Power BI Coffee Quality Analysis

Introduction

Coffee is one of the most popular drinks in the world, and its quality has a significant impact on both market value and customer satisfaction. Numerous steps are involved in the process from farm to cup, and each one affects the taste, aroma, and overall quality of the finished product. The main goal of this project is to analyse extensive data on the production of coffee, processing techniques, sensory analysis, and defect occurrences. Through the use of this data, the project hopes to unearth insights that will benefit all parties involved in the coffee business, from growers and producers to roasters and retailers, in order to enhance coffee quality, optimize production procedures, and eventually provide customers with an improved coffee experience.

Objective

This research aims to provide a comprehensive analysis of the several elements that affect the quality of coffee. Examining the connections between sensory characteristics, defect rates, processing techniques, and place of origin is part of this. Finding patterns that can direct enhancements in the processes of coffee manufacturing and quality control, comprehending the interplay between many variables, and identifying the critical factors influencing coffee quality are the objectives. The research aims to offer practical insights that can improve the overall worth and marketability of coffee products through this analysis.

Tools used

- Power BI
- Excel

Reseach Questions

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

Dataset Overview

Attributes: The dataset comprises in-depth sensory analyses that address features including Body, Acidity, Aroma, Flavour, Aftertaste, Clean Cup, Uniformity, and Sweetness. Each of these characteristics has a score, and the sum of these scores determines the overall quality of the coffee.

Total Cup Points: Derived from each sensory attribute separately, this is the cumulative score that indicates the overall quality of the coffee. It is the main measure of coffee quality.

Class One Defects: These are the main flaws that can be seen with the naked eye. Examples of these are black beans, sour beans, beans with insect damage, and beans with fungal damage.

Category Two Defects: These are milder flaws that can be identified by taste, such as overfermentation, staleness, rancidity, and chemical flavour.

Country and Origin Region: The dataset includes the geographic details of the coffee's growing region. A key element in defining flavor profiles and general quality is the place of origin.

Processing Method: Information on the type of processing used, such as natural, honey-processed, or washed, is provided. The flavor and quality of coffee are greatly influenced by the processing techniques used.

Variety: The particular type of coffee (such as Arabica or Robusta) is mentioned because each variety has distinct qualities that add to the overall quality.

Moisture Percentage: This measure shows the amount of moisture in the coffee beans, which has an impact on roasting, storage, and overall quality.

Bag Weight and Bag Count: These measurements shed light on the amount of coffee produced and aid in determining the production's size.

height: A crucial element influencing the flavor and caliber of the coffee is the height at which it is cultivated. Information on this point is given.

Grading Date: These indicate when the coffee was analyzed and harvested, which makes it possible to analyze trends over time.

Important Features

1. Numeric Columns

- Total Cup Points: A numeric aggregate score representing overall coffee quality.
- Sensory Attributes:
- Aroma
- Flavor
- Acidity
- Aftertaste
- Balance
- Body
- Clean Cup
- Sweetness
- Uniformity
- Defects:
- Category One Defects (e.g., black beans, sour beans)
- Category Two Defects (e.g., staleness, rancidness)
- Altitude: Numeric, typically recorded in meters above sea level.
- Moisture Percentage: Numeric, indicates moisture content in beans.
- Bag Weight: Numeric, weight of coffee bags.
- Number of Bags: Numeric, total number of bags produced or sold.

2. Categorical Columns

- Country of Origin: Categorical, indicates where the coffee was grown.
- Region: Categorical, specifies the region within the country.
- Processing Method: Categorical, such as washed, natural, honey-processed.
- Variety: Categorical, type of coffee plant (e.g., Arabica, Robusta).
- Color: Categorical, indicating the color of the beans.
- Lot Number: Categorical, unique identifier for each batch of coffee.
- Harvest Year: Categorical, as it represents a specific time period.
- Grading Date: Categorical, though it could also be treated as a date column.
- In-Country Partner: Categorical, represents local partners involved in production.
- Expiration Date: Categorical, though this could also be treated as a date.

Data preprocessing

Removed the following Columns :

Lot Number: This field should be removed because it is solely used as a unique identifier and has no analytical value.

Status: A row can be eliminated if all of its values (such as "Completed") are the same since it lacks variation.

Expiration Date: Please delete this column as it has no bearing on the coffee's quality at the time of grading.

Defects: Delete the column if it merely has zeros in it and offers no meaningful information.

Harvest Year: this should be eliminated because the grading date is more precise and significant than the harvest year in terms of coffee quality.

Split Data:

Split the Altitude column using a delimiter -

Max and Min Altitude:

Convert data types of Max Altitude and Min Altitude to whole numbers.

Replace null values in Max Altitude with 0.

Calculate Average Altitude:

Create a new column for Average Altitude as follows:

Average Altitude = (Max Altitude + Min Altitude)/2

Cleaning Columns:

Region has two empty rows which are also the errors in altitude, so I am removing them: 207 to 205 rows.

Removed errors in average altitude which were caused due to caused due to the split in values like :1200 A 1400 or using different symbols other than hyphen 205 to 203 rows.

Bag Weight: Extracted text before delimiter space and removed kg .

Renamed to Bag Weight(kg) converted datatype to whole number.

grading date : Replaced the st, nd, rd, th with space and changed the grading date to date datatype, also replaced rows containing Augu with August.

Removed Empties in the variety column and now has 205 to198 rows.

For aroma, flavour, Aftertaste, Acidity, Body, Balance, Overall, Total Cup Points , Moisture Percentage changed the delimiter "." to " , " and changed the datatype to decimal.

Changed Uniformity, Clean Cup, Sweetness, to whole number data type

Cleaned the Region column: Translated rows containing Chinese and French words. Then replaced many duplicate rows containing similar regions:

- o   By lowering the whole column and capitalizing each word in the column again
- o   Manually replacing the rows that have similar names but contained commas or other delimiters

Cleaned the Color column:

- o   Capitalized words in the colour column
- o   Yellow Green, Yello-Green, Yellow- Green to Yellow-Green
- o   Bluish-Green to Blue-Green
- o   Greenish to Green

Removed errors in Uniformity Column194 to192 rows

Data Modelling

Splitting of table into three other tables for Data Modelling:

Fact Table: Sensory_Evaluation

Primary Key: ID

Metrics: Acidity, Aftertaste, Aroma, Balance, Body, Clean Cup, Flavor, Overall

Dimension Table 1: Coffee_Preprocessing

Primary Key: ID

Foreign Key: ID (from Sensory_Evaluation)

Attributes: Bag Weight, Category One Defects, Category Two Defects, Color, Grading Date, Moisture Percentage, Number of Bags, Quakers

**Dimension Table 2: Coffee_Production**

Primary Key: ID

Foreign Key: ID (from Sensory_Evaluation)

Attributes: AvgAltitude, Country, In-Country Partner, MaxAltitude, MinAltitude, Processing Method, Region, Variety

Relationship : The above three tables have one-one relationship with each other with cross-filter direction as both

Pivoted Attributes:

This table is created for ease of use with visuals when using the Sensory attributes, this table contains attributes in one column and scores(values) in another column
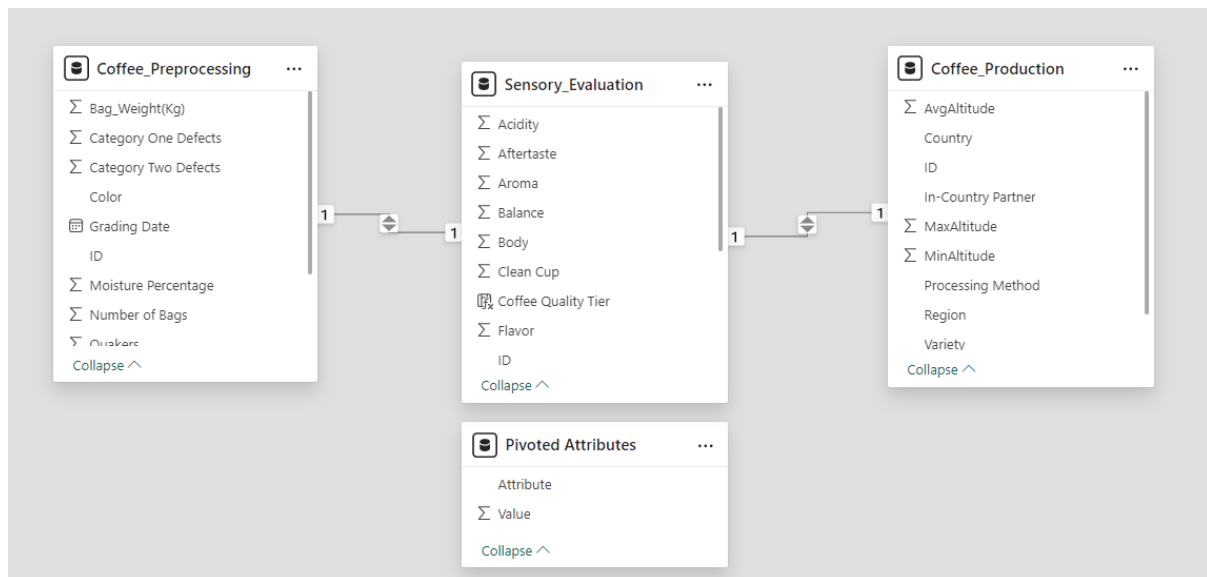
Data Analysis

Calculated Measures:

Sum of Cup Points = SUM(Sensory_Evaluation[Total Cup Points])

Average of Cup Points = AVERAGE(Sensory_Evaluation[Total Cup Points])

Total Bags = SUM(Coffee_Preprocessing[Number of Bags])

Total Defects = SUM(Coffee_Preprocessing[Category One Defects]) +

SUM(Coffee_Preprocessing[Category Two Defects])

Processing_Methods = COUNT(Coffee_Production[Processing Method])

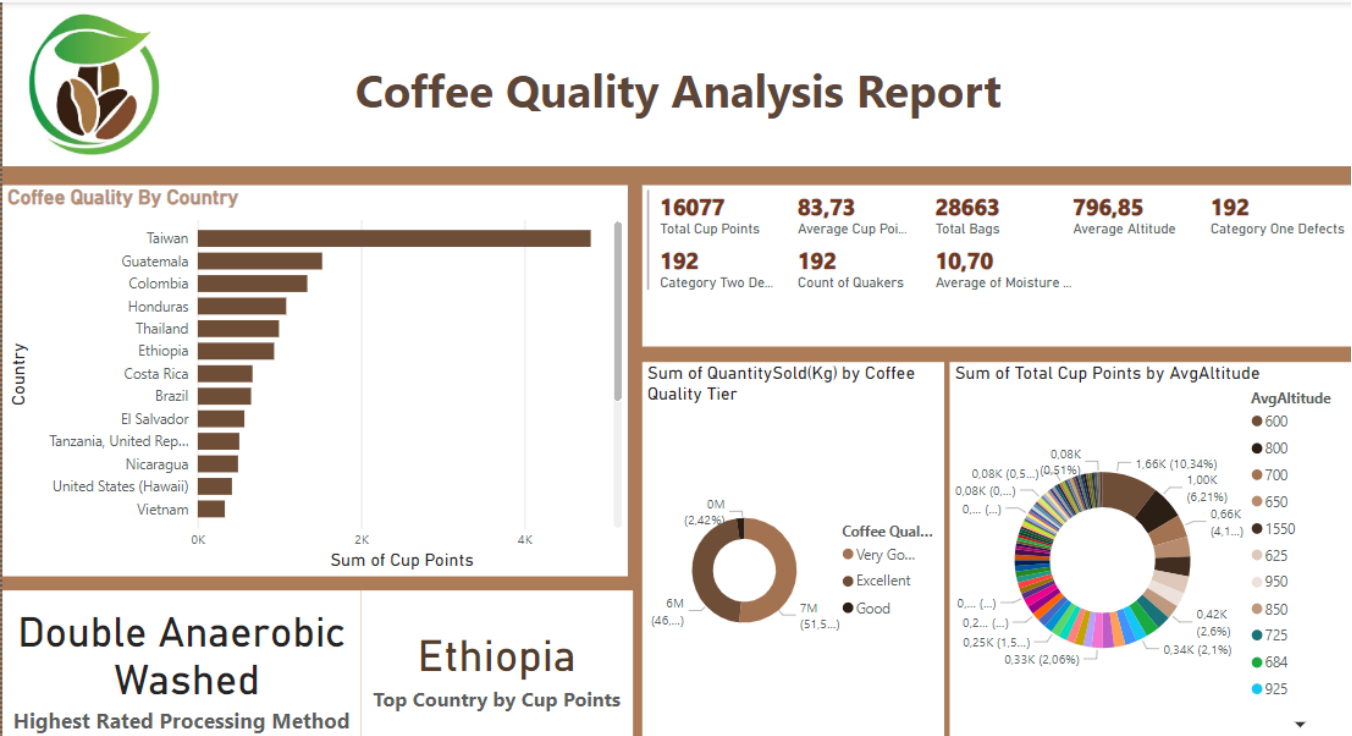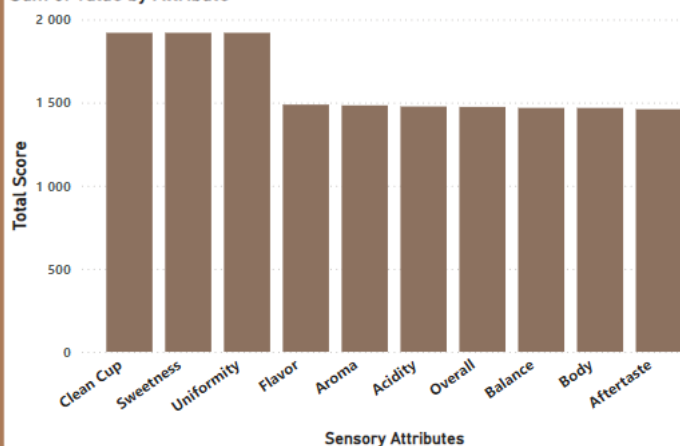**Calculated Columns:**

Coffee Quality Tier =

SWITCH(

    TRUE(),

    Sensory_Evaluation[Total Cup Points] >= 85, "Excellent",

    Sensory_Evaluation[Total Cup Points] >= 82, "Very Good",

    Sensory_Evaluation[Total Cup Points] >= 70, "Good",

    "Average"

)

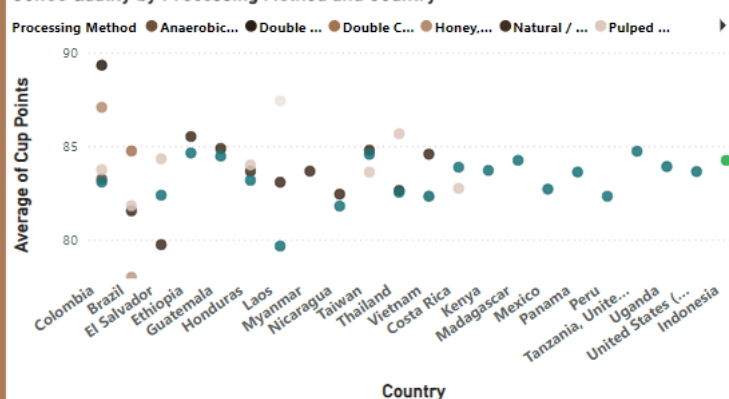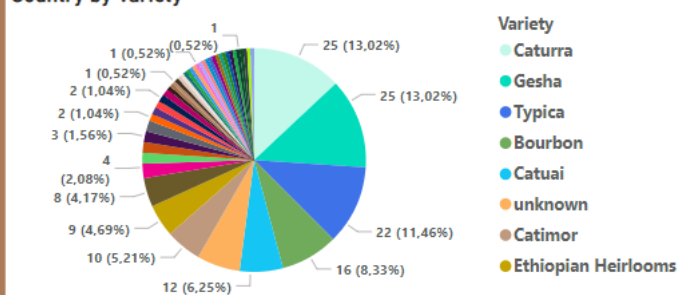QuantitySold(Kg) = Coffee_Preprocessing [Number of Bags] * Coffee_Preprocessing[Bag_Weight(Kg)]

Report

Sum of Value by Attribute

Total Score vs. Sensory Attributes

Clean Cup, Sweetness, Uniformity, Flavor, Aroma, Acidity, Overall, Balance, Body, Aftertaste

Coffee Quality by Processing Method and Country

Processing Method ● Anaerobic... ● Double ... ● Double C... ● Honey,... ● Natural / ... ● Pulped ... ▶

Average of Cup Points vs. Country

Colombia, Brazil, El Salvador, Ethiopia, Guatemala, Honduras, Laos, Myanmar, Nicaragua, Taiwan, Thailand, Vietnam, Costa Rica, Kenya, Madagascar, Mexico, Panama, Peru, Tanzania, Unite..., Uganda, United States (..., Indonesia

Country by Variety

Variety
- ● Caturra
- ● Gesha
- ● Typica
- ● Bourbon
- ● Catuai
- ● unknown
- ● Catimor
- ● Ethiopian Heirlooms

25 (13,02%), 25 (13,02%), 22 (11,46%), 16 (8,33%), 12 (6,25%), 10 (5,21%), 9 (4,69%), 8 (4,17%), 4 (2,08%), 3 (1,56%), 2 (1,04%), 2 (1,04%), 1 (0,52%), 1 (0,52%), 1 (0,52%), 1

Country by Processing Method

1 (0,52%), 23 (11,98%), 44 (22,92%), 118 (61,46%)

Processing Method
- ● Washed / Wet
- ● Natural / Dry
- ● Pulped natural / honey
- ● Anaerobico 1000h
- ● Double Anaerobic Wa...
- ● Double Carbonic Mac...
- ● Honey,Mossto
- ● Semi Washed
- ● SEMI-LAVADO
- ● Wet Hulling
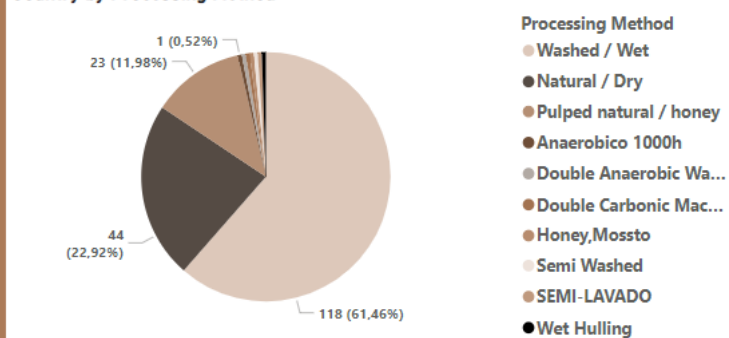
Coffee Quality by Processing Method and Region

Processing Method ● Anaerobico 1000h ● Double Anaerobi... ● Double Carboni... ● Honey,Mossto ● Natural / Dry ● Pulped natural ... ● Semi Washed ● SEMI-LAVA... ● Washed / ... ● Wet Hulling
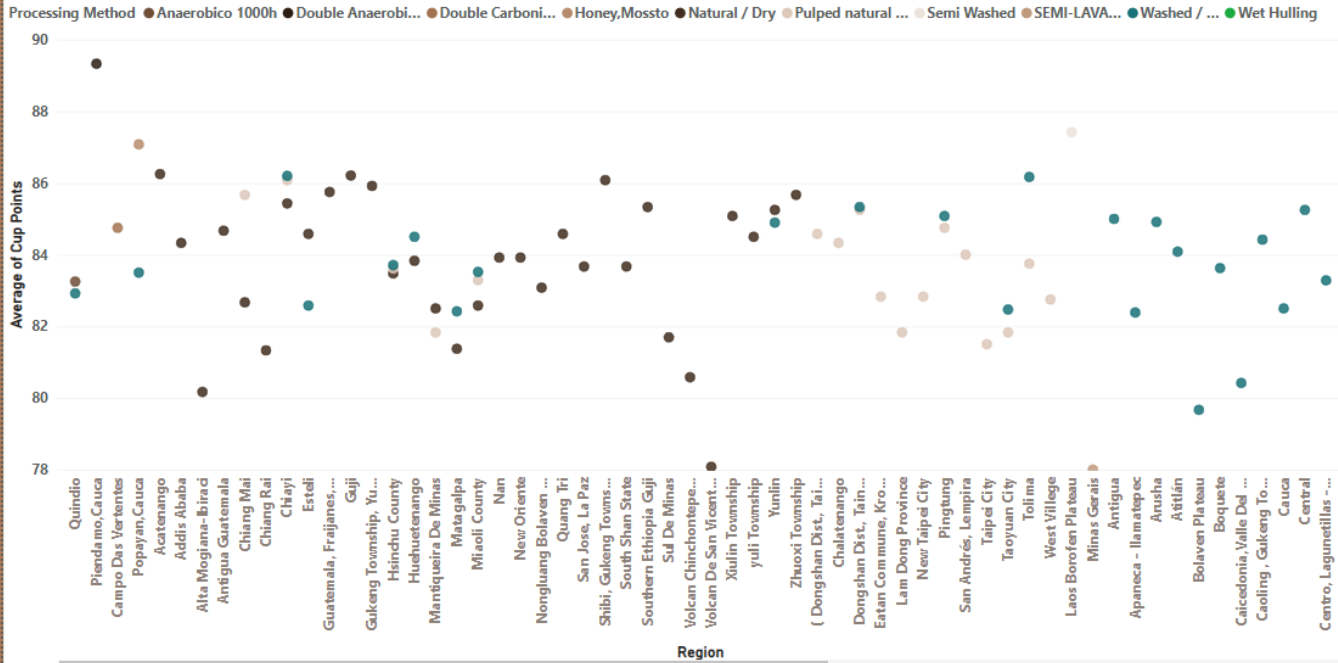
Correlation Between Sensory Attributes and Coffee Quality

Key influencers   Top segments

What influences Flavor to   Increase   ?

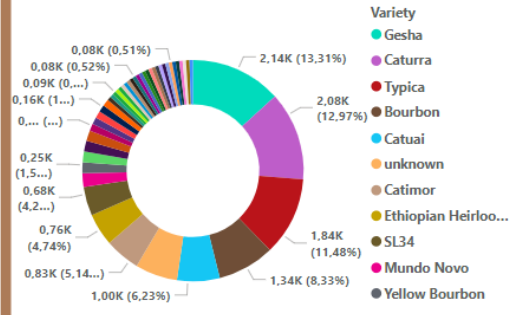When...   ...the average of Flavor increases by

Sum of Cup Points goes up 1.74   0,26

Sensiry Attribute

Attribute
- ☐ Acidity
- ☐ Aftertaste
- ■ Aroma
- ☐ Balance
- ☐ Body
- ☐ Clean Cup
- ☐ Flavor
- ☐ Overall
- ☐ Sweetness
- ☐ Uniformity

Total Cup Points by Variety

Variety
- ● Gesha
- ● Caturra
- ● Typica
- ● Bourbon
- ● Catuai
- ● unknown
- ● Catimor
- ● Ethiopian Heirloo...
- ● SL34
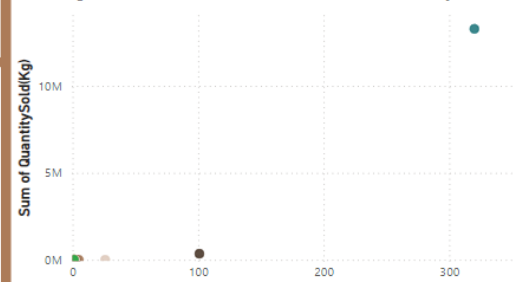- ● Mundo Novo
- ● Yellow Bourbon

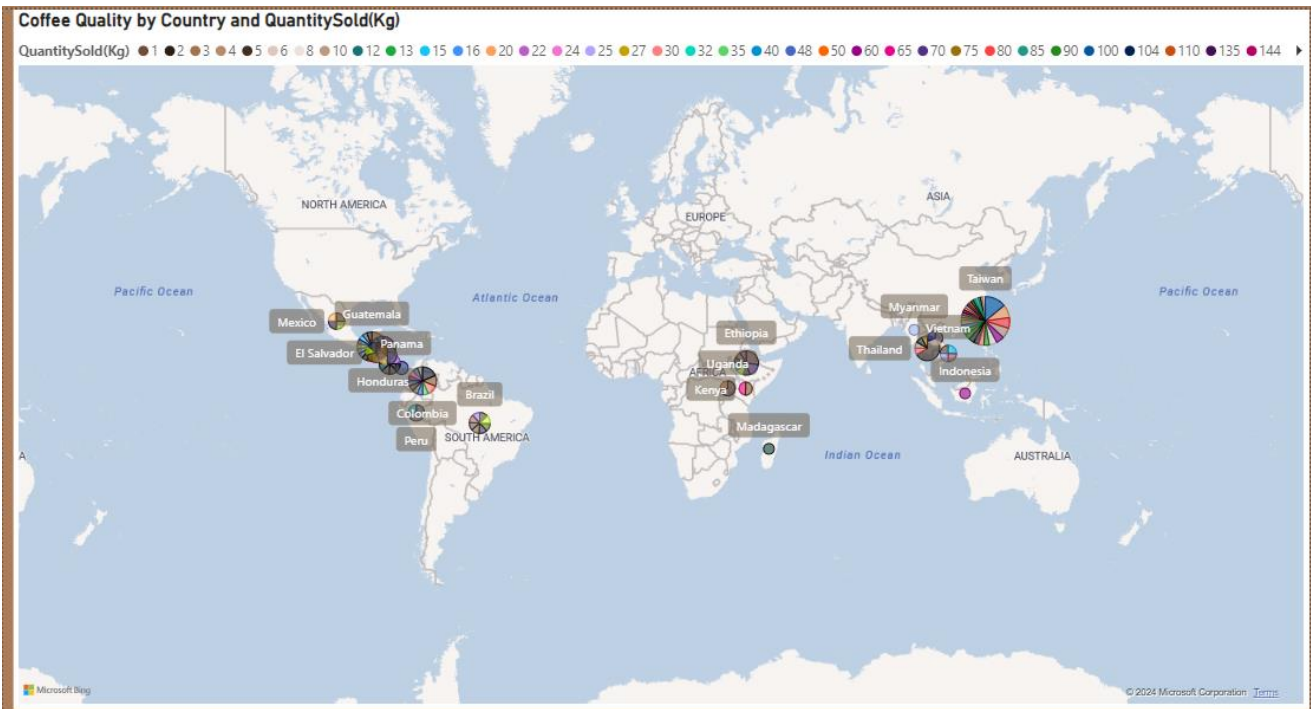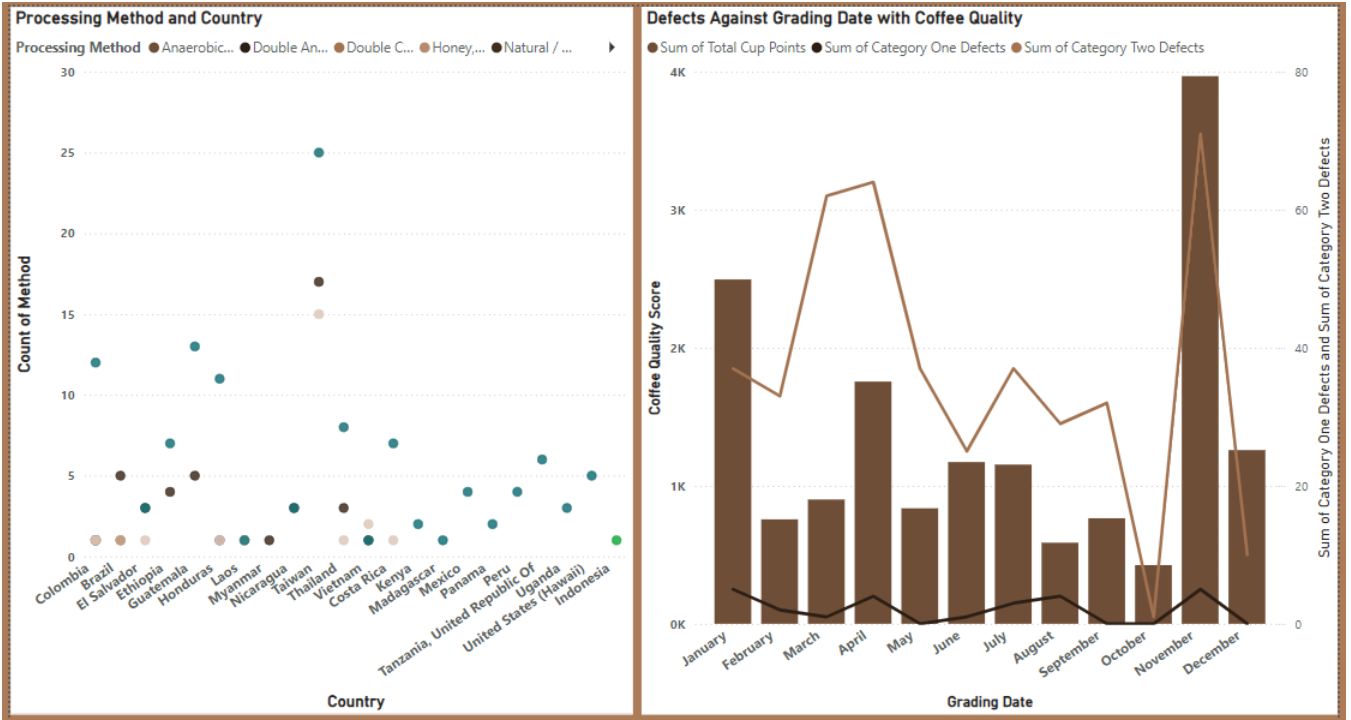## Category Two Defects and Category One Defects by Grading Date

● Sum of Category Two Defects ● Sum of Category One Defects
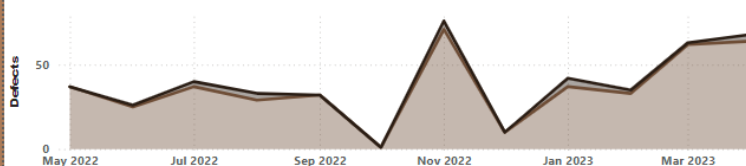
Total Defects and Sum of QuantitySold(Kg) by Processing Method

Processing Method ● Anaerobi... ● Double ... ● Double ... ● Honey,... ▶

Processing Method and Country

Processing Method ● Anaerobic... ● Double An... ● Double C... ● Honey,... ● Natural / ... ▶

Count of Method

30
25
20
15
10
5
0

Country: Colombia, Brazil, El Salvador, Ethiopia, Guatemala, Honduras, Laos, Myanmar, Nicaragua, Taiwan, Thailand, Vietnam, Costa Rica, Kenya, Madagascar, Mexico, Panama, Peru, Tanzania, United Republic Of, Uganda, United States (Hawaii), Indonesia

Defects Against Grading Date with Coffee Quality

● Sum of Total Cup Points ● Sum of Category One Defects ● Sum of Category Two Defects

Coffee Quality Score

4K
3K
2K
1K
0K

Sum of Category One Defects and Sum of Category Two Defects

80

Grading Date: January, February, March, April, May, June, July, August, September, October, November, December

Coffee Quality by Country and QuantitySold(Kg)

QuantitySold(Kg) ●1 ●2 ●3 ●4 ●5 ●6 ●8 ●10 ●12 ●13 ●15 ●16 ●20 ●22 ●24 ●25 ●27 ●30 ●32 ●35 ●40 ●48 ●50 ●60 ●65 ●70 ●75 ●80 ●85 ●90 ●100 ●104 ●110 ●135 ●144 ▶

NORTH AMERICA — Pacific Ocean — Atlantic Ocean — EUROPE — ASIA — Pacific Ocean

Mexico, Guatemala, El Salvador, Panama, Honduras, Colombia, Peru, Brazil, SOUTH AMERICA

Ethiopia, Uganda, Kenya, Madagascar

Myanmar, Taiwan, Vietnam, Thailand, Indonesia

Indian Ocean — AUSTRALIA

Microsoft Bing

© 2024 Microsoft Corporation Terms

Insights

- **Top Countries**: Guatemala, Ethiopia, and Colombia lead in coffee quality, with Guatemala having the highest sum of cup points.

- **Average Scores**: The average cup points are 83.73, indicating a generally high quality of coffee across the dataset.

- **Quality Tiers**: The pie chart shows a distribution of coffee quality tiers, with a significant portion being premium quality.

- **Processing Methods**: The "Double Anaerobic Washed" method is highlighted as the highest-rated processing method.

- **Countries:** Ethiopia routinely earns high quality rankings for coffee, making it stand out as the top country by average cup points.

- The distribution of coffee quality across various places is depicted in the map visualization, which raises the possibility that some regions are linked to higher-quality coffee because of factors like height, a good environment, or customary farming methods.

- According to the bar chart, the sensory criteria with the greatest overall scores are Clean Cup, Uniformity, and Sweetness. These characteristics most likely have a significant impact on the overall quality of the cup.

- The scatter plot demonstrates how various processing techniques affect coffee quality differently depending on the location. The association suggests that certain processing techniques, such as anaerobic and natural/dry, may be responsible for higher cup points in particular geographical areas.

- The association between flaws and coffee quality throughout time is shown in the bar chart. The data supports the idea that faults have a detrimental effect on

the overall quality of coffee because months with higher defect counts typically have lower total cup points.

- The line graph displays patterns in faults over time, which has an effect on quality ratings.

- While the sensory qualities are not evenly distributed throughout the datasets, the most highly regarded characteristics that affect coffee quality are flavor, aroma, and balance. This implies that raising these characteristics may lead to improved ratings for coffee quality overall.

- The most popular and potentially most profitable variations are shown in the charts with QuantitySold(Kg) and Total Cup Points by Variety. Sorts like Typica, Bourbon, and Gesha, for instance, are very prized.

- While some processing techniques, such Washed and Natural/Dry, predominate in coffee production, the impact of these techniques on quality differs by area, as illustrated by the pie charts.

Conclusion

This project effectively examined the elements that affect coffee quality, paying particular attention to sensory characteristics, processing techniques, and geographic locations. Important conclusions were obtained from the data analysis, demonstrating that certain processing techniques and geographical areas, such as Ethiopian coffee and the Double Anaerobic Washed procedure, consistently produce higher quality scores. Furthermore, it was shown that sensory characteristics including flavor and aroma play a significant role in overall quality. The analysis emphasizes how crucial these elements are to raising coffee quality and directing manufacturing methods.