Fitness Analysis Final Report :

**Summary :**

The aim of this project is to analyze health and fitness tracking data to uncover insights into user activity patterns, health behaviors, and potential areas for lifestyle improvements. This includes understanding trends in exercise frequency, calorie expenditure, sleep patterns, and other fitness metrics.

**Data source :**

' [https : //www.kaggle.com/datasets/jijagallery/fitlife-health-and-fitness-tracking-dataset](https://www.kaggle.com/datasets/jijagallery/fitlife-health-and-fitness-tracking-dataset) '

**Description :**

**Demographic Information :**

participant\_id : Unique identifier for each participant  
age : Age of participant (18-65 years)  
gender : Gender (M/F/Other)  
height\_cm : Height in centimeters  
weight\_kg : Weight in kilograms  
bmi : Body Mass Index calculated from height and weight

**Activity Metrics :**

activity\_type : Type of exercise (Running, Swimming, Cycling, etc.)  
duration\_minutes : Length of activity session  
intensity : Exercise intensity (Low/Medium/High)  
calories\_burned : Estimated calories burned during activity  
daily\_steps : Daily step count

**Health Indicators :**

avg\_heart\_rate : Average heart rate during activity  
resting\_heart\_rate : Resting heart rate  
blood\_pressure\_systolic : Systolic blood pressure  
blood\_pressure\_diastolic : Diastolic blood pressure  
health\_condition : Presence of health conditions  
smoking\_status : Smoking history (Never/Former/Current)

**Lifestyle Metrics :**

hours\_sleep : Hours of sleep per night  
stress\_level : Daily stress level (1-10)  
hydration\_level : Daily water intake in liters  
fitness\_level : Calculated fitness score based on cumulative activity

**Technologies used :**

1) **Python :**  Data cleaning, transformation, and exploratory data analysis (EDA)

2) **MySQL :** Structured data storage and querying

3) **Excel :**  Additional data cleaning and verification

4 ) **Tableau / Power BI :**  Data visualization and dashboard creation for presenting insights

**Problem Statements :**

**1) Which activity type is the most useful for the following goals :**

a) Burning Maximum amount of Calories.

b) Relationship between activity type , intensity and an optimal / healthy BMI.

2) General Health Analysis :

a) How many users meet the recommended 10,000 steps/day?

b) What percentage of users get less than 6 hours of sleep daily?

c) What percentage of users are at optimal hydration level?

3) Health Condition Analysis :

a) Which gender is most prone to which health condition?

b) Relationship between fitness level and health conditions?

**Data Cleaning :**

Using Python pandas library :

The raw data was sourced from Kaggle’s *FitLife Health & Fitness Tracking Dataset*, such as daily activity.

The cleaning was done using **Python (Pandas)**.

Used the following libraries:

* **pandas for data manipulation**
* **numpy for handling numerical data**
* **datetime for date conversions**
* **warnings to suppress unnecessary output**

Initial Rows **: 687,701 (sample size)**

Null Values Removed **: 490,275 rows (71.3% of data)**

**Cleaned Dataset Rows : 197,426 (28.7% retained)**

**Output File : cleaned\_health\_fitness\_data.xlsx**

**Functions Used :**

| **Function** | **Purpose** |
| --- | --- |
| **import pandas as pd** | Imports the **Pandas** library for data manipulation. |
| **import numpy as np** | Imports **NumPy** for handling numerical operations. |
| **import openpyxl** | Enables reading and writing of .xlsx Excel files. |
| **from datetime import datetime** | Imports the **datetime** class to manage and manipulate date/time values. |
| **pd.read\_excel()** | Reads an Excel file into a Pandas DataFrame. |
| **df.isnull()** | Returns a DataFrame showing True where values are missing. |
| **df.isnull().any(axis=1)** | Checks each row for any null (missing) values. |
| **df[df.isnull().any(axis=1)]** | Filters and displays rows that contain at least one null value. |
| **df.dropna()** | Removes rows that contain any null values. |
| **df.columns.str.strip()** | Removes leading/trailing spaces from column names. |
| **df.to\_excel()** | Saves the DataFrame back to an Excel file. |
| **print()** | Displays output, values, or messages in the console. |

## EDA Report:

## Key Health Performance Indicators (KPIs):

1. **Sleep Deprived Users:**
   * 12.19% of users are sleep deprived.
   * Indicates a moderate level of sleep deprivation which could impact overall health and wellness**.**
2. **Users Meeting Step Goal:**
   * 25.35% of users meet their daily step goal.
   * Suggests that physical activity among the majority of users needs improvement.
3. **Users with Normal Blood Pressure (BP):**
   * 22.75% of users have normal BP.
   * Highlights a concern, as over 77% may be at risk of hypertension or other cardiovascular conditions.
4. **Users with Optimal Body Mass Index (BMI):**
   * 64.08% maintain an optimal BMI.
   * Indicates that the majority of users fall within a healthy weight range, which is a positive sign.
5. **Users with Optimal Hydration:**
   * 77.49% are well-hydrated.
   * Demonstrates good hydration habits among users, which supports metabolic and overall body function.

### Top 5 Activities for Maximum Calorie Burn:

* **HIIT tops the list with the highest average calories burned (31.02 Cal from high-intensity sessions).**
* **Followed by:**
  + **Running: 25.36 Cal (High Intensity)**
  + **Cycling: 21.88 Cal (High Intensity)**
  + **Basketball: 20.49 Cal (High Intensity)**
  + **Dancing: Lower total compared to others.**
* **High-intensity activities consistently burn more calories, reinforcing the benefit of high-intensity workouts.**

### Health Condition by Gender:

* **Hypertension is the most prevalent condition, particularly among Females, followed by Males.**
* **Diabetes shows a higher concentration among Males.**
* **Asthma appears comparatively lower but is present across all genders.**
* **Suggests gender-specific health intervention strategies may be needed.**

### Activity Intensity vs. BMI Category:

* **High Intensity activities are associated mostly with Healthy BMI levels.**
* **Medium Intensity shows a mix of Healthy and Overweight participants.**
* **Low Intensity activities include more participants in the Overweight and Obese categories.**
* **Indicates a strong correlation between higher intensity exercise and healthier BMI outcomes**.

### Fitness Level vs Health Condition:

* **Advanced fitness level is linked with fewer cases of chronic conditions (Hypertension, Diabetes, Asthma).**
* **Intermediate and Beginner levels show a higher number of participants with Hypertension and Diabetes.**
* **Points to a clear relationship between fitness levels and better health outcomes.**

## Insights & Recommendations:

1. **Increase Physical Activity:**

**With only 25.35% meeting step goals, promoting daily activity through challenges or reminders may help.**

1. **Encourage High-Intensity Workouts:**
   * **HIIT and similar high-intensity exercises significantly boost calorie burn and promote healthier BMI.**
2. **Targeted Health Programs:**
   * **Design gender-sensitive programs especially focusing on Hypertension in females and Diabetes in males.**
3. **Promote Better Sleep:**
   * **With 12.19% sleep-deprived users, interventions to improve sleep hygiene can enhance overall health.**
4. **Leverage Hydration Strength:**
   * **High hydration rates are a strength—these habits should be maintained and promoted further.**
5. **Customized Fitness Plans:**
   * **Fitness plans tailored by user fitness level could reduce chronic health conditions.**