Scooby-Doo IMDb Ratings

By Regression Rockstars:

James Cai, Steph Reinke, Sarah Wu, Michael Zhou

December 7, 2023

Introduction

- Scooby-Doo is an animated TV show following a group of crime-solving teenagers
- First aired on CBS from 1969 - 1976, with many subseries releasing after
- We wanted to analyze the relationship between features of a Scooby-Doo episode and the IMDb rating
    - Motivation: aid future episodes in being more successful

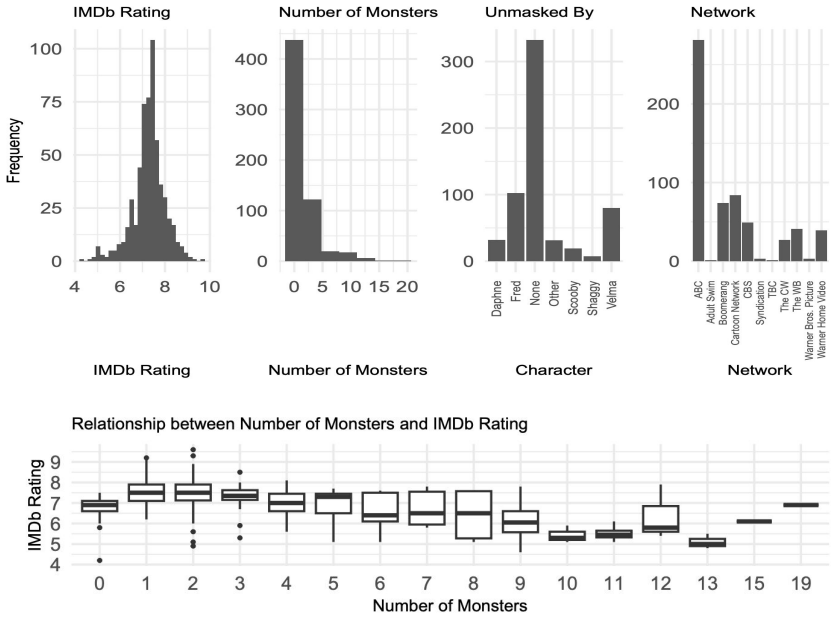
The Data

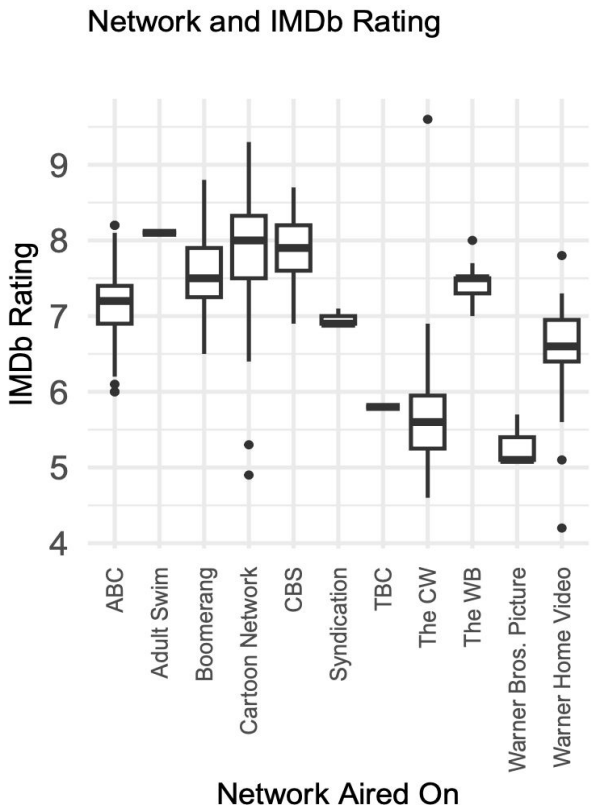
- Data set was found on Kaggle and was manually aggregated
- The curator watched every Scooby-Doo iteration while tracking variables
- Each observation is an episode from the Scooby-Doo franchise up until February 25, 2021, including movies and specials
- Relevant variables: network, monster.amount, unmask_villain

variable	class	description
index	double	Index ordering based on Scoobypedia
series_name	character	Name of the series in which the episode takes place or in movies' cases the Scoobypedia's grouping classification
network	character	Network the TV series takes place in, if it is a movie will use similar grouping as series.name variable
season	character	Season of TV Series, if not TV Series will default to the format
title	character	Title of Episode/Movie
imdb	character	Score on IMDB (NULL if recently aired)
engagement	character	Number of reviews on IMDB (NULL if very recently aired)
date_aired	double	Dated aired in US
run_time	double	Run time in min
format	character	Type of media
monster_name	character	Name of monster
monster_gender	character	Binary monster gender
monster_type	character	Monster type

Exploratory Data Analysis

- Aided with data cleaning
- Possible relationship between IMDb and # of monsters and network?





Final Model

Model Selection:

- Multiple linear regression
- Training Set (75%)
- Testing Set (25%)
- 3-fold cross validation
  - ⇨ Adjusted R^2
  - ⇨ AIC
  - ⇨ BIC

some significant p-values!

term	estimate	std.error	statistic	p.value
(Intercept)	7.469	0.144	51.764	0.000
monster.amount	-0.146	0.057	-2.560	0.011
monster.amount_x_unmask_villainFred	0.052	0.084	0.615	0.539
monster.amount_x_unmask_villainNone	0.113	0.058	1.949	0.052
monster.amount_x_unmask_villainOther	0.152	0.077	1.955	0.051
monster.amount_x_unmask_villainScooby	0.130	0.177	0.735	0.463
monster.amount_x_unmask_villainShaggy	0.272	0.218	1.248	0.213
monster.amount_x_unmask_villainVelma	0.184	0.085	2.171	0.031
network_Boomerang	0.438	0.102	4.274	0.000
network_Cartoon.Network	0.686	0.082	8.390	0.000
network_CBS	0.691	0.107	6.471	0.000
network_The.WB	0.202	0.115	1.756	0.080
network_other	-0.872	0.103	-8.439	0.000
unmask_villain_Fred	-0.059	0.187	-0.315	0.753
unmask_villain_None	-0.310	0.151	-2.057	0.040
unmask_villain_Other	-0.273	0.224	-1.220	0.223
unmask_villain_Scooby	-0.282	0.327	-0.863	0.389
unmask_villain_Shaggy	-0.378	0.446	-0.848	0.397
unmask_villain_Velma	-0.170	0.185	-0.921	0.358

imdb ~ network + monster.amount + unmask_villain + monster.amount * unmask_villain

Interesting Findings

Network:

- Baseline: ABC
- Episodes fared best on CBS and Cartoon Network

network_Boomerang	0.438
network_Cartoon.Network	0.686
network_CBS	0.691
network_The.WB	0.202
network_other	-0.872

Monster Amount:

- Episodes fared best when there were fewer monsters
- Unless Velma unmasks villain!

monster.amount	-0.146
monster.amount_x_unmask_villainVelma	0.184

Prediction:

- R^2 and RMSE lower for testing set!

Dataset	R-Squared	RMSE
Training	0.483	0.554
Testing	0.506	0.423

Conclusion

- Ratings are hard to predict!
- Network and monster.amount seem to be significant predictors for IMDb ratings for Scooby-Doo
- Limitations:
  - Manually aggregated data
  - Limited variables to work with
- Further Work:
  - More variables!
  - Explore data from other shows