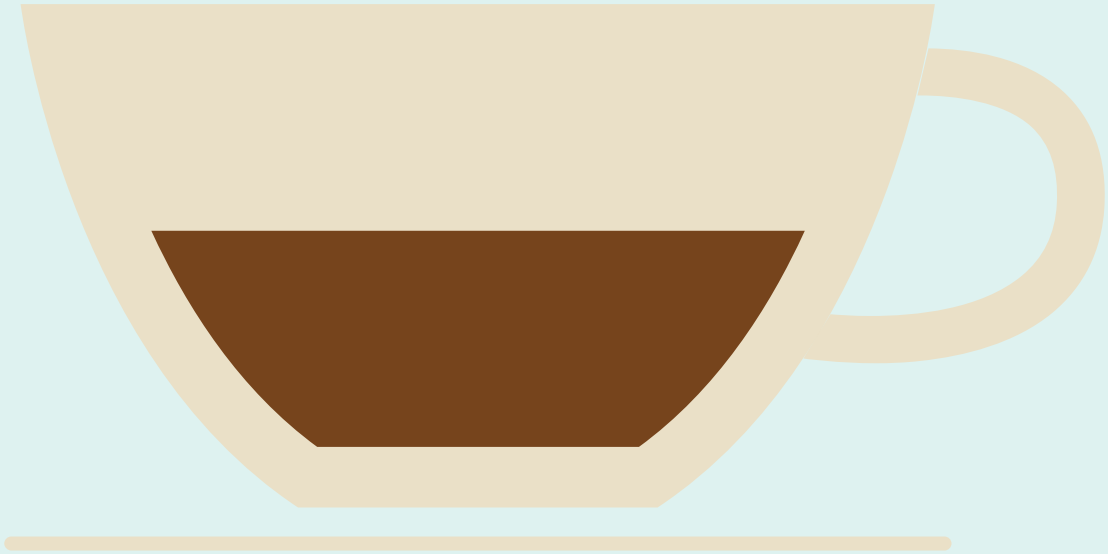
Caffeine Productivity & Rest Analysis

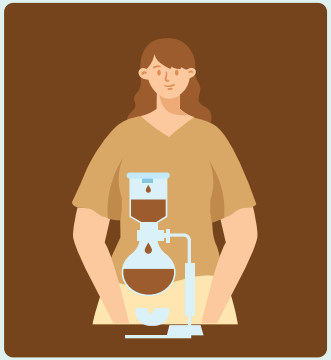
Group Better Buzz: Cooper Jumamil, Havanna Robbins, Haylee Marks-Mitchell, June Phillips

Table of Contents



Introduction

Background



- Lots of people rely on caffeine via coffee, tea, or energy drinks on a daily basis to:
  - Stay awake
  - Study fuel
  - Get through long days
- It is not always clear how each type of caffeinated beverage actually affects:
  - Focus during the day
  - Sleep quality at night
- We wanted to look at the data and see how caffeine truly affects people

Research Questions

1.

Does the type of caffeinated beverage (coffee, tea, or energy drink) affect focus levels?

2.

Does beverage type affect sleep outcomes, including sleep quality and whether sleep was negatively impacted?

3.

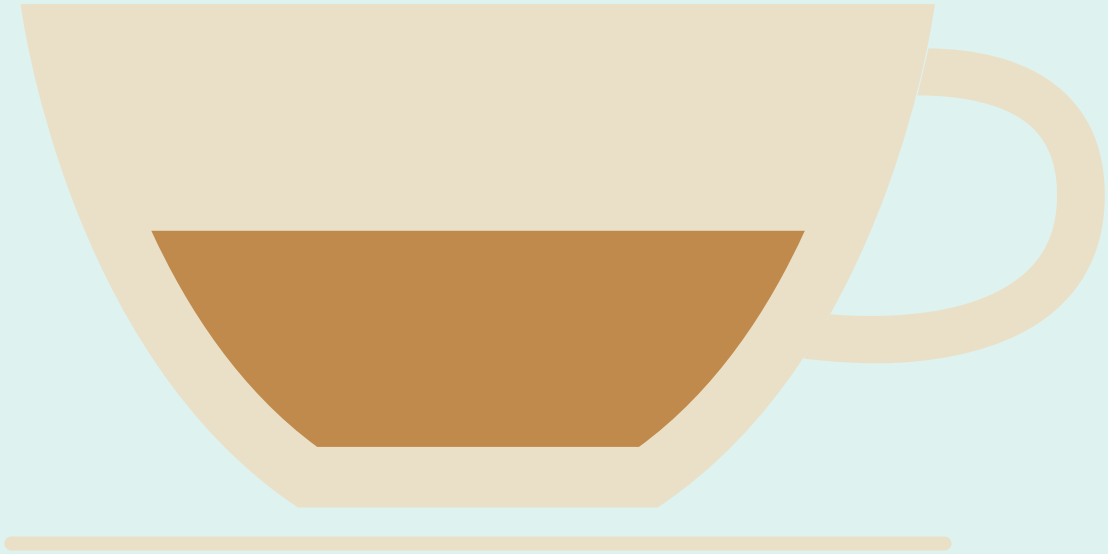
How do other factors like caffeine amount and time of day interact with beverage type when it comes to focus and sleep?

4.

(Stretch goal) Can we build a simple predictive model to estimate the probability that a person's sleep will be impacted, based on their caffeine intake pattern?

Hypothesis

- Coffee and energy drinks will have *higher focus levels* on average, but will have a *negative impact* on sleep and sleep quality as opposed to tea.
- More specifically, we expect:
  - Focus: Coffee and energy drink drinkers will report higher focus_level than tea drinkers.
  - Sleep quality: Tea drinkers will have higher average sleep_quality scores.
  - Sleep impact: Coffee and especially energy drinks will show higher rates of sleep_impacted = 1 than tea.



Data Sourcing & Processing

Data Set

- Found on Kaggle
- “Caffeine Intake Tracker”
- Collected by Preksha Dewoolkar
- 500 observations, 13 features
- Data collected through a comprehensive survey study involving 500 participants who tracked their caffeine consumption patterns and subsequent effects
- Data represents a snapshot of caffeine consumption habits:
    - impact on focus and sleep quality
    - Collected through standardized self-reporting methods

Data Columns

caffeine_mg

Amount of caffeine consumed (normalized)

age

Age of the participant (normalized)

focus_level

Self-reported focus level on a 0–1 scale

gender_female, gender_male

One-hot encoded gender variables

sleep_quality

Self-reported sleep quality on a 0–1 scale

sleep_impact

Binary flag (1 = sleep was negatively impacted, 0 = no impact)

beverage_coffee, beverage_tea, beverage_energy_drink

One-hot encoded beverage type

time_of_day_morning, time_of_day_afternoon, time_of_day_evening

When the person consumed caffeine

Data Processing

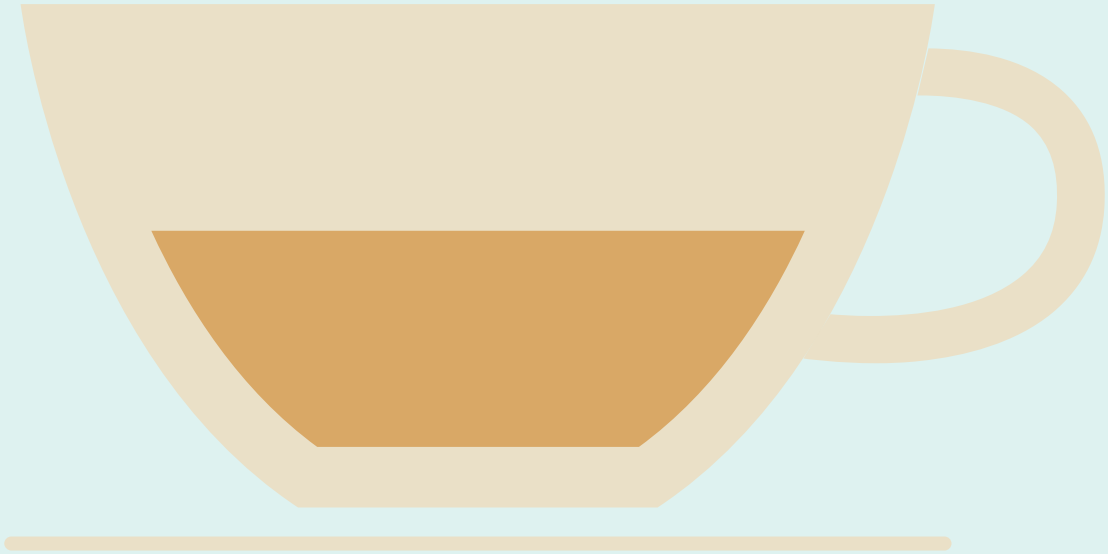
Loaded dataset using Pandas

Check for:
no missing values and correct data types

Combined Coffee, Tea, Energy Drink:
beverage_type

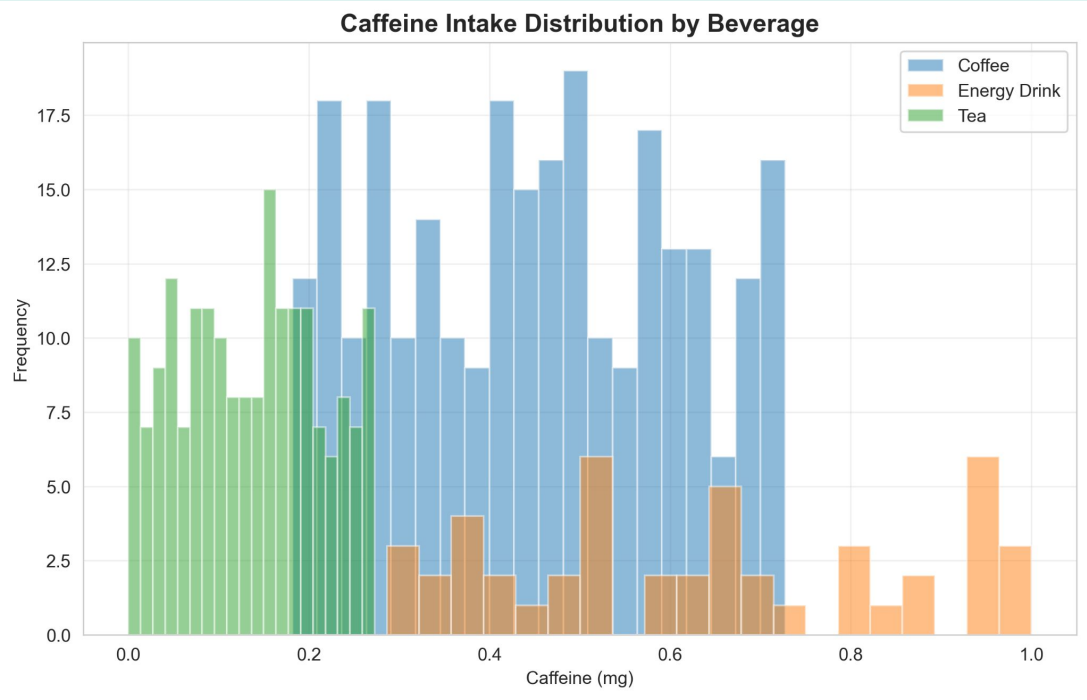
Verified distributions and normalized numerical fields when necessary

Prepared features for modeling:
caffeine_mg, age, one-hot beverage and time-of-day variables. gender indicators



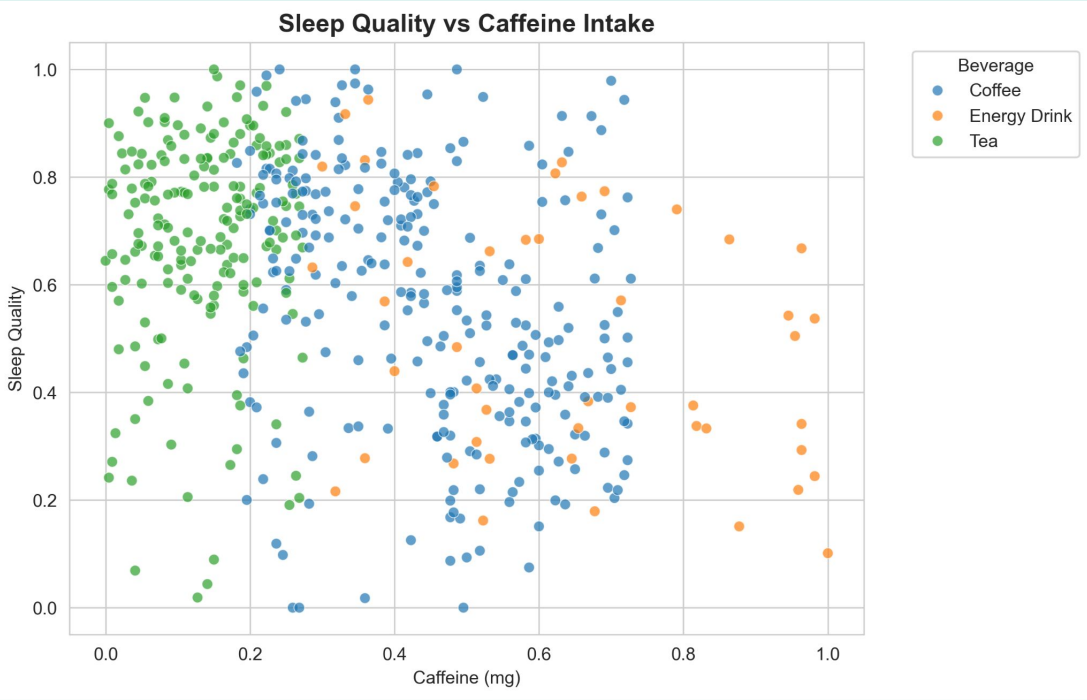
Exploratory Data Analysis

Caffeine Intake Distribution by Beverage



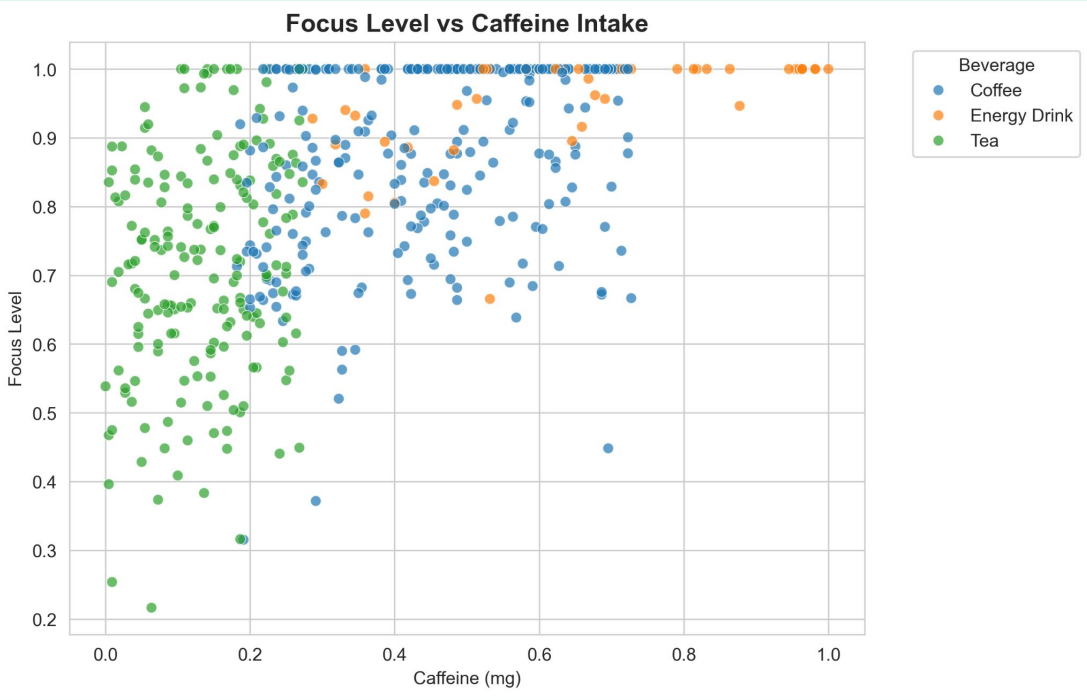
- Bar chart
- X axis = Caffeine (in milligrams)
- Y axis = Frequency of consumption

Sleep Quality vs. Caffeine Intake



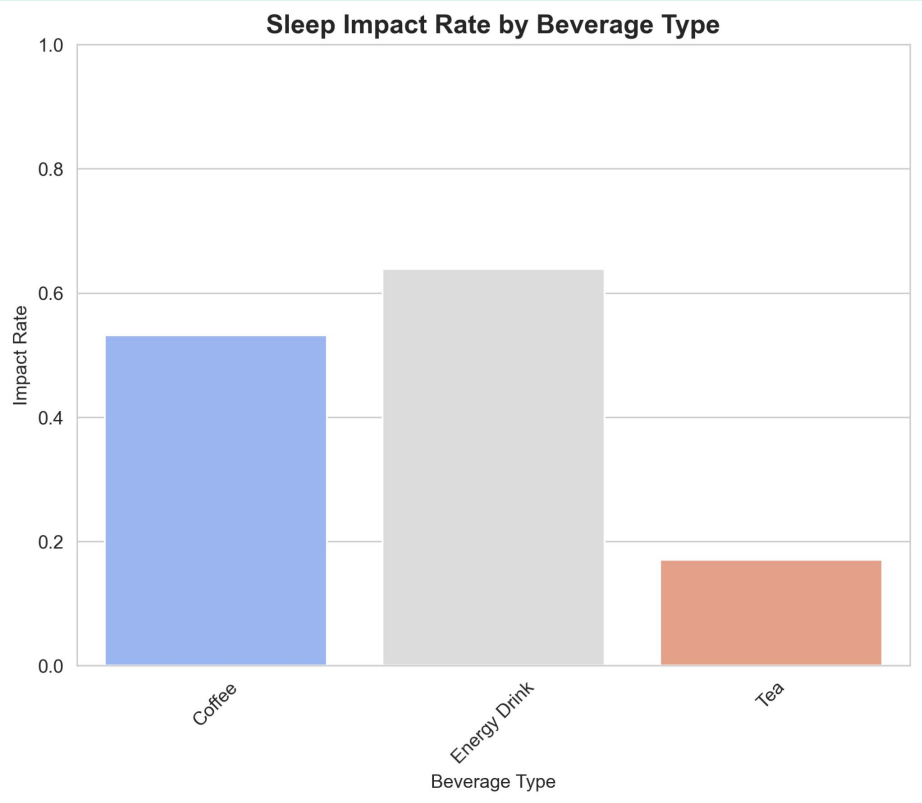
- Scatterplot
- X axis = Caffeine (in milligrams)
- Y axis = Sleep quality (scale of 0-1)

Focus Level vs. Caffeine Intake



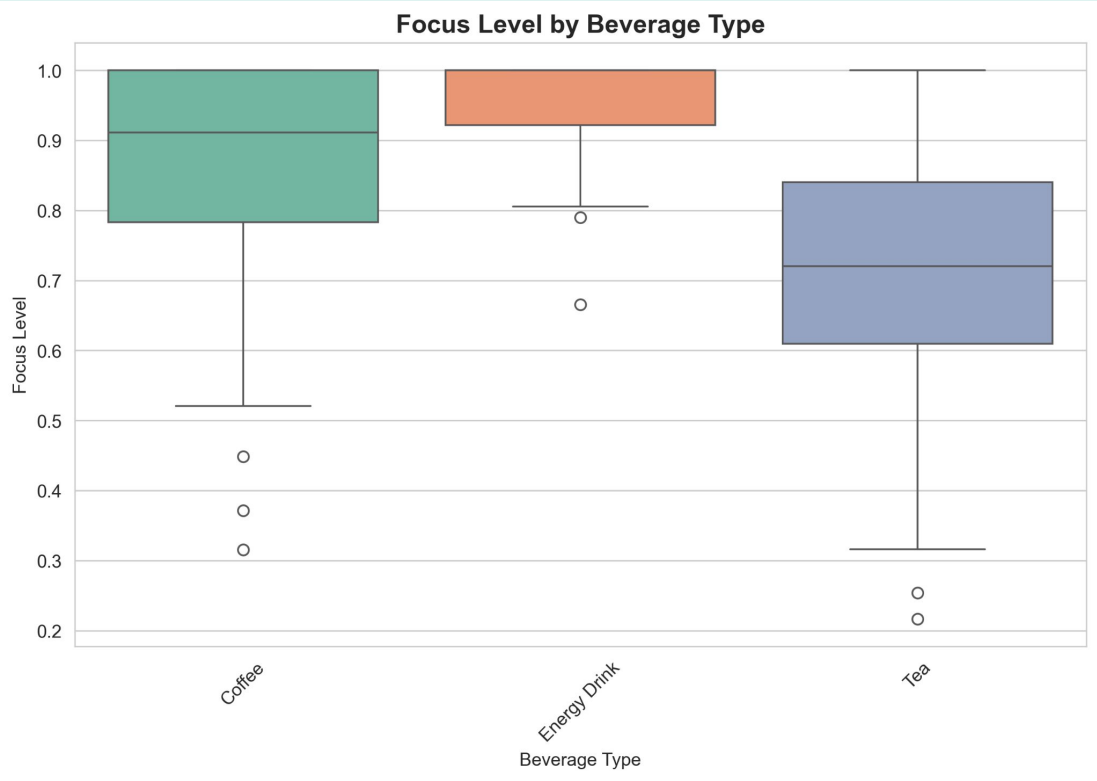
- Scatterplot
- X axis = Caffeine (in milligrams)
- Y axis = Focus Level (scale of 0-1)

Sleep Impact by Beverage Type



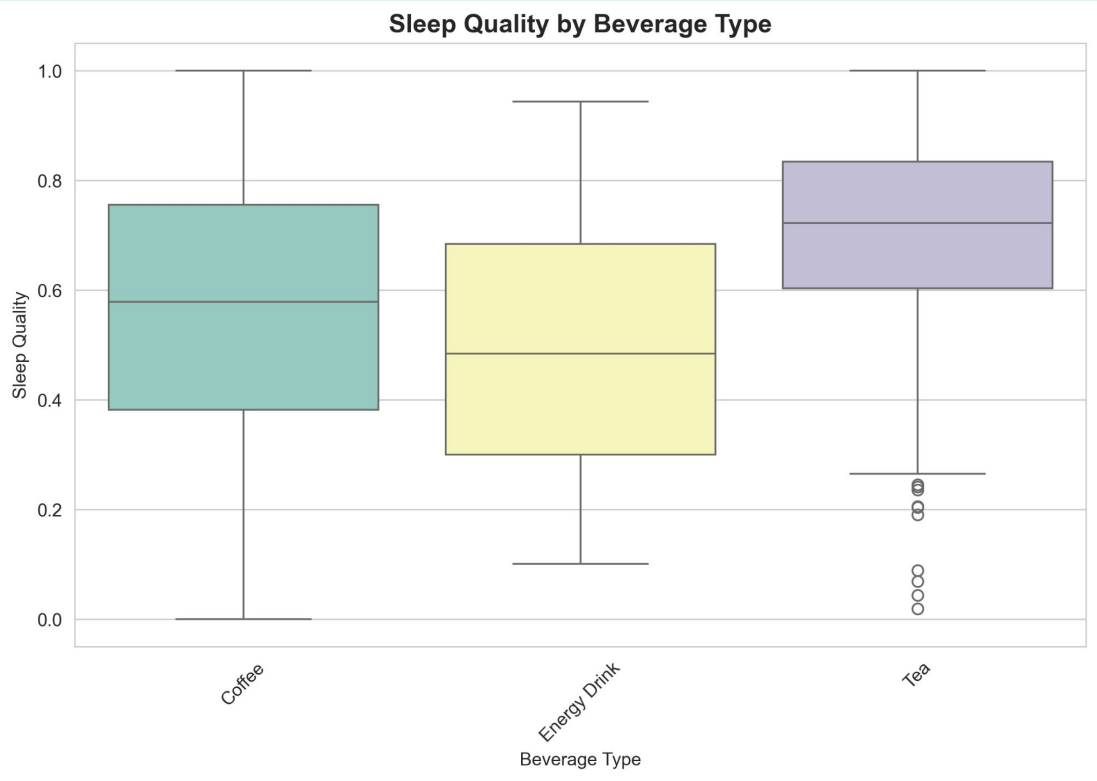
- Bar chart
- X axis = Beverage Type
- Y axis = Impact Rate (scale of 0-1)

Focus Level by Beverage Type



- Box plot
- X axis = Beverage Type
- Y axis = Focus Level (scale of 0-1)

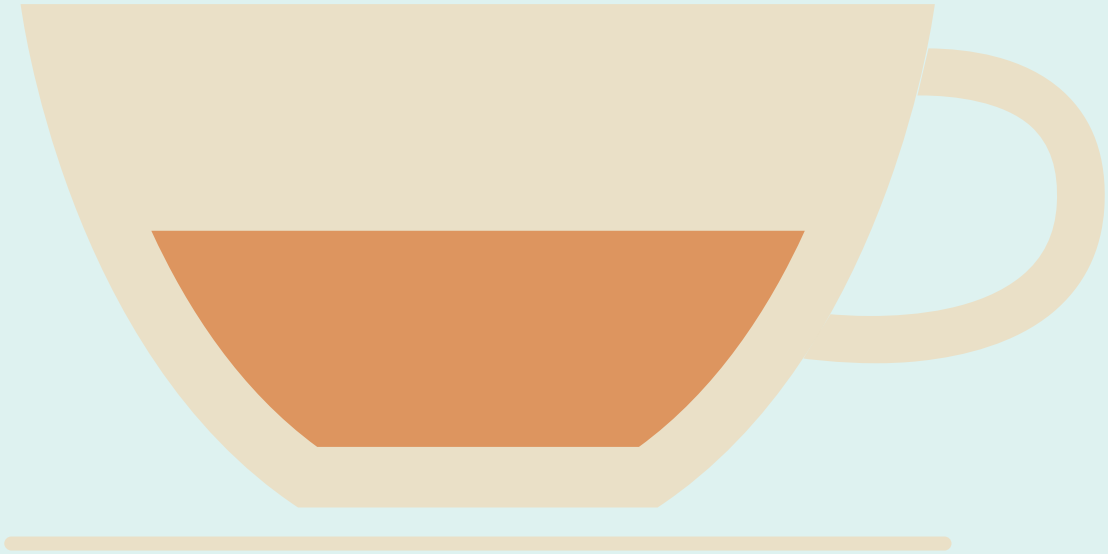
Sleep Quality by Beverage Type



- Box plot
- X axis = Beverage Type
- Y axis = Sleep quality (scale of 0-1)

Graph Interpretations

- Energy drink users tend to cluster at lower sleep quality scores
- Coffee has the widest spread of focus levels
- Tea appears to have the least negative impact on sleep



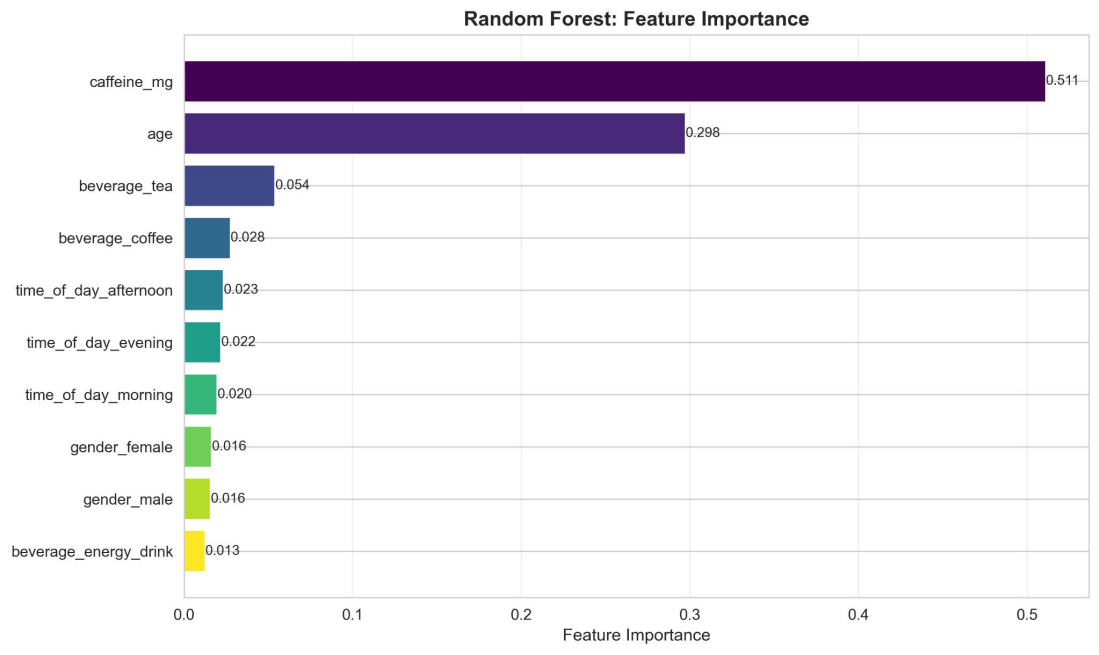
Modeling & Analysis

Models

Logistic Regression
Fast and interpretable
Good for understanding feature impact
Suitable for binary classification (sleep impacted: yes/no)

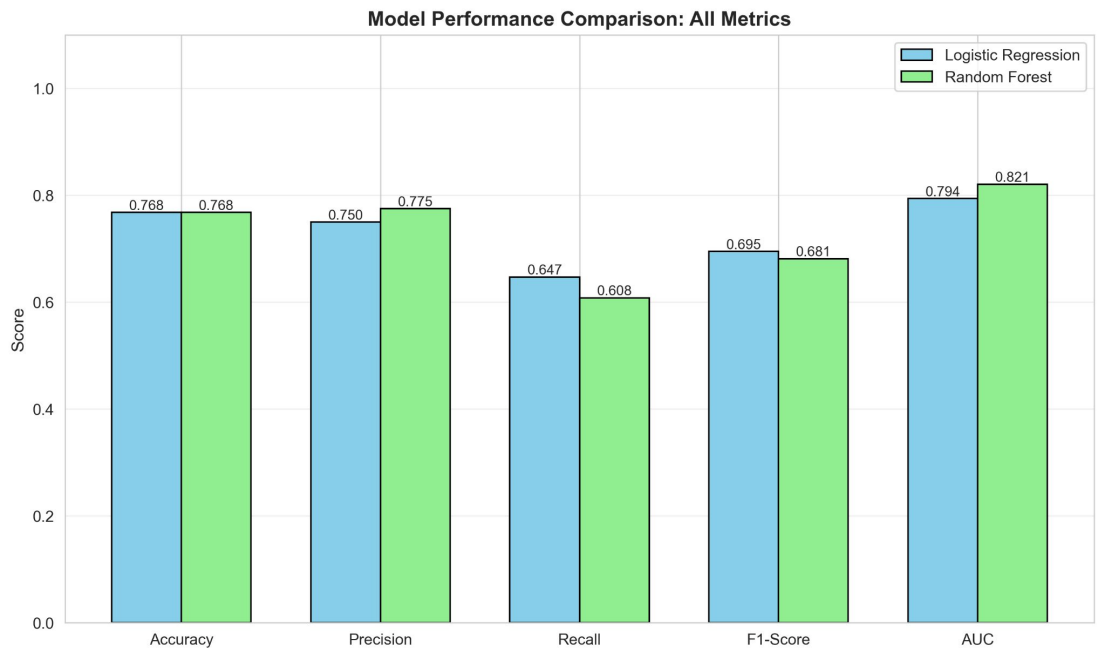
Random Forest
Captures complex interactions between features
Provides feature importance rankings
Less prone to overfitting

Random Forest: Feature Importance



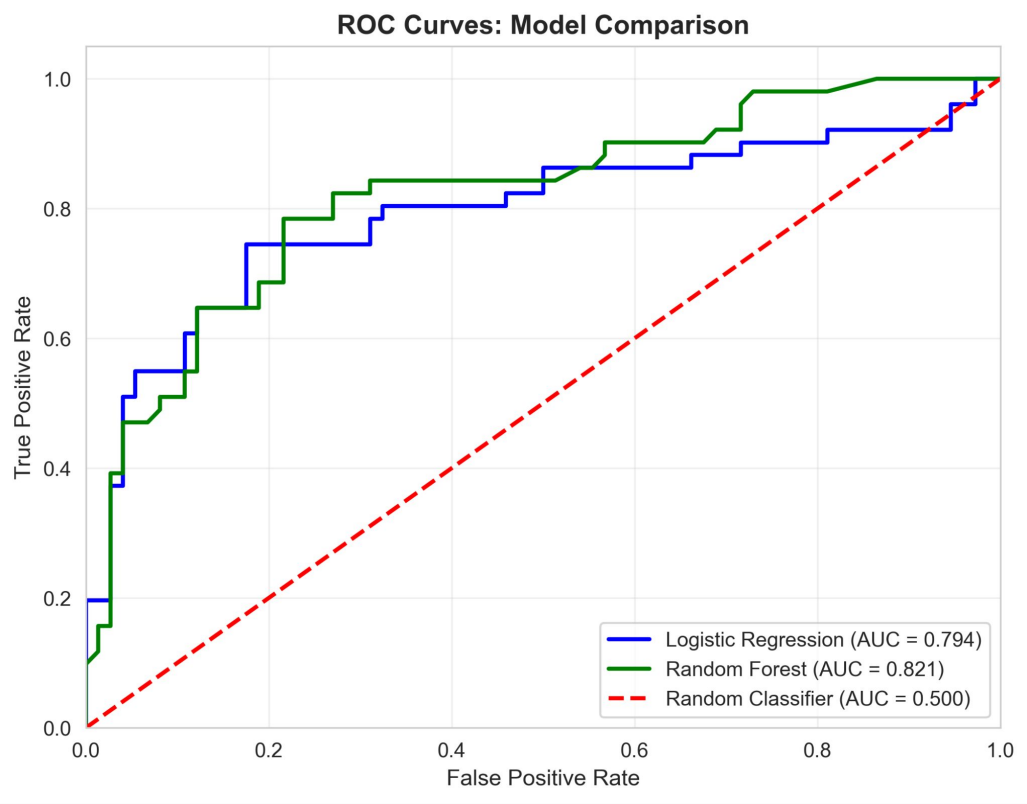
- Bar chart ranking feature importance from Random Forest
- Top Findings:
  - Caffeine (mg): 51.12% - Dominant predictor (as expected)
  - Age: 29.75% - Second most important
  - Beverage type: Tea (5.4%), Coffee (2.8%), Energy (1.3%)
  - Time of day & Gender: <2% combined importance

Model Performance Comparison: All Metrics

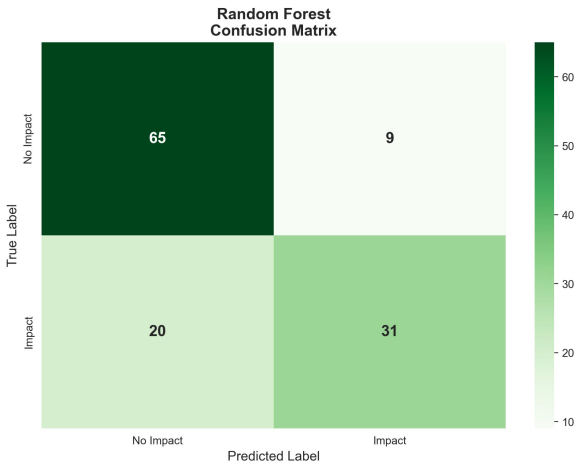


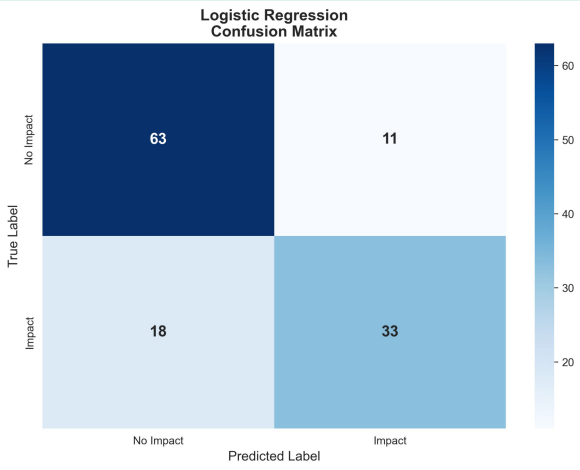
- Side-by-side confusion matrices for both models
- Interpretation:
  - Logistic Regression: 63 TN, 11 FP, 18 FN, 33 TP
  - Random Forest: 65 TN, 9 FP, 20 FN, 31 TP
  - RF has fewer false positives, better at identifying true negatives

ROC Curves



- ROC curves comparing both models
- Interpretation:
  - Logistic Regression AUC: 0.7939
  - Random Forest AUC: 0.8207
  - RF's higher AUC indicates superior ability to distinguish between sleep impact and no impact





Confusion Matrix

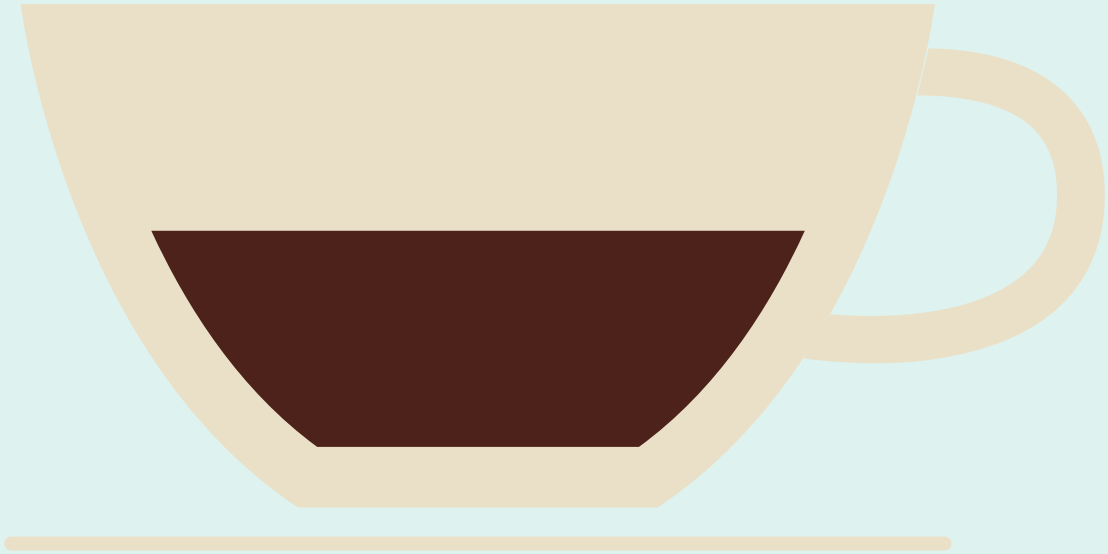
- Bar chart comparing 5 metrics (Accuracy, Precision, Recall, F1, AUC)

Key Metrics	Logistic Regression	Random Forest
Accuracy	76.80%	76.80%
Precision	75.00%	77.50%
Recall	64.71%	60.78%
F1-Score	69.47%	68.13%
AUC	79.39%	82.07%

Best Model?

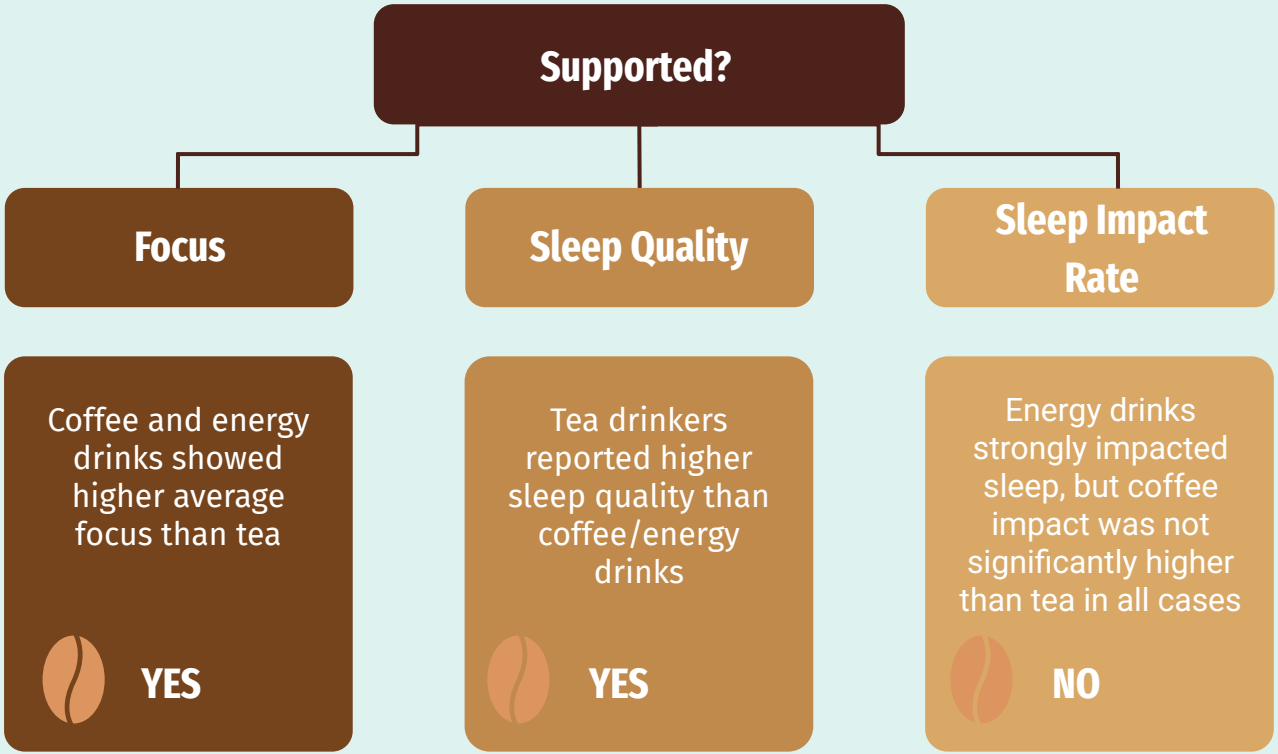
Random Forest

- Superior AUC (0.8207 vs 0.7939)
  - Better overall discrimination
- Higher precision (77.5% vs 75%)
  - Fewer false positives
- Captures non-linear caffeine-sleep relationships
- Feature importance reveals caffeine is the dominant factor (51%)
  - Confirms our hypothesis
- Most Important Predictive Factors
  - Caffeine mg
  - Evening consumption
  - Energy drink consumption
  - Coffee consumption



Discussion & Conclusion

Hypothesis Evaluation



Conclusions

01 Coffee and energy drinks do improve focus more than tea.

02 Tea is consistently associated with better sleep outcomes

03 Energy drinks carry the highest risk of sleep disruption

04 Caffeine amount and time of day significantly influence sleep impact

05 A Random Forest model can reasonably predict sleep disruption

Limitations

Self-reported metrics introduce subjective bias

Dataset size is limited (n = 500)

One-hot beverage data limits nuance (e.g., no dosage or brand type)

Future Work

Expand dataset with more daily logs

Explore clustering to identify caffeine-use behavior segments

Include physiological measures like heart rate or sleep-tracker data

Build a Streamlit dashboard

Daily caffeine intake logs over weeks

Thank You!

Q&A