Introduction

- Academy Awards - world's most prestigious film festival
- Type of Problem: Binary Classification
  - Can be scaled to multi-class model
- Origin of Data
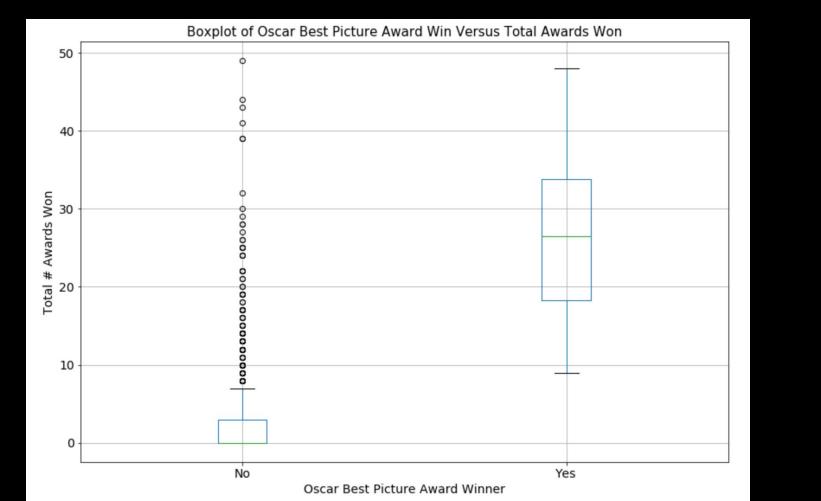 - shorturl.at/juzIN

Introduction

- Contains Numerical Info and 20 other Film Awards Info
  - E.g. Golden Globes, BAFTA, Directors Guild, etc.
- Before Preprocessing Shape: (1235, 69)
  - Containing films from 2000-2018
  - Total Oscar Best Picture Winners: 18
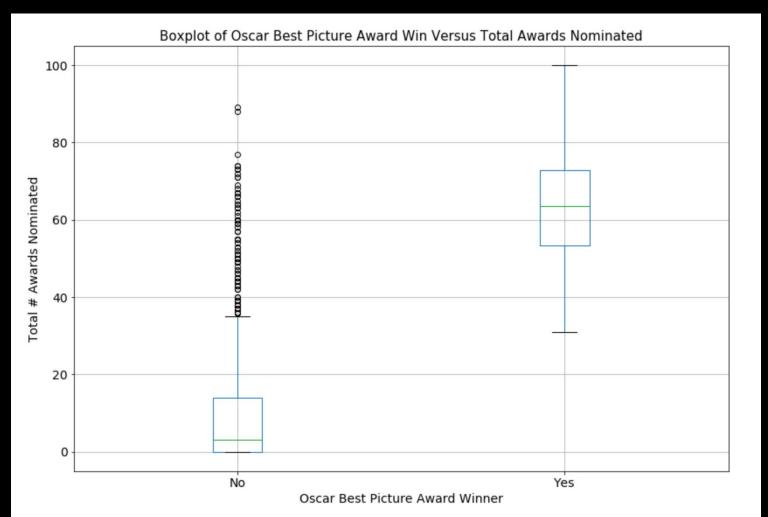- After Preprocessing Shape: (1235, 1017)

Preprocessing

- Missing Data:
  - Replace missing with "NaN"
- Numerical Features
  - E.g. Gross, # of User Reviews, IMDB rating
  - Standard Scaling
- Oscar Categorical Features (Yes/No)
  - Convert to binary 1/0
- Genre, Nomination and Won Categories
  - Data in the form of
    - Action|Adventure|Sci-Fi
    - Best Song|Best Composer|Best Director|Best Picture
  - Split data, one-hot encoded
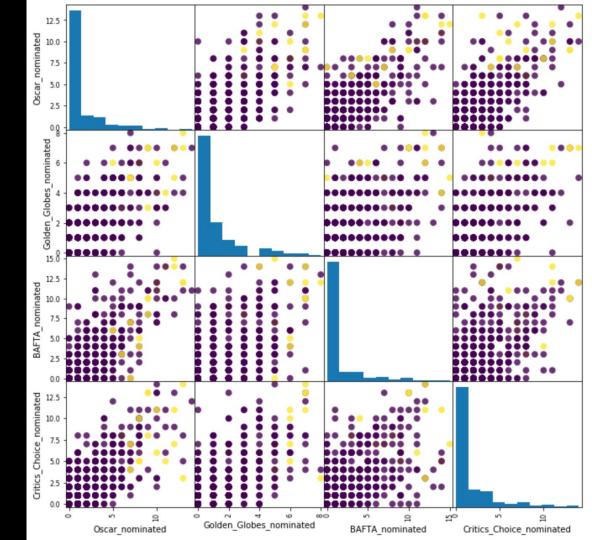
Exploratory Data Analysis

- Difficulty with EDA
  - Largely categorical data,
  - Very sparse matrix
- How to encapsulate large amounts of categorical 1's and 0's within graphs?
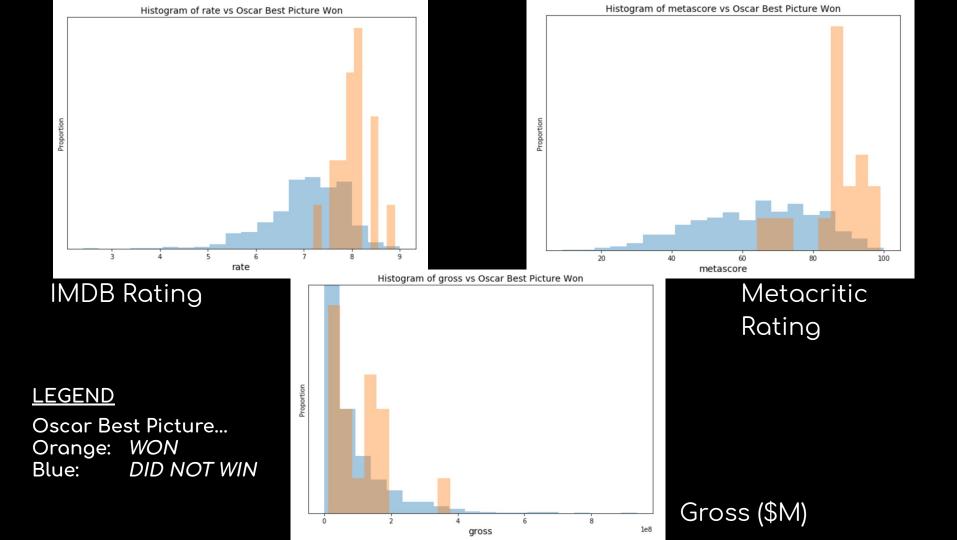 - Must narrow down question
- Need to consult Film Experts
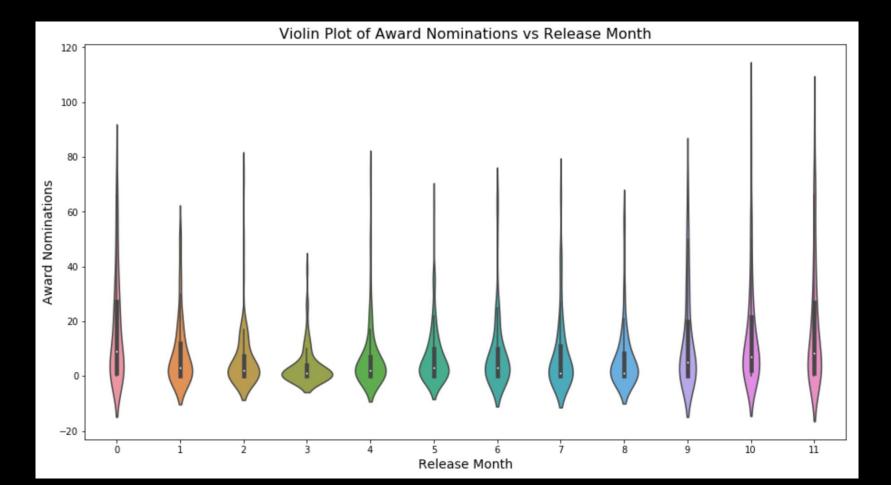
Boxplot of Oscar Best Picture Award Win Versus Total Awards Won

Boxplot of Oscar Best Picture Award Win Versus Total Awards Nominated

Scatter Matrix of Important Film Award Nominations

- Oscars
- Golden Globes
- BAFTA
- Critics Choice

Histogram of rate vs Oscar Best Picture Won

IMDB Rating

Histogram of metascore vs Oscar Best Picture Won

Metacritic Rating

Histogram of gross vs Oscar Best Picture Won

LEGEND

Oscar Best Picture...
Orange: WON
Blue: DID NOT WIN

Gross ($M)

Violin Plot of Award Nominations vs Release Month

predicting

THE

OSCARS®

THANK YOU!