STRAVA Fitness App Statistics Project Report

Project Name: "STRAVA Fitness Data Analytics Case Study"

Project Type:

SQL Analysis + Power BI Dashboard + Pandas EDA + Data Cleaning

Contribution: Individual

Project Summary:

The STRAVA Fitness App Analytics project entailed analyzing personal health and activity data tracked by fitness trackers. The aim was to pull information on user behavior — daily steps, sleep, calories consumed, and health factors such as BMI — using SQL, Python (Pandas), and Power BI. The ultimate aim was to take raw data and turn it into a clean, interactive, and informative dashboard that not only uncovers health trends but also assists in making intelligent decisions.

The analysis was done on three basic datasets:

1. Daily Activity: steps, walking distance, active minutes, and calories burned

2. Sleep Logs: total sleep minutes, bed time, and sleep efficiency

3. Weight Logs: user body weight and BMI for different dates

These dataframes were merged on `Id` and `Date` to create a master aggregate dataset. Data cleaning and transformation tasks like handling missing values, date format conversion, and feature engineering like Sleep Efficiency and SleepEfficiencyGroup (Excellent, Good, Average, Poor) were conducted using Pandas.

Insights Uncovered:

Through Exploratory Data Analysis (EDA), important patterns were revealed: - Weekends and holidays experienced step counts typically increase - Calorie burned were strongly correlated with total steps and Very Active Minutes - Sedentary individuals at high intensities would tend to have low sleep efficiency - Sleep habits were also quite varied among users, even frequent activity users - User BMI was from normal to obese with a high correlation with steps taken and activity level

Tool-by-Tool Analysis:

- ## SQL:
  - Imported master dataset into MySQL Workbench
  - Executed queries to get average steps per user, top active users, and identify anomalies
  - Combined BMI values to track user health ranges and outliers

- ## Power BI:
  - Created a slicer-enabled, business-class dashboard with:
  - Area, Bar, Donut, KPI graphics, Line, and Scatter
  - Included filtering by Date, UserID, and SleepEfficiencyGroup - Created graphics to track activity, calorie, and BMI patterns over time

- ## Pandas
  - Joined different datasets with common keys
  - Created calculated fields: SleepEfficiency and SleepEfficiencyGroup - Graphs generated: sleep efficiency histogram, BMI histogram, steps calories scatterplot, and a correlation heatmap

Project Statement:

This project is focused on analyzing user wellness trends through wearable fitness data. It converts individual daily logs into meaningful reports with data cleansing, analysis, and narrative techniques. With the inclusion of SQL, Python, and Power BI,

it is a complete solution for users and fitness professionals to monitor and enhance health metrics.

Business Goal:

To develop an analytical system able to:

- Monitoring short-term and long-term trends in fitness
- Identification of top and bottom performers
- Connecting sleep quality and physical activity
- Emphasizing health markers such as BMI and caloric burn -
Providing a data-driven dashboard for regular health monitoring

Dataset Overview:

Three critical datasets:

1. Daily Activity – Steps, Distance, Active Minutes, Calories

2. Sleep Logs – Minutes Asleep, Time in Bed, Sleep Efficiency

3. Weight Logs – Weight (kg), BMI, Date

Merged on `Id` and `Date`. New features obtained are:

- SleepEfficiency = TotalMinutesAsleep / TotalTimeInBed

- SleepEfficiencyGroup = Classified by category

- Time fields: Day, Week, Month

Variable Description:

- Id: Unique User Identifier

- ActivityDate: Date of Activity

- TotalSteps, Calories, VeryActiveMinutes: Activity Metrics Core -

TotalMinutesAsleep, TotalTimeInBed: Measures of sleep

- SleepEfficiency: Quantitative measure of sleep quality - WeightKg, BMI: Body Composition Measures

- SleepEfficiencyGroup: Categorical value (Excellent, Good, Average, Poor)

Conclusion:

The STRAVA Fitness Analytics project was able to demonstrate the worth of fitness data when combined with formal analysis and visualization.

Using SQL for query, Pandas for data manipulation and EDA, and Power BI for dashboards helped the project deliver actionable health data. BMI was also analyzed alongside other activity measures in an effort to assess user health profiles. Outliers and trends were graphed so individual-level observations of sleeping quality, activity level, and caloric performance might be made.

The resulting dashboard is a health monitoring GUI that may be helpful to personal use, personal trainers, or product development teams for fitness products.

This approach is a platform for future enhancement, including predictive modeling (e.g., weight change or sleep quality) and the inclusion of additional biometric data.