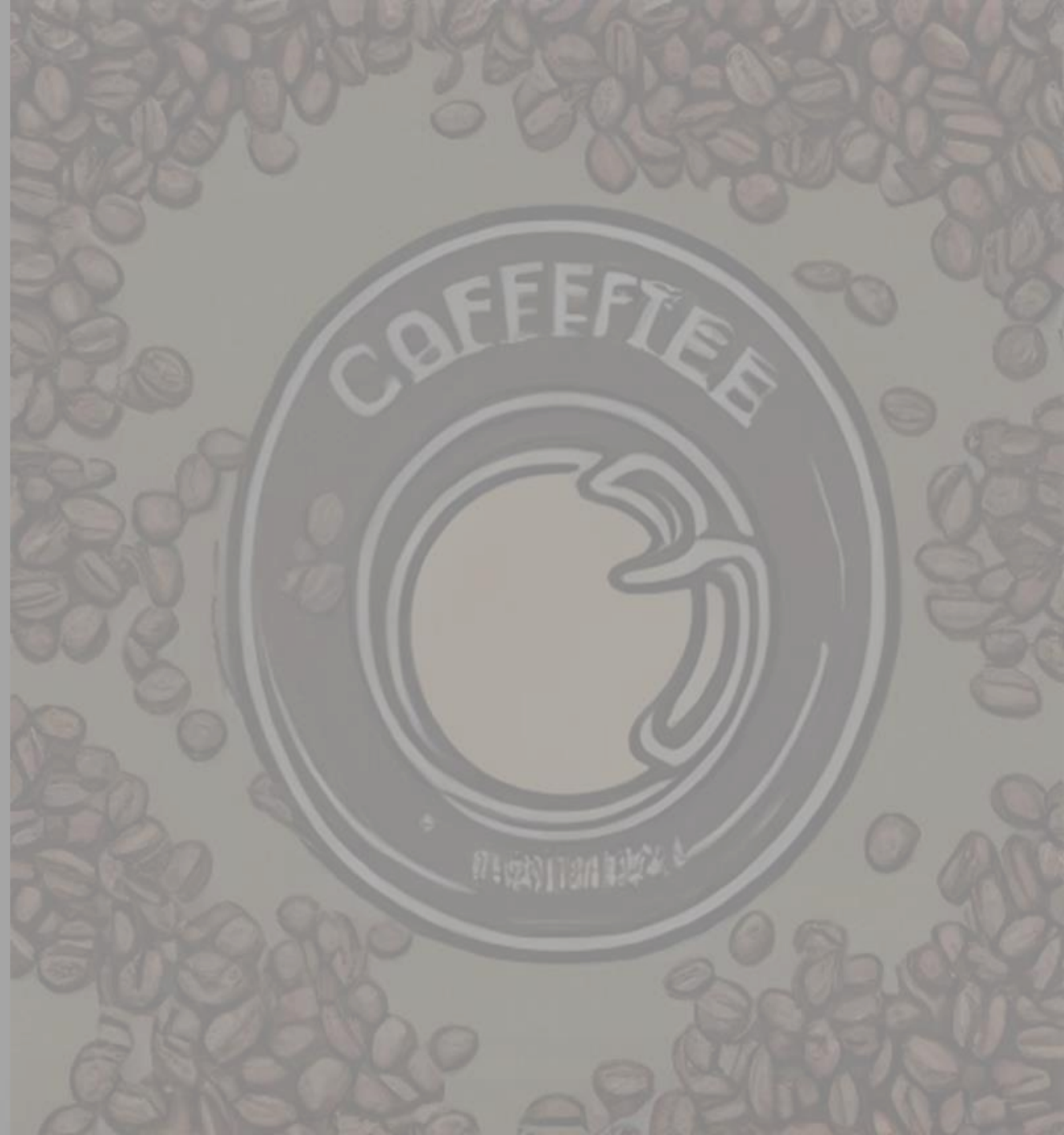




ARABICA COFFEE QUALITY DASHBOARD

CONTENTS

- Business Problem
- Tools used
- About Dataset
- About Coffee
- Sensory evaluations
- Data Understanding
- Data Dictionary
- Model View
- Steps, Methods and Calculations
- Visualization (Dashboard)
- Insights
- Recommendations



BUSINESS PROBLEM

The primary goal of this project is to leverage the rich dataset provided by Coffee Quality Institute (CQI) to understand the factors that contribute to coffee quality.



TOOLS USED

- Microsoft Power BI Desktop



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.

ABOUT COFFEE

- A process known as **Coffee Cupping** is used to find the quality score of a coffee.
- Every coffee in the world is given a quality score out of 100, which is broken down into sections.
- The final coffee quality score is the sum of the total score of each cup, minus the defects.
- The scores in every category (Fragrance/aroma, Flavour, Aftertaste, Acidity, Body, Uniformity, Balance, Cleanliness, Sweetness, Overall score) are then added together (subtracting any defects) to achieve the final cup score.
- Cuppings are typically carried out by **Q graders**, who are trained by the Coffee Quality Institute to analyse and grade coffee.
- A specialty coffee must score at least 80 points.
- Coffee that scores closer to 80 points is really good but can contain many more faults than the sample that enters the 90+ hall of fame.

| TOTAL COFFEE CUPPING QUALITY SCORE | | |
|------------------------------------|--------------------------------|----------------------|
| 90 - 100 | OUTSTANDING | SPECIALTY COFFEE |
| 85 - 89.99 | EXCELLENT | |
| 80 - 84.99 | VERY GOOD | |
| < 80.0 | BELOW SPECIALTY COFFEE QUALITY | NOT SPECIALTY COFFEE |

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.

'**Total Cup Points**' is the total of 9 features given above and 1 **Overall** parameter where the graders can show their personal consideration. The more the sample reflects the typical features based on its origin, the higher the coffee score.

DATA UNDERSTANDING

- **Dataset file type:** Comma Separated File.
- **Columns provided:** ID, Country of Origin, Lot Number, Altitude, Region, Number of Bags, Bag Weight, In-Country Partner, Harvest Year, Grading Date, Variety, Status, Processing Method, Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup, Sweetness, Overall, Defects, Total Cup Points, Moisture Percentage, Category One Defects, Quakers, Color, Category Two Defects, Expiration
- **Added columns:** Life Span, Species, Quality, Speciality Coffee, Moisture Range, Harvest Year, Expiration Year
- **Data Dimensions:** 207 rows and 31 columns.

DATA DICTIONARY

- **ID:** A unique identifier for each entry in the dataset.
- **Country of Origin:** The country where the coffee beans were grown.
- **Lot Number:** The specific lot number associated with the batch of coffee beans.
- **Altitude:** The altitude at which the coffee beans were grown.
- **Region:** The specific region within the country where the coffee beans were cultivated.
- **Number of Bags:** The quantity of bags containing coffee beans in the batch.
- **Bag Weight:** The weight of each bag of coffee beans.
- **In-Country Partner:** The local partner or organization involved in the coffee production process within the country of origin.
- **Harvest Year:** The year in which the coffee beans were harvested.
- **Grading Date:** The date when the coffee beans were graded or evaluated.
- **Variety:** The variety or species of coffee beans, such as Arabica or Robusta.

DATA DICTIONARY

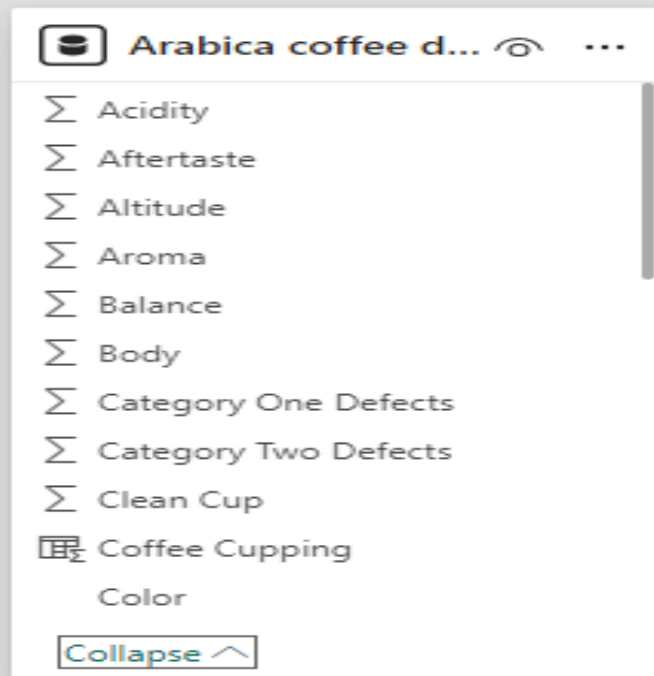
- **Status:** The status of the batch, indicating whether it's completed or ongoing.
- **Processing Method:** The method used to process the coffee beans, such as washed/wet, natural/dry, or semi-lavado.
- **Defects:** Defects are the detected unpleasant flavors warranting a negative score.
- **Total Cup Points:** The total points awarded to the batch based on its sensory evaluation.
- **Moisture Percentage:** The percentage of moisture present in the coffee beans.
- **Category One Defects:** The number of category one defects found in the batch.
- **Quakers:** The presence of quaker beans, which are underdeveloped or defective beans.
- **Color:** The color of the coffee beans.
- **Category Two Defects:** The number of category two defects found in the batch.
- **Expiration:** The expiration date associated with the evaluation or certification of the coffee beans.

DATA DICTIONARY

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
|----|----|---------------------|------------|-----------|------------|----------|-----------|------------|------------|------------|-----------|----------|-------------|-------|--------|------------|---------|------|---------|-----------|-------|
| 1 | ID | Country o | Lot Numb | Altitude | Region | Number o | Bag Weigh | In-Country | Harvest Ye | Grading D | Variety | Status | Processing | Aroma | Flavor | Aftertaste | Acidity | Body | Balance | Uniformit | Clean |
| 2 | 0 | Colombia | CQU20220 | 1700-1930 | Piendamc | 1 | 35 kg | Japan Cof | 2021 / 202 | Septembe | Castillo | Complete | Double Ar | 8.58 | 8.5 | 8.42 | 8.58 | 8.25 | 8.42 | 10 | |
| 3 | 1 | Taiwan | The 2022 F | 1200 | Chiayi | 1 | 80 kg | Taiwan Cc | 2021 / 202 | Novembe | Gesha | Complete | Washed / | 8.5 | 8.5 | 7.92 | 8 | 7.92 | 8.25 | 10 | |
| 4 | 2 | Laos | The 2022 F | 1300 | Laos Borot | 19 | 25 kg | Taiwan Cc | 2021 / 202 | Novembe | Java | Complete | Semi Was | 8.33 | 8.42 | 8.08 | 8.17 | 7.92 | 8.17 | 10 | |
| 5 | 3 | Costa Rica | CQU20220 | 1900 | Los Santos | 1 | 22 kg | Japan Cof | 2022 | Septembe | Gesha | Complete | Washed / | 8.08 | 8.17 | 8.17 | 8.25 | 8.17 | 8.08 | 10 | |
| 6 | 4 | Colombia | CQU20230 | 1850-2100 | Popayan, C | 2 | 24 kg | Japan Cof | 2022 | March 6th | Red Bourk | Complete | Honey, Mc | 8.33 | 8.33 | 8.08 | 8.25 | 7.92 | 7.92 | 10 | |
| 7 | 5 | Guatemala | The 2022 F | 1668 | Chimalter | 5 | 30 kg | Taiwan Cc | 2022 | Novembe | Gesha | Complete | Washed / | 8.33 | 8.33 | 8.25 | 7.83 | 7.83 | 8.17 | 10 | |
| 8 | 6 | Taiwan | The 2022 F | 1250 | Chiayi | 1 | 27 kg | Taiwan Cc | 2021 / 202 | Novembe | Gesha | Complete | Washed / | 8.33 | 8.17 | 8.08 | 8 | 7.83 | 8.25 | 10 | |
| 9 | 7 | Taiwan | The 2022 F | 1200 | Chiayi | 1 | 90 kg | Taiwan Cc | 2021 / 202 | Novembe | SL34+Gesh | Complete | Natural / I | 8.25 | 8.25 | 8.17 | 8 | 7.92 | 8.08 | 10 | |
| 10 | 8 | Taiwan | The 2022 F | 1250 | Chiayi | 1 | 30 kg | Taiwan Cc | 2021 / 202 | Novembe | SL34 | Complete | Washed / | 8.08 | 8.08 | 8.25 | 8.08 | 7.92 | 8 | 10 | |
| 11 | 9 | Tanzania, CN 412723 | 1400-1700 | KILIMANJA | | 320 | 60 kg | Kenya Cof | 2022 / 202 | February | Bourbon | Complete | Washed / | 8.08 | 8.17 | 8.08 | 8.17 | 8 | 8 | 10 | |
| 12 | 10 | Ethiopia | 010/0296/ | 1800-2200 | Guji | 10 | 30 kg | Japan Cof | 2021 / 202 | May 31st, | Ethiopian | Complete | Natural / I | 8.08 | 8.25 | 8 | 8.08 | 7.92 | 7.92 | 10 | |
| 13 | 11 | Guatemala | The 2022 F | 2000 | Acatenang | 5 | 15 kg | Taiwan Cc | 2021 | Novembe | Gesha | Complete | Natural / I | 8.08 | 8 | 8 | 7.75 | 8.25 | 8.17 | 10 | |
| 14 | 12 | Taiwan | The 2022 F | 1250 | Yunlin | 1 | 60 kg | Taiwan Cc | 2021 / 202 | Novembe | Gesha | Complete | Washed / | 8.08 | 8 | 8.08 | 8.08 | 8 | 8 | 10 | |
| 15 | 13 | Ethiopia | Grade 1, G | 1900-2000 | Guji | 40 | 60 kg | METAD Ag | 2021 / 202 | August 26 | Gesha | Complete | Natural / I | 7.67 | 8.17 | 8 | 8.33 | 8 | 8 | 10 | |
| 16 | 14 | Colombia | The 2022 F | 1850 | tolima | 70 | 35 kg | Taiwan Cc | 2021 / 202 | Novembe | Caturra | Complete | Washed / | 8.08 | 8 | 8.08 | 7.92 | 8.08 | 8 | 10 | |
| 17 | 15 | Taiwan | The 2022 F | 1100 | Chiayi | 1 | 60 kg | Taiwan Cc | 2021 / 202 | Novembe | SL34 | Complete | Pulped na | 8.17 | 8.08 | 8 | 7.92 | 8 | 7.92 | 10 | |
| 18 | 16 | Ethiopia | CQU20230 | 1900-2100 | Gedeb, Yir | 8 | 5 kg | Japan Cof | 2022 | April 7th, | Wolishalo | Complete | Washed / | 8.17 | 8.08 | 7.92 | 8.17 | 7.75 | 7.92 | 10 | |
| 19 | 17 | Taiwan | 202203 | 1300 | Shibi, Guk | 5 | 2 kg | Blossom V | 2022 | October 2 | Gesha | Complete | Natural / I | 8 | 8.17 | 8 | 7.92 | 7.92 | 7.92 | 10 | |
| 20 | 18 | Taiwan | 202112 | 1200 | Gukeng To | 8 | 1 kg | Blossom V | 2021 / 202 | October 2 | Gesha | Complete | Natural / I | 8.08 | 8.17 | 7.75 | 7.92 | 7.83 | 8 | 10 | |
| 21 | 19 | Tanzania, CN 412723 | 1570-1600 | Arusha | | 200 | 30 kg | Kenya Cof | 2022 / 202 | February | Bourbon | Complete | Washed / | 8.17 | 8 | 7.92 | 7.92 | 8.17 | 7.75 | 10 | |
| 22 | 20 | Guatemala | The 2022 F | 1900 | Guatemala | 8 | 30 kg | Taiwan Cc | 2021 / 202 | Novembe | Gesha | Complete | Natural / I | 8 | 7.92 | 8.08 | 7.92 | 7.75 | 8 | 10 | |
| 23 | 21 | Taiwan | 2022/02 | 950 | Chiayi | 1 | 20 kg | Taiwan Cc | 2022 | December | Typica | Complete | Natural / I | 8.08 | 8 | 7.92 | 7.92 | 8 | 7.82 | 10 | |

Source_File(df_arabica_clean)

MODEL VIEW



STEPS, METHODS AND CALCULATIONS

Data Integration/Extraction:

- The Case study is related to Coffee Quality.
- The file type of dataset provided is Microsoft Excel Comma Separated Values. Its extension is .csv.
- Name of the dataset file is df_arabica_clean.
- The file origin type of dataset is 65001: Unicode (UTF-8).
- The delimiter is comma in dataset file as it is the default delimiter for a comma separate value (CSV) file.
- Opened Power BI desktop, Scrolled down and selected the Text/CSV data source in Get Data option in data ribbon on Home tab then select source file from desired folder and clicked okay.

STEPS, METHODS AND CALCULATIONS

Data Cleaning and Preparation:

- Selected the table of concern in Navigator and clicked on transform prior to which dataset get loaded in the power query editor for transformation.
- The dataset contains 1 table named 'df_arabica_clean'.
- Shape of the dataset: 207 rows and 31 columns.
- Checked the datatypes of all columns.
- The dataset contains 18 numerical and 13 categorical columns as per the Power Query automatic datatype detection.
- Renamed the df_arabica_clean table to Arabica coffee dataset.
- Removed the ID, Lot Number, Region, Number of Bags, Bag Weight, Grading date, Status, Total Cup Points, Quakers column using Remove column option in the manage columns menu of Home tab.
- Renamed Country Of Origin column with Country, Expiration with Date of Expiry.
- Checked the datatypes of all fields.

STEPS, METHODS AND CALCULATIONS

- Extracted the first 4 characters from Harvest Year using extract option in text column menu of Transform tab. Also, changed the datatype from text to whole number type.
- The dataset contained values in multiple language to resolve this issue in order for better analysis, with the help of Google Sheet and using GOOGLETRANSLATE(text, [source_language], [target_language]) changed the language to English. Cell address was used in place of text, “auto” in place of source_language and “en” in place of target_language.
- Replaced 台灣咖啡研究室 with space.
- Removed Blossom Valley International宸嶧國際 using filter.
- Changed the values from mixed language to English in the In-Country Partner column.
- Replaced United States (Hawaii) with United States in Country column.
- Replaced Tanzania, United Republic Of with United Republic of Tanzania in Country of Origin column.
- Removed blank from Altitude column.

STEPS, METHODS AND CALCULATIONS

- Replaced ~and A with – in Altitude column.
- Extracted 4 characters from beginning from Altitude column and then replaced – with space. Also changed the datatype from text to whole number type.
- Set the format to Capitalize each word in the In-Country Partner column using format option.
- Removed blanks, Double Carbonic Maceration / Natural, Honey Mossto from Processing Method column.
- Replaced Semi Washed with Pulped Natural/Honey from Processing Method column.
- Set the format to Capitalize each word in the Processing Method column using format option.
- Set the format to Capitalize each word in the Variety column using format option.
- Removed Blanks, Unknow and Unknown, Shg values in Variety column using filter.

STEPS, METHODS AND CALCULATIONS

- Replaced bluish-green with Blue-Green, greenish with Green, yello-green with Yellow-Green, yellow green with Yellow-Green, yellow- green with Yellow-Green, yellowish with Yellow.
- Extracted Year from Expiration column using extract and last characters option in transform tab.
- Renamed Last Character column to Expiration Year.
- Replaced catuai with Catucaí and Gesha with Geisha, Maragogype with Maragogipe in Variety column.
- Filtered out the varieties other than Arabica.
- The duplicates from removed from all the tables using reduce rows menu and remove duplicates option in Home tab.
- Country column was ordered in ascending order. Defects column was removed.

STEPS, METHODS AND CALCULATIONS

DAX:

- Created a calculated columns by clicking on new column option in table view and using formula's:
 - a. Life Span = 'Arabica coffee dataset'[Expiration Year] - 'Arabica coffee dataset'[Harvest Year]
 - b. Total Defects = 'Arabica coffee dataset'[Category One Defects] + 'Arabica coffee dataset'[Category Two Defects]
 - c. Total Score = 'Arabica coffee dataset'[Aroma] + 'Arabica coffee dataset'[Flavor] + 'Arabica coffee dataset'[Aftertaste] + 'Arabica coffee dataset'[Acidity] + 'Arabica coffee dataset'[Body] + 'Arabica coffee dataset'[Balance] + 'Arabica coffee dataset'[Uniformity] + 'Arabica coffee dataset'[Clean Cup] + 'Arabica coffee dataset'[Sweetness] + 'Arabica coffee dataset'[Overall]
 - d. Coffee Cupping = 'Arabica coffee dataset'[Total Score] - 'Arabica coffee dataset'[Total Defects]
 - e. Species = "Arabica" f. Speciality Coffee = SWITCH(TRUE(), 'Arabica coffee dataset'[Coffee Cupping] > 79.99, "Speciality badge", 'Arabica coffee dataset'[Coffee Cupping] > 59.99, "Commercial grade", "Below commercial grade") g. Moisture Range = SWITCH(TRUE(), 'Arabica coffee dataset'[Moisture Percentage] > 12, "Excessive moist", 'Arabica coffee dataset'[Moisture Percentage] > 8.99, "Correct moisture", "Excessive dry")

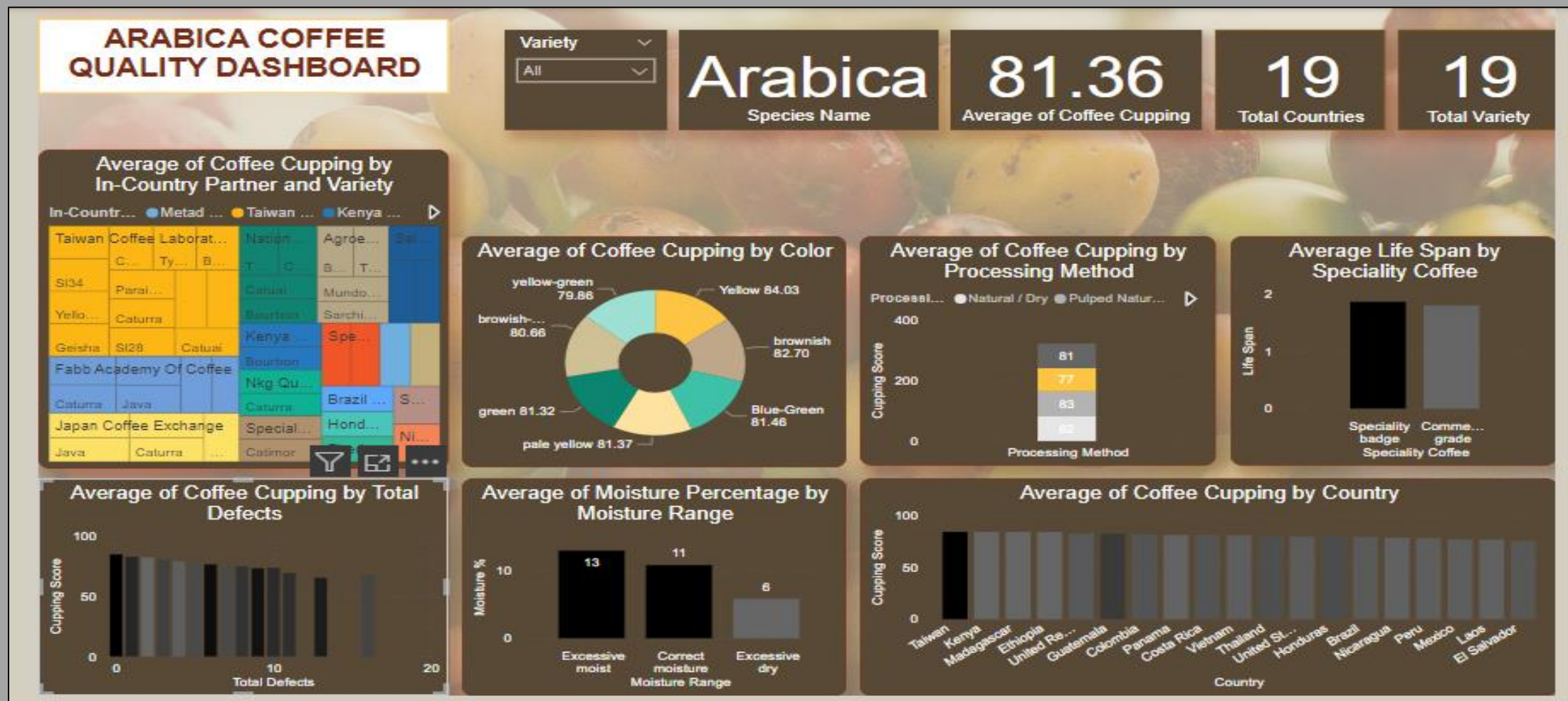
STEPS, METHODS AND CALCULATIONS

- Created a calculated measures by clicking on New Measure option in table view and by using the formula's:
a. Total Variety = `DISTINCTCOUNT('Arabica coffee dataset'[Variety])`
- Created a calculated column by the name Moisture Percentage using the formula: `Moisture Percentage = 'Coffee Quality'[Moisture] / 100`. After that, changed the format of the Moisture Percentage from general to percentage in column tools in the table view and changed the number of decimal places from 2 to 0.
- Replaced plus (+) with comma (,) in variety column.
- Replaced bluish-green with blue-green, greenish with green, yello-green with yellow-green, yellow green with yellow-green, yellow- green with yellow-green, yellowish with yellow.
- Extracted Year from Expiry date column using extract and text after delimiter option in transform tab.
- Changed the datatype of date of expiry column from text to whole number type and changed column name to Expiry Year.
- Created a calculated column names Total Defects using formula: `Total Defects = 'Coffee Quality'[Category One Defects] + 'Coffee Quality'[Category Two Defects]`
- Removed blanks from variety column using filter.

STEPS, METHODS AND CALCULATIONS

- To crosscheck the correctness of Total Cup Points columns, created a calculated column named Total Score and the previous Total Cup Points was deleted. The formula used: Total Score = 'Coffee Quality'[Aroma] + 'Coffee Quality'[Flavor] + 'Coffee Quality'[After taste] + 'Coffee Quality'[Acidity] + 'Coffee Quality'[Body] + 'Coffee Quality'[Balance] + 'Coffee Quality'[Uniformity] + 'Coffee Quality'[Clean Cup] + 'Coffee Quality'[Sweetness] + 'Coffee Quality'[Overall]
- Created a calculated column Cupping Score to get the final Score. The Formula used was: Coffee Cupping = 'Arabica coffee dataset'[Total Score] - 'Arabica coffee dataset'[Total Defects]
- Filtered out the varieties other than arabica.
- Replaces catuai with Catucaí and Gesha with Geisha, Maragogype with Maragogipe in variety column.
- Removed blanks from Processing method column. Created a calculated column names Species using the formula: Species = "Arabica"
- Created a calculated measured by name Total Defects Sum using formula: Total Defects Sum = SUM('Coffee Quality'[Total Defects])
- Created a calculated measured by name Total Countries using formula: Total Countries = DISTINCTCOUNT('Arabica coffee dataset'[Country])

VISUALISAZATION



INSIGHTS

- The species taken into consideration is **Arabica**.
- The total number of countries taken into consideration is **19**.
- The total number of varieties taken into consideration is **19**.
- **Taiwan** is the best-working country when it comes to upgrading Arabica coffee quality.
- Highest number of variety are grown by **Taiwan Coffee Laboratory**.
- **Defects** are inversely proportional to coffee quality.
- **Yellow** beans should be plucked to obtain good quality beans.
- Coffee beans should contain **9-12 % moisture**. The moisture present in the coffee beans of the dataset is correct.
- The **Pulped Natural/Honey** processing method should be adopted.
- **Taiwan Coffee Laboratory** is the best in-country partner.
- The better the coffee quality, the **longer** the lifespan of the coffee.



RECOMMENDATIONS

- **Focus on Quality Control:** Implement stringent quality control measures throughout the production process to minimize defects and maintain high coffee quality standards.
- **Collaborate with Taiwan's Laboratory:** Strengthen partnerships with Taiwan's laboratory to leverage their expertise and resources in coffee research, analysis, and quality assurance.
- **Prioritize Pulped Natural/Honey Processing:** Emphasize the use of the Pulped Natural/Honey processing method for coffee beans to ensure cleaner, brighter flavors and higher cup scores.
- **Control Moisture Content:** Monitor and regulate the moisture content of coffee beans to fall within the optimal range of 9-12%, as this contributes to superior cup scores and overall coffee quality.
- **Quality Yellow Coffee Beans:** Source and select yellow coffee beans, indicating fully developed beans as they give the coffee its best flavors and ensures a better starting point for coffee production.
- **Invest in Training and Equipment:** Provide training for coffee producers on best practices for cultivation, processing, and quality control. Additionally, invest in equipment and infrastructure to support these efforts and enhance overall coffee quality.

THANK YOU