Dark Chocolate Bars

Outline

Data
Problem
Methods
Results
Conclusions
Improvements

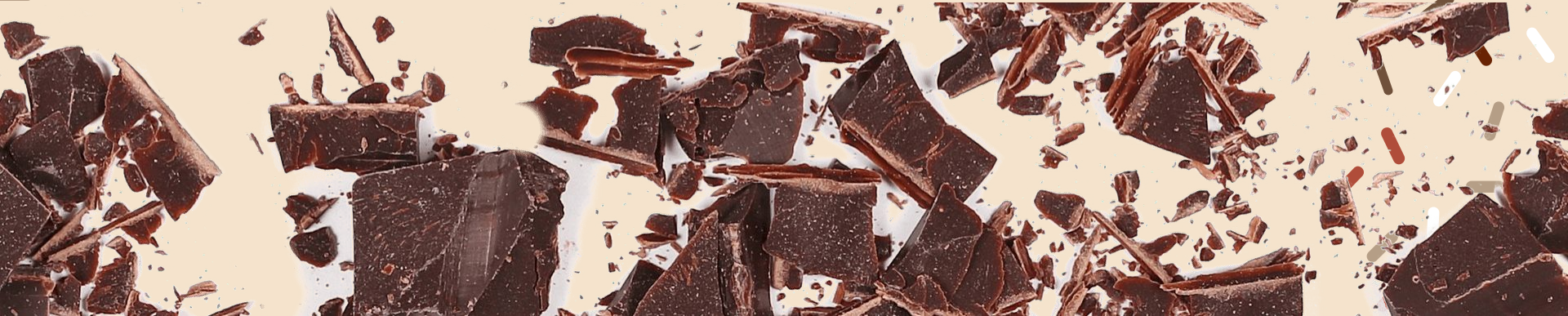# Data

- Dark Chocolate Review (2006-2021)
- 10 Variables
- Decrease in Chocolate Reviews

Problem

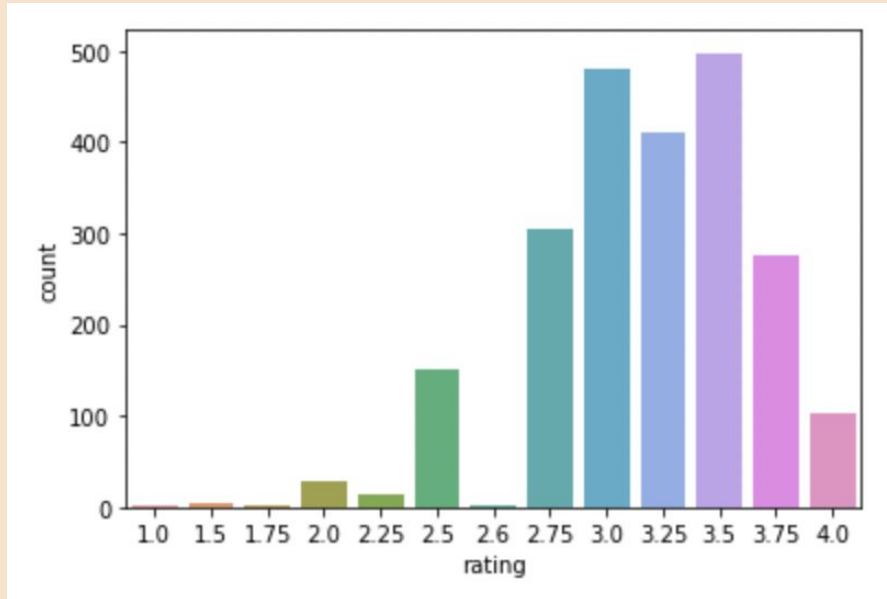What features best predict dark chocolate rating?

Methods

- Clean and Filter Collected Data
- Analyze Data
- Narrowed Down Variables
- Visualize Data
- Summarize Data
- Altered Data to get Best Model

Results

Counts of Rating

Rating Scale

4.0 - 5.0   = Outstanding
3.5 - 3.9   = Highly Recommended
3.0 - 3.49 = Recommended
2.0 - 2.9   = Disappointing
1.0 - 1.9   = Unpleasant

Results
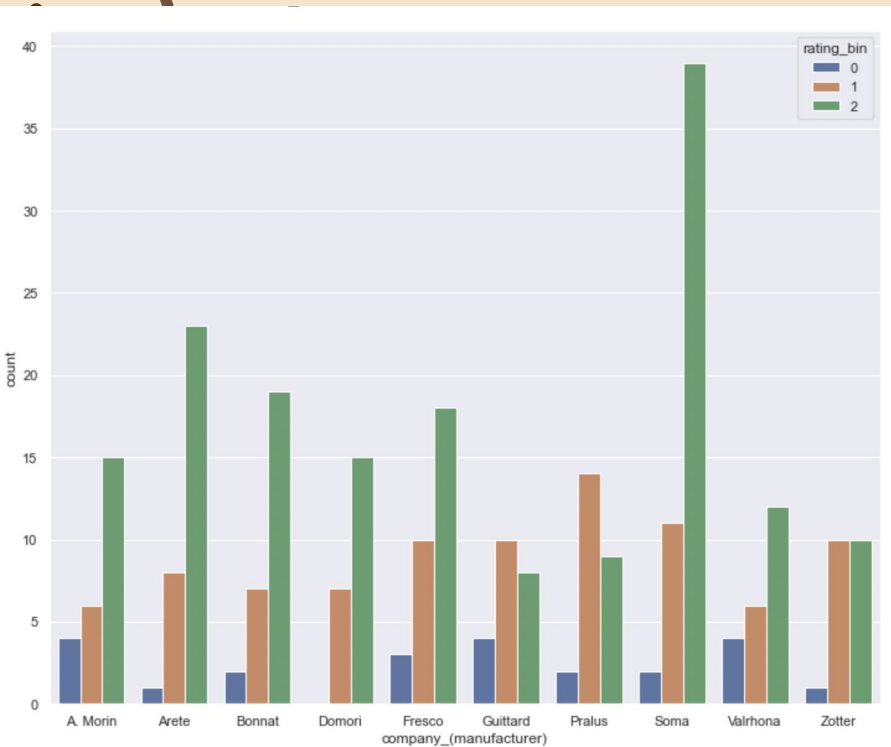
Manufacturer Rating Count

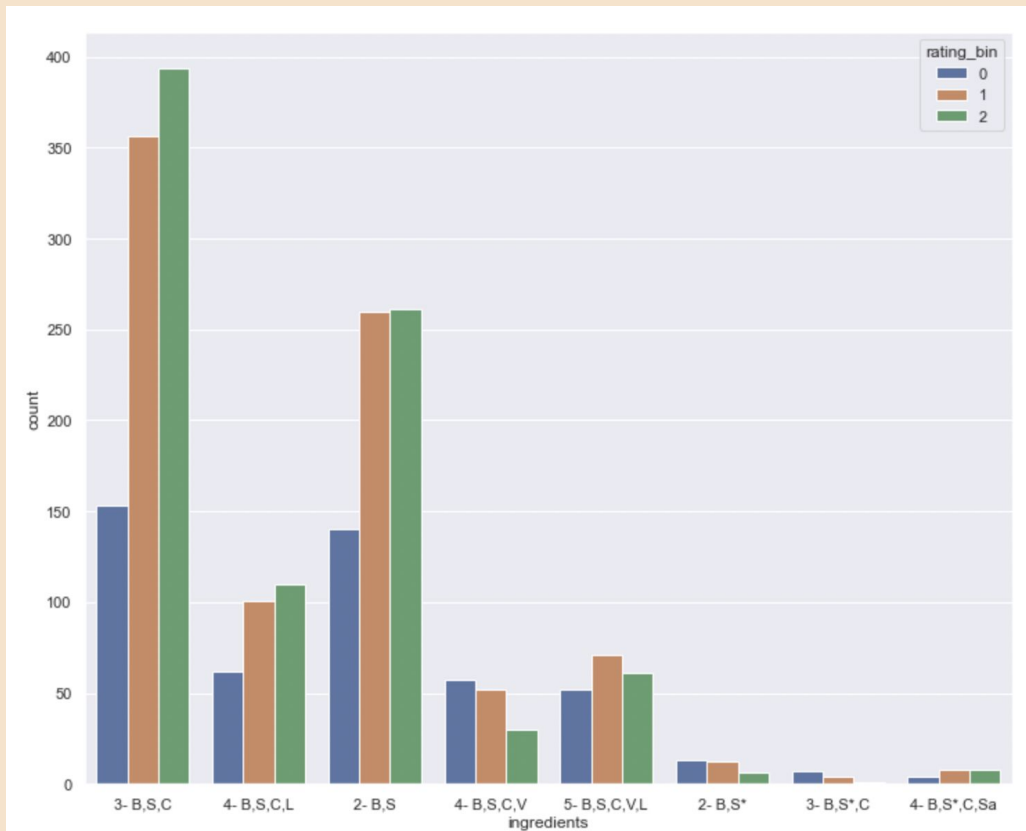## Bean Origin Rating Count
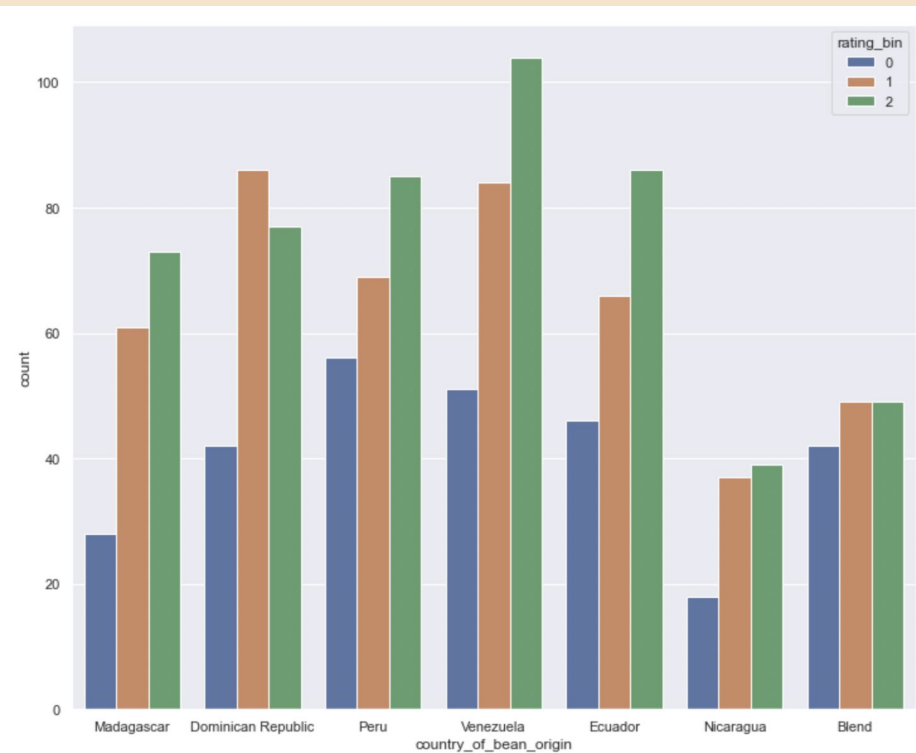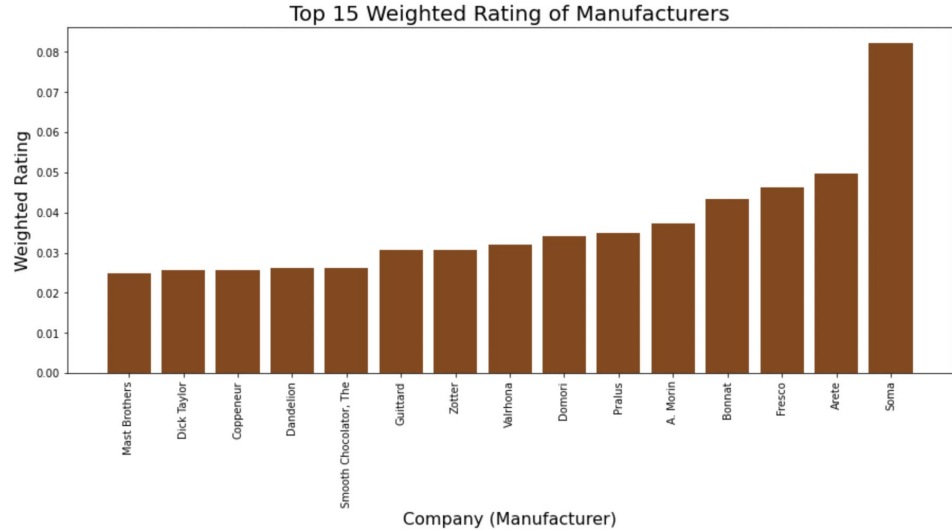
Ingredients Rating Count

Rating Scale

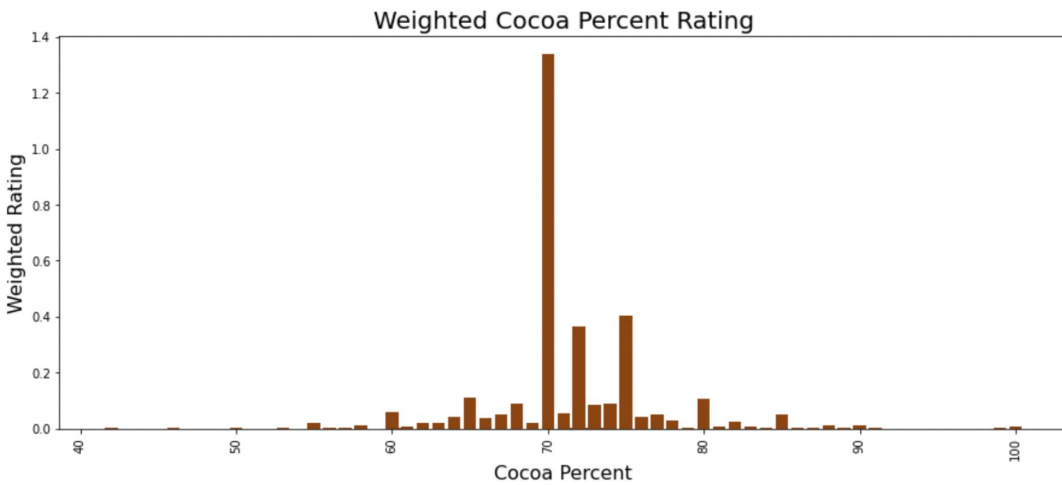0 = Disappointing & Unpleasant
1 = Recommended ←··←··←··←··↑
2 = Outstanding and Highly ↑

Boxplot, Rating by Company location

Top 15 Weighted Rating of Manufacturers

Top 15 Weighted Chocolate Company Location Rating

Top 15 Weighted Chocolate Bean of Origin Rating

Weighted Cocoa Percent Rating
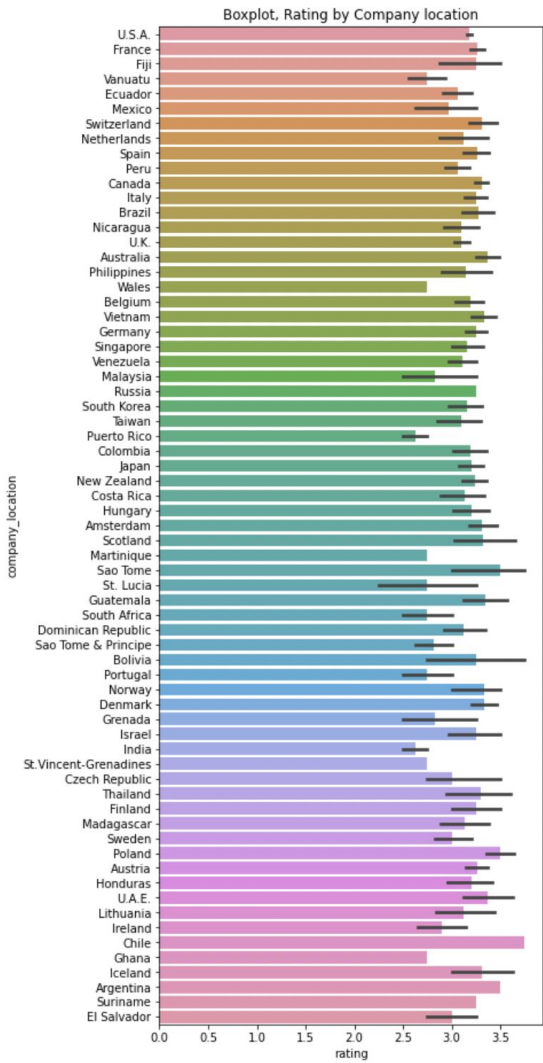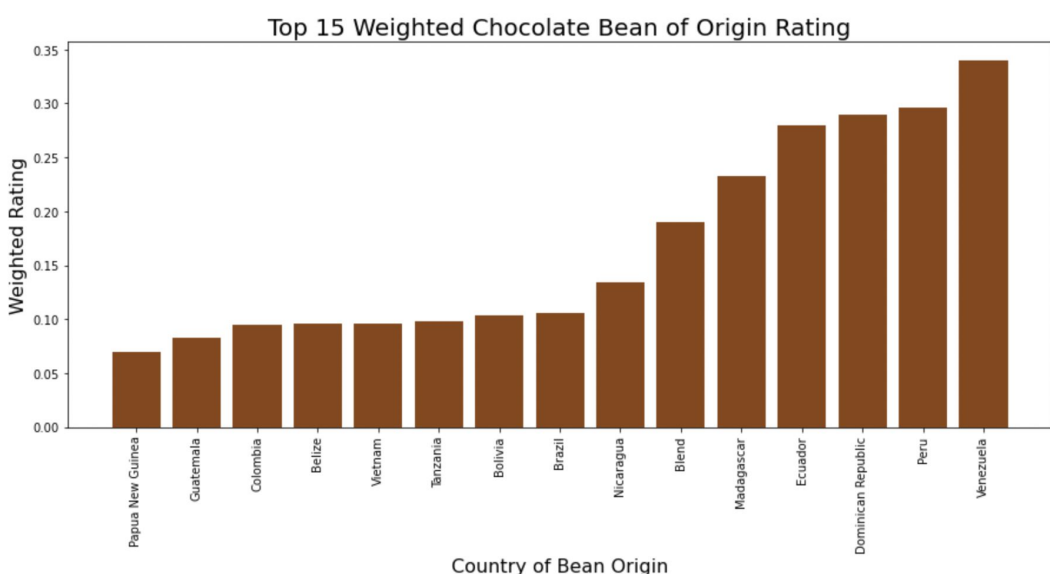
Results

First Model

```
Training accuracy:  0.7879499217527387
Validation accuracy:  0.6018735362997658
```

```
1  accuracy_score(y_test, test_preds)
```

```
0.6221441124780316
```

Drop:
Columns:
        S.B.O.B.N., Char. , & ref #

Second Model

```
Training accuracy:  0.8380281690140845
Validation accuracy:  0.7728337236533958
```

```
accuracy_score(y_test, test_preds)
```

```
0.7803163444639719
```

Drop:
Columns:
          S.B.O.B.N. & Char.

Best Model

```
Training accuracy:   0.8227016885553471
Validation accuracy:    0.8033707865168539
```

```
accuracy_score(y_test, test_preds)
```

```
0.8210526315789474
```

Drop:
Columns:
        S.B.O.B.N., Char. , & ref #
VC:

        Company>1
        Com_loc >10
        Bean_ori>10
        Ingredients>10

Word Cloud

Conclusions

- Model accuracy 80%
- Best in Prediction of Plain Dark Chocolate Bars
- Our recommendation:
  - Manufacturer: Soma
  - Bean of Origin: Venezuela
  - Cocoa Percent: 70%
  - Ingredients: Bean, Sugar, Cocoa Butter
  - "Cocoa, Creamy, Nutty"

Improvements/ Next Step

- Web scrape more data
- Try to incorporate characteristics
- Do more feature engineering
- Understand the untouched column

Thank You

For any additional questions, please contact us via email.

Email:
jrjohnso@bsc.edu
qsou@bsc.edu
jtjones1@bsc.edu