

# Project title

## Brilliant Cassowary - Report

Nidhi Soma (ns848)      Joice Chen (jc3528)      Jinpeng Li (jl3496)  
Stephen Syl-Akinwale (sis33)

## Introduction

One of the most widely known beverages in today's age is coffee. It is present in many different settings, ranging from students and employees to casual coffee shop frequenters and critical coffee enthusiasts. Our current research questions are: - How does self-perceived coffee experience affect actual and stated preferences? - How does coffee preference and consumption habits differ by demographic attributes such as age, gender, education level, race, and political identity?

## Data description

*What are the observations (rows) and the attributes (columns)?* The observations represent an individual respondent to the survey. There are 4,042 rows. The columns are questions that they answered, ranging from demographic data to coffee preferences. There are 98 of these columns.

*Why was this dataset created?*

To understand the general public's preferences as consumers for coffee.

*Who funded the creation of the dataset?*

World champion barista James Hoffmann and Cometeer – a subscription service that makes flash-frozen coffee capsules.

*What processes might have influenced what data was observed and recorded and what was not?*

The survey quickly was sold out, and Hoffman's audience in general is coffee specialists. That will likely skew the population surveyed to be people who likely prefer specialty coffee, so it may be a biased sample. Additionally, this survey was conducted through people ordering tasting

kits online, which were then sent to the participants to prepare and complete voluntarily, so there may have been differences in that. One example is that because participants were following a livestream to demonstrate how to do their taste test, their coffees may have been out long enough to have cooled, which could be another unaccounted variable that affected their taste preferences.

*What preprocessing was done, and how did the data come to be in the form that you are using?*

Zip codes and geographic data seemed to have been removed. Participants were anonymized to protect their privacy. It wasn't disclosed how Hoffmann and his team collected all the taste test results that participants filled out, but once they got that data, they made it into a spreadsheet to be shared with the public.

*If people are involved, were they aware of the data collection and if so, what purpose did they expect the data to be used for?*

People were involved, and they were made aware of the data collection by a YouTube video stating that the purpose of this taste test was to understand coffee preferences in the USA. The participants had to order the coffee tasting kit on their own in order to participate, showing their willingness to accept these terms. Hoffmann also made his intentions clear in his video with why he wanted to collect the data, and that he was planning to publicize the raw data later on.

*Notes*

Coffee A - Light roast, Washed Coffee B - Medium Roast Coffee C - Dark roast Coffee D - Fermented, Natural, Fruity

## **Data analysis**

Addressing our first question on actual self-perceived coffee experience compared to stated preferences. We use examples of people stating their preferences and look at the difference in data where they have a taste test.

*Pre-registered hypotheses*

Our first pre-registered hypothesis was that younger people 18 - 25 prefer their coffee with lighter roasts.

**Evaluation of significance**

**Interpretation and conclusions**

**Limitations**

**Acknowledgments**