



# Exploring Coffee Quality Data with Power BI

Presented by Anjali

# ABOUT PROJECT

The Coffee Quality Institute (CQI) is a non-profit organization dedicated to enhancing the quality and value of coffee globally. Established in 1996 and headquartered in California, USA, CQI's mission is to advance coffee quality through various initiatives such as research, training, and certification programs. The organization collaborates with coffee growers, processors, roasters, and other industry stakeholders to elevate quality standards, promote sustainability, and support the growth of the specialty coffee sector.



# OBJECTIVE

- Identify Key Determinants of Coffee Quality
- Assess the Impact of Geographic and Environmental Factors
- Analyze the Relationship Between Processing Methods and Coffee Quality
- Examine Defect Occurrences and Their Impact
- Explore Correlations and Interactions Between Variables
- Provide Actionable Insights for Coffee Producers and Buyers

# TOOLS USED



**POWER BI**

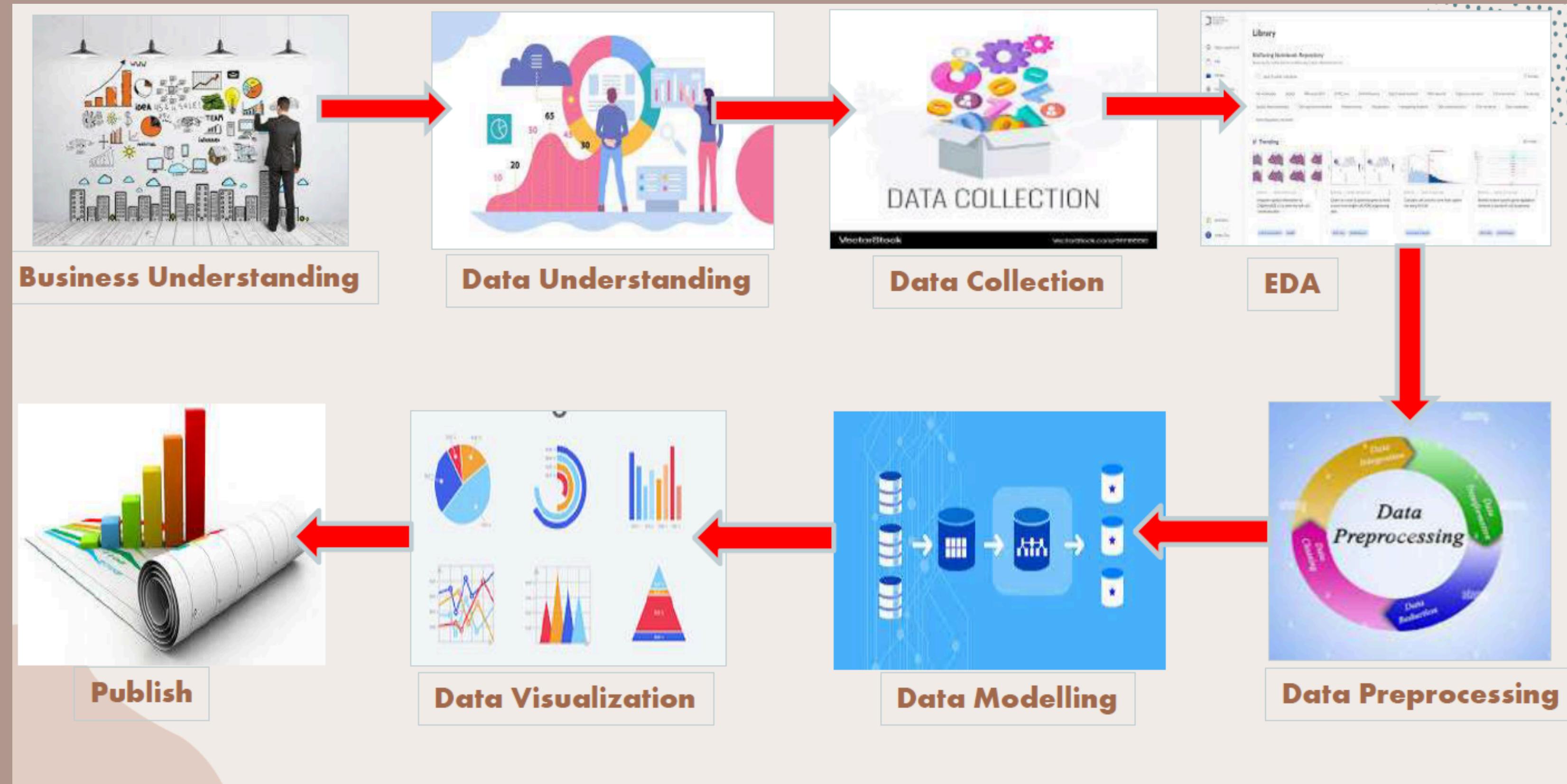


**Microsoft  
Excel**



**Microsoft  
Power Point**

# Project Architecture



# TABLE CONTENT

- EDA
- DATA PREPROCESSING AND SCALING
- DATA MODELING
- VISUALIZATION
- ANALYSIS REPORT
- INSIGHTS

# Exploratory Data Analysis



The EDA process on the coffee quality dataset helps uncover patterns, relationships, and potential anomalies. This foundational understanding is crucial for subsequent steps, such as building predictive models, identifying key drivers of quality, and making informed decisions in the coffee production and purchasing process.

## 1. Understand the Dataset Structure

**Columns:** Identify the columns in the dataset and their data types (e.g., numeric, categorical).

**Key Variables:** Common variables in a coffee quality dataset might include:

**Origin Region/Country:** The region or country where the coffee is grown.

**Variety:** The coffee bean variety (e.g., Arabica, Robusta).

**Processing Method:** The method used to process the coffee (e.g., washed, natural).

**Altitude:** The altitude at which the coffee is grown.

**Total Cup Points:** The overall quality score of the coffee.

**Sensory Attributes:** Scores for aroma, flavor, acidity, body, etc.

**Defects:** Information about any defects in the coffee beans.

## 2. Data Cleaning

**Missing Values:** Identify and handle missing values. For example, you could impute missing values, drop rows/columns with missing data, or flag them for further analysis.

**Outliers:** Detect outliers in numeric variables, especially in sensory attributes and Total Cup Points.

**Data Consistency:** Check for inconsistencies in categorical variables (e.g., spelling errors in "Origin Region" or "Variety").

## 3. Analysis

### A. Distribution of Total Cup Points

**Histogram:** Create a histogram to visualize the distribution of the Total Cup Points. This will help you understand the spread of coffee quality scores.

**Visual:** Use visual card for counts of total cup points.

### B. Analysis of Categorical Variables

**Bar Charts:** For categorical variables like Variety, Processing Method, and Origin Region, create bar charts to show the frequency or count of each category.

**Variety:** Display the count of coffee samples by variety (e.g., Arabica, Robusta).

**Processing Method:** Show the distribution of processing methods (e.g., washed, natural, honey).

**Origin Region:** Display the number of samples from each origin region.

### C. Relationship Between Total Cup Points and Sensory Attributes

Tree map : explore the relationship between Total Cup Points and various sensory attributes such as Aroma, Flavor, Acidity, etc.

column: Sensory attribute (e.g., Aroma). vs Total cup points

### D. Comparing Total Cup Points Across Categories

Bar graphs/line charts : compare the distribution of Total Cup Points across different categories (e.g., Variety, Processing Method, Origin Region).

X-Axis: Categorical variable (e.g., Processing Method).

Y-Axis: Total Cup Points.

# Dashboard Layout

**Title:** "Exploratory Data Analysis of Coffee Quality Dataset"

**Sections:**

**Distribution Analysis:** Display histograms and bar charts showing the distribution of Total Cup Points and categorical variables.

**Correlation Analysis:** Include line charts for analysis.

**Geographical Insights:** Use heatmaps, Tree maps visuals to present regional data.

**Interactive Elements:** Provide slicers and interactive visuals for user-driven exploration.

**Conclusions and Insights:** Summarize key findings from the EDA, including patterns, correlations, and outliers.



**Data preprocessing in Power Query Editor prepares your coffee quality dataset for effective analysis and visualization in Power BI. By cleaning, transforming, and structuring the data, you ensure that your subsequent analysis is based on accurate, consistent, and relevant data. This leads to more meaningful insights and better decision-making.**

## **1. Load Data into Power Query Editor**

### **2. Initial Data Cleaning**

#### **A. Remove Unnecessary Columns**

#### **B. Rename Columns**

#### **C. Handle Missing Data**

#### **D. Data Type Conversion**

#### **E. Create New Calculated Columns/Measures**

### **3. Data Transformation**

#### **A. Unpivot Data**

#### **B. Pivot Data**

### **4. Data Filtering**

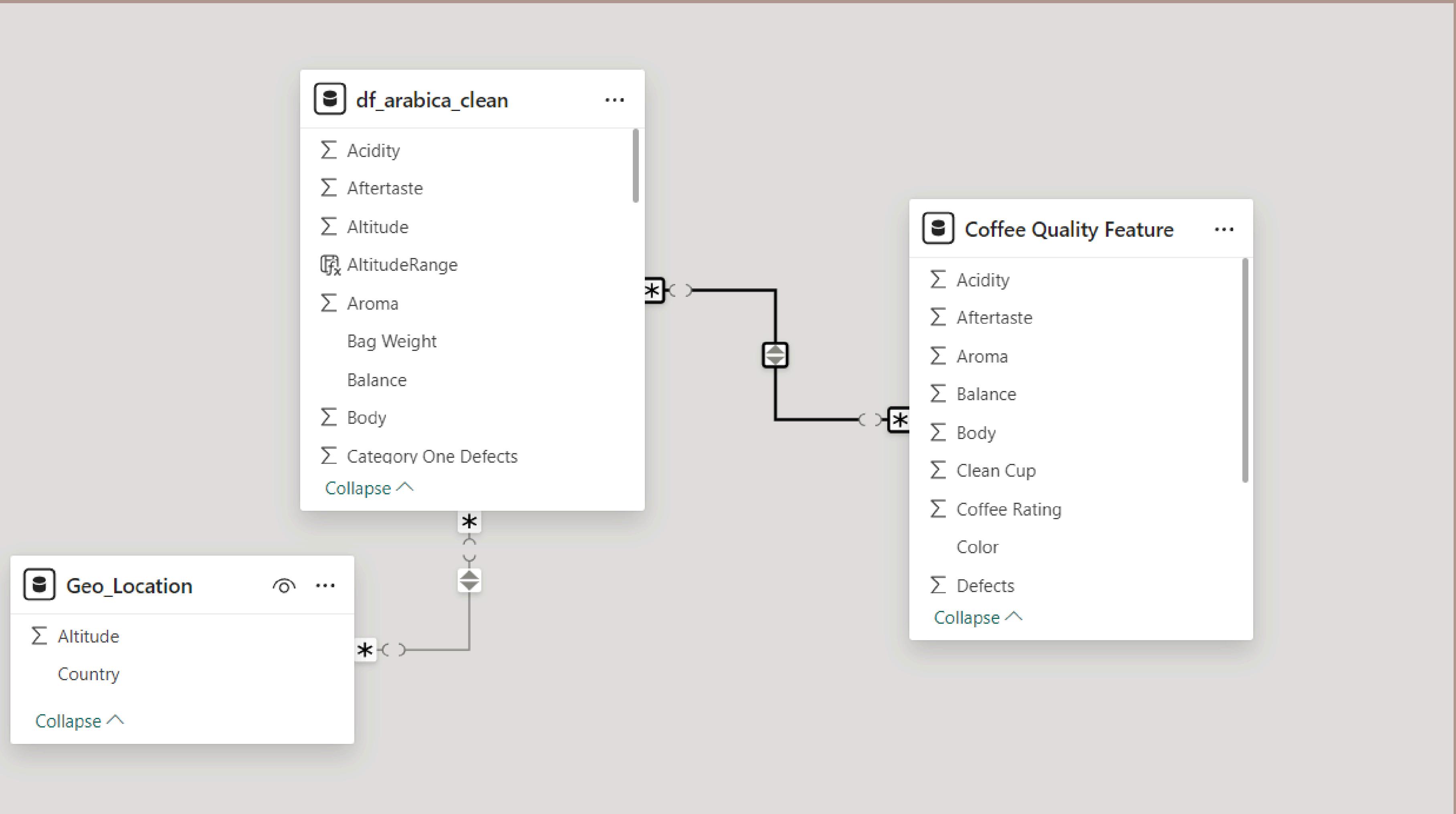
### **5. Remove Duplicates**

### **6. Data Validation and Final Checks**

### **7. Close & Apply**

# Data Modelling!





# visualization



**Heatmap : correlations between all sensory attributes and overall quality.**

Aroma	Average of Acidity	Average of Aftertaste	Average of Body	Average of Balance	Flavor	Average of Sweetness	Average of Uniformity	Coffee Quality	Average of Total
9	7.00	6.71	7.13	6.92	7.08	10.00	10.00	6.88	79.00
7	7.08	7.04	7.27	7.15	7.17	10.00	10.00	7.11	80.09
6	7.30	7.21	7.17	7.21	7.25	10.00	10.00	7.17	80.75
2	7.49	7.42	7.56	7.48	7.42	10.00	10.00	7.47	82.25
5	7.53	7.38	7.45	7.43	7.50	10.00	10.00	7.44	82.27
7	7.67	7.17	7.83	7.50	7.33	10.00	10.00	7.33	82.50
0	7.51	7.45	7.54	7.56	7.58	10.00	10.00	7.49	82.73
7	7.64	7.48	7.57	7.55	7.67	10.00	10.00	7.59	83.16
1	7.75	7.62	7.73	7.70	7.75	10.00	9.96	7.75	84.06
5	7.75	7.68	7.73	7.74	7.83	10.00	10.00	7.87	84.35
2	7.83	7.77	7.73	7.78	7.92	10.00	9.93	7.85	84.64
9	7.89	7.82	7.90	7.84	8.00	10.00	10.00	7.92	85.26
2	7.92	7.88	7.79	7.83	8.08	10.00	10.00	7.91	85.43
1	8.11	7.98	7.92	8.06	8.17	10.00	10.00	8.11	86.38
7	8.04	8.09	7.92	8.00	8.25	10.00	10.00	8.04	86.50
3	8.17	8.08	7.92	8.17	8.42	10.00	10.00	8.33	87.42

83.81

Average of Total Cup Points

Harvest Year

All

Region

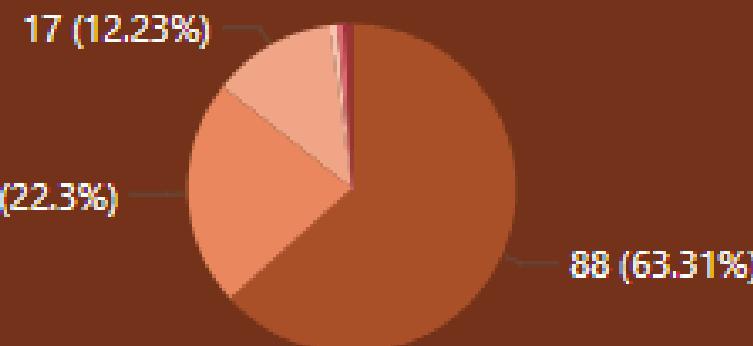
All

7.70

Average of Coffee Rating

# Impact of Processing Method on Coffee Quality

Count by Processing Method



Processing Method

Washed / Wet

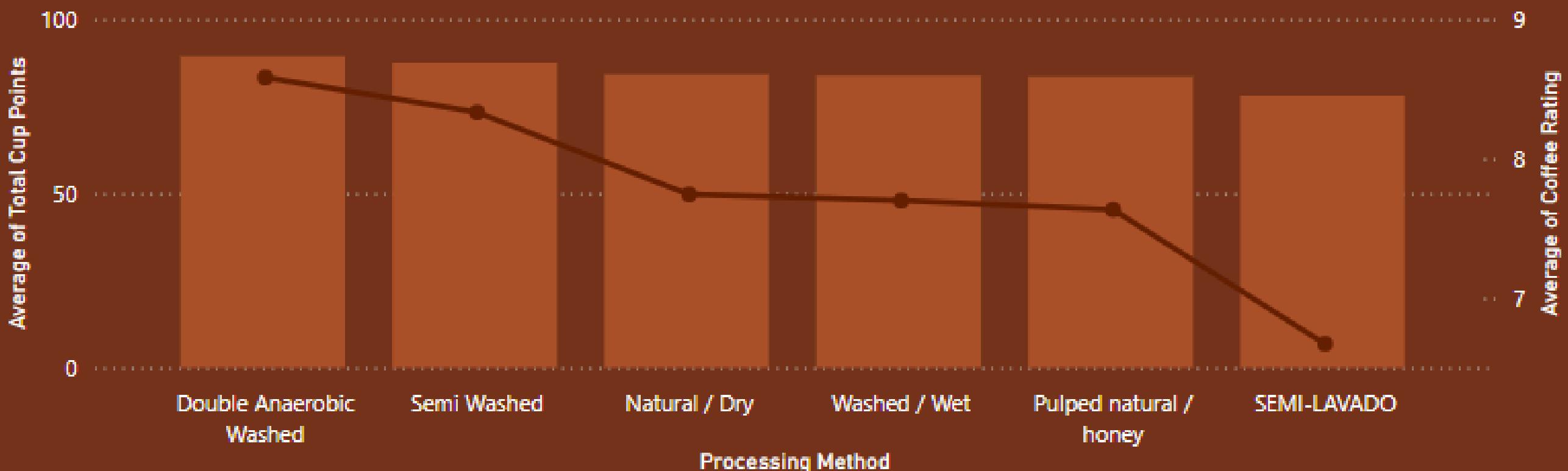
Natural / Dry

Pulped natural / honey

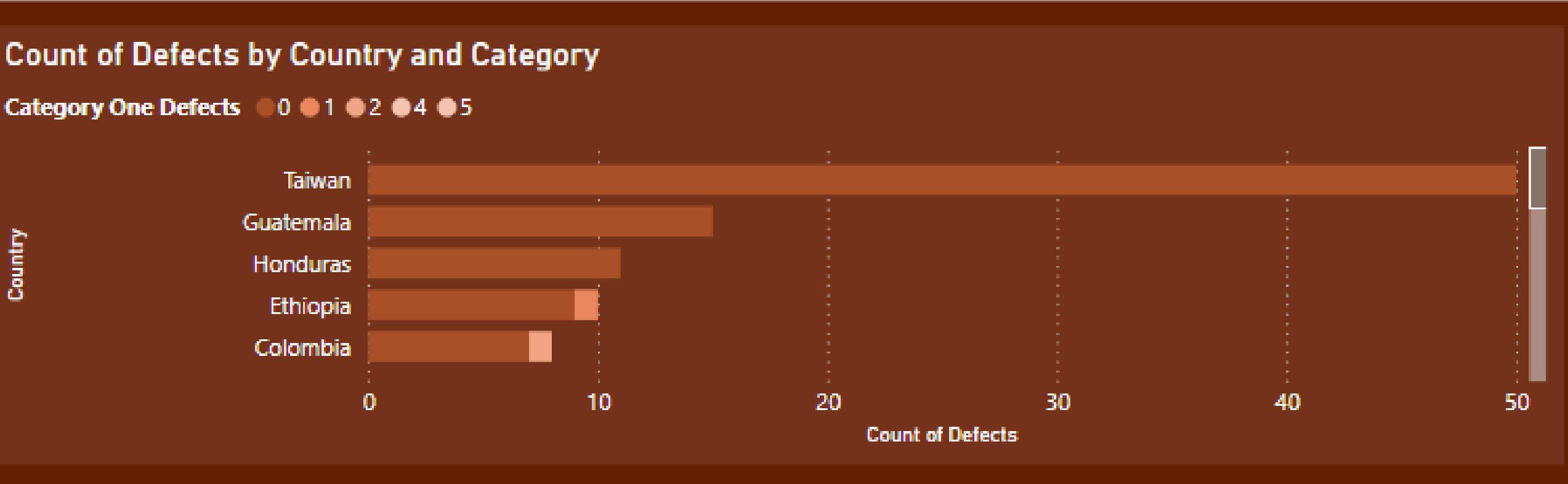
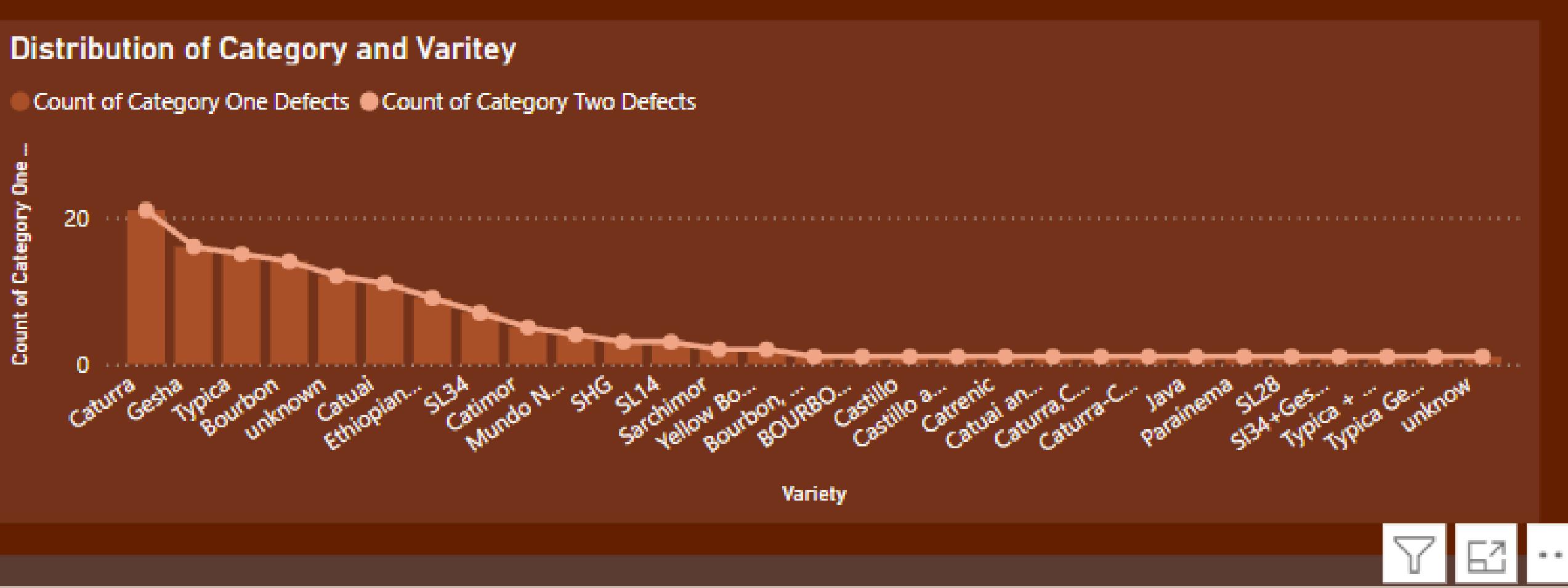
Double Anaerobic Washed

Average Total Cups point and Rating by Processing Method

Average of Total Cup Points ● Average of Coffee Rating



# Impact of Defects on Coffee Quality

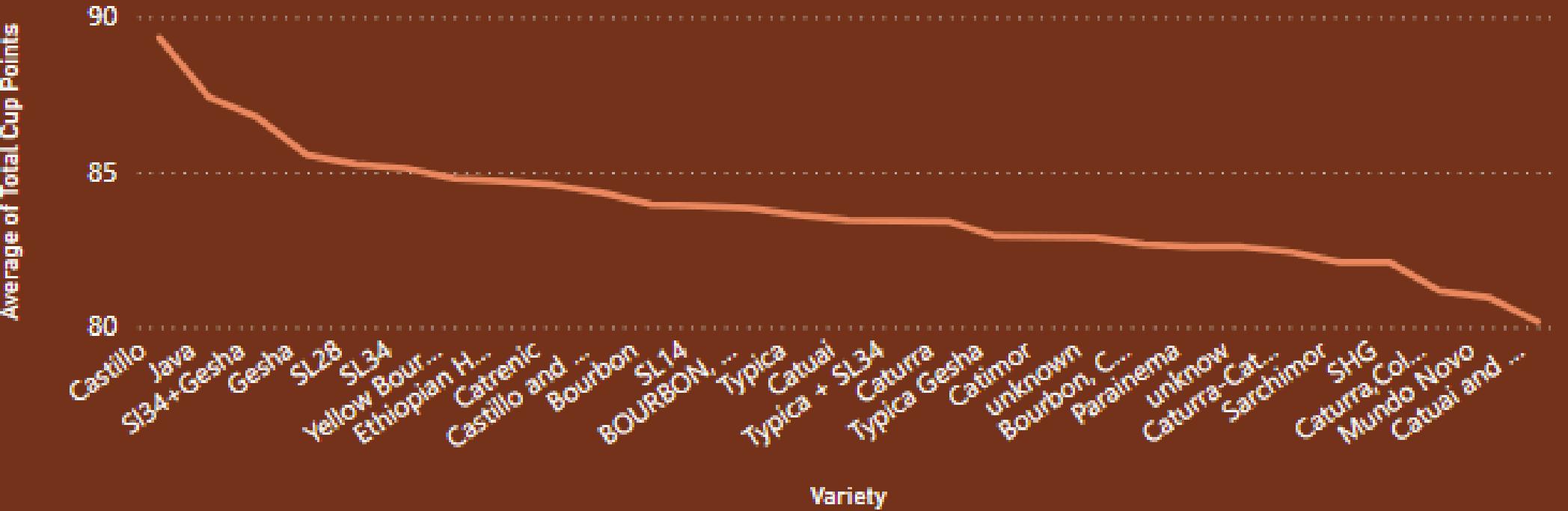


139

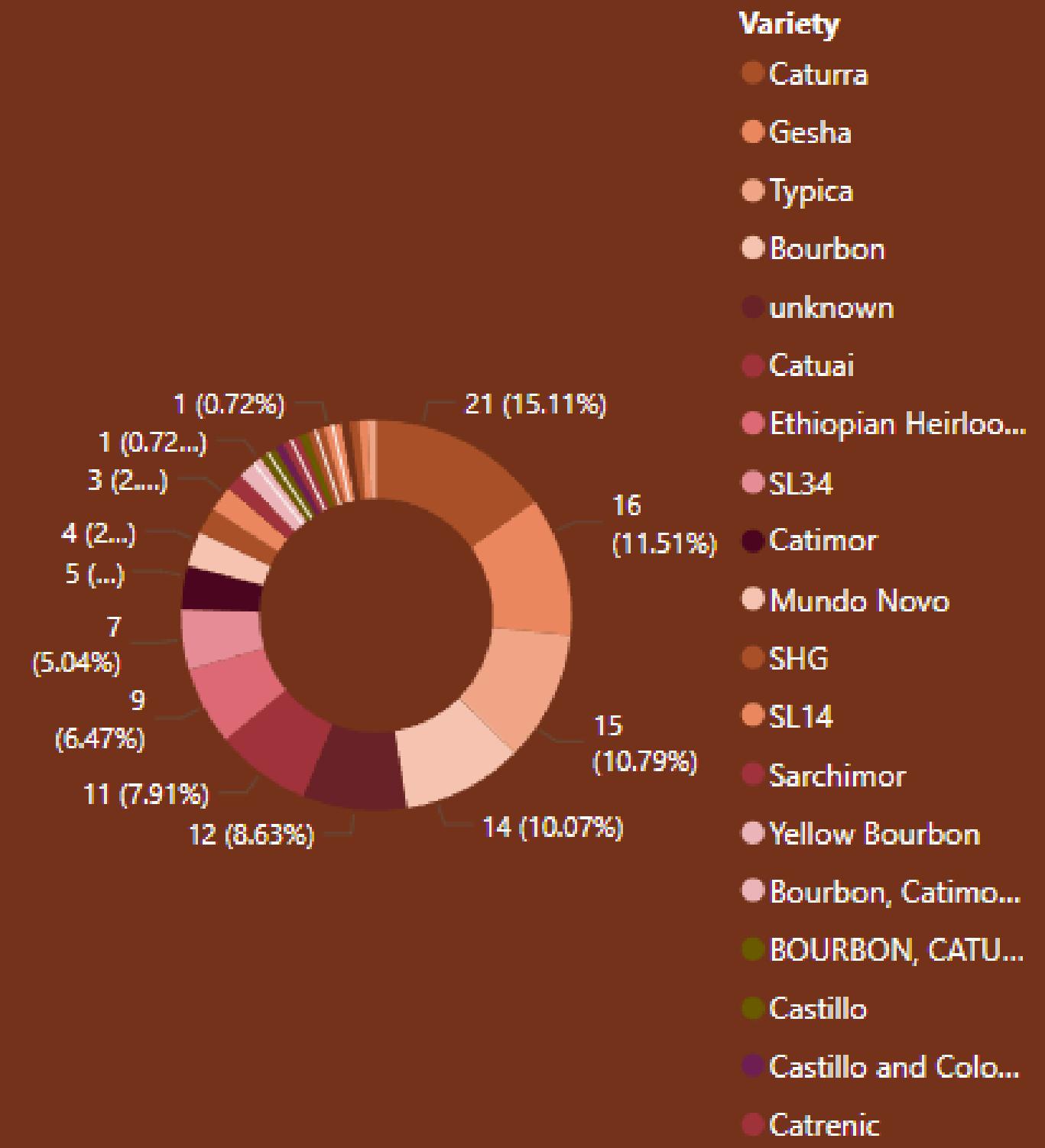
Count By Variety

# Impact of Variety on Coffee Quality

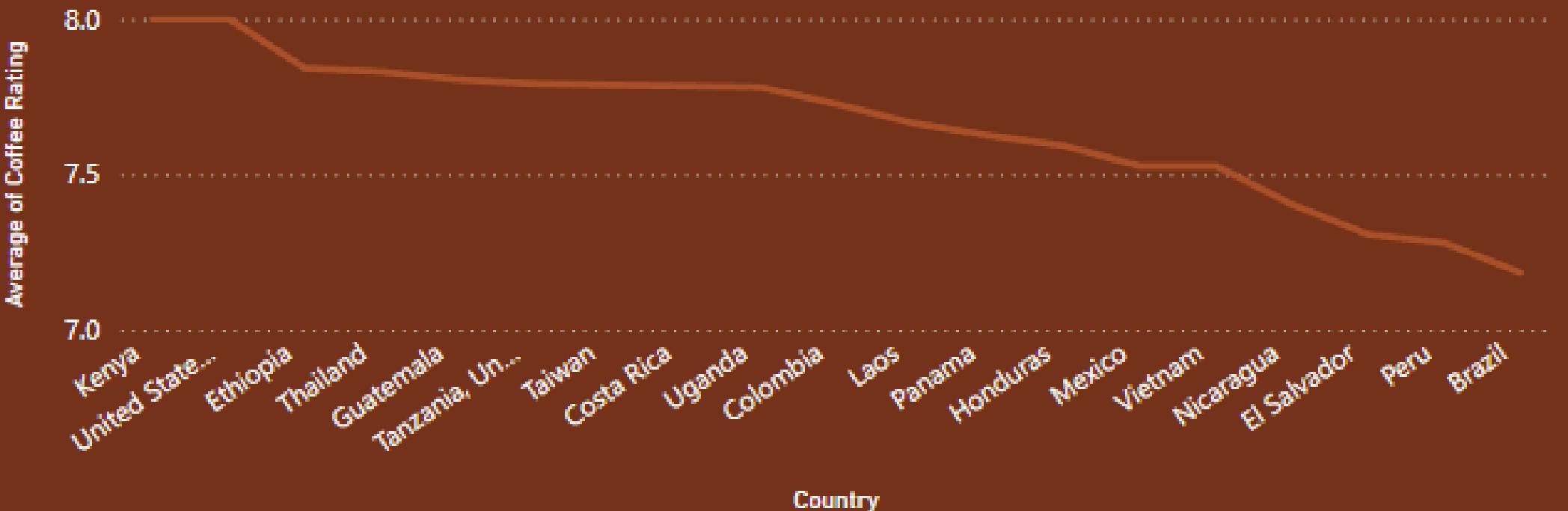
## Average of Total Cup Points by Variety



## Count by Variety



## Average of Coffee Rating by Country



# Impact of Origin Regions on Coffee Quality

Processing Method

All

89.33

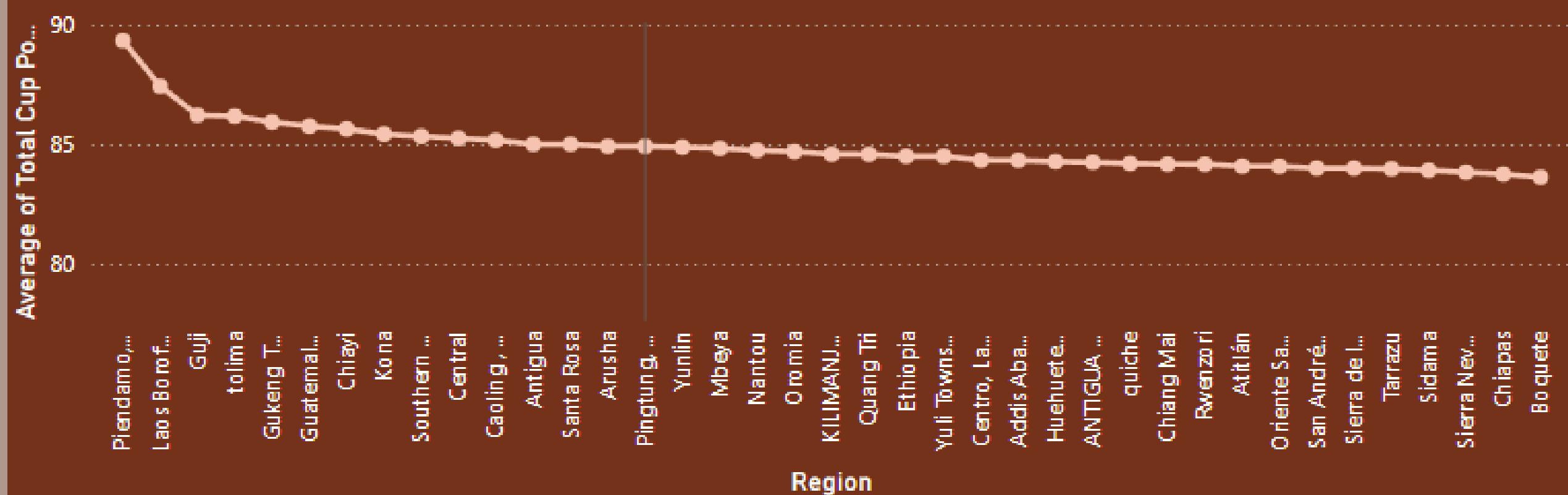
MaxCupPointsByRegion

78.00

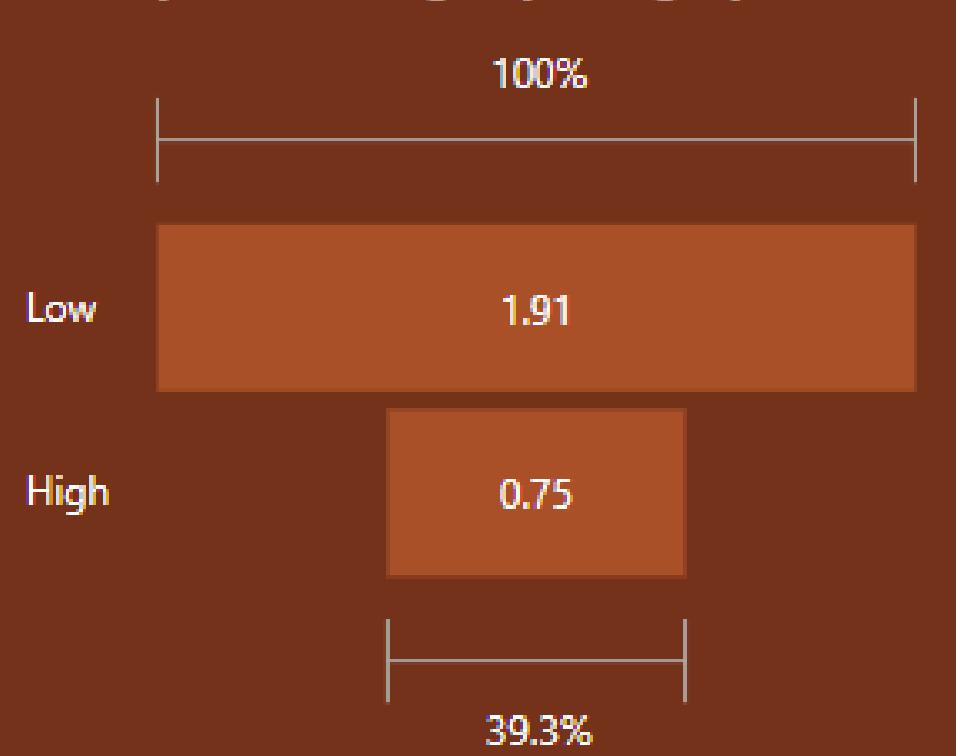
MinCupPointsByRegion



Average of Total Cup Points by Region

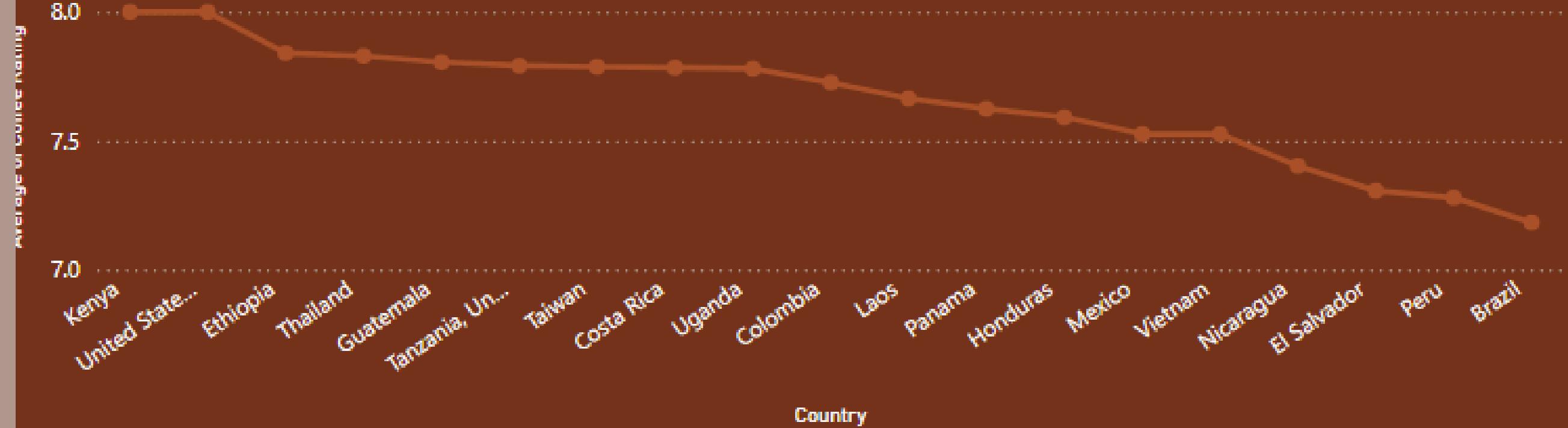


Quality Score Range by category



# Impact of Origin Country on Coffee Quality

## Average of Coffee Rating by Country



Moisture Percentage

0.00 13.50

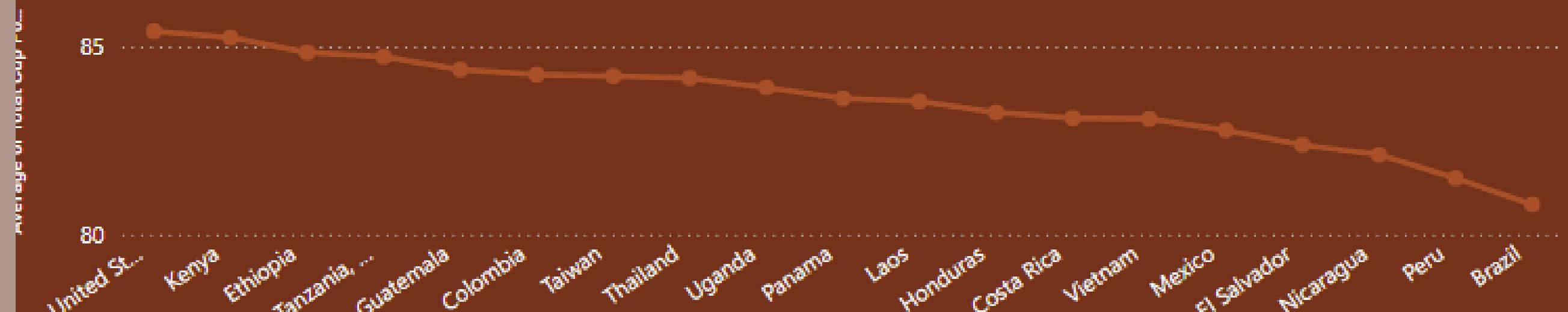
Harvest Year

All

Grading Year

11-05-2022 28-04-2023

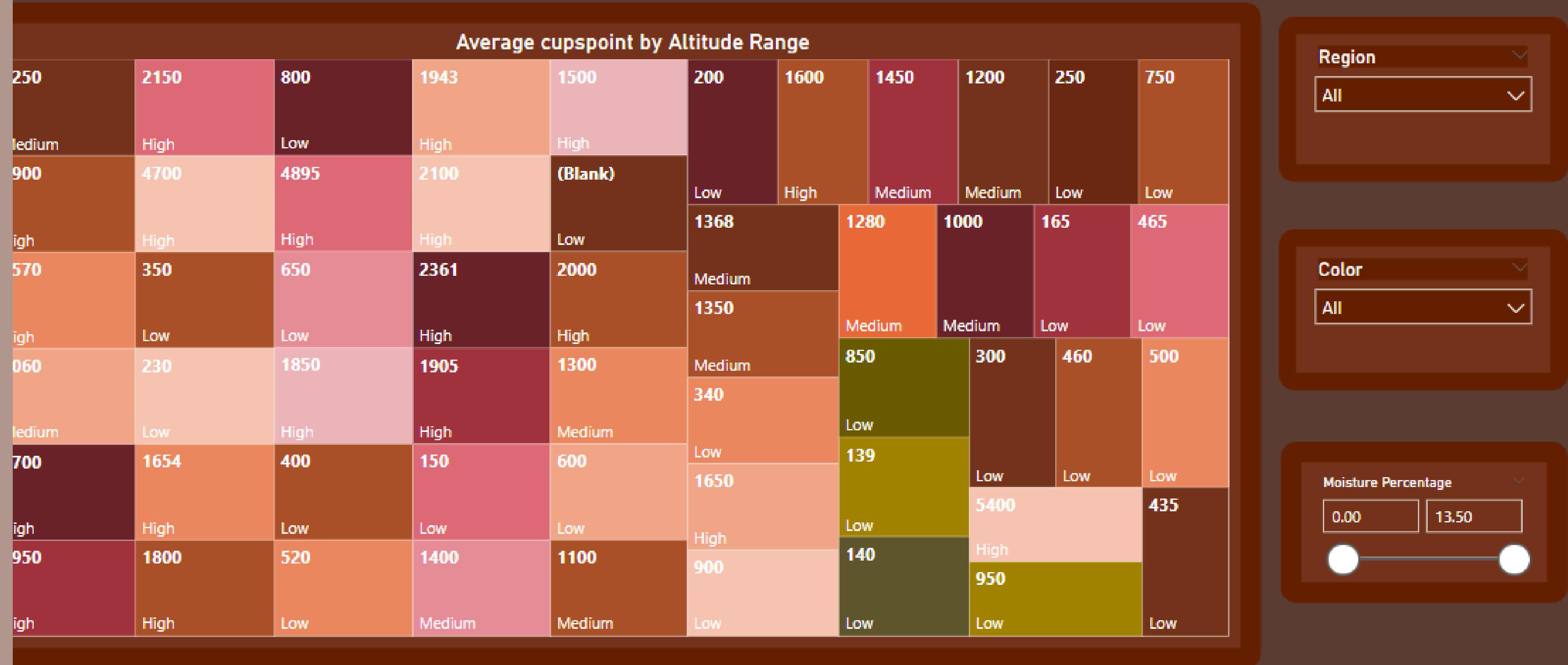
## Average cup points by country



In-Country Partner

All

# Impact of Altitude on Coffee Quality



# insights

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

Aroma, flavor, Acidity, Body, Balance, Sweetness, Aftertaste, Uniformity are the key attributes. A coffee with a high total cup score will typically excel across multiple sensory dimensions, indicating a well-rounded and high-quality coffee.

- **Is there a correlation between processing methods, origin regions, and coffee quality scores?**

Yes, there is often a correlation between processing methods, origin regions, and coffee quality scores. The processing method can enhance or detract from the inherent qualities of the coffee depending on how well it is executed and the characteristics of the region. Coffees from certain regions or wet/washed process method tend to score higher due to the favourable interplay of these factors. Understanding these correlations can help in predicting and optimizing coffee quality.

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

Category 2 defects are more compared to category 1 defects.

- Caturra variety has more defects.
- country Taiwan has more defects and Honduras has low defects.
- Harvest year 2021/2022 has more defected beans.

harvesting practices for specific varieties or countries to reduce defect occurrences and improve overall quality.

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

The Total Cup Points represent the culmination of how well these various factors work together. A coffee that excels in most or all of these areas will achieve a high score, while one that suffers in one or more areas will see its score reduced accordingly. Variety has high Cup Points. - Castillo, Java& Sl34+Gesha are top 3 Quality when compared to Cup Points. - Ethiopia, USA, Thailand, Gautemela, Kenya are the top 5 Countries.

# Conclusion

## Final Outcome:

The ultimate goal of this analysis is to develop a comprehensive understanding of what makes high-quality coffee and to use this information to improve coffee production, sourcing, and purchasing decisions. This dataset can also help coffee producers and industry stakeholders benchmark their products against industry standards and identify areas for improvement.

Thank you