# Arabica Analysis in PowerBi

## Data Pre-processing

1) Changed data type of bag weight to number after removing kg from the values

2) Removed Status column as it had only one value-Completed.

3) Removed Harvest year, Grading and Expiration date, Lot Number, In-country partner, as they don't answer the questions

4) Removed empty rows from Region, Processing method, Variety

5) Removed rows with 'Unknown' value in Variety column

6) Standardized values in Colours column

7) Removed Defects column. Created a new column called total defects which adds up the defects values in Category 1 and Category 2 defects columns

## Results of Analysis

Q1) What are the key determinants of coffee quality as evaluated through sensory attributes such as aroma, flavor, acidity, etc.?

- Based on correlations coefficients, it is evident that sensory attributes like Flavour, Aftertaste, Balance, Acidity, Aroma and Body have strong correlation with the overall quality measured by Total cup points.
- Uniformity, Clean cup and Sweetness show no correlation with Total cup points points as they remain almost constant throughout the data.

Q2) Is there a correlation between processing methods, origin regions, and coffee quality scores?

- Considering processing methods used, Double Anaerobic Washed had highest Average Total cup points while Semi-LAVADO had the least. Average Total cup points showed moderate variation with processing methods used.
- Average Total cup points did not vary much with origin country. Ethiopia had the highest Average Total cup points while El Salvador had the least.
- Map showing the top 3 regions with average Total cup points, for each country was also plotted for better visualization.

Q3) Can we identify any trends or patterns in defect occurrences and their impact on overall coffee quality?

- The Total cup points had a low negative correlation with both Total Defects and Quakers. Quakers had a low positive correlation with Total Defects.

- Beans with greenish tint was found to have more defects compared to beans with yellow-brown tint.
- Double Carbonic Maceration/ Natural processing method showed highest average total defects while Wet Hulling showed the least.
- Sarchimor variety of beans showed the highest average total defects.

Q4) How do different variables interact to influence the Total Cup Points, which represent an overall measure of coffee quality?

- Pairplots between the various coffee attributes (Flavour, Aftertaste, Balance, Acidity, Aroma and Body) showed a strong positive correlation among them. This implies that coffee beans with a high value for one attribute tend to have high values for all of them. This thereby leads to an overall high coffee quality score.