

Project Timeline:

Lifestyle & Wellbeing dataset from Kaggle (self-reported survey data)

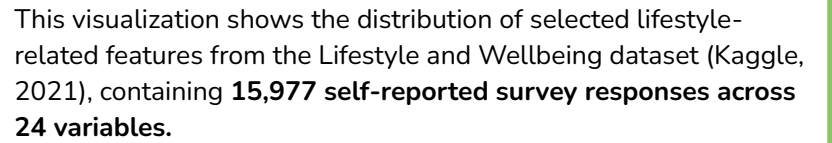
Handle missing values, convert types,
prepare categorical variables

Exploratory analysis, visualizations,
summary statistics

Normalize values, create “health” and “lifestyle” scores

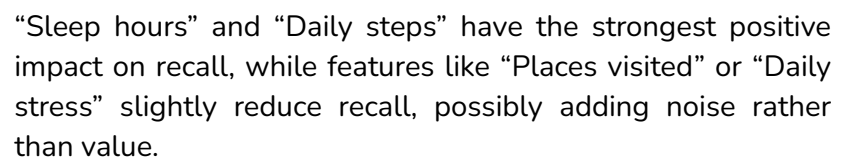
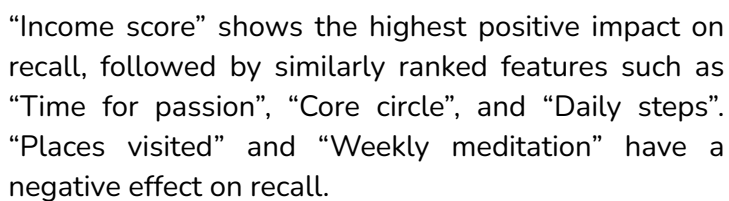
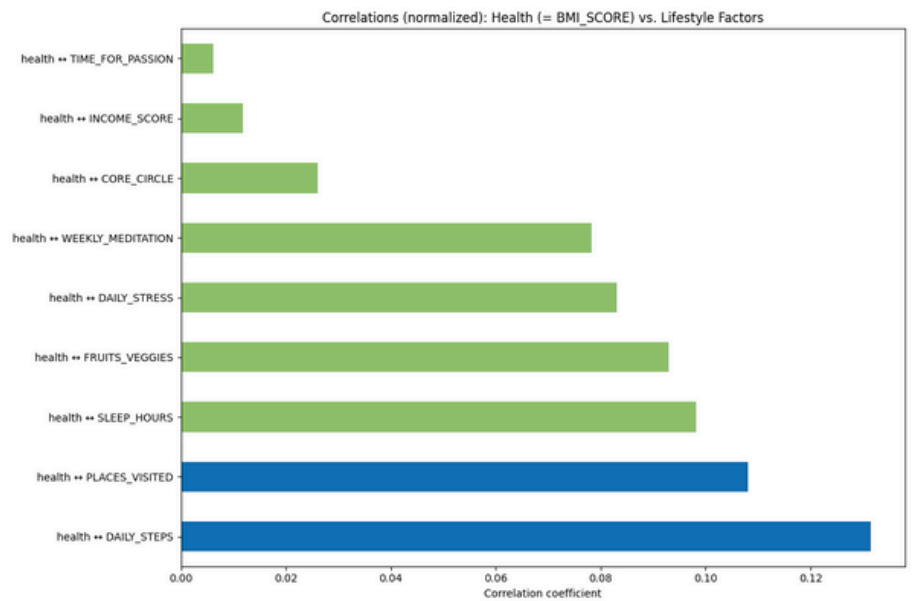
Identify behavioral patterns using various clustering methods

Predict BMI category (healthy vs overweight)



- 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.

- **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.



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.

