W203 Lab 2 – Group 2
Date: 03-23-2022
Ken Trinh, Michelle Lee, Tanmay Mahapatra

Research Proposal

Coffee is one of the most popular non-alcoholic beverages globally prized for its aroma and caffeine content. Coffee is brewed from roasted beans of the plant species Coffea, which is native to sub-Saharan Africa and individual islands in the Indian Ocean. Ever since its discovery, coffee has grown to more than 70 tropical countries, making up a multi-billion dollar market.

According to "Topic: Coffee market in the United States" published in 2002 by Statista, the United States has the largest coffee market worldwide in terms of exporting and consumption with revenue reaching over 80 billion U.S. dollars (Ridder, 2002). Despite being consumed for over 120 years, there are still new trends entering the market every year, shifting the popularity of different coffee types. Some of these trends include different beans, different caffeine concentration, and different brewing techniques. Because of the constant changes in coffee trends, everything from the variety of the plant, the chemistry of the soil, the weather, the amount of rainfall and sunshine, and even the precise altitude at which the coffee grows can affect the taste of the final product.

In an effort to understand what drives the popularity of coffee, our team proposed a study into a variety of factors attributed to the coffee bean itself and how they associate with the rating of different types of coffee. In this research, we will be addressing the following question:

How does the country of origin affect customers' rating of coffee?

In order to answer this question, we will use the 2020 coffee rating dataset gathered by tidytuesday (Rfordatascience, 2020). We plan to tackle this problem by building several models to capture the question. The first model will capture the effect of country of origin on coffee ratings. Because there could be other confounding variables, we want to also look into other covariates including flavor and processing method. These features will be added to a second model to capture their effect. Additional models will be needed in the study to provide a more conclusive result.

Our research team believes that by answering this question would enable us to identify how different factors affect the rating of coffee. Ultimately, the result of our analysis would help Acme with identifying new coffee products to include in the store by prioritizing features as well as with improving our marketing strategy.

References

Rfordatascience. (2020). *Tidytuesday/readme.md at master · rfordatascience/tidytuesday*. GitHub. Retrieved March 23, 2022, from https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-07-07/readme.md

M. Ridder (2002). *Topic: Coffee market in the United States*. Statista. Retrieved March 23, 2022, from https://www.statista.com/topics/1248/coffee-market/#topicHeader__wrapper