



Predict Global Sales of Video Games





Introduction



IMDb is an online database of information related to video games, movies, television series, Etc...

IMDb

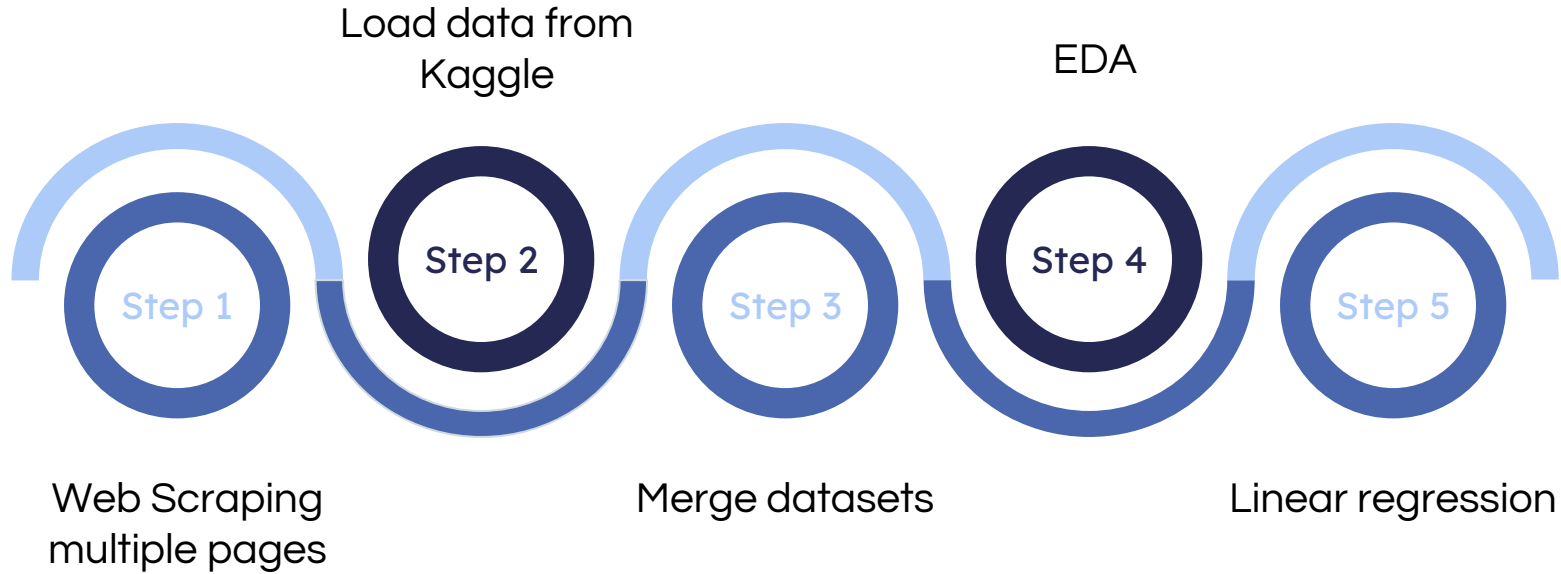


Objective

The goal of this project is to predict global sales of the video games using linear regression model based on features of the data from [IMDb](#) website and [Kaggle](#).



WorkFlow



Data Discreption

We scrap 10 pages from [IMDb](#) then marge it with another dataset from [Kaggle](#) after that we got 658 rows and 7 columns .

Data :

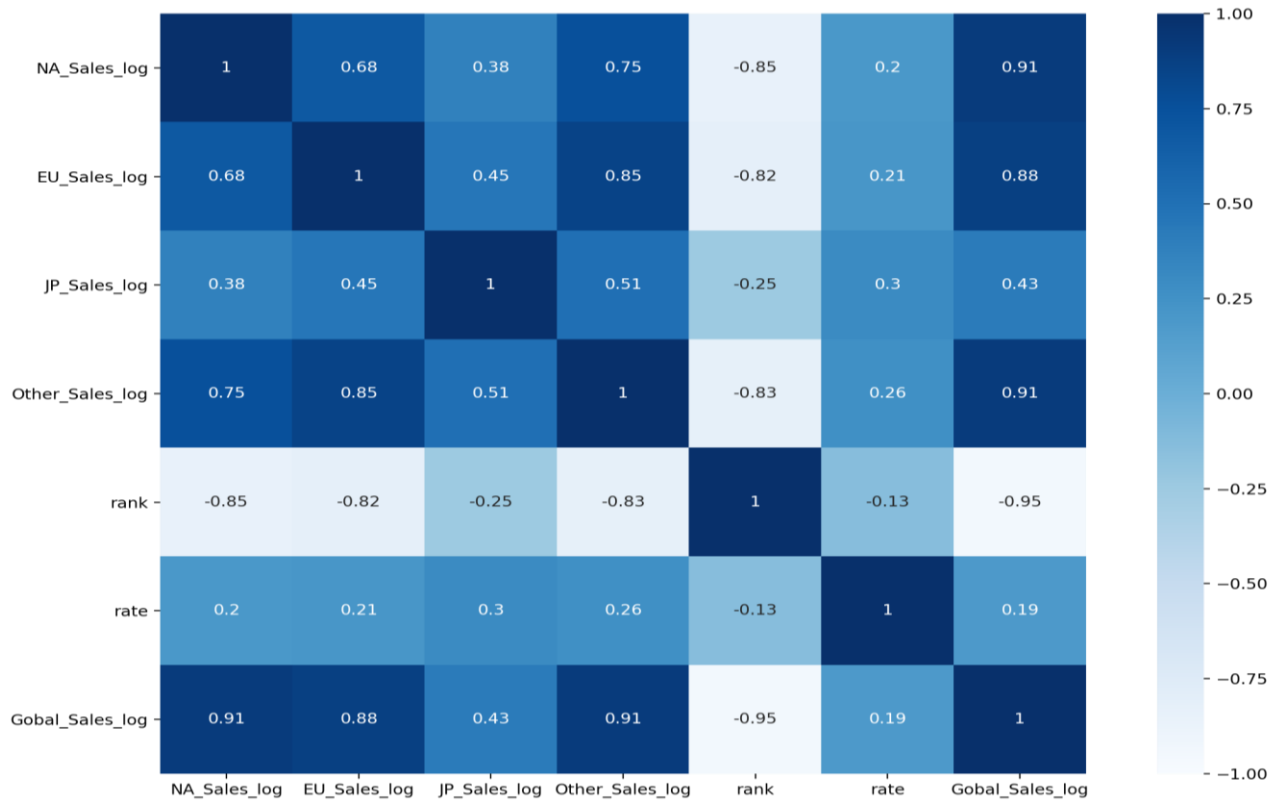
➤ Features:

- rate - Number of rate.
- Rank - Ranking of overall sales.
- NA_Sales - Sales in North America (in millions).
- EU_Sales - Sales in Europe (in millions).
- JP_Sales - Sales in Japan (in millions).
- Other_Sales – Sales in the rest of the world (in millions).

➤ Target:

- Global_Sales - Total worldwide sales.

EDA





Linear Regression

Features Engineering

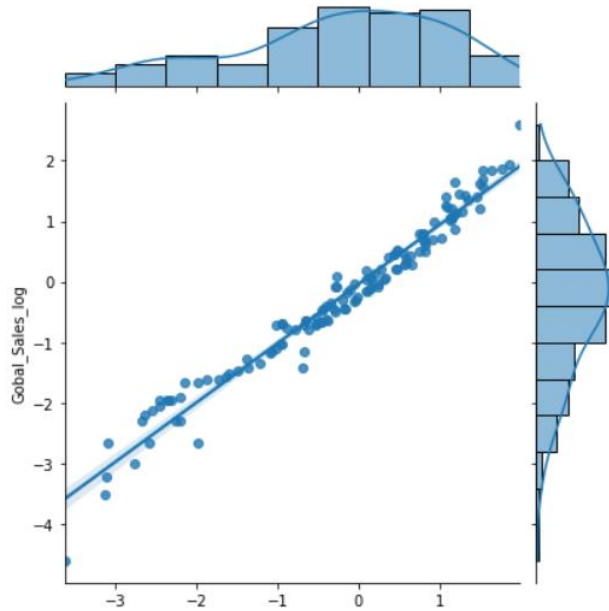
Using data transformation to make data
more normalized.



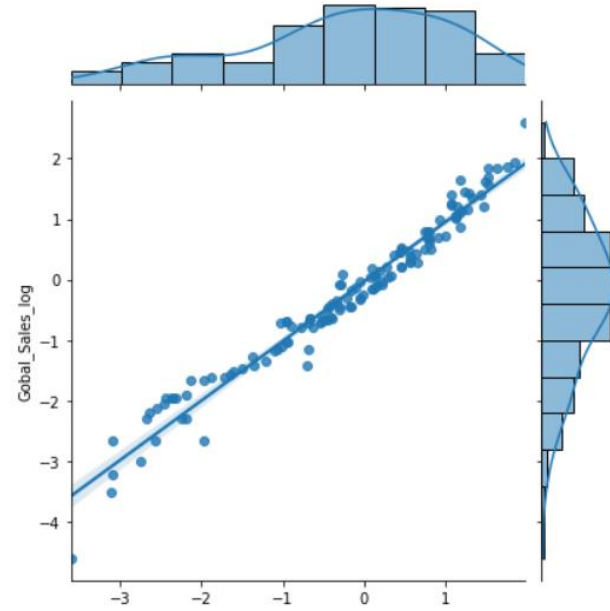
experiments

	Baseline model	Standard Scale model	MinMax Scale model	Polynomial model
Train	0.967	0.967	0.967	0.997
validation	0.973	0.973	0.973	0.997

Models

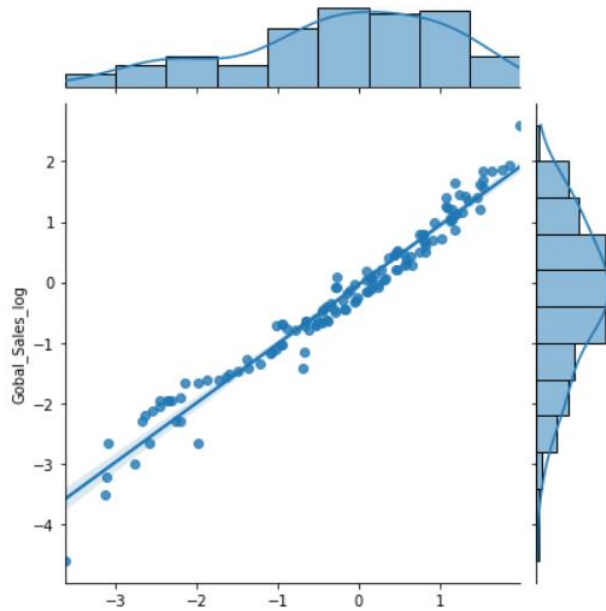


Baseline model

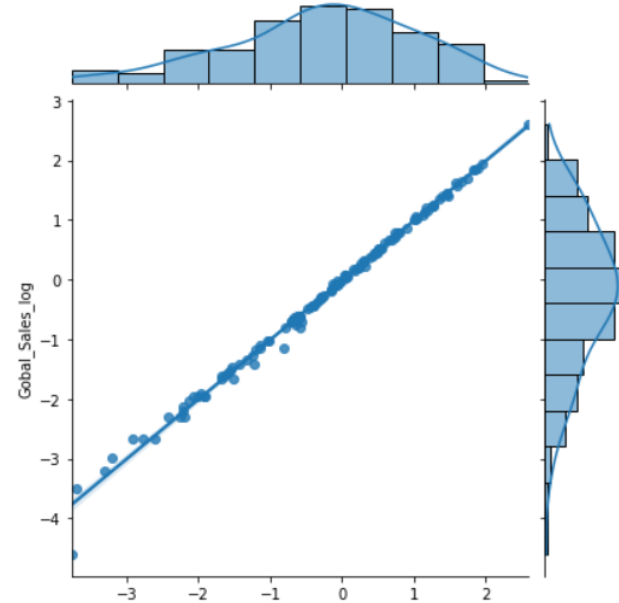


Standard Scale model

Models



MinMax Scale model



Polynomial model

Test Model (Unseen data)

After we do experiment, we noticed the best experiment from r^2 results is polynomial model we got $r^2 = 0.99$ which mean the model well trained on the data .



Lasso & Ridge



Lasso	Ridge
0.9614	0.9613

Conclusion

At the end from R-Squared Test results, we can assume that the best model is polynomial model because they have the highest r^2 in both test and train



Thank you