

Team CatDog: All About Chocolate

An Executive Summary

Motivation and Questions

Chocolate has always been associated with joy and love across the world as one of the most popular food types. Indeed, the global chocolate consumption is about 7.2 million metric tons per year: that's 2.2 pounds of chocolate for everyone! Determining which factors may have a positive impact on the quality of a chocolate bar is crucial to chocolate manufacturers; as consumers, we can find the best-tasting chocolate in the store with these features in mind. We aim to answer the following questions:

Question 1: Which factors are the most important to the manufacturers?

Question 2: Given the information of a chocolate bar, what is the prediction of its rating?

Data

The information of each chocolate includes the company information, chocolate bean information, the (percentage) ingredients and its taste profile. Our models are fitted using the *ratings*. The rating of a chocolate bar is a score given by a person who tries this bar, and it reflects to what extent this person likes it. We see ratings as an indicator of popularity.

Value

Our study offers the following insights. Chocolate bars with added sugar tend to have higher ratings. Cocoa beans from certain Asian and African regions were rated higher. For chocolate manufactures, cocoa beans from certain places and various flavorings could raise the popularity and therefore higher overall sales. The rating is also an indicator of which chocolate a company should boost production. Companies can send out surveys with chocolate samples before launching it. Based on the ratings received, they can decide the production scale more precisely before moving to mass production.

Take-aways

- According to our models, the taste profile is the most important feature. Particular flavors such as different fruit flavors, vanilla, creamy, etc. have the greatest impact on ratings.
- The cocoa percent of a chocolate bar is positively correlated to its rating.
- Cocoa beans and companies from Africa and Asia give better ratings.
- One of our models, XGBoost, has the best performance in predicting ratings.