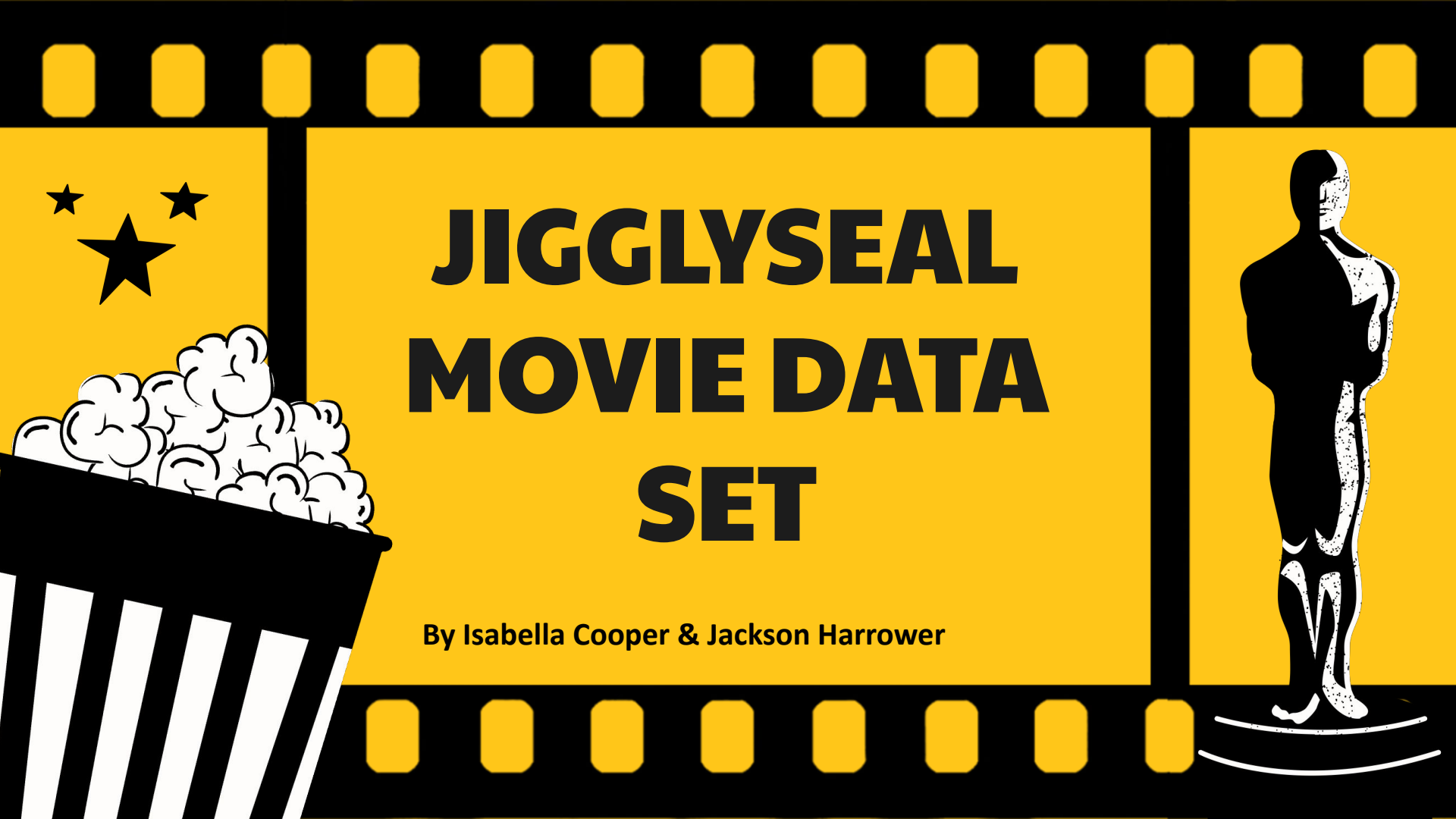

JIGGLYSEAL
MOVIE DATA
SET
By Isabella Cooper & Jackson Harrower

EASILY MISUNDERSTOOD FEATURES

numVotes

The number of IMDb votes for a movie indicates how many registered IMDb users have rated that specific movie

---

primaryTitle

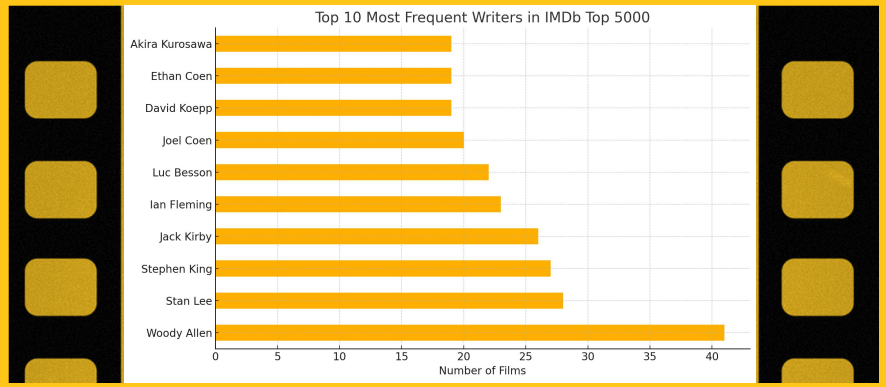
Just the title of the movie

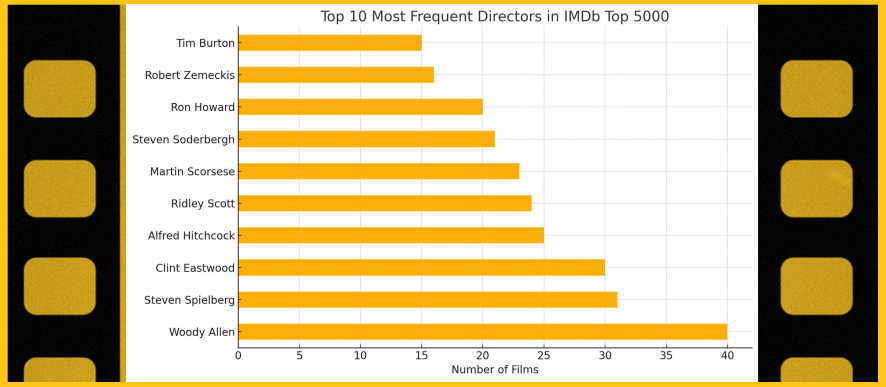
startYear

Just the release date of the movie

The questions

MAIN	What writers and directors are most likely to appear on the IMDb Top 5000 list?
SIDE ONE	Which directors tend to get the best average rating?
SIDE TWO	what is the most uncommon genre?



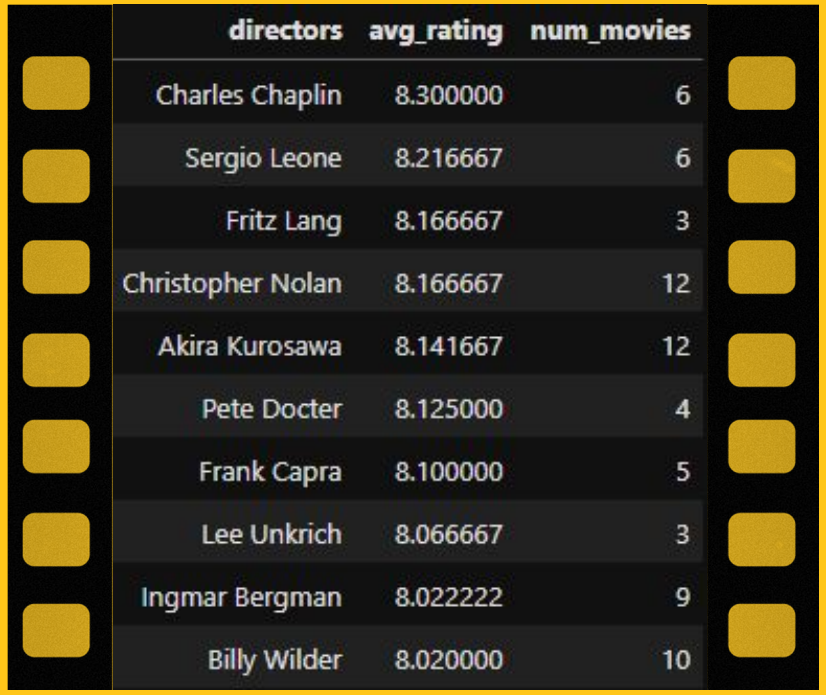


MAIN QUESTION

What writers and directors are most likely to appear on the IMDb Top 5000 list?

FINDINGS

To figure out which writers and directors are most likely to appear on the IMDb Top 5000 list, we looked at how often certain names showed up in the data. Not surprisingly, directors like Christopher Nolan, Steven Spielberg, and Martin Scorsese came up the most, which makes sense considering their strong reputations with both critics and audiences. On the writing side, Christopher Nolan, Fran Walsh, and Peter Jackson stood out. Their work on major franchises like The Dark Knight trilogy and The Lord of the Rings helped secure their spots. Overall, the data supports the idea that a small group of influential creatives consistently shape many of the top-performing films. The visualizations made these patterns easy to see and helped show how much of an impact these individuals have on a film's success.



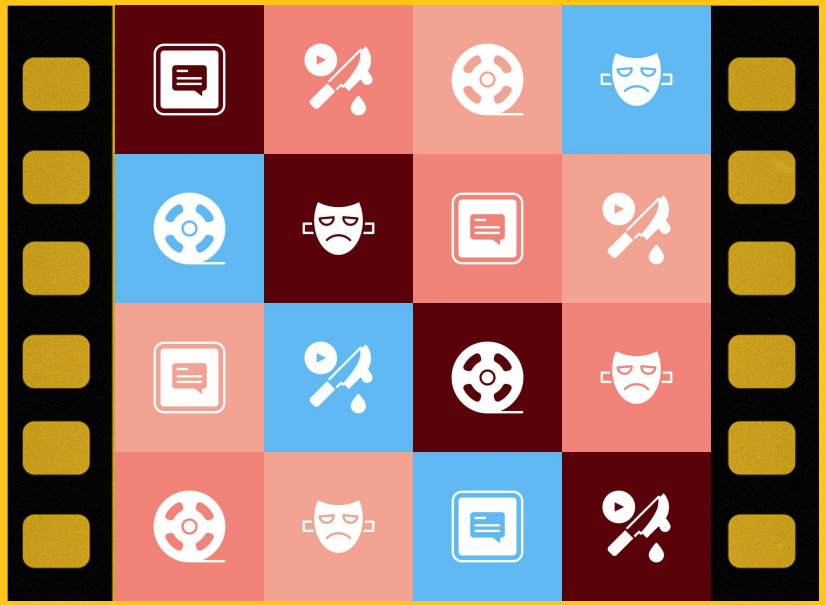
directors	avg_rating	num_movies
Charles Chaplin	8.300000	6
Sergio Leone	8.216667	6
Fritz Lang	8.166667	3
Christopher Nolan	8.166667	12
Akira Kurosawa	8.141667	12
Pete Docter	8.125000	4
Frank Capra	8.100000	5
Lee Unkrich	8.066667	3
Ingmar Bergman	8.022222	9
Billy Wilder	8.020000	10

SIDE QUESTION 1

Which directors tend to get the best average rating?

FINDINGS

The image shows the top ten directors based on their average IMDb ratings, focusing only on those who have multiple films in the dataset. Charles Chaplin comes out on top with an impressive 8.3 average across six movies, which really speaks to how consistently his work has been appreciated. Directors like Christopher Nolan and Akira Kurosawa also stand out, each with 12 films averaging around 8.16 and 8.14. That says a lot not just about the quality of their work, but also about how active they've been in shaping great cinema. What's interesting is that the list includes both classic and modern directors, showing that consistently high ratings aren't tied to any specific time period. All in all, this reinforces the idea that some directors have a real knack for creating films that stick with both critics and audiences.



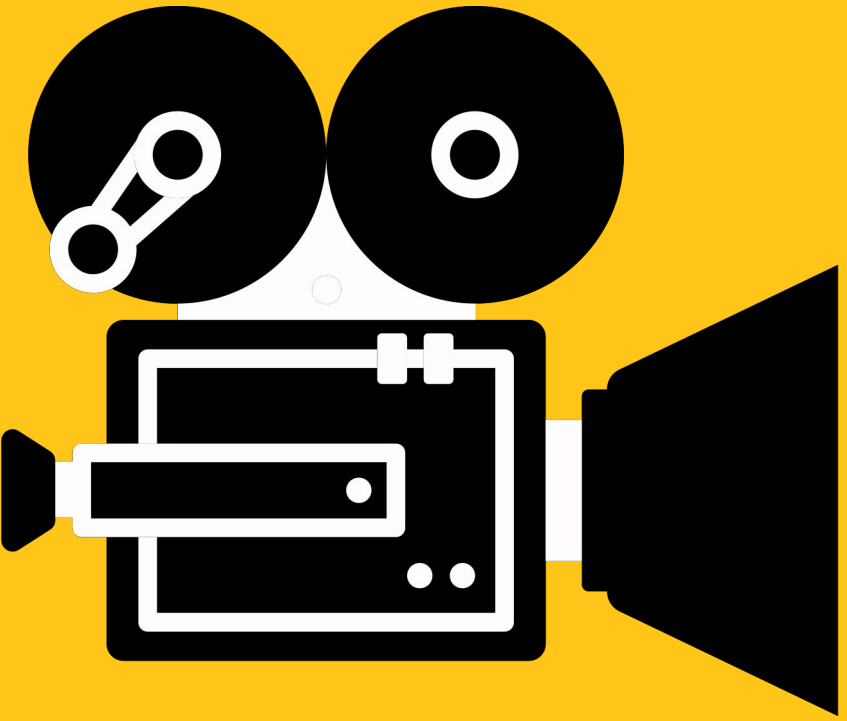
SIDE QUESTION 2

what is the most uncommon genre?

FINDINGS

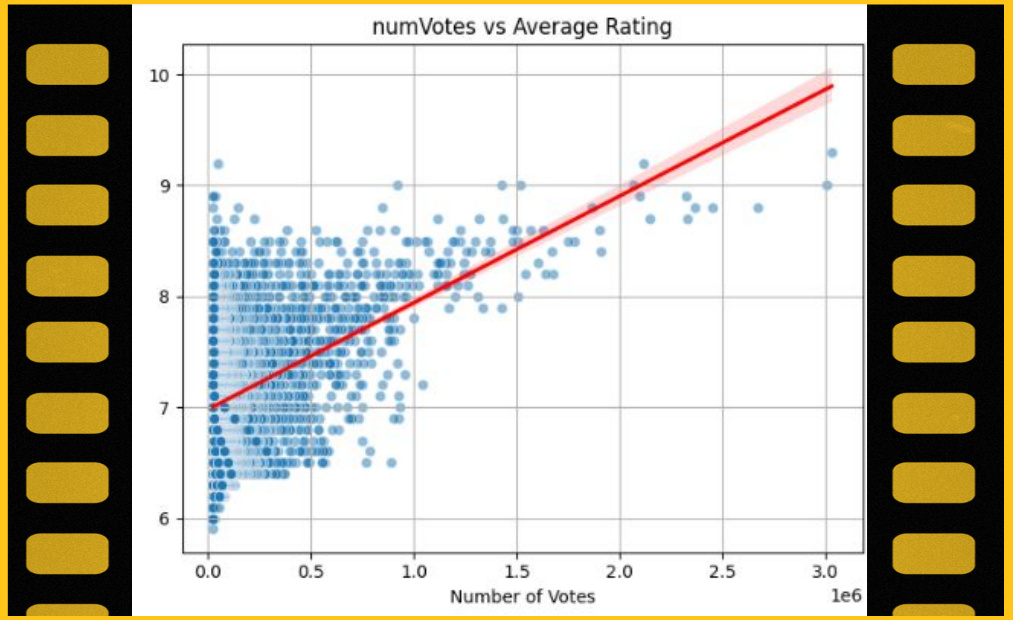
The 5 least common movie genres were as follows, with news having a very large lead over the others
News: 2
Film-Noir: 27
Western: 58
Musical: 65
Documentary: 77
For Contrast the top 5 most popular genres were as follows
Drama: 3136
Comedy: 1640
Action: 1129
Crime: 1112
Adventure: 917

Correlation



Linear Correlation with average Rating

startYear: -0.276
numVotes: 0.387
runtimeMinutes: 0.287



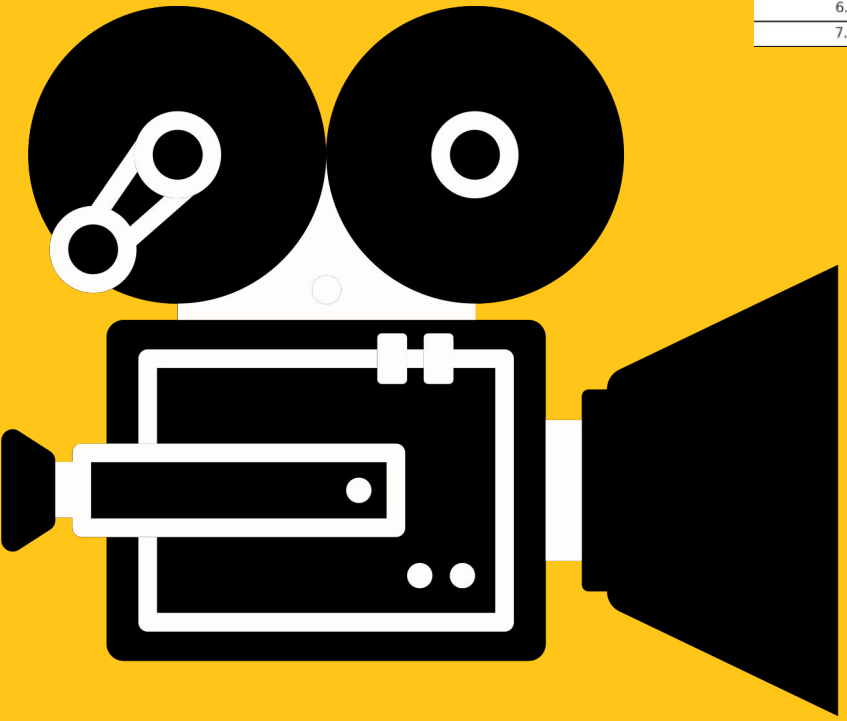
Correlation Analysis

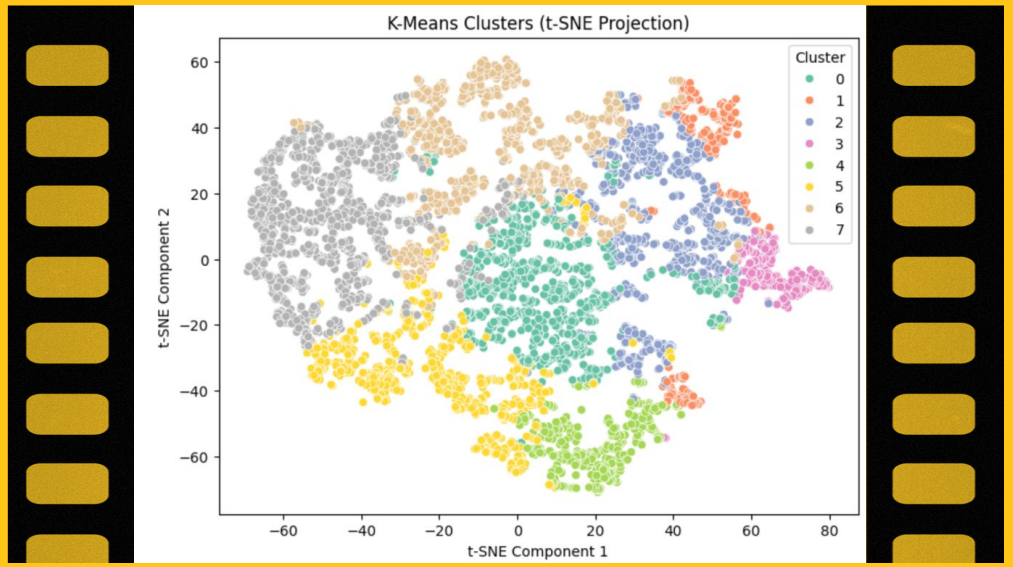
EXPLANATION

- The correlation analysis looks at how three factors—release year (startYear), number of votes (numVotes), and runtime (runtimeMinutes)—relate to a film's average IMDb rating.
- StartYear has a negative correlation of -0.276, suggesting that older films tend to have slightly higher ratings. This might be because only the most iconic older films remain well-regarded over time.
- NumVotes shows a positive correlation of 0.387, meaning movies with more votes generally receive higher ratings. This suggests that broader audience engagement tends to boost average scores, possibly due to a more balanced pool of opinions.
- RuntimeMinutes has a weaker positive correlation of 0.287, indicating that longer films might be rated a bit higher, though the effect is less strong.
- The scatter plot highlights the connection between vote count and rating, with the red regression line showing a clear upward trend. Overall, these patterns give us a better idea of which traits are loosely linked to success on IMDb's Top 5000 list.

Cluster Graph

Cluster	Avg Year	Avg Rank	Avg Rating	Avg Votes	Avg Runtime (min)
0.0	2007.67	1679.44	7.43	144560.0	98.85
1.0	2001.04	1196.13	7.72	126987.0	175.55
2.0	2007.53	1151.81	7.64	196530.0	129.6
3.0	2004.06	444.58	8.06	1020217.0	131.56
4.0	1958.21	1054.16	7.78	96105.0	106.52
5.0	1986.51	3132.4	6.86	74782.0	105.78
6.0	2013.02	3389.68	6.79	142561.0	125.92
7.0	2011.0	4123.08	6.49	82669.0	99.82





Link to notebook

THANK YOU!