**Methodology, Insights and Recommendations**

**Business Problem:**

The project aims to analyze rich dataset provided by Coffee Quality Institute (CQI) to gain insights into the factors influencing coffee quality. This information is helpful for various stakeholders in coffee industry, including growers, processors, roasters and retailers.

**Objective:**

1. To find the factors that are influencing sensory attributes such as Aroma, Aftertaste, Flavor, Acidity, Body, Balance, Uniformity, Clean cup and Sweetness.

**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.

1. Is there a correlation between processing methods, origin regions, and coffee quality scores?
2. Can we identify any trends or patterns in defect occurrences and their impact on overall coffee quality?
3. How do different variables interact to influence the Total Cup Points, which represent an overall measure of coffee quality?

**Tools Used:**

The project is build using Microsoft Power Bi, the dataset is in .csv format.

**Dataset:**

The dataset used for the project has 207 rows and 31 columns which is in the shape of (207,31).

Out of 31 columns 18 columns are numeric data and 13 columns are categorical data.

**Methodology:**

1. The dataset is loaded.
2. **Transformation and manipulation of data-**

* Understand the data
* Dealing with headers and column names
* Check column quality (valid, errors, empty, null values)
* Remove duplicates & replace values

1. For visualization Slicer, Card visuals, Donut chart, Scatter plot, clustered column chart, Stacked bar chart, Line chart and Decomposition Tree is used.

**Insights & Recomendations:**

* There are **22 countries** with sum of total cup points (coffee quality) as **17.33K** and with 10 processing methods, sum of category 1 defects is **28** and sum of category 2 defects is **466**.
* **Altitude at 1200 Altitude** is best for growing coffee because it records best Total cup points (coffee quality) of **1.91k Total cup points** (most contribution).
* The **sensory attributes like Sweetness, Clean cup** records highest average coffee quality rating of **10.00**.
* **Washed / Wet method** is the most used processing method in majority of regions.
* **Double Anaerobic washed** processing method from Piendamo, Cauca region brought **great values to coffee quality** with average total cup points of **89.33**.
* **Double Anaerobic washed** processing method from **Colombia** had a major impact for achieving the most coffee quality score on an average amoung all the other countries.
* **Washed / Wet processing method** brought most **category 1 defects of “20”** and also **most category 2 defects of “325”**. So better avoid using Washed / Wet processing method.
* **Mexico** country records the **most category 1 defects of “8”.**
* **Ethiopia and Honduras** country records the **most category 2 defects of “50”**.