Viewing the Data Visualisation

You can open the index.html file and everything should work. If not you can also access the visualisation at <https://bryrant93.github.io/OscarVisualisations/>.

Data Analysis Folder

dataAnalysis.py

I did all my data analysis in VScode using the jupyter notebook extension to easily see the charts. As long as dataAnalysis.py is opened in VScode with the jupyter notebook extension enabled, the analysis should be viewable by clicking “Run Cell” at the top of the file.

Just in case that fails to work I have included all the same material at the bottom of this README.

dataManipulation.py

Additionally I have included the file I used to manipulate my downloaded CSVs and create the CSVs I needed for my visualisations. But as it was an iterative process and I didn’t save every step of the process the code does not run or provide an exact impression of the route I took.

CSV files

* metacritic\_movies & Oscars\_df: The two original CSV I downloaded from Kaggle.
* reviewNumberFrame1990 & reviewNumberFrame1996
* winnerLine: Simple CSV for my first visualisation.
* oscarFrameFinal: The CSV I constructed for my second visualisation

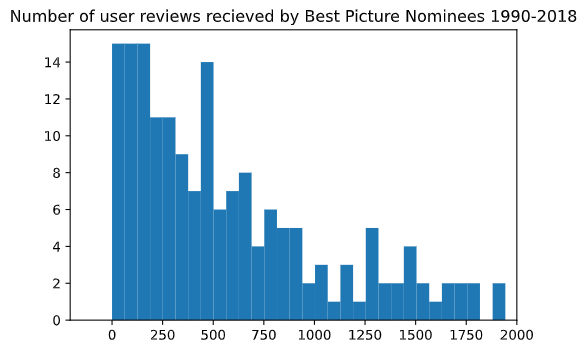
Backup Data Analysis

As the data I have chosen covers a large time period, and the further back you went the more incomplete the data became I had to decide on the period which had the most accurate and truthful data.

The following two histograms display the distribution of films from 1990-2018 by the number of reviews they received.

Each bin in the critic reviews is approximately 20, while each bin in the user reviews in approximately 60.

Please note the user review histogram has been dramatically truncated, as the second half of the chart was sparsely populated.

 Chart, histogram

Description automatically generatedA picture containing chart

Description automatically generated

These were not hugely enlightening, so I decided to examine the data in a scatter graph so I could see the distribution of number of films by year.

Chart

Description automatically generatedChart, scatter chart

Description automatically generated

In the scatter graphs above I noticed that there was a change in 1996 for Critic reviews, as the films in that period stopped averaging fewer than 20 reviews and jumped up to high 20s and only rose from there.

There was not as noticeable a trend for user reviews, but when displayed on the histogram below (1996-2018) it became apparent that there were approximately only 7 films in the period 1996-2018 that had 60 or fewer reviews. As in the entire dataset there was no film with a number of critical reviews higher than 60 I decided this was an acceptable period of time to explore. Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

I made histograms for the following section, but they showed nothing and have chosen to omit them. The scatter graphs below were far more enlightening and are the skeleton of what my final visualisation became.

Each grey dot represents the aggregate rating of a film by either group, while the coloured dots represent the aggregate rating of the Best Picture winner for that year.

There is a trend in this period that before 2009 critics and users were around equal at 'predicting' the winner, with users holding a slight advantage. (Critics: 3 correct, Users: 5 correct). Then on and after 2009, the critic's accuracy jumps up and user's accuracy drops down. (Critics: 7 correct, Users: 1 correct)

There were changes made to Oscar voting procedures in 2009, which some claimed would cause more generic films to win the Best Picture, so I knew I had found something worth visualising and began from here.

A picture containing chart

Description automatically generatedChart, scatter chart

Description automatically generated