

# Chocolate Bar Analysis

**Source of Data:** The data is available on [Kaggle](#) and downloaded from [this link](#).

**Data Collection:** The data was collected from [Flavors of Cacao](#) and is narrowly focused on plain dark chocolate.

**Data Content:** The data set contains expert ratings of over 1,700 individual chocolate bars, along with information on their regional origin, percentage of cocoa, the variety of chocolate bean used and where the beans were grown.

Flavors of Cacao Rating System:

- 5= Elite (Transcending beyond the ordinary limits)
- 4= Premium (Superior flavor development, character, and style)
- 3= Satisfactory (3.0) to praiseworthy (3.75) (well made with special qualities)
- 2= Disappointing (Passable but contains at least one significant flaw)
- 1= Unpleasant (mostly unpalatable)

**Data Profile:**

- The data set contains 1795 rows and 9 columns. After data cleaning, it has 8 columns.
- Since there are 888 null values, which is nearly half of the data, in the bean\_type column, this column has been dropped.
- There is no duplicate value.

**Limitation and Ethics:**

- Limitation: The data set does not include all chocolate bars, but only a selected subset and the ratings reflect a subjective opinion.
- Ethical issue: The data contains no PII values or discriminative variable.

**Questions to Explore:**

1. Where are the best cocoa beans grown?
2. Which countries produce the highest-rated bars?
3. What's the relationship between cocoa solids percentage and rating?
4. Which company's chocolates are of better quality?
5. Which cocoa beans are the highest-rated bars made from?

Index	Column	Description	Time Variant/Invariant	Data type
1	company_name	Name of the company manufacturing the bar	Invariant	Qualitative
2	specific_bean_origin	The specific geo-region of origin for the bar	Invariant	Qualitative
3	ref	A value linked to when the review was entered in the database. Higher = more recent	Invariant	Quantitative
4	review_date	Date of publication of the review	Variant	Quantitative
5	cacao_percentage	Cocoa percentage of the chocolate bar being reviewed	Invariant	Quantitative
6	company_location	Manufacturer base country	Invariant	Qualitative
7	rating	Expert rating for the bar	Invariant	Quantitative
8	broad_bean_origin	The broad geo-region of origin for the bean	Invariant	Qualitative