The chocolate bars

Statistical test to prove the obvious
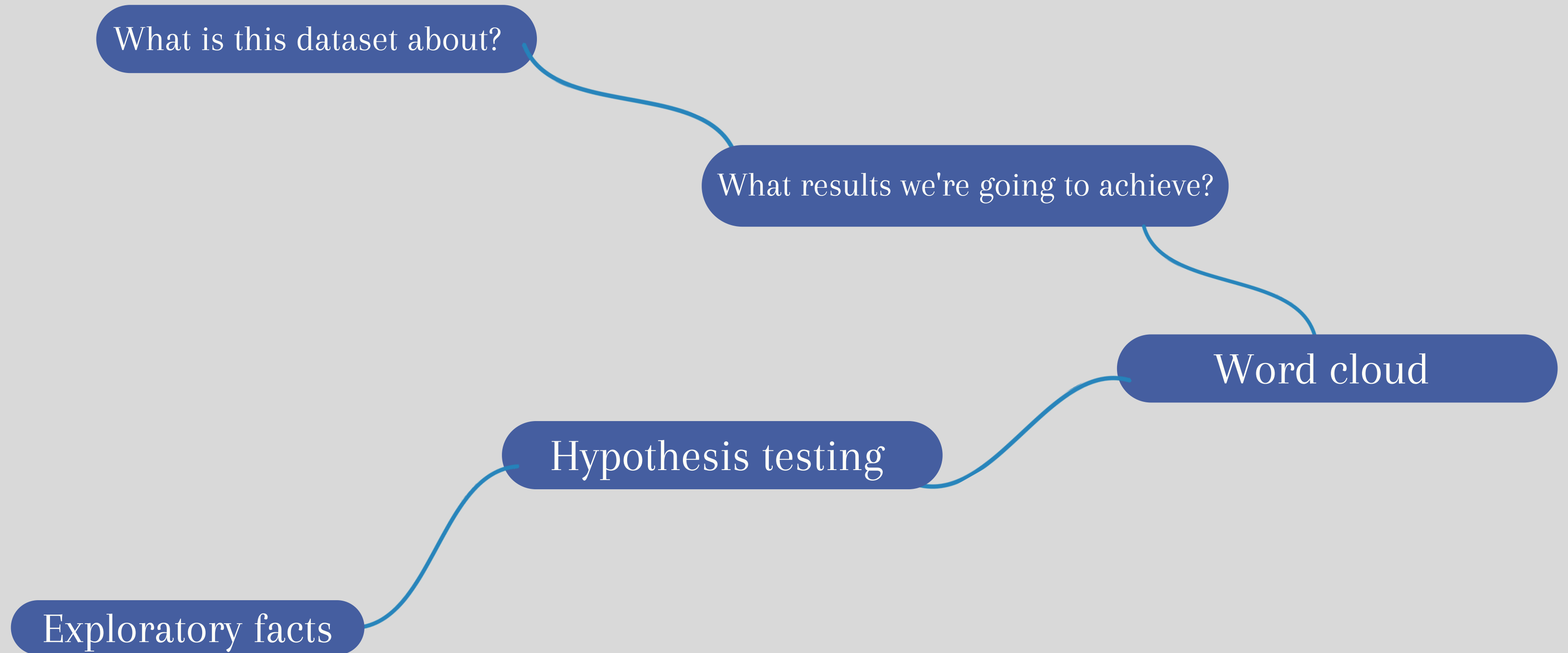
By: Magno Lira

Summary

What is this dataset about?

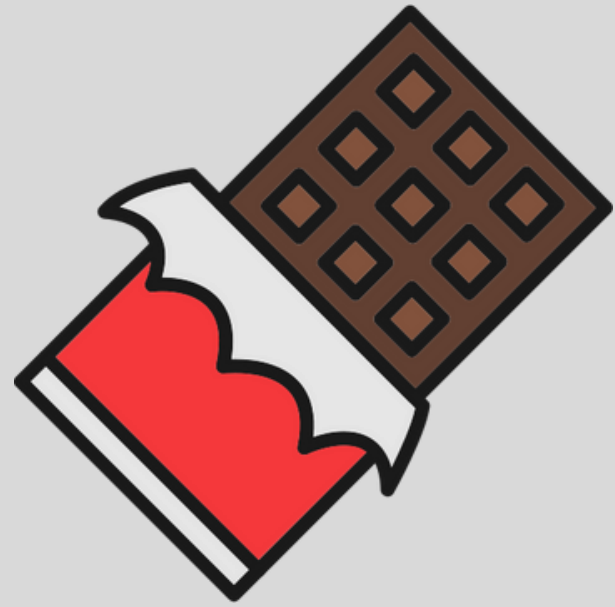What results we're going to achieve?

Word cloud

Hypothesis testing

Exploratory facts

What is this dataset about?

To begin with, the dataset is composed of a range of information about chocolate industries, such as rating, cocoa percentage, and company location. In addition, we are interested in five of the ten columns, which are:

Rating

Cocoa Percent

Ingredients

Country of bean origin

Most memorable characteristics

What results we're going to achieve?

Word cloud

A word cloud with the most memorable characteristics

Statistical test

It has proved that chocolate's rating with sugar is greater than chocolate's rating with non-sugar

Exploratory facts

the top 5 countries with best rating in chocolate production and also the number of occurrences

Chocolate with sugar and non-sugar: distribution of cocoa

The word cloud

Most memorable characteristics

Hypothesis testing

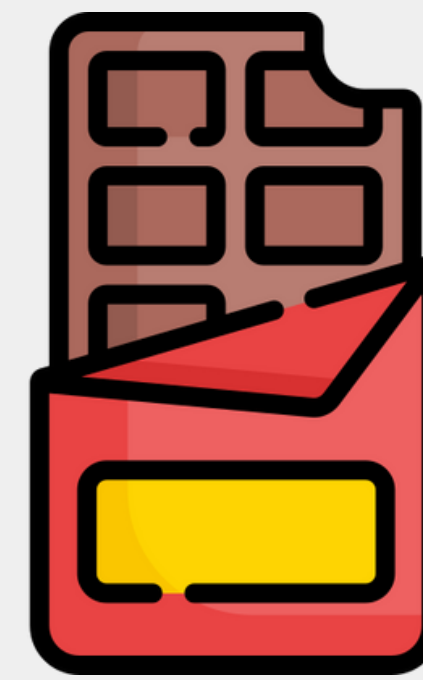It used the Mann-Whitney U test, which the null hypothesis is:

HO: Sugar chocolate's average rating=Non-sugar chocolate's average rating

The samples were the same size(=20) and were defined in an aleatory way

The statistical significance: 0.10

p_value: 0.0678

Top 5 countries with the most occurrences and highest rating in chocolate production
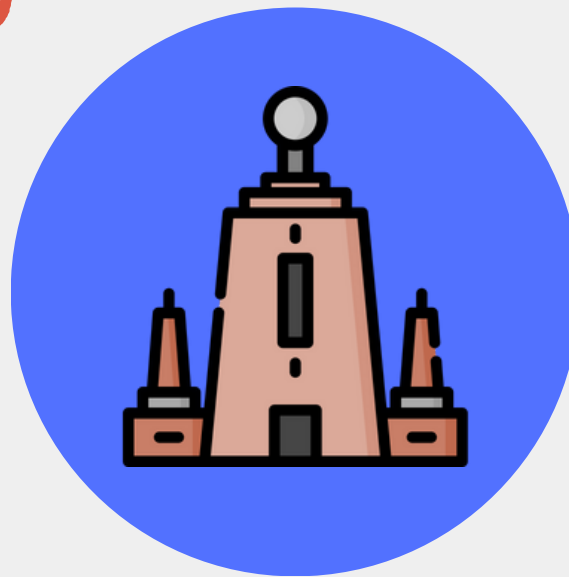
VENEZUELA

N°:20

PERU

N°:19

MADAGASCAR

N°:11

ECUADOR

N°:9

BRAZIL

N°:5

Chocolate with sugar and non-sugar: distribution of cocoa

Sugar

| 70% cocoa | 72% cocoa | 75% cocoa | 74% cocoa | 64% cocoa |

No sugar

| 70% cocoa | 75% cocoa | 72% cocoa | 80% cocoa | 65% cocoa |