**Final Project Summary**

**Project Summary**

**Tools Used**

* **Software:** Jupyter Notebook
* **Programming Language:** Python
* **Python Libraries:**
  + pandas for data manipulation and cleaning
  + matplotlib and seaborn for data visualization
  + numpy for generating placeholder data where necessary

The combination of these tools allowed for efficient data exploration, cleaning, and visualization, ensuring insights were derived systematically.

**Questions Asked**

1. How have average movie ratings changed over the years?
2. How does movie runtime correlate with ratings?
3. How does the number of votes correlate with average ratings?
4. How does the average rating vary across different ranges of number of votes?
5. How do ratings vary by year for the most popular movies?

These questions aimed to explore IMDb ratings comprehensively, highlighting trends in audience reception and movie characteristics.

**Insights Discovered**

* **Question 1:** Movie ratings have remained stable over the years, with no significant upward or downward trend.
* **Question 2:** Runtime has no significant correlation with ratings (-0.0004), suggesting that movie length does not affect audience reception.
* **Question 3:** There is a very weak positive correlation (0.0609) between the number of votes and ratings, meaning movies with more votes are only slightly more likely to have higher ratings.
* **Question 4:** Movies with the most votes (100,001+) have the highest average ratings (7.02), while those with fewer votes (101-1,000) have the lowest (5.76).
* **Question 5:** For the most popular movies (top 10% by votes), ratings fluctuate significantly over time, with standout films driving trends in certain years.

Overall, the analysis highlights audience voting patterns and rating stability across the dataset, offering actionable insights.

**Recommendations**

If this data were presented to a client (e.g., a streaming platform or production studio):

* **Focus on Popular Movies:** Movies with higher votes tend to perform better; leveraging marketing to boost audience engagement could improve visibility and ratings.
* **Optimize Content:** Runtime does not influence ratings significantly; focus on quality storytelling rather than adhering to specific runtime norms.
* **Analyze Trends by Year:** Invest in analyzing standout films in high-performing years to replicate success factors.
* **Encourage Audience Participation:** Platforms should emphasize user engagement (e.g., incentivizing reviews) to increase vote counts, as this can marginally improve perceived ratings.

**Future Work**

* **Incorporate Additional Features:** Include data on genres, directors, and budgets to expand analysis and refine insights.
* **Sentiment Analysis:** Use audience reviews to understand factors influencing ratings beyond numeric data.
* **Compare Across Platforms:** Examine how IMDb ratings compare to other platforms (e.g., Rotten Tomatoes) for cross-platform insights.
* **Year-on-Year Analysis:** Delve deeper into the factors influencing rating fluctuations for the most popular movies.

This extended analysis can guide production and distribution strategies, supporting decisions based on trends, audience behavior, and content performance.