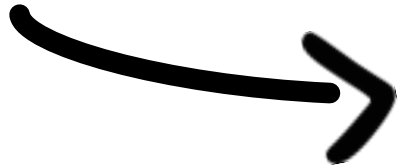
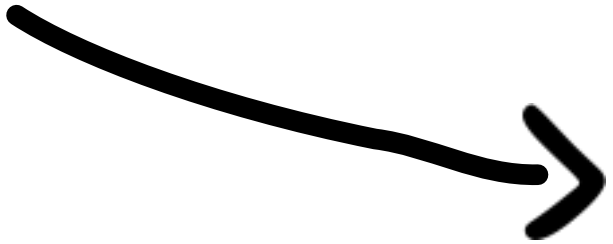
The Anatomy of The Simpsons

Deconstructing IMDB Ratings with Linear Regression

Turn this



Into this



Season Airdate Featured characters Director Writing credits Percentage of lines in the script Season premier or not Position in season schedule Is Milhouse in it Written by John Swartzwelder or not Number of credited writers Number of guest stars Season score Who spoke the TV viewership Number of ratings from males Number of ratings from females Number of ratings from males under 18 Rating among males Rating among females Number of ratings from males 18 to 29 Breakdown of secondary characters screen time Historical context

.....

"Homer at the Bat"
Aired: February 20, 1992
Season 3

Predicted IMDB rating

8.84

Actual IMDB rating

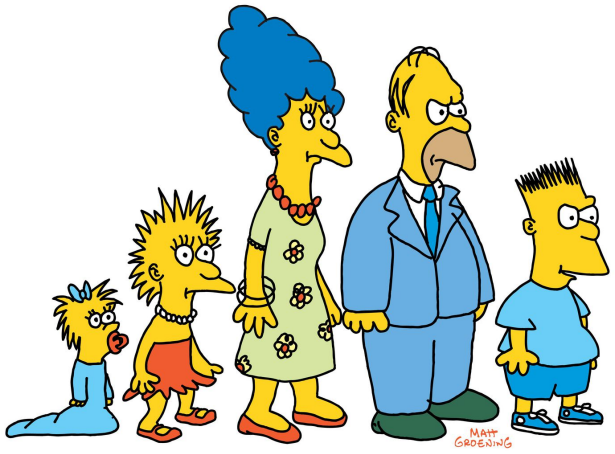
8.7

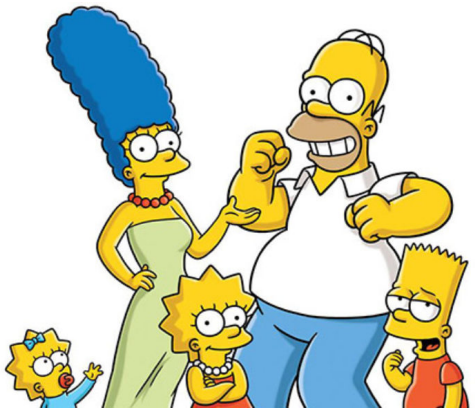
The Rundown

1. What and Why
2. The Process
3. The Model

1 Why The Simpsons?

- Longest-running American sitcom and scripted series
- A cultural touchstone
- Syndication deal (FXX) in 2013 worth $750 million
- $4.2 billion in consumer sales in first 25 seasons

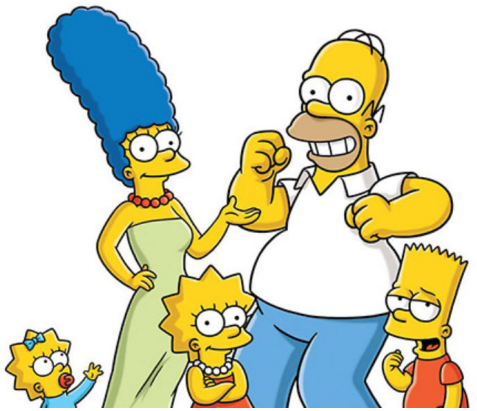




1 WHY IMDB RATINGS?

- Episode “quality” is a subjective measure
- 32 seasons and 684 episodes
- 357,627 ratings on IMDb

17TH AMONG TV SHOWS



2 Primary Data Sources

IMDb

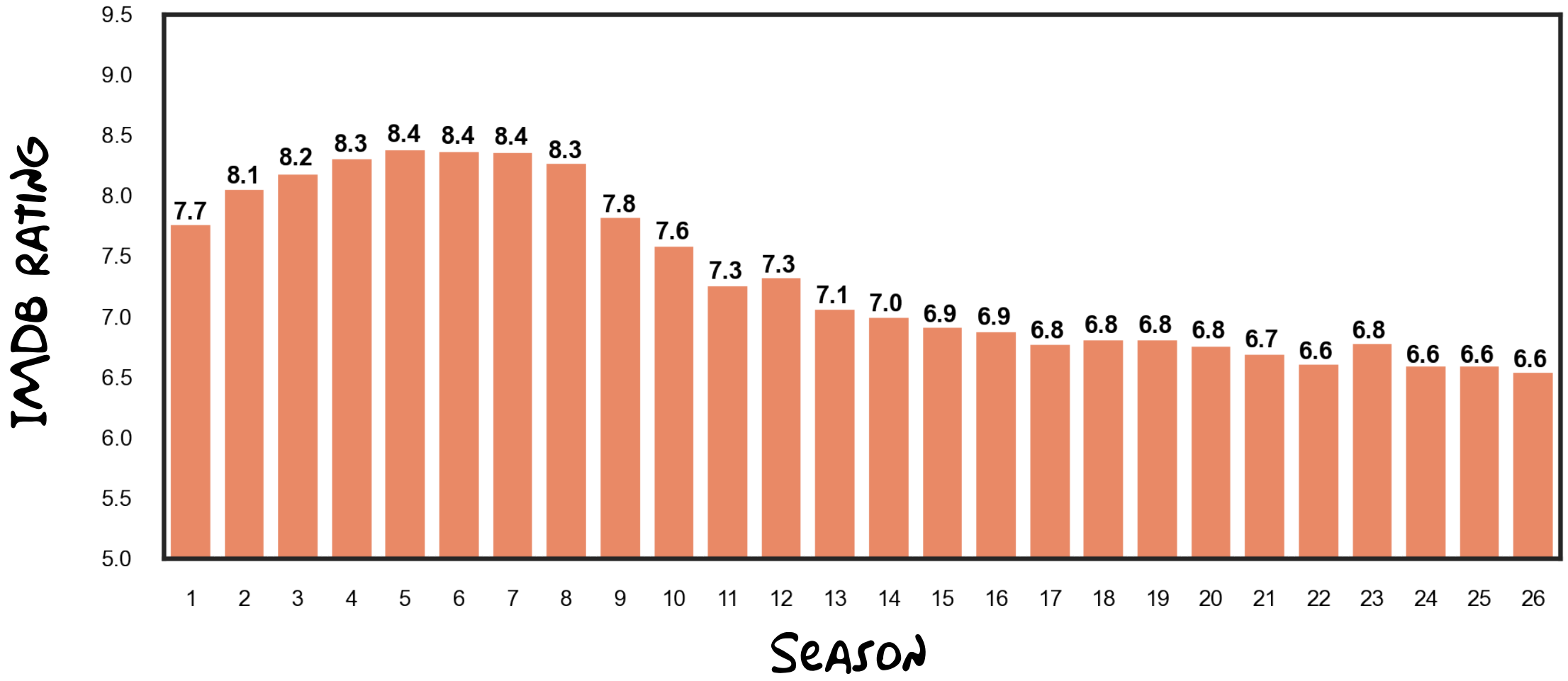
kaggle

Wikipedia
The Free Encyclopedia

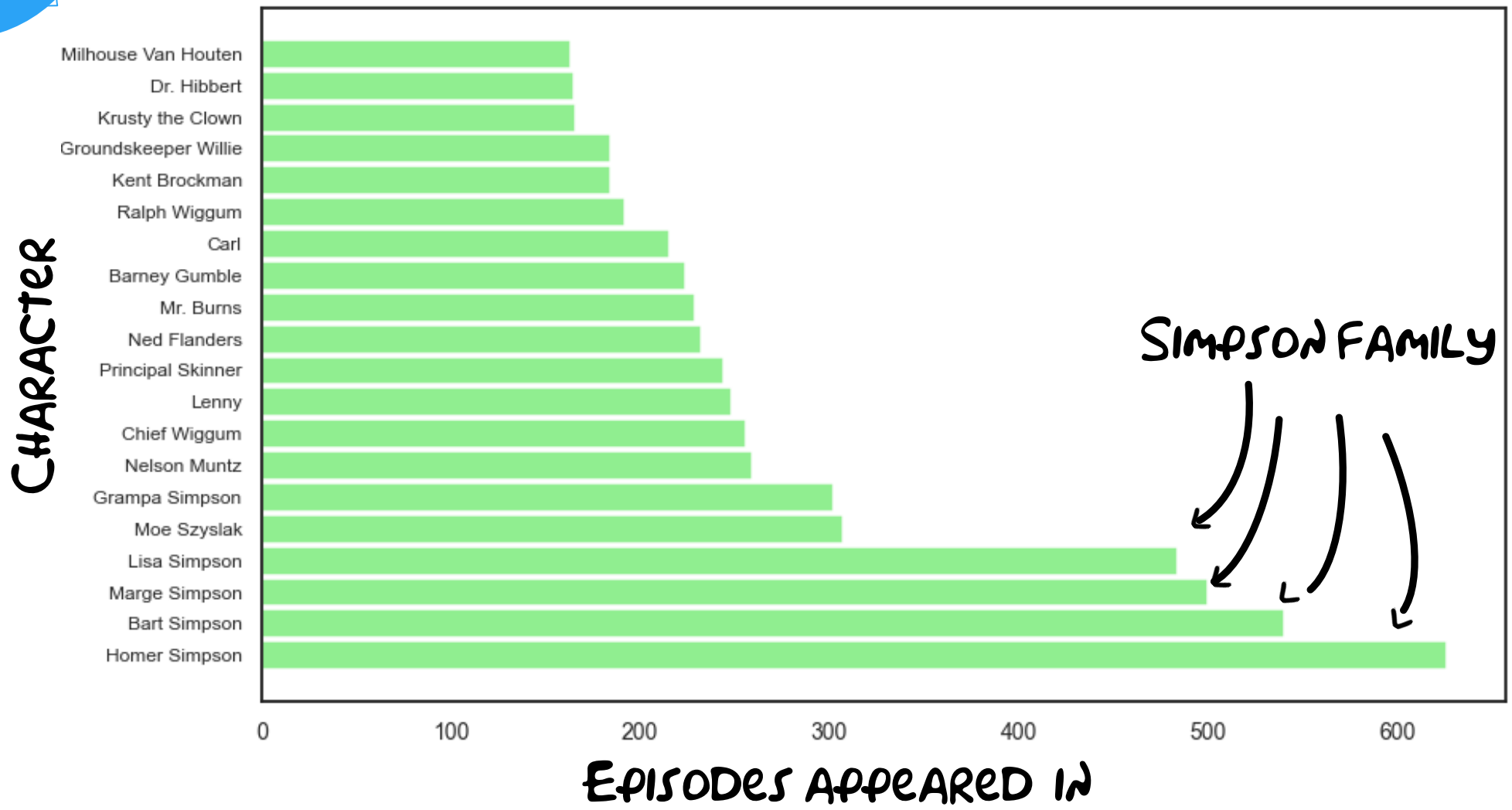
THE tvdb

Simpsons Wiki

2 IMDB Rating By Season



2 Episodes appeared in



2 Data Cleaning

3+ seasons with no script data

NaNs

“Lisa Goes Gaga”

2 Example Data Point

“Flaming Mo’s”

Season: 3, **Episode:** 10
Airdate: November, 21, 1991
Director: Rich Moore
Writer: Robert Cohen

Description: Moe claims a drink that Homer invented on his own, with the drink making the bar a local hot spot, but threatens his friendship with Homer.
Rating: 8.8
Number of ratings: 2851
Number of ratings from:

- **<18:** 5
- **18-29:** 624
- **30-44:** 1,203
- **45+:** 241
- **Male:** 1937
- **Female:** 209

Characters: Homer Simpson, Barney Gumble, Barfly #1, Krusty the Clown, Mayor Quimby, Marge Simpson, Selma Bouvier, Patty Bouvier, Bart Simpson, Wanda, Nelson Muntz, Maggie Simpson, Lisa Simpson, Drederick Tatum, Moe Szyslak, Quimby's Aide, Professor Frink, Kent Brockman, Coughing Man, Harv Bannister, Bernard, Lenny, Hugh Jass, Martin Prince, Self, Self, Lionel Hutz, Edna Krabappel

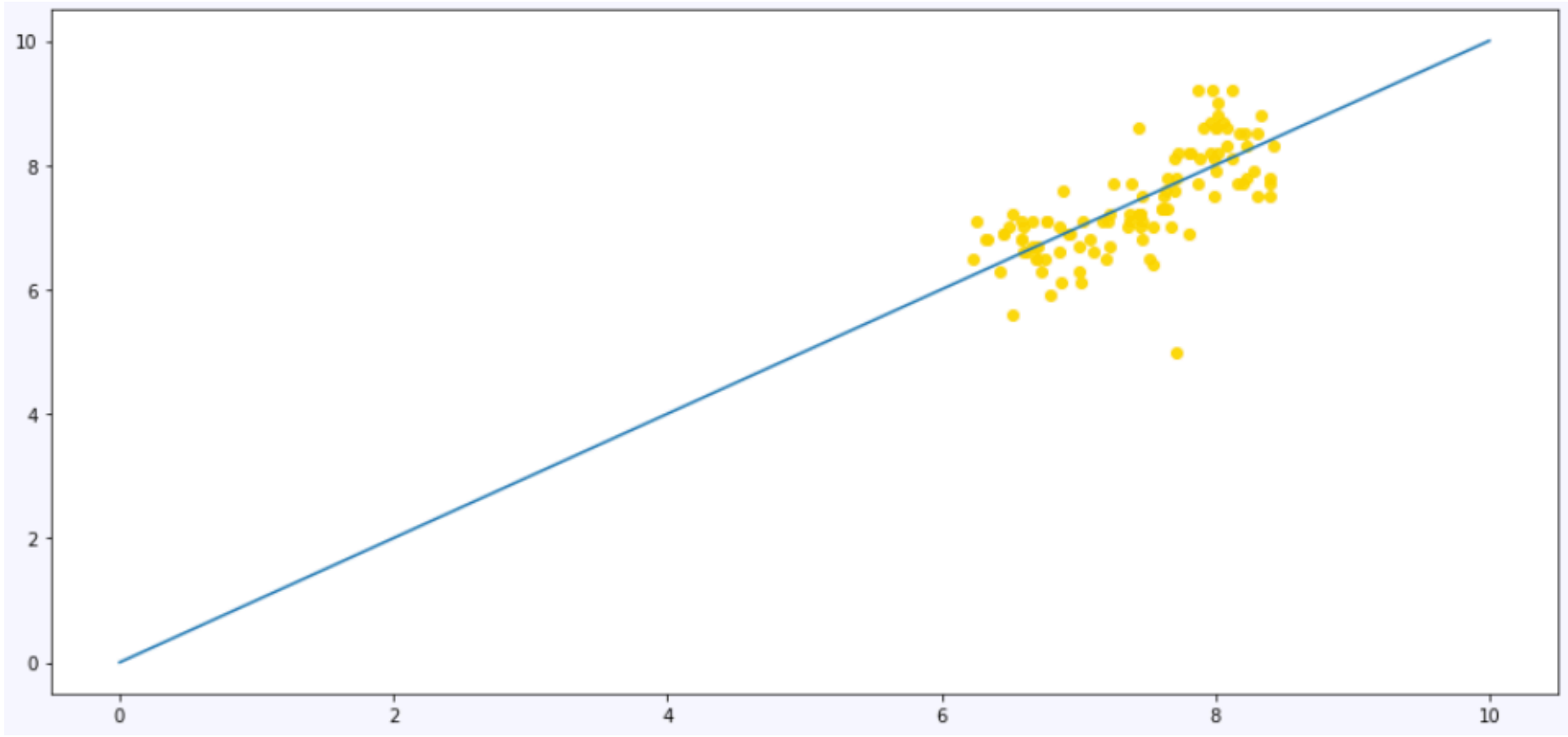
TV viewers: 23.9 million
Number of guest stars: 2
% of total words in script:

- **Homer:** 20.8477%
- **Bart:** 5.57465%
- **Lisa:** 2.44368%
- **Marge:** 3.2837%
- **Family total:** 32.1497%

% of non-family words in script:

- **Moe:** 37.8728%
- **Milhouse, Mr. Burns, Grampa, Flanders, Principal Skinner:** 0%

2 First crack: R2 score of .51



2 Feature Engineering

Rating	1.000000	1.000000
Season score	0.827049	0.827049
Number of ratings	0.791112	0.791112
Season	-0.759537	0.759537

Season score

John Swartzwelder multiplied	0.329695	0.329695
Grampa multiplied	0.306548	0.306548
Number of writers	0.303381	0.303381
Written by John Swartzwelder	0.240544	0.240544
Number of guest stars	-0.133082	0.133082

Written by John Swartzwelder
X
season score

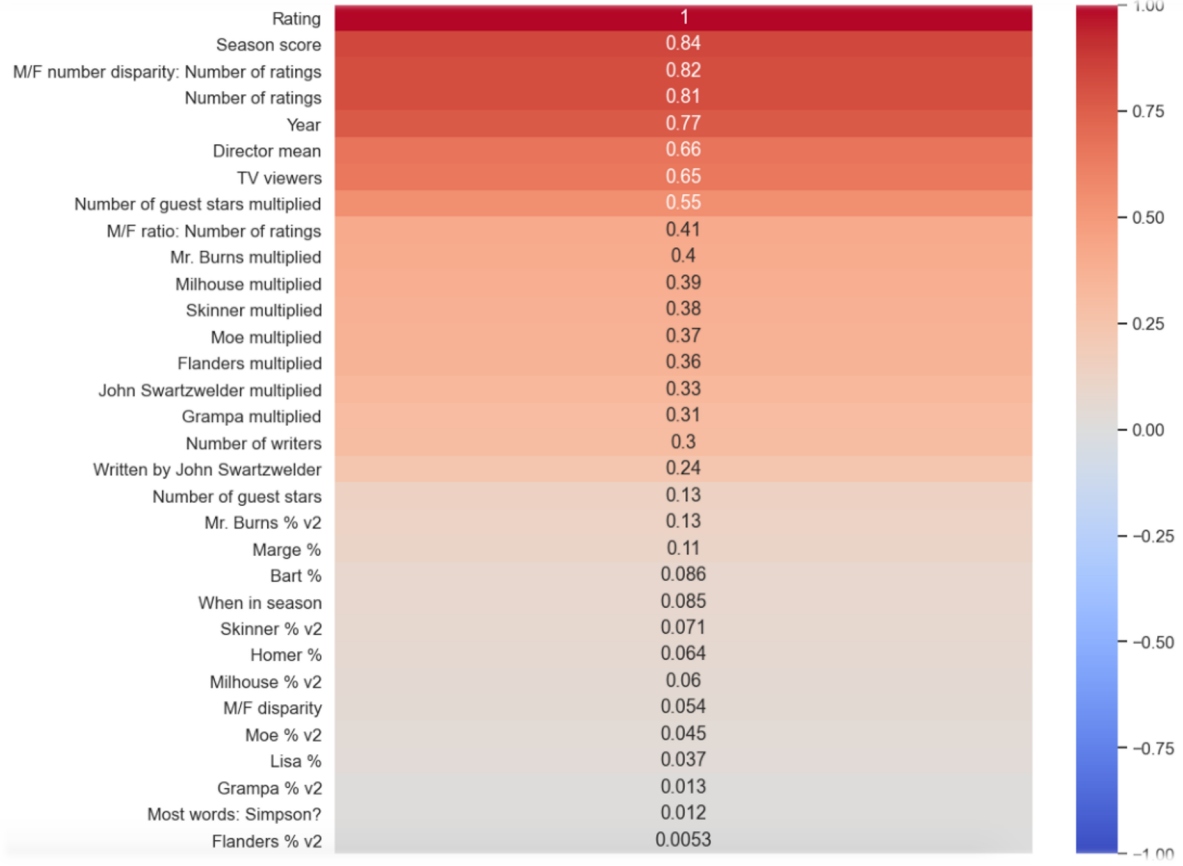
Treehouse of Horror?	0.094814	0.094814
Bart %	0.085727	0.085727
When in season	0.085034	0.085034
Lisa story	-0.074475	0.074475
Skinner % v2	0.071131	0.071131
Homer %	0.064171	0.064171

When in season

2 Progress!

```
1  lr_newest.score(X_train_lisa, y_train_lisa)
```

```
0.8341859975028593
```



R2 score of

.83

2 CV and Regularization

.79

R2 score after 3-fold cross-validation

+

.28

MAE

Cut down to 13 features

.78

R2 score after 3-fold cross-validation

+

.24

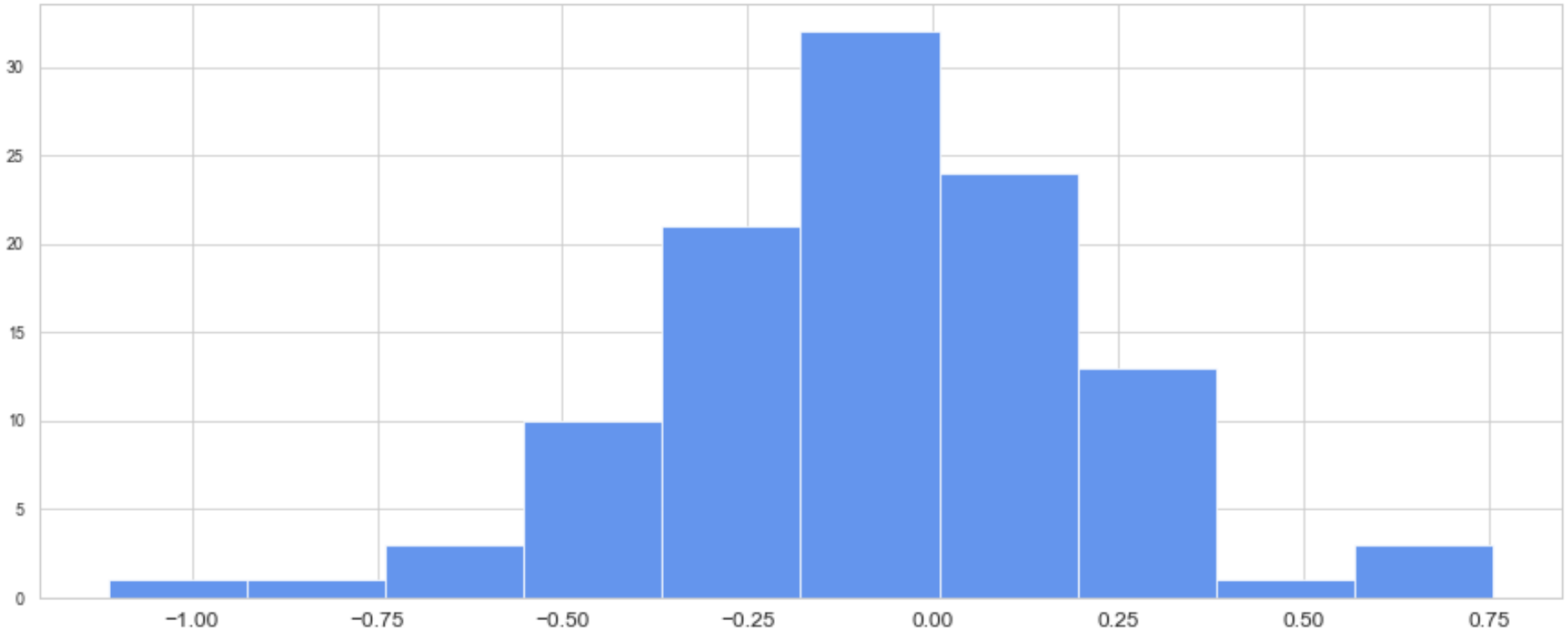
MAE

3 Final Model

$$y = 6.5675 + 0.0005X1 + 0.0658X2 - 0.0160X3 + 0.3051X4 + 0.4700X5 - 0.6102X6 + 0.2448X7 + 0.3527X8 - 0.0145X9 + 0.5066X10 + 0.0909X11 + 0.7849X12 - 0.1305X13 \pm 0.2377$$

X1 = “Number of ratings” X2 = “Director mean X3 = “Number of writers” X4 = “Homer %” X5 = “Lisa %” X6 = “Marge %” X7 = “Mr. Burns multiplied” X8 = “Flanders multiplied” X9 = “Number of guest stars”
X10 = “Season score” X11 = “Most words: Simpson?” X12 = “M/F disparity” X13 = ”M/F ratio: Number of ratings”

3 Residual distribution



3 Performance

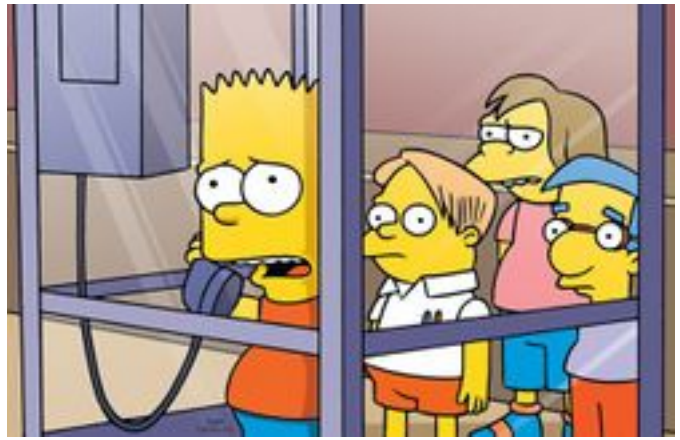
"Krusty Gets Busted"
Aired: April 29, 1990
Season 1

Predicted IMDB rating

8.14

Actual IMDB rating

8.4



"Bart on the Road"
Aired: March 31, 1996
Season 7

Predicted IMDB rating

8.52

Actual IMDB rating

8.6

References

- “IMDb,” IMDb.com
- “TVDB,” TheTVDB.com
- “Wikipedia,” Wikipedia.com
- “Wikisimpsons,” https://simpsonswiki.com
- “Homer – The Simpsons Cartoon Color Scheme,” https://www.schemecolor.com/homer-the-simpsons-cartoon-colors.php
- “Dialogue Lines of the Simpsons,” Pierre Megret, https://www.kaggle.com/pierremegret/dialogue-lines-of-the-simpsons
- “‘The Simpsons’ Lands $750 Mil Cable Syndication, VOD Pact with FXX,” Cynthia Littleton, https://variety.com/2013/tv/news/the-simpsons-lands-750-mil-cable-vod-syndication-pact-with-fxx-1200837036
- “Putting the $ in $pringfield: ‘Simpsons’ Merch a Cash Cow for 20th TV,” Marc Graser, https://variety.com/2013/biz/news/putting-the-in-pringfield-simpsons-merch-a-cash-cow-for-20th-tv-1200725561