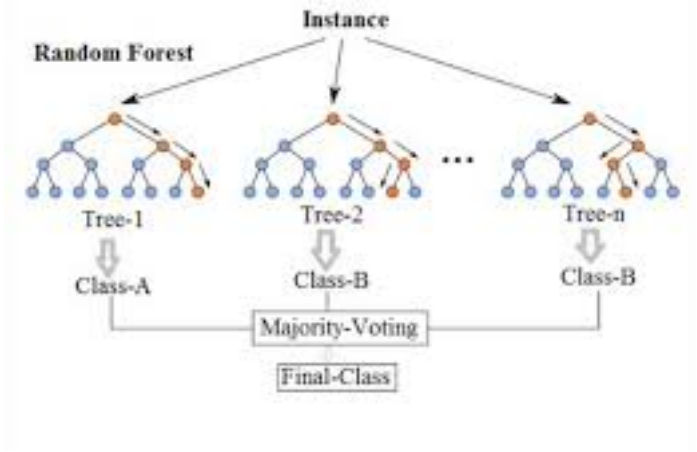
Will There Be A Next Season?!

Predicting television series renewal with IMDb data and categorical models

Nora May
Metis, Project McNulty

Outline

- Data Collection and Cleaning
- Feature Engineering
- Model Selection
- Predictions
- Future Directions



Data

Collection

- IMDb datasets, 4 relational

Cleaning

- Separate Series and Episodes
- English Language only
- Fill missing values for start year, total seasons

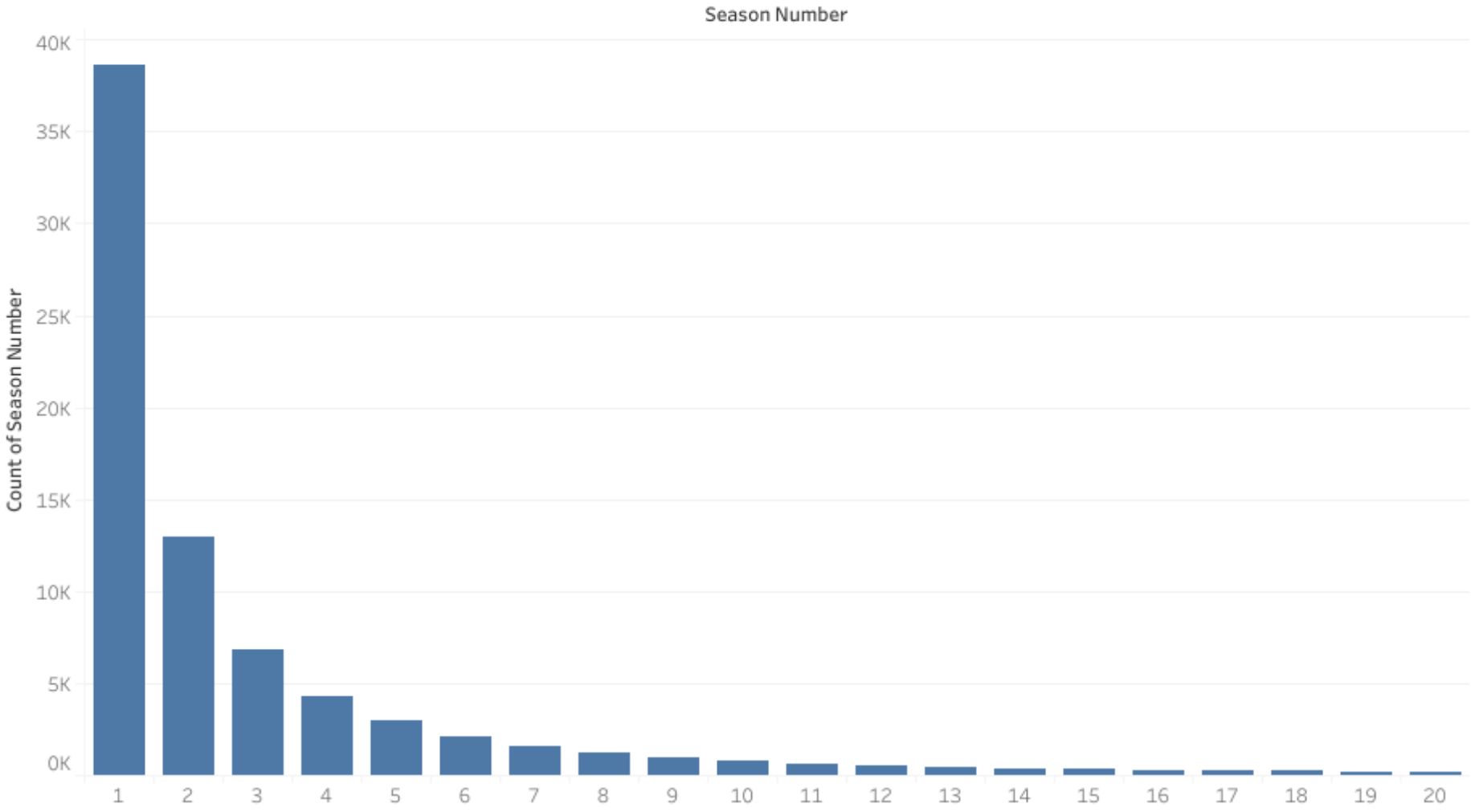
Construction

- Aggregate information per episode for "Season" database
- Categorical True or False for "is there a next season?"

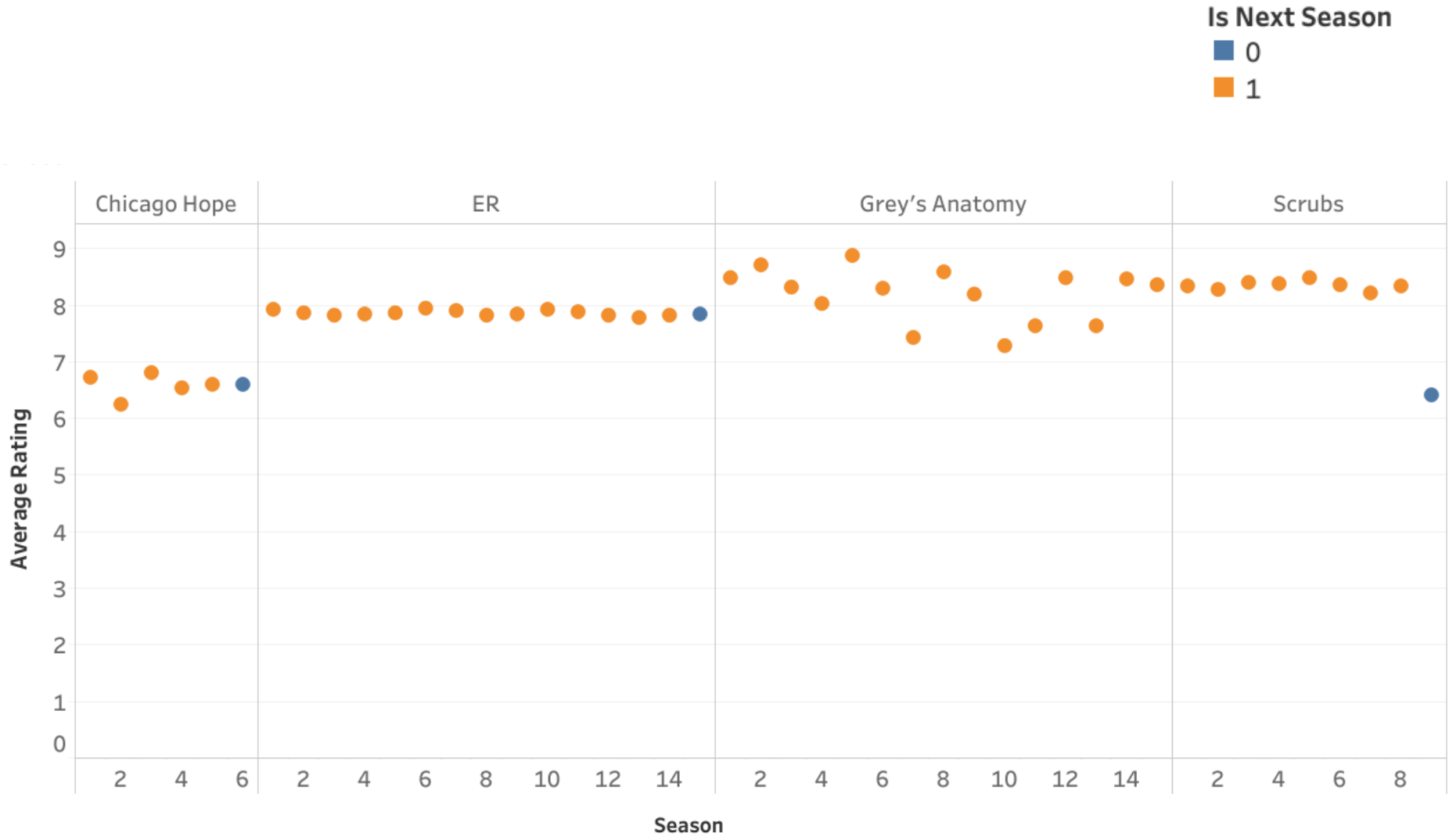
Selection

- Test/Train split on series code
- Separate out seasons that are currently airing/unknown

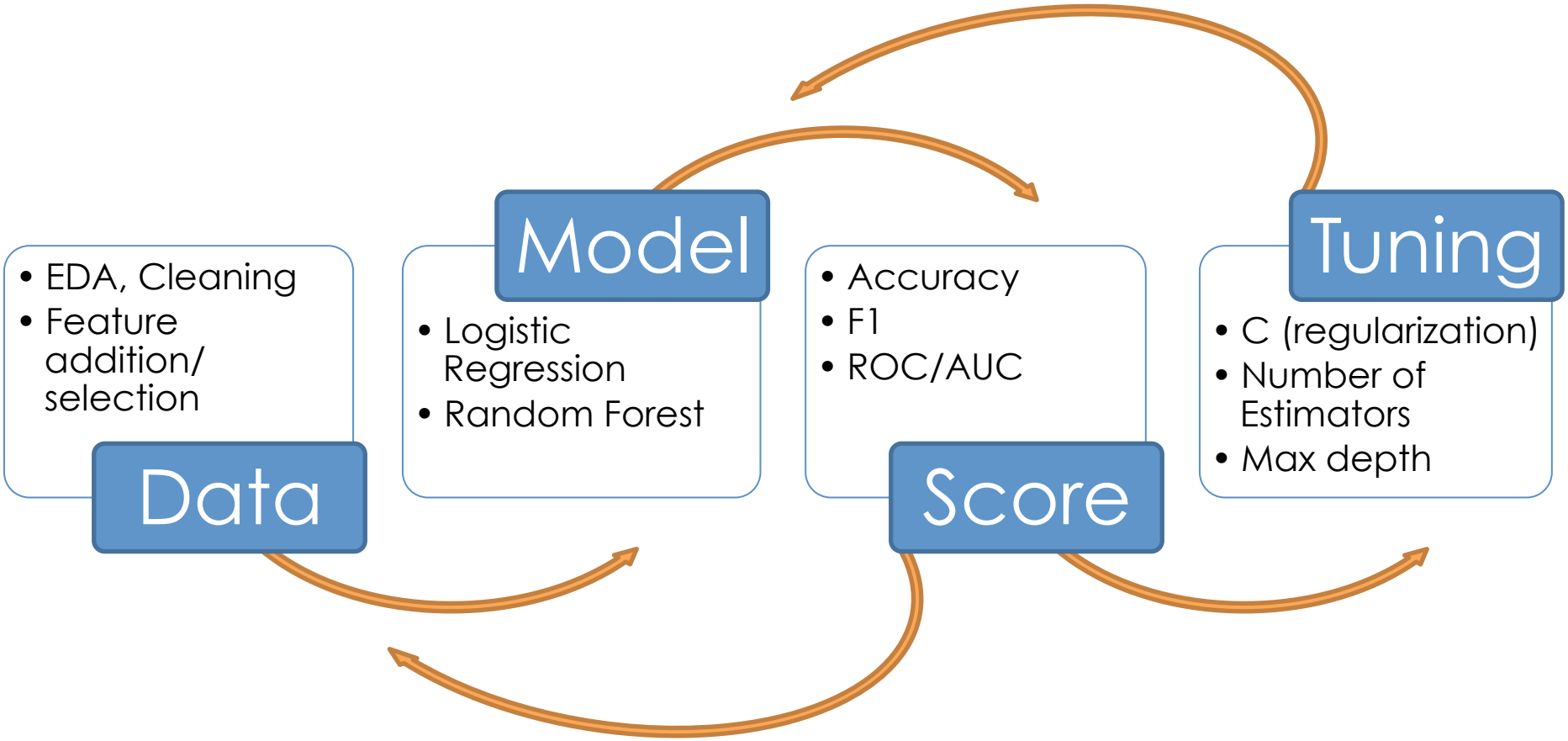
EDA


Season Number
Count of Season Number
40K
35K
30K
25K
20K
15K
10K
5K
0K
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

EDA

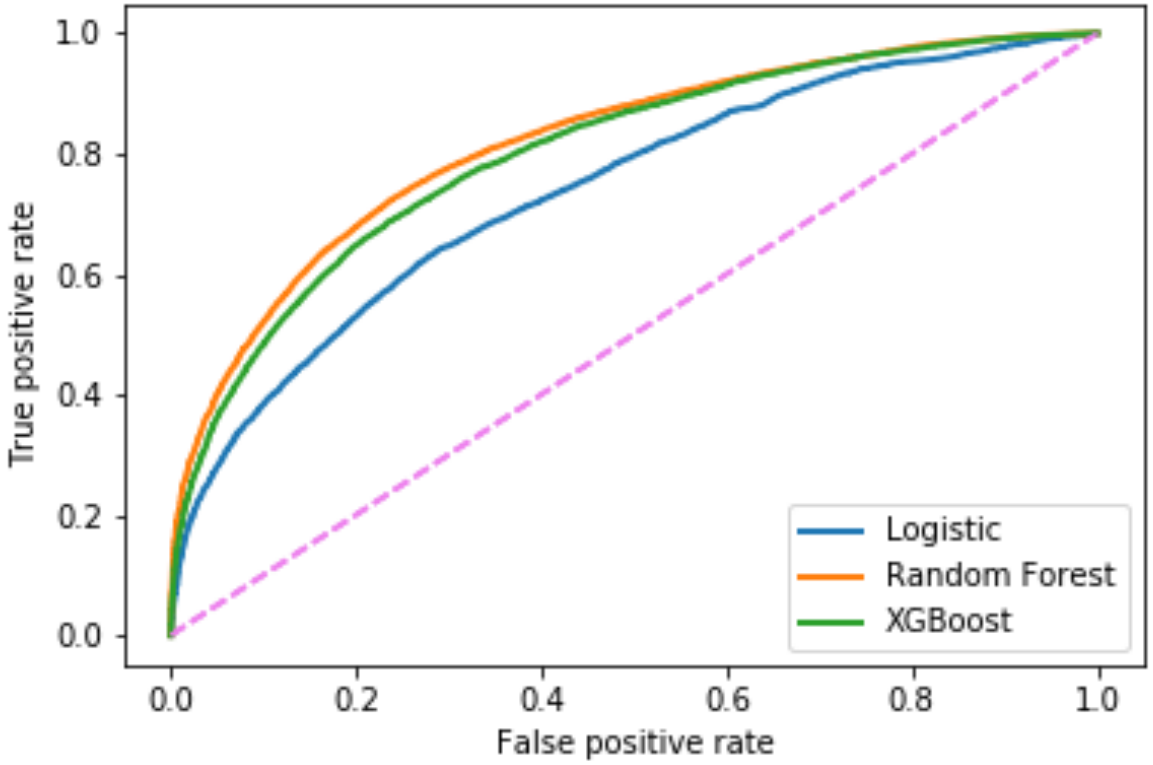

Is Next Season
0
1
Chicago Hope
ER
Grey's Anatomy
Scrubs
Average Rating
0
1
2
3
4
5
6
7
8
9
2
4
6
8
10
12
14
Season

Modeling


Data
• EDA, Cleaning
• Feature addition/ selection
Model
• Logistic Regression
• Random Forest
Score
• Accuracy
• F1
• ROC/AUC
Tuning
• C (regularization)
• Number of Estimators
• Max depth

ROC AUC Score:

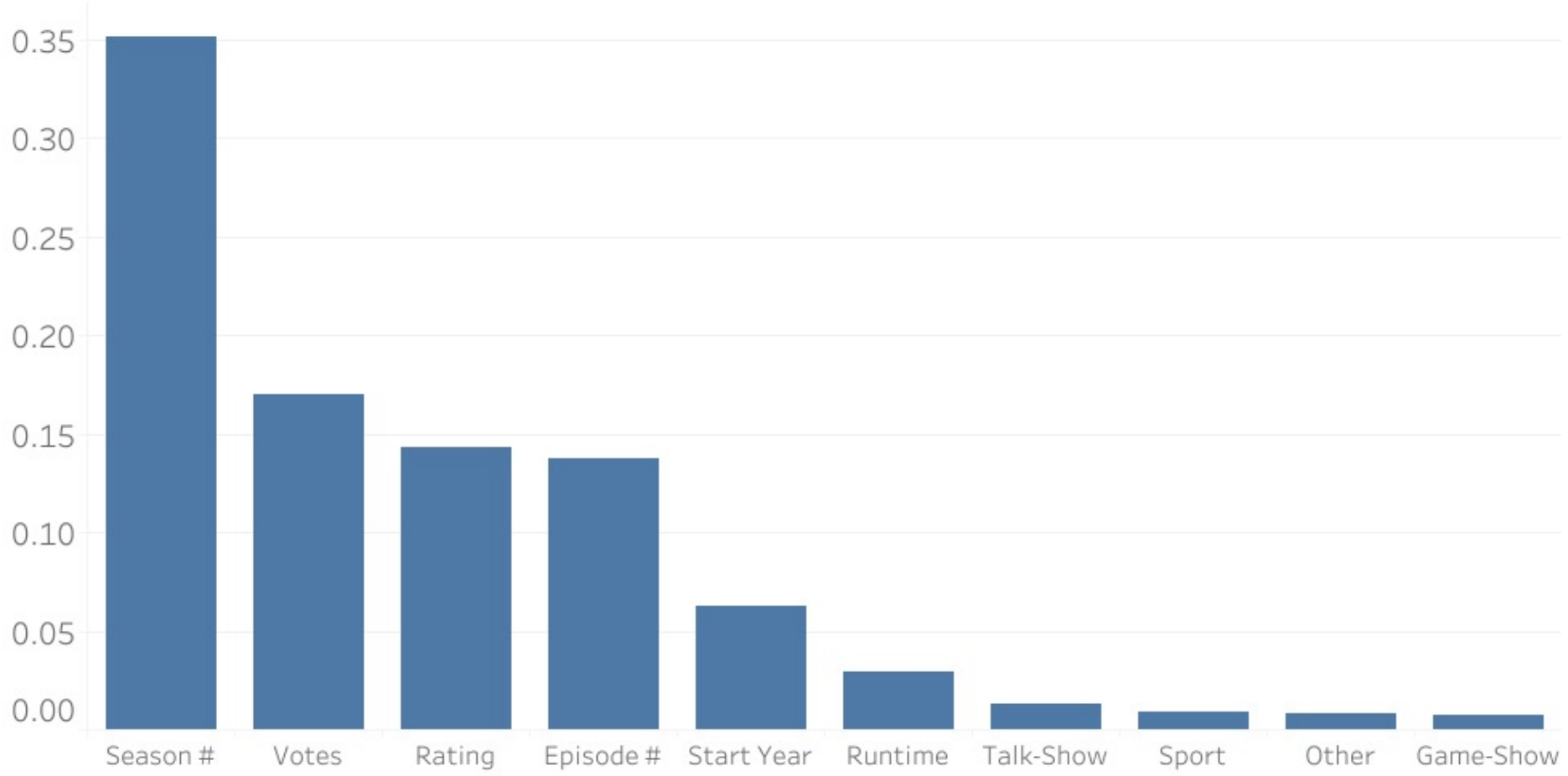
Logistic	=	0.74
Random Forest	=	0.81
XGBoost	=	0.79



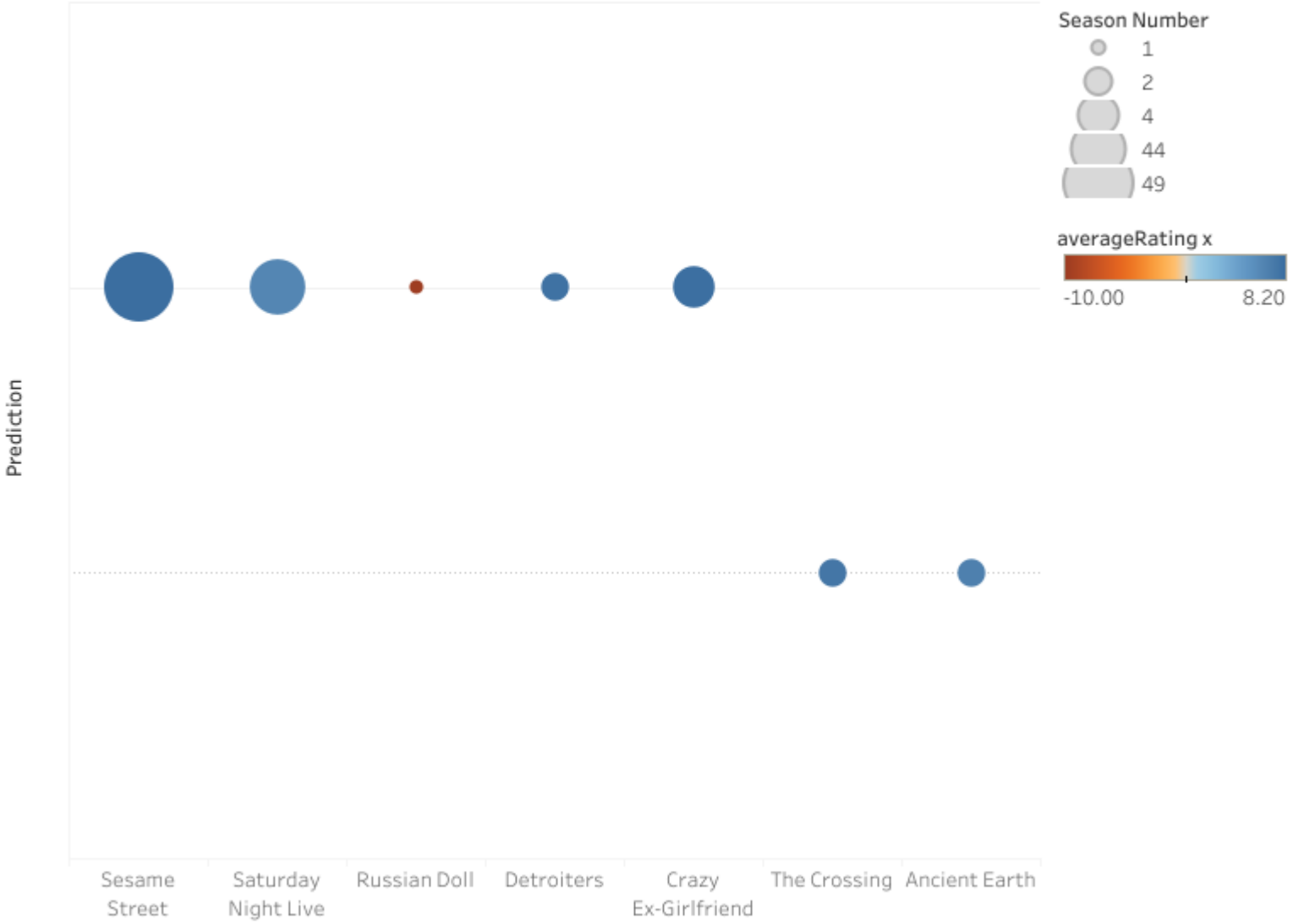
Random Forest Model: Feature Importances

50,000 observations (train), 100 estimators, max depth 11

71% accuracy (test), F1 Score: 0.73



Predictions



Future Work

- Further tuning of XGBoost model
- Features for cast, budget, network, viewership
- Began a web app where you can find out if your favorite show will be on


That's all Folks!