Which lifestyle patterns have the strongest impact on health?

This project investigates how different aspects of lifestyle influence personal health (BMI score). Using a structured dataset, we applied various data science techniques to uncover patterns, visualize relationships, and use predictive algorithms. You can explore our full analysis and results directly on this page.

Project Timeline:

1. Data collection
Lifestyle & Wellbeing dataset from Kaggle (self-reported survey data)

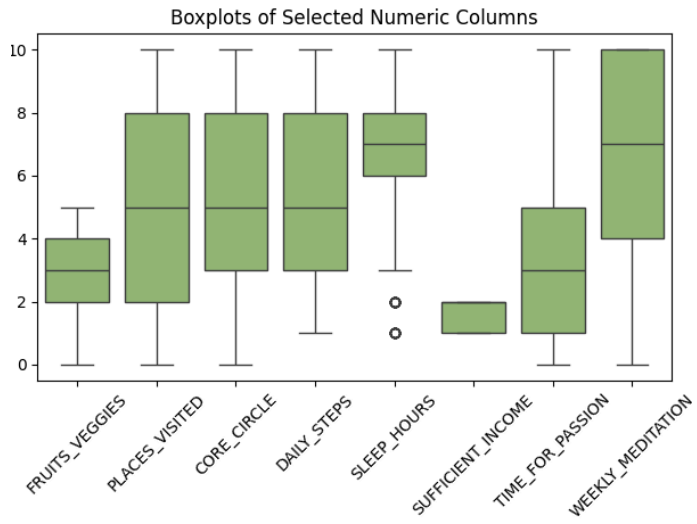
2. Data cleaning
Handle missing values, convert types, prepare categorical variables

3. Data Analysis
Exploratory analysis, visualizations, summary statistics

4. Feature Engineering
Normalize values, create “health” and “lifestyle” scores

5. Clustering
Identify behavioral patterns using various clustering methods

6. Classification
Predict BMI category (healthy vs overweight)



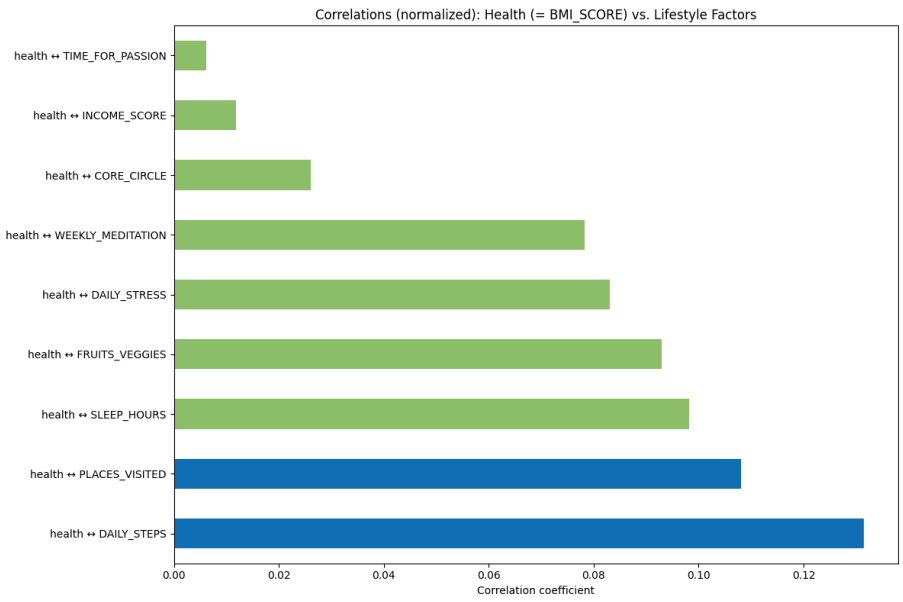
This visualization shows the distribution of selected lifestyle-related features from the Lifestyle and Wellbeing dataset (Kaggle, 2021), containing **15,977 self-reported survey responses across 24 variables.**

The boxplot highlights variation across lifestyle habits:

- Features like Sleep Hours are tightly clustered around a median value, suggesting similar routines.
- In contrast, Daily Steps, Weekly Meditation, and Time for Passion show wide distributions, reflecting large behavioral differences between individuals.
- Binary-like variables such as Sufficient Income have limited variance.

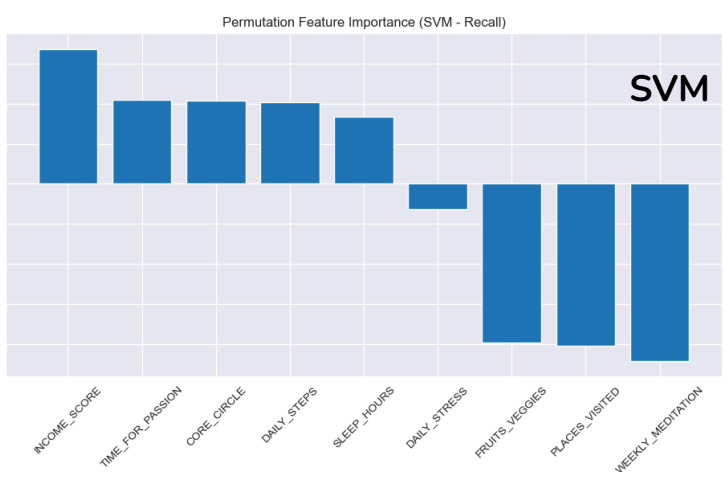
Correlations:

- **Positive Correlations:** Bars extending right indicate that as a lifestyle factor increases, so does the "Health" score (healthier BMI).
- **Strongest Link:** The DAILY_STEPS score shows the most significant positive correlation, highlighting its strong association with better health.
- **Key Positive Factors:** SLEEP_HOURS, PLACES_VISITED also show positive correlations with improved health outcomes.
- **Less Impactful Factor:** Some positive habits, (e.g. TIME_FOR_PASSION, CORE_CIRCLE) show a positive but weaker link to our "Health Score" (healthy BMI) in this data. They're beneficial, but their direct impact on BMI isn't as pronounced as factors like SLEEP_HOURS or DAILY_STEPS.



1

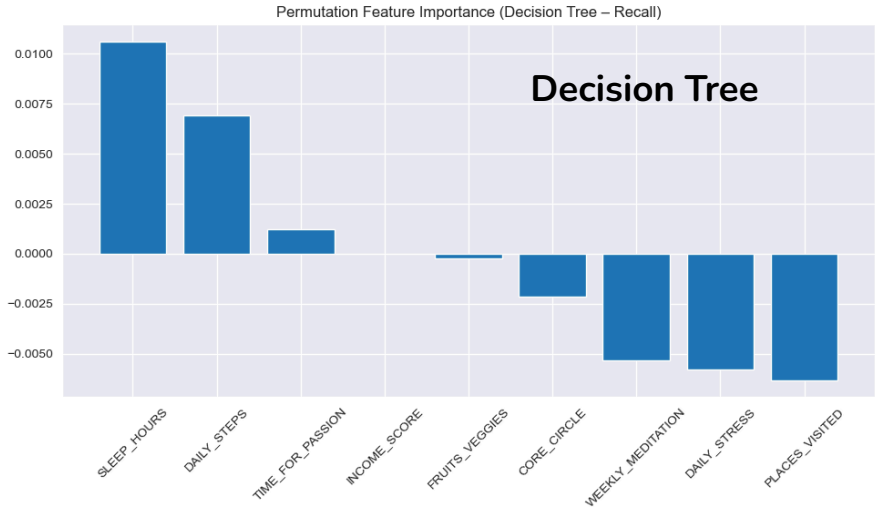
Accuracy: **0.607** Precision: **0.616**
Recall: **0.881** F1 : **0.725**



“Income score” shows the highest positive impact on recall, followed by similarly ranked features such as “Time for passion”, “Core circle”, and “Daily steps”. “Places visited” and “Weekly meditation” have a negative effect on recall.

2

Accuracy: **0.593** Precision: **0.615**
Recall: **0.827** F1: **0.705**



“Sleep hours” and “Daily steps” have the strongest positive impact on recall, while features like “Places visited” or “Daily stress” slightly reduce recall, possibly adding noise rather than value.

Conclusion:

In this project, we explored which lifestyle factors have the strongest impact on health by analyzing a dataset with various lifestyle subcategories. After preprocessing the data and performing feature selection, we applied different classification models, including Decision Trees and Support Vector Machines (SVM). The SVM achieved the best performance, identifying "Income score", "Time for passion", "Core circle", and "Daily steps" as the most impactful positive factors for health prediction. The Decision Tree highlighted "Sleep hours" and "Daily steps".These findings show that health is not determined by a single factor, but rather by a combination of physical activity, sleep quality, social relationships, and economic conditions.

Learn More:

HAW HAMBURG

Data Science Project SoSe25. Nicolas Wenzel, Alina Wolf, Tonmoy Majumder, MD. Sun