

NETFLIX

Movies And Series Awards Classification

By: Tarfah Alabbad & Rawabi Alharbi





contents



INTRODUCTION



METHODOLOGY



TOOLS



EDA



PREPROCESSING



MODELLING



CONCLUSION

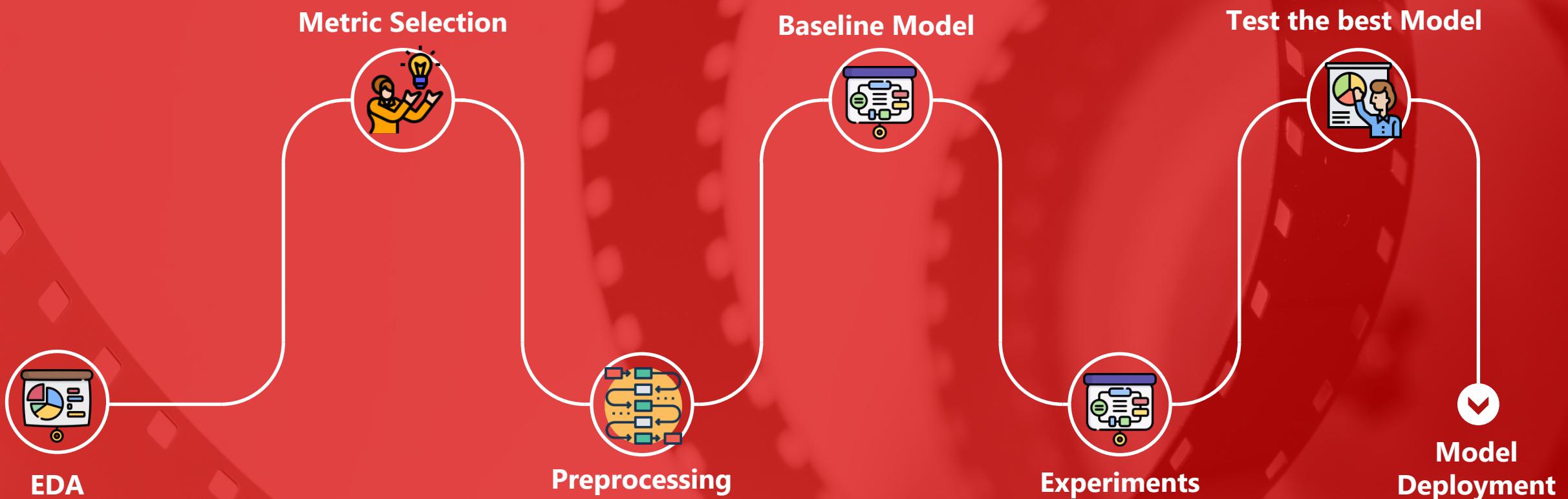
Introduction

- Who doesn't know Netflix?
- Problem Statement
- Objective
 - What factors that plays a role in awarding the movie/series?

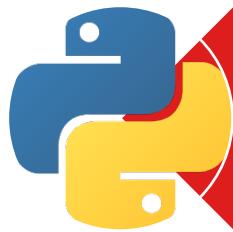
Introduction

- **Dataset**
 - Downloaded from Kaggle
 - +15k observations => +8k
 - 29 features => 10

Methodology



TOOLS



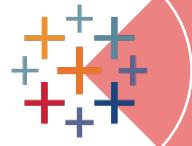
Python, Jupyter notebook



VS Code, HTML, CSS, Flask



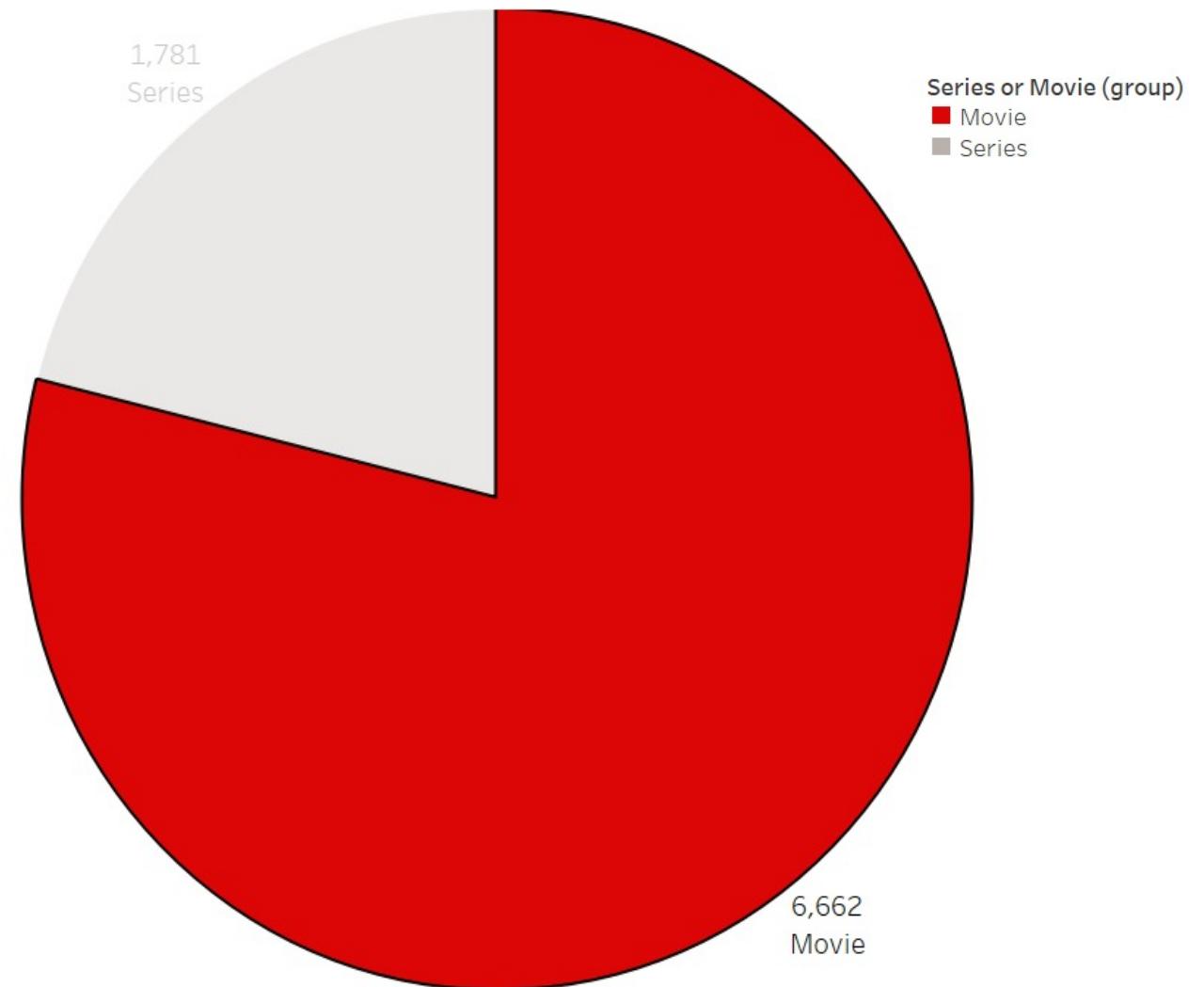
NumPy, Pandas, SKLearn



Matplotlib, Seaborn,
Tableau



the number of movies and series in the data set



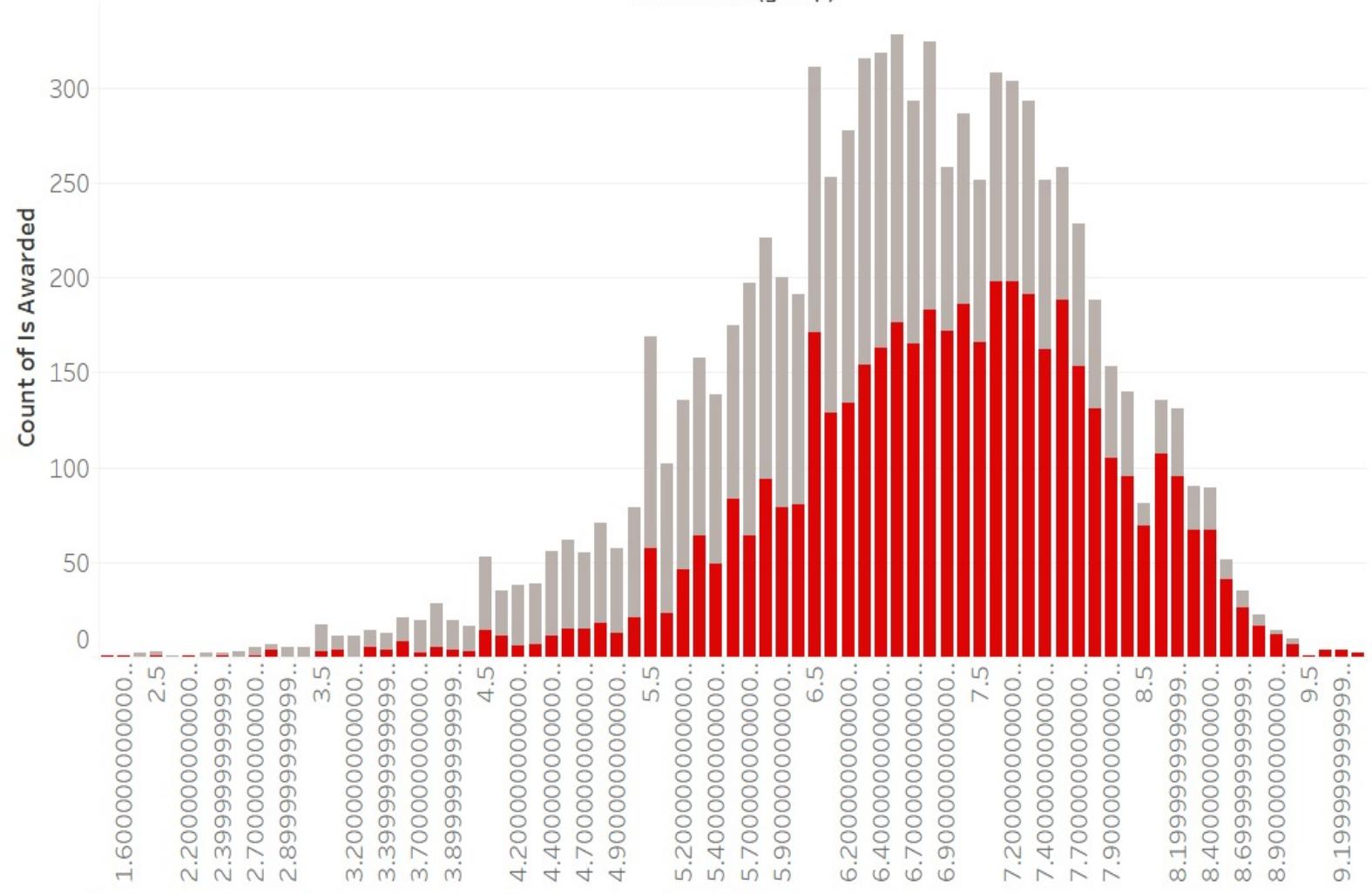
EDA

NETFLIX

The IMDb score of the target

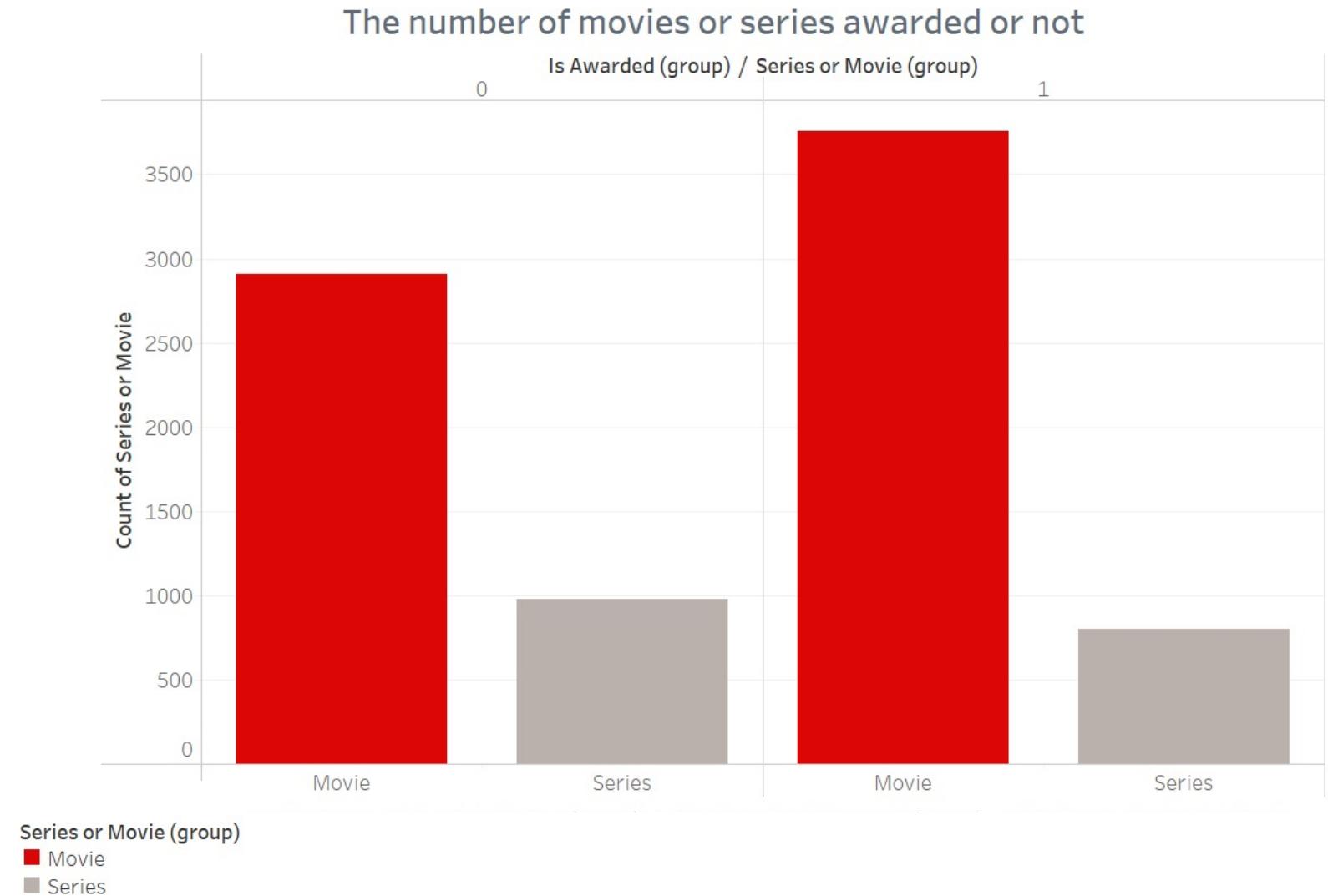
IMDb Score (group)

Is Awarded (group)
0
1



EDA

NETFLIX



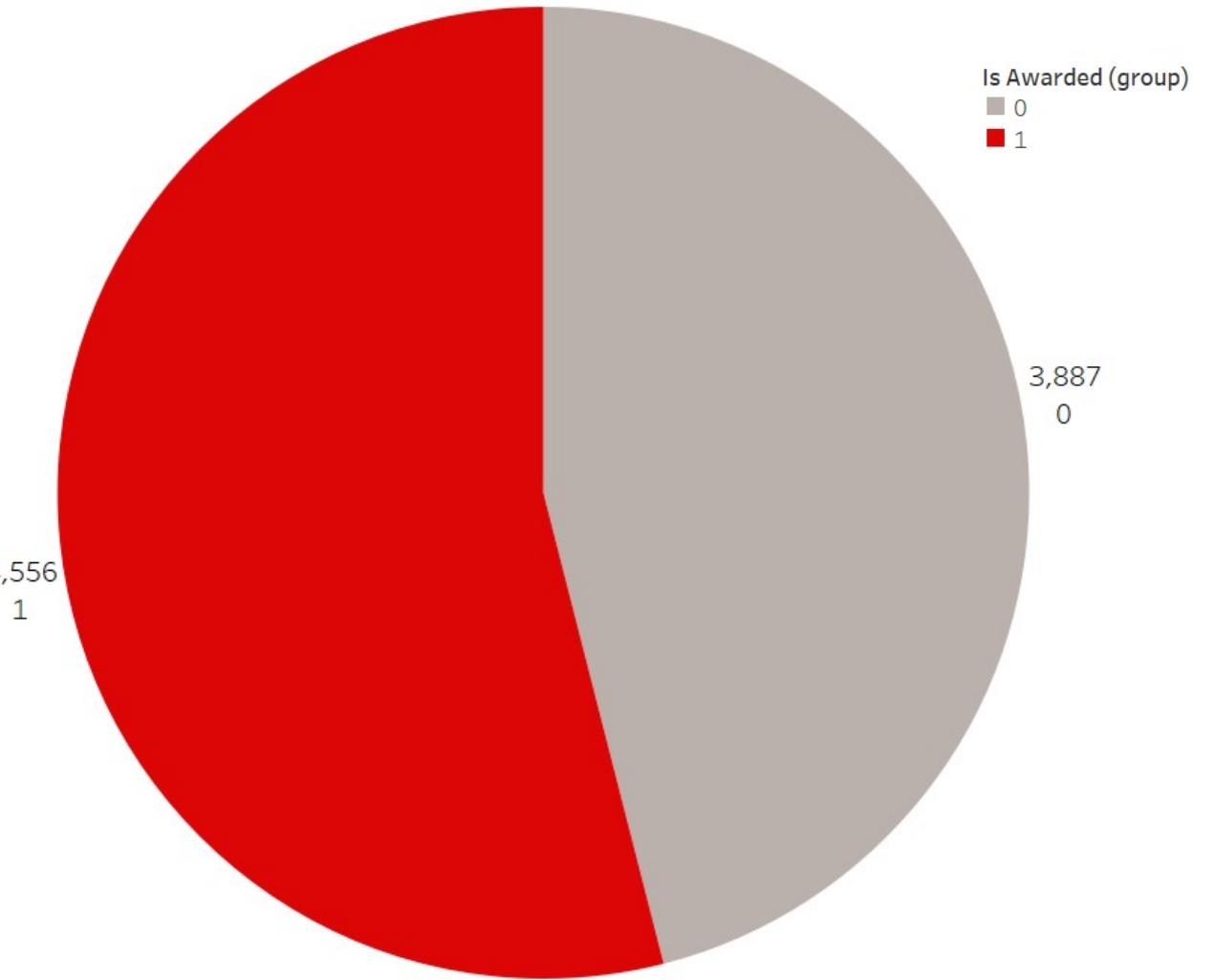


EDA



NETFLIX

The distribution of target data

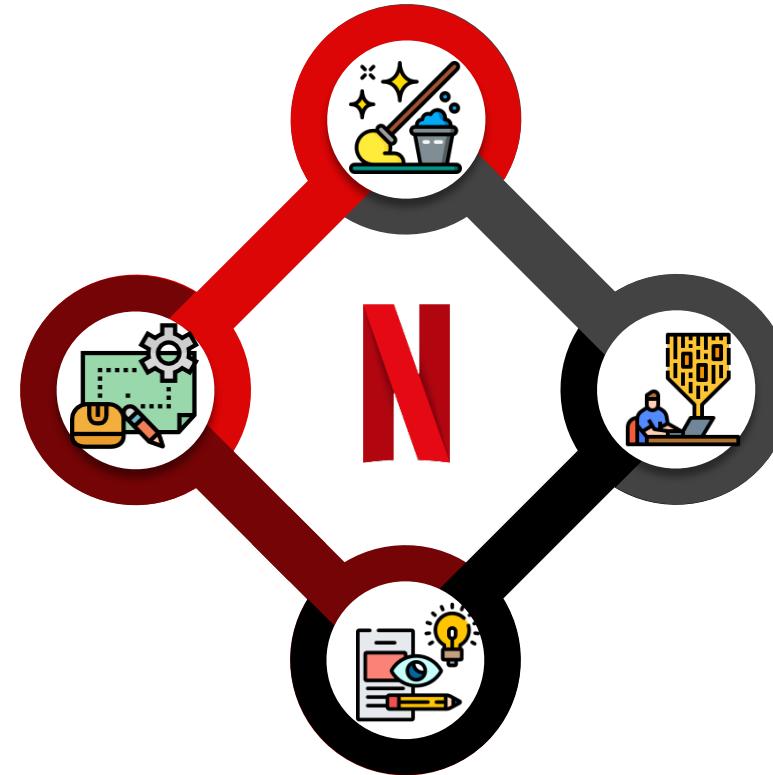


PRE-PROCESSING



MISSING VALUES

- Dropping rows
- Filling with mean



Columns Type

- IMDb Votes
- IMDb Score

Categorical Values

- LabelEncoder
- get_dummies

- **Creating New Columns**
- **Dropping Unneeded Column**

- + Is Awarded
 - + Is Nominated
- Summary, Poster, Trailer Site ,Netflix Link, IMDb Link...

Modelling

- **Training set:** 50%
- **Validation set:** 20%
- **Testing set:** 30%
- **Classification evaluation metric:** F1 score

Modelling

Tunned Models	F1 Score
K-nearest Neighbors	78%
Logistic Regression	79%
Decision Tree	79%
Random Forest 	81%
XGBoost Classifier	80%
Stacking	80%

Modelling

Tunned Models	F1 Score
K-nearest Neighbors	78%
Logistic Regression	79%
Decision Tree	79%
Random Forest	81%
XGBoost Classifier	80%
Stacking	79%

Result



Best Model Results (Random Forest)

Dataset	Accuracy	F1 Score
Training	85%	86%
Testing	82%	84%

Conclusion



The model works with good performance to classify the Movie/Series as awarded or not.



Future Work

Explore more features, such as director, main actor/actress, main genre.



MODEL DEPLOYMENT





Thank
You

