

Start

- Data Preprocessing
- Load CSV File
  - One-hot Encoding for Genre
  - Separate Target Variable (Votes)

- EDA Insights
- Genre affects votes strongly
  - Votes distribution is skewed
  - Runtime has weak positive correlation

- Objectives & Data
- Q1: Predict popularity → Use Regression Models
  - Q2: Does Genre matter? → Yes, confirmed in EDA

- Techniques Applied
- Visual Analysis: Boxplots, Histograms, Heatmaps
  - Models Used: Linear Regression, Random Forest, XGBoost

- Model Evaluation
- Metrics: MAE, RMSE,  $R^2$
  - Best Performing Model: XGBoost

End