CONCLUSIONS: COFFEE QUALITY PROJECT

QUESTION 1:

Where in the world does the best coffee come from?

An interactive map was created to display the regions and countries that produce the greatest amount of highest rated coffee. Based on this map a majority of the best coffee comes from countries near Central America and closer to the Equator worldwide. An accompanying bar graph further breaks down this metric to show that Columbia, by far, does in fact produce the highest rated coffee in the world according to our data set. The closest runner up to Columbia is Guatemala, with Mexico, Ethiopia and Brazil in a dead heat for third place.

QUESTION 2:

Is Colombian coffee all that better from the rest of the world?

H0: High coffee score distribution across all quality metrics from Colombia is not very different from the world

H1: High coffee score distribution across all quality metrics from Colombia is different from the world

1. Aroma, Aftertaste, Body and Balance all have greater proportion of "High Scores" in Colombia vs. the World - as apparent from the stacked bar charts

2. A chi-square test was performed on the distribution of "high" scores across various coffee quality metrics for Colombia and the World.

   With 95% confidence interval - the Chi-Square statistic for Colombia(44.48) is higher than critical value (14.06) leading to a low p-value (1.7236707541760223e-07)

   Hence, we accept the alternate hypotheses that the distribution of high scores across coffee quality metrics in Colombia is indeed different from the World and significant.

QUESTION 3:

1. The pulped natural/honey processing method has the least aroma and flavor, but highest body.

2. We used means to normalize the data because there were more records of some processing methods than others.

3. ANOVA write-up -- "Does the processing method significantly impact the flavor, body or aroma of coffee?"

4. We wanted to determine if there was a statistically significant difference between the impacts of processing methods on key coffee attributes: flavor, aroma, and body. We had read that these were the attributes that were most impacted by the processing method, and wanted to use our dataset to evaluate if that were true or not. Given that we had 5 categorical groups that we wanted to test (each of the 5 processing methods), we used an ANOVA test in place of a t-test.

QUESTION 4:

1. First we tested the impact of processing method on coffee flavor. Here our null hypothesis was that processing method has no impact on coffee flavor. Our alternative hypothesis was that processing method has significant impact on coffee flavor. The p-value that resulted (0.0007457874514009814) is small enough that we can reject the null hypothesis and confidently say that the processing method has an impact on coffee flavor. Through the boxplot we also see that natural/ dry and washed/ wet methods have more outliers, while the pulped natural/ honey method has a smaller distribution.

2. Next, we used an ANOVA to assess the impact of processing methods on coffee aroma.

3. Our null hypothesis was that the processing method has no impact on coffee aroma, and our alternative hypothesis is that processing method has significant impact on coffee aroma. In this case, the p-value is large enough (0.1948126243804556) that we have to accept the null hypothesis that the processing method doesn't have an impact on aroma.

4. Lastly, we tested the impact of processing method on the body of coffee using an ANOVA. Our null hypothesis was that the processing method has no impact on the body of coffee, while our alternative hypothesis is that the processing method does have a significant impact on the body of coffee. For this test, the p-value was very small (1.2368895749915461e-06) leading us to reject the null hypothesis and accept that the processing method has significant impact on the body of coffee.

QUESTION 5:

1. We wanted to compare the origin and flavor profiles of the two types of coffee bean species: Arabic and Robusta. We have read that Arabica represents a majority of coffee that's out there (especially when it comes to single farm coffee) and is known for it's

smoother taste; compared to Robusta which is known to have a harsher and more bitter taste.

2. In our dataset, almost 98% of coffee beans were Arabica:

3. While Arabica is grown all over the world, Robusta primarily comes from India and Uganda. The country with the greatest number of Arabica coffee bean farms is Mexico, representing 18% of all Arabica coffee bean farms. Columbia, Ethiopia and Brazil are also home to a significant percentage of Arabica coffee bean farms.

4. That said, what we really wanted to know is how species play a role in the quality of coffee and its individual attributes. We see that the overall coffee scores in our dataset were very similar for Arabica and Robusta. However, despite being known for having a harsher taste, single origin Robusta beans in our dataset scored higher for flavor, aroma, aftertaste and acidity. This could be a matter of the sample of Robusta in our particular data set being so small, that it isn't truly representative of robusta as a species. That said, according to our analysis, robusta should not be looked over when on the hunt for good quality beans."