



BEANalytics: A Data-Driven Exploration of Coffee Quality Determinants



DS4A Team 21: Celine Hsu, Elena Song, Jiawen Chen, Stella Tong, Yulia Yan

Why Predict Coffee Quality?

Brand Reputation

Maintain brand reputation with consistent coffee quality



Consumer Satisfaction

Match consumer preference with higher accuracy

Market Segmentation

Target specific regions and customers

Quality Control

Find good coffee bean sources and evaluate quality

Sustainable Sourcing

Support efforts to protect unique coffee-growing regions

Data Sources

Dataset

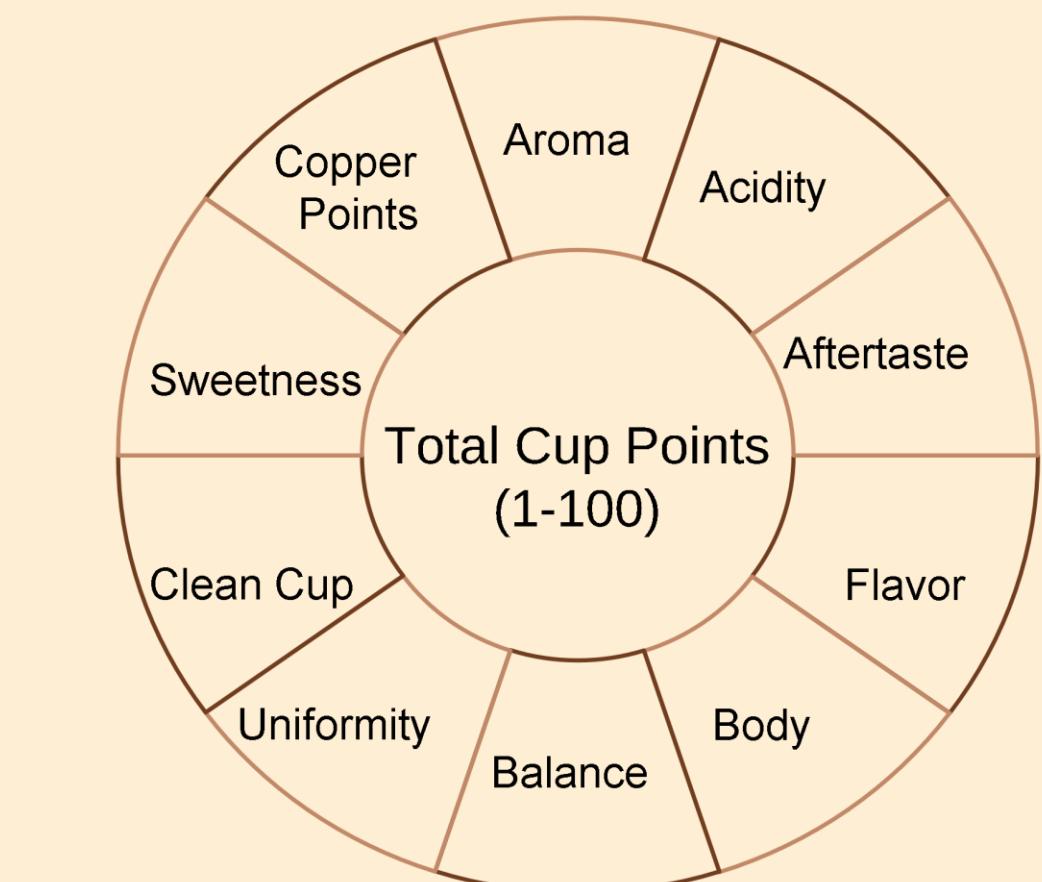
- 1338 Data Points
- 44 variables
- Collected in 2018



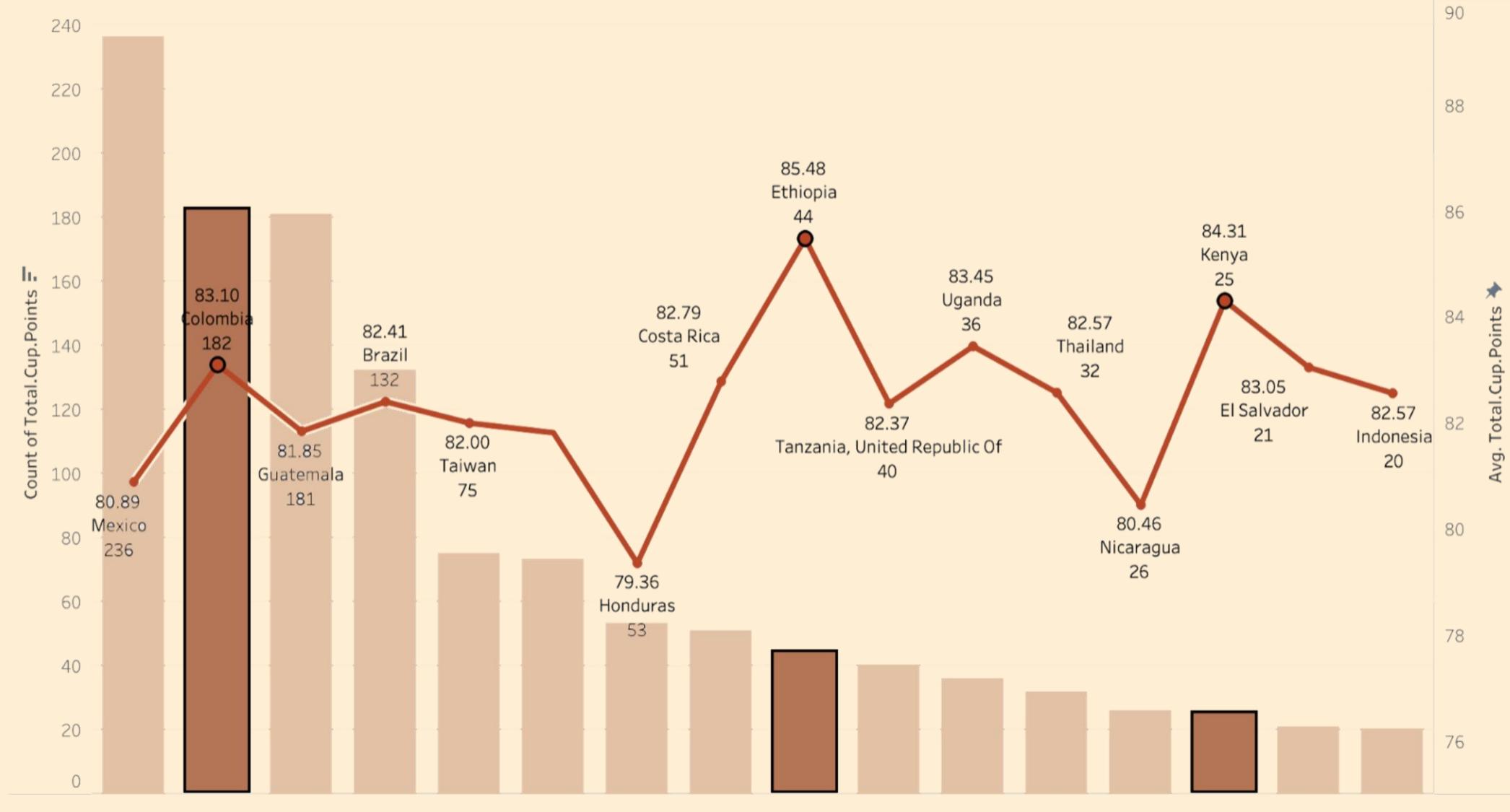
Output

Input

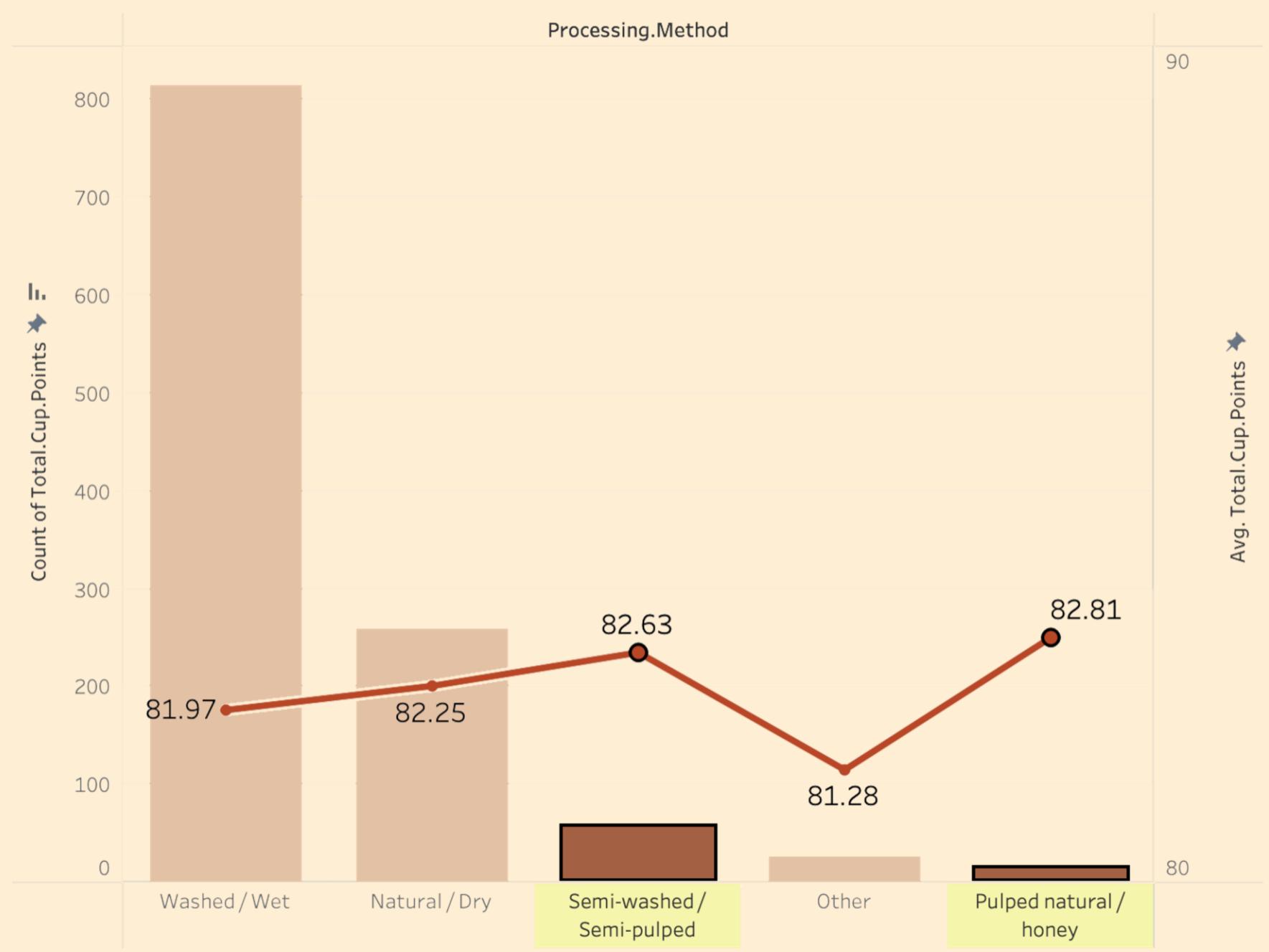
- Bean features
- Farm features



Countries of Origin

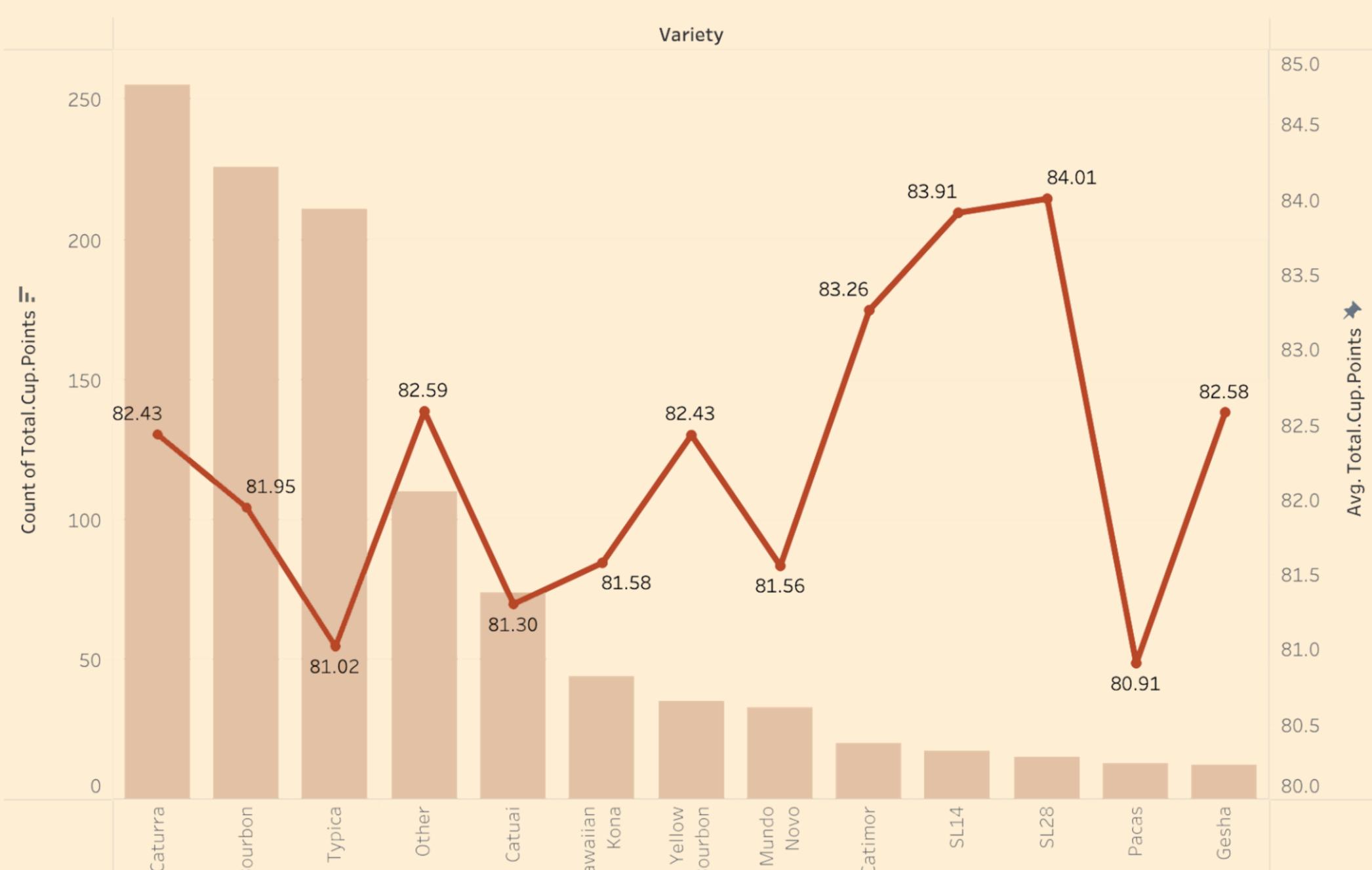


Processing Methods



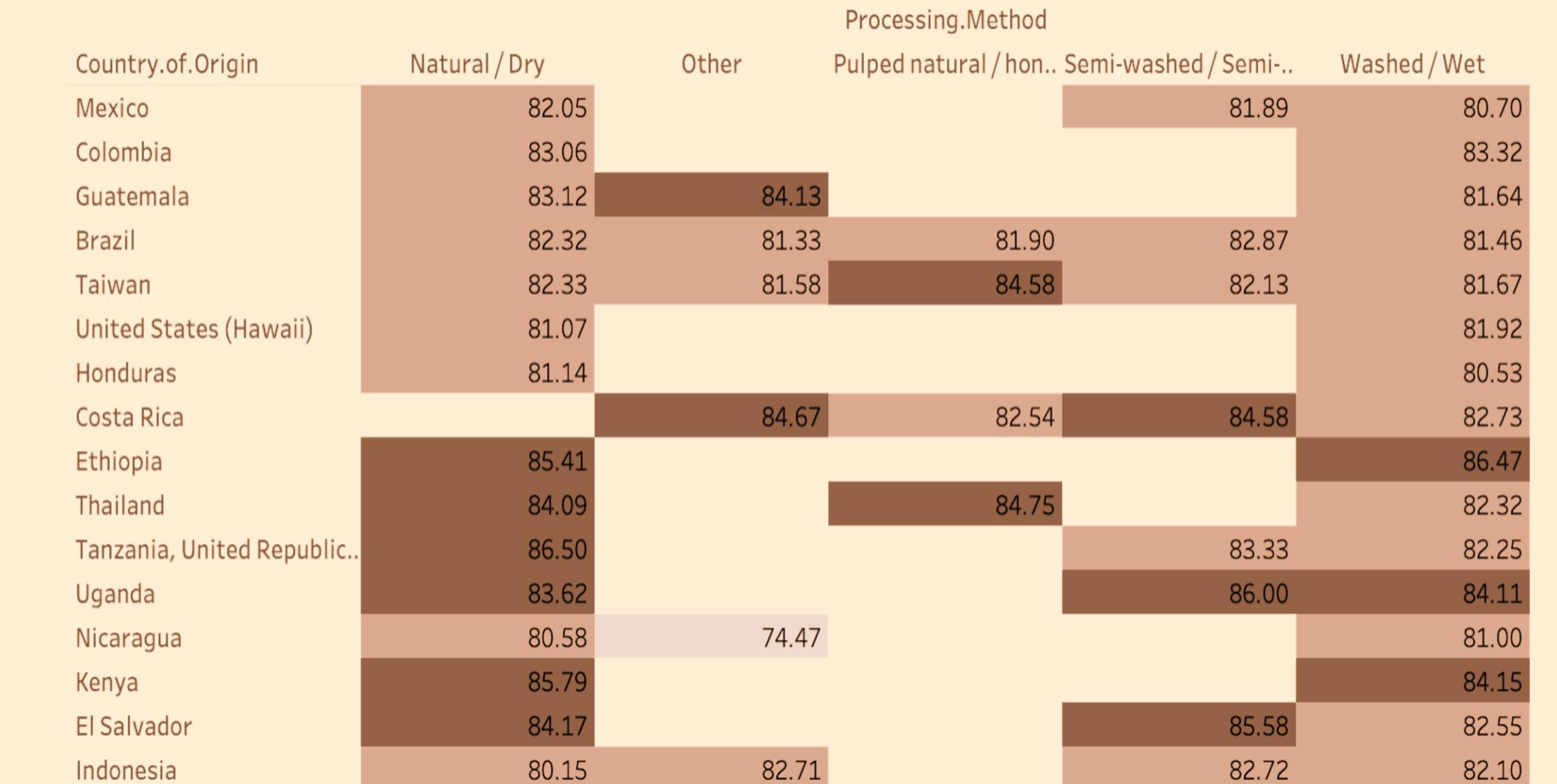
- Washed/Wet Method:**
 - The most common way
 - Coffee beans are immersed in water during processing.
- Natural/Dry Method**
 - The most traditional way
 - Coffee beans are spread out and sun-dried.
 - Yields a rich and earthy flavor profile.
- Semi-washed Method**
 - Partially washed, leaving sticky substance
 - Enhances sweetness and complexity
- Pulped Natural Method**
 - A modern twist of semi-washed method
 - Exhibits a harmonious flavor balance

Varieties of Beans



- Caturra, Bourbon, and Typica are the most common coffee varieties, each having over 200 counts.
- Despite their popularity, these three varieties do not guarantee high quality, particularly Typica..
- The highest average quality belongs to SL28, originated from Kenya, and SL14, found mostly in Kenya and Uganda. (potentially explaining the high average quality of these two countries.)

Processing Methods by Countries



- Washed/wet methods are the most commonly used, but its average score fluctuates across countries, leading to lower overall quality.
- Not every country utilizes the semi-washed and pulped natural methods, perhaps because they are relatively new methods.
- Which method works the best vary across countries.

Research Methods

01 Data Scoping

Select relevant variables (44->11) using domain knowledge

02 Data Cleaning

Impute missing values
Manage inconsistencies and outliers

03 Modeling

Automated model selection("LazyPredict")
Interpretability as a criterion

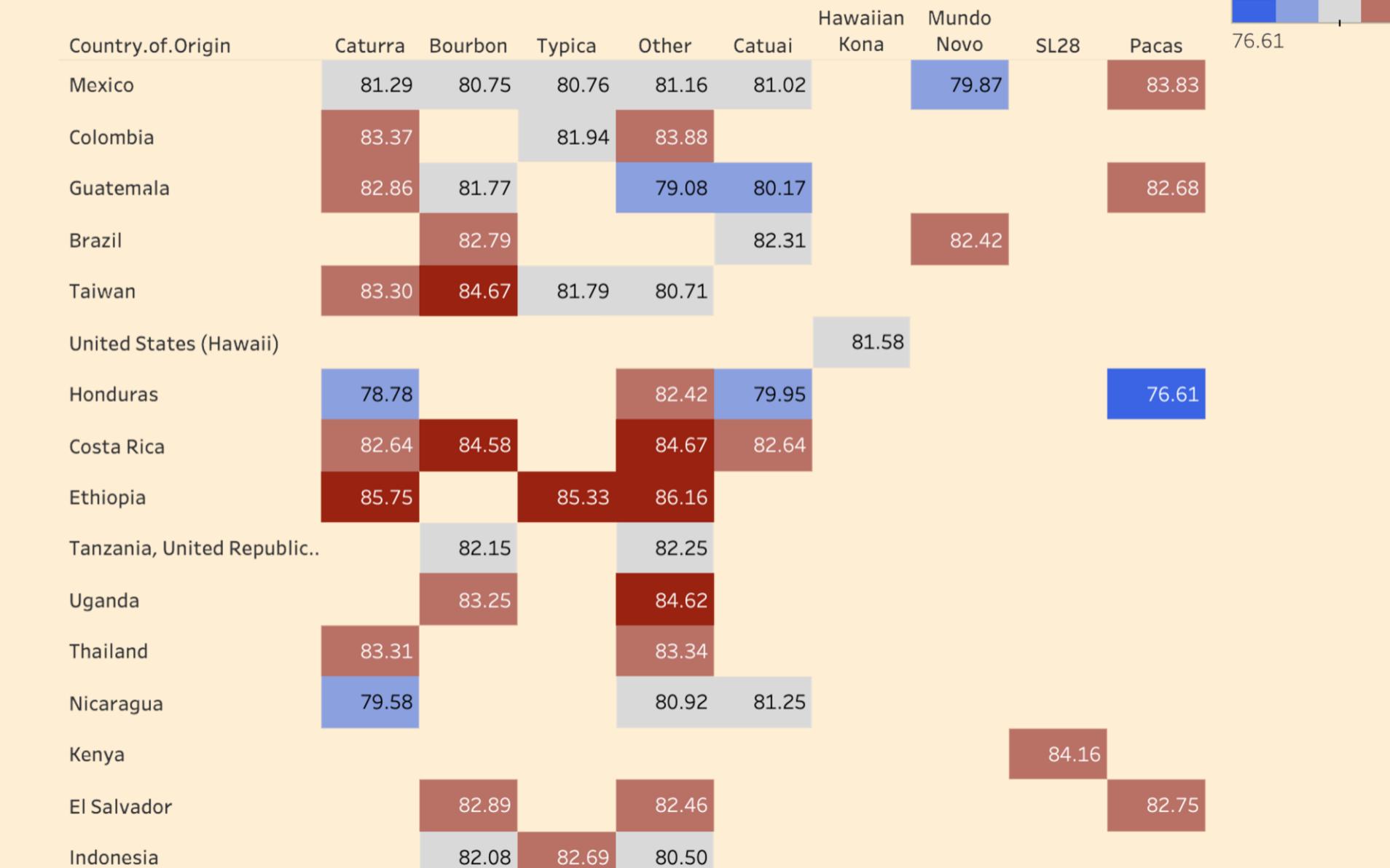
04 Evaluation

Using MSE to evaluate models



Models	MSE
Linear Regression	4.293e+25
Lasso	0.687
Random Forest	0.979
Gradient Boost	0.692
NuSVR	0.624

Varieties of Beans by Countries



- On average, each country produces only 4-5 types of varieties, while there are 40+ varieties in total.
- The average score of each variety type fluctuates significantly across different countries.

Tips for Coffee Lovers

- What makes an exceptional cup of coffee?

Geographic factors, processing methods, and the variety of beans are intertwined in determining good coffee.

- Which coffee-growing regions produce higher-quality coffee?

Ethiopia, Uganda, Kenya and Columbia produce good coffee in general.

Mexico and Guatemala provide more choices, but the quality varies.

- What do we know about varieties and processing methods?

Some common coffee varieties may not guarantee quality (be aware of Typica).

Semi-washed and pulped natural methods are the best performed, highlighting the importance of ongoing advancement in coffee processing techniques.

