#### **Power BI - Capstone Project**

**Capstone Project: Exploring Coffee Quality Data with Power BI  
About Dataset**

**Coffee Quality Institute:**

The Coffee Quality Institute (CQI) is a non-profit organization that works to improve the quality and value of coffee worldwide. It was founded in 1996 and has its headquarters in California, USA.

CQI's mission is to promote coffee quality through a range of activities that include research, training, and certification programs. The organization works with coffee growers, processors, roasters, and other stakeholders to improve coffee quality standards, promote sustainability, and support the development of the specialty coffee industry.

**Data:**

The data includes a range of information on coffee production, processing, and sensory evaluation. It also contains data on coffee genetics, soil types, and other factors that can affect coffee quality.

Sensory evaluations (coffee quality scores)

* Aroma: Refers to the scent or fragrance of the coffee.
* Flavor: The flavor of coffee is evaluated based on the taste, including any sweetness, bitterness, acidity, and other flavor notes.
* Aftertaste: Refers to the lingering taste that remains in the mouth after swallowing the coffee.
* Acidity: Acidity in coffee refers to the brightness or liveliness of the taste.
* Body: The body of coffee refers to the thickness or viscosity of the coffee in the mouth.
* Balance: Balance refers to how well the different flavor components of the coffee work together.
* Uniformity: Uniformity refers to the consistency of the coffee from cup to cup.
* Clean Cup: A clean cup refers to a coffee that is free of any off-flavors or defects, such as sourness, mustiness, or staleness.
* Sweetness: It can be described as caramel-like, fruity, or floral, and is a desirable quality in coffee.

**PLEASE NOTE: 'Total Cup Points' is literally the total of 10 features given above. There were some notebooks trying to predict the total cup points given these features. We know the exact function underlying the total cup points.**

**Defects:**

Defects are undesirable qualities that can occur in coffee beans during processing or storage. Defects can be categorized into two categories: Category One and Category Two defects.

Category One defects are primary defects that can be perceived through visual inspection of the coffee beans. These defects include Black beans, sour beans, insect-damaged beans, fungus-damaged beans, etc.

Category Two defects are secondary defects that are more subtle and can only be detected through tasting. These defects include Over-fermentation, staleness, rancidness, chemical taste, etc.

**Objective:**

The primary goal of this project is to leverage the rich dataset provided by CQI to understand the factors that contribute to coffee quality. Specifically, we aim to explore the following research questions:

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

2. Is there a correlation between processing methods, origin regions, and coffee quality scores?

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

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

**NOTE: Go through the dataset and documentation thoroughly. Feel free to explore more scenarios based on What – IF analysis. Experiment with wide range of visualizations and formulas**

**Documentation**

The project which we are dealing now is related to Coffee quality data using power BI and the data includes a range of information on coffee production, processing, and sensory evaluation. It also contains data on coffee genetics, soil types, and other factors that can affect coffee quality.  
  
Also we see there are coffee quality scores like Aroma, Aftertaste, flavour, acidity, sweetness, bitterness etc .The data set consists of 31 columns and 207 rows with 13 categorical columns and 18 numerical columns.  
  
Step 1:

**Load data set:**

1. When the data set is downloaded from the odin portal and we see the file is in csv format.
2. Opened power bi desktop under get data < selected csv. File and uploaded and loaded the data in power bi desktop.
3. The data consists of single table which was checked under model view.

**Data Transforming:**

1. By using transform data, we can check whether we have any null values, empty columns and any unnecessary columns to remove the duplicated columns and to get precise data.
2. I have seen few unnecessary columns which are not useful for data visualization and noticed few blank columns which are not helpful for our analysis and tried to remove the columns i.e. Lot Number, Altitude
3. I have noticed that the data type for ”Grading year” and “ expiration year” was text and have created dax function to change the data type from text to date.

Modelling:  
  
Created new table as Attribute groups and divided each sensory attributes under three categories like Taste, Quality, Body

1.What are the key determinants of coffee quality as evaluated through sensory attributes such as aroma, flavour, acidity, etc.?  
  
As we examine the trend for sensory attributes related to taste quality and body, it becomes evident that varieties such as **Gesha** and **Caturra** consistently exhibit higher scores across different countries.

2.Is there a correlation between processing methods, origin regions, and coffee quality scores?  
  
We are examining the correlation between coffee quality scores for Aroma, Acidity, and Flavor, as well as considering processing methods. Regarding origin, we observe that Taiwan stands out compared to other countries. Additionally, the coffee production in 2022 surpasses that of 2023

3. Can we identify any trends or patterns in defect occurrences and their impact on overall coffee quality?  
  
As we observe, the occurrence of defects from both Category One and Category Two has increased over the years for the wash/wet processing method, while it remains lower for the wet hulling method. Additionally, these trends vary based on the country of origin  
  
4. How do different variables interact to influence the Total Cup Points, which represent an overall measure of coffee quality?  
  
We examined the total cup points based on processing methods and found that washed/wet processing has the highest score, while double anaerobic washed processing has the lowest score. Additionally, we analysed moisture percentages for bag weight across different years, observing that 60 kg bags have the highest moisture content, whereas 1 kg bags have the lowest