

Final_Product

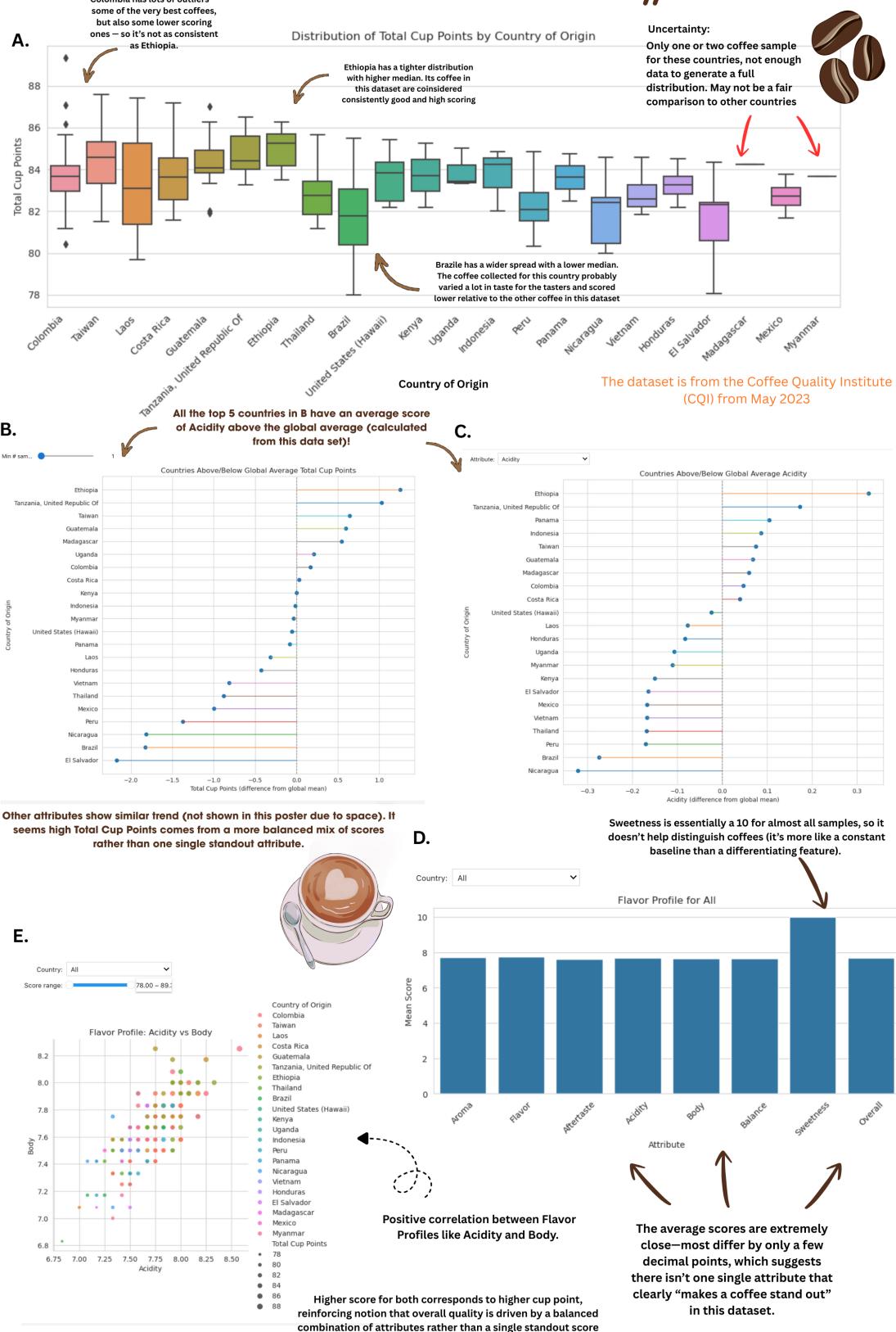
December 12, 2025

1 Final Product - Coffee Quality Visualizations

Name: Angel Lai

Computing ID: Bpy2nc

Does one attribute make a coffee stand out?



Visualization A:

This boxplot shows how Total Cup Points (overall quality scores assigned by professional tasters) vary by Country of Origin. Each point in the underlying dataset is one coffee lot, and its Total Cup Points is a summary of multiple sensory attributes like aroma, flavor, acidity, body, and balance. Grouping by country lets us compare the quality distributions of coffees from different origins.

Each box represents the distribution of scores for one country. The horizontal line inside the box is the median (the typical score), the top and bottom of the box show the 75th and 25th percentiles (the middle 50% of coffees), and the “whiskers” extend toward the more extreme values. Any individual dots beyond the whiskers would be outliers, representing unusually high- or low-scoring lots. For some countries (like Madagascar and Myanmar in this dataset), there is only one coffee sample; in those cases, the boxplot collapses to a single horizontal line because there isn’t enough data to form a full distribution (only one data point).

This visualization helps compare both the typical quality and the consistency of coffees by origin. Countries with higher medians tend to produce higher-scoring coffees overall, while countries with taller boxes or longer whiskers show more variation in quality. Countries with shorter boxes and whiskers are more consistent. Interpreting the plot this way allows us to see which origins not only reach high scores but also how reliably they do so across different lots.

Visualization B & C:

In conjunction with Visualization C, we can check whether the countries that rank highest in Total Cup Points (like Ethiopia, Tanzania, Taiwan, Guatemala, and Madagascar) also score above the global average on specific flavor attributes (for example, Acidity). In the attribute-difference plot, countries to the right of 0 have higher-than-average attribute scores, while countries to the left are below average. Comparing the two plots helps us see whether “top quality” countries are consistently above average in certain attributes, or if high Total Cup Points comes from a more balanced mix of scores rather than one single standout attribute.

Visualization D:

Because the attribute means are so tightly clustered, differences in overall quality are more likely driven by small trade-offs across multiple attributes rather than one standout score. Visualization E reinforces this belief. (Read Annotation)

```
[10]: import pandas as pd

resources = [
    ["df_arabica_clean.csv", "Original Coffee Quality Institute May 2023",
     ↴dataset", "Angel/df_arabica_clean.csv"],
    ["Rashed Sumon's Kaggle Post", "Description of Coffee Quality Data from May",
     ↴2023", "https://www.kaggle.com/code/rashedsumon/
     ↴coffee-quality-data-cqi-may-2023/input"],
    ["EstablishingData.ipynb", "Notebook for establishing the data", "Angel/
     ↴Notebooks/EstablishingData.ipynb"],
    ["ExploringData.ipynb", "Exploratory data analysis notebook", "Angel/
     ↴Notebooks/ExploringData.ipynb"],
    ["Code_4_FinalProduct.ipynb", "Code for visualizations on infographic",
     ↴"Angel/Notebooks/Code_4_FinalProduct.ipynb"],
```

```
["Coffee Quality Visualization.png", "Final Infographic", "Angel/CoffeeQuality Visualization.png"],  
["CQI Website", "Background information about certification and scoring",  
 "https://coffeeinstitute.org"],  
 ["ChatGPT", "For help in coding", "https://chatgpt.com"],  
]  
  
df_resources = pd.DataFrame(resources, columns=["Resource Name", "Description",  
 "Link / File Path"])  
df_resources  
from IPython.display import display, HTML  
  
display(HTML(df_resources.to_html(index=False)))
```

<IPython.core.display.HTML object>