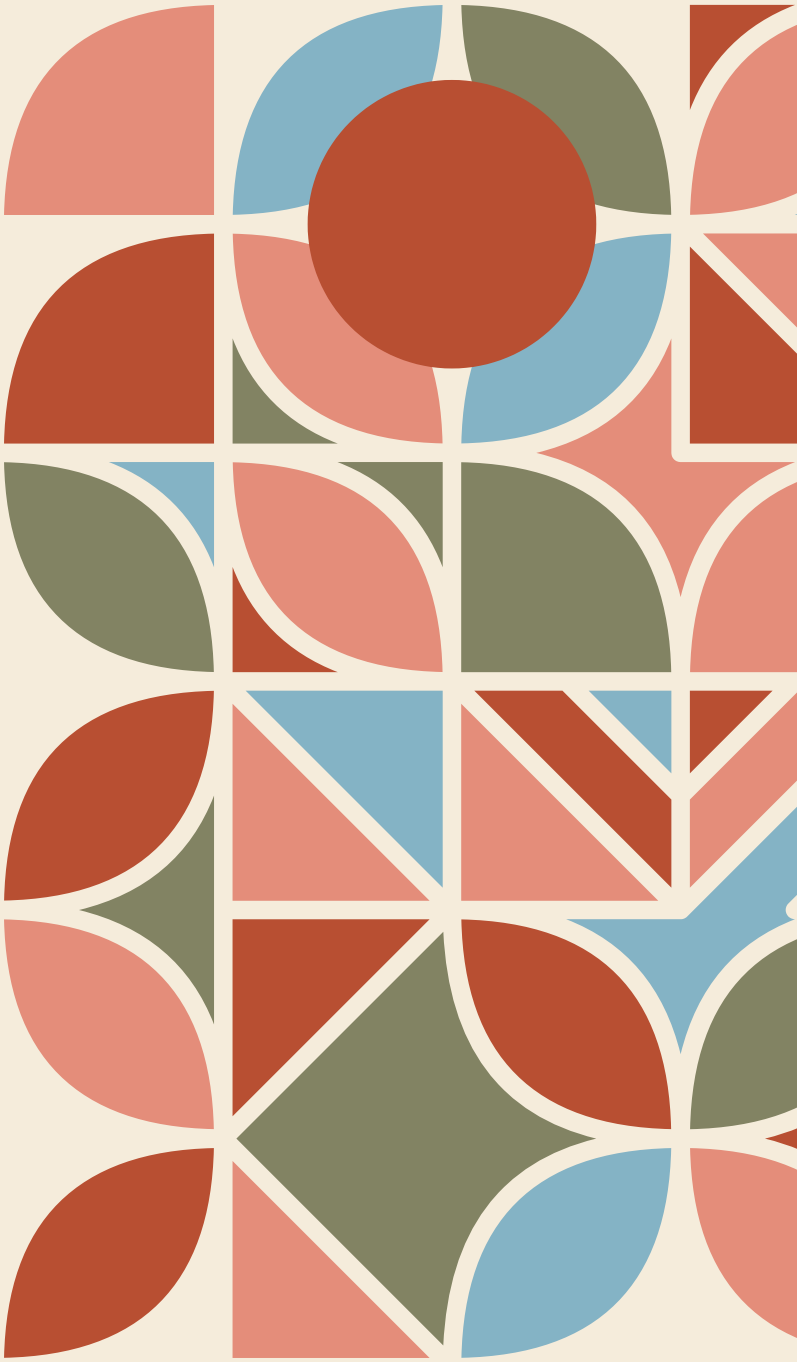
PROJECT 4 PROPOSAL 2023



Prepared By :

Cayley Morrow

Dominique Spencer

Jesslyn Lengkong

For:

UWA Data Analytics Bootcamp 2023

TABLE OF **CONTENTS**

PROJECT 4
PROPOSAL

PROJECT OVERVIEW

BACKGROUND

Coffee is one of the most popular beverages globally and is derived from the seeds of coffee berries. Two main species are Arabica and Robusta, and within each of these, there are numerous varieties.

Major coffee-producing countries include Brazil, Vietnam, Colombia, and Ethiopia. Factors like weather conditions, economic trends, and consumer preferences influence the coffee market.

The flavour, aroma, acidity, body, and other characteristics of coffee beans are influenced by the region where they are grown, altitude, soil composition, and processing methods.

Coffee enthusiasts often have specific preferences regarding the type of coffee they enjoy. Understanding and catering to these preferences is crucial in the coffee industry.

COFFEE RECOMMENDATION MODEL

Users will select specific criteria from our dataset to define their preferences for a coffee bean. Subsequently, our model will generate the top 5 recommendations tailored to their choices. Through a user-friendly drop-down menu, individuals can make their selections, and the model will present the top 5 recommendations, accompanied by additional details about each suggested coffee bean.

THE QUESTIONS WE WANT TO ANSWER ARE:

1. Isn't coffee just coffee?
    - Many variables can affect how a coffee tastes. This can be confusing to someone and make it difficult to articulate their preferences. Creating a platform that simplifies this but specifies these details (so people can learn!) is the goal!
2. We want to see if our model can accurately predict five (5) recommendations that match our end user's selection (~80%)
3. What contributions could this information have on the coffee market?
    - Drive decision-makers about green coffee markets
    - Influence farming practices/decisions
4. Would this model contribute to any positive impacts on a personal and global level?
    - Promote lesser-known coffee regions and beans - increase profit and funding
    - Positively influence an individual consumption experience
5. What would we want to do with this model in an ideal world?
    - Creation of an app?

QUESTIONS TO ANSWER

DATASETS TO BE USED

THE DATASETS THAT WILL BE USED:

- Arabica and robusta coffee scraped from the Coffee Quality Institute (CQI) database in January 2018.
- Data has come from https://github.com/jldbc/coffee-quality-database/tree/master.
    - If time permits, we will look at potential web scraping of 2023 CQI data.
- Our model will have 1340 rows of usable data that feature quality measures, bean metadata, and farm metadata.

DELEGATION OF TASKS

CAYLEY:

- API Creation
- HTML Help
- Machine Learning Help
- PowerPoint Presentation

DOMINIQUE:

- Data Cleaning
- Machine Learning
- ReadMe

JESSLYN:

- HTML/JavaScript/Streamlit
- Machine Learning Help
- ReadMe