Strava Fitness Data Analysis Case Study Python Analysis and Insights

Data Preprocessing in Python

Following SQL process, the cleaned datasets were saved and loaded as data frames in python. In addition to this, the minutes and seconds data are also loaded. The datatypes were verified, and the date and datetime columns were set to the right format.

After that, the minute data which are spread across several data frames were combined (as shown below) to make one dataset for easy analysis in the future stages.

```python
#merging minutes data to one df for analysis
df_minutesactivitynarrow = df_minutecaloriesnarrow
df_minutesactivitynarrow = pd.merge(df_minutesactivitynarrow, df_minuteintensitiesnarrow, on=['Id', 'ActivityMinute'], how='outer')
df_minutesactivitynarrow = pd.merge(df_minutesactivitynarrow, df_minutestepsnarrow, on=['Id', 'ActivityMinute'], how='outer')
df_minutesactivitynarrow = pd.merge(df_minutesactivitynarrow, df_minuteMETsnarrow, on=['Id', 'ActivityMinute'], how='outer')
df_minuteactivitynarrow = df_minutesactivitynarrow
df_minuteactivitynarrow.head()
```

✓ 3.8s Python

	Id	ActivityMinute	Calories	Intensity	Steps	METs
0	1503960366	4/12/2016 10:00:00 AM	0.9438	0	0	12
1	1503960366	4/12/2016 10:00:00 PM	1.8876	1	6	24
2	1503960366	4/12/2016 10:01:00 AM	0.9438	0	0	12
3	1503960366	4/12/2016 10:01:00 PM	1.8876	1	2	24
4	1503960366	4/12/2016 10:02:00 AM	2.5168	1	31	32

Data Analysis and Visualization in Python

After preprocessing, the data was carefully analysed to derive valuable insights and patterns

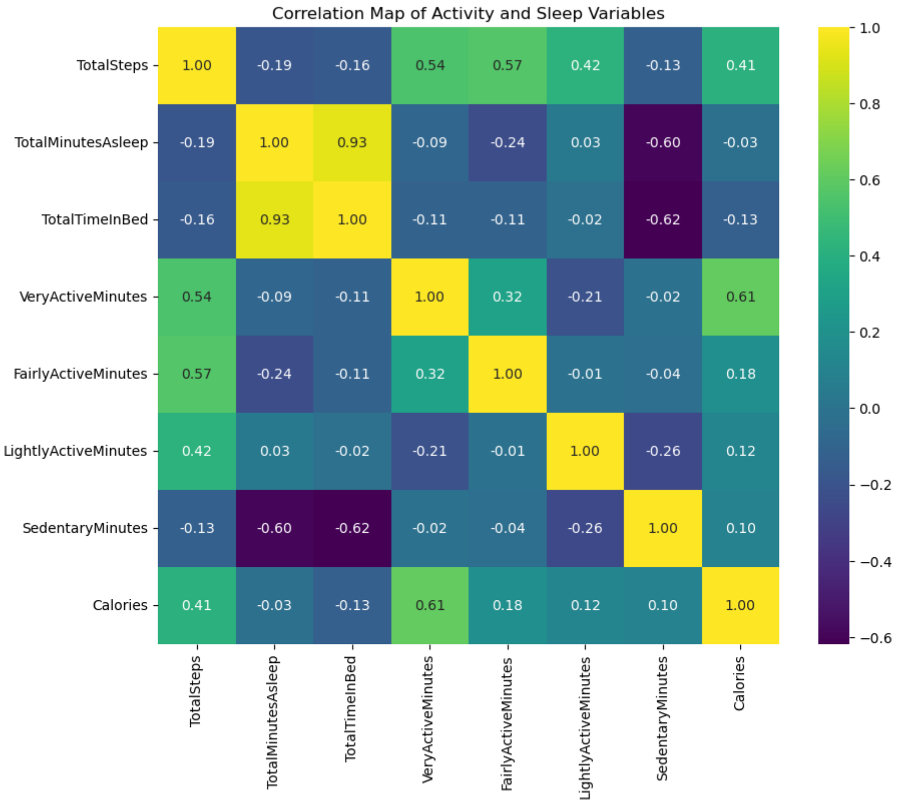
Analysing Sleep and Daily Activity Correlation

The sleep patterns are analysed against the daily activity to identify the correlation between them. A correlation matrix was also visualised to understand the correlation between various features.

```python
df_dailyactivity['date'] = df_dailyactivity['ActivityDate'].dt.date
df_sleepday['date'] = df_sleepday['SleepDay'].dt.date

df_dailyactivity_and_sleep = pd.merge(
    df_dailyactivity,
    df_sleepday,
    on=['Id', 'date'],
    how='inner'
)

correlation = df_dailyactivity_and_sleep['TotalSteps'].corr(df_dailyactivity_and_sleep['TotalMinutesAsleep'])
print('Correlation between Total Steps and Minutes Asleep:', correlation)
```

✓ 0.0s

```
Correlation between Total Steps and Minutes Asleep: -0.18686649892545953
```



The Correlation between Total Steps and Minutes Asleep was found to be approx. -0.19. And in the correlation matrix, it is negative for the features that indicate activity time and sleep. From both the analysis, it is clear that there is almost no linear relationship between the activity and sleep data. This could be because of various other factors which are not available in the dataset; for instance - age, gender, profession, etc. Previously on SQL, it has already been analysed with the sleep next day. Even then, there was no direct correlation. This indicates that we require more data to suggest good sleeping patterns to the user.

Analysing correlation between METs and sleep

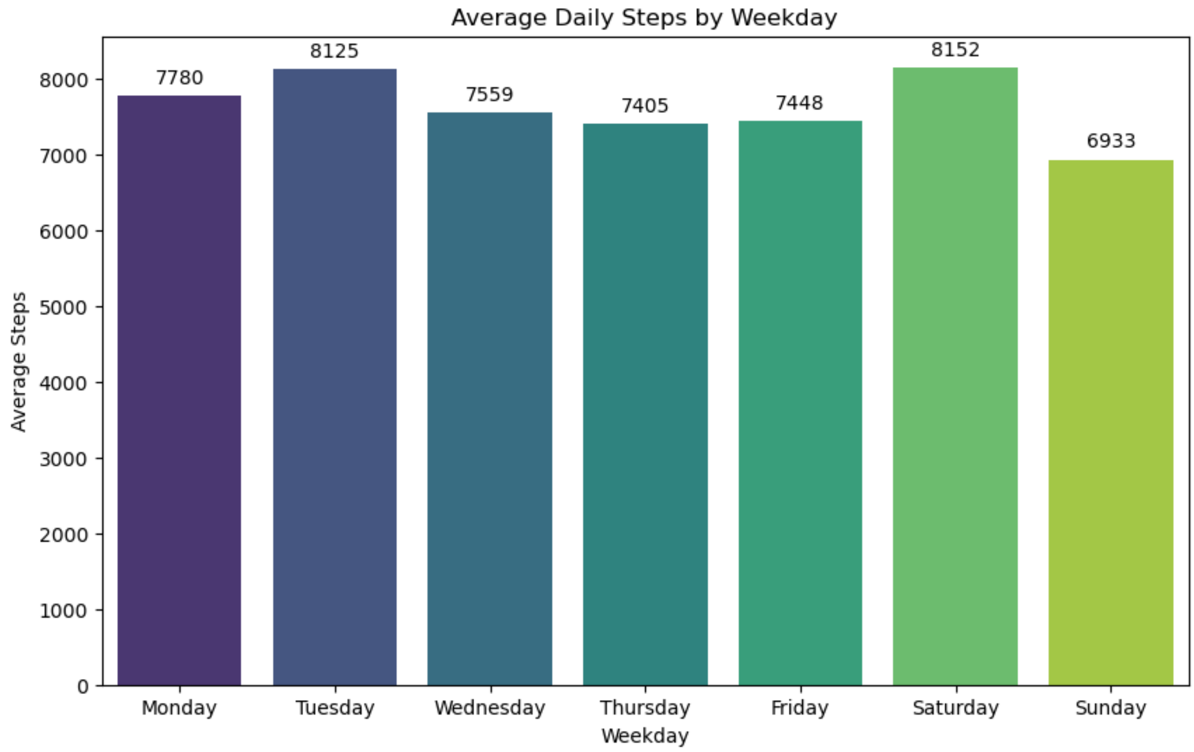
MET stands for Metabolic Equivalent of Task. It is a unit used to estimate the amount of energy expenditure during physical activity compared to resting metabolic rate. In simple terms, an activity with a MET value of 3 means you are burning three times more energy than when at rest.

```
daily_METs = df_minuteMETsnarrow.groupby(['Id', 'ActivityMinute'])['METs'].mean().reset_index()

daily_sleep = df_minutesleep.groupby(['Id', 'date'])['value'].sum().reset_index()

merged_data = pd.merge(daily_METs, daily_sleep, left_on=['Id', 'ActivityMinute'], right_on=['Id', 'date'])
print(merged_data[['METs', 'value']].corr())
✓ 1.2s

           METs     value
METs   1.000000  0.347416
value  0.347416  1.000000
```

The correlation coefficient between METs and sleep is approx. 0.35 between METs and sleep value. This means there is a moderate positive correlation. Activities are usually categorized by METs based on their intensity: light, moderate, or vigorous. for instance, sitting quietly is below 1.5 MET, walking at a moderate pace is about 3-6 METs, and running or intense activity: often 6+ METs. So, if there is a high intensity workout performed, the user can be exhausted and can have a better sleep. So, suggesting high intensity workout to help with sleep improvement can be considered. However, this needs to be done considering the age, gender, and physical limitations of the user.

Analysing the most active days

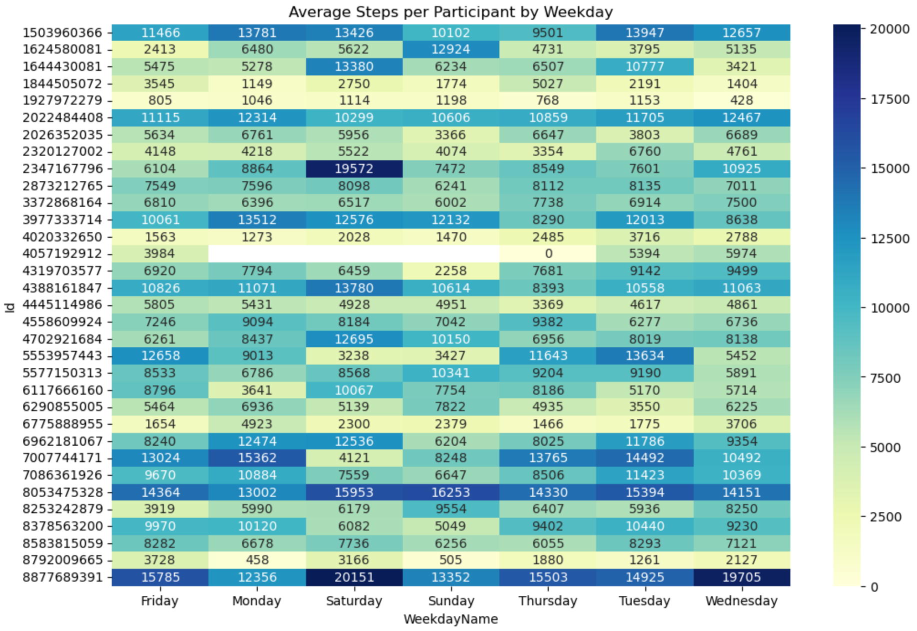
An additional weekday column is created to analyse how the activity is over different days of the week. This can help in providing more motivational inputs on days that lack activity.



From the analysis, it is observed that Sundays have the lowest activity with 6933 steps, and the highest is on Saturdays and Tuesdays with 8152 and 8125 average steps respectively.

Analysing individual user activity on weekdays

A heatmap is generated to understand how the activity is distributed across the week for different users.

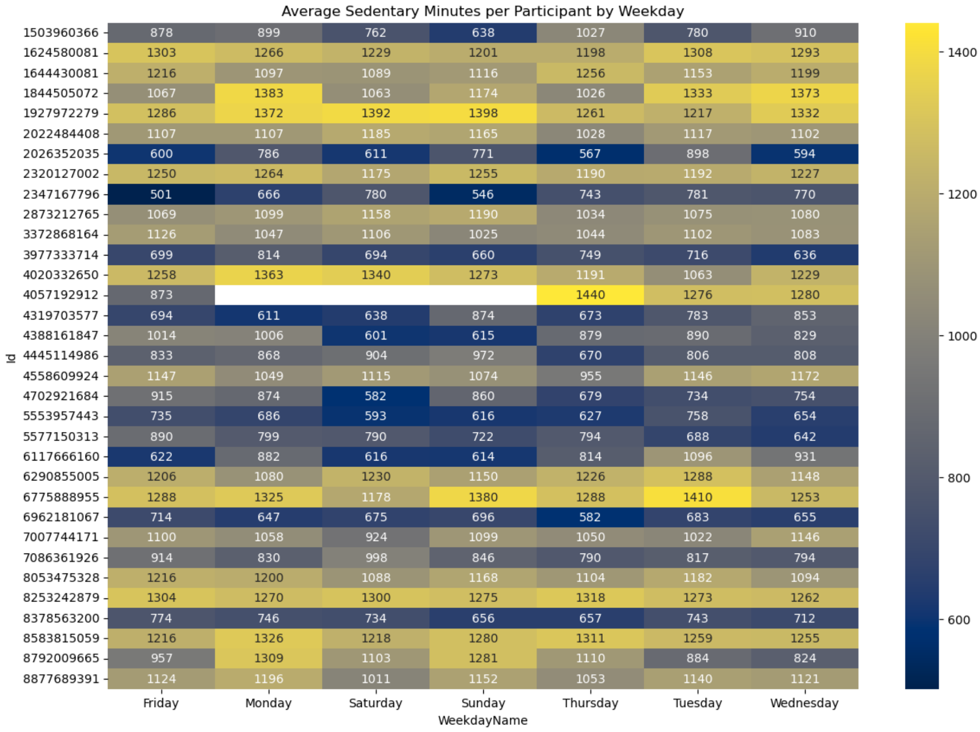


From the heatmap, it can be observed that, some users have consistently high activity across the week. Even they have their highest on Tuesday and Saturday mostly. Some of the users have consistently low usage across the week indicating that they aren't using the tracker properly. A dip in the mid-week can also be observed which could be due to work related commitments. Suggestions to do brisk activities during these periods can engage the user in using the tracker and still keeping the steps consistent. Sundays are the lowest as most users take a break and engage in leisurely activities that day. This can be changed into a rest day or low activity day to improve the motivation on Mondays.

Analysing average sedentary minutes per user per weekday

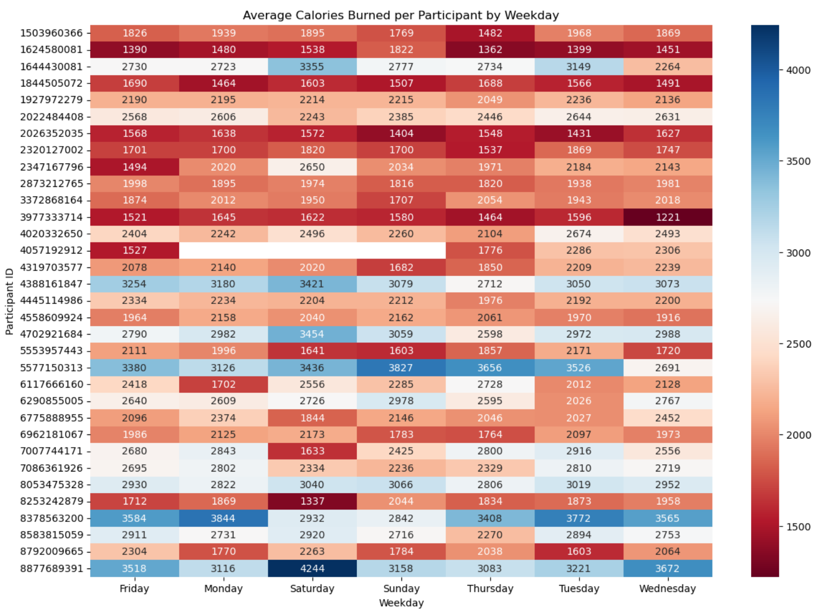
The users with really high sedentary minutes close to 1400 (almost equal to 24 hrs a day) on all days may not be wearing it throughout the day. There must be some sensor to verify this. Additionally, user can be notified via mail or text to use the tracker. Also, the blue shaded columns show the ones with lowest sedentary minutes. They must be wearing it throughout day and night. This way, their sleep is also monitored properly. Motivate the users to do this to get better data to improve sleep related insights too. The white gap in the heatmap is for the user with only 4 rows of data.

From the heatmap below, it is evident that most users aren't consistent with the usage.



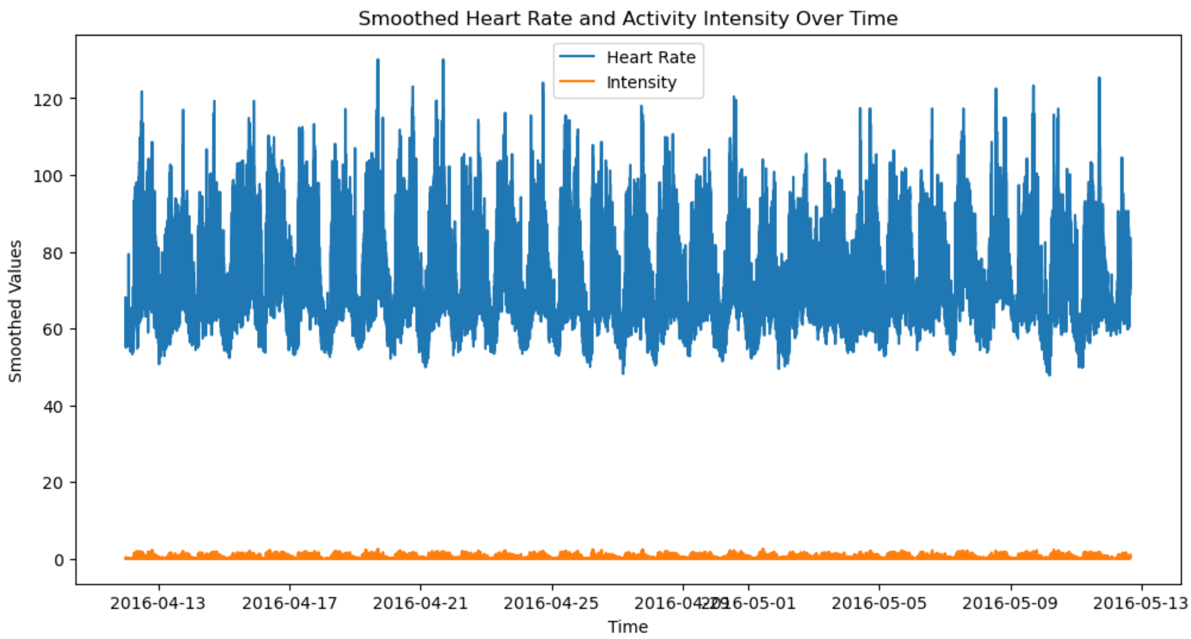
Analysing the calories burned by the user on weekday basis

The results are obvious in the below heatmap. The one who are consistent in using the tracker and have high average steps burn higher calories. However, most of them burn the minimum requirement of calories for an average adult which is good. The calories burned really doesn't matter according to the Harvard university studies. The calorie intake is what matters. If the user is taking extra calories, they need to minimize that or do extra activity to burn them. Without the information on calorie intake, this analysis won't be complete. Stakeholders need to be informed so as to include this information along with the other missing details.



Analysis of correlation between heartrate and intensity of workout

The heart rate and intensity per minute shows a moderate to strong positive correlation of 0.69, indicating that the heart rate increases with increase in intensity and the values taken by the tracker is right. This is helpful in giving useful insights to the user. according to the Harvard studies, for a 40-year-old person with a maximum heart rate of 180, the target heart rate falls somewhere between 117-135 beats per minute for moderate exercise, or 139-167 for vigorous exercise. So, if there is some setting to track the heart beat and notify the user in case of any abnormalities, that would be helpful. Strava fitness uses a third-party ring to track the heartrates. Not all user may have this. This is probably why there aren't much features on the app to monitor the heartbeat. however, keeping track of the heartbeat and consistently checking for abnormalities would be helpful to detect heart problems earlier (similar to the apple watch feature).



Key Insights from Python Analysis

Based on the comprehensive analysis of user activity, sleep, and activity data, the following key insights and strategic suggestions have been identified to enhance user engagement and promote healthier habits:

Personalized Motivation Reminders:

- Use activity and sleep pattern insights to send targeted notifications encouraging users to meet daily step goals or engage in more vigorous workouts.
- Highlight the benefits of higher intensity workouts in improving sleep quality, tailored to individual capabilities.

Promote Consistent Device Usage: Implement reminders or alerts to notify users if they haven't worn the device for significant portions of the day. Additionally, provide regular tracking with rewards, badges, or gamification to boost everyday usage.

Encourage Regular Sleep and Activity Logging: Provide educational content and motivation to improve sleep patterns and increase daily physical activity, based on users' current patterns.

Feature Development and Health Monitoring:

- Develop features that enable heartbeat monitoring, abnormal heart rate detection, and personalized alerts, similar to advanced health wearables.
- Promote these features as premium or safety benefits to incentivize device adoption and ongoing engagement.

Targeted Campaigns for Inactive Users: Identify and re-engage users with low activity or inconsistent tracking through personalized challenges, social sharing, or community-based activities.

Leverage Behavioural Data for Content and Challenges: Design weekly or monthly activity challenges based on observed low-activity days (e.g., Sundays) to motivate users to stay active throughout the week.

Additional Data: Include data regarding profession and personal preference to improve suggestions and keep the user engaged

By leveraging these insights and tailored marketing strategies, the app can increase user engagement, promote healthier habits, and build long-term user loyalty.