Report: Data Analysis of HealthApp Dataset

Introduction:

The HealthApp dataset contains log data collected from a single device, most likely a smartwatch, capturing 35 different measurement features related to physical activities. The dataset contains around 2.5 lakh data points collected over two weeks (12 days) from December 23, 2017, to January 3, 2018. In this report, we present the key insights from the dataset.

Data Insights:

1. **Device Information:** The ID feature in the dataset has the same value for all data points, indicating that the log data was collected from a single device, most likely a smartwatch.

2. Measurement Features: The dataset contains 35 different measurement features, including Step_LSC, Step_SPUtils, Step_ExtSDM, Step_StandReportReceiver, and others. More than 82% of the data points are covered by these trackers. For all modulus, I choose Steo_LSC, step_SPUtils, Step_ExtSDM

3. **Dataset Size:** The size of the dataset is around 1.9 lakhs, making it a medium-sized dataset.

4. **Duplicate Values:** The dataset contains around 425 duplicate values, which can be removed to reduce data redundancy.

5. **Time Period:** All the data was collected in December 2017 and January 2018, and the dataset covers a two-week period from December 23, 2017, to January 3, 2018.

6. **Activity Capture:** In 2017, most of the activities related to calories burned, step count, and altitude were captured by the device in the 24, 30, 31 December. In 2018, most of the activities were captured by the device in 2 of January.

7. **Peak Activity Times:** Users were most active around 8 a.m. in the morning, indicating that users are more likely to start their workouts or activities in the morning.

8. **Activity Trend:** The dataset shows that users were more active in the first week but became less active in the second week except one day (2^{nd} Jan).

Limitations:

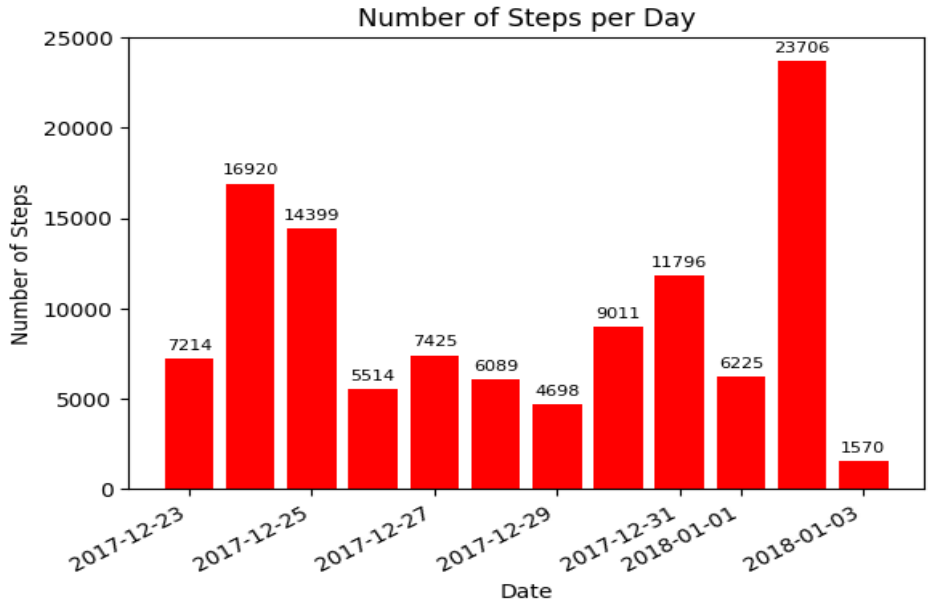
- The dataset only covers a two-week time period, limiting the scope of the analysis.

- The data is collected from a single device, making it difficult to generalize the findings to a larger population.

Conclusion:

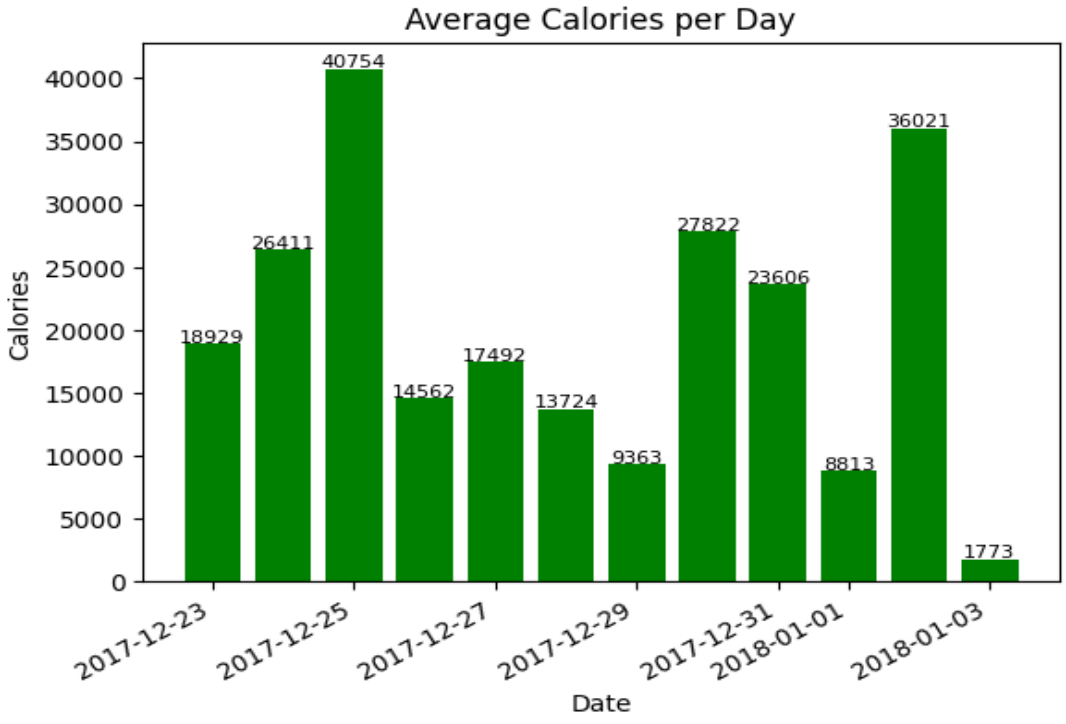
The HealthApp dataset provides valuable insights into the user's activities and behaviour’s during a two-week period (12 days) in 2017 and 2018. These findings could be useful for developing targeted interventions to promote physical activity and healthy behaviours among users of HealthApp. However, further research is needed to confirm these findings and generalize them to a larger population.

Explaining my point with Visualization



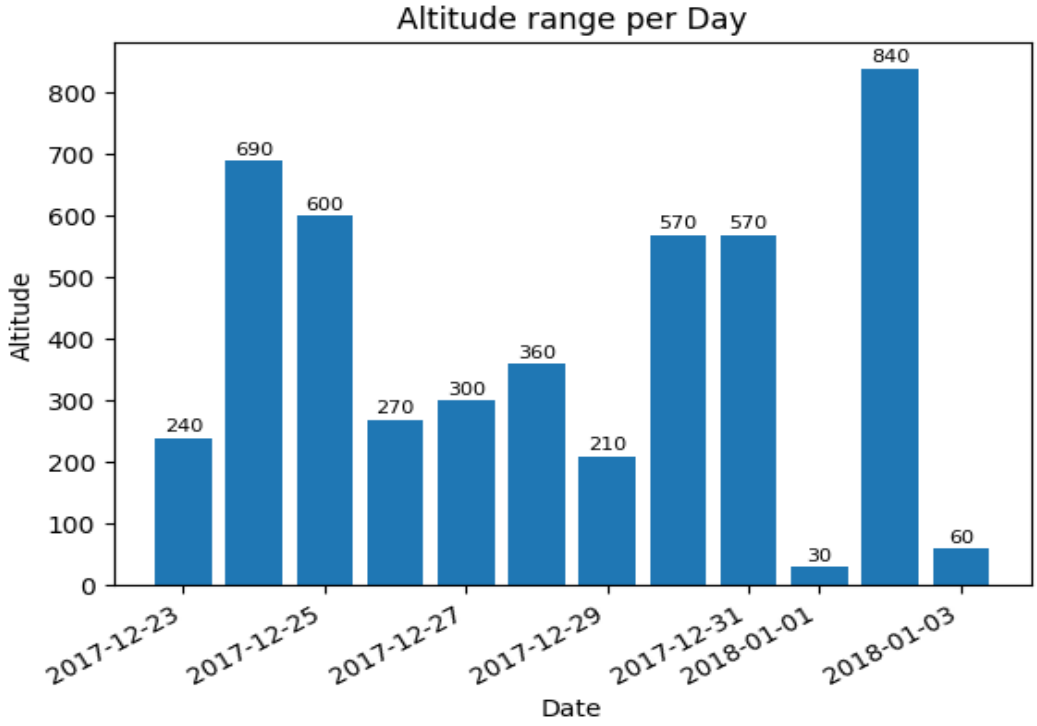
Observations:

- Maximum number of steps cover by a user are 23706.
- User are mostly active in a 12 days data on a 02 Jan 2018, 24 Dec 2017, 25 Dec 2017
- User is mostly unactive on 29 Dec 2017 and 03-01-2018



Observations:

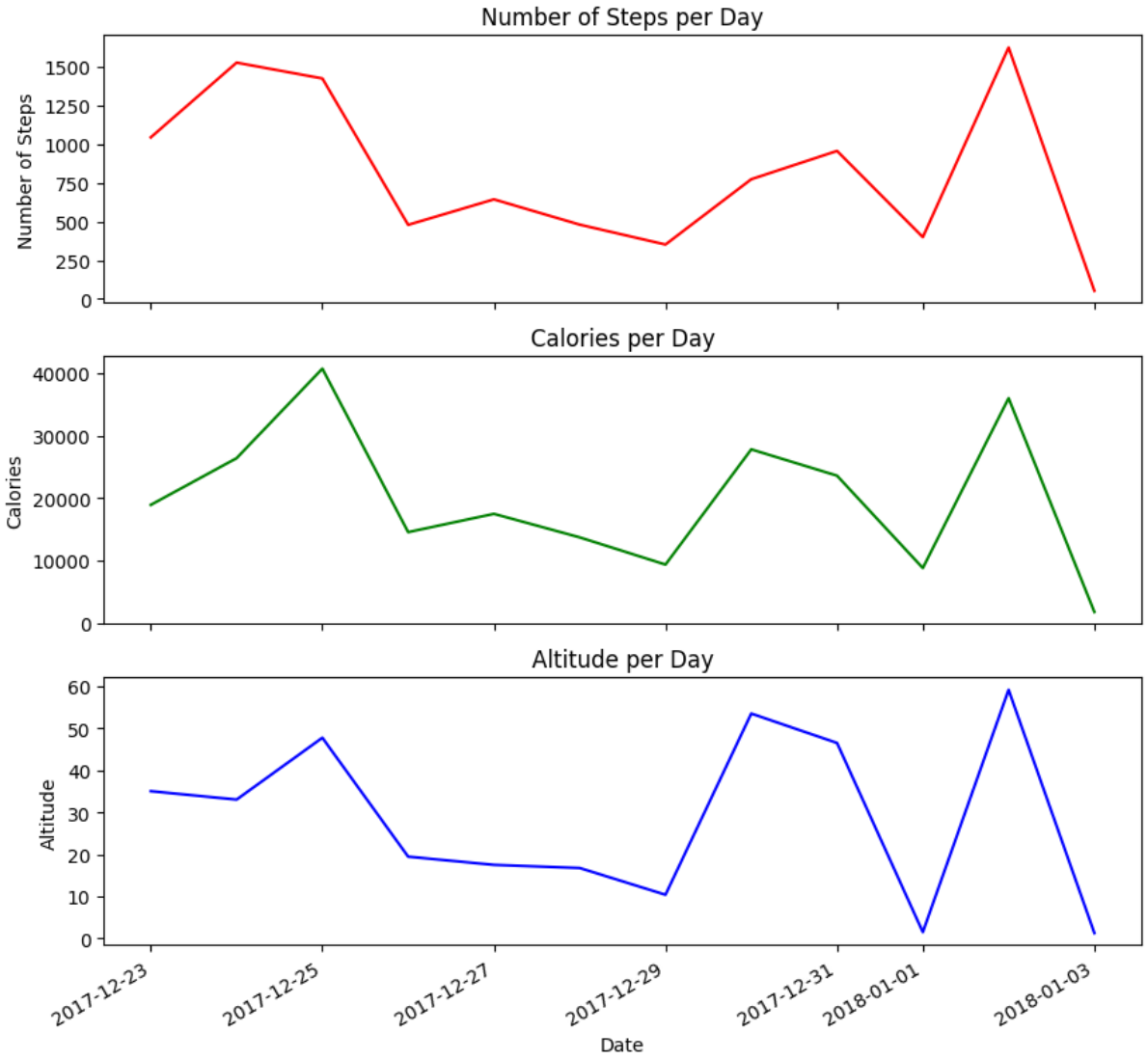
- As we see a graph, we found out that, user maximum calories are burn on 24,25,30,31 Dec 2017 and 2 Jan 2018.
- Minimum calories on 3 Jan 2018 and 29 Dec 2017



Observations:

- Maximum altitude that user reach is 840 on 2 Jan 2018
- Minimum altitude is 30 on 1 Jan 2018
- highest value of altitude that user achieve are 690, 600, 570

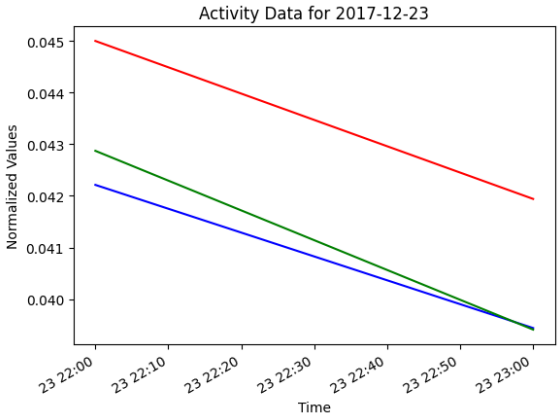
Analyse the all three parameter at once to get more insights.

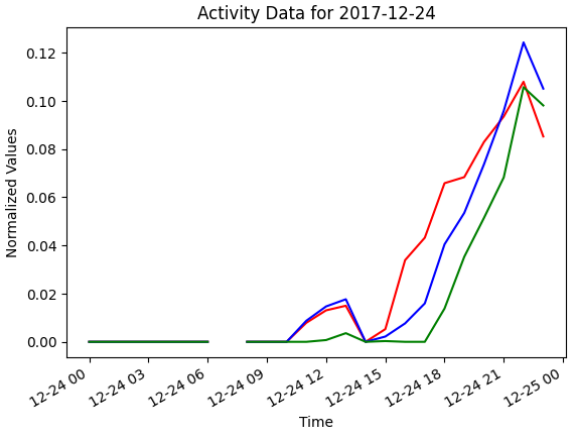


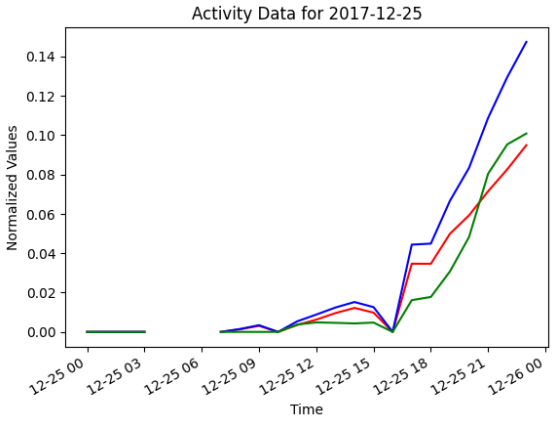
Observations

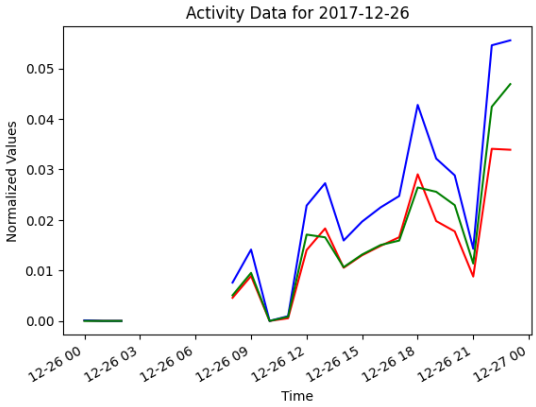
- When a Number of steps is high and altitude is low then calories burn are low.
- When a steps and altitude is low calories, burn are also low.
- When a steps and altitude is high calories, burn is also high.
- User are mostly active (means his step count, altitude and calories all are highest in 12-day period) on 25 Dec 2017 and 2 Jan 2018
- User are mostly unactive (means his step count, altitude and calories all are lowest in 12-day period) on 26 Dec 2017 and 1, 3 Jan 2018

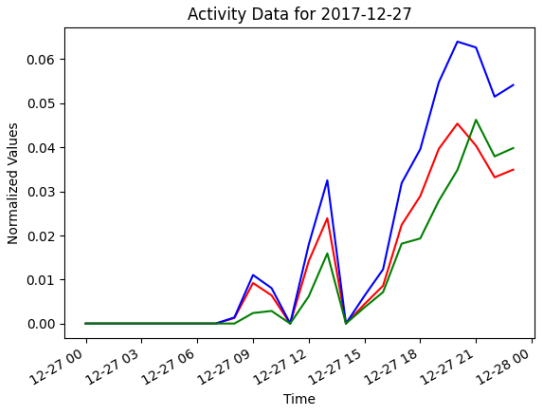
Data analysis of per day on hourly bases

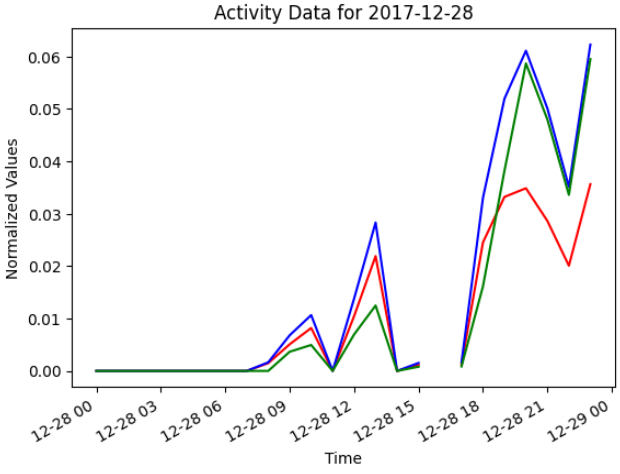


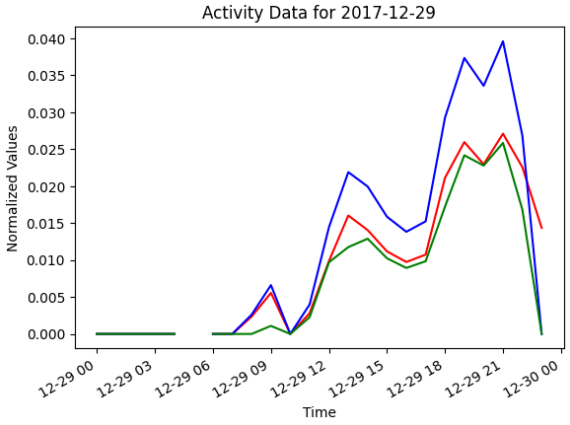


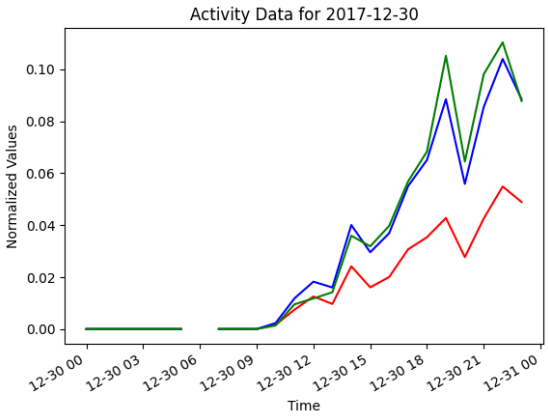


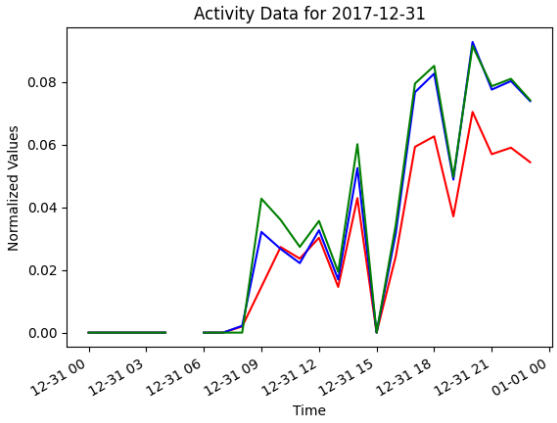


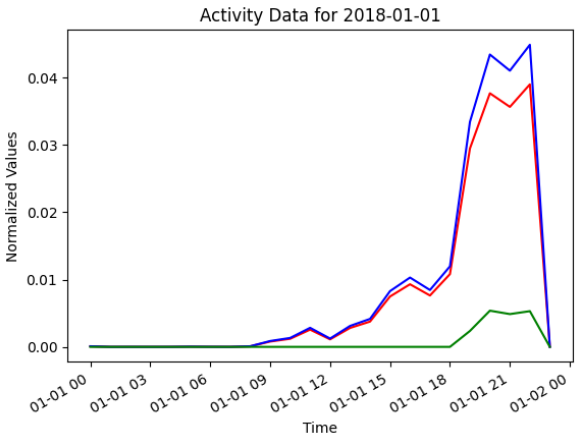


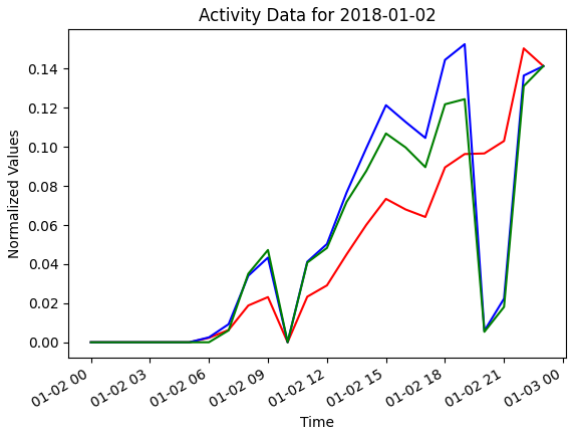


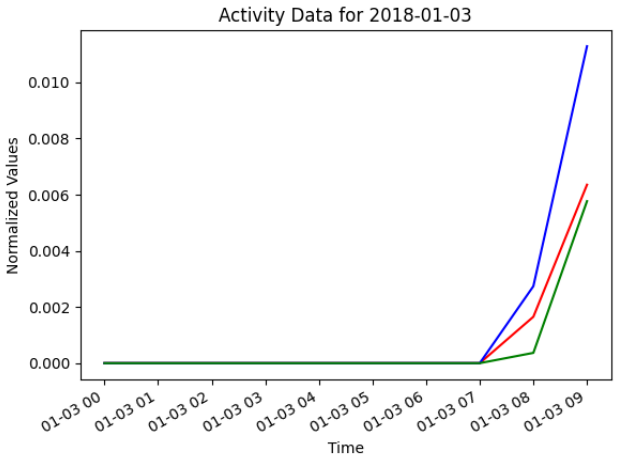












Observations:

- After analysis these graphs, I learn about a behaviour of user.
- User wake up or start there day around 8am most of his weekdays.
- On weekend user wakeup or start his day around 10am
- User mostly active on 2 Jan because that day he wakes up around 6am and burn high calories, reach at highest altitude and cover highest number of steps also.