EXPLORING COFFEE QUALITY DATA WITH POWER BI

---

1. Introduction

Coffee quality assessment is a multi-factor evaluation involving sensory attributes, processing methods, geographical influences, and defect levels. This project analyses a global coffee dataset using Power BI to understand:

- **What determines overall coffee quality?**

- **How do sensory attributes influence Total Cup Points?**

- **How do processing methods, origin, and altitude affect quality?**

- **How do defects impact coffee grades?**

- **How do variables interact collectively to influence Total Cup Points?**

The study integrates DAX measures, interactive dashboards, Key Influencer analysis, and Decomposition Trees to uncover deep insights.

---

2. Dataset Overview

The dataset includes:

2.1 Sensory Attributes (10 attributes)

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

Note: **Uniformity, Sweetness, and Clean Cup maintain a constant score of 10 for almost all samples — these do not meaningfully influence quality variation and were excluded from KPI-based comparisons.**

2.2 Non-Sensory Attributes

- **Region, Processing Method, Variety, Color, Moisture %,Altitude, Quakers, Defects (Category 1 and Category 2)**
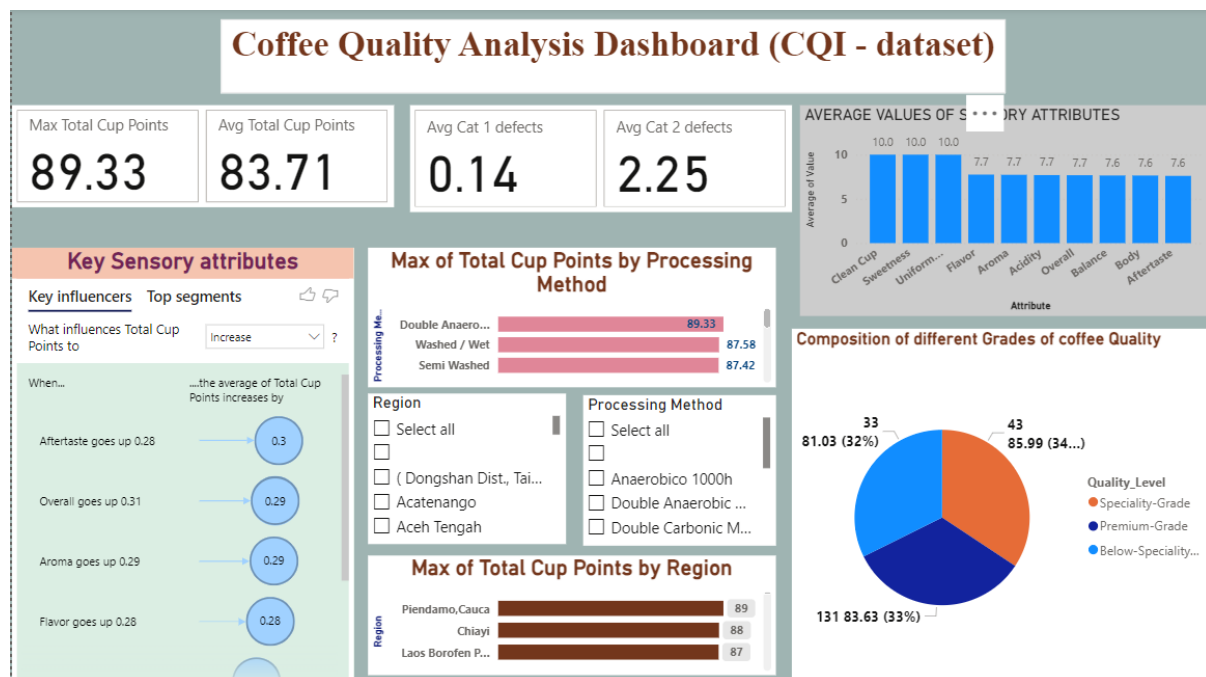
2.3 Output Variable

- **Total Cup Points (primary indicator of coffee quality)**

---

3. Methodology

3.1 Power BI Visualizations Used:

- Key Influencers
- Decomposition Trees
- Scatter plots
- Bar & Line charts

---

4. Insights and Analysis

Objective 1: Key Determinants of Coffee Quality as Evaluated Through Sensory Attributes

1. Overview

The CQI coffee quality evaluation framework uses 10 sensory attributes to determine the Total Cup Points.

These are:

In this dataset, Sweetness, Uniformity, and Clean Cup were constant at 10/10 for almost all coffee samples. Because they show no variation, Power BI correctly excluded them from influencer modelling—they do not contribute to explaining differences in Total Cup Points.

Thus, the analysis focuses on the seven sensory attributes that actually vary and drive score differences.

2.1 Key Influencers for Increasing Total Cup Points
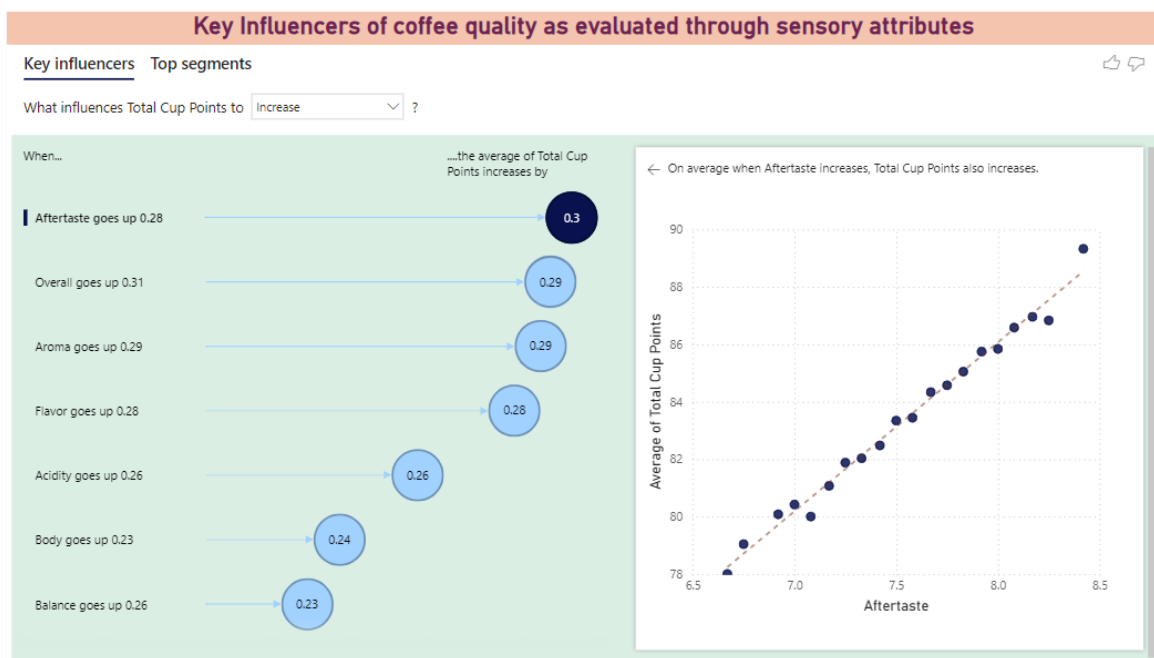
Based on Power BI's Key Influencer model, the sensory attributes that most significantly increase Total Cup Points are:

Primary Drivers (Strongest Influence)

Aftertaste ↑ 0.30

Overall ↑ 0.29

Aroma ↑ 0.29

Flavor ↑ 0.28

These four attributes show the strongest positive linear relationship with Total Cup Points.
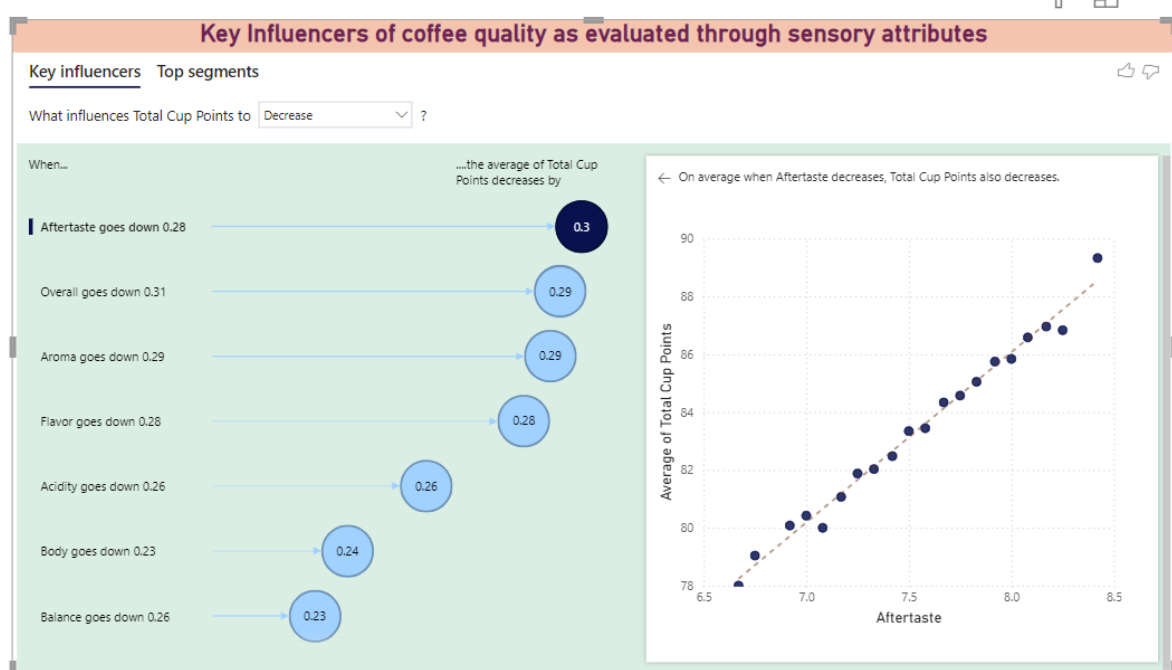
Secondary Drivers (Moderate Influence)

Acidity ↑ 0.26

Body ↑ 0.24

Balance ↑ 0.23

Interpretation

Coffees with high scores in aftertaste, aroma, flavor, and overall consistently achieve higher Total Cup Points.

2.2 Key Influencers for Decreasing Total Cup Points

When the sensory attributes decline, the following show the strongest effect in lowering Total Cup Points:

Aftertaste ↓ −0.30

Overall ↓ −0.29

Aroma ↓ −0.29

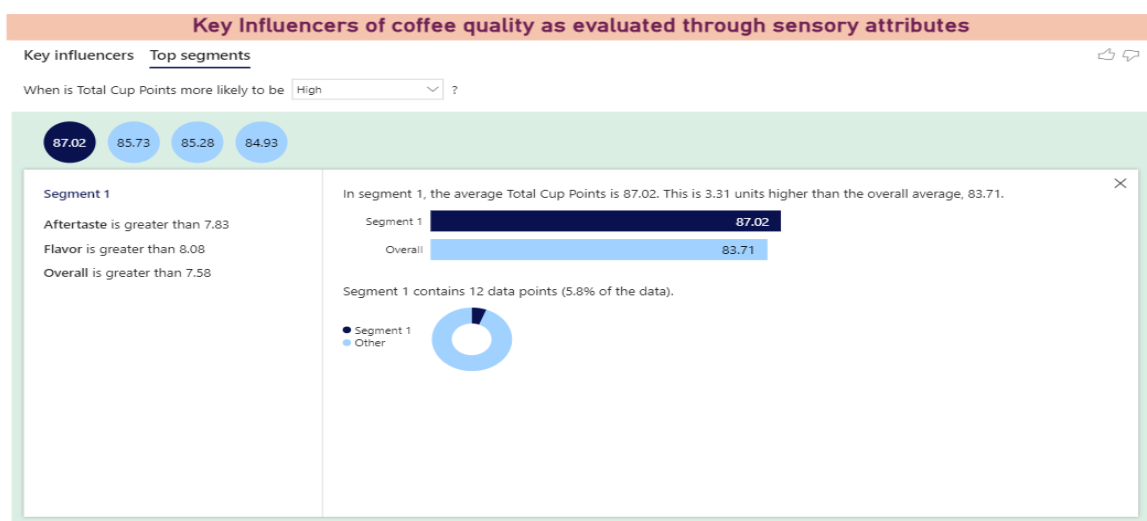Flavor ↓ −0.28

Acidity ↓ −0.26

Body ↓ −0.24

Balance ↓ −0.23

Interpretation

Declines in aftertaste, aroma, and flavor are the clearest early indicators of quality deterioration.

3. Top Segments Analysis (High-Score Conditions)

Power BI detected clusters ("segments") of sensory conditions that consistently produce above-average cup scores.

Segment 1 (Highest Scoring Profile)

- **Aftertaste > 7.83**

- **Flavor > 8.08**

- **Overall > 7.58**

- **Average Score: 87.02 (3.31 points above global mean)**

Meaning:

Coffees with very strong aftertaste and flavor consistently fall into the highest quality tier.

Segment 2 (High Quality Cluster)
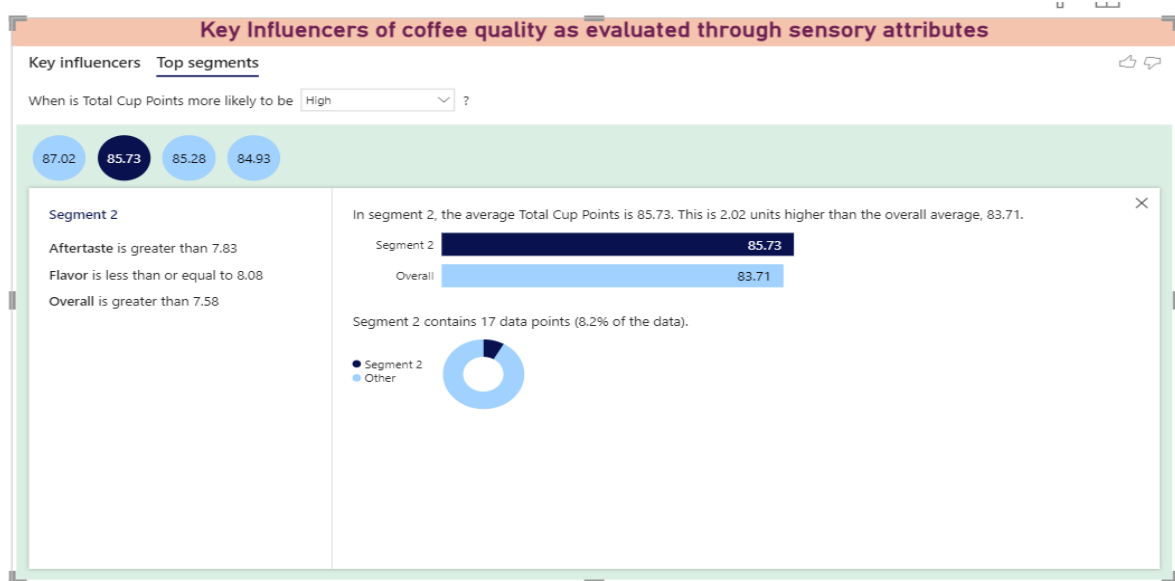
- Aftertaste > 7.83

- Flavor ≤ 8.08

- Overall > 7.58

- Average Score: 85.73

Meaning:

Even when flavor is slightly lower, strong aftertaste ensures high scoring potential.

Segment 4
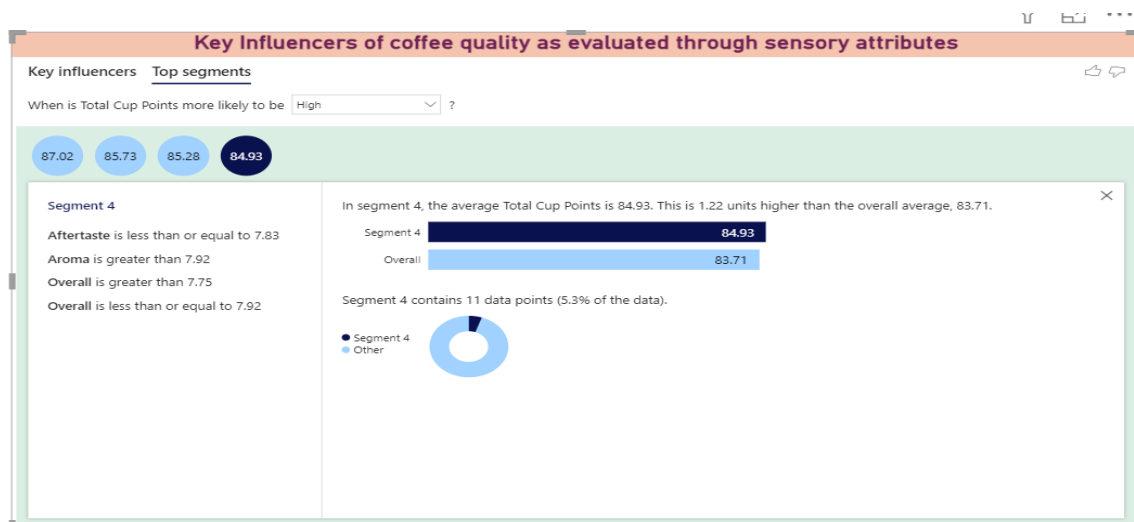
- Aftertaste ≤ 7.83

- Aroma > 7.92

- Overall > 7.75

- Average Score: 84.93

Meaning:

Aroma acts as a compensating sensory factor when aftertaste is below peak levels.

4. Top Segments Analysis (Low-Score Conditions)

Low Score Segment 1
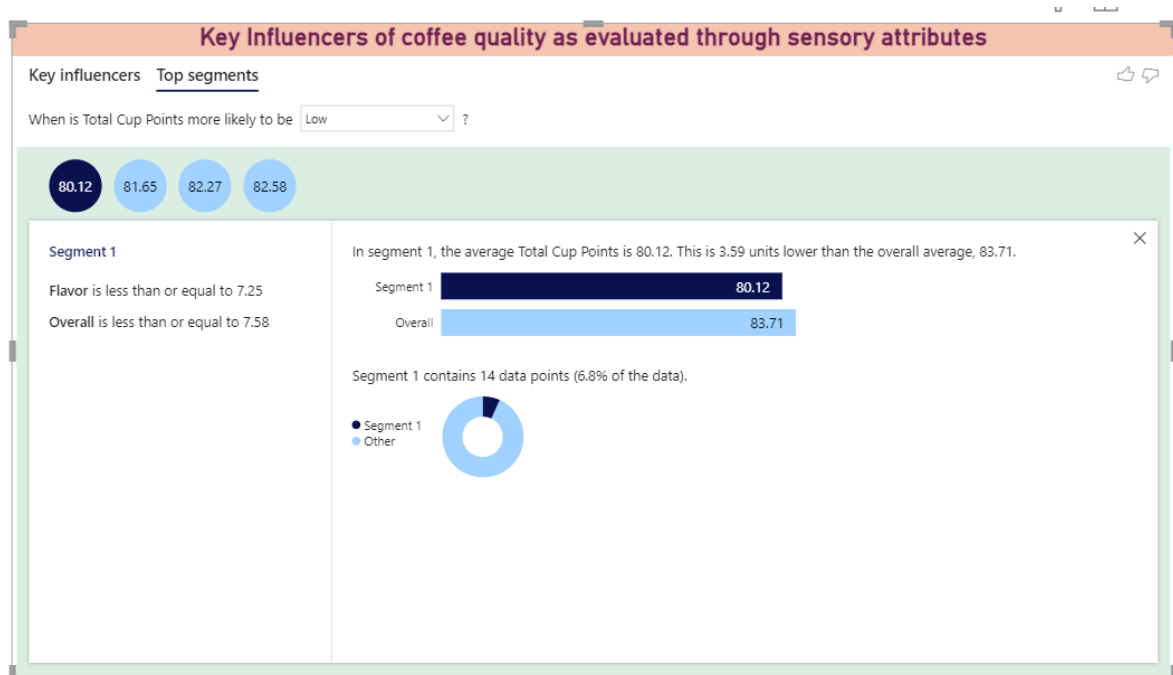
- Flavor ≤ 7.25

- Overall ≤ 7.58

- Average Score: 80.12 (3.59 points below global mean)

Meaning:

Flavor and overall impression are the weakest attributes in low-quality coffees.

Low Score Segment 2
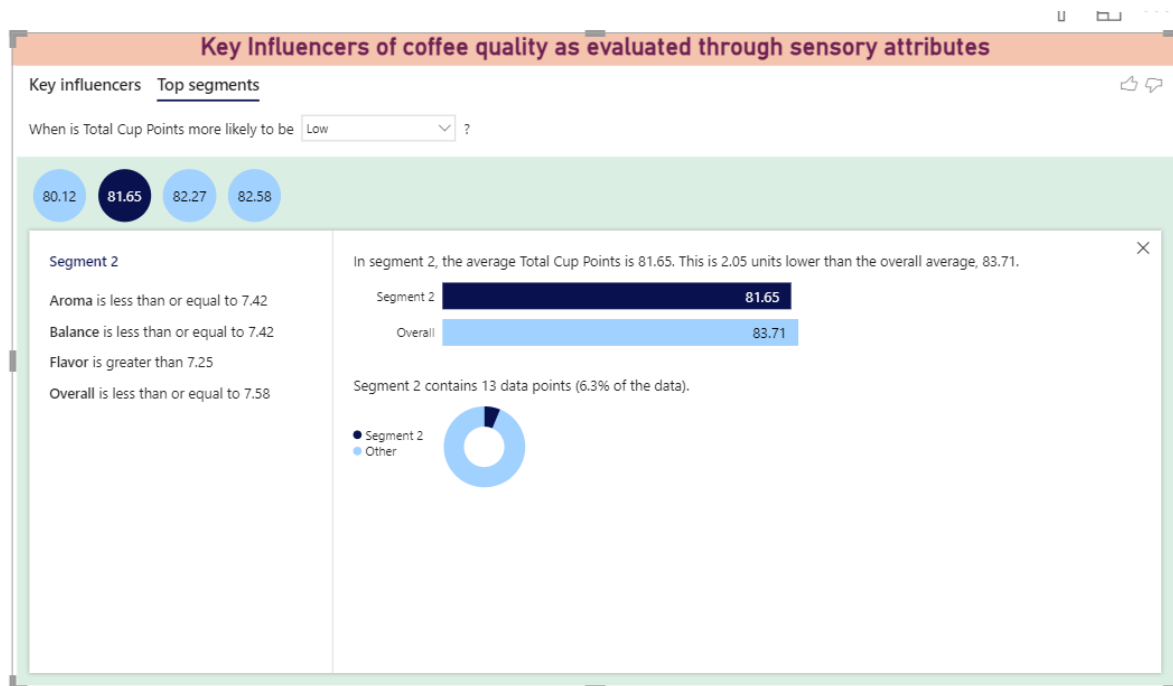
- Aroma ≤ 7.42

- Balance ≤ 7.42

- Flavor > 7.25

- Average Score: 81.65

Meaning:

Aroma/Balancing defects lower quality more than flavor defects alone.

Low Score Segment 3
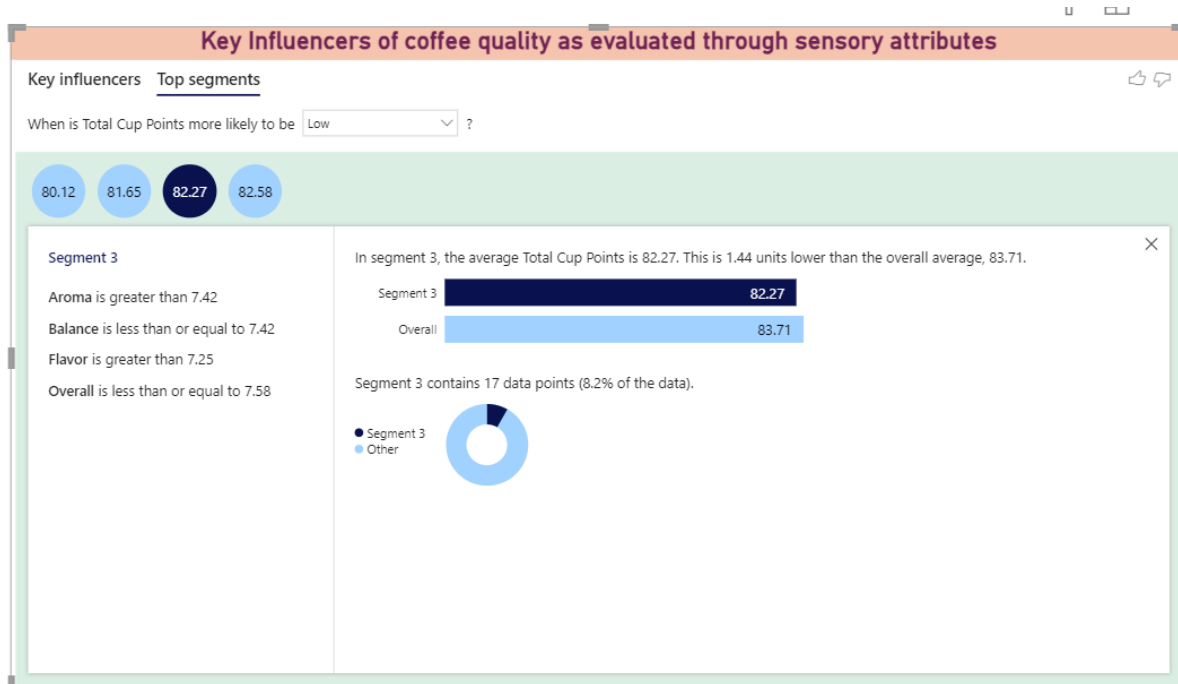
- Aroma > 7.42

- Balance ≤ 7.42

- Flavor > 7.25

- Overall ≤ 7.58

- Average Score: 82.27

Meaning:

Even with decent aroma/flavor, poor balance and weak overall perception lead to lower scores.

5. Final Summary :

In specialty coffee evaluation, aftertaste, aroma, and flavor are foundational. These attributes capture the bean's intrinsic quality, processing accuracy, and roasting effectiveness.

The Power BI model confirms that these are the core sensory pillars that elevate coffee into higher scoring brackets. Conversely, deterioration in the same attributes is the earliest indicator of quality decline.

Objective 2: Is there a correlation between processing methods, origin regions, and coffee quality scores?

1.Processing Method vs Coffee Quality Scores

Processing method is one of the strongest determinants of final cup quality because it directly influences fermentation, sugar development, acidity structure, and clarity of flavor.
The evaluation considers:

- **Average Total Cup Points**

- **Maximum Total Cup Points achieved**

- **Processing_Count** (number of samples for each method)

Average of Total Cup Points by Processing Method and Processing_Count

Processing_Count ●1 ●5 ●25 ●46 ●124

Processing Method	Average of Total Cup Points
Double Anaerobic Washed	89.33
Semi Washed	87.42
Honey,Mossto	87.08
Double Carbonic Maceration / Natural	84.75
	84.42
Wet Hulling	84.25
Natural / Dry	83.70
Washed / Wet	83.65
Pulped natural / honey	83.55
Anaerobico 1000h	83.25
SEMI-LAVADO	78.00

1.Highest Scoring Methods — Based on Averages (Consistency)

Processing Method	Avg Cup Points	Processing_Count	Insight
Double Anaerobic Washed	**89.33**	Low	Extremely high-quality, controlled fermentation micro-lot
Semi Washed	**87.42**	Low	Clean, high-clarity cup profile
Honey / Mossto	**87.08**	Low	Enhanced sweetness + fruit-forward notes

Processing Method	Avg Cup Points	Processing_Count	Insight
Double Carbonic Maceration	**84.75**	Low	Advanced fermentation drives complexity

Insight:
These top-scoring methods almost always have **very low Processing_Count**, meaning they are **micro-lots** designed for maximum cup quality with precision fermentation. They deliver consistently high averages due to their focus on premium batches.
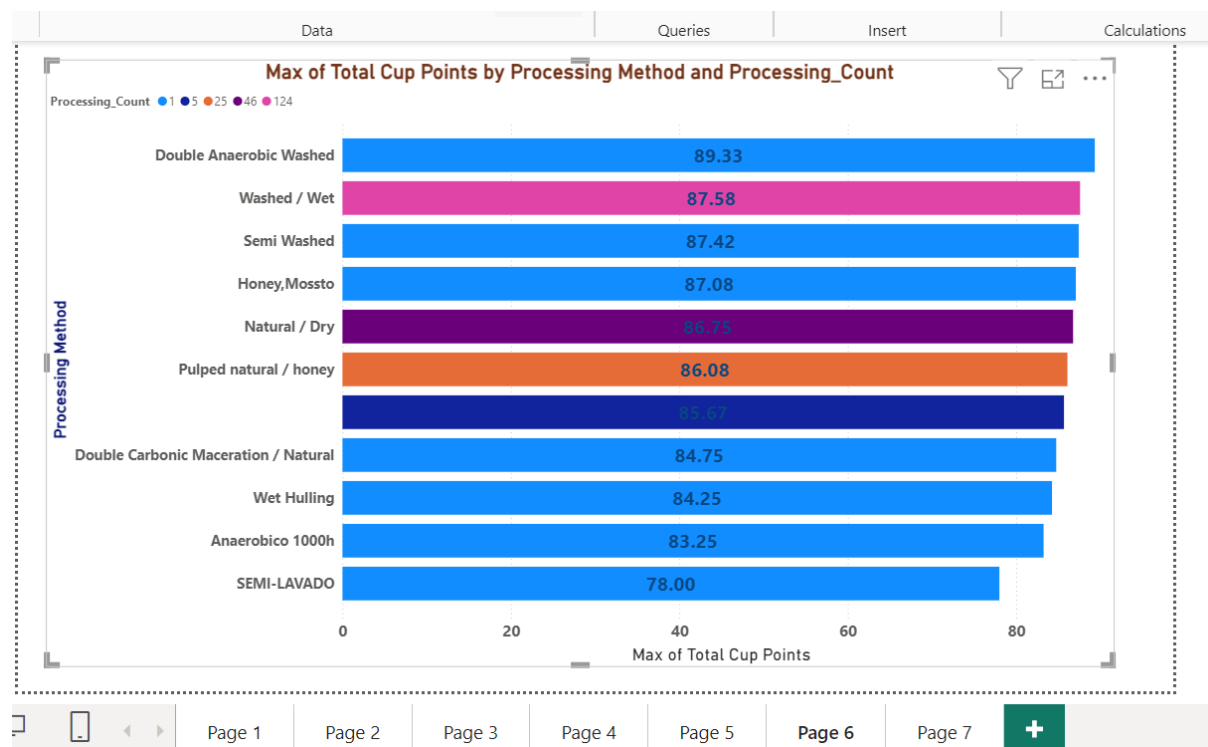
2. Highest Scoring Methods — Based on Maximum Cup Points (Peak Potential)

The maximum scores reveal the **best possible outcome** each method can achieve.

Processing Method	Max Cup Points	Insight
Double Anaerobic Washed	**89.33**	Same as average → extremely consistent

Processing Method	Max Cup Points	Insight
Washed / Wet	**87.58**	Shows the method can produce exceptional cups even if average is moderate
Semi Washed	**87.42**	High consistency, high ceiling
Honey / Mossto	**87.08**	Always produces high-caliber coffees
Pulped Natural / Honey	**86.08**	High potential when processed carefully
Natural / Dry	**86.75**	Peaks very high but averages lower due to variability

Insight:

Even if a method like **Washed/Wet** or **Natural/Dry** has **moderate average scores**, their **maximum score is high**, meaning these methods have *high potential* but also *high variability*.

3. Lower Performing Methods

Method	Avg Score	Processing_Count	Insight
Anaerobico 1000h	83.25	Low	Long fermentation introduces risk
Wet Hulling	84.25	Medium	Common in Indonesia; less acidity and clarity
SEMI-LAVADO	**78.00**	High	Lowest performance due to basic processing, high variability

Insight:

Lower scores typically occur in **high-count, large-scale processes**, where quality varies widely across farms.

4. Processing_Count (Sample Size) Effects

a. When Processing_Count is LOW (1–5 samples)

- These methods appear **high-scoring** because they represent:
    - Premium, selective lots
    - Experimental fermentation
    - High control and low variation

Result: Highest averages and maximum scores, but small representation.

b. When Processing_Count is HIGH (25–124 samples)

- These methods show **lower averages** due to:
    - Wider farm-level variation
    - Mixed quality levels
    - Commercial-scale production

Result: The "true" performance emerges with more data → averages slightly decrease.
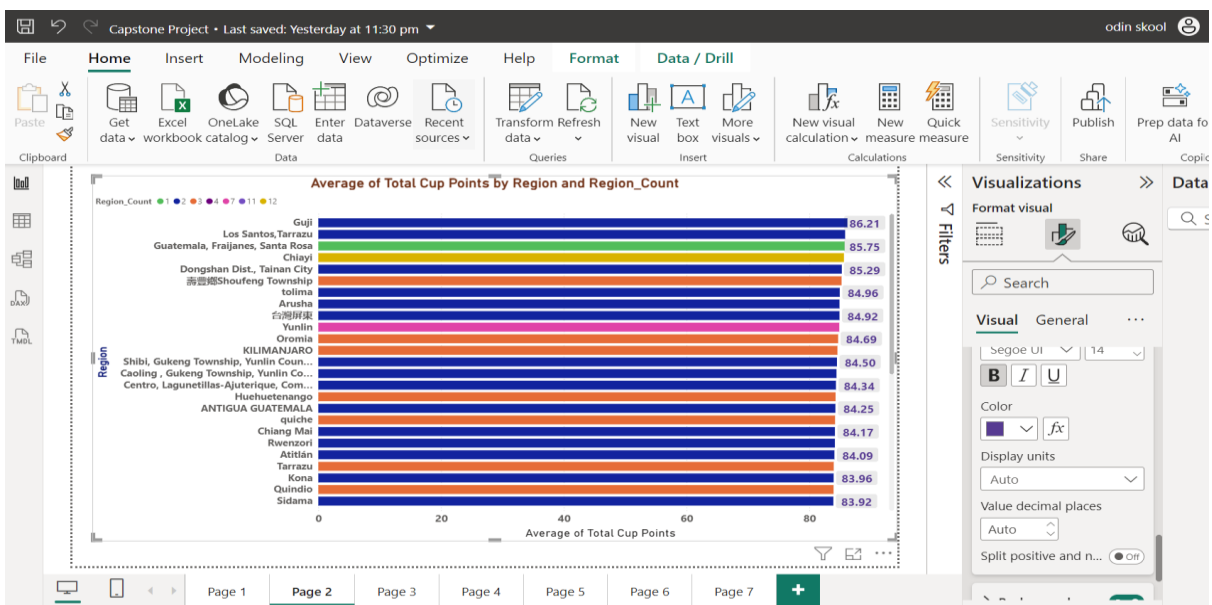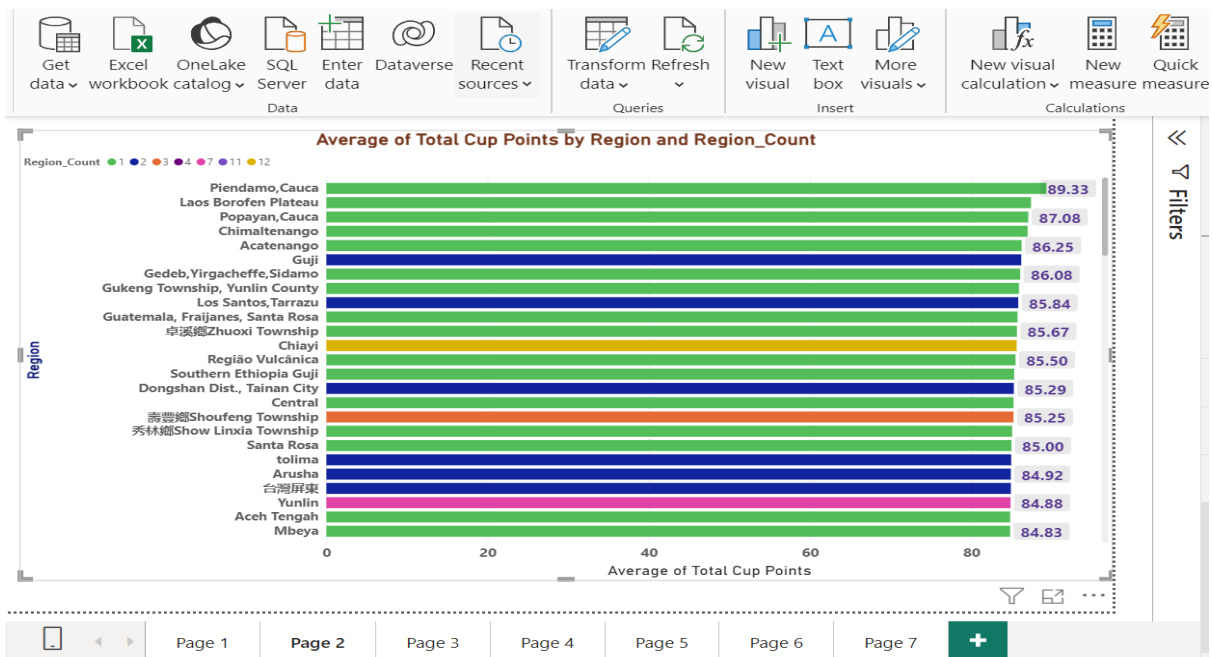
Processing Summary

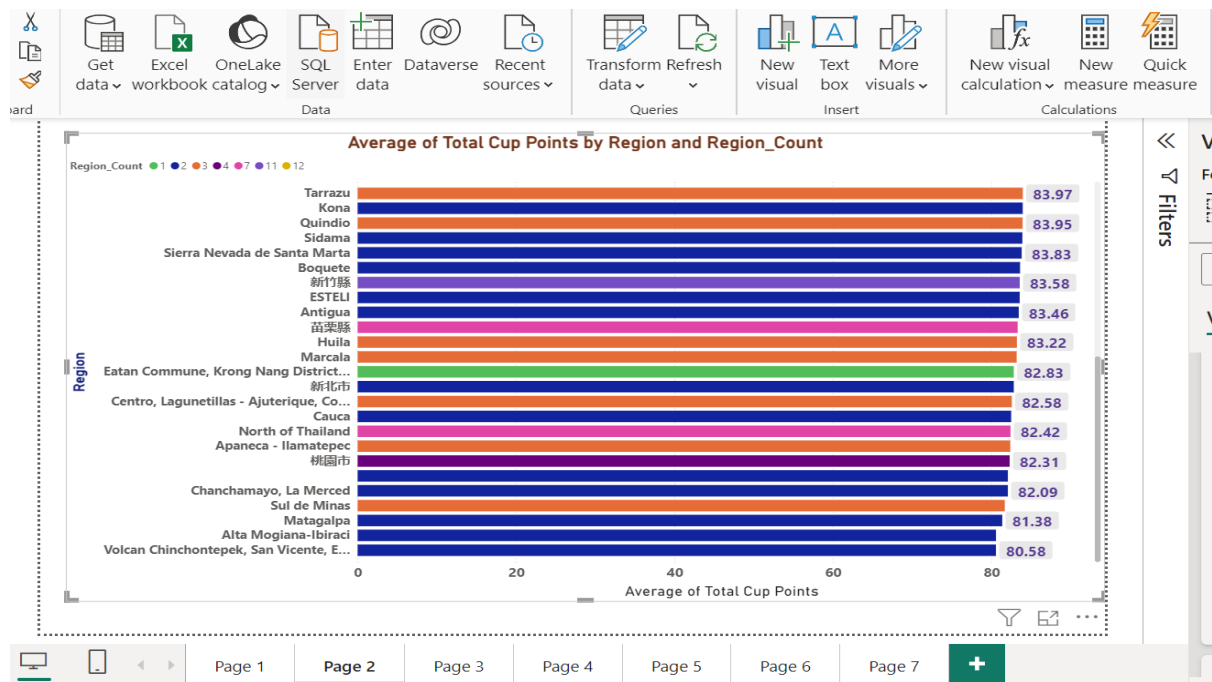A method with high potential (high max score) but lower average suggests variability.
A method with both high average and high max indicates consistent excellence.

2.Origin Region vs Coffee Quality Scores:

Region is a strong indicator of coffee quality due to altitude, climate, soil composition, and traditional processing techniques.

Average of Total Cup Points by Region and Region_Count

Region_Count ●1 ●2 ●3 ●4 ●7 ●11 ●12

Region	Average of Total Cup Points
Tarrazu	83.97
Kona	
Quindio	83.95
Sidama	
Sierra Nevada de Santa Marta	83.83
Boquete	
新竹縣	83.58
ESTELI	
Antigua	83.46
苗栗縣	
Huila	83.22
Marcala	
Eatan Commune, Krong Nang District...	82.83
新北市	
Centro, Lagunetillas - Ajuterique, Co...	82.58
Cauca	
North of Thailand	82.42
Apaneca - Ilamatepec	
桃園市	82.31
Chanchamayo, La Merced	82.09
Sul de Minas	
Matagalpa	81.38
Alta Mogiana-Ibiraci	
Volcan Chinchontepek, San Vicente, E...	80.58

Pages: Page 1 | **Page 2** | Page 3 | Page 4 | Page 5 | Page 6 | Page 7 | +

A. High-Performing Regions (85–89 Cup Points):

Examples include:

- **Piendamo, Cauca (89.33)**

- **Laos Borofen Plateau (87.08)**

- **Popayan, Cauca (86.25)**

- **Chimaltenango (86.08)**

- **Guji, Yirgacheffe, Sidamo cluster (85.5+)**

Insight:
These are **high-altitude, specialty-focused regions**, producing coffees with pronounced acidity, floral complexity, and balanced sweetness.

B. Mid-Performing Regions (83–85):

Examples:

- **Arusha**

- **Tolima**

- **Região Vulcânica**

- **Dongshan District**

Insight:
Stable specialty-grade coffees with good character but less extreme elevation or advanced processing.

C. Lower-Scoring Regions (78–82):

Examples:

- **Kericho (82.17)**

- **North of Thailand (82.31)**

- **Matagalpa (81.33)**

- **Minas Gerais (78.00)**

Insight:
Large-scale or lower-altitude production tends to reduce acidity complexity and cup clarity.

Region_Count (Sample Size) Effects:

When Region_Count is LOW (1–3 samples)

- Scores appear **extremely high** because samples come from **elite, curated micro-lots**.

- Used often for competitions or limited releases.

When Region_Count is HIGH (7–12 samples)

- Averages naturally **decrease** due to:

  - Mixed processing methods

  - Inclusion of multiple farms

  - Diverse quality levels

Insight:
High region counts reveal the *true average* of the origin, while low counts reflect *highlighted premium lots*.

Final Summary :

Coffee quality shows a strong correlation with **both processing methods and origin regions**, and this relationship becomes clearer when the **number of samples (Processing_Count and Region_Count)** is considered.

Advanced fermentation methods such as **Double Anaerobic**, **Carbonic Maceration**, and **Honey/Mossto fermentations** demonstrate the highest average and maximum cup scores, though they come from **small sample sets**, indicating selective micro-lot production. Conventional methods such as **Natural/Dry** and **Washed/Wet** show high maximum potential but lower averages due to higher variability, especially when sample size increases.

Similarly, regions with exceptional climatic and altitudinal advantages—such as **Cauca (Colombia)**, **Guji/Yirgacheffe (Ethiopia)**, **Laos Plateau**, and **Guatemala's volcanic belt**—consistently achieve the highest cup scores. However, regions with **larger sample counts** display reduced averages because they include wider farm-level variability and mixed processing practices.

Overall, **processing precision**, **environmental terroir**, and **sample-size variability** collectively shape the observed coffee quality. High scores are driven by selective micro-lots and advanced fermentations, while broader datasets reveal the underlying quality of large-scale production regions and methods.

Objective 3: Can we identify any trends or patterns in defect occurrences and their impact on overall coffee quality

Defects play a crucial role in determining the physical and sensory quality of coffee. This objective evaluates **Category 1 defects** (severe physical defects) and **Category 2 defects** (less severe defects) and assesses how their presence influences **Total Cup Points**, which represent overall coffee quality.

1.Category 1 Defects vs Total Cup Points

Trends/Patterns in Category 1 defect occurrences and their Impact on Overall Coffee Quality
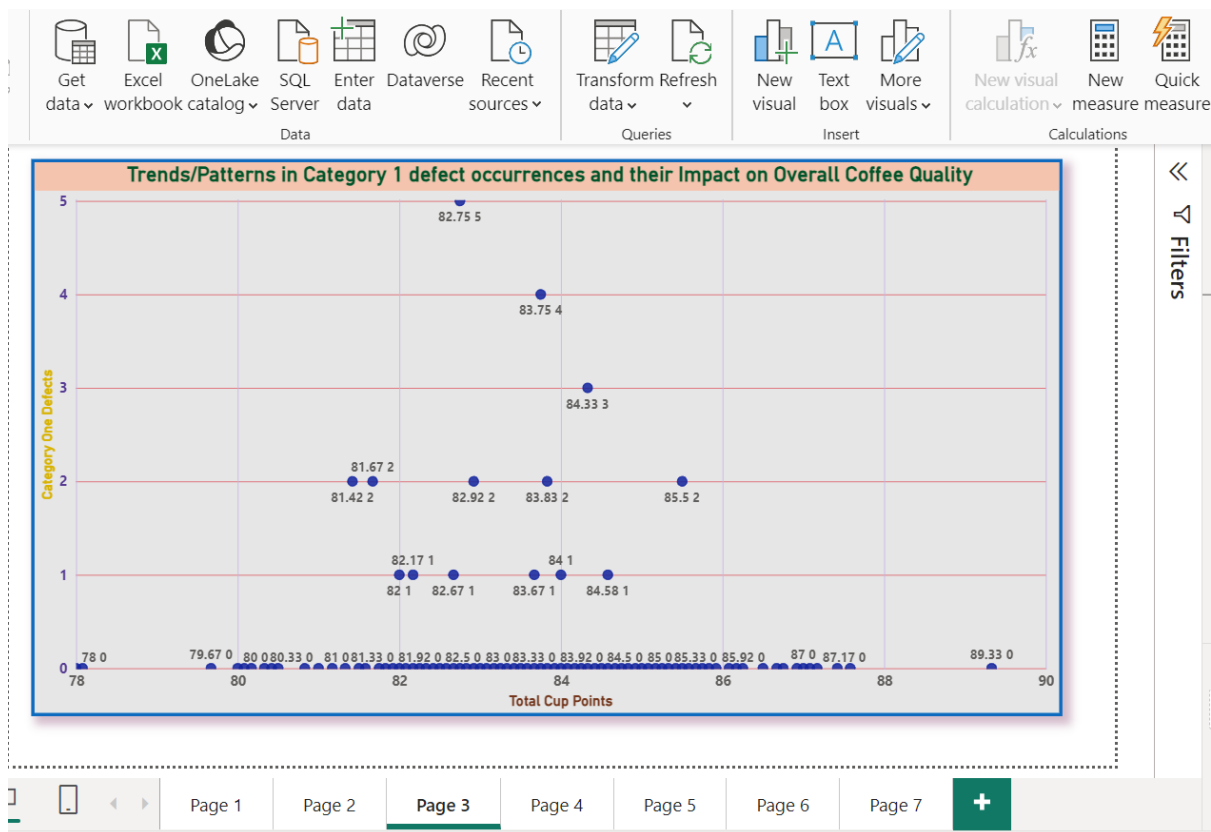
A. Overall Pattern

The scatterplot reveals a **very strong inverse relationship** between Category 1 defects and coffee quality:

- Almost **all coffees scoring above 85** have **zero Category 1 defects**.
- Category 1 defects are extremely rare in high-scoring specialty coffees.
- Even small counts (1–2 defects) noticeably **lower cup scores** into the 81–84 range.

B. Specific Trends Identified

1. **Scores 87–90:**
   - **0 Category 1 defects only.**
   - Perfect beans with no primary defects.

2. **Scores 84–86:**
   - Mostly **0 defects**, rarely **1 defect**.
   - Even a single primary defect pulls score downward.

3. **Scores 81–83:**

   o Commonly **1–2 defects**.

   o These coffees still have good sensory quality but lack premium consistency.

4. **Scores 78–80:**

   o Frequently **2–5 defects**.

   o Strong downward impact visible for coffees containing multiple Category 1 defects.

C. Interpretation :

Category 1 defects have a severe, immediate, and non-linear negative impact on overall coffee quality.
Even a single Category 1 defect can reduce scores by **2–4 points**, and multiple defects bring quality down sharply.

This category is the strongest predictor of **whether coffee remains "specialty grade" or drops into commercial grade**.

2.Category 2 Defects vs Total Cup Points
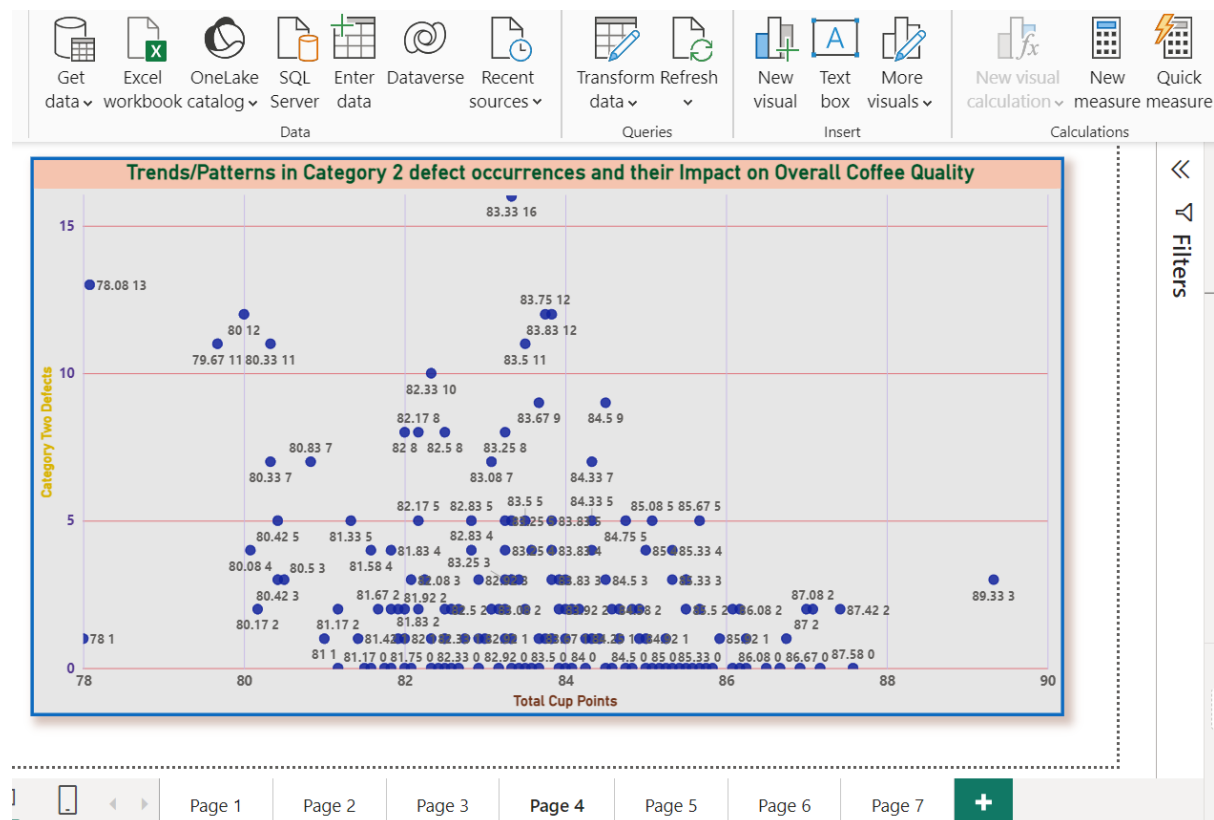
A. Overall Pattern

Unlike Category 1, Category 2 defects:

- Are **much more frequent** in the dataset

- Show a **wider dispersion** across different scoring ranges

- Have a **weaker, but still visible** inverse relationship with quality

B. Specific Trends Identified

1. **High-scoring coffees (85–90):**

   o Contain **0–2 Category 2 defects**.

   o Some exceptional lots show up to **3**, but rarely beyond that.

2. **Mid-range coffees (82–84):**

   o Commonly exhibit **2–6 defects**.

   o These coffees are still specialty but less refined.

3. **Low-scoring coffees (78–81):**

   o Show **7–15 defects**.

   o Clear clustering of high-defect lots in the lower scoring band.

4. **Outliers:**

   o Most extreme defect cases (10–15 defects) consistently fall below **83 points**.

   o No coffee with more than 7 defects crosses the **85-point** boundary.

C. Interpretation

Category 2 defects produce a gradual reduction in cup quality. They do not destroy cup integrity like Category 1 defects, but they accumulate and degrade sweetness, clarity, and uniformity.

This creates a **quality gradient**, where more Category 2 defects gently but consistently lower the score.

Final Summary :

The analysis reveals clear and measurable patterns between defect occurrences and overall coffee quality. Category 1 defects, although infrequent, exert a

severe negative influence, with even a single defect sharply reducing Total Cup Points and preventing coffees from reaching high specialty grades. In contrast, Category 2 defects are more common and contribute to a gradual decline in quality. Coffees scoring above 85 points consistently maintain zero Category 1 defects and minimal Category 2 defects, demonstrating that defect-free physical integrity is a prerequisite for achieving exceptional sensory performance.

Overall, the findings confirm that **both defect categories show strong inverse correlations with Total Cup Points**, and their presence significantly affects whether coffee qualifies as specialty grade or falls into lower commercial grades.


Objective 4 : How do different variables interact to influence Total Cup Points (overall coffee quality)?

1. Sensory Attribute Interactions (Core Drivers)

The Decomposition Tree shows that **Total Cup Points rise or fall only when multiple sensory attributes move together**. No single attribute alone pushes a coffee into the highest or lowest scoring brackets.

1.1 Interaction Patterns at the Top-End (87–89 Total Cup Points)
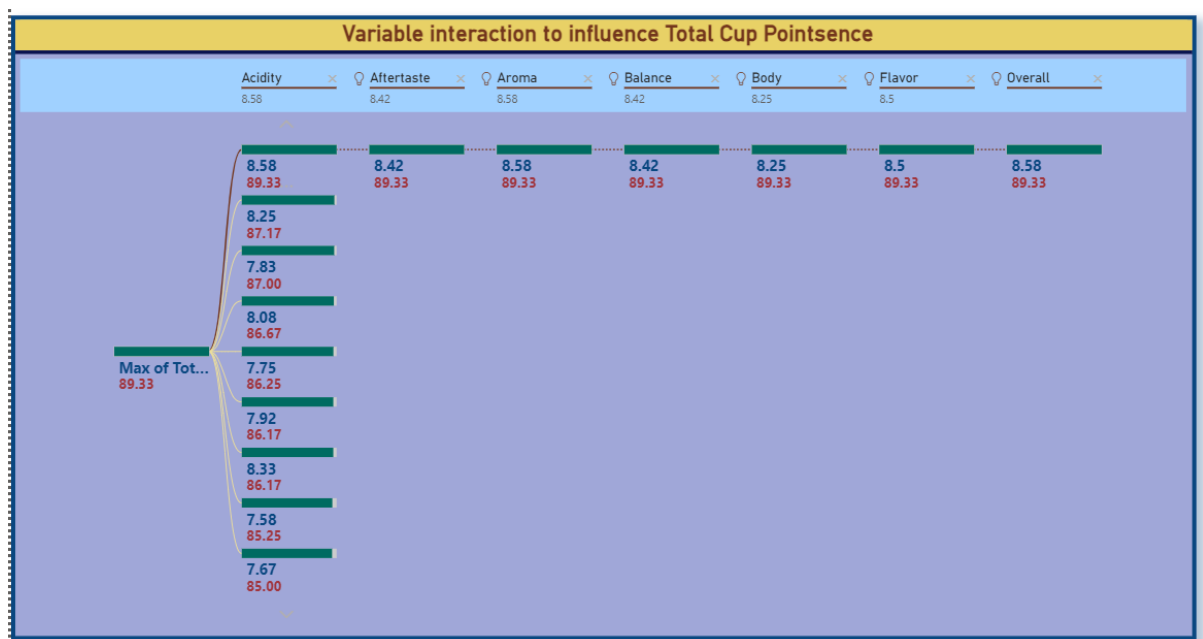
All key sensory attributes cluster at the high end simultaneously:

- **Acidity:** 8.25 – 8.58

- **Aroma:** 8.25 – 8.58

- **Aftertaste:** 8.25 – 8.42 *(strongest predictor)*

- **Flavor:** 8.33 – 8.50

- **Body:** 8.00 – 8.25

- **Balance:** 8.17 – 8.42

- **Overall:** 8.25 – 8.58

Interpretation:

High-scoring coffees require **sensory harmony** across multiple attributes.

When these attributes align, they push the Total Cup Points upward like a **reinforcing chain**. Even if one attribute drops slightly, the others compensate

1.2 Interaction Patterns at the Bottom-End (78–82 Total Cup Points)
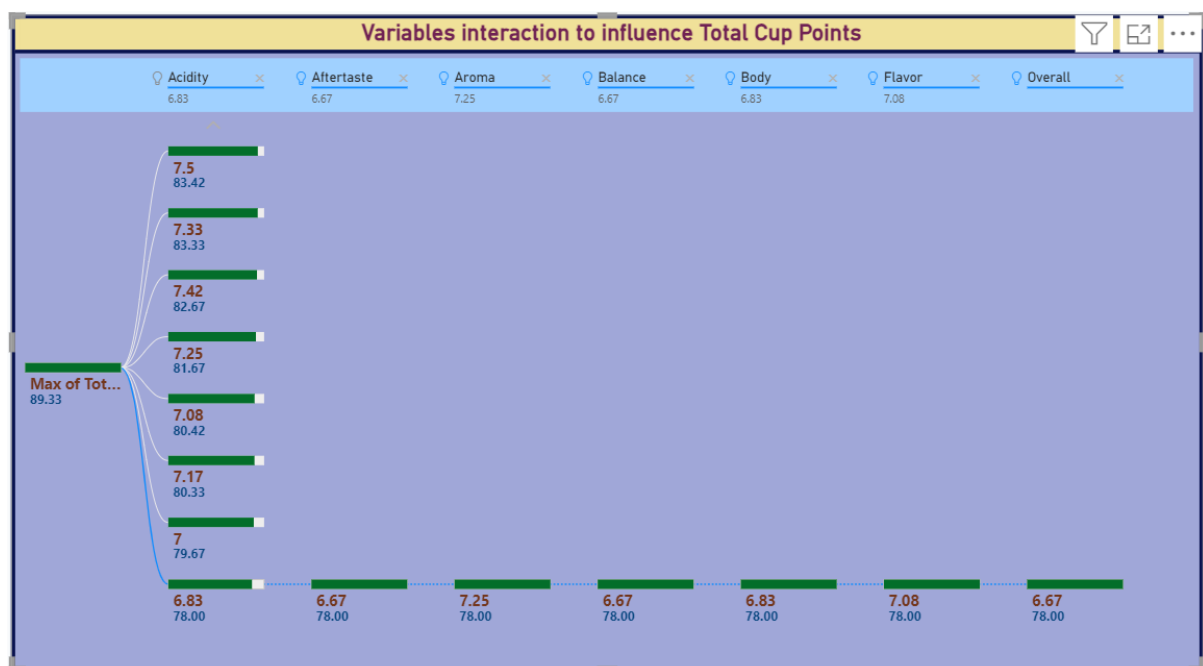
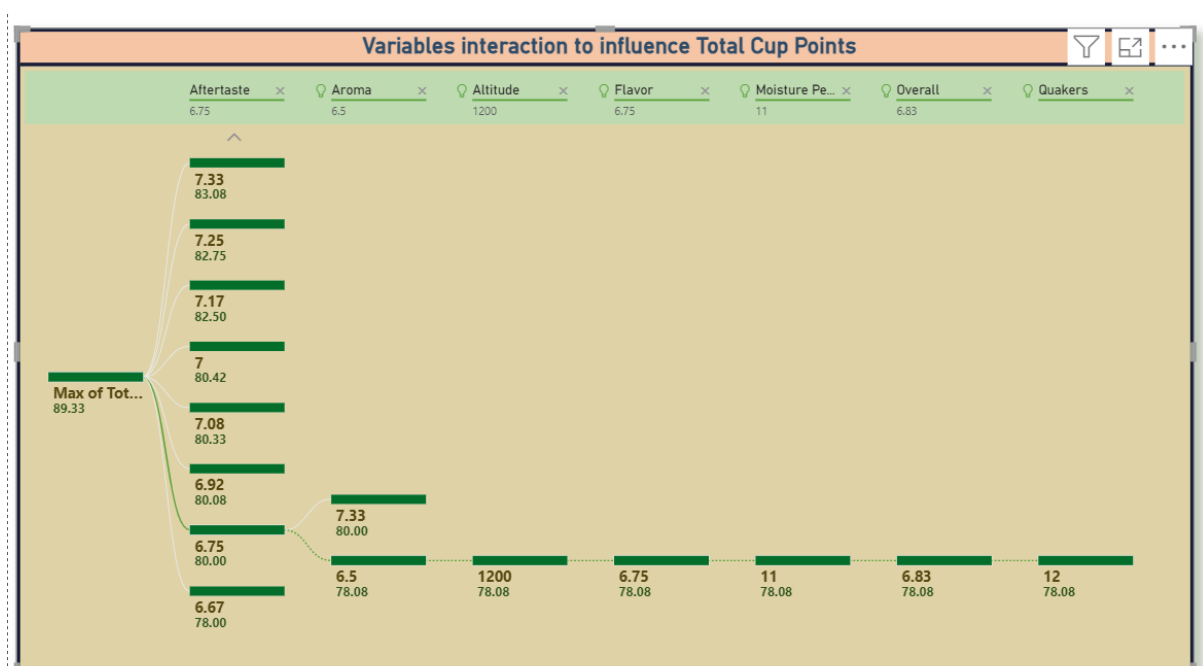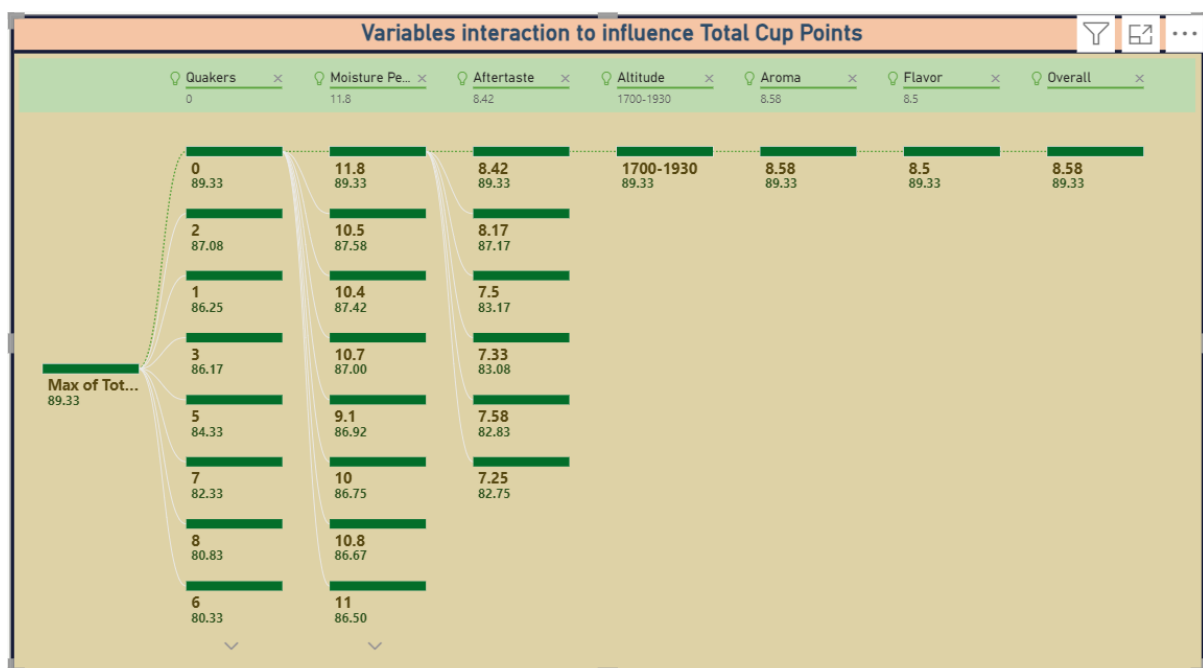Multiple sensory traits fall together into the low range:

- **Acidity:** 6.67 – 7.25

- **Aroma:** 6.50 – 7.42

- **Aftertaste:** 6.75 – 7.25

- **Flavor:** 6.75 – 7.42
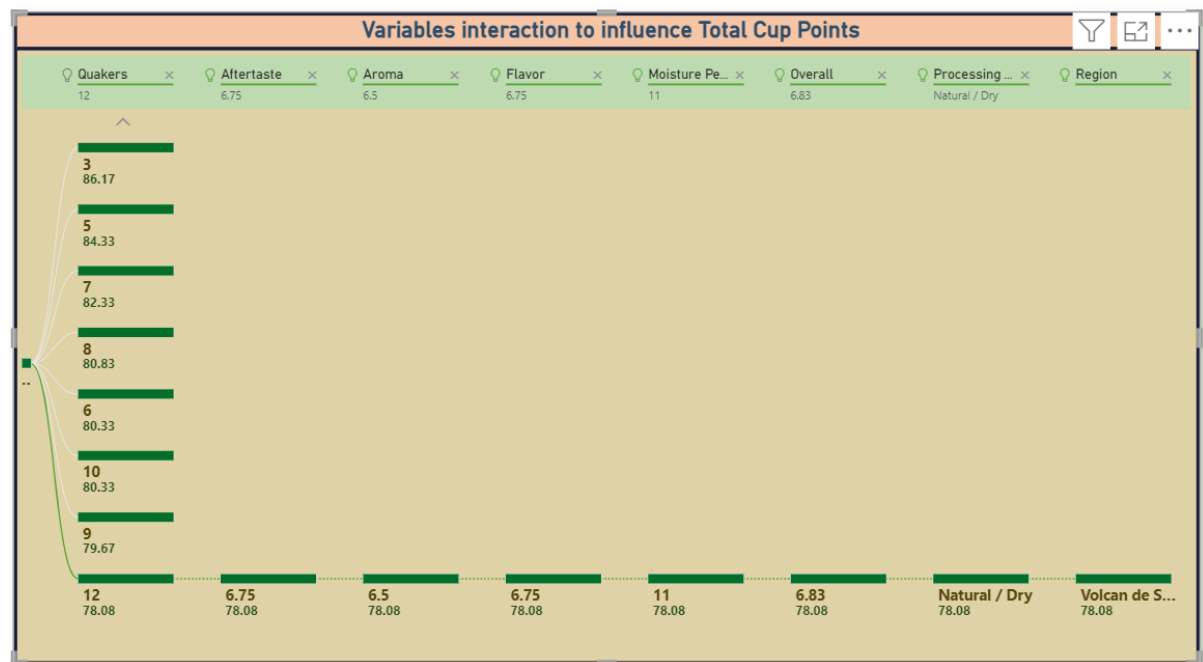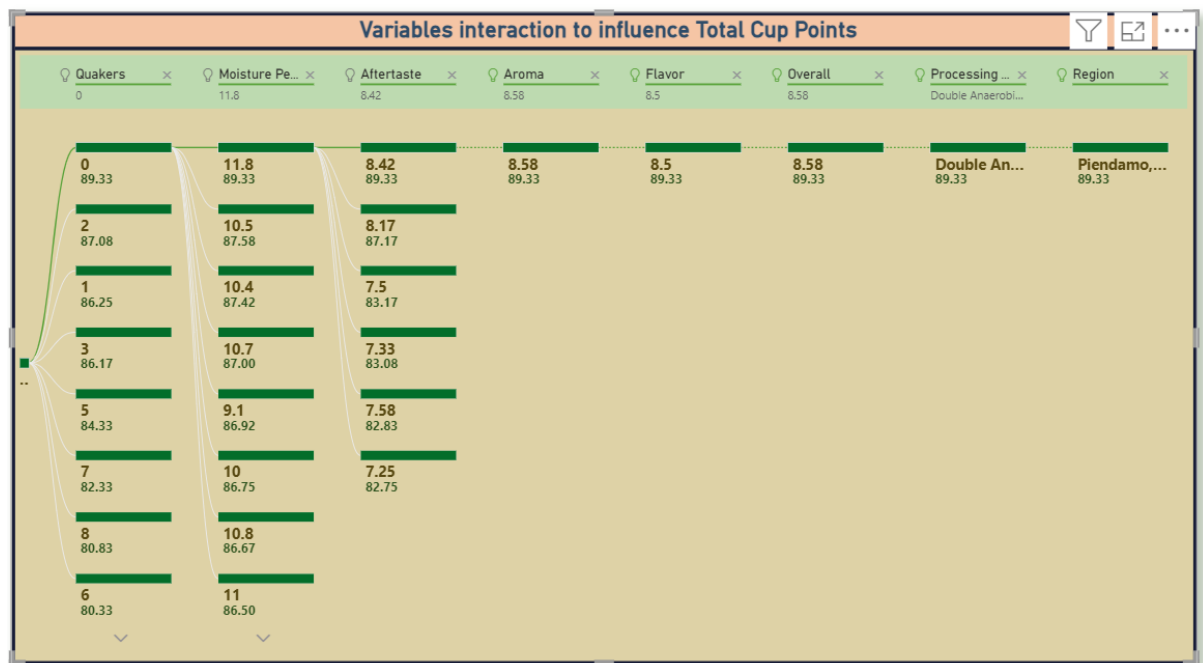
- **Body:** 6.67 – 7.17

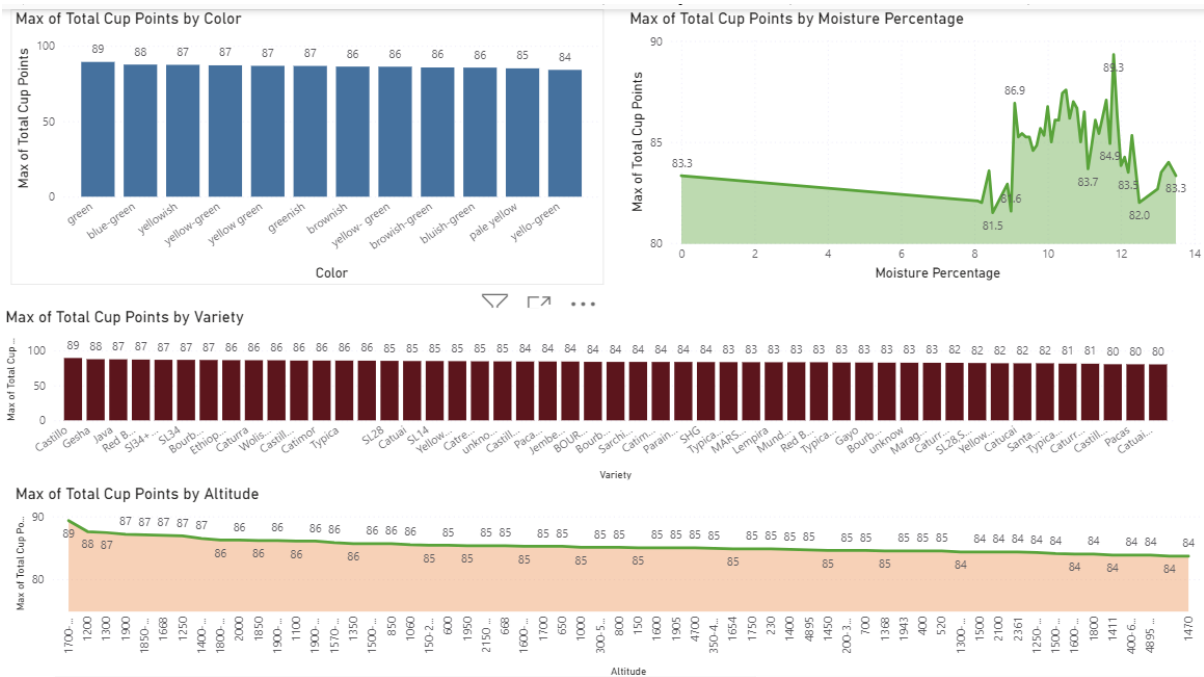- **Balance:** 6.67 – 7.25
- **Overall:** 7.00 – 7.42

Interpretation:

Low quality happens when **weakness accumulates** across multiple attributes.

2. Interaction of Sensory Attributes with External Variables

Variables interaction to influence Total Cup Points

Quakers	Moisture Pe...	Aftertaste	Aroma	Flavor	Overall	Processing ...	Region
0	11.8	8.42	8.58	8.5	8.58	Double Anaerobi...	
0 — 89.33	11.8 — 89.33	8.42 — 89.33	8.58 — 89.33	8.5 — 89.33	8.58 — 89.33	Double An... — 89.33	Piendamo,... — 89.33
2 — 87.08	10.5 — 87.58	8.17 — 87.17					
1 — 86.25	10.4 — 87.42	7.5 — 83.17					
3 — 86.17	10.7 — 87.00	7.33 — 83.08					
5 — 84.33	9.1 — 86.92	7.58 — 82.83					
7 — 82.33	10 — 86.75	7.25 — 82.75					
8 — 80.83	10.8 — 86.67						
6 — 80.33	11 — 86.50						

Variables interaction to influence Total Cup Points

Quakers	Aftertaste	Aroma	Flavor	Moisture Pe...	Overall	Processing ...	Region
12	6.75	6.5	6.75	11	6.83	Natural / Dry	
3 — 86.17							
5 — 84.33							
7 — 82.33							
8 — 80.83							
6 — 80.33							
10 — 80.33							
9 — 79.67							
12 — 78.08	6.75 — 78.08	6.5 — 78.08	6.75 — 78.08	11 — 78.08	6.83 — 78.08	Natural / Dry — 78.08	Volcan de S... — 78.08

Max of Total Cup Points by Color

Max of Total Cup Points by Moisture Percentage

Max of Total Cup Points by Variety

Max of Total Cup Points by Altitude

1. High-Quality Pathway (Top Scores ≈ 89.33)

Across all tree configurations, the highest-scoring coffees follow this pattern:

(A) Zero Quakers

- *Quakers = 0* is the strongest base indicator for premium coffee lots.
- Coffee samples with **zero defective beans** consistently anchor the highest scoring path.

(B) Optimal Moisture Content (≈ 11–12%)

- *Moisture ≈ 11.8%* maximizes score stability and flavor expression.
- Slight changes (10.4–10.7%) still yield strong outcomes but peak values sit near **11.8%**.

(C) High Sensory Attributes

- Aftertaste ≈ **8.4+**
- Aroma ≈ **8.5+**
- Flavor ≈ **8.5+**
- Overall ≈ **8.5+**

These four sensory metrics form the **core driver cluster**.
When all four cross the ~8.4 threshold, Total Cup Points almost always reach **89+**.

(D) High-Altitude Origins

- *Altitude: 1700–1930 meters*
  This altitude range repeatedly appears as the highest-scoring branch. Higher altitudes → slower cherry maturation → denser beans → richer flavor.

(E) Processing Method

- **Double Anaerobic Fermentation**
  Among all methods, this one shows the most consistent link with premium outcomes.

(F) Region Correlation

- Regions like **Piendamó (Cauca)** and a few Colombian micro-lots dominate the upper path.

- These regions are known for volcanic soil, stable microclimates, and advanced fermentation innovation.

2. Low-Quality Pathway (Scores ≈ 78–80)

The lowest-scoring coffees follow the opposite interaction pattern:

(A) High Quaker Levels (9–12)

- This is the **single strongest predictor** of low quality.

- Lots with 10+ quakers never escape the low-score range, even if other factors improve.

(B) Lower Sensory Attribute Values

- Aftertaste ≈ 6.7–7.1

- Aroma ≈ 6.5–7.2

- Flavor ≈ 6.7–7.2

- Overall ≈ 6.8–7.0

All these fall below specialty-grade thresholds.

(C) Suboptimal Moisture Content

- Moisture ≈ **10.4–11.2%** but with instability or poor handling (e.g., fermentation issues).

- Deviations outside the stable 11.5–12% range correlate with reduced quality.

(D) Lower Altitude Ranges

- Altitude ≈ 600–1100 m
  Produces beans with lower density → flatter flavor.

(E) Basic Processing Methods

- Washed/Wet processes appear frequently in low paths when paired with high quakers and poor sensory marks.
  (Important: washed method is not bad by itself, but in your dataset, low lots tend to be washed.)

(F) Regions With Lower Performance

- Regions like **Bolaven Plateau** and some low-altitude Latin American zones appear in low-score paths.

3. Which Variables Have the Strongest Interaction Effects?

Across all decomposition trees:

Tier 1 — Foundational Quality Drivers (Most Impactful)

1. Quakers (Defects)
2. Aftertaste, Aroma, Flavor, Overall (core sensory cluster)
3. Altitude

Tier 2 — Secondary Enhancers (Moderate Impact)

4. Moisture Percentage
5. Processing Method

Tier 3 — Contextual Modifiers (Lower Impact)

6. Region
7. Variety
8. Bean Color

Final Summary :

Coffee quality is primarily shaped by defect-free beans (Quakers), strong sensory attributes (aftertaste, aroma, flavor, overall), optimal moisture content, high-growing altitudes, and advanced processing methods; these variables interact in sequence to produce the highest Total Cup Points.

---

5. Conclusions

- Aftertaste, Aroma, Flavor, and Overall are the strongest sensory drivers.

- Process type—especially anaerobic fermentation—significantly enhances quality.

- Altitude and region play major roles.

- Defects (especially Category 1) sharply reduce quality.

- Moisture, quakers, and variety also influence cup points when combined with sensory strengths.

---

6. Recommendations

- Focus on optimizing aftertaste, aroma, flavor, and overall balance.

- Adopt advanced processing: Double Anaerobic / Honey / Semi Washed.

- Control moisture strictly between 10.5–11.8%.

- Maintain quakers = 0–2.

- Reduce Category 1 defects to zero.

- Prioritize lots from 1700–1900m altitude.

- Look for Castillo, Gesha, SL28 varieties.

- Select origins with consistent high scores like Colombia, Ethiopia, Costa Rica.

Power BI File (PBIX) :

Coffee_Quality_Analy
sis_Capstone_Project.