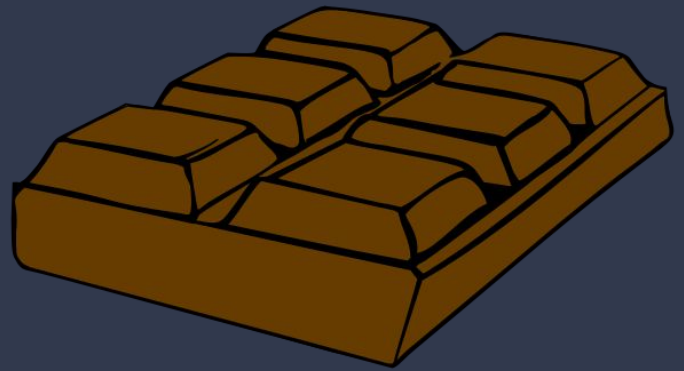
Chocolate Rating Analysis

Authors:
Dorothy Alexander, James Chandler O'Neal,
Mallory Wilson

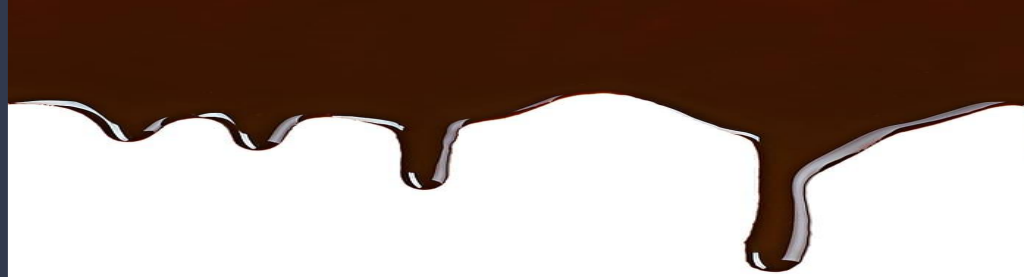


Business Problem

Hu Kitchen is creating a new chocolate bar and looking to determine what factors play a role in producing the highest rated chocolate bars on the market.

In order to determine the reasoning behind these outcomes, this project has used logistic regression, decision tree classifiers, and other models to predict the most impactful features for chocolate bar ratings.

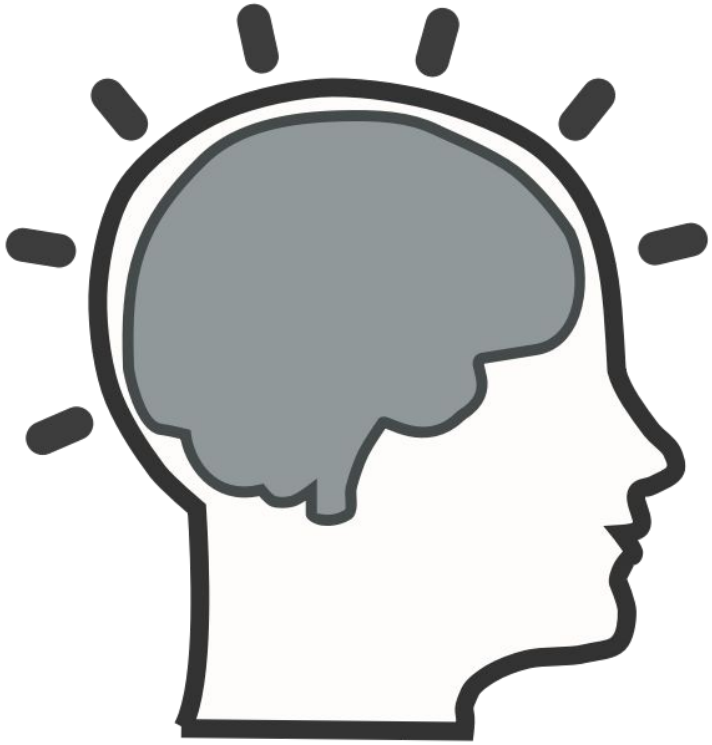
Overview



- Hu kitchen sought to determine how to increase their dark chocolate bar sales.
- Focus: To analyze and determine the features that are most impactful towards the rating of chocolate bars.
- Observed Flavors of Cacao Dark Chocolate Dataset.
- Dataset information: company location, bean origin, cocoa percent, ingredients, memorable characteristics, and rating.
- This analysis could be useful for Hu kitchen to determine the best .

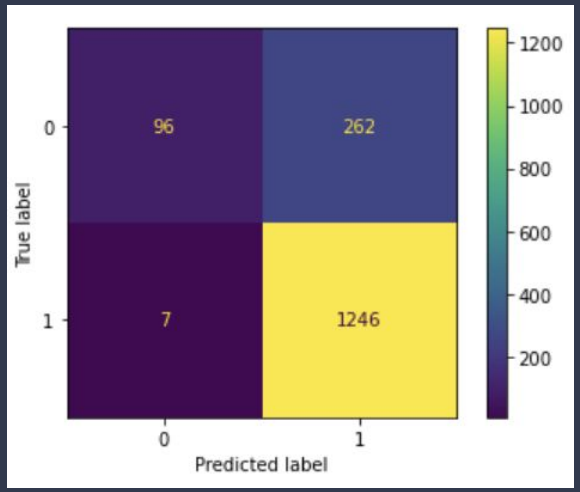
Data Understanding

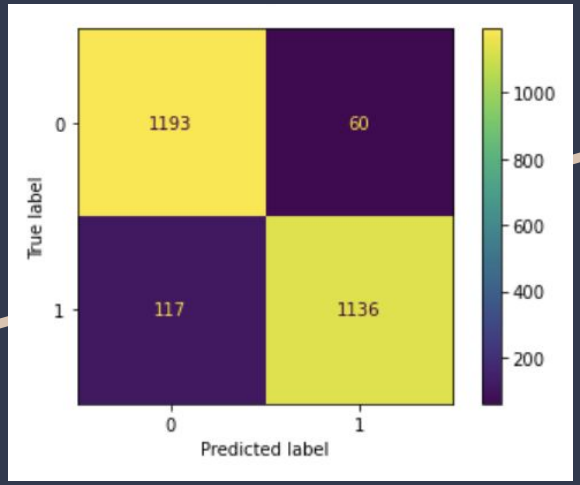
- Flavors of Cacao dataset
  - Target: Rating
    - Features
  - One Hot Encoding



Methods

- Exploratory Data Analysis
- Feature Engineering
- Baseline Model - Logistic Regression
- Decision Tree Classifier
- Grid Search
- Hyperparameter tuning
- Random Forest Classifier
- Pipeline Model





Model 1: Logistic Regression

Input:

- Parameters passed:
  - Default

Output:

- Training score:
  - 0.8571428571428571
- Validation score:
  - 0.7853658536585366

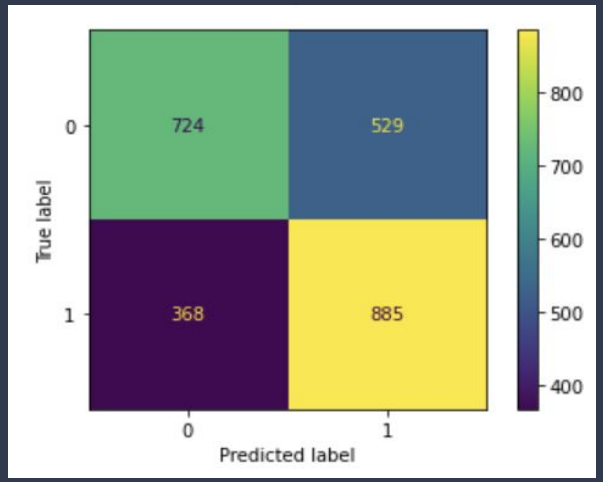
Model 2: Logistic Regression With Pipeline

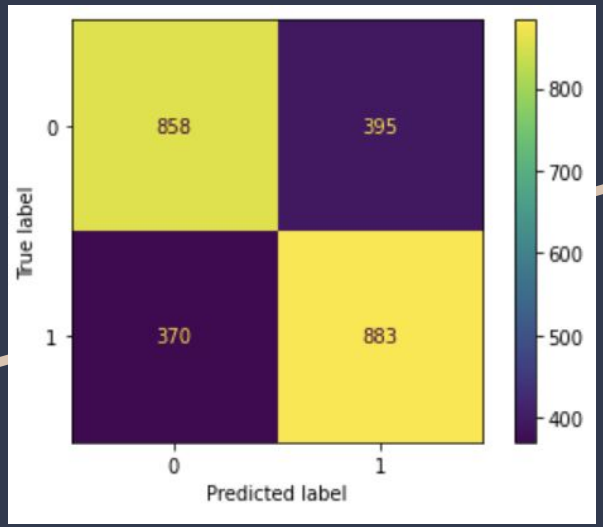
Input:

- Parameters passed:
  - C=1, max_iter=1000 penalty='l2'
  - solver= 'newton-cg'

Output:

- Training score:
  - 0.9565922920892495
- Validation score:
  - 0.8356164383561644





Model 3:
SMOTE DecisionTreeClassifier

Input:

- Parameters passed:
  - default

Output:

- Training score:
  - 0.6571207430340558
- Validation score:
  - 0.7244897959183674

Model 4:
DecisionTreeClassifier With Pipeline

Input:

- Parameters passed:
  - criterion': 'entropy'
  - max_depth': 5
  - max_leaf_nodes': 14

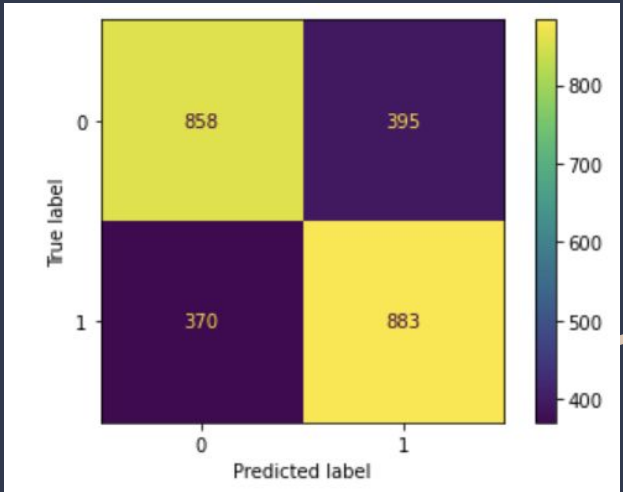
Output:

- Training score:
  - 0.8780316344463973
- Validation score:
  - 0.8818897637795275

Final Model

Pipelining a Decision Tree Classifier

Using a GridSearchCV



SMOTE
DecisionTreeClassifier

Input:

- decisiontreeclassifier_
_criterion': 'entropy',
- 'decisiontreeclassifier_
_max_depth': 6,
- 'decisiontreeclassifier_
_max_leaf_nodes': 14

Output:

- Test score:
  - 0.8754208754208753

Feature Importance

1. **Company Manufacturer:** Bin D
2. **Ingredient Combination:** B, S, C
3. **Company Manufacturer:** Bin C
4. **Specific Bean Origin**: High Weighted Ratings
5. **Ingredient Combination:** B, S, C, V, L

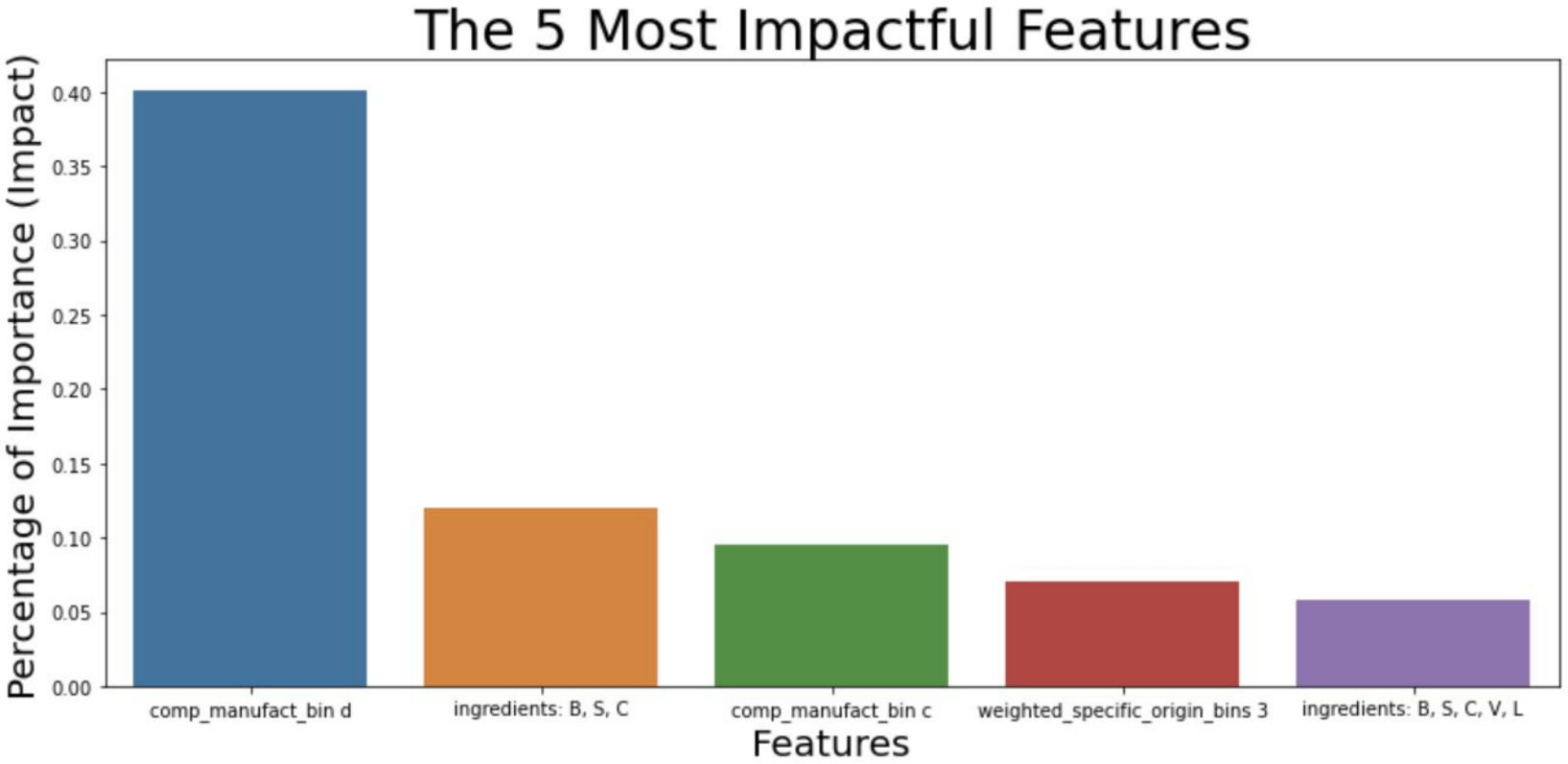
Key:

* B=bean, S=sugar ,C=cocoa butter, V=vanilla, L=lecithin

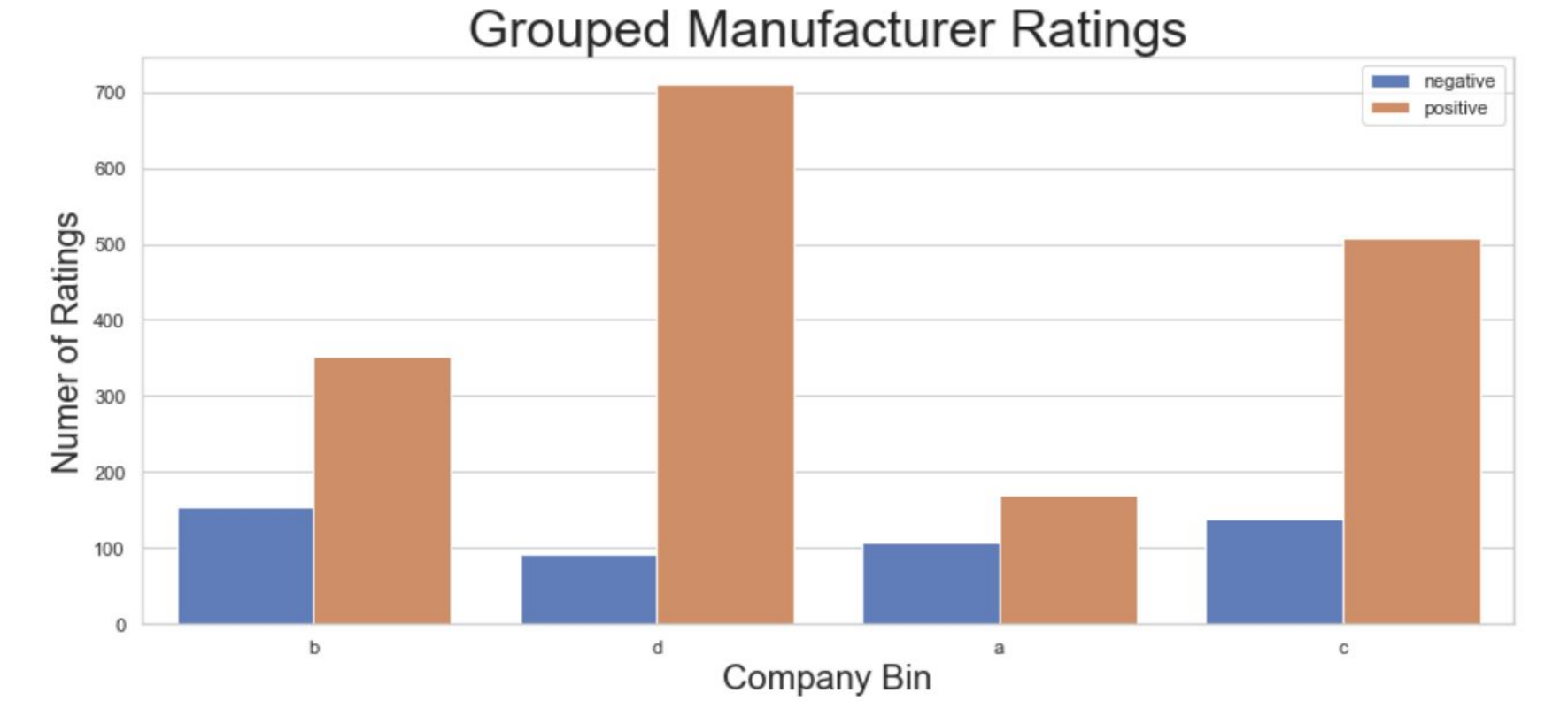
*Bin D: companies with the greatest number of positive ratings

*Bin C: Second group of companies with the greatest number of positive ratings

Top 5 Feature Importance

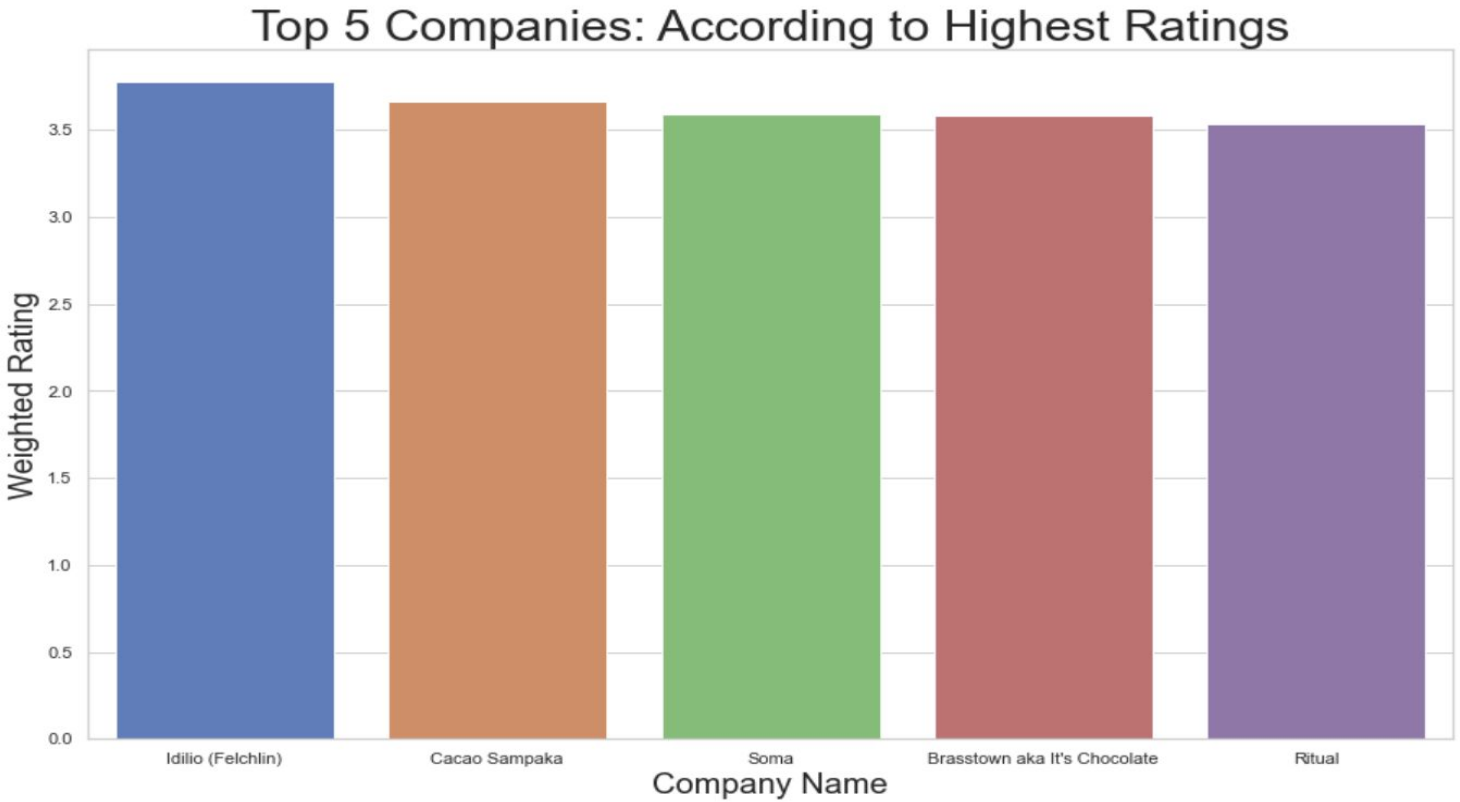


Company Manufacturer

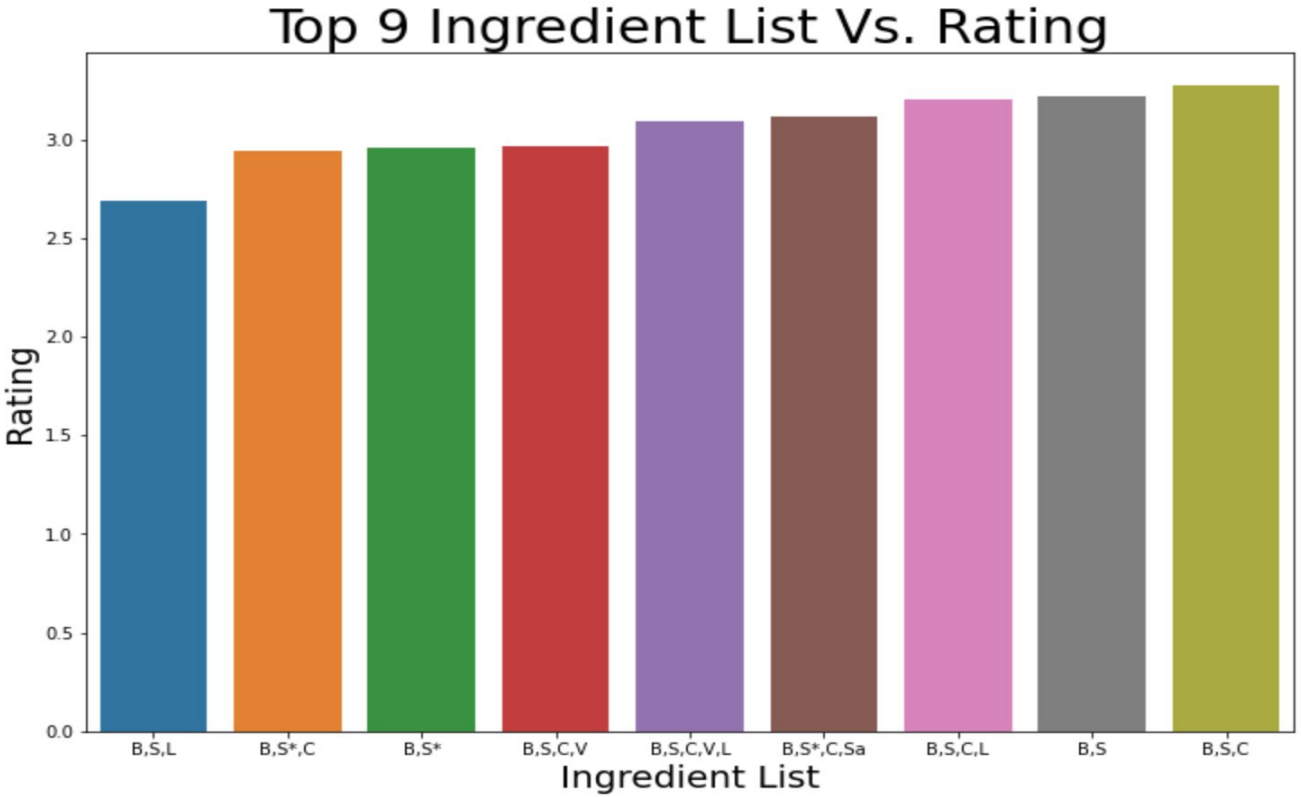


Company Manufacturer

From Bin D (top 5)



Ingredients



Conclusions

- Suggestions to Hu kitchen:
    - To create a chocolate bar which includes one of the following **ingredient combinations**: B,S,C or B,S,C,V,L.
    - **Partner** with one of the following businesses: Idilo, Cacao S., Soma, Brasstown A.I.C, or Ritual.

Next Steps

- This database only contained information on dark chocolate so looking into other kinds of chocolates such as milk and white chocolate.

- Creating a GUI (Graphical user interface) would be helpful for companies to be able to input their features they are looking at and the GUI would produce a prediction of what their chocolate rating as an output. This would help companies produce the best chocolate they possibly could.

For Any Additional Information:

Please contact us via email or Github:

Mallory Wilson :

Email - mallorye1103@gmail.com

Github - malloryewilson

Dorothy Alexander :

Email - dorothy408@gmail.com

Github - dcalexan

James Chandler O'Neal :

Email - jchandleroneal@gmail.com

Github - jchandleroneal