The background of the slide is a photograph of coffee-related items. In the foreground, a light brown cardboard coffee tray holds two white disposable coffee cups with black lids. One cup is partially visible on the left, and another is more centrally located. In the background, slightly out of focus, is another white cup filled with a latte, featuring a white milk foam design on top. The entire scene is set against a dark, blurred background, likely a cafe counter.

# Coffee Recommendation System

---

By Joel Thomas



---

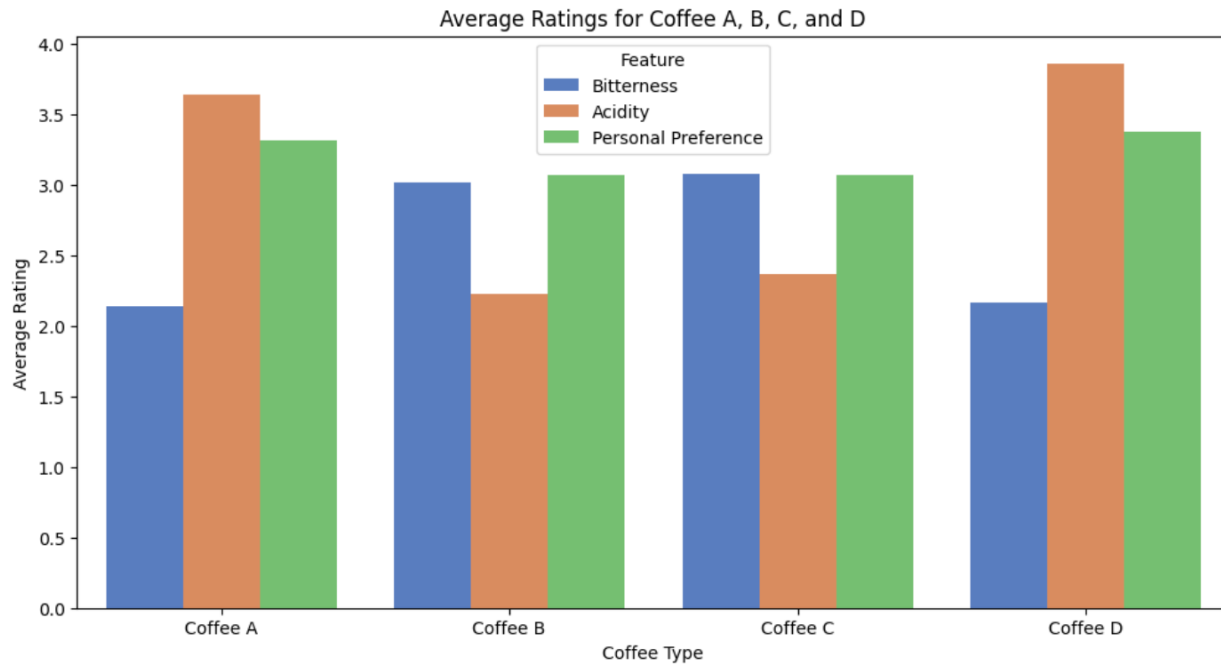
## Data set of interest

- `coffee_survey.csv` from [kaggle.com](https://www.kaggle.com)
- In October 2023, world barista champion James Hoffman and coffee company Cometeer held the "Great American Coffee Taste Test" on YouTube, during which viewers were asked to fill out a survey about 4 coffees they ordered from Cometeer for the tasting.
- I have decided to use this data to develop a model that can predict a person's preferred "style" of coffee based on bitterness and acidity ratings.



# Research Question

- What are most peoples preferred levels of bitterness and acidity ?
- How do these features relate with coffee notes?
- People who regularly drink coffee are looking to try new things, one way they approach this is by trying coffee from different farms and different roaster. But what if the approach was most taste based?
- It is important to me to answer the questions “I want something more acidic than usual today” or similar.
- A model that can output “notes” that are clearly visible on a bag of coffee based on how much acidity or bitterness a person is looking for in their cup of coffee.



- While not properly visible in the graph, Coffee D was most preferred. The high acidity and mild bitterness to balance it out is known to be one of the most popular styles of coffee.

- Preprocessing the data

- Dropping null values
- Only selecting relevant columns
- Then plotting to compare acidity, bitterness and rating.



# Model

- Random Forest Classifier was chosen to make a model since the data had numerical and categorical values.
- As seen in the figure the model is able to predict what “style” of coffee is preferred.
- This makes it much easier when recommending coffee especially if this is being used in an app or website to advertise to people preferences based on past purchases and ratings.

Accuracy: 0.3210161662817552

	precision	recall	f1-score	support
Bold	0.00	0.00	0.00	14
Bright	0.24	0.12	0.16	34
Caramalized	0.06	0.03	0.04	35
Chocolatey	0.28	0.45	0.35	71
Floral	0.00	0.00	0.00	20
Fruity	0.43	0.77	0.56	117
Full Bodied	0.12	0.09	0.10	47
Juicy	0.21	0.09	0.13	33
Nutty	0.17	0.12	0.14	26
Sweet	0.33	0.06	0.10	36
accuracy			0.32	433
macro avg	0.19	0.17	0.16	433
weighted avg	0.26	0.32	0.26	433

# Conclusion

- While this model is very useful in predicting “style”/ notes of coffee. It can easily be extended for more uses. Such as preferred roast levels, brewing methods etc.
- With a bit of data manipulation for higher accuracy and change in target variable a lot of new opportunities to study this data becomes available.
- It can easily be used in stores and online to help people when shopping for coffee