45 1/2 / 50

Project Grading Rubric

Points Comments

Group #: 10

- Notebooks and Code (25 Points) 23 1/2 / 25

  See HTML Feedback for Details.

  Points in nbgrader are out of 50, so divided by 2 for points recorded here.

  Explora: 14/14
  Project: 32.25/36

- Essay and Submission (25 Points) 22/25

  See comments within essay below

  - Well Written and Coherent Essay (8 points) 6 1/2 / 8

    - Sets Context — Lacking foundation of source data, but motivation/intro was good.
    - Leads Reader and uses structure to organize — No structure, transitions abrupt. (looks like "you do this, I'll do that, and we will put it together")
    - Hits Target Audience of Non-Expert — Good job here, but discussion of 'popularity' needs to be up front
    - Paragraph Structure and Grammar 3 1/2 / 5 Mostly very good on paragraphs and grammar

  - Develops Interesting Questions (5 points) — more of a "more info on the data and what it tells us"

  - Effective Visualizations (5 points) 5/5 Terrific here.

    - Titles, axis labels, useful legends ✓
    - Allows reader to effectively compare data and glean information ✓

  - Interpretation of Visualizations to Support Answering of Questions (5 points) 5/5 Yes.

  - Submission and Following Instructions (2 points) 2/2 Full slate of files submitted.

Student Assessment
- While not identical, the two partners assessed each other consistently with "red flags" re

Grading

Summarizing the deliverables

1. The first deliverable is the current notebook, which encompasses phases 1, and 2. It should be executable by me and by our TAs, and should not be dependent on your own file system and its paths.
2. The set of CSV files corresponding to the tidy data result of your processing in this notebook.
3. The `movie_explore.ipynb` is another notebook deliverable.
4. A PDF of your essay comminicating what you learned in the project.

The project will be graded out of 50 points, split evenly with 25 points for assessment of the two notebooks (and their output), and 25 points for assessment of your essay and the work entailed.

Assessment of the Notebooks

- Correctness of the Functional Dependencies determined for the source data,
- Correctness of code,
- Appropriate design and creation of functions, including use of appropriate parameters,
- **Use** of those functions in higher level functions,
- Avoidance of global code or global variables outside of functions,
- Answers in markdown and code solution cells,
- Code documentation:
  - docstrings
  - inline comments
- Correct output (the CSVs)

Assessment of the Essay

- Presence of two or three interesting questions that have sufficient depth and can be answered by the data,
- **Good** visualizations that help understand the data and answer the questions.
  - Must avoid being too busy, or in other ways inappropriate to what can be well conveyed in a figure in an essay,
  - Must be appropriated titled, with axis labels and clear units,
  - Must allow the reader to compare data and interpret results for themselves.
- Essay must be well written, including:
  - Good use of headers and structures, giving it a clear organization,
  - Good grammar, punctuation, use of terms,
  - Set a background/context/introduction,
  - **Develop** the questions to be answered,
  - Present the visualizations and describe what the reader is seeing and help **interpret** the results,
  - Hits the target audience of a non-expert.

Henry Robinson and Mary Price

Professor Bressoud

CS181

9 April 2021

The Pandemic's Effect on Film

The SARS-COV-2 (COVID-19) pandemic has caused many extreme effects on our society. Some businesses have suffered severely from a lack of customers while others thrive from people's inability to leave the comfort of their home. One such industry that has had its change from the pandemic discussed a lot in media is the film industry. Are more people watching movies because they are stuck inside, or is their incapability to go to a theater caused them to pick up other hobbies? Also, is there now a change in demand due to the type of film? In order to answer some of these questions, we went through movie data and created charts that show the pandemics effect on different aspects of film.

The first data table we created (Fig 1.1), we can see three different sections represented on the table, divided among the year, quarter, and month of the release of a film. The top row represents the total popularity between all of the films for that month, the one in the middle shows the total number of votes the films received, and the final row shows

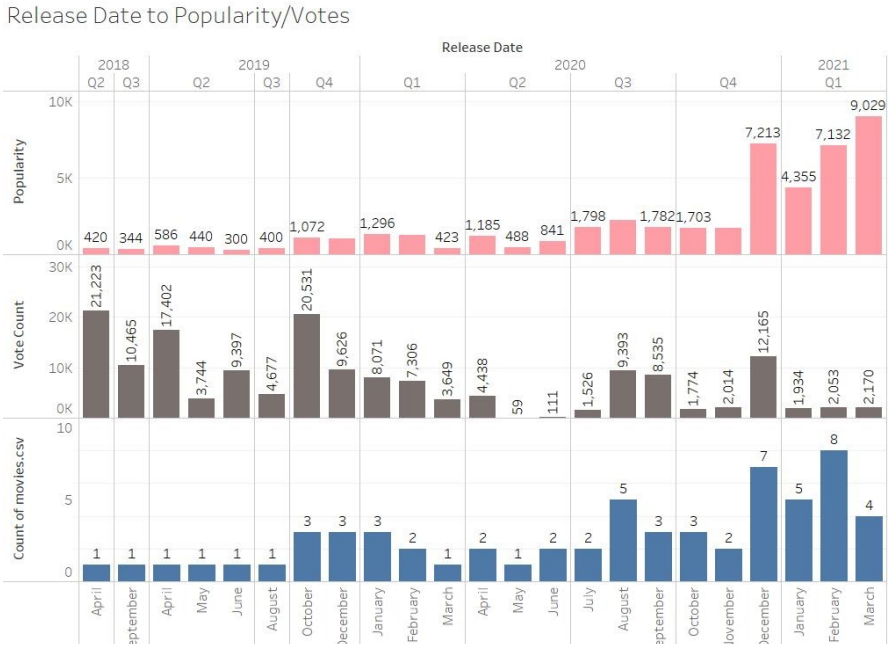


Fig 1.1

Sum of Popularity, sum of Vote Count and count of movies.csv for each Release Date Month broken down by Release Date Year and Release Date Quarter. The data is filtered on Release Date, which ranges from 6/10/2014 to 3/23/2021.

the total number of films represented in the data.

Through that lens, we can see the fluctuation of data for the films over the years, and just how much the 2020 COVID-19 pandemic affected the film industry. While the data may be a bit skewed, each month before 2020 only having roughly one to three films accounted for in the data, the popularity and votes show a decline. This is seen with how films in 2019 had large spikes in popularity, particularly in April and October of that year, despite the data only accounting for one to two films during those periods. That fluctuation in popularity shows that people were still interested in going to the theater before the pandemic started.

I do think this is talking of pandemic related.

Moving ahead in the data to around March 2020, when the lockdown began to hit, the number of films being released and seen plummeted hard. This was due to factors such as theaters being shut down and film studios not wanting to release films that would lose out on revenue. These actions are reflected in how popularity and releases are low at the beginning of quarantine. Due to factors such as streaming services and films began to release again around the end of 2020, we can see an increase in films towards the end of 2020.

An over statement since popularity was only factor for what movies are in the set.

Although we see an increase at the end of 2020 and beginning of 2021, there are still negative effects for votes, despite the release of films being higher than the larger months of 2019. We can see that there are lower vote counts overall in late 2020 and early 2021. Even with the lower number of vote counts, the popularity of these films had a higher total overall. This may be due to the data totaling popularity when the data is extracted from the website, such that the number of mentions and searches for these recent films are higher, while the actual engagement of voting for data is lower.

This should be up front, not trailing

Moving onto a different set of data, we analyzed how many actors were in certain genres from 2019-2021, which is recorded in Fig 1.2. The left side of the table has the number of actors

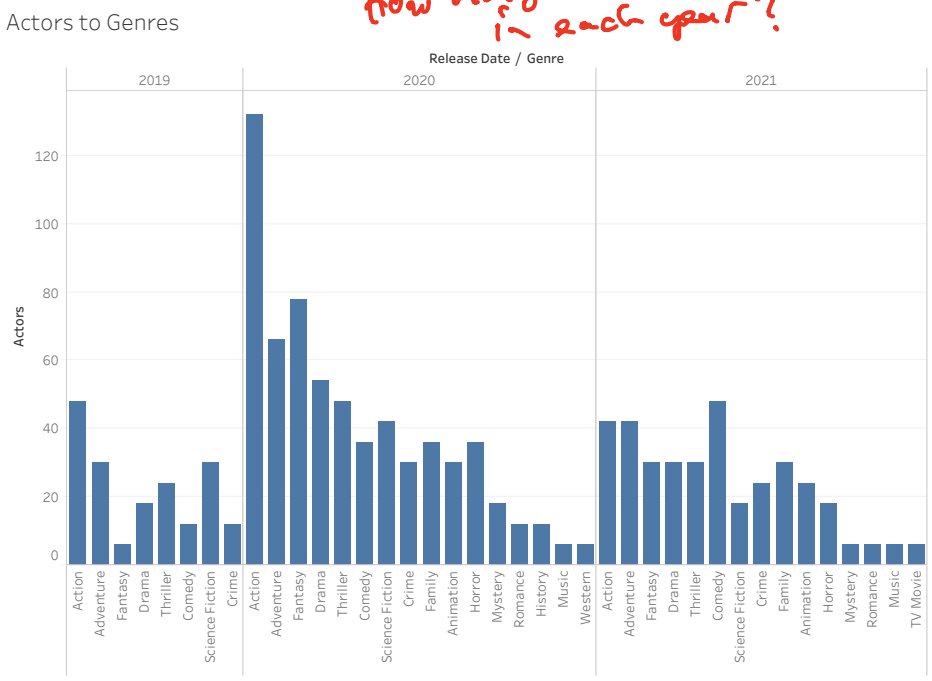
bar chart

Again need to know processing yielded six actors per each of the 64 movies in the sample

Another good visual. Allows comparison.

How many movies are represented in each genre?

in each of the bars. The graph is split up into three different sections, one for each year, and each of those sections have the genres for that year.



Count of actors.csv for each Genre broken down by Release Date Year. The view is filtered on Genre and Release Date Year. The Genre filter keeps 17 of 17 members. The Release Date Year filter keeps 2019, 2020 and 2021.

Fig 1.2

Looking at the table, we can see that there were not many actors in 2019 overall, this being seen with how low every bar is for that year. This is most likely due to the same error in our data that we had previously, which was that there are less movies in the 2019 dataset than in the 2020 dataset. Looking past the small number of actors, we can still see that "Action" is the genre with the most actors. This remains the same in 2020, but changes in 2021 with "Comedy" containing more actors.

Can we "normalize" for this effect to get proportions?

This swap in genre from 2019/2020 to 2021 can be due to either not having enough data for the year, because we are not even half way done with 2021, or a higher demand for "Comedy" films. Looking at the former, it is possible that we do not have enough movies from the year to determine which genre has the most actors for the year. Although this is possible, if we look at "Action" for 2019 and 2020, we can see that this genre has almost double the number of actors compared to the second highest genre for their years. Going back to 2021, the "Action" genre is not far behind "Comedy", but it's also tied with "Adventure", 2019's second highest genre and 2020's third highest. Hence, we can see that there is probably another factor causing

Good work on interpretation

this change in genre and that factor could be a higher demand for “Comedy” films. This would make sense given that 2020 was an extremely hard and taxing year for everyone, so a changing demand in genre could be due to people wanting to watch more comedic films in their spare time. The only way to determine which genre has the most demand for actors in 2021 would be to wait until more movies released in the year.

These data tables showed two things, that the film industry started off rough at the beginning of the pandemic but is now thriving and that the demand for actors in genres could be changing due to the pandemic. With these data tables, we can answer our previous questions, how is the film industry doing compared to previous years and is the demand for certain types of films changing. The beginning of the pandemic caused a small number of films being released, which lead people to believe that industry was dying, but now many films are being released. With the option to either stream movies in your home or watch them in the theater, newly released films are easier than ever to access. Also, although we do not have much data for 2021 at the moment, there is a shift in demand for more actors in “Comedy” genres now which might reflect people wanting to see more comedic films after the pandemic. Therefore, the pandemic has not had a negative effect on the film industry like we were led to believe at the beginning of quarantine.