



# **“Pitcher Perfect” Beer Rating Model**

Served by Devra Alper

# Introduction

## “Pitcher Perfect”

- Aims to predict beer ratings on ratebeer.com
- Can help breweries determine what features they should focus on when brewing a new beer
- Positive reviews could improve visibility on the site, which may increase sales and revenue



# Methodology

## Web scraping

BeautifulSoup



## Data manipulation



pandas



NumPy

## Modeling



## Plotting

matplotlib



seaborn



Sloop Brewing Company

# Sloop Juice Bomb

94

85

3.77 ★★★★★  
221 Ratings 150 Reviews

us East Fishkill, New York, United States

IPA - Hazy / New England (NEIPA) • 6.5%



[What these scores mean](#) [More statistics](#)

Juice bomb is a hazy, golden, unfiltered IPA. The low bitterness and late hopping showcases the citrusy, juicy notes that many American hops contain. A blast of tropical aroma is followed by a resinous and balanced flavor. Juice bomb features all the aspects and nuances of full hop flavor, not just bitterness — providing an upfront but easy drinking beer.

## Served in

Shaker Tulip

## Estimated calories

195 cal per 355ml

## Availability

Available Bottled Unknown On Tap Regional Distribution

## Tags

Edit

#cloudy-ipa

Show less ^

## More statistics

### Ratings

Average Rating	3.77
Review Average	3.67

### Overall Score

These scores are calculated using an algorithm and are weighted using a Bayesian formula; it takes into account both the reviews given by each user and the total number of reviews for the beer.

Score	94/100
Weighted Average	3.66

### More Information

Date added	February 27, 2016
Last updated	September 18, 2018

**1260**

Beer ratings

**57**

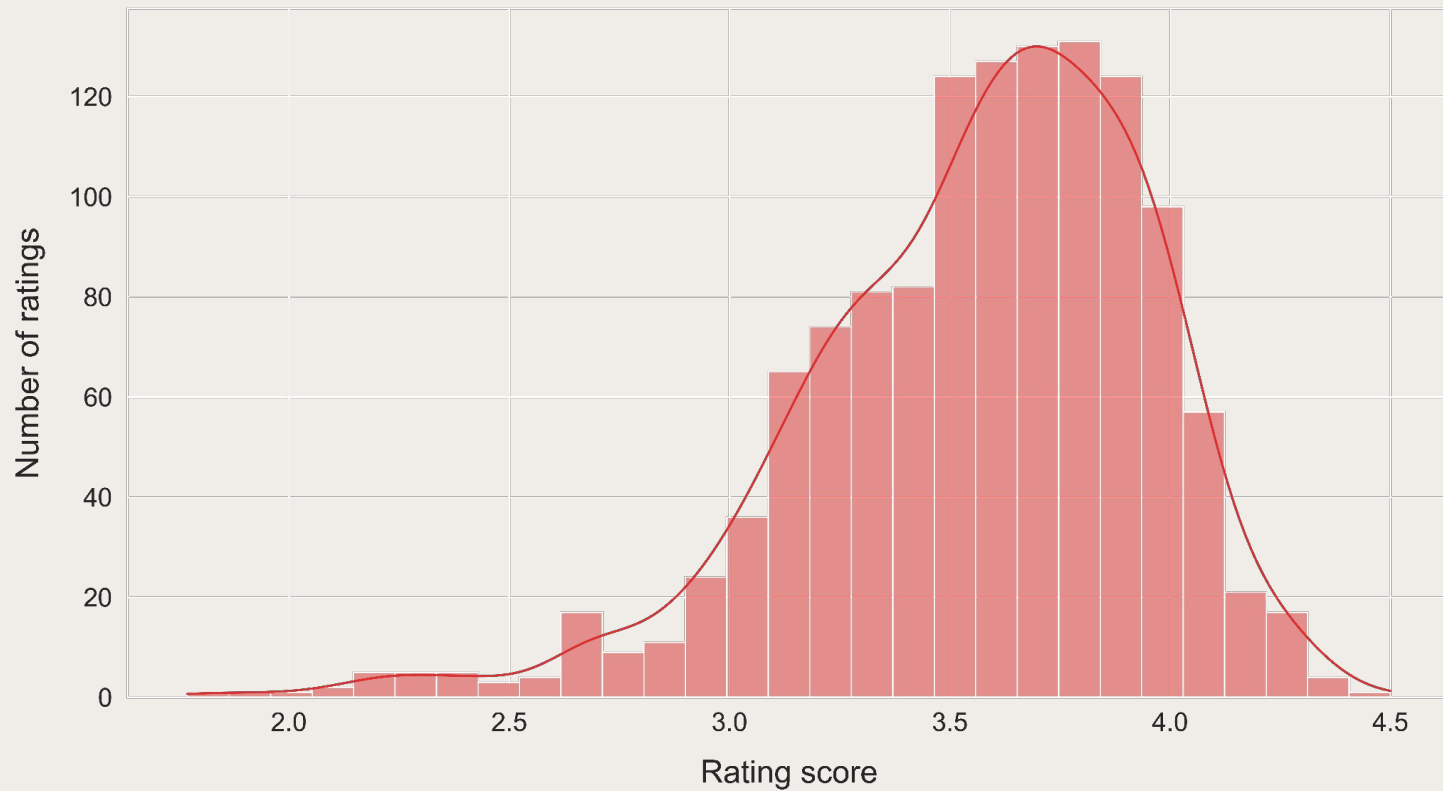
Features of the beer ratings

**33**

Features used in the final model



# Distribution of Ratings



# Modeling



## Linear regression

With all features  
included

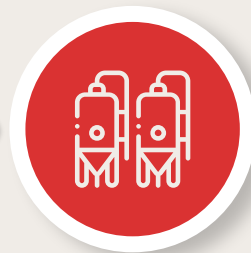
Train  $R^2 = 0.46$   
CV  $R^2 = 0.38$   
RMSE = 0.29



## Lasso

**cross-validation**  
For feature selection

Train  $R^2 = 0.45$   
CV  $R^2 = 0.32$   
RMSE = 0.29



## Lasso

With selected  
features only

Train  $R^2 = 0.35$   
CV  $R^2 = 0.32$   
RMSE = 0.31



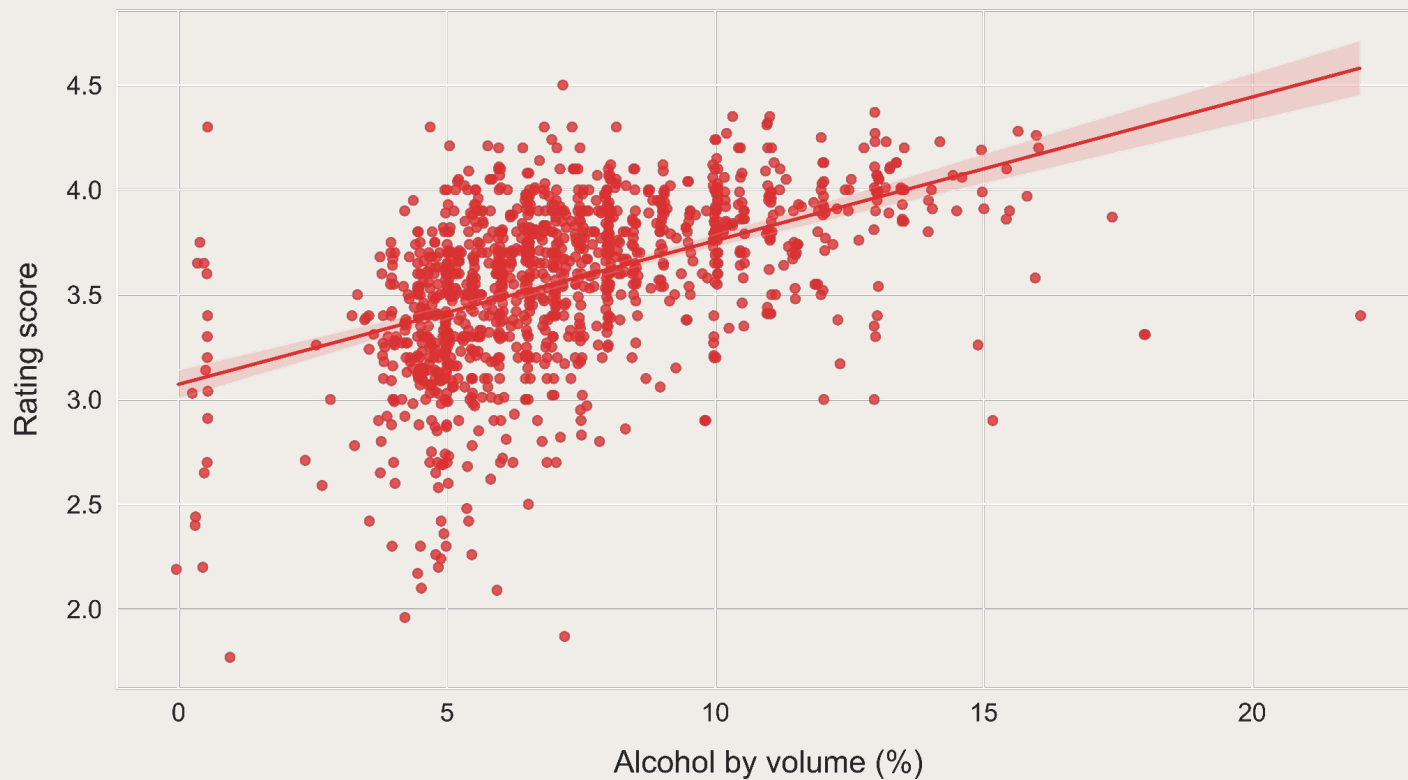
## Linear regression

With selected  
features only

Train  $R^2 = 0.45$   
CV  $R^2 = 0.40$   
RMSE = 0.29

**Coefficient = 0.16**

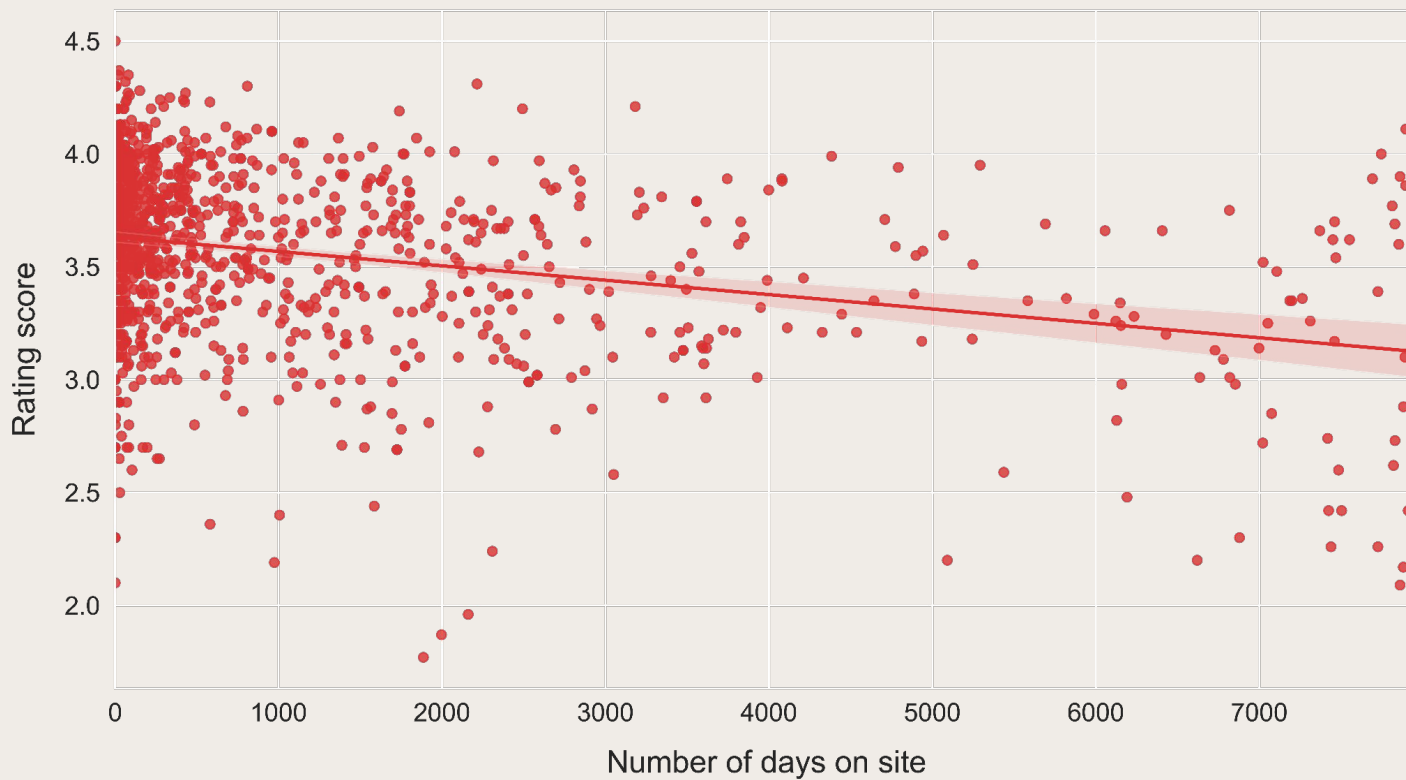
## Linear Regression between Rating & ABV





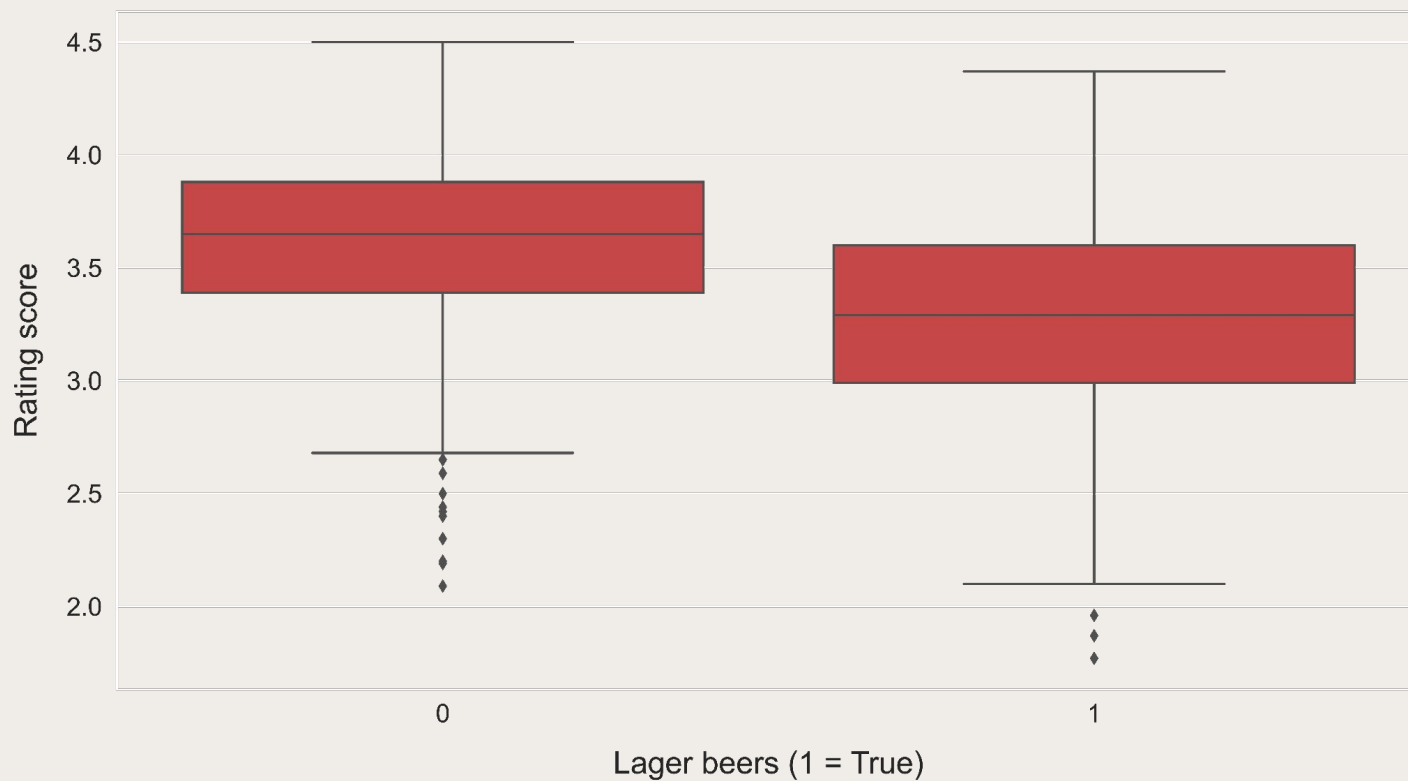
**Coefficient = -0.09**

## Linear Regression between Rating & Days on Site



**Coefficient = -0.07**

## Relationship between Rating & Lager Beers



# Other features

## Location of brewery

United States  
Russia  
Finland  
England  
France  
New Caledonia  
Poland  
Japan  
Germany  
Netherlands

## Overall style

Cider, Mead, Sake  
Other styles  
Wheat beer  
Stout and Porter  
Sour beer

## Served in

Yard  
Flute  
Weizen  
Tulip  
English pint  
Stein  
Lager glass  
Stem glass  
Trappist glass  
Shaker  
Masu  
Ochoko  
Guinomi

## Other

Number of ratings  
Number of reviews



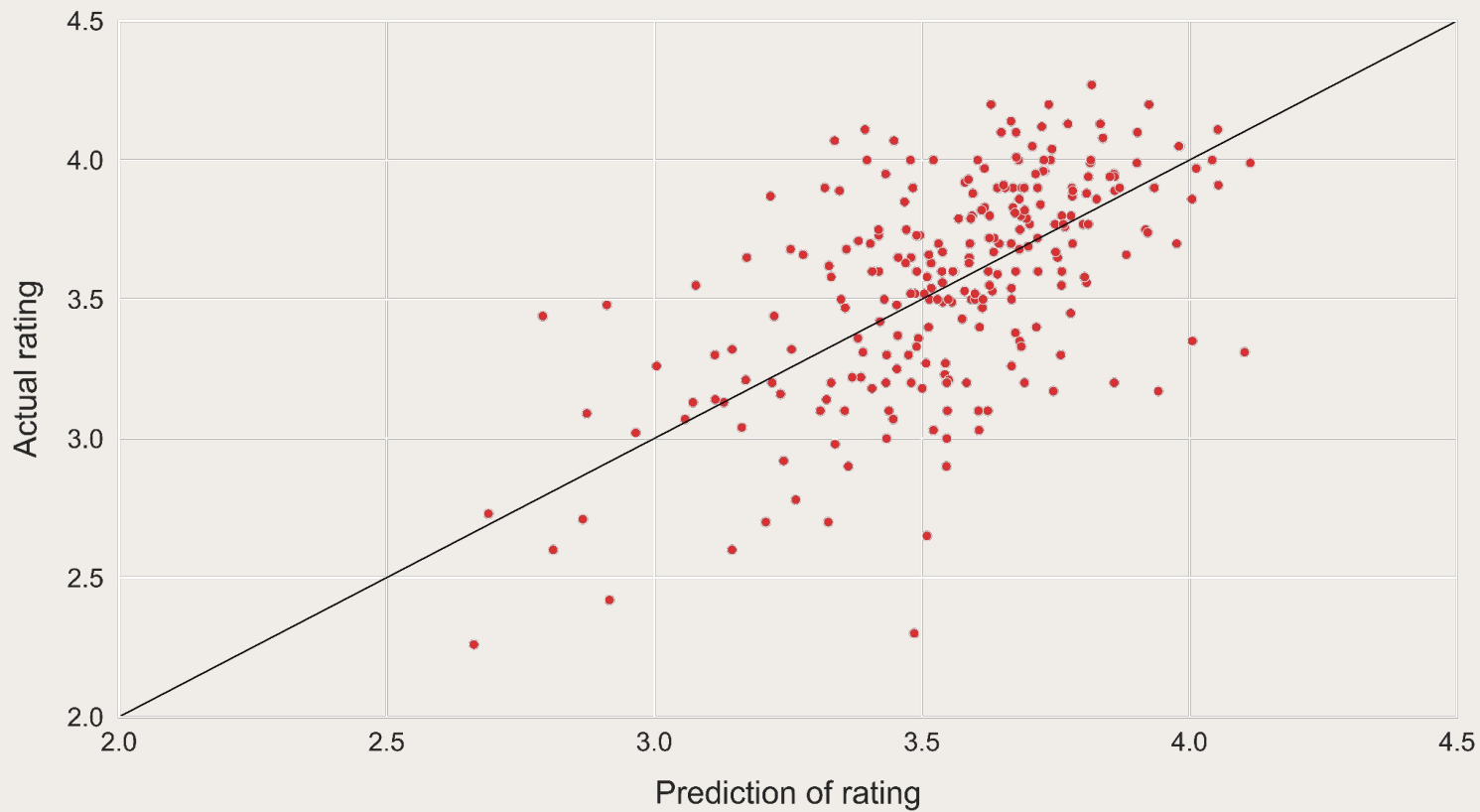
## **Final model**

Results on test data

$$\mathbf{R^2 = 0.34}$$

$$\mathbf{RMSE = 0.30}$$

# Actual vs. Predicted Rating



# Recommendations

**When brewing a new beer, keep this in mind:**

- The higher the **ABV**, the better
- Don't brew a lager! **Stouts, Porters, & Sour beers** are the only styles positively correlated with rating
- **Expect the rating to decrease** a bit after beer is initially added to ratebeer.com

**Next steps:**

- Analyze “description” and written reviews on site to find patterns & any association with rating

# Cheers!

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik

Please keep this slide for attribution