

# COMPUTER GAMES RATING: ANALYSIS OF THE BIGGEST DIGITAL DISTRIBUTION SERVICE



Artem Khomytskyi: 20221686      Tymofii Kuzmenko: 20221690

## OBJECTIVE

In our analysis of Steam games timeseries, we aimed to explore the factors that influence video game ratings. We utilized a comprehensive dataset containing various game attributes such as genre, price, supported languages, and platform support.

We examined how these attributes affect game ratings. By analyzing these factors, we try to get valuable insights into what drives positive ratings for games on Steam.

## METHODOLOGY

Firstly, we addressed any missing values in the dataset, particularly removing rows where the target variable 'Metacritic score' was missing and entries with a rating of 0 to ensure the quality of our data.

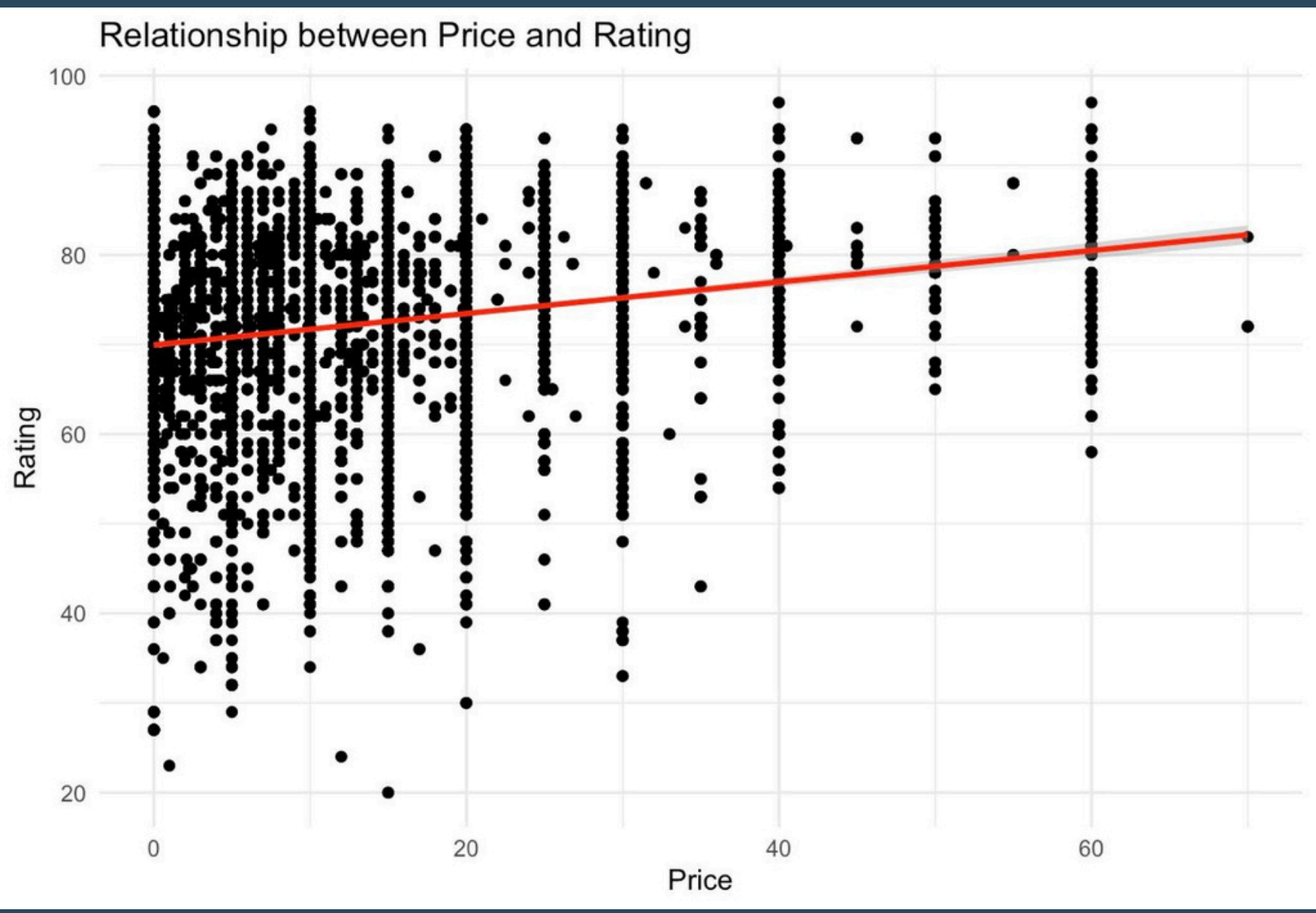
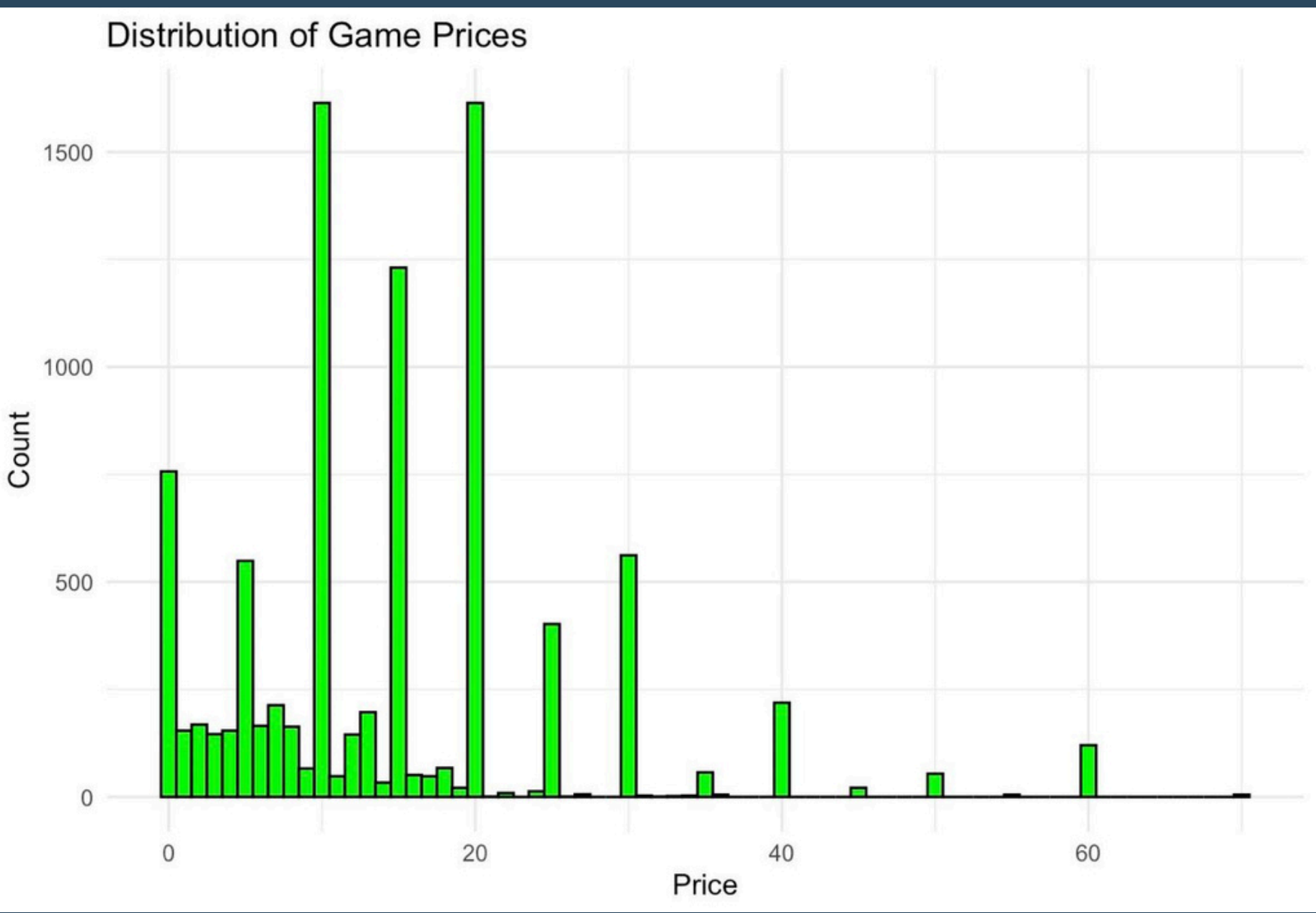
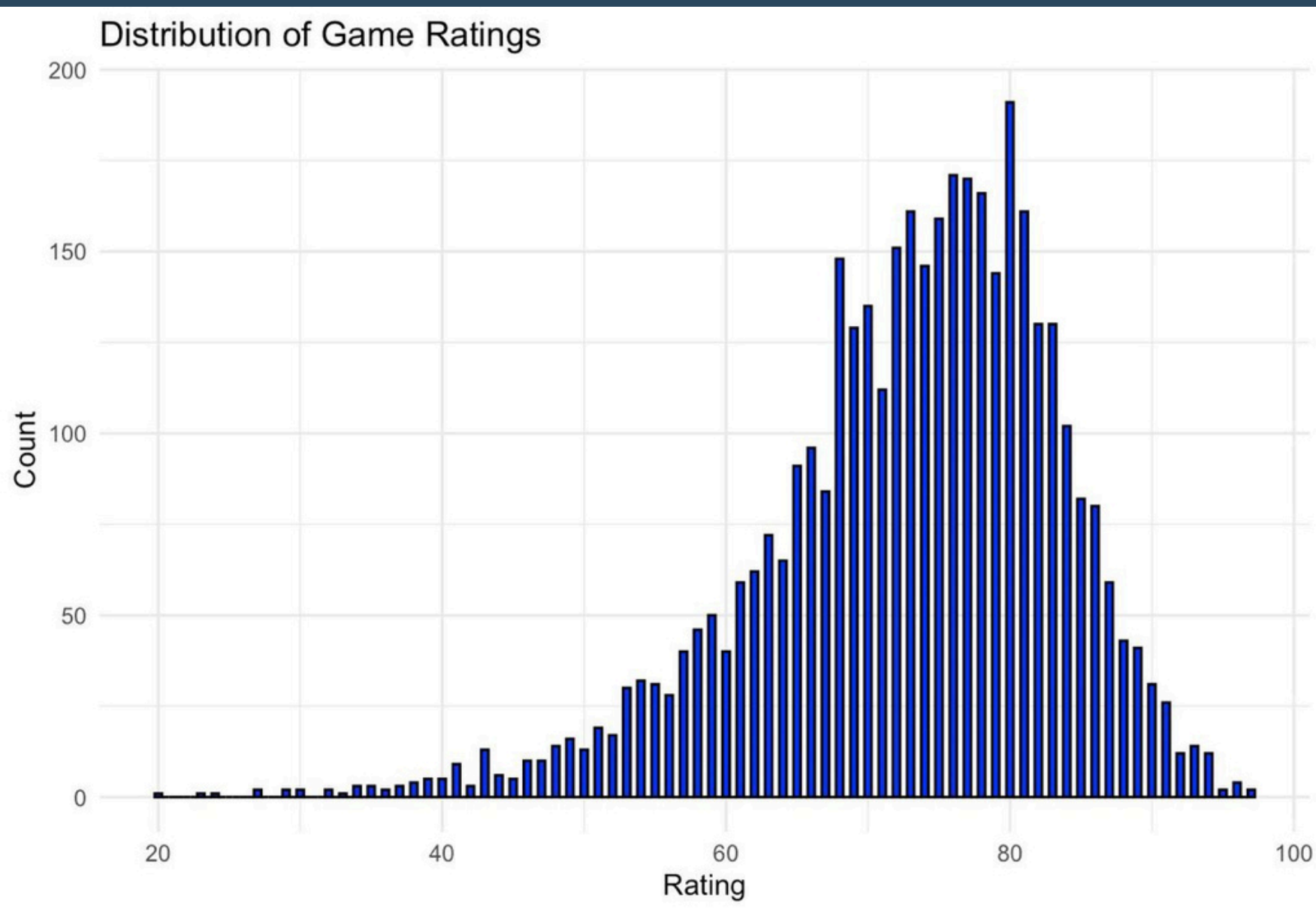
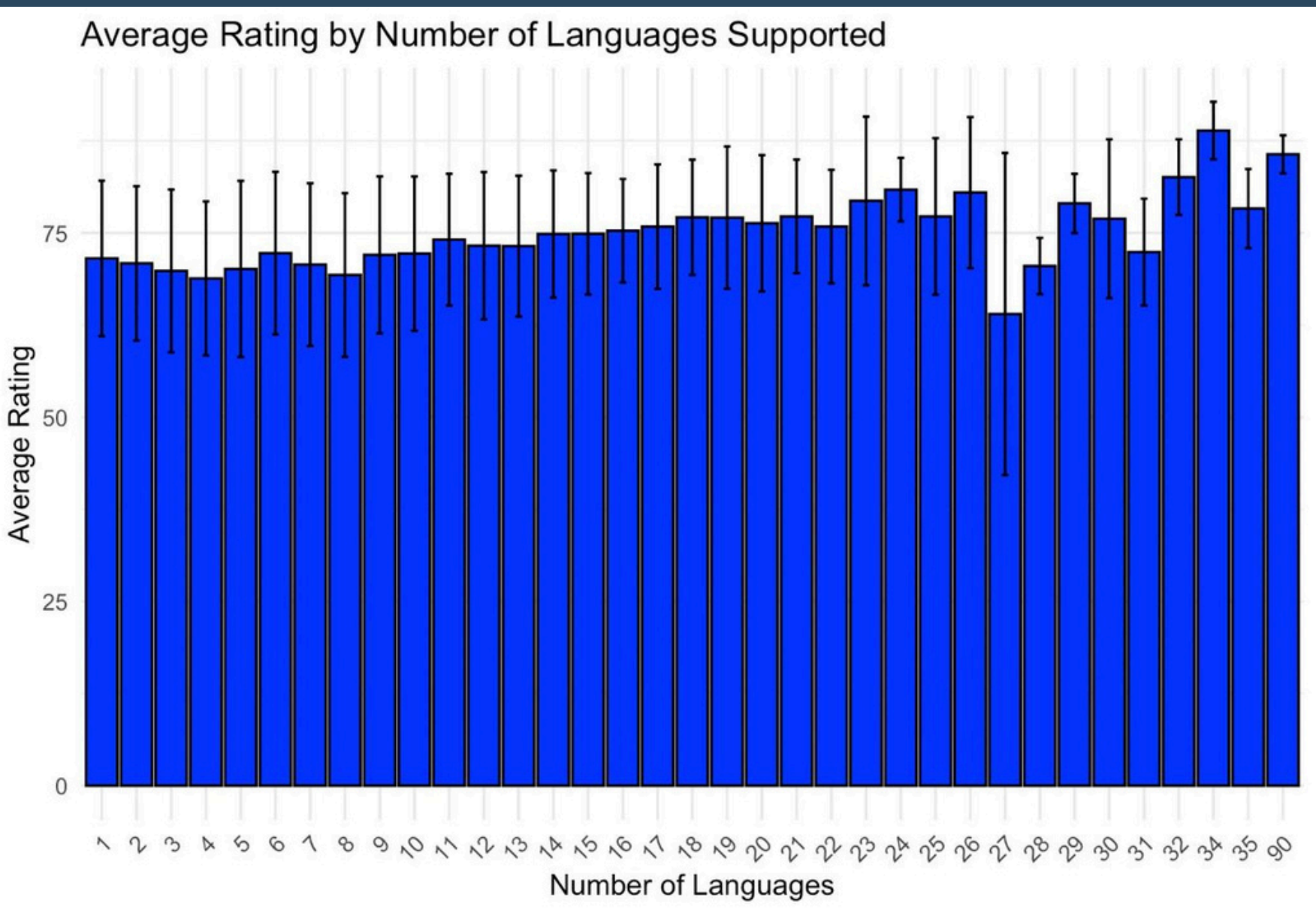
To facilitate analysis, we transformed the 'Genres' column from a string of comma-separated genres to an array of individual genres. This allowed us to create binary columns for each unique genre, indicating the presence or absence of each genre in a given game.

Our exploratory data analysis (EDA) included visualizing the distribution of game ratings, examining the average rating by genre, and exploring the distribution of game prices. We also investigated the relationship between game price and rating using bar chart.

Next, we extracted and preprocessed the languages supported by each game, creating binary columns for each unique language.

We also explored how platform support (Windows, MacOS, and Linux) affected game ratings. Boxplots showed the distribution of ratings based on whether games supported these platforms.

Finally, we specified and fitted a linear regression model to quantify the impact of various factors on game ratings. Our model included predictors such as platform support, number of supported languages, genre, and price.



## CONCLUSION

Using our model, we determined that MacOS support, the number of supported languages, specific genres, and price significantly impact game ratings. MacOS support, number of languages, and price positively influence ratings, while genres like 'Simulation' and 'Violent' negatively affect them. The inclusion of platform support, genre, and price in our model highlights their significant roles in determining game ratings.

Further analysis revealed that certain genres like 'Free to Play' and 'Early Access' are associated with higher ratings, suggesting a trend towards these genres in popular games. Additionally, industry-specific trends and consumer preferences should be considered. For example, the significant positive impact of 'Free to Play' games might reflect a shift in consumer behavior towards cost-effective gaming options. Understanding these dynamics could help developers and marketers better cater to their audience's preferences.

Finally, considering the evolving landscape of gaming platforms and genres, continuous monitoring and analysis will be essential. As the gaming industry grows and diversifies, keeping track of emerging trends and their impact on game ratings will provide valuable insights for stakeholders.

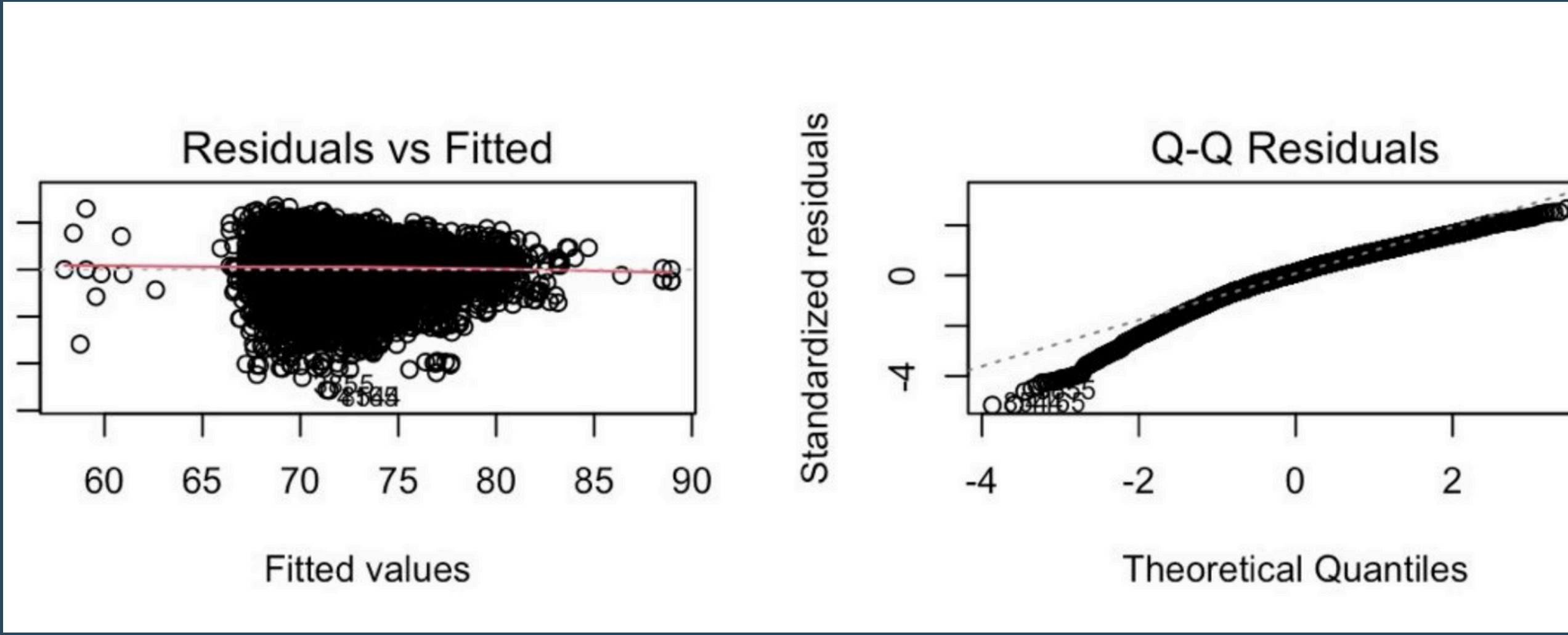
## RESULTS/FINDINGS

The results from our linear regression model suggest varying levels of significance among the predictors for game ratings. MacOS support, number of supported languages, certain genres, and price were found to be significant factors. Specifically, MacOS support had a positive impact on ratings (Estimate = 2.10332,  $p < 2.73e-13$ ), as did the number of supported languages (Estimate = 0.18861,  $p < 2e-16$ ) and price (Estimate = 0.16216,  $p < 2e-16$ ).

Attribute	Estimate	Std. Error	t value	Pr(> t )
MacOS	2.103	0.287	7.318	2.73E-13
Number of Languages	0.189	0.016	11.736	< 2E-16
Early Access	1.745	3.541	0.493	0.622
Free to Play	3.398	1.03	3.3	0.001
Massively Multiplayer	1.139	1.151	0.989	0.322
Sports	0.89	0.908	0.979	0.327
Violent	-10.341	3.54	-2.921	0.003
Price	0.162	0.01	16.179	< 2E-16

Interestingly, several genres had significant effects. 'Free to Play' (Estimate = 3.39760,  $p < 0.00097$ ) and 'Early Access' (Estimate = 1.745,  $p < 0.622$ ) were associated with higher ratings, while 'Simulation' (Estimate = -1.37299,  $p < 0.00369$ ) and 'Violent' (Estimate = -10.34135,  $p < 0.00349$ ) genres had negative impacts.

Our model accounted for potential serial correlation and heteroskedasticity by using robust standard errors, enhancing the reliability of our estimates. The model's overall fit was modest (R-squared = 0.0695), indicating that while these factors are significant, other variables not included in the model also influence game ratings.



Data:

<https://www.kaggle.com/datasets/fronkongames/steam-games-dataset>