Chocolate Bar Ratings

03/31/2023

Titi Adigun

Data Analytics

Summary.

Chocolate, which is one of the most popular candies in the world, does not have all the bars created equally. Over 1700 individual chocolate bars were rated along with data on their regional origin, percentage of cocoa, the variety of chocolate bean used and where the beans were grown.

Data Source.

This data is publicly available (open source) and can be downloaded from [Chocolate Bar Ratings](#).

Data cleaning and consistency checks.

* The data set had 2 blank spaces in columns 'Bean Type' and 'Broad Bean Origin', which were removed.

* In column 'Company Location', the countries U.K. and U.S.A. were renamed as 'United Kingdom' and 'United States.

* In column 'Broad Bean Origin', some data has more than country and were reduced to one each.

* Checked for duplicate data and for mixed data.

* Renamed the column headings properly.

* Renamed '\xa0' as 'xao' in 'Bean Type' column.

* Converted the date to object.

Data profile.

Number of rows and columns in the cleaned dataset: 1793, 9

(Original data: 1795, 9)

Limitation and Ethics

* Some of the values in column 'Broad Bean Origin' has more than one country, I had to pick one of the countries for the analysis.

* In terms of ethics, the data does not contain any personal information.

Column details.

Column	Column Description	Time variant	Data type
CompanyÂ	Company that grew the Bean	Invariant	Qualitative, Nominal
Specific Bean Origin or Bar Name	Name of Bean	Invariant	Qualitative, Nominal
REF	Reference number	Invariant	Qualitative, Discrete
Review Date	Date of last review	variant	Qualitative, Discrete
Cocoa Percent	Percentage of cocoa in each bean	Invariant	Quantitative, Continuous
Company Location	Location of chocolate manufacturer	Invariant	Qualitative, Nominal
Rating	Each chocolate rating	Invariant	Quantitative, Continuous
Bean Type	Type of each bean	Invariant	Qualitative, Ordinal
Broad Bean Origin	Country of original of bean	Invariant	Qualitative, Ordinal

Quantitative variables descriptive statistics.

	REF	Review\Date	Rating
Count	1795.000000	1795.000000	1795.000000
Mean	1035.904735	2012.325348	3.185933
Std	552.886365	2.927210	0.478062
Min	5.000000	2006.000000	1.000000
25%	576.000000	2010.000000	2.875000
50%	1069.000000	2013.000000	3.250000
75%	1502.000000	2015.000000	3.500000
Max	1952.000000	2017.000000	5.000000

	Cocoa\Percent
Count	1793
Min	42%
Max	100%
Mode	70%
Median	70%

Questions to explore.

- What are the top Bean Company Maker with the highest rating and top 5 Company Maker with the least rating of Bean?
- What are the top Broad Bean Origin with the highest rating and top 5 Broad Bean Origin with the least rating of cacao flavors?
- Is the Bean quality dependent on the Company Maker, Bean Type or Broad Bean Origin?
- Is the chocolate rating dependant of the percentage of bean in the chocolate?