STRAVA FITNESS APP – PYTHON (PANDAS) ANALYSIS

Overview:

Python (with Pandas, Seaborn, Matplotlib) were employed for initial data wrangling, joining, feature engineering, and developing initial EDA plots.

- ## Tools Used:
  - pandas
  - numpy
  - seaborn
  - matplotlib
  - Jupyter Notebook

Steps Taken:

- Loaded 3 datasets:
  - dailyActivity_merged.csv
  - sleepDay_merged.csv
  - weightLogInfo_merged.csv

- Converted date fields
  - pd.to_datetime() for 'ActivityDate', 'SleepDay', and 'Date'

- Merged datasets on ['Id', 'ActivityDate']
  - Combined data for activity, sleep, and weight

- New derived features:
  - SleepEfficiency = TotalMinutesAsleep / TotalTimeInBed
  - SleepEfficiencyGroup = categorized version (Excellent/Good/Average/Poor).

- Handled missing values and dropped non-essential columns

Visualizations Developed:

- Bar plot: Average Steps per User
- Line plot: Calories Burned Over Time

- Histogram: Sleep Efficiency
- Heatmap: Correlation between activity and sleep measures
- Scatter plot: Steps and Calories
- Stacked Bar: Activity Minutes Breakdown
- Histogram: BMI Distribution

Insights:

- Sedentary minutes worsened sleep quality
- The majority of users had a normal activity pattern with mixed sleep efficiency
- Correlation between steps and caloric burn is established
- Sleep Efficiency Grouping was useful in offering healthy categories to users

Why Pandas Was Important:

- Supported preprocessing before modeling or dashboarding
- Applied feature engineering with ease
- Visualized relationships prior to delving into SQL or Power BI