Angela Hamilton

Career Foundry Project 6

Sourcing Open Data

11.29.2023

Chocolate Bar Ratings

**Data Sourcing:**  This is an external data source from Kaggle, a subsidiary of Google. The data source is an online platform for data scientists and machine learning enthusiasts, that provides a wide collection of datasets that are available to the public.

**Data Collection**: Open-Source data that is free and available to the public. Since this dataset is considered, open data collected by third-party sources, there could be bias in the data collection. The last update for this dataset was back in 2017, which causes there to be a time lag.

**Data Content**:

The data contains ratings for the different flavors of cocoa of over 1,700 chocolate bars from 2006 – 2017. The data is categorized by cocoa bean types for various regions and contains cocoa bean origins, flavor ratings, manufacturers name, reference numbers, % of cocoa, and the year of the review.

**Data Limitations:**

The data was collected from a Public Doman with open licenses to allow creators to share their work, which presents an opportunity for bias and data inaccuracies.

**Data Relevance:**

This data set meets the data requirements for this project, as it is open source, contains a geospatial component, and it meets the size and variable requirements. This project is older than three years old but was provided by Career Foundry as an available dataset that meets the specific criteria for this achievement.

Clean your data in Python

[The **Flavors of Cacao** dataset contains **9 columns**](http://flavorsofcacao.com/chocolate_database.html)[1](http://flavorsofcacao.com/chocolate_database.html)[2](https://medium.com/@Redox1/my-first-data-analysis-project-cocoa-flavors-dataset-42fa36825927)[3](https://medium.com/@maranjevivian/my-first-excel-project-cocoa-flavor-analysis-f2909a29b40c). Here is a brief description of each column:

1. **Company (Manufacturer)**: The name of the company that produced the chocolate bar.
2. **Specific Bean Origin or Bar Name**: The specific name of the chocolate bar.
3. **REF**: A reference number that links to the rating source.
4. **Review Date**: The date the chocolate was reviewed.
5. **Cocoa Percent**: The percentage of cocoa in the chocolate bar.
6. **Company Location**: The location of the company that produced the chocolate bar.
7. [**Rating**: The expert rating of the chocolate bar on a scale of 1 to 5, where 5 is considered “Elite” 4](https://www.kaggle.com/).
8. **Bean Type**: The variety of chocolate bean used.
9. **Broad Bean Origin (Country of Bean Origin)**: The region where the beans were grown.

The **Flavors of Cacao Rating System** is a rating scale used to evaluate the quality of chocolate bars. [The rating system was developed by Brady Brelinski, a member of the Manhattan Chocolate Society, and is based on a combination of objective and subjective criteria 1](http://flavorsofcacao.com/review_guide.html). The rating scale ranges from 1 to 5, with 5 being the highest rating. Here is a brief description of the rating scale:

* **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)