EXPLORATORY DATA ANALYSIS (EDA)

CHOCOLATE BAR RATINGS

QUESTION 1

GROUP 5

NAMES OF GROUP MEMBERS

PIUS SARFO BUOBU

MUMUNI SAMDEEN

GBODZI SELASI YAW

ANDREWS ACHEAMPONG

OUTLINE

Understanding the business

Understanding the data

Cleaning the data

Questions

Analysis and Presentation of findings

1. Understanding the business

Upon taking the various parameters of the given dataset into consideration, we realized that the dataset is owned by the Manhattan Chocolate Society.

The Manhattan Chocolate Society, through research and tastings of fine chocolate, examines what is responsible for specific characteristics in chocolate whether they are due to specific growing regions, cacao genetics, manufacturer, or some other cause.

The principal objective of the Society is to taste, compare and discuss the qualities of fine chocolate.

2.0 Understanding the data

After critical observation of the dataset we were given, we realized that there were 1795 rows and 9 columns.

The 9 columns are listed below.

Ratings of chocolate- chocolate bars were rated on a scale of 5.

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)

* Percentage of cocoa- it represents the amount of cocoa expressed in %, found in chocolate
* Company names- indicate the names of the array of chocolate manufacturing companies
* Location of company- suggests where the chocolate producing companies are situated
* Origin of cocoa bean- indicates where each of the various types of cocoa bean is emanating from
* Type of bean= suggests the specific type of cocoa bean used in producing chocolate

3.0 Questions

For us to be able to explore the dataset to draw some insights from the data, the following questions were considered.

* Which country has the highest chocolate producing companies?
* Which country/countries has/have the least number of chocolate manufacturing industries?
* What is the most occurring “ratings” of the chocolate bars?
* Find the relationship between the % of cocoa in chocolate and ratings of chocolate bars.

Analysis and Presentation of Findings

After going through the above processed, we used matlibplot.pyplot codes to visualize the data with respect to the various questions that were asked above.

* Which country has the highest chocolate producing companies?

The information gathered from our analysis indicates that the United States of America harbors majority (over 750) of the chocolate producing companies followed by France, Canada, UK and Italy in that order.

Histogram

Description automatically generated with medium confidence

* Which country/countries has/have the least number of chocolate manufacturing industries?

In trying to find out the last five countries with the least number of chocolates producing companies, we observed that Ecuador, Wales, Nicaragua, India, and Philippines had one company each.

A picture containing icon

Description automatically generated

* What is the most occurring “ratings” of the chocolate bars?

Most of the chocolate bars (about 380) were rated 3.5. Almost 350 bars were rated 3.0. a few bars were rated 5.

