

# PANDAS ANALYSIS

## 1. Objective:

I Observed that the purpose of this analysis was to explore the user's fitness tracking data (steps, sleep, heart rate, calories, etc.) and derive meaningful insights using Python.

## 2. Tools & Libraries i Used:

Python 3.x

pandas (for data manipulation)

matplotlib and seaborn (for visualizations)

Jupyter Notebook in VS Code

## 3.Steps used in the pandas analysis

### A. Data Loading:

- All 18 CSV datasets were loaded using pandas.
- Each dataset was assigned to a dictionary called `dataframes` for organized access.

### B. Data Cleaning:

- Checked for missing values and duplicates across all datasets.
- Removed duplicates using `.drop\_duplicates()` and handled missing values using `.dropna()` where appropriate.
- Converted date/time fields (e.g., `ActivityDate`, `SleepDay`, `ActivityHour`) into datetime format.

### C. Data Visualization:

- Created all visualizations for the given datasets.
- Plots included:
  - Steps vs Calories Burned
  - Sleep Duration Distribution
  - Heart Rate Trends
  - Daily and Hourly Activity Summaries
  - Weight vs BMI Analysis and etc

- Used matplotlib and seaborn for line plots, scatter plots, histograms, bar charts, and boxplots.

#### D. Insights Extracted:

- Users with higher active minutes generally burned more calories.
- Most users had consistent sleeping patterns between 350-450 minutes per night.
- Heart rate patterns showed peaks in early mornings and evenings.
- Step and calorie activity varied hourly, suggesting consistent workout habits.

#### Conclusion:

The Python EDA provided a comprehensive understanding of user fitness behaviors. The combination of visual insights and cleaned data enables further analysis or dashboarding in tools like Power BI.

## Python Code Used for Data Cleaning

```
import pandas as pd
```

```
file_names = [
    'dailyActivity_merged.csv', 'dailyCalories_merged.csv', 'dailyIntensities_merged.csv',
    'dailySteps_merged.csv', 'heartrate_seconds_merged.csv', 'hourlyCalories_merged.csv',
    'hourlyIntensities_merged.csv', 'hourlySteps_merged.csv', 'minuteCaloriesNarrow_merged.csv',
    'minuteCaloriesWide_merged.csv', 'minuteIntensitiesNarrow_merged.csv', 'minuteIntensitiesWide_merged.csv',
    'minuteMETsNarrow_merged.csv', 'minuteSleep_merged.csv', 'minuteStepsNarrow_merged.csv',
    'minuteStepsWide_merged.csv', 'sleepDay_merged.csv', 'weightLogInfo_merged.csv'
]
```

```
dataframes = {}
for file in file_names:
    df = pd.read_csv(file)
    name = file.replace('.csv', '')
    dataframes[name] = df
    print(f'Loaded: {file} Shape: {df.shape}')
for name, df in dataframes.items():
    print(f'\n Dataset: {name}')
    print(f' Original Shape: {df.shape}')
    print(f' Null values: \n{df.isnull().sum()}')
    print(f' Duplicates: {df.duplicated().sum()}')
    df = df.drop_duplicates().dropna()
    dataframes[name] = df
    print(f' Cleaned Shape: {df.shape}')
```