Coffee Ratings Analysis

jorge gonzalez gonzalez

data analytics

2020

CONTENTS:

[**1.INTRODUCTION** 2](#_Toc57299196)

[**2.ANALYSIS** 3](#_Toc57299197)

[**3.INSIGHTS** 9](#_Toc57299198)

## **1.INTRODUCTION**

The objective of this report is to analyse the coffee features from data collected from the Coffee Quality Institute.

There is data both Arabica and Robusta beans across many countries and professionally rated on a 0-100 scale.

The ratings are based on several coffee metrics like aroma, flavour, aftertaste, acidity, body, balance, uniformity, and sweetness.

**Dataset Structure**

Full description / examples can be found at: <https://database.coffeeinstitute.org/coffee/357789/grade>

|  |  |
| --- | --- |
| **Variable** | **Description** |
| total\_cup\_points | Total rating/points (0 - 100 scale) |
| species | Species of coffee bean (arabica or robusta) |
| owner | Owner of the farm |
| country\_of\_origin | Where the bean came from |
| farm\_name | Name of the farm |
| lot\_number | Lot number of the beans tested |
| mill | Mill where the beans were processed |
| ico\_number | International Coffee Organization number |
| company | Company name |
| region | Region where bean came from |
| producer | Producer of the roasted bean |
| number\_of\_bags | Number of bags tested |
| bag\_weight | Bag weight tested |
| in\_country\_partner | Partner for the country |
| harvest\_year | When the beans were harvested (year) |
| grading\_date | When the beans were graded |
| variety | Variety of the beans |
| processing\_method | Method for processing |
| aroma | Aroma grade |
| flavour | Flavour grade |
| aftertaste | Aftertaste grade |
| acidity | Acidity grade |
| body | Body grade |
| balance | Balance grade |
| uniformity | Uniformity grade |
| clean\_cup | Clean cup grade |
| sweetness | Sweetness grade |
| cupper\_points | Cupper Points |
| moisture | Moisture Grade |
| color | Color of bean |
| expiration | Expiration date of the beans |
| certification\_body | Who certified it |
| altitude\_mean\_meters | Altitude mean meters |

## **2.ANALYSIS**

**Coffee Species**

Coffee species are divided into two main species: Arabica and Robusta. These are further divided into multiple varieties (We will look into them in the next section).

Arabica is cultivated more around the world than Robusta, because of its better flavor.

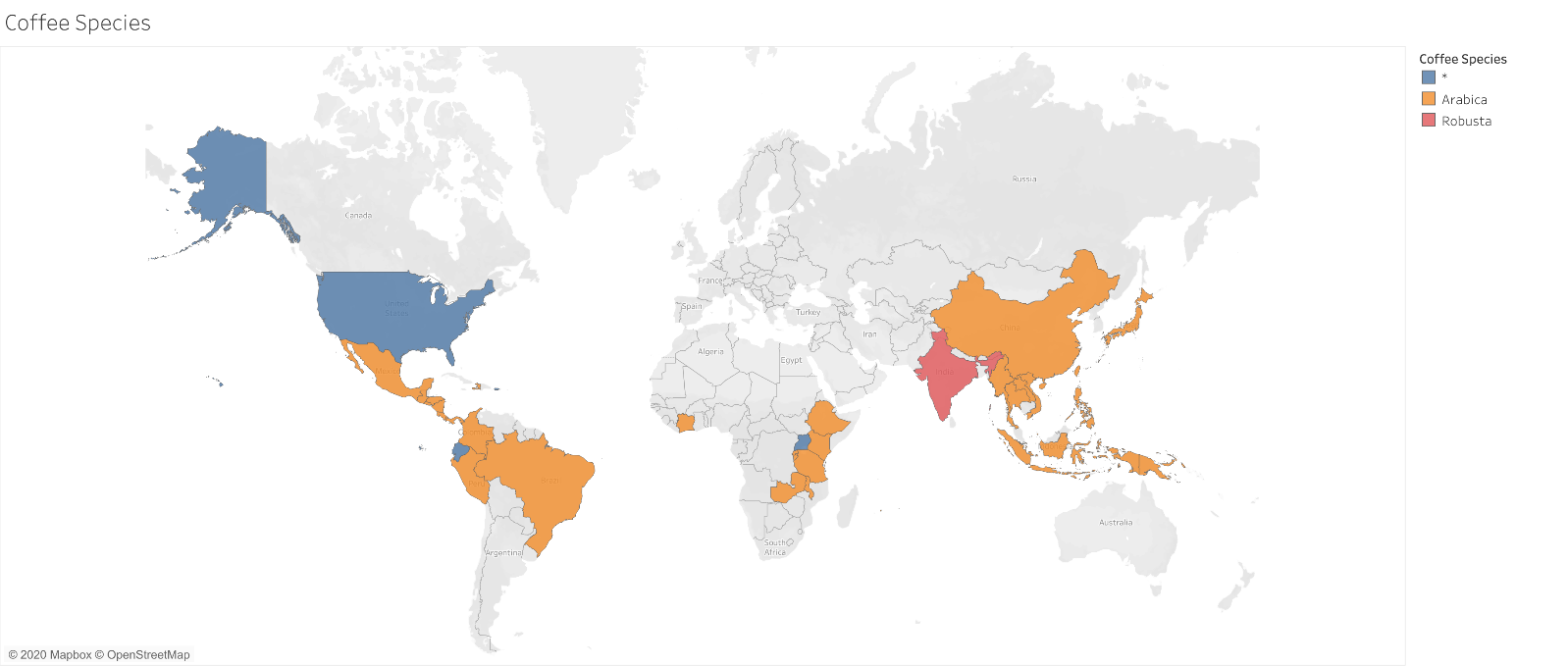
60% of the coffee produced worldwide is Arabica and 40% is Robusta.

Robusta is more productive and less susceptible to plant diseases, such as leaf rust.

However, Robusta taste is one-dimensional and bitter.

Arabica beans consist of 0.8-1.4% caffeine and Robusta beans consist of 1.7-4% caffeine.

*World Map – Coffee Species*



**Coffee Variety**

Coffee varieties are the diverse subspecies derived through selective breeding or natural selection of coffee plants.

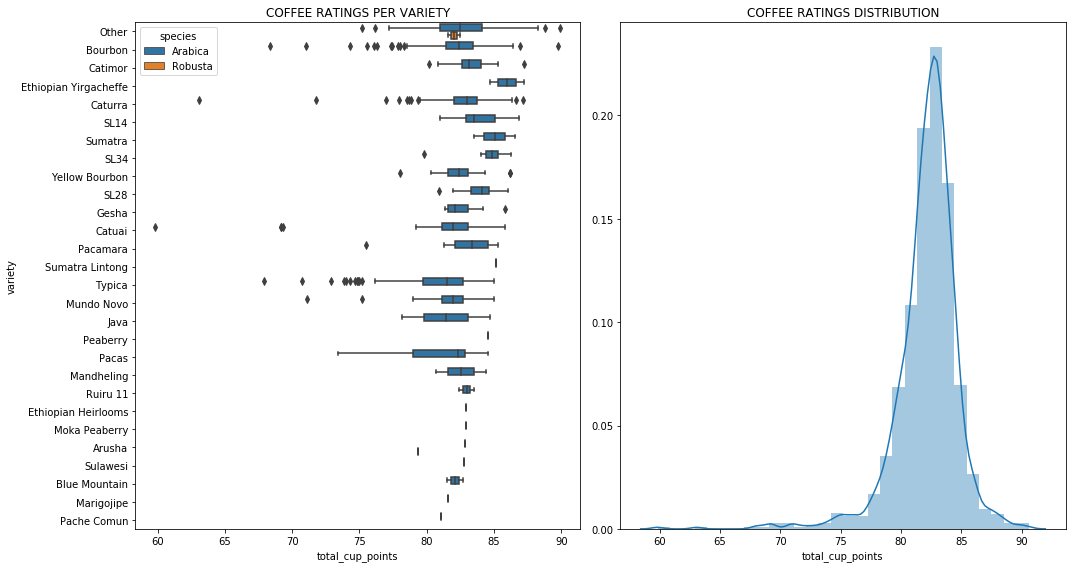
These are the coffee varieties we have in our dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| Coffee Variety | N0 of Samples | Coffee Variety | N0 of Samples |
| Caturra | 237 | Ruiru 11 | 2 |
| Typica | 208 | Mandheling | 2 |
| Bourbon | 207 | Sumatra | 2 |
| Other | 102 | Blue Mountain | 2 |
| Catuai | 69 | Arusha | 2 |
| Yellow Bourbon | 33 | Java | 2 |
| Mundo Novo | 28 | Ethiopian Yirgacheffe | 2 |
| Catimor | 20 | Sulawesi | 2 |
| SL14 | 16 | Peaberry | 1 |
| SL28 | 14 | Moka Peaberry | 1 |
| Pacas | 13 | Pache Comun | 1 |
| Gesha | 12 | Sumatra Lintong | 1 |
| Pacamara | 8 | Ethiopian Heirlooms | 1 |
| SL34 | 8 | Marigojipe | 1 |

We plot the data to understand value rating distribution, localize extreme values and understand which kind of distribution we have.

Most of the values fall in a range between 75 and 95 and we have a left – skewness distribution.

There are few outliers for Bourbon, Caturra and Typica varieties. This may happen because we have analysed more samples for these variety.

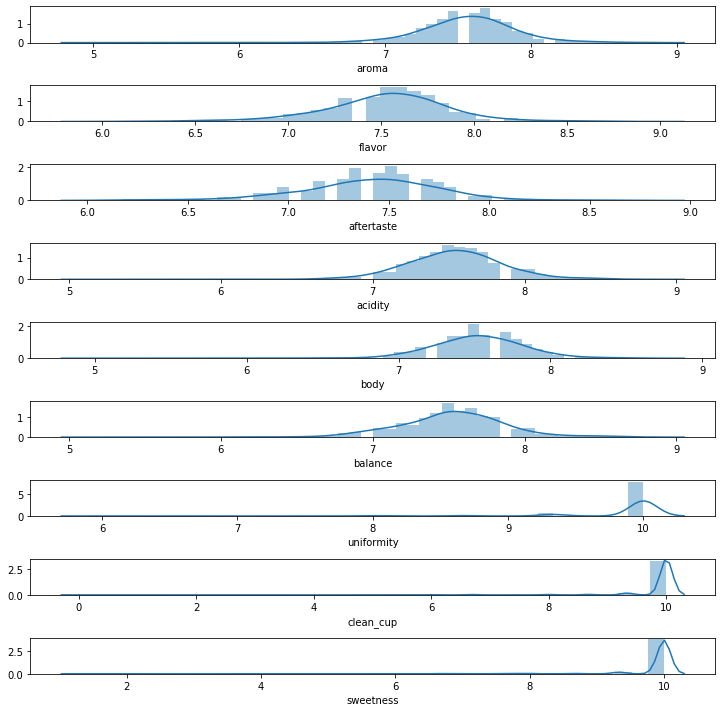


**Coffee Metrics**

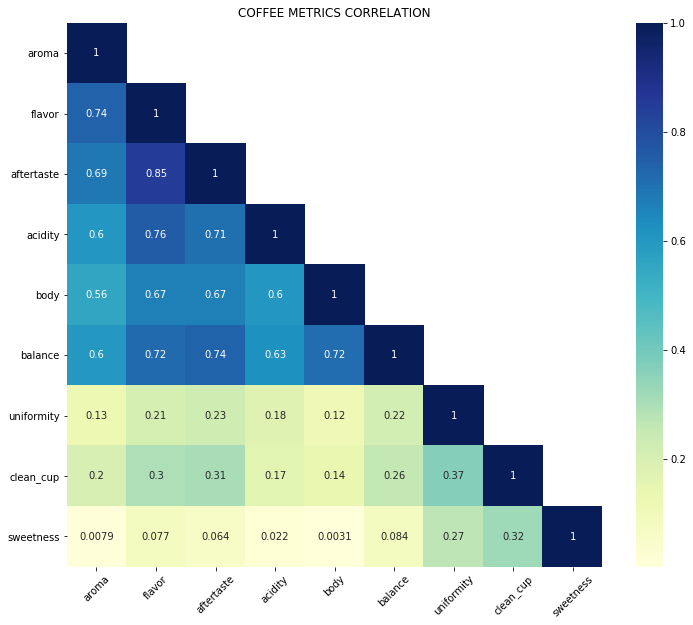
A standard coffee cupping procedure involves deeply sniffing the coffee, then loudly slurping the coffee so it spreads to the back of the tongue.

The coffee taster attempts to measure aspects of the coffee’s taste, specifically the body, sweetness, acidity, flavour and aftertaste.

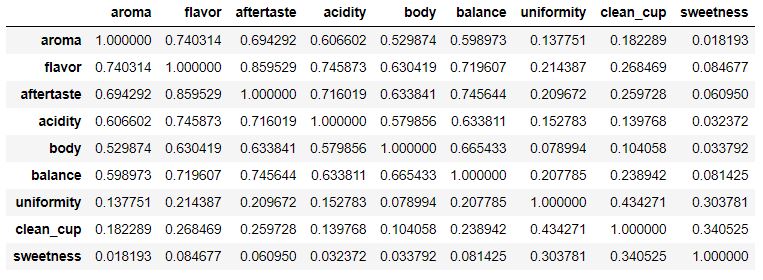
*Coffee Metrics – Distribution Plot*



*Coffee Metrics – Heatmap Correlations*

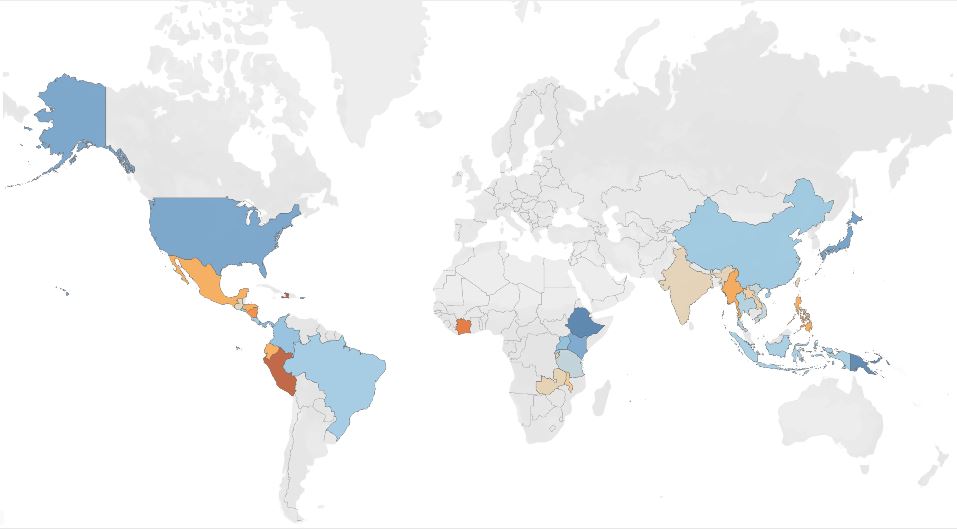
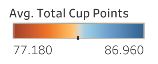
****

*Coffee Metrics – Correlation Matrix*

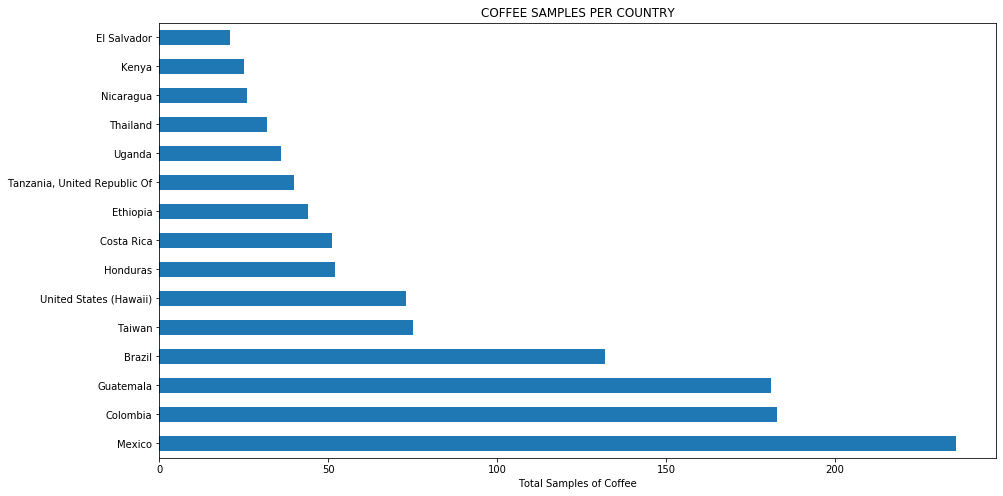
****

**Coffee Ratings per Country**

*World Map – Coffee Ratings*

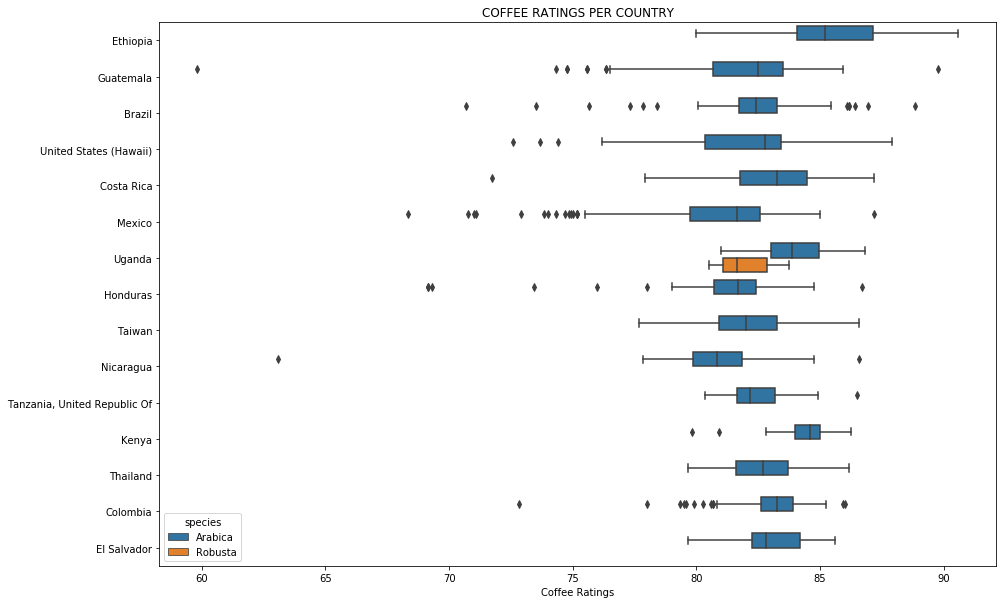
**

We will be analysing countries where we have at least 20 samples of coffee.



We plot the data to understand value rating distribution per country, localize outliers and visualize where values are concentrated.

Most of the values fall in a range between 75 and 90 and we have a left – skewness distribution.

****

## **3.INSIGHTS**

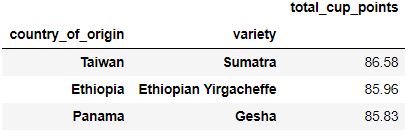
**Coffee Variety**

The best rated coffee varieties are Sumatra, Ethiopian Yirgacheffe and Gesha.

Asian countries tend to produce a few coffee varieties but they offer high quality coffee.

We have seen as well that coffee varieties from Japan and Indonesia have high ratings.

*Mean Ratings for top 3 coffee Varieties*

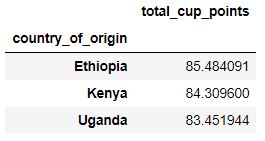


**Coffee Ratings Per Country**

The best rated countries are Ethiopia, Kenya, and Uganda.

Ethiopia and Kenya cultivate Arabica coffee while Uganda cultivate both Arabica and Robusta.

*Mean Ratings per Country*



Coffee from Africa has high ratings.

Our next steps would be to analyse if there is any factor which make coffee from these areas better rated.