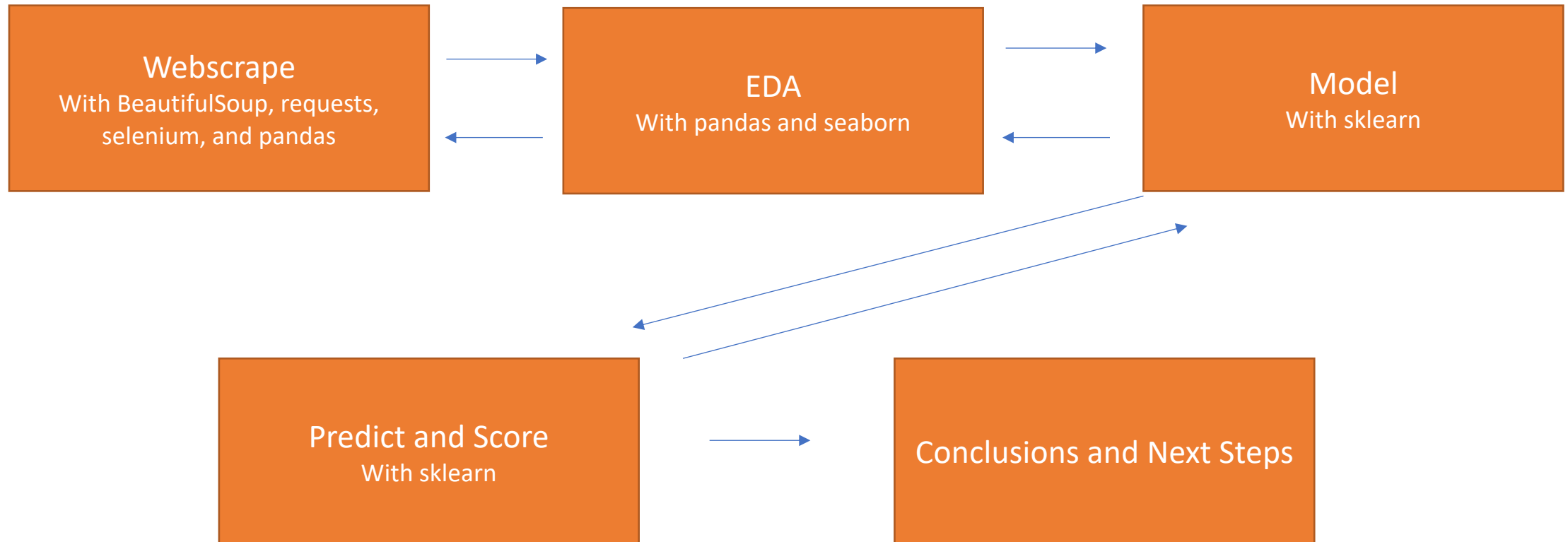# CAN WE PREDICT SCORES AT RATEBEER.COM

RateBeer

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

Webscrape
With BeautifulSoup, requests, selenium, and pandas

EDA
With pandas and seaborn

Model
With sklearn

Predict and Score
With sklearn

Conclusions and Next Steps

Webscraping Data

- Multiple layers of links

- Tables only load when webpage loads

- Available Tags

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
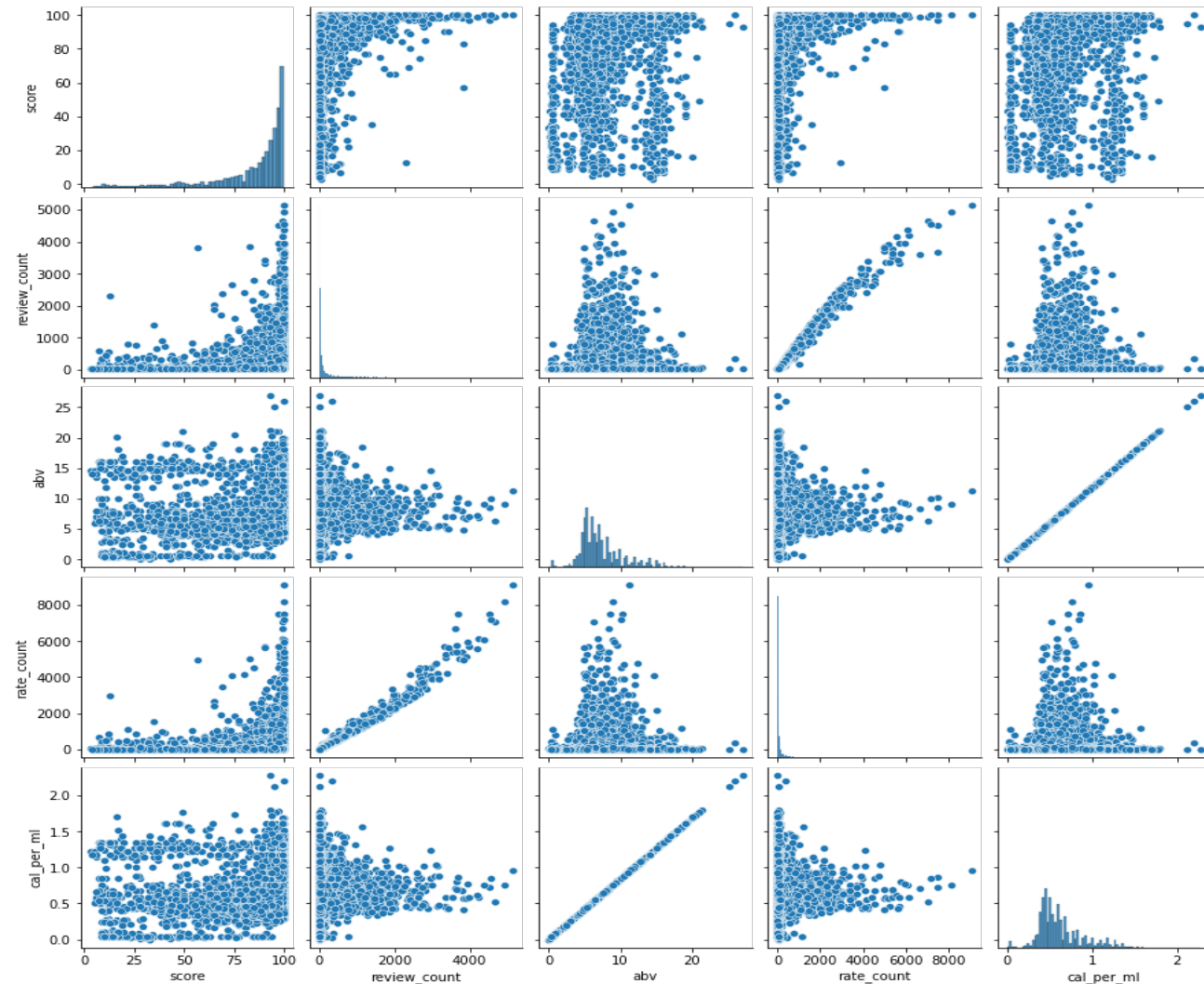
Adding Nonlinear Features and Finalize Model

- 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.