**STRAVA**

**PYTHON – EDA REPORT**

**INTRODUCTION**

**ABOUT THE APP**

Strava is a fitness tracking and social networking app for athletes, primarily focused on recording and sharing outdoor activities like running and cycling. It uses GPS data to track various metrics such as distance, time, pace, and elevation. Strava also allows users to connect with friends, join clubs, participate in challenges, and discover new routes and segments.

**OBJECTIVE**

To explore how users utilize smart devices like the Strava and provide data-driven recommendations to improve marketing strategy and customer engagement.

**DATA COLLECTION**

The dataset was collected via Fitbit fitness trackers and includes data on physical activity, sleep patterns, calories burned, steps taken, and heart rate.

**Files Used**:

* dailyActivity\_merged.csv
* dailyCalories\_merged.csv
* hourlyCalories\_merged.csv
* sleepDay\_merged.csv
* hourlySteps\_merged.csv
* heartrate\_seconds\_merged.csv
* minuteMETsNarrow\_merged.csv
* weightLogInfo\_merged.csv

**IMPORTED LIBRARIES**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import plotly.express as px

**DATA CLEANING AND PREPROCESSING**

 Converted date columns to datetime format.

 Removed duplicates and null values.

 Ensured consistent user IDs across different datasets.

 Created additional features such as:

* Active Minutes per Day
* Steps Category (Low, Moderate, High)
* Calories per Minute

heartrate\_seconds\_merged.csv

Id Time Value

0 2022484408 4/12/2016 7:21:00 AM 97

1 2022484408 4/12/2016 7:21:05 AM 102

2 2022484408 4/12/2016 7:21:10 AM 105

3 2022484408 4/12/2016 7:21:20 AM 103

4 2022484408 4/12/2016 7:21:25 AM 101

heartrate\_df.isnull().sum()

Id 0

Time 0

Value 0

dtype: int64

weight\_df

Id Date WeightKg WeightPounds Fat \

0 1503960366 5/2/2016 11:59:59 PM 52.599998 115.963147 22.0

1 1503960366 5/3/2016 11:59:59 PM 52.599998 115.963147 NaN

2 1927972279 4/13/2016 1:08:52 AM 133.500000 294.317120 NaN

3 2873212765 4/21/2016 11:59:59 PM 56.700001 125.002104 NaN

4 2873212765 5/12/2016 11:59:59 PM 57.299999 126.324875 NaN

BMI IsManualReport LogId

0 22.650000 True 1462233599000

1 22.650000 True 1462319999000

2 47.540001 False 1460509732000

3 21.450001 True 1461283199000

4 21.690001 True 1463097599000

weight\_df.isnull().sum()

Id 0

Date 0

WeightKg 0

WeightPounds 0

Fat 65

BMI 0

IsManualReport 0

LogId 0

dtype: int64

weight\_df = weight\_df.drop('Fat', axis = 1)

**Exploratory Data Analysis (EDA)**



Observations & Insights

**1. Time Span**

* Data covers **April 16 to May 9, 2016**.
* This data is captured for the user with the id 6117666160
* Heart rate is captured **at a high frequency**

**2. Heart Rate Range**

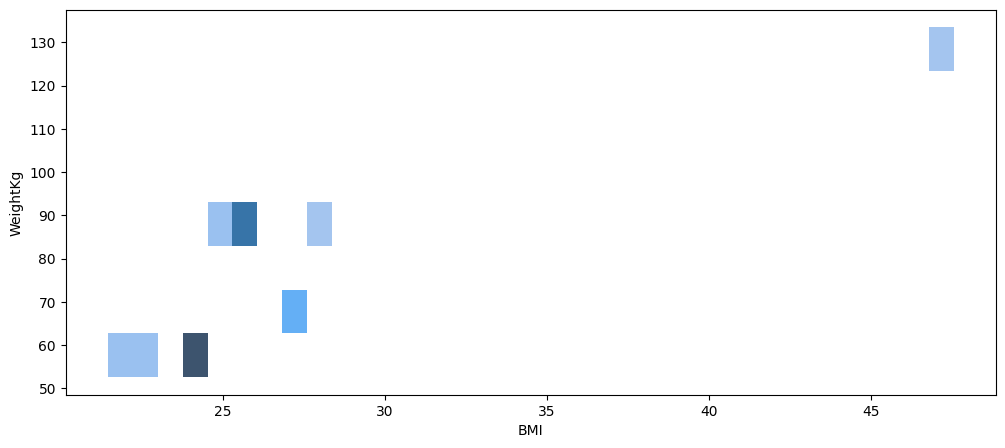
* Most heart rate values lie between **60 bpm and 130 bpm**.
* **Spikes up to ~190 bpm** occur occasionally → could indicate:
  + High-intensity workouts (e.g., running or HIIT),
  + Sudden stress episodes,

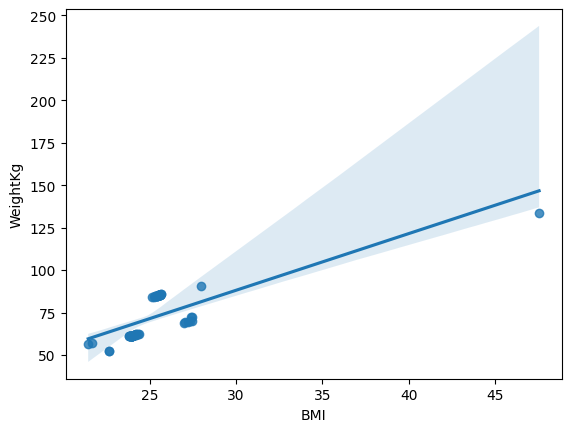
**3. Data Gaps / Flat Lines**

* Noticeable **gaps or plateaus** in the data:
  + Around **April 24**, **May 2**, and **May 4–5**.
* This suggests periods of:
  + No tracking (device off/unworn),
  + Sensor failure,
  + Sleep or inactivity periods with stable HR.

**4. Variability Patterns**

* **High frequency fluctuations** show an **active cardiovascular profile**.
* The user seems:
  + Physically active (frequent spikes in HR),
  + Possibly working out regularly,
  + Having quick recovery periods (drops after spikes).





**Strong Positive Correlation**

* As expected, **higher weight leads to higher BMI**.
* The correlation appears mostly **linear**, consistent with the BMI formula:

BMI=Weight (kg)Height (m)2\text{BMI} = \frac{\text{Weight (kg)}}{\text{Height (m)}^2}BMI=Height (m)2Weight (kg)​

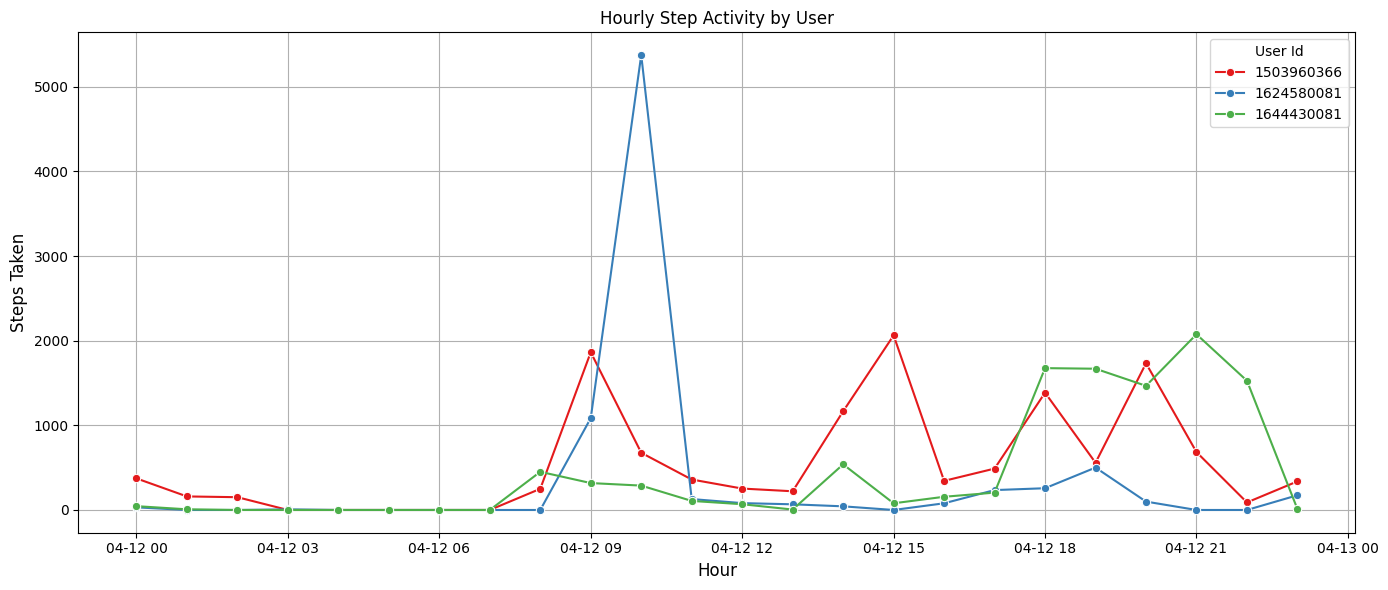
Since height remains constant per user, BMI grows with weight.

**Outlier Present**

* One user has a **BMI of ~47.5** and **weight ~133.5 kg**.
* This is well outside the normal range, possibly indicating:
  + Obesity
  + Incorrect data input

**Most Data Points Cluster Around Normal BMI**

* A majority of users have BMI values between **22 and 28**, which corresponds to:
  + **Normal weight** (18.5–24.9)
  + **Overweight** (25–29.9)



This line chart shows the **steps taken per hour on April 12th** by three users:

* **1503960366**
* **1624580081**
* **1644430081**

**Low Early Morning Activity**

* All users show **almost no steps before 8 AM**, indicating:
  + Night sleep hours
  + Consistent rest behavior

**3. User-Specific Patterns**

**1503960366 (Red)**

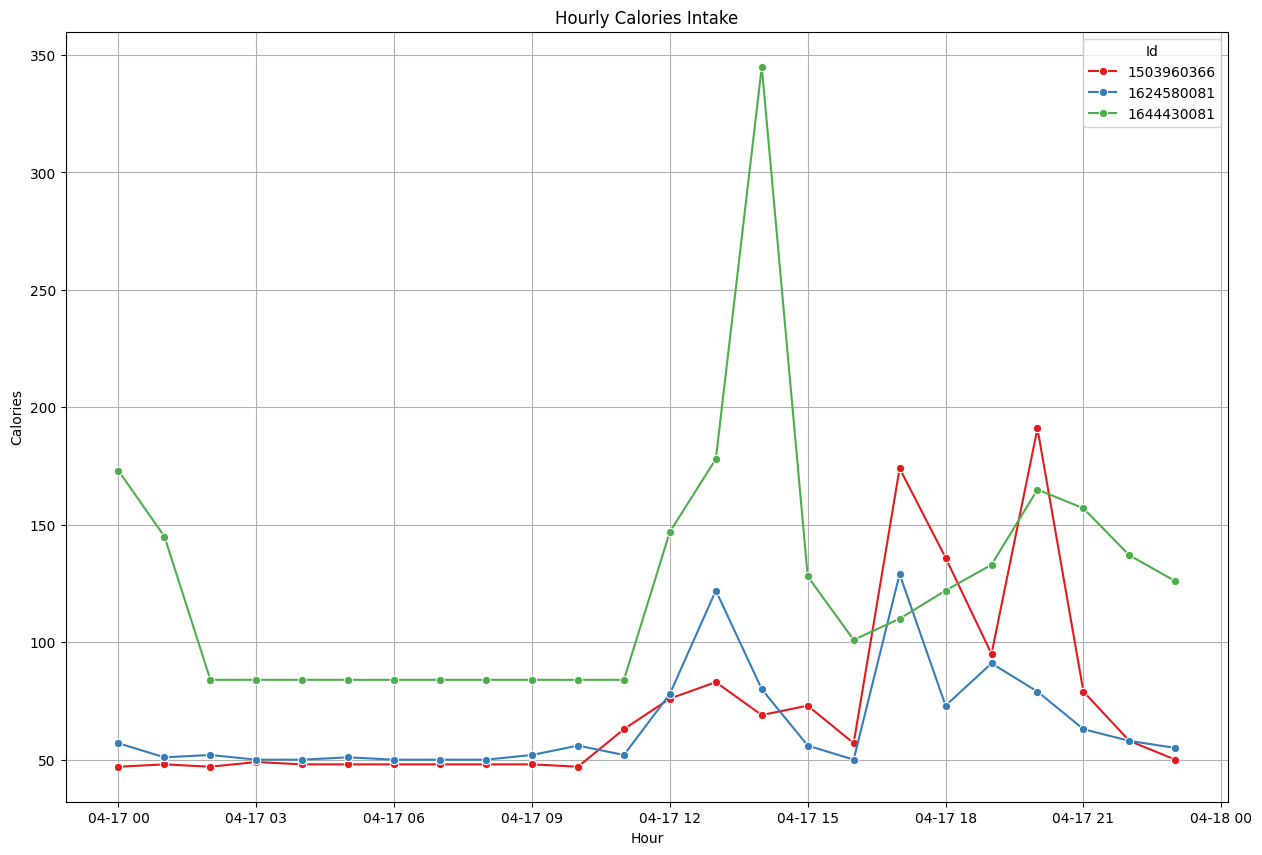
* Moderate activity throughout the day.
* Multiple bursts: ~9 AM, ~3 PM, ~7–8 PM.
* Likely active in **short intervals**, possibly spread-out walking or chores.

**1624580081 (Blue)**

* Almost no activity **except one huge spike** at **10 AM**.
* Possibly did a **single workout or walk**, then remained inactive.

**1644430081 (Green)**

* Very little daytime activity.
* Becomes very active from **6 PM to 10 PM**.
* Might suggest a **night owl routine** or **evening workout habit**.



This chart shows **calories burned per hour** (not intake as the title says; Fitbit data typically logs *calories burned*) for 3 users:

* 🔴 **1503960366**
* 🔵 **1624580081**
* 🟢 **1644430081**

**1. Baseline Calories (Resting Metabolism)**

* All users burn **~45–80 calories/hour** steadily throughout the day (resting + light activity).
* Indicates Fitbit includes **basal metabolic rate** even in low activity periods.

**3. Activity Distribution**

**🔴 1503960366 (Red)**

* Low calories until **late afternoon**, then active post 4 PM.
* Consistent with **evening exercise routine** or job activity after work hours.

**🔵 1624580081 (Blue)**

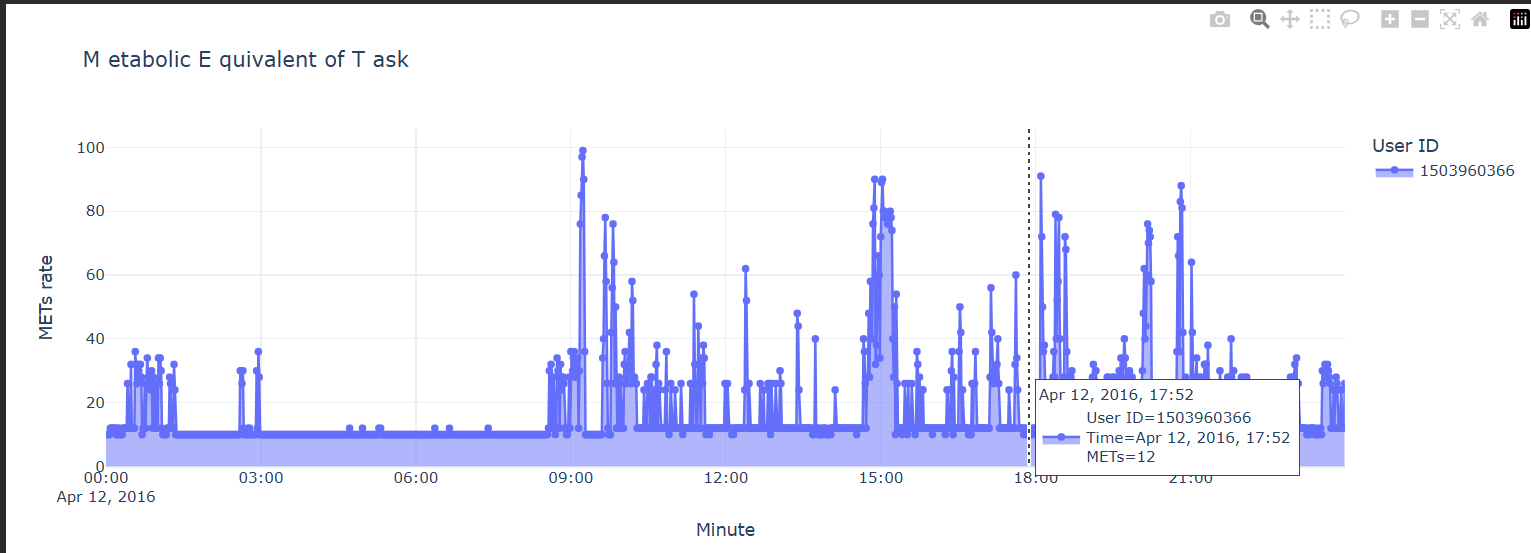
* Relatively **steady** but with minor afternoon peaks.
* Moderate lifestyle with **occasional active bursts**.

**🟢 1644430081 (Green)**

* High burn **very early** (midnight ~ 170 calories), again around **2 PM**.
* Likely very active or possibly device misread/sync anomaly during these hours.

**4. Inactive Early Mornings**

* Between **12 AM to 6 AM**, all users show low calorie burn, suggesting:
  + Sleep or very minimal movement
  + Consistent circadian rest cycles



This interactive Plotly line chart visualizes the **METs (Metabolic Equivalent of Task) rate** on a **minute-by-minute basis** over **April 12, 2016** for User ID 1503960366.

* **METs** measure the **intensity of physical activity**.
* 1 MET = resting state
* 3–6 METs = moderate activity
* 6+ METs = vigorous activity

**1. Frequent Fluctuations in METs**

* The user exhibits **consistent variation in activity** throughout the day.
* Spikes often reach **20–40+ METs**, indicating **brief bursts of high-intensity movements**.

| **Time** | **Activity Intensity** |
| --- | --- |
| ~09:00 | Peaks up to 100 METs 🔥 |
| ~15:00 | Sustained bursts 40–80 |
| ~20:00–21:00 | Repeated high activity |

**3. Low or No Activity Periods**

* Flat lines at or near **10 METs** suggest:
  + Sedentary behavior (e.g., sitting, resting)
  + Baseline activity
* Gaps or flat stretches (e.g., 03:00–06:00) may indicate **sleep or minimal movement**.

**EDA Report Aligned with Power BI Dashboard**

**🔗 Dashboard Integration Highlights**

- Sleep analysis in the dashboard is directly supported by the processed `sleepDay\_merged.csv` data.

- METs and heart rate data from `minuteMETsNarrow\_merged.csv` and `heartrate\_seconds\_merged.csv` power time-series visualizations in Power BI.

- Weight and BMI metrics analyzed here match with BMI category visuals in the dashboard.

- Steps and calorie insights align with hourly activity plots for user behavior over time.

