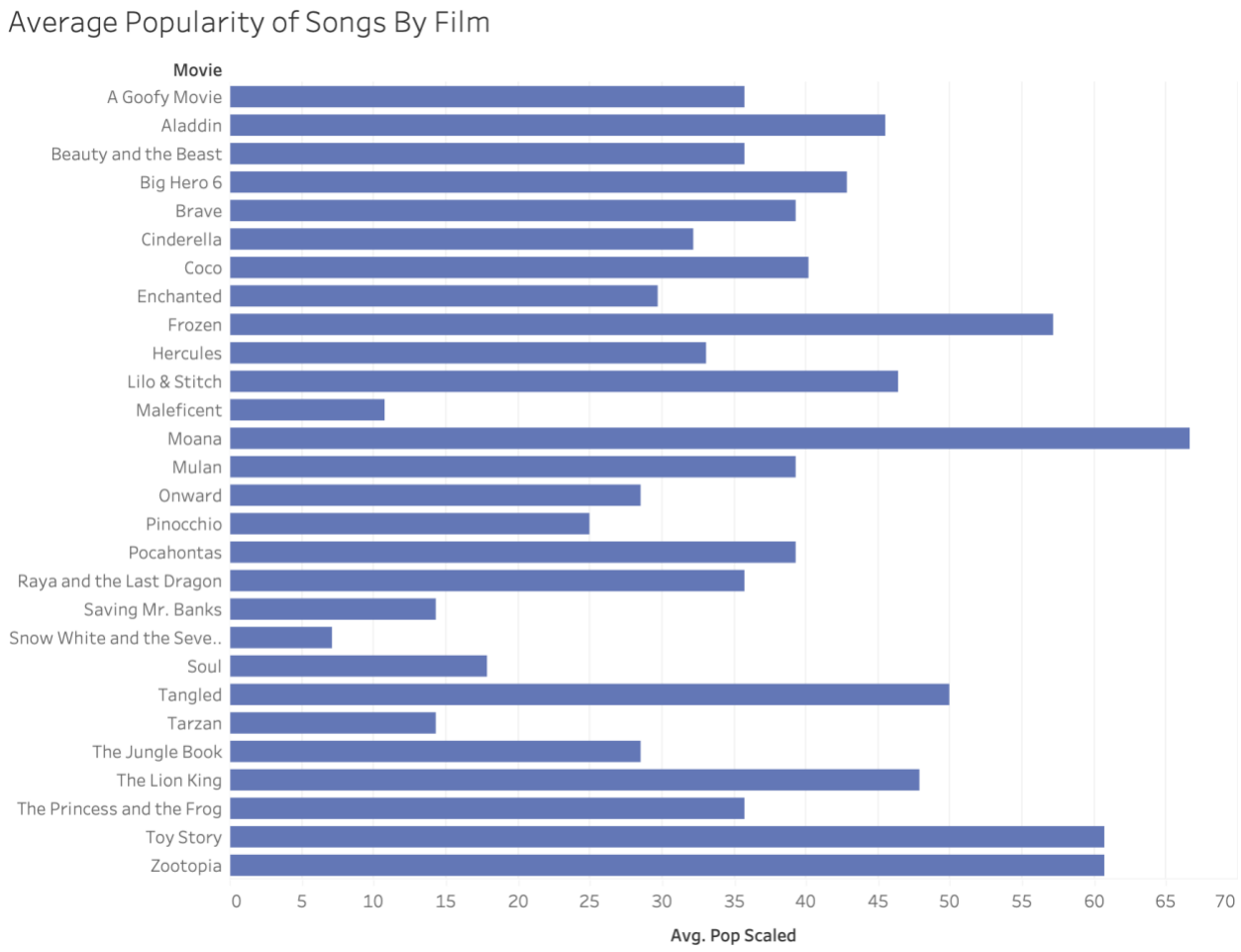


Year. Color shows sum of Box office (float). Size shows sum of Box office (float). The marks are labeled by Year.

We can see that 2019 was the most profitable year for The Walt Disney Company in terms of Box Office Revenue. This was the year that the live action versions of both Aladdin and the Lion King were released. Years in which no films were released are not included in this data.

3. Average Popularity of Songs By Film

This third visualization takes a look at the average popularity of the songs from each film. This computes the average scaled popularity of every song from a film and measures it against other films. This data is specifically looking at the songs on Spotify's Disney Hits playlist and therefore does not include less popular songs or scores.

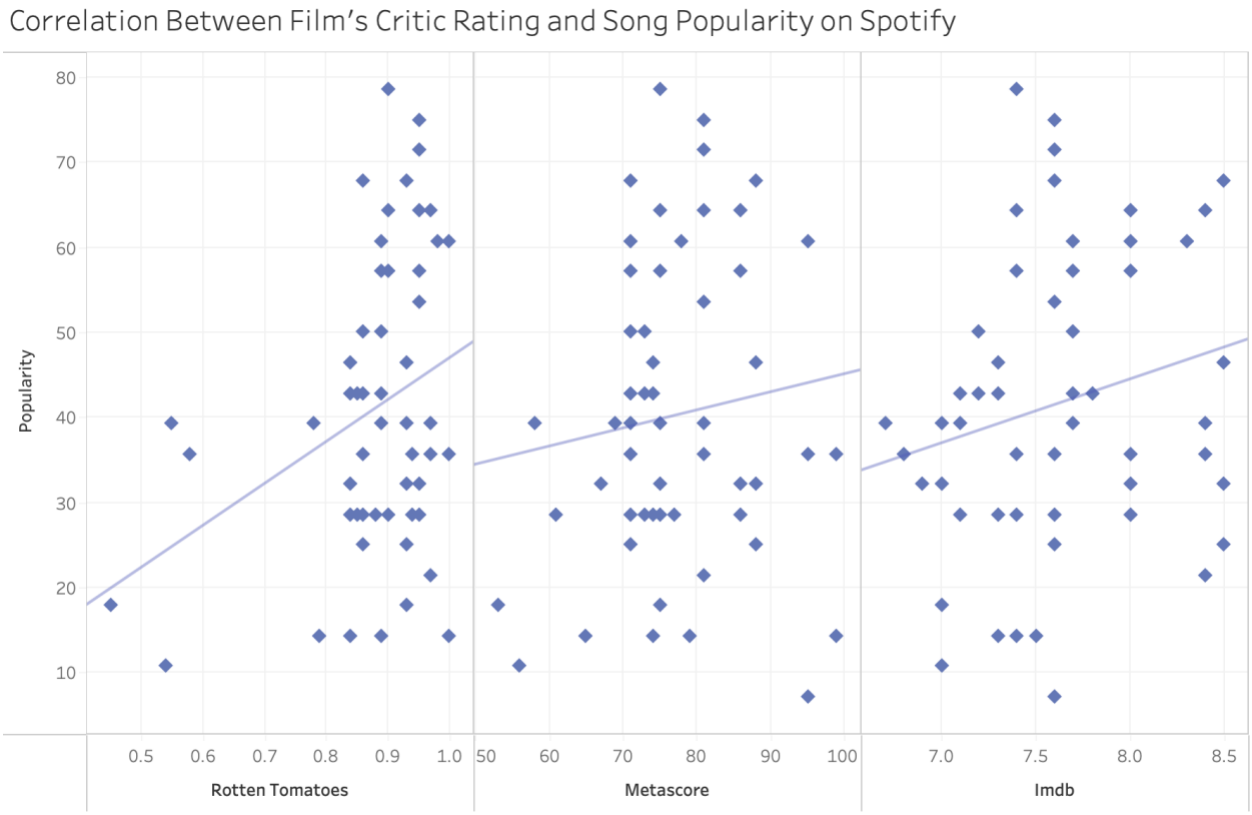


Average of Pop Scaled for each Movie.

We can see that the songs from Moana have the highest average popularity on Spotify. This is followed closely by Toy Story and Zootopia. The songs from Snow White and the Seven Dwarfs have the lowest average scaled popularity.

4. Film's Critic Rating Vs. Song Popularity

Our next visualization looks at critic ratings from a few different sources: Rotten Tomatoes, Metascore and IMDB. These scores are cross-referenced with the popularity of the songs from the films.

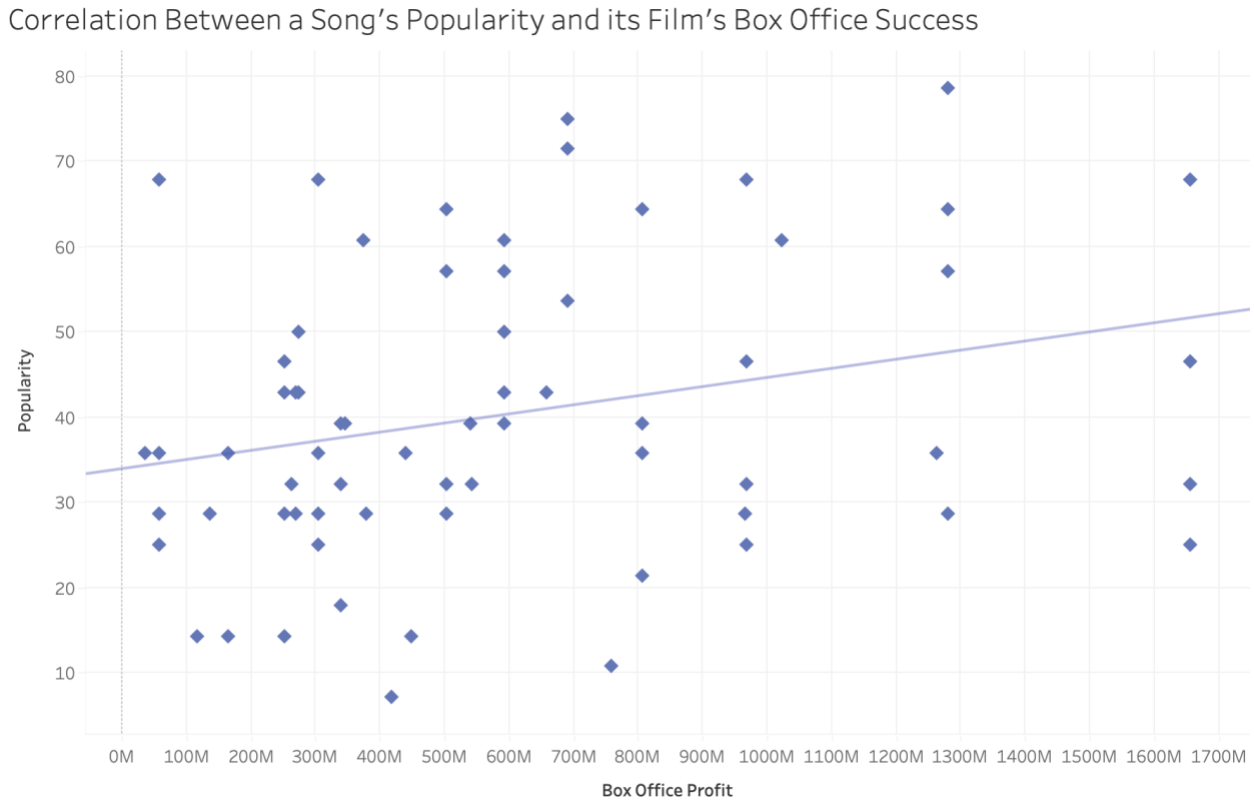


Rotten Tomatoes, Metascore and Imdb vs. Pop Scaled.

We can see that the ratings from Rotten Tomatoes are the most highly correlated (p=0.01) with the film's songs' popularity on Spotify. The ratings from IMDB (p=0.13) and Metascore (p=0.4) show a slight positive trend but are not statistically significant. This data show us that a high score from critics on Rotten Tomatoes is more likely to be correlated with a song having a higher amount of listens on Spotify.

5. Song Popularity Vs. Box Office Rating

Our last visualization takes a look at the correlation between a song's popularity on Spotify and its film's box office success in theaters.



Box office (float) vs. Pop Scaled.

There is a statistically significant positive correlation (p=0.03) between song popularity and box office success. This is intuitive, as patrons who are seeing movies in theaters are also likely to stream their music. The correlation could possibly go both ways, as consumers who hear songs they enjoy might also be more likely to see the movie it is from. Proving causation in either direction would require additional research into the subject matter.

Project Summary

For this project, I collected data from three different sources in order to research popularity and profit of various Walt Disney films. I collected a CSV file, an HTML table, and an API source. I used the tools available to me, including BeautifulSoup and Spotipy, in order to parse this data and turn it into a functional dataset.

Throughout this course I have learned how to interact with data from novel sources. Using different packages in Python, we are able to collect data from sources that are not nicely formatted CSV files. I learned how to connect to an API and collect necessary information, and I also learned how to visually inspect HTML to find the pieces that I need.

Once the data was cleaned, it was uploaded into a database using SQL. Each dataset was uploaded individually and then merged into one final dataset to use to perform analysis and create visualizations.

The main question I was investigating was whether song popularity and film success were correlated with each other. There is a statistically significant positive trend in the data, and we also found correlations between song popularity and film critic scores. I was able to create visualizations to support these claims, as well as some additional visualizations that graphically presented information regarding the individual films and box office profit by year.

The datasets regarding film information are public and there are no ethical implications that would necessarily arise. The third dataset I acquired using the Spotify API does present the possibility of running into complications of that nature. The Spotify API allows us to access a user's profile, liked songs, and much more. For the purposes of this project, no user data was collected, however in a project that did focus on individual users, steps would have to be taken to protect their privacy and ensure that all data was collected consensually.