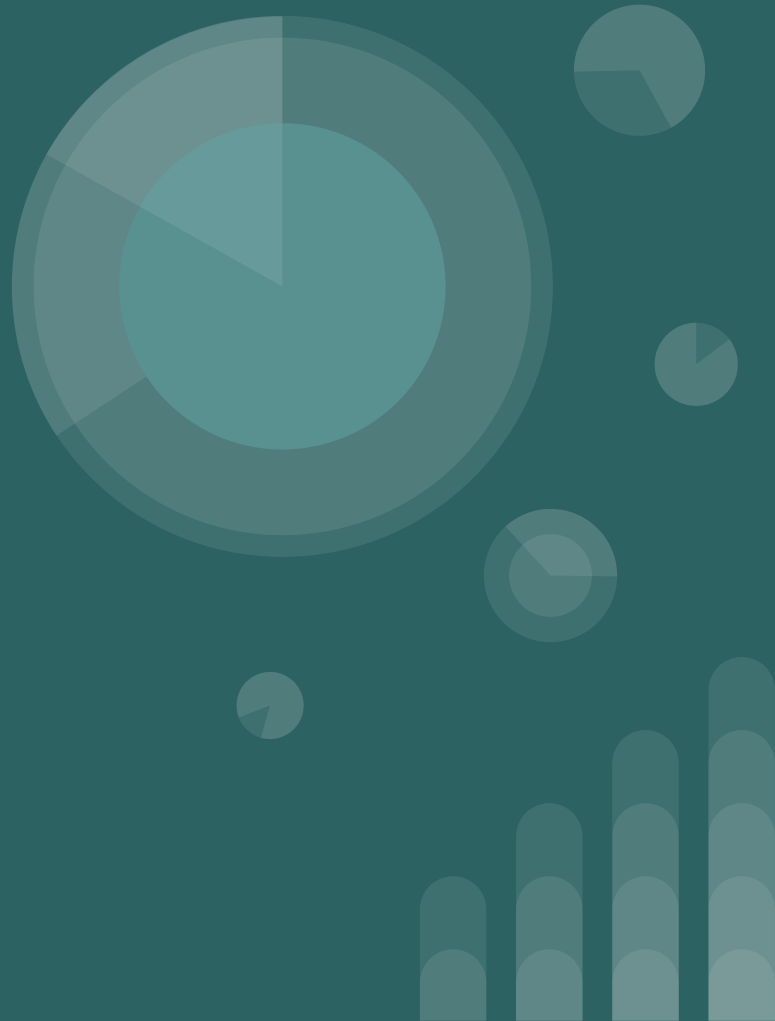


Steam Dataset

By: Spencer Whan, Edward Madaire, Nathan Mitchell,
Tim Callahan, & Harryson Belizaire





What Is The Steam Dataset?

- Dataset from Kaggle - relating to “Steam” developed by the Valve corporation.
 - Popular digital video game distribution platform.
- Includes public data that’s displayed via steam pages related to users, games, and recommendations.
 - Primarily review / recommendation based data.
 - Due to privacy policies, specific information on a users game library, game sales, etc..., are unavailable.

Steam Dataset Contents

- games.csv
- recommendations.csv
- users.csv
- Compiled into 'Steam_DB.xlsx' for simplicity



The dataset was stated as already cleaned. After further analysis, this was found to be true.



Important Terminology

'rating' -

- Overwhelming Positive
- Very Positive
- Mostly Positive
- Positive
- Mixed
- Negative
- Mostly Negative
- Very Negative
- Overwhelmingly Negative

'price_final' - The current listed price of the game.

'price_original' - Original price on game release date.

'discount' - Discounted price due to the game being on sale.



Descriptive Statistics





Descriptive Statistics: games.csv

- 98.43% of games have Windows operating system accessibility.
- 25.6% of games have Macintosh operating system accessibility.
- 17.78% of games have Linux operating system accessibility.
- 100% of games are steam deck compatible.

positive_ratio:

- Mean: 77
- Median: 81
- Mode: 100

user_reviews:

- Mean: 1826
- Median: 48
- Mode: 10

price_final:

- Mean: 8.6
- Median: 4.99
- Mode: 0



Descriptive Statistics: recommendations.csv

helpful:

- Mean: 3

funny:

- Mean: 1

hours:

- Mean: 200.16
- Median: 99.2
- Mode: 0.3



Descriptive Statistics:

USERS.CSV

products:

- Mean: 122
- Median: 58
- Mode: 11

reviews:

- Mean: 3
- Median: 1
- Mode: 1



Power BI Dashboards



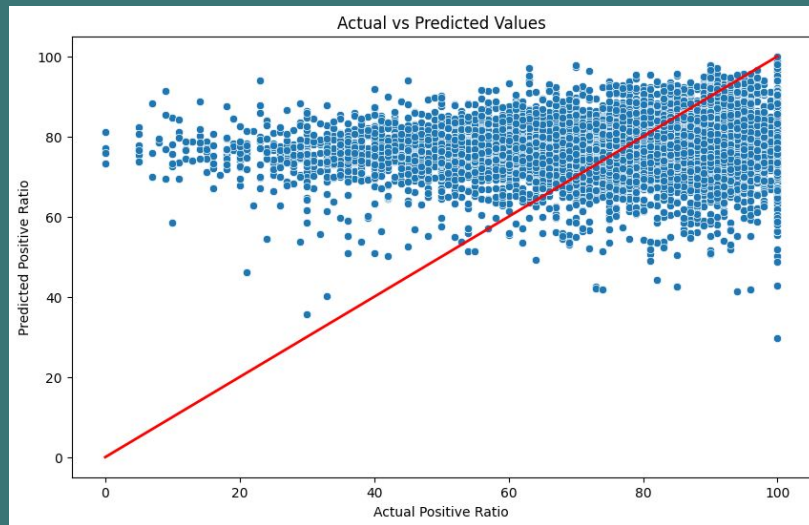


Predictive Statistics



Predictive Statistics

- Attempting to predict ratio of positive reviews from number of reviews, price and discount
- Random forest regression model
- Third Party Data = High Mean Average Error





Conclusion and Recommendations

- Top categories of all the reviews are mostly positive.
 - Users are more likely to give positive reviews than negative reviews.
 - Concluded that people are more willing to review something they have enjoyed compared to something they have not.
- Gaming market is mostly dominated by the Windows operating system.
 - 98.43% of games listed on steam have Windows operating system accessibility.
 - More games should become compatible with other operating systems (Mac, Linux).
- Difficult to create accurate conclusions using predictive statistics due to the data.
 - Data is entirely public information - without the private backend information.
 - Backend information would improve accuracy.



Thanks For Watching

Questions?

