

Project description including dataset.

Link to dataset that I used <https://www.kaggle.com/datasets/rtatman/chocolate-bar-ratings/data>

Summary of data source

In this dataset, you can find expert evaluations for more than 1,700 distinct types of chocolate bars. It includes information on their geographical origins, cocoa content, types of chocolate beans utilized, and the specific locations where the beans were cultivated. Each individual chocolate bar undergoes assessment on a scale ranging from 1 to 5, indicating different tiers of flavor excellence. The ratings take into account various aspects such as the range of flavors, harmony, intensity, texture, aftermelt, and overall impressions. This dataset proves invaluable for diving into the diverse flavors and attributes of different chocolate bars, discovering regions renowned for cultivating superior cocoa beans, pinpointing nations known for producing top-rated bars, and investigating the correlation between cocoa content percentage and ratings.

Data Collection

A comprehensive evaluation was conducted on more than 1,700 distinct chocolate bars by experts. The assessment included intricate details regarding their geographical source, cocoa content, types of cocoa beans utilized, and the regions where the beans were cultivated. These evaluations were meticulously carried out by Brady Brelinski, one of the founding members of the Manhattan Chocolate Society. The ratings were determined through a thorough examination of taste, consistency, aftermelt, and general impressions.

Data collection checks

I have decided to keep the ^ value I tried to look up what ^ was on google and nothing came up. It is in the bean type columns and scattered throughout the other columns. In the Company(Maker if known) word had ^ in front of it as shown with original vs cleaned data set it seemed it was a filler letter. Other than that there weren't that many errors or special characters scattered around like that. It might've been a filler letter that was mistakenly put there.

Data Limitations

This dataset unveils insights about the flavors and qualities of over 1,700 variants. However, it's essential to acknowledge its limitations. The dataset primarily focuses on plain dark chocolate, which might not capture the entire array of chocolate types out there. The ratings reflect personal experiences rather than universal standards. Aspects like health benefits, social missions, or organic status are not factored in the dataset, potentially impacting consumer preferences. Despite these boundaries, this dataset serves as a unique tool for investigating cacao flavors and grasping what influences chocolate quality.

Data Ethics

When it comes down to handling the chocolate dataset ethically, there are several key factors to consider. Transparency, fairness, informed consent, data privacy, accountability, integrity, responsible use, and social impact all play a crucial role in this process.

Transparency is essential as it involves providing clear explanations about how data is collected, processed, and used while also addressing any potential biases that may exist. Fairness and mitigating bias are important to ensure that ratings are not unduly influenced.

Obtaining informed consent from individuals and safeguarding their data privacy are fundamental steps in this ethical journey. Accountability is necessary to hold both individuals and organizations responsible for their actions.

Maintaining integrity in handling the data guarantees its accuracy and reliability. Responsible use aims to minimize any potential harm that could arise from working with the dataset.

By upholding these principles persistently, one can promote fairness, transparency, and privacy protection when dealing with the chocolate dataset.

Key Questions

1. What is the distribution of chocolate ratings across different cocoa percentages?
2. How does the regional origin of chocolate relate to its flavor ratings?
3. Are there any trends in chocolate ratings based on the variety of chocolate bean used?
4. How does the aftermelt experience vary among chocolate bars with different flavor ratings?
6. What is the relationship between the percentage of cocoa and the flavor intensity of chocolate?
8. Are there any noticeable patterns in the flavor profiles of chocolates with elite ratings?
9. How do the ratings of chocolates from different countries compare?
10. Where are the best cocoa beans grown?
11. Which countries produce the highest-rated bars?
12. What's the relationship between cocoa solids percentage and rating?
13. What is the average rating throughout the years?