# **Technical Report (Digital Diet - Predicting Wellness)**

**Problem statement**

How do the digital diet habits of participants affect their anxiety levels? What are the best practices of those with positive anxiety scores?

**Setting the scene and audience**

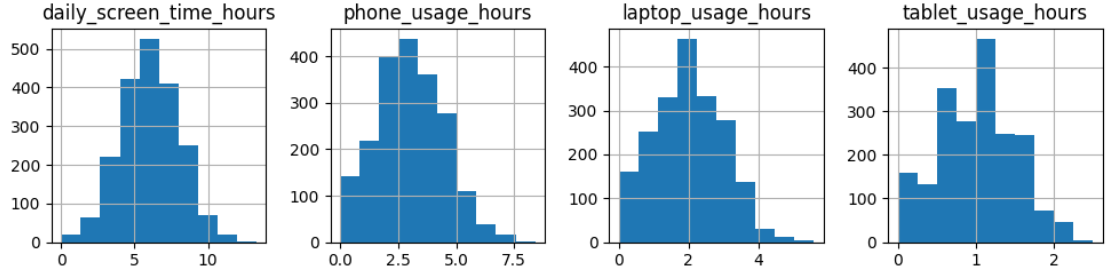
This dataset has been collected as part of an advertising campaign to increase awareness of the impact of digital habits and encourage positive change.

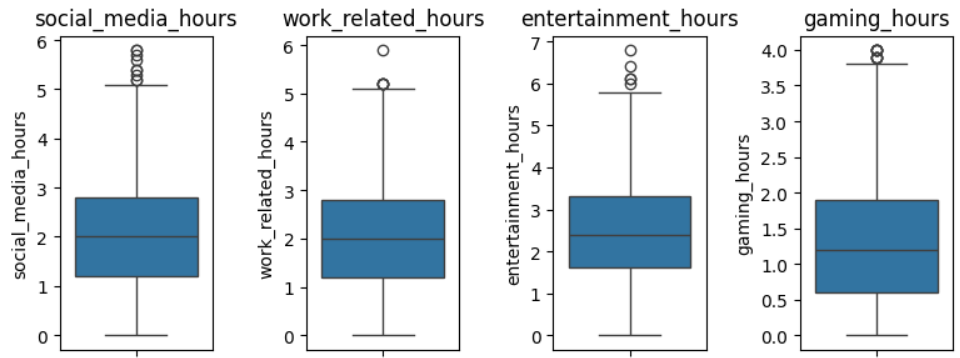
**Goal**

An analysis of the relationship between digital consumption habits and wellness.

1. **Introducing the data**

Data was relatively clean (nulls or duplicates). However, several features were skewed by outliers.





Decision made to exclude outliers after inspecting histograms, boxplots, quartiles, thresholds.

After cleaning there were 1915 rows/unique participants (of 2000).

1. **EDA & Visuals**

**High-level demographics:**

A close-up of a graph

AI-generated content may be incorrect.

**Anxiety score is the main metric.**

It consists of integers from 0-20. A high score is a good score.

**Demographic breakdown (by Average anxiety score):**

A graph of different colored bars

AI-generated content may be incorrect.

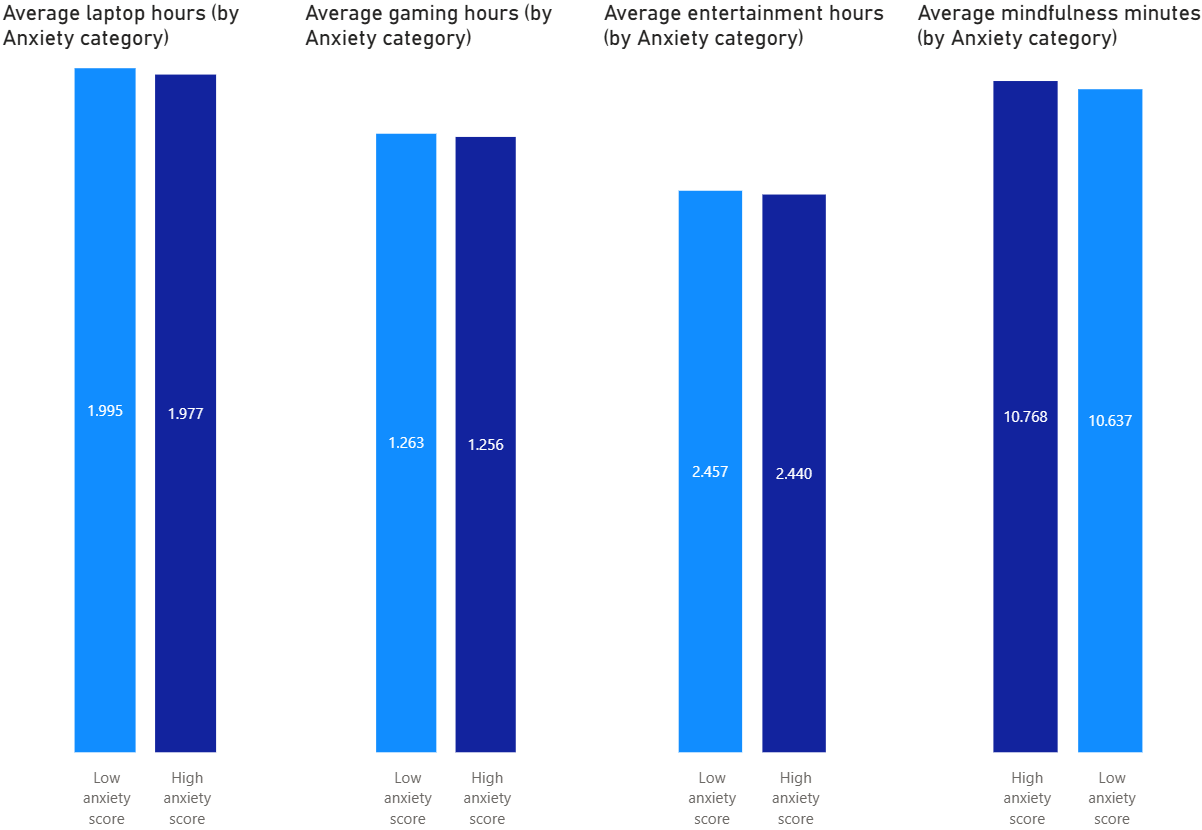
For gender and locale: Non-binary and rural-located participants scored higher.

Main takeaway: Wellness app users scored higher than non-users.

**Deeper look (high vs low scorers):**

High scorers: those between 11-20.

Comparing the habits of high vs low scorers. What sets high scorers apart?



Main takeaways:

* High scorers (in dark blue) spent less time on their laptops, gaming, and entertainment.
* They also clocked more mindfulness minutes.

1. **Predictive modelling**

**Target variable (y)** is if participant has a Good anxiety score (11-20), or not.

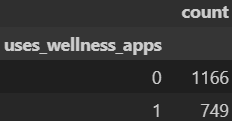
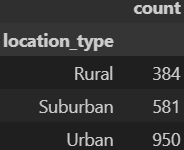
Goal is binary classification.

Success criteria: Build a model that can accurately predict a participant’s anxiety score.

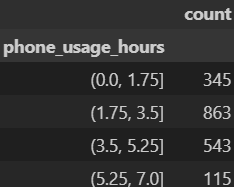
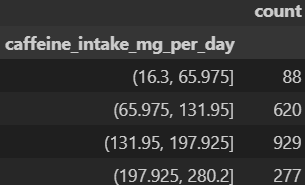
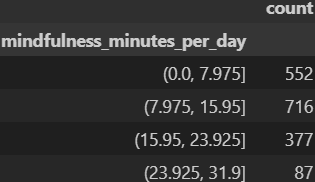
Target audience: Potential app users. App designers. Digital natives/immigrants interested in positive outcomes.

**Feature selection**

2 categorical: location type, uses wellness apps.

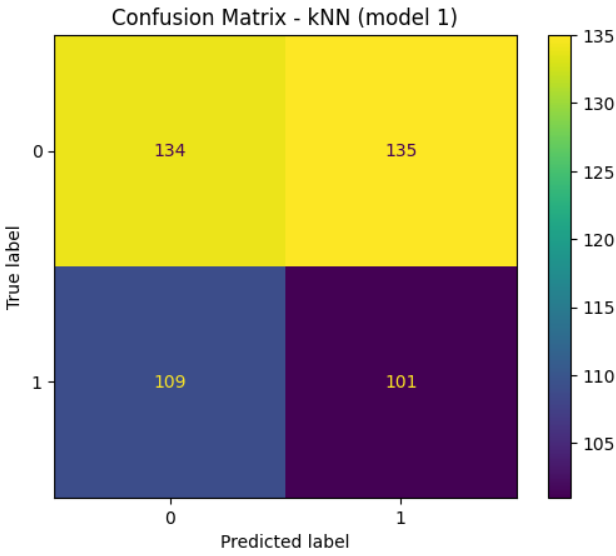
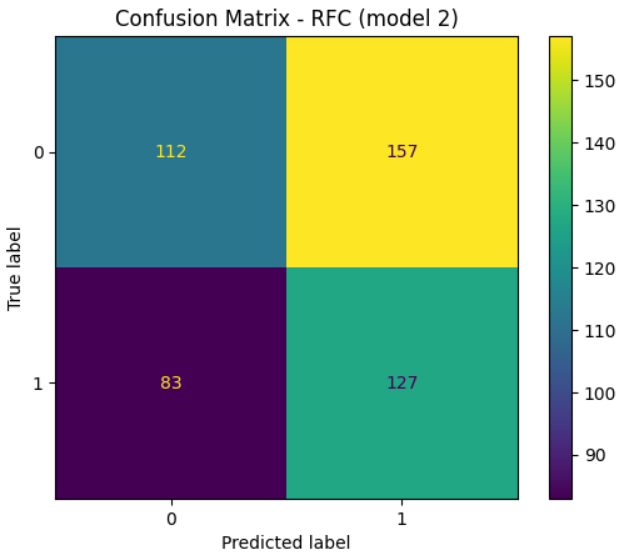


3 numerical: phone usage, caffeine intake, mindfulness minutes.

All 5 features are distinctively different. Categories and bin thresholds also have distinguished counts.

**Comparative analysis**

k-Nearest Neighbors (model 1) versus Random Forest (model 2) ****

Random Forest made more positive predictions than k-Nearest Neighbors.

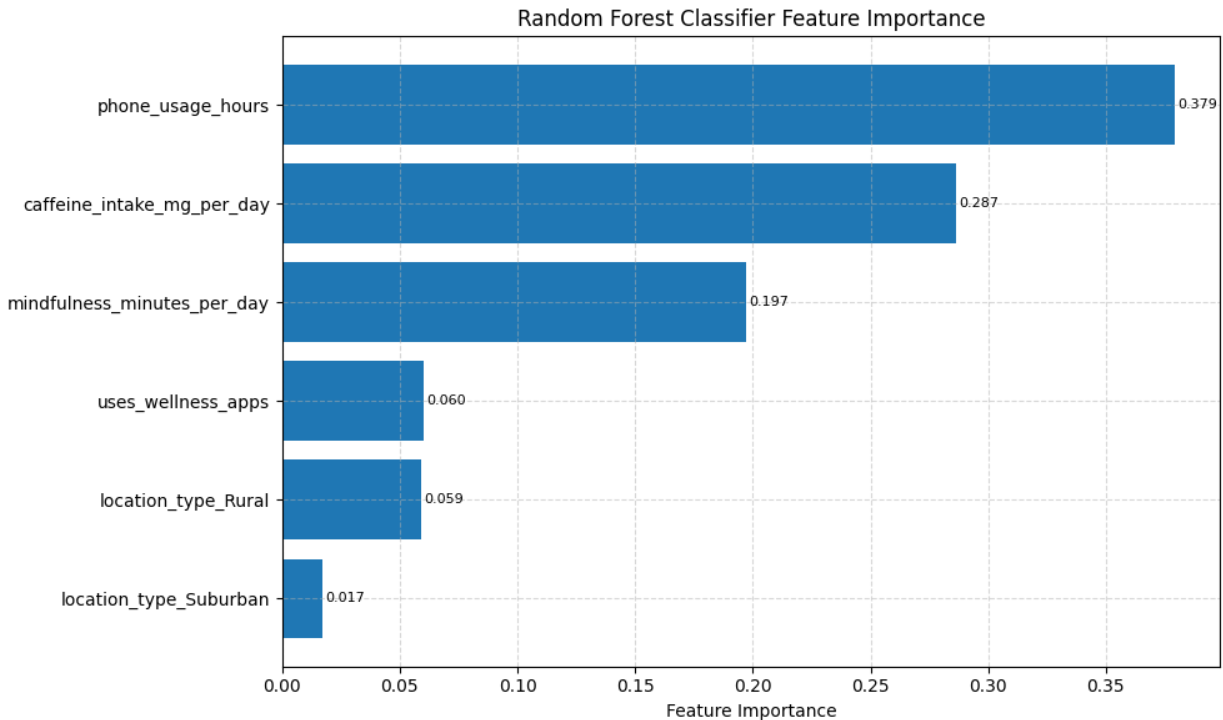
A black screen with white text

AI-generated content may be incorrect. Against a baseline accuracy of 56%, both models were slightly behind. However, relying only on accuracy can be misleading. Precision and recall score both focus on true positive prediction (correctly predicted positive instances). F1-score combines them into a single value, providing a balanced assessment of model performance.

Context-wise, precision and recall are both important. We can also rely on F1-score as a performance indicator.

**Comparatively, Random Forest outscored k-Nearest Neighbors.**

**Performance analysis (with Feature importance)**



Phone usage, caffeine intake, and mindfulness minutes contributed most to model performance and efficiency (more than 86% combined). Wellness app usage came in next.

1. **Conclusion & reflections**

Insights and recommendations (ordered for optimization):

* Take screen-time breaks
* Watch caffeine intake
* Clock mindfulness minutes
* And for synergy, use wellness apps to help track these new habits

The granularity of the data implies that small daily changes can have a big impact on wellness.

In reflection, while the dataset was relatively clean, identifying trends and insights wasn’t a linear process. This was the case when visualizing metrics and for feature selection. The decision to handle and exclude outliers did help the process.

This complexity of the data reflected the effect of small differences in digital diet habits. The granularity of the data sometimes came down to that of minutes and days, ultimately affecting participants’ overall wellness indicators.