
Chocolate Bar Ratings

Analysis project



MÉLISSA TOURÉ

2023

Project overview

Chocolate holds a special place as one of the world's most beloved confections. Crafted from the seeds of the cacao tree, its production varies across different regions, resulting in a rich tapestry of flavors.

Notably, *The Manhattan Chocolate Society* has curated a comprehensive dataset, encompassing the ratings of over 2,600 chocolate bars spanning from 2006 to 2009. This invaluable resource, accessible on their website '[Flavors of Cacao](#)' serves as the foundation for our analysis. In this exploration, we will delve into the characteristics of these ratings, with a particular focus on understanding the relationship between the cacao percentage and the corresponding ratings.

Data summary

- **Data Sourcing**

The dataset was extracted from the open data source [Kaggle](#). However, originally, the data was collected on the website '[Flavors of Cacao](#)'. The 'Flavors of Cacao' dataset is a comprehensive collection of information about chocolate bars and their flavor characteristics, and was created by the Manhattan Chocolate Society. The dataset is reunited in one unique file.

- **Data Collection:**

The data was collected from different sources, including evaluations and testing groups of professionals of more than 2,600 chocolate bars over the past few years. Additionally, it includes information related to the chocolate such as its geographic origins, percentage of cacao or the ingredients, which, we can assume, generally can be found on the chocolate packaging or companies websites.

Since, the data is manually collected and created from user-generated content, we can assume that the accuracy and consistency of the data may vary. Also, we can assume that the dataset may contain errors, bias and limitations as it is based both on concrete (ex. ingredients) and subjective information (ex. reviews or rating), and research are based on a sample of a whole population, since it includes only professionals testing and excluded the rest of the U.S consumer population.

- **Data Content:**

The dataset has 10 columns : REF, Company (Manufacturer), Company Location, Review Date, Country of Bean, Origin Specific Bean, Origin or Bar Name, Cocoa Percent, Ingredients, Most Memorable Characteristics and Rating. It also encompasses 2588 rows. Regarding the timeline of the data, the information dates back from 2006 to 2022. Regarding the rating scale, it is from 1 to 5.

- **Data Limitations:**

Since the dataset relies on exclusively evaluations by professionals, it excludes consumer perspective. Also, flavor and quality ratings are subjective and can vary among the excluded individuals, which limits the insights for the chocolate industry. Finally, tastings are made on a selection of bar chocolate coming from specific regions of the world, which excludes the other types of chocolate bars or other geographical representations.

- **Data Ethics:**

All data collected are coming from open-sources (ex. company of the chocolate bar or ingredients) or individuals who agreed to participate in the testing and study. All information about the testing events history, interview questions, tastings groups or guests are displayed on the website of the organization. However, some information provided seems not to be up-to-date.

Then, in terms of bias, there may be collection, sample, exclusion and measurement bias since all data are made by individuals and there is a subjective dimension with the reviews and rating score. As for privacy and security, it seems there are no ethical issues.

Hypothesis

The cocoa percentage is expected to have a significant impact on the chocolate rating. We anticipate that there is a correlation, meaning that as the cocoa percentage increases, the chocolate rating will also increase, or conversely, as the cocoa percentage decreases, the rating will decrease.

Business Questions

1. Which chocolate bars are the best rated ?
2. Which countries produce the best rated chocolate ?
3. Is there a relationship between the percentage of cocoa and the rating of the chocolate ?
4. Which ingredients are the best rated of the chocolate ?
5. Which origin of the beans is best rated ?
6. Which company is the most popular ?
7. What are the 'key words' in the reviews describing a chocolate of quality (high rate) ?

Deliverables

- Tableau presentation [[Here](#)]
- Python Jupyter notebook scripts