

CAN WE PREDICT SCORES AT RATEBEER.COM



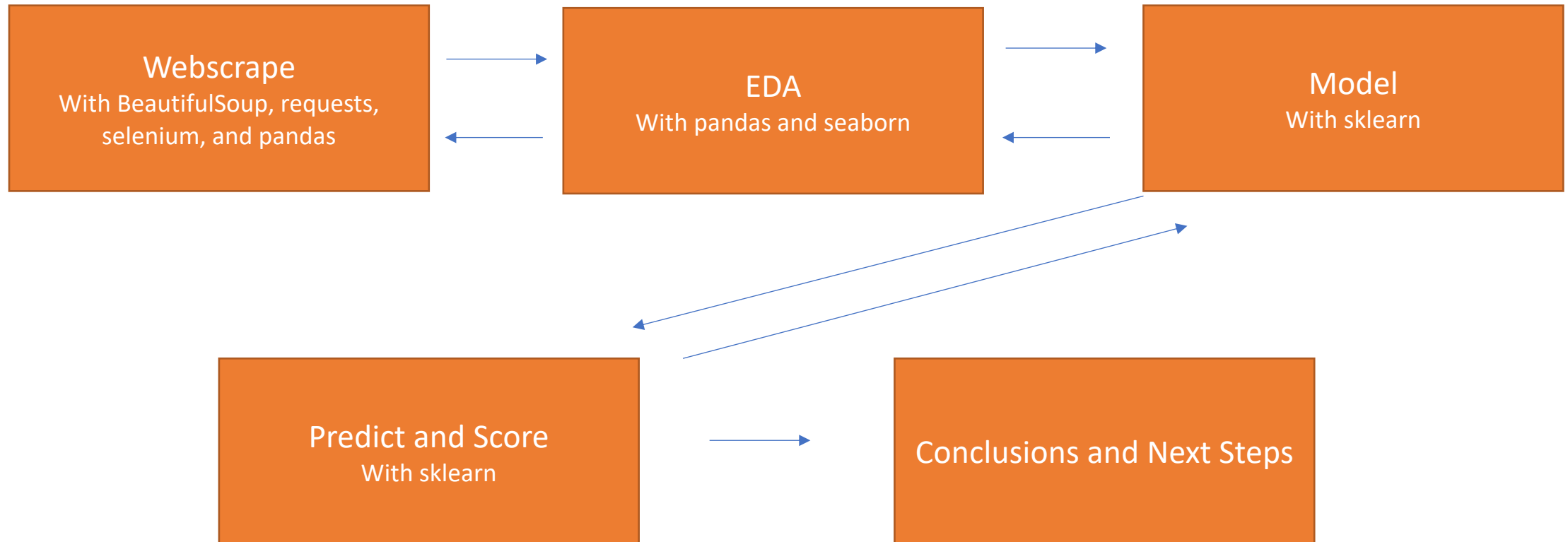
Motivation:

What influences a beer's score on ratebeer.com and can you predict what the score of a beer will be based on some of its basic attributes?

Objectives:

Pull data from ratebeer.com and use linear regression to analyze it.

Methodology



Webscraping Data

- Multiple layers of links
- Tables only load when webpage loads
- Available Tags

RateBeer

Beer Forums Places

Home / Beer Styles

Beer Styles

Pending Update This page is undergoing updates and might have a few omissions or parentage inas meantime for your convenience.

Anglo-American Ales

- Alt**ier** / Stic**ke** Alt
- Barley Wine / Wheat Wine / Rye Wine
- Bitter - Ordinary / Best
- Bitter - Premium / Strong / Extra Special (ESB)
- Blonde Ale / Golden Ale
- Brown Ale
- Brown Ale - Imperial
- California Common / Steam Beer
- Cream Ale
- IIPA DIPA - Imperial / Double Hazy (NEIPA)
- IIPA DIPA - Imperial / Double IPA
- IPA
- IPA - Belgian / Saison
- IPA - Black / Brown / Cascadian Dark
- IPA - Rye

Lagers

- Amber Lager - Czech / Polotmavý
- Amber Lager - International / Vienna
- Bock - Doppelbock
- Bock - Dunkler Bock
- Bock - Eisbock
- Bock - Heller Bock / Maibock
- Dark Lager - Dunkel / Tmavý
- Dark Lager - India / Hoppy
- Dark Lager - International / Premium
- Helles / Dortmunder Export
- Malt Liquor
- Märzen / Oktoberfest Bier
- Pale Lager - American
- Pale Lager - Hoppy / India (IPL)
- Pils
- Pils - International / Premium

RateBeer

Beer Forums Places

Home / Beer Styles / Alt**ier** / Stic**ke** Alt

Alt**ier** / Stic**ke** Alt

How Many Have You Had? [Alt**ier** / Stic**ke** Alt](#)

Join Here and Start Rating!

The Alt**ier** is a well-balanced, well-attenuated, bitter yet malty, clean, and smooth, amber to copper colored to style beer. The bitterness is balanced by the malt richness, but the malt intensity and character can range from and malty than international amber lagers. The Stic**ke** Alt and Doppelstic**ke** Alt are a malter and hopper versic

Serve in [Kölsch/Alt**ier**](#) ☐ Include Retired [Show O](#)

	NAME	COUNT
1	Uerige Doppel Sticke	946
2	Schumacher 1838er	122
3	Godspeed Oberkassel	19
4	Vrooden Doppelsticke Altier	16
5	Holy Mountain / E9 Donnylana Doppelsticke	11
6	Kernker Aoltbeer	33
7	Uerige Sticke	703
8	Schlüssel Sticke	17
9	Widmer Brothers Winternacht	69

RateBeer Best Update: Our annual beer awards have been

RateBeer

Search for Beers, Breweries, or Bars...

Home / Beer Styles / Alt**ier** / Stic**ke** Alt

VERIFIED TOP 50

Uerige Doppel Stic**ke**

3.78 100 1,269 Ratings 946 Reviews [Alt**ier** / Stic**ke** Alt](#) · 8.5%

[What Does a Score Mean](#) [More Statistics](#)

Uerige DoppelStic**ke** might just be "Germany's interpretation of a barley-wine", ie a third definition next to England's "malt-driven" idea (Nevada's BigFoot).
OG - 1085; IBU - 75; Ingredients; Hops:Spalt Spalter, Hallertauer Perle, Spalt Hallertauer; Malts: Pilsner malt, caramel malt, roasted mal

Served in [Kölsch/Alt**ier**](#) Estimated calories 255 cal per 355ml

Availability [Unknown Bottled](#) [Unknown On Tap](#) [Regional Distribution](#)

Tags [Edit](#)

[#1001-beers-you-must-taste-before-you-die](#) [#doppelstic**ke**](#) [#dusseldorf-alt**ier**](#) [#hallertau](#) [#perle](#) [#spalt](#)

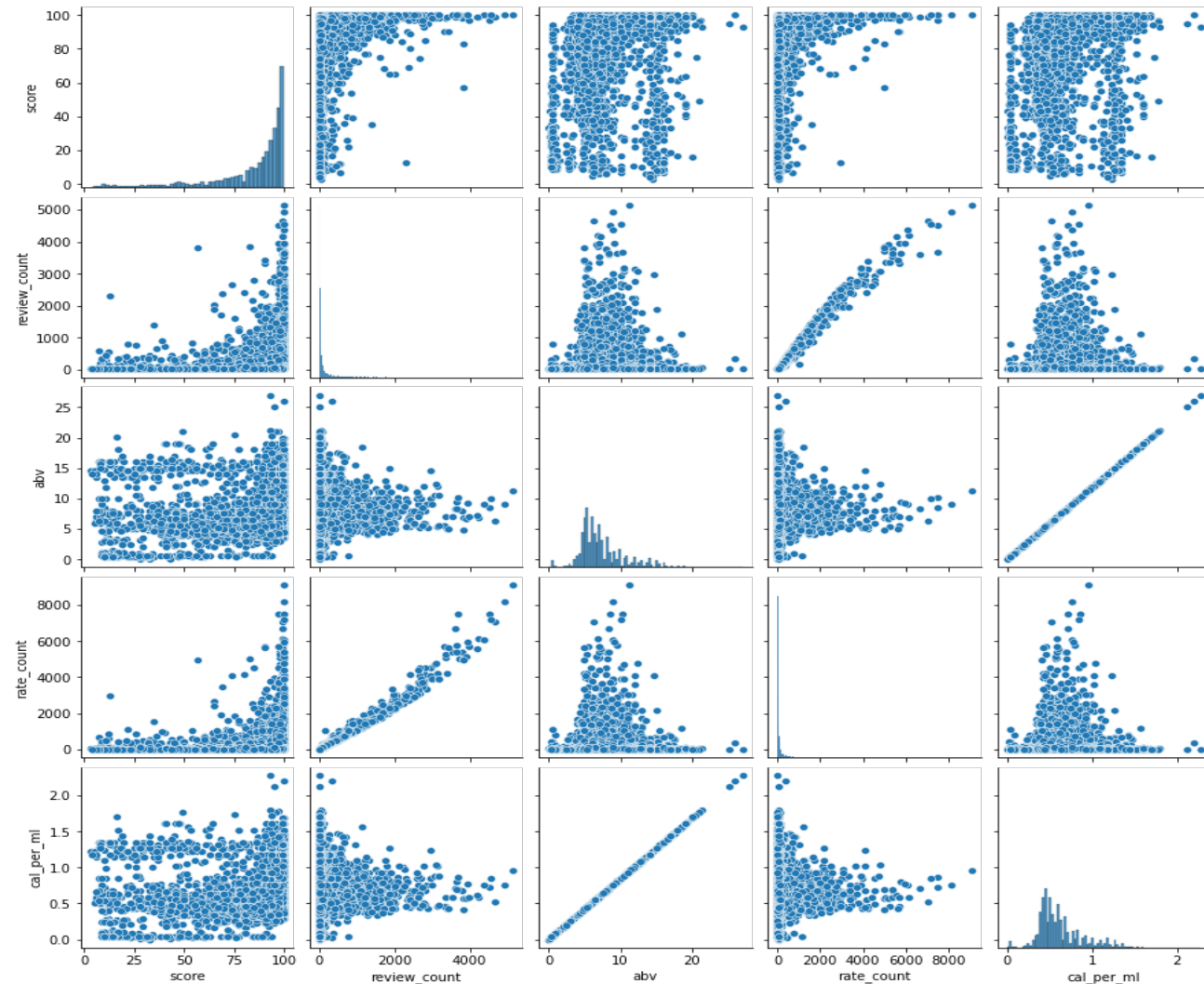
Show less

Reviews (954)

[Friends](#) [Score breakdown](#) [Quick ratings](#)

Baseline Model

- Multi-collinearity with abv and calories and also with rate count and review count
- Reduced linear features to 2
- R-squared for baseline was Only 0.0508



Adding Nonlinear Features and Finalize Model

- Tried feature engineering
- Attempted improvement with regularization.
- Final model RMSE is 12.3
- R-squared on test is 0.54

Non-Linear Feature	R-Squared Improvement on Training Data
Style	.55
Country	.16
Other	.06
Total	.57

Conclusions

- The model is underfitted and needs additional or different data.
- Root Mean Squared Error of 12.3 points on a scale of 1-100 is too high.
- Webscraped data was not sufficient to build a good model

Next Steps

- Webscrape the data again using a deeper process pulling words from reviews instead of tags as well as pulling more/all beers.
- Explore additional data sources to get data such as price and availability.