

PREDICTING RATING OF VIDEO GAMES





OUTLINE:

Motivation.

Data collection & cleaning.

EDA.

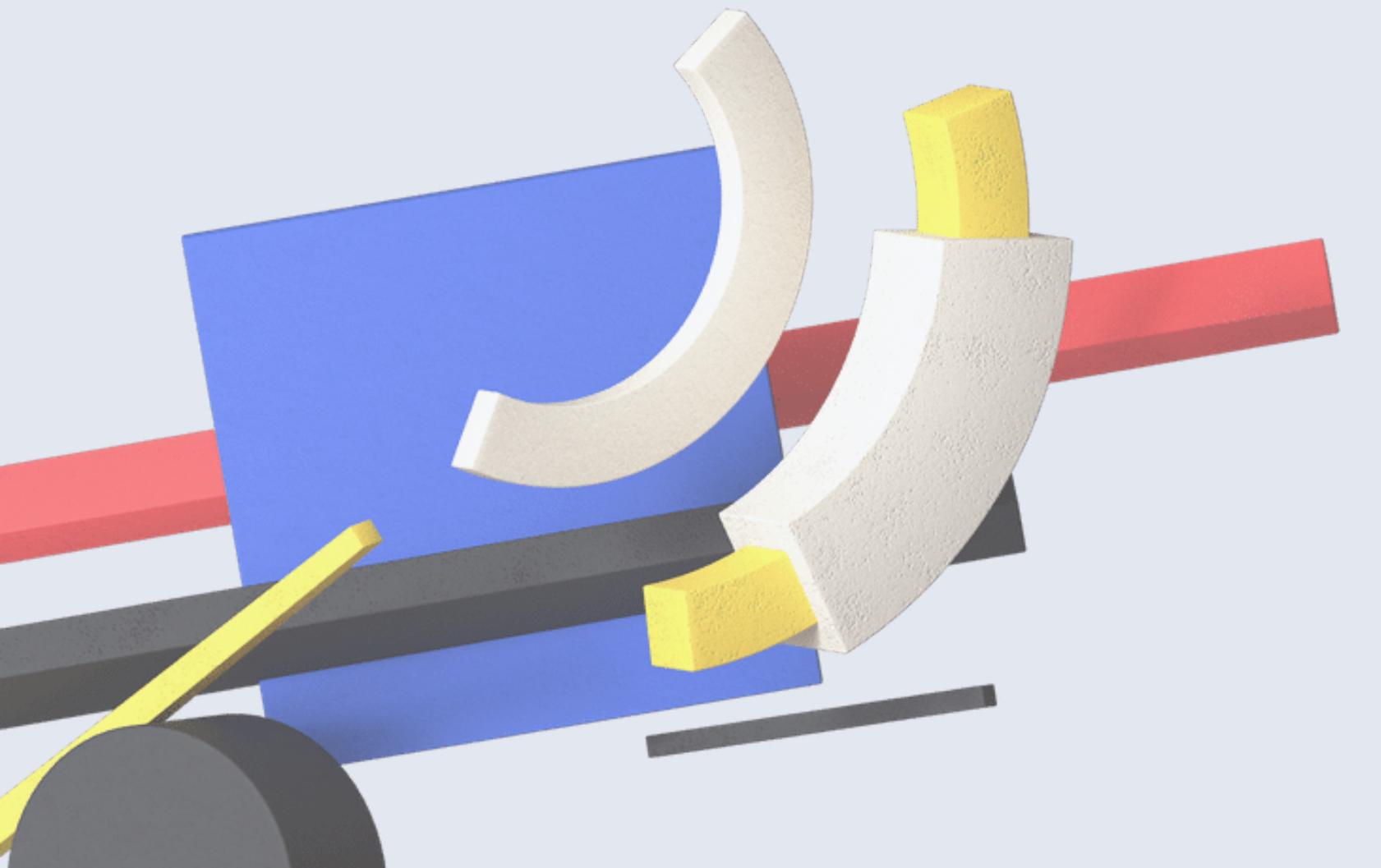
METHODOLOGY

Result.

CONCLUSION

MOTIVATION

The scope of this project is to **classify** each video game in the dataset by **ESRB rating**, to do this we used Logistic Regression, Random Forest, k-NN, Naive bayes, and Decision Tree.



Data Collecting

The data was collected from the Kaggle webpage

Taggart : Rating (T,M,E) ,It has 15 features:

Name - the name of the game

Platform - platform

Year_of_Release - year of release

Genre - game genre

Publisher - studio

NA_sales - North American sales (millions of copies sold)

EU_sales - sales in Europe (millions of copies sold)

JP_sales - sales in Japan (millions of copies sold)

Other_sales - sales in other countries (millions of copies sold)

Global_sales - sales in other countries (millions of copies sold)

Critic_Score - Critic score (maximum 100)

Critic_count - The number of critics used in coming up with the Criticscore

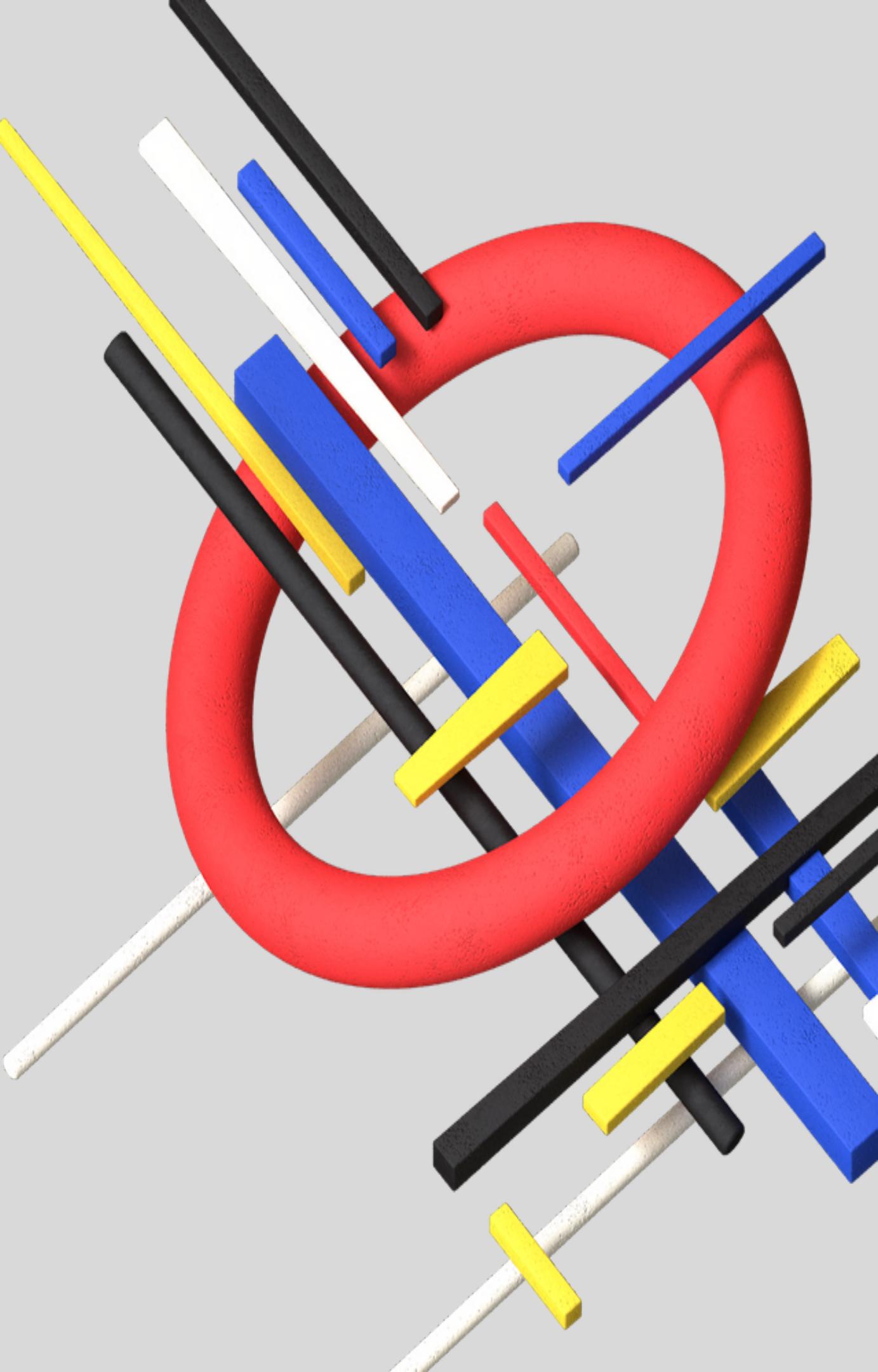
User_Score - user score (maximum 10)

Rating - rating from the organization ESRB (English Entertainment Software Rating Board).

Developer - developer

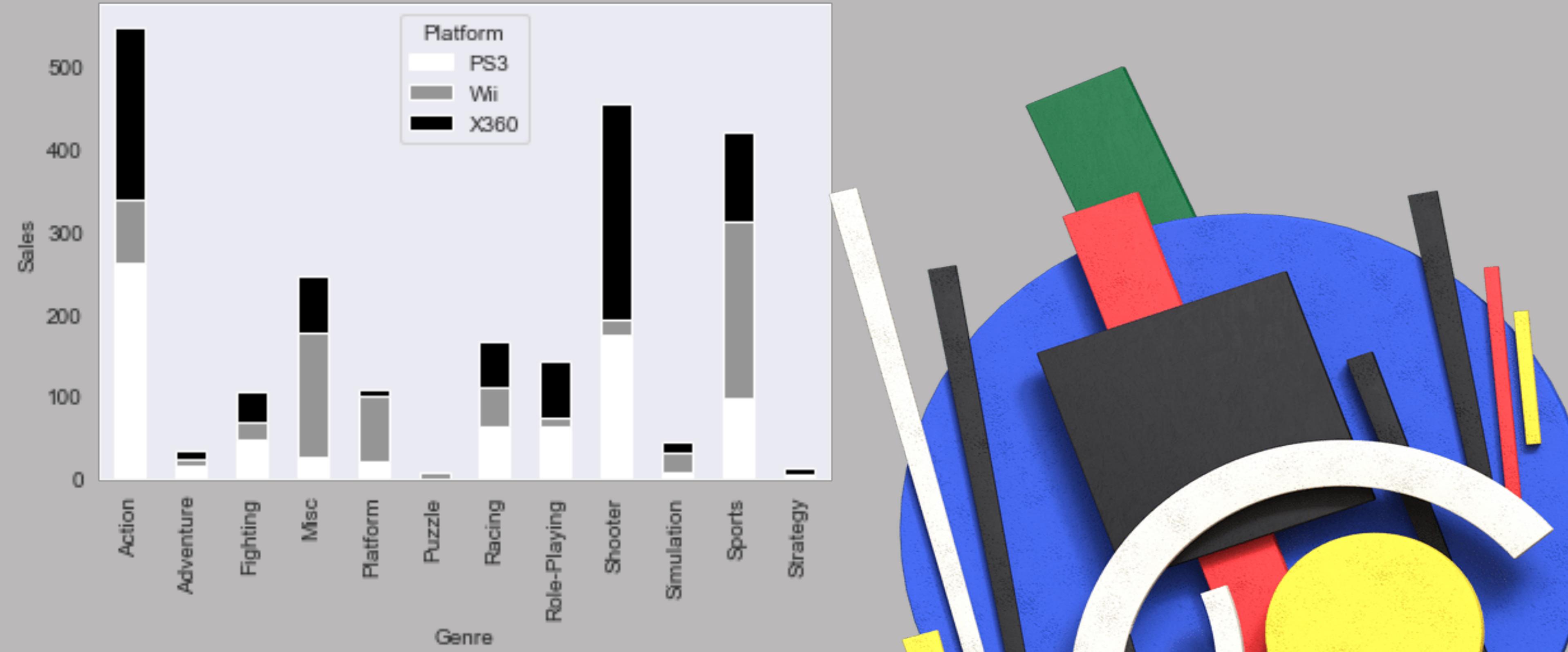
Data Cleaning

- Renaming columns for ease of use
- Dropping games without a year of release or genre
- Convert(User_Score, Year_of_Release, User_Count, Critic_Count) to int
- Eliminating all rows with NaN value

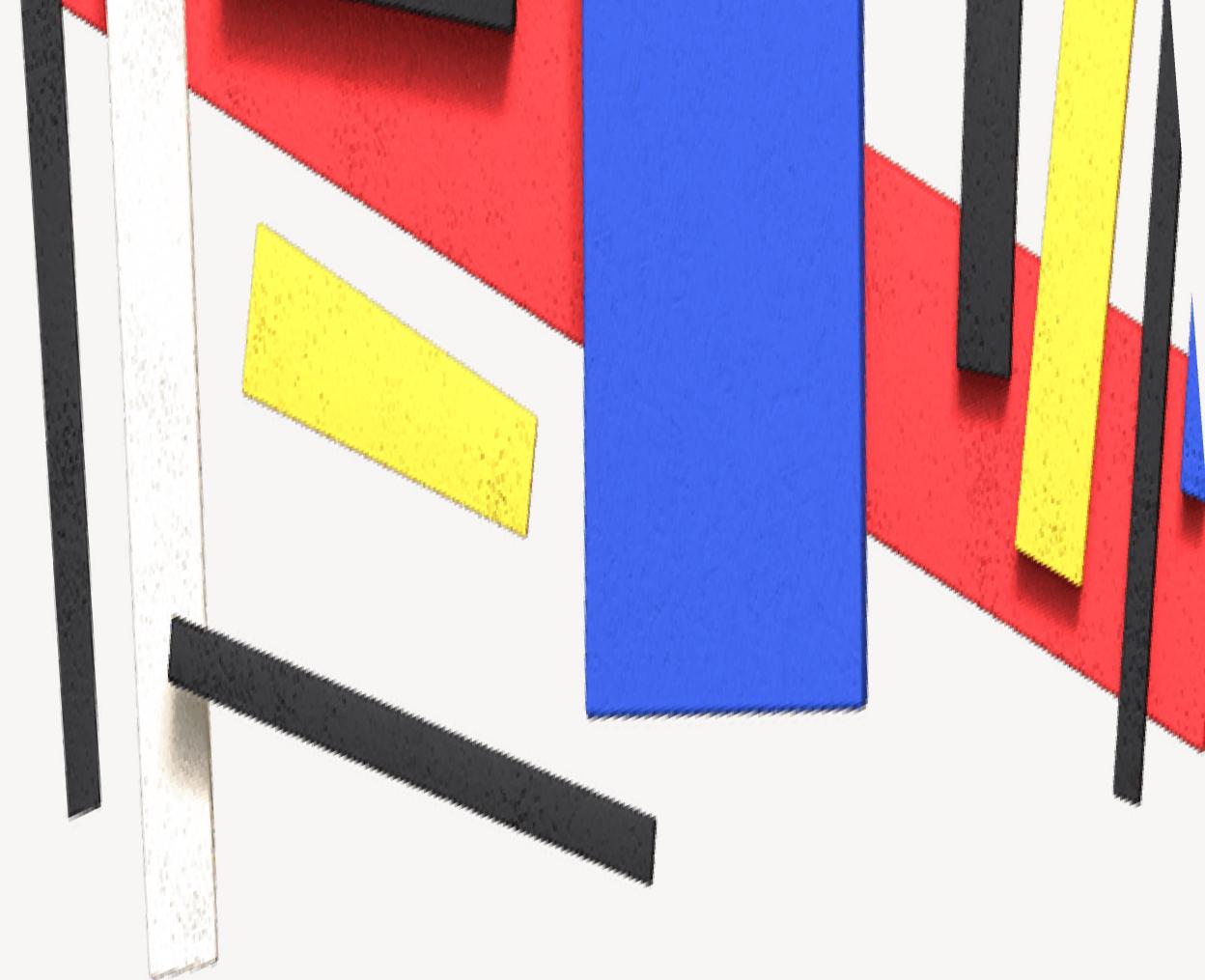
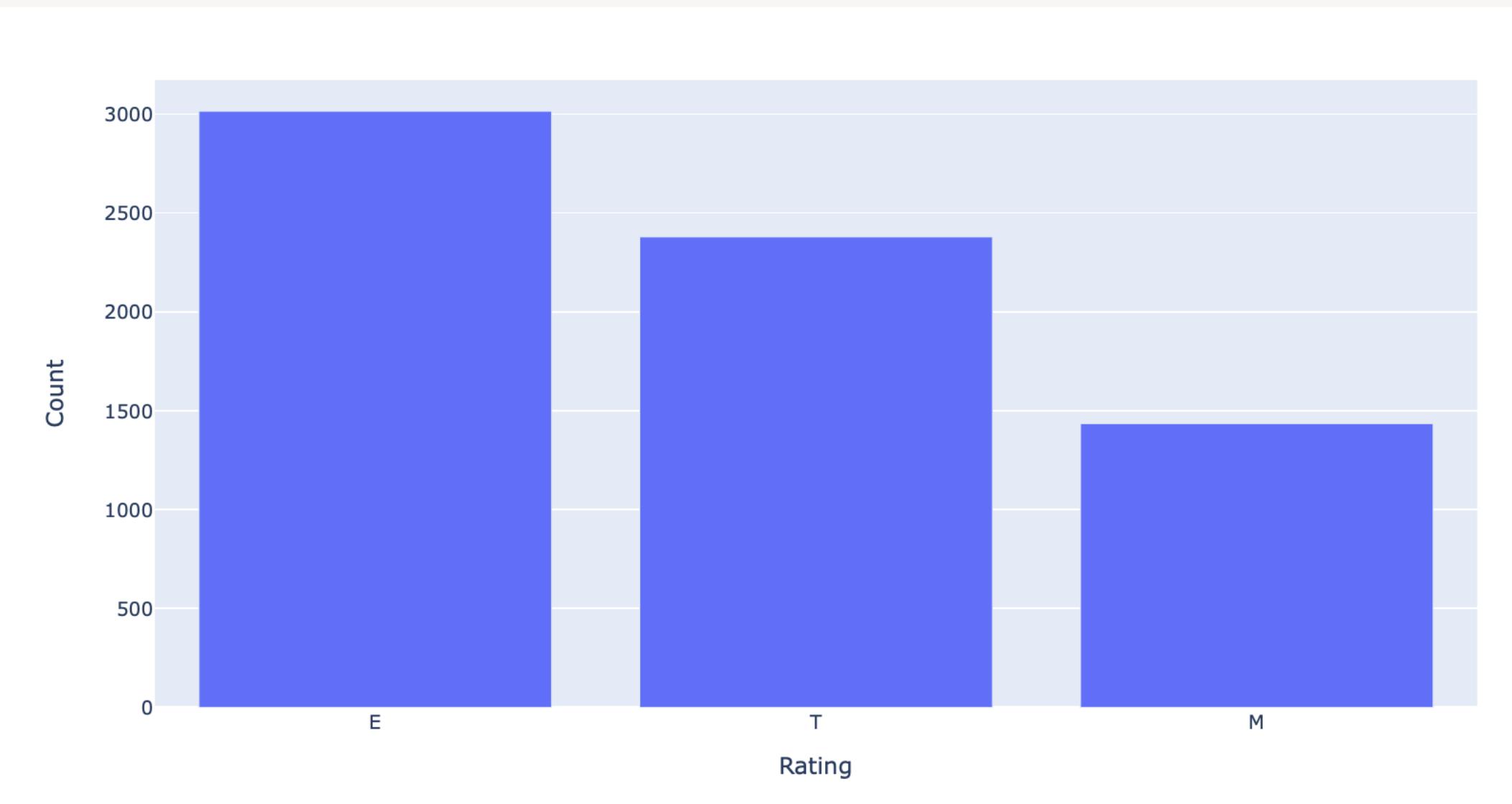


EDA

Stacked Barplot of Sales per Game Genre



EDA



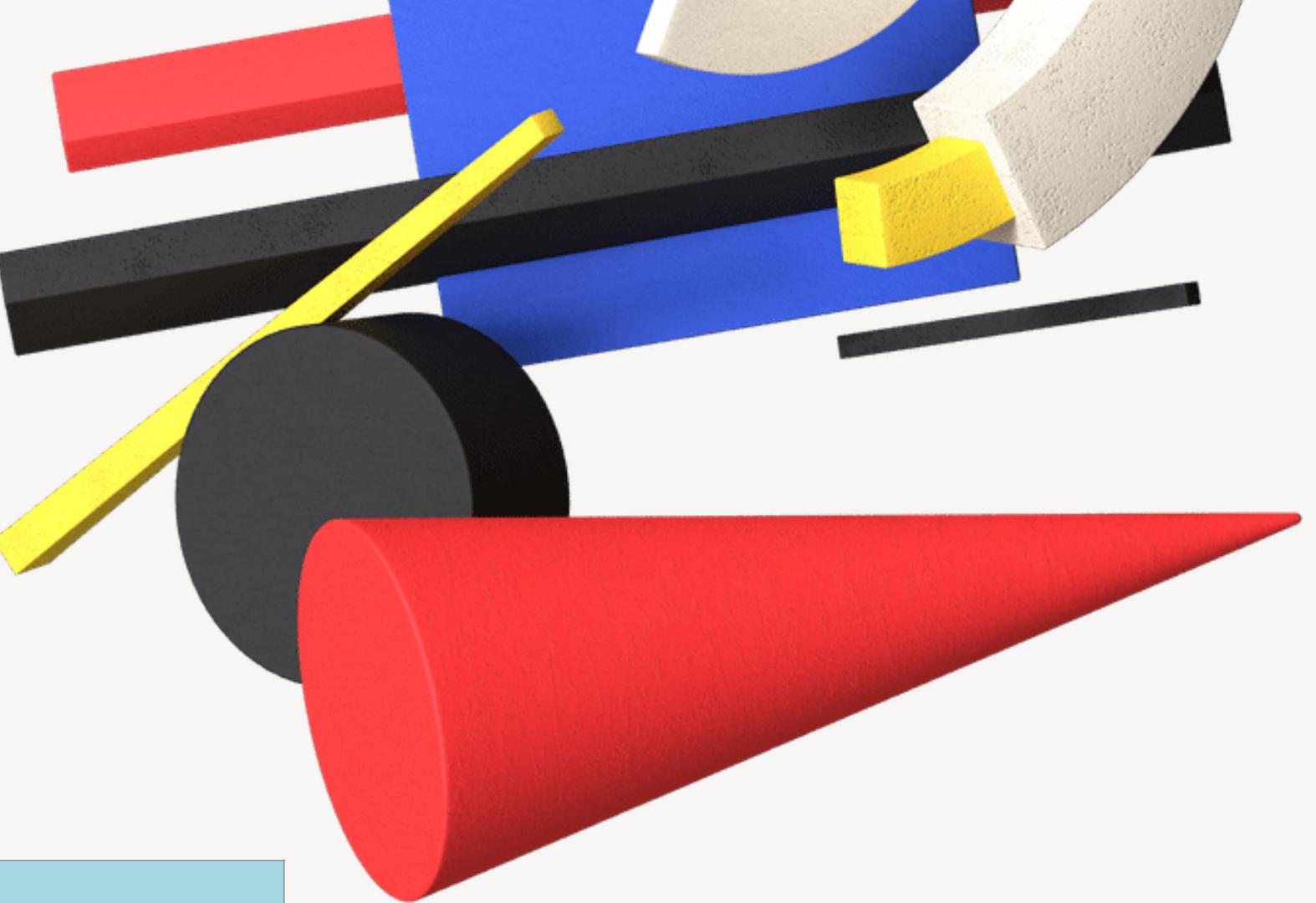


METHODOLOGY

Five different models were applied:

- The Linear Regression Model.
- The Decision Tree Model.
- Random Forests Model.
- K-Nearest Neighbors Model
- Naive Bayes Model
- Voting Classifier

RESULT



Logistic Regression

Accuracy	F1
0.53	0.36

Random Forest

Accuracy	F1
0.57	0.43

Decision Tree

Accuracy	F1
0.55	0.43

Naive Bayes

Accuracy	F1
0.51	0.33

K-Neigbor

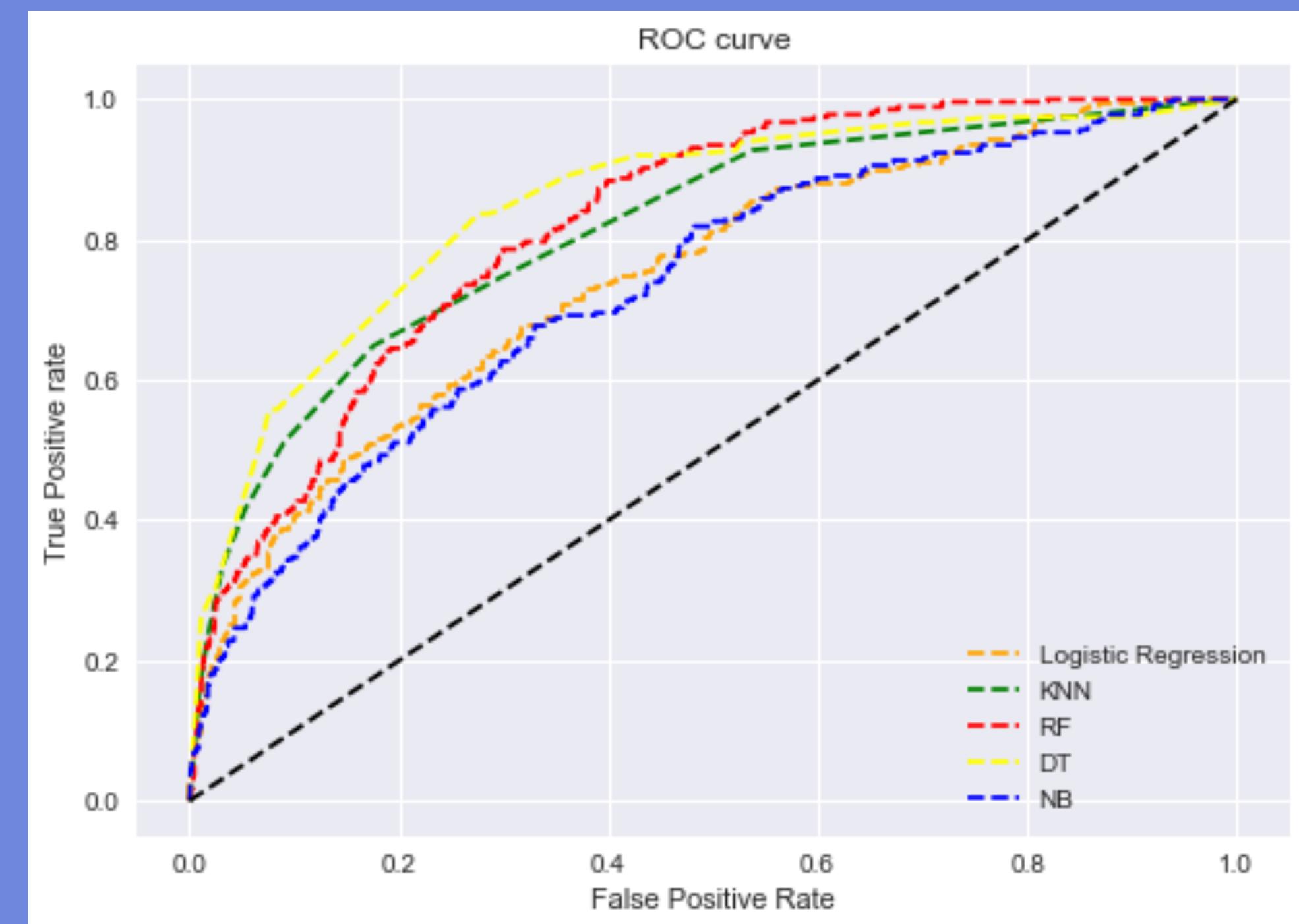
Accuracy	F1
0.60	0.52

Voting Classifier

Accuracy	F1
0.672	0.46



ROC



CONCLUSION

- ▶ The results show that **Random Forest & knn** provide the best results with **F1 score of 0.5,0.6**
- ▶ **For further research:**
 - a bigger dataset could be beneficial for **unsupervised learning methods**.
 - New features like review text, number of uninstalls, and number of updates could provide new insight that can improve the modeling results



THANK YOU

Presented by:

Rana Alqahtani

Ohoud Albabtain

Khulud Amer

