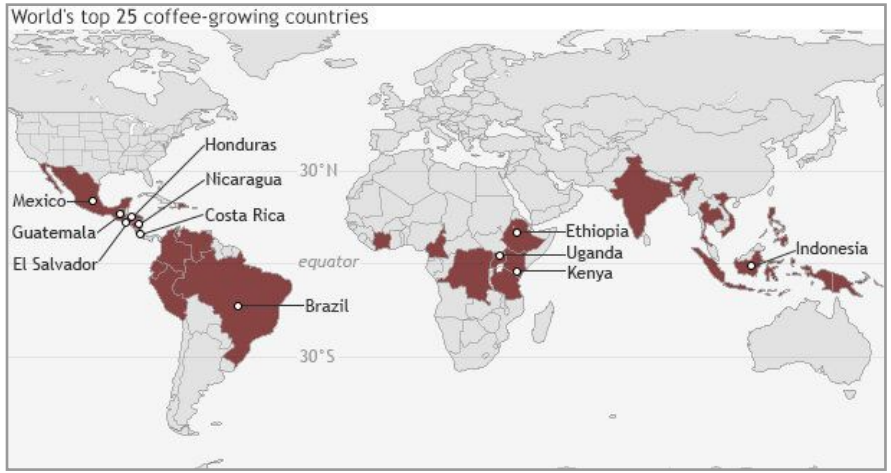
Coffee Quality Analysis

Pablo Aganza, Moez Hudda, Mohammed Elzubeir, Sonia Song

History of Coffee

- Originated in Ethiopia in the 11th century according to legends
- Became popularized in Europe and subsequently the US in the 17th Century
- According to NOAA Climate.gov, coffee lovers consume more than 2.25 billion cups of coffee a day. It is among the most valuable tropical exports on the planet.
- Ideal conditions to grow coffee: cool to warm tropical climates and rich soil. For this reason, the top coffee-growing countries tend to concentrate in countries along the Equator.



Hypothesis & Question

Hypotheses: Given origins in highlands and due to marketing, altitude has a positive relation to coffee quality
Null:

- Altitude has no effect on coffee quality
- Variables defined in data section have no effect on coffee quality

Main Question
Which attributes of coffee influence the ranking of Coffee?

Sub Questions:
Which Countries produce the highest ranked coffee?
Which Countries produce the lowest ranked coffee?
Are there patterns we see in characteristics about coffee which reinforce why a country may be ranked higher or lower?

Data

Leveraged 3 Data Sets

Coffee Distribution

Coffee Quality

Weather & Climate

Merged tables, dropped columns and began our exploration

Ran a number of calculations and created new tables for Means of each characteristic by country

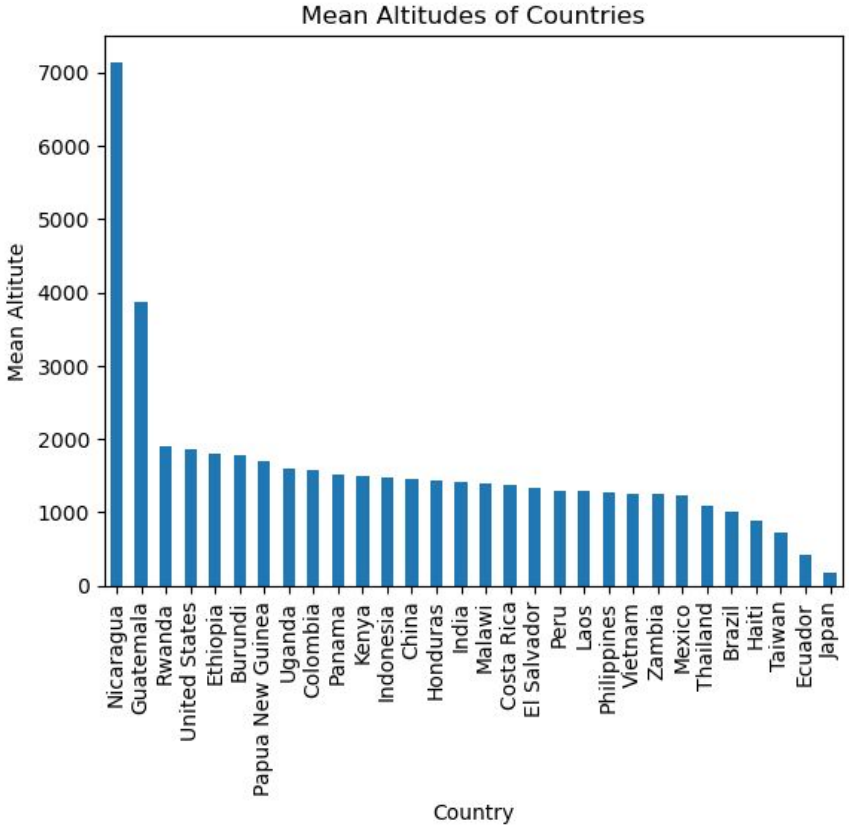
Variable Definitions:

- Acidity: Acidity in coffee refers to the brightness or liveliness of the taste.
- Altitude: Height of the coffee farm within the country
- Balance: Balance refers to how well the different flavor components of the coffee work together.
- Moisture: The amount of water within the coffee beans
- Sweetness: It can be described as caramel-like, fruity, or floral, and is a desirable quality in coffee.
- Uniformity: Uniformity refers to the consistency of the coffee from cup to cup.
- Ranking: Score given on the coffee from the data set.

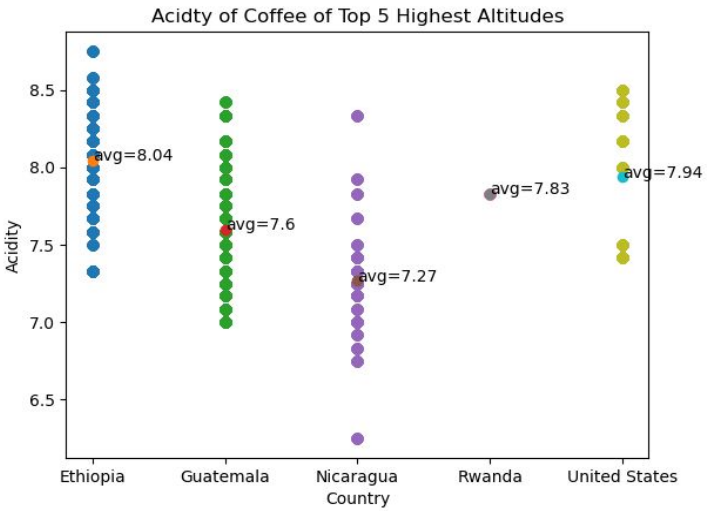
Graphs and Discussion

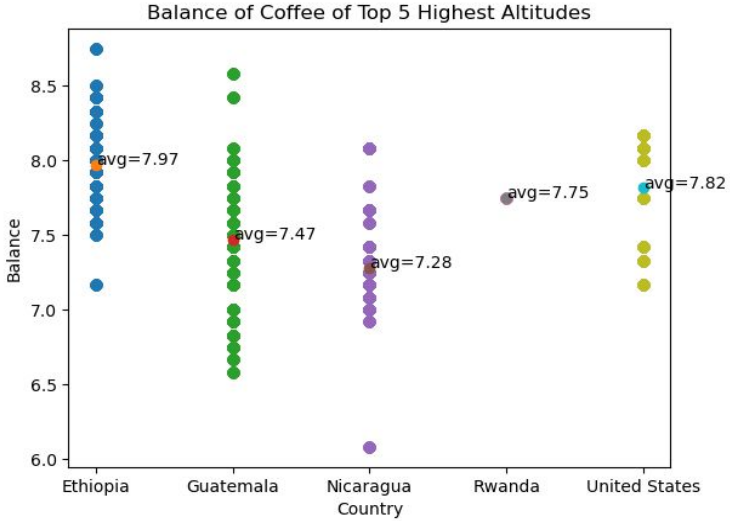
- From our hypothesis, we looked at the countries with the highest altitudes
- From this, we see the countries with the highest mean altitudes are Nicaragua, Guatemala, Rwanda, United States (Hawaii), and Ethiopia.
- Notable countries are Nicaragua and Guatemala as they are far above the average mean altitude of all countries

- Next, we can look at the other six variables defined in the data section (Acidity, Balance, Moisture, Sweetness, Uniformity, and Ranking) for these five high altitude countries.



Acidity and Balance

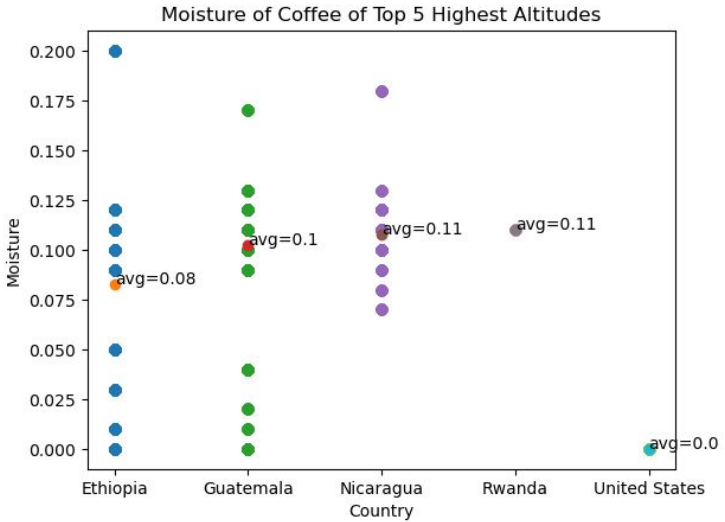


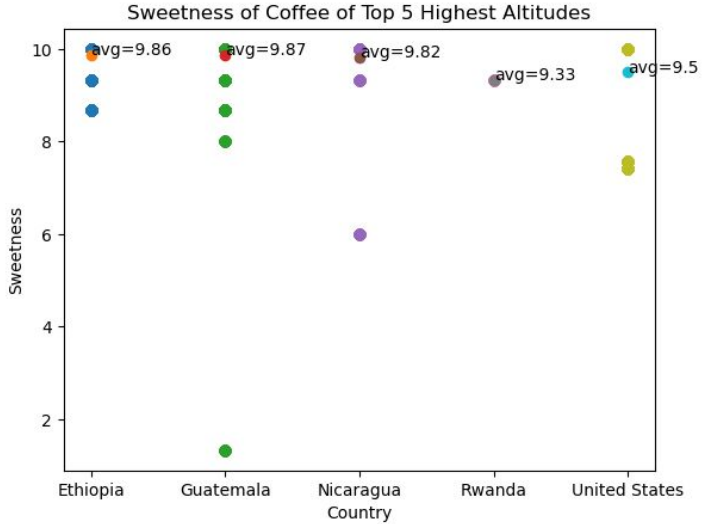


- Notice how Nicaragua and Guatemala have noticeably lower averages in both Acidity and Balance than the other countries

Moisture and Sweetness

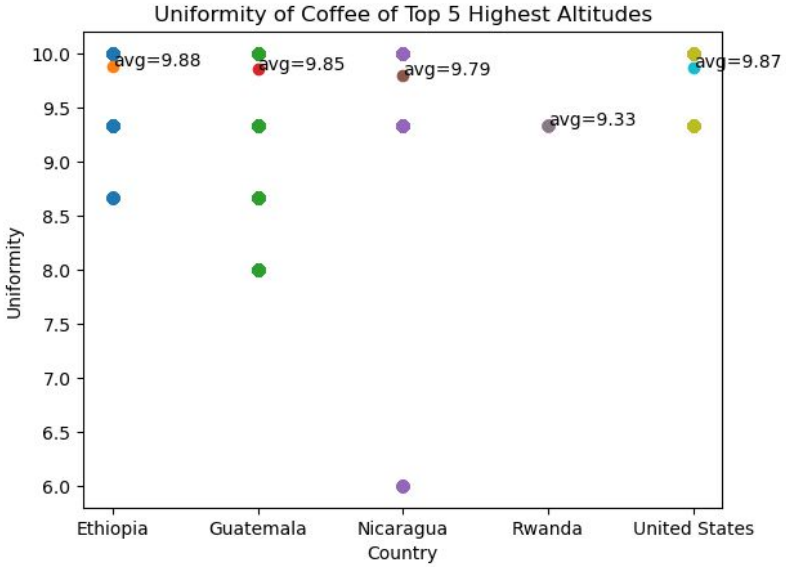
- For both Moisture and Sweetness, averages are about the same amongst all countries

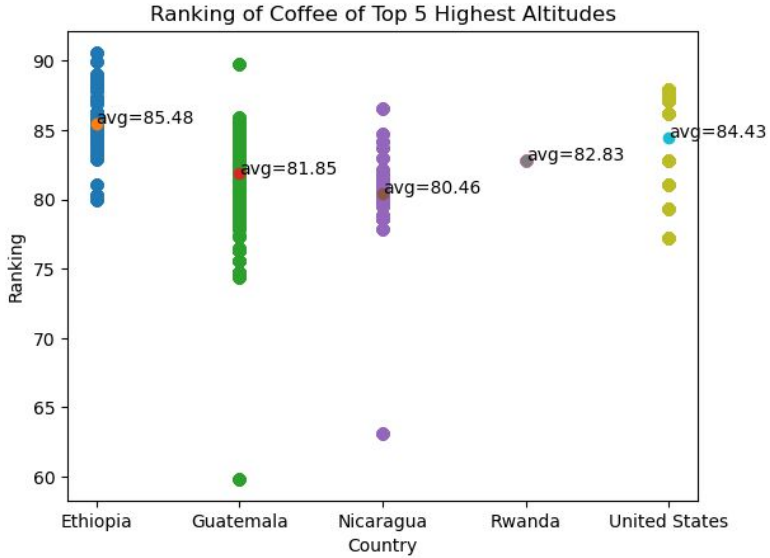




- United States (Hawaii) has zero for Moisture, most likely indicating that this metric was not collected for these coffee beans

Uniformity and Overall Ranking

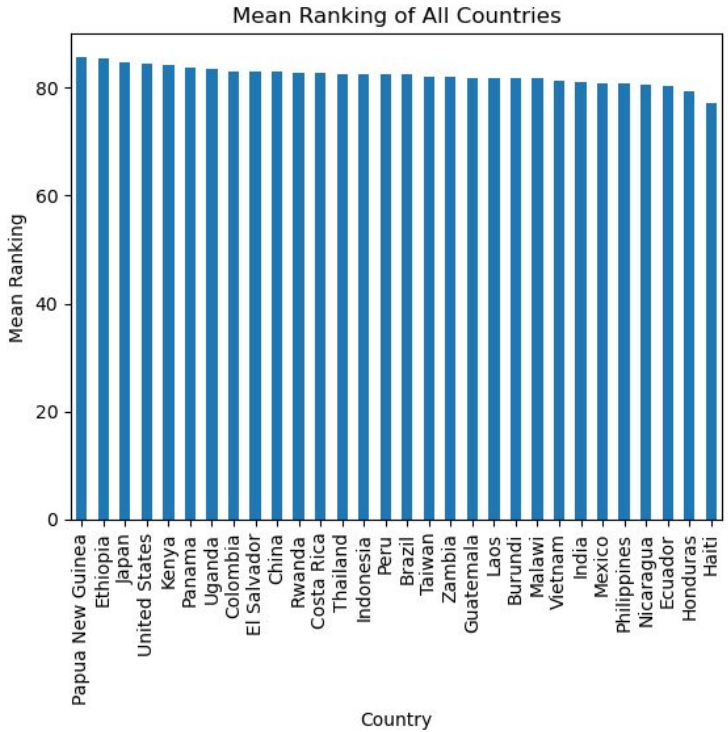




- For Uniformity, averages are all fairly close as well, excluding Rwanda which has a slighter lower average than the rest
- Based on this, we see that Nicaragua, being the country with the highest altitude by far of over 7000 meters, has the lowest ranking of these five countries.

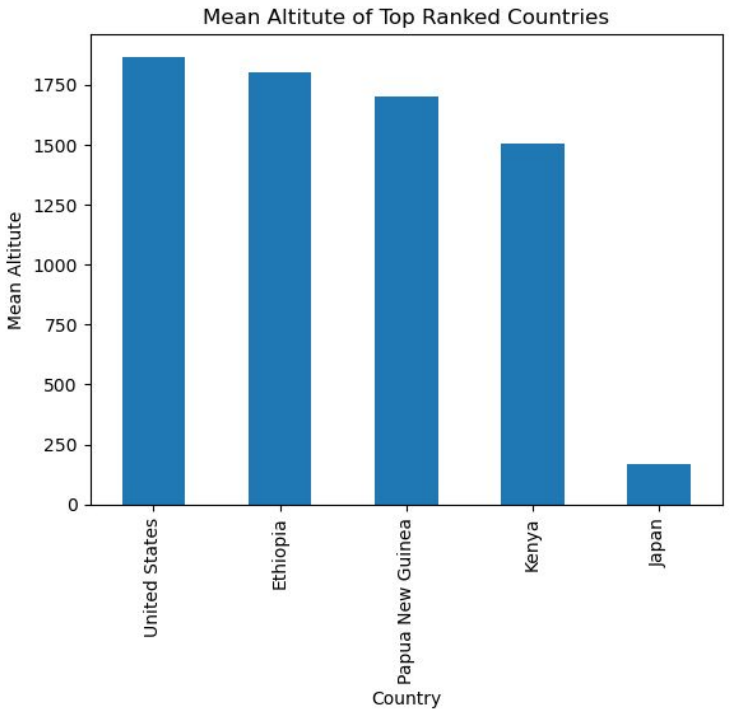
Rank of All Countries

- From here, we see that the top five ranked countries are Papua New Guinea, Ethiopia, Japan, United States (Hawaii), and Kenya. Comparing this with the highest altitude countries, we see that rankings out of 30 positions are:
    - Nicaragua - 27th
    - Guatemala - 19th
    - Rwanda - 11th
    - United States (Hawaii) - 4th
    - Ethiopia - 2nd



Altitude of Top Rank

- From this, we see our hypothesis of altitude having a direct positive correlation to the coffee quality is in question.
- Infer that average altitude in the range of 1500-1750 meters is most ideal for growing quality coffee, and having too high of an altitude can result in lesser quality coffee



Regression

$$CQI = \beta_0 + \beta_1 \cdot Sweetness + \beta_2 \cdot Moisture + \beta_3 \cdot Balance + \beta_4 \cdot Acidity + \beta_5 \cdot Altitude + \beta_6 \cdot Body + \beta_7 \cdot Uniformity + \epsilon$$

Dep. Variable:	Total Cup Points	**R-squared:**	0.947
Model:	OLS	**Adj. R-squared:**	0.930
Method:	Least Squares	**F-statistic:**	55.63
Date:	Sat, 13 Apr 2024	**Prob (F-statistic):**	1.57e-12
No. Observations:	30	**Log-Likelihood:**	-15.533
Df Residuals:	22	**AIC:**	47.07
Df Model:	7	**BIC:**	58.28
Covariance Type:	nonrobust		

	coef	std err	t	P> |t|	[0.025	0.975]
const	-1.2929	7.456	-0.173	0.864	-16.756	14.170
Sweetness	0.9060	0.198	4.581	0.000	0.496	1.316
Moisture	-0.6192	3.027	-0.205	0.840	-6.896	5.657
Balance	3.5673	1.156	3.087	0.005	1.171	5.964
Acidity	2.9133	0.818	3.563	0.002	1.218	4.609
Altitude (mean meters)	-3.648e-05	7.55e-05	-0.483	0.634	-0.000	0.000
Body	0.2879	0.582	0.494	0.626	-0.920	1.495
Uniformity	2.4100	0.572	4.215	0.000	1.224	3.596

Omnibus:	7.791	**Durbin-Watson:**	2.444
Prob(Omnibus):	0.020	**Jarque-Bera (JB):**	8.528
Skew:	-0.556	**Prob (JB):**	0.0141
Kurtosis:	5.363	**Cond. No.**	1.74e+05

Regression Discussion

- After re-reading the documentation on the dataset we discovered that in fact Total Cup Points was merely the numerical sum of the other factors (exact function underlying the CQI variable is already known).
- This would be akin to predicting a person's weight using Body Mass Index (BMI) which is defined as weight/height^2.
- Future analysis would entail possibly finding an alternative Coffee Quality Indicator (CQI) for the regression e.g the average scores given by experts in a blind test.
- With a P-value of 0.634, Altitude does not appear to have a statistically significant effect in determining coffee quality within the context of our current model. The altitude variable is not collinear with CQI as it isn't an input factor in the construction of the CQI

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.74e+05. This might indicate that there are

strong multicollinearity or other numerical problems.

References

- A Brief History Of The Price of Coffee. Sprudge Special Projects Desk. https://specialprojects.sprudge.com/?p=353.
- Climate zone finder. Global Historical Weather and Climate Data | Weather and Climate. 2024. https://weatherandclimate.com/.
- Coffee Prices - 45 Year Historical Chart. Macrotrends. April 15, 2024. https://www.macrotrends.net/2535/coffee-prices-historical-chart-data.
- Fatih B. Coffee Quality Data (CQI May-2023). Kaggle. https://www.kaggle.com/datasets/fatihb/coffee-quality-data-cqi.
- Scott, Michon. Climate & Coffee. June 19, 2015. https://www.climate.gov/news-features/climate-and/climate-coffee#:~:text=Optimal%20coffee%2Dgrowing%20conditions%20include,the%20Middle%20East%3B%20and%20Asia.
- The History of Coffee. NCA. https://www.ncausa.org/about-coffee/history-of-coffee#:~:text=An%20Ethiopian%20Legend&text=There%2C%20legend%20says%20the%20goat,want%20to%20sleep%20at%20night.