What Make Coffee Popular?

Ken Trinh, Michelle Lee, Tanmay Mahapatra

03/30/2022

Contents

1   Introduction

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. 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 an investigation following the question:

> *How does the country of origin affect the popularity of a coffee?*

Prior to looking at the data set, we conceptualized the popularity of a coffee as a rating point, out of 100, given to a cup of coffee and the country of origin as the country that the coffee bean came from. Some covariates that we planned to consider in the study are grouped into two categories: sense and taste and processing methods. Here we conceptualized the sense and taste category as any rating given to the cup of coffee that related to sense and taste factors of coffee. Finally, the processing methods as the methods that the coffee bean is processed by.

Our analysis will use the 2020 coffee data set gather by tidytuesday[1]. We planned to use some of the observational data collected as follows:

- First, we use column named `total_cup_points` that came with the data set as our measurement for our conceptualization for the popularity of coffee. This column is already given as a numerical value out of 100, so the closer to 100 means the more popular that specific cup of coffee.

- Second, we will apply one hot-encoding technique to `country_of_origin` and `processing_methods` to transform the string type data into integer type. Depending on the result of our exploratory analysis, we will selectively leave out one country and one processing method as dummy variable prior to performing our modeling. These two steps are performed in order to ensure we are fitting numerical values and also enhancing the interpretability of our model.

- Finally, we identified `aroma`, `flavor`, `aftertaste`, `acidity`, `body`, `balance`, `uniformity`, and `sweetness` as columns belonging to the sense and taste covariates group. In order for these variables to be meaningful, we plan to transformed them according to the result of our exploratory analysis. We hypothesize that controlling for these effect will provide some more insight into our initial questions

After the data is collected and processed, 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

---

[1]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

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 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.