Business Problem:

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

Objective:

1) To find the determinants/factors that are influencing the sensory attributes of coffee quality viz. Aroma, Flavor, Aftertaste, 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.

2) To find the correlation between processing methods, origin regions, and coffee quality scores.
3) To find trends or patterns in defect occurrences and their impact on overall coffee quality.
4) To observe how different variables interact to influence the Total Cup Points, which represent an overall measure of coffee quality.

Tools used:

The project is entirely made using Power BI and the inbuilt features. The dataset is in .csv format and MS Excel was used for the readability.

Shape of the dataset:

The dataset that is used for the project has 206 rows and 31 columns i.e in the shape (206,31).

Out of the 31 columns, 18 columns are numerical data and 13 columns are categorical data.

Steps performed:

1) The dataset is loaded, then all the blank rows were removed.
2) To have distinct values, dimension tables were created viz. Altitude DIM, Color DIM, Expiration DIM, Grading Date DIM, Harvest Year DIM, In-Country Partner DIM, Processing Method DIM, Region DIM and Variety DIM.
3) Some names in the “Color” column had typo error, which had been replaced with correct naming.
4) Calculated columns were created for visualization purposes such as the one “Quantity” for visual representation of quantities by countries.
5) For visualization Scatter Plot, Card, Stacked Bar Chart, Donut Chart, Clustered Column Chart, Line Chart and Map were used.

Insights:

1) There are 22 countries, 120 regions, 21 In-country partners and 11 Processing Methods in the entire dataset.
2) In the case of the sensory attribute **Aroma**:
    - United Republic of Tanzania made the highest with average rating of 7.9.
    - In-country partner "Kenya Coffee Traders Association" had produced the best aroma among all the In-country partners.
3) In the case of the sensory attribute **Flavor**:
    - Madagascar made the highest with average rating of 7.92.
    - In-country partner "Asociacion Nacional Del Cafe" had produced the best flavor among all the In-country partners.
4) In the case of the sensory attribute **Aftertaste**:
    - Ethiopia made the highest with average rating of 7.76.
    - In-country partner "METAD Agricultural Development plc" had produced the best aftertaste among all the In-country partners.
5) In the case of the sensory attribute **Acidity**:
    - Ethiopia made the highest with average rating of 8.02.
    - In-country partner "METAD Agricultural Development plc" had produced the best acidity among all the In-country partners.
6) In the case of the sensory attribute **Body**:
    - United Republic of Tanzania made the highest with average rating of 7.92.
    - In-country partner "METAD Agricultural Development plc" had retained the best body among all the In-country partners.
7) In the case of the sensory attribute **Balance**:
    - United Republic of Tanzania made the highest with average rating of 7.92.
    - In-country partner "Blossom Valley International宸嶧國際" and "Kenya Coffee Traders Association" had produced the best balance among all the In-country partners.
8) In the case of the sensory attribute **Uniformity**:
    - All the countries had the same ratings.
    - In-country partner all the In-Country partners had equally managed to contribute towards Uniformity.
9) In the case of the sensory attribute **Clean Cup**:
    - All the countries had the same ratings.
    - In-country partner all the In-Country partners had equally managed to contribute towards Clean Cup.
10) In the case of the sensory attribute **Sweetness**:
    - All the countries had the same ratings.
    - In-country partner all the In-Country partners had equally managed to contribute towards Sweetness.
11) At an altitiude 1700-1930 , coffee qualities were greatly affected except in Uniformity, Clean Cup and Sweetness where Total Cup Points were uniform in values.
12) Among the varieties of coffee beans "Castillo" happened to be the best type of coffee beans except in Uniformity, Clean Cup and Sweetness where Total Cup Points were uniform in values.
13) In Sensory attribute "Aroma" particulary "Brownish" colored coffee beans happened to affect the smell of coffee quality while in the other sensory attributes "Brownish-green"

colored coffee beans affected the coffee qualities; except in Uniformity, Clean Cup and Sweetness where Total Cup Points were uniform in values.

14) “Double Anaerobic Washed” processing method brought great values to the coffee qualities except in Uniformity, Clean Cup and Sweetness where Total Cup Points were uniform in values.
15) Speaking of the coffee quality attributes based on the grading dates, the year “2022” , quarter wise the 4th quarter, month wise particularly “September and October” and “3rd and 4th week” of every month on average happened to contribute the most to all the coffee attributes, except in Uniformity, Clean Cup and Sweetness where Total Cup Points were uniform in values.
16) “Double Anaerobic Washed” processing method for the country **Colombia** had a major impact for achieving the most Coffee Quality Score on an average among all the other countries.
17) From the year of Harvest i.e. 2017/2018, the defects became significantly less towards 2023 but also the Cup Points drastically decreased along.
18) Ethiopia happens to be the country with highest defects.
19) The “Washed/Wet” processing method brought the most defects to coffee quality indicating that it should be avoided.