Bellabeat: Trends on Smartwatch Users

By Miguel Antonio Li

The Business Task

Bellabeat is a high-tech company that manufactures health-focused smart products. These products helps inform and inspire women around the world. The founders of the company believe that there are more opportunities to grow their business. With this, they have discussed with the marketing team on how they could gain insight on their customer's usage of their products. As part of the marketing analytics team, I was tasked to look into one of the products and analyze the the smart device usage data in order to gain new findings on how their customers are already using their device. The goal in my analysis is to find trends in the smart device usage and present how my findings can help influence the company's marketing strategy.

One of the stakeholders in this task is Urska Srsen who is the Bellabeat's cofounder and Chief Creative Officer. Another stakeholder that I must take under consideration is Sando Mur who is a Mathematician and Bellabeat'scofounder; she is also a member of the Bellabeat executive team.

Data Sources Used

For this analysis, I will be using public data that explores smart device users' daily habits. The data I will be using is the FitBit Fitness Tracker Data which is a public domain which is made available through Mobius. This dataset contains personal fitness track from around 33 fitibit users. These users are consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitory. The data also contains information on the users' about daily activity, and steps. This data is organized in a long format as there are multiple entries with the same user IDs.

Due to the fact that the data is based on actual data from fitbit and is approved to be available through Mobius, I am confident to say that the data provided is reliable, original, and cited. As I have mentioned previously, the data provided contains all sorts of data that were taken from the users' device usage. Provided the extensive information, I would say that this data is comprehensive. Lastly, the modifications of these files were made of last year which is to say that the data that I will be using for my analysis is current. Although the data is in accordance to ROCCC, I found a couple issues of the data. Regarding the data on the weight_log, it only contains the data of eight users. As for the data on Sleep, there are only twenty-six users that are listed down. Due to of users in these data tables, I will not be combining them with the other data sets as it would bring inconsistent results when it comes to weight analysis and sleep analysis.

Cleaning and Manipulation of Data

The tools that I have used for cleaning and manipulation of data are the following – Excel, Python, and PostgreSQL. To start, I duplicated the current data tables so that I have a back up file in case I encounter an error while working on my analysis. I used python pandas library to extract the data from the excel files and uploaded into the database. Through the python script, I checked if there are any duplications in the rows of data before uploading it into the database. When all the data has been uploaded into the database, I double check if each table has the same number of rows. This is to make sure that there will not be any irregularities once I join the different tables together. I have joined all the tables except for "daily_activty", "weight_log", and "sleep_day". The data table is not used to join the other tables because it is summary of all the other tables. As mentioned earlier, the data sets on weight_log and sleep_analysis had a lack of data which means that they cannot be joined with all the other tables but they will be used as an different analysis. While cleaning the data, I had not found any irregularities or inaccuracies in the data set. After cleaning and sorting my data, I was able to validate my joined data tables as I compared my results to the data table of "daily_activity" which the resulted to no results once executed.

A summary of your analysis

To start of my analysis, I checked the correlation between the different distances to the average calories. After setting up the query, I find out that the very_active_distance does have a higher calorie count among the other distances as the correlation between the average calories and the very active distance is 0.5. This means that more calories are burned more if users went on to more active distances.

```sql
SELECT
    CORR(very_active_distance, avg_calories) AS corr_very_active_calories,
    CORR(moderately_active_distance, avg_calories) AS corr_moderate_active_calories,
    CORR(light_active_distance, avg_calories) AS corr_light_active_calories,
    CORR(sedentary_active_distance, avg_calories) AS corr_sedentary_calories
FROM fitbit_summary;
```

Data Output | Messages | Notifications

Showing rows:

	corr_very_active_calories double precision	corr_moderate_active_calories double precision	corr_light_active_calories double precision	corr_sedentary_calories double precision
1	0.51161932557632	0.061978654696414406	0.35110002554182707	-0.02632542680389649

The next step I did is the examination between the correlation between average total steps and average total calories burned. After manipulating the data, the result was 0.43, indicating a moderate positive relationship between the two variables. In other words, as the number of steps increases, the number of calories burned also tends to increase.

To explore this further, I checked whether there were users who had high calorie burn but lower step counts. The analysis revealed that seven out of thirty-three users burned more calories than average despite having fewer steps.

To understand why, I examined their activity levels to see how strenuous their workouts were. Interestingly, six out of the seven users fell under the "Low Very Active Distance" category, meaning they didn't engage in high-intensity activities.

```sql
SELECT "User Id", avg_total_steps, avg_calories,activity_level
FROM fitbit_summary
WHERE avg_total_steps > (SELECT AVG(avg_total_steps) FROM fitbit_summary)
  AND avg_calories < (SELECT AVG(avg_calories) FROM fitbit_summary);
```

Data Output Messages Notifications

	User Id bigint	avg_total_steps numeric	avg_calories numeric	activity_level text
1	4558609924	7685.1290322580645161	2033.2580645161290323	Low Very Active Distance
2	1503960366	12116.741935483871	1816.4193548387096774	Moderate Very Active Distance
3	3977333714	10984.566666666667	1513.6666666666666667	Low Very Active Distance
4	6962181067	9794.8064516129032258	1982.0322580645161290	Low Very Active Distance
5	2873212765	7555.7741935483870968	1916.9677419354838710	Low Very Active Distance
6	5553957443	8612.5806451612903226	1875.6774193548387097	Low Very Active Distance
7	2347167796	9519.6666666666666667	2043.4444444444444444	Low Very Active Distance

Since this result seemed counterintuitive, I looked deeper into their active minutes. It turns out that these users, despite their low active level, had higher Lightly Active Minutes, suggesting that they burned more calories through prolonged, lower-intensity activity rather than short bursts of intense exercise.

```sql
SELECT "User Id", avg_total_steps, avg_calories, lightly_active_minutes, very_active_minutes
FROM fitbit_summary
WHERE avg_total_steps > (SELECT AVG(avg_total_steps) FROM fitbit_summary)
  AND avg_calories < (SELECT AVG(avg_calories) FROM fitbit_summary);
```

Data Output Messages Notifications

	User Id bigint	avg_total_steps numeric	avg_calories numeric	lightly_active_minutes numeric	very_active_minutes numeric
1	4558609924	7685.1290322580645161	2033.2580645161290323	284.9677419354838710	10.3870967741935484
2	1503960366	12116.741935483871	1816.4193548387096774	219.9354838709677419	38.7096774193548387
3	3977333714	10984.566666666667	1513.6666666666666667	174.7666666666666667	18.9000000000000000
4	6962181067	9794.8064516129032258	1982.0322580645161290	245.8064516129032258	22.8064516129032258
5	2873212765	7555.7741935483870968	1916.9677419354838710	308.0000000000000000	14.0967741935483871
6	5553957443	8612.5806451612903226	1875.6774193548387097	206.1935483870967742	23.4193548387096774
7	2347167796	9519.6666666666666667	2043.4444444444444444	252.5000000000000000	13.5000000000000000

This finding suggests that calorie burn is influenced not just by step count but also by the type and duration of activity performed. Further analysis could explore whether longer but lower-intensity movement is as effective as shorter, high-intensity workouts for burning calories. To execute this analysis, I checked the avg calories burned per minute in the different activity levels. Based on the data, the users who are in the category of low activity burns more calories per minute compared to those who are very active.

```sql
SELECT
    activity_level,
    AVG(avg_calories / NULLIF((lightly_active_minutes + fairly_active_minutes + very_active_minutes), 0)) AS avg_calories_per_minute
FROM fitbit_summary
GROUP BY activity_level
ORDER BY avg_calories_per_minute DESC;
```

Data Output Messages Notifications

Showing rows: 1 to 3 Page No: 1 of 1

	activity_level text	avg_calories_per_minute numeric
1	Low Very Active Distance	13.48887641430208273002
2	High Very Active Distance	11.4984312008416216
3	Moderate Very Active Distance	10.5938151344984521

I investigated my findings further by checking the total duration of the users' activities. This is to check whether users in the category of low activity engage in lower activities. The results I got from my analysis is that users in the low active category tend to have more active minutes than the other categories. In other words, longer durations of low-intensity movement can result in a lower calorie burn.

```sql
SELECT
    activity_level,
    SUM(lightly_active_minutes + fairly_active_minutes + very_active_minutes) AS total_active_minutes,
    SUM(avg_calories) AS total_calories_burned
FROM fitbit_summary
GROUP BY activity_level
ORDER BY total_calories_burned DESC;
```

Data Output Messages Notifications

Showing rows: 1 to 3

	activity_level text	total_active_minutes numeric	total_calories_burned numeric
1	Low Very Active Distance	5098.97555844847279810288	50933.6203625382157085
2	Moderate Very Active Distance	1772.5849810630795352	18020.9559139784946237
3	High Very Active Distance	556.4193548387096775	6366.0645161290322581

I then checked the relationship of sedentary minutes to the different categories of users. I then find that users who are highly active tend to have a higher average of sedentary minutes. This means that that the highly active users are engaging in intense, but sporadic, bursts of activity, with long recovery periods between them. while sedentary minutes are high, highly active users might still benefit from more structured recovery or less inactivity throughout the day. Encouraging short bursts of movement throughout the day (e.g., walking or stretching) could help these users improve overall health.

Query Query History

```sql
SELECT
    activity_level,
    AVG(sedentary_minutes) AS avg_sedentary_minutes,
    AVG(very_active_minutes) AS avg_very_active_minutes,
    AVG(fairly_active_minutes) AS avg_moderate_active_minutes
FROM
    fitbit_summary
GROUP BY
    activity_level
ORDER BY
    avg_sedentary_minutes DESC;
```

Data Output Messages Notifications

	activity_level text	avg_sedentary_minutes numeric	avg_very_active_minutes numeric	avg_moderate_active_minutes numeric
1	High Very Active Distance	1130.4354838709677420	75.6129032258064517	9.7580645161290323
2	Low Very Active Distance	1003.6107173076583533	8.49197476914384589878	11.81351657575934550968
3	Moderate Very Active Distance	946.3529269824685784	45.0223107396282269	19.2210551191875470

The next thing I wanted to analyze is if the users' sleep duration affects the activities of the users. After manipulating the data, I later find that theres sleep duration doesn't the activities of the users because users with less sleep that seemed to burn more calories and have more steps compared to those with more sleep.

```sql
SELECT
    f."User Id",
    AVG(s."TotalMinutesAsleep") AS avg_sleep_minutes,
    AVG(f.avg_calories) AS avg_calories_next_day,
    AVG(f.avg_total_steps) AS avg_steps_next_day
FROM fitbit_summary f
JOIN sleep_day s ON f."User Id" = s."Id"
GROUP BY f."User Id"
ORDER BY avg_sleep_minutes DESC;
```

Data Output Messages Notifications

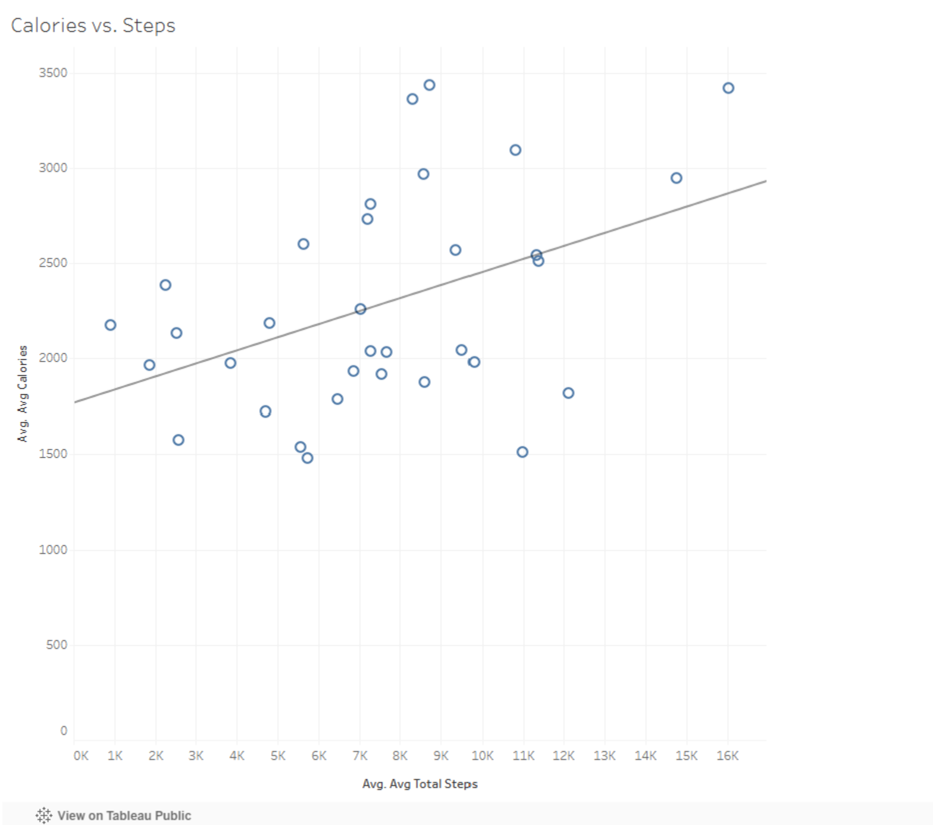
	User Id bigint	avg_sleep_minutes numeric	avg_calories_next_day numeric	avg_steps_next_day numeric
1	1844505072	652.0000000000000000	1573.4838709677419355	2580.0645161290322581
2	2026352035	506.1785714285714286	1540.6451612903225806	5566.8709677419354839
3	6117666160	478.7777777777777778	2261.1428571428571429	7046.7142857142857143
4	4319703577	476.6538461538461538	2037.6774193548387097	7268.8387096774193548
5	5553957443	463.4838709677419355	1875.6774193548387097	8612.5806451612903226
6	7086361926	453.1250000000000000	2566.3548387096774194	9371.7741935483870968
7	6962181067	448.0000000000000000	1982.0322580645161290	9794.8064516129032258
8	2347167796	446.8000000000000000	2043.4444444444444444	9519.6666666666666667
9	8378563200	445.1290322580645161	3436.5806451612903226	8717.7096774193548387
10	8792009665	435.6666666666666667	1962.3103448275862069	1853.7241379310344828
11	5577150313	432.0000000000000000	3359.6333333333333333	8304.4333333333333333
12	4702921684	417.4814814814814815	2965.5483870967741935	8572.0645161290322581

The analysis of user activity levels, sedentary time, and calorie burn has provided valuable insights into the relationship between activity, rest, and performance. Users with high activity levels (very active or moderately active) tend to have higher average sedentary minutes compared to those with lower activity levels. This suggests that highly active individuals may engage in intense exercises or workouts that require extended periods of rest and recovery afterward. Despite the increased sedentary time, these active users still manage to burn significantly more calories during their workout sessions, indicating that short bursts of high-intensity activity are highly effective for calorie burning.

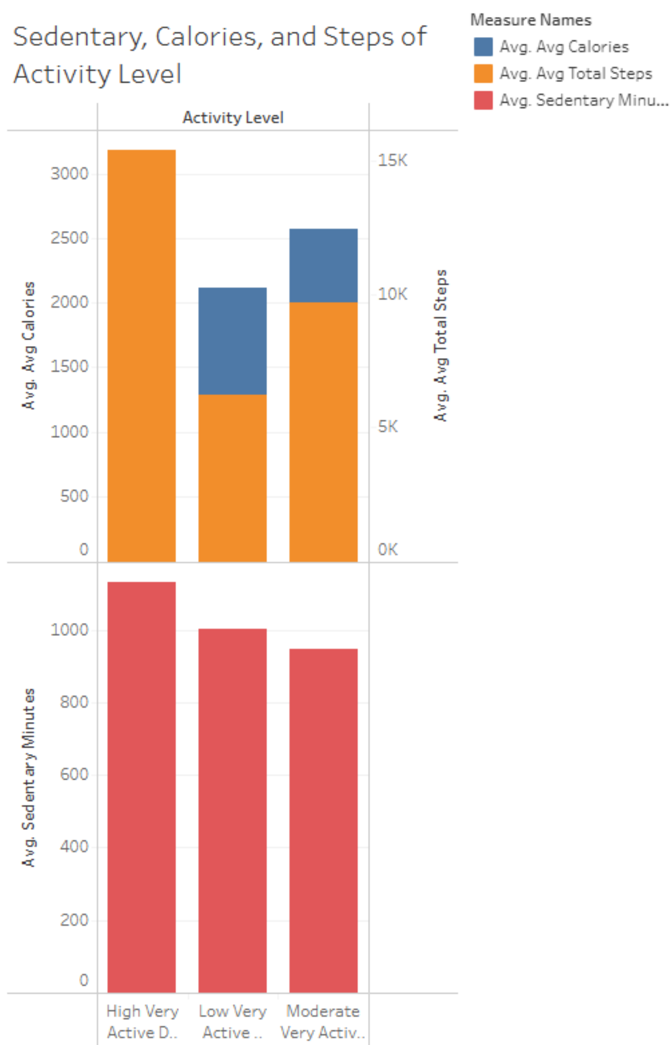
Further analysis revealed a positive correlation between steps and calories burned, where users with more steps generally burn more calories. However, users with fewer steps can still burn more calories due to prolonged, lower-intensity movement, such as lightly active minutes. Interestingly, users who engage in higher levels of lightly active minutes appear to burn more calories per minute compared to those who engage in high-intensity exercise. This suggests that longer durations of moderate-intensity activity can be just as effective, if not more, than short bursts of high-intensity exercise for calorie burn.

The investigation into the correlation between sleep and activity levels showed no significant relationship. In fact, users with less sleep sometimes burned more calories and had higher steps compared to those who slept more. This indicates that while sleep is crucial for recovery, activity levels and calorie burn seem to be more strongly influenced by the type and intensity of physical activity rather than sleep duration alone. Additionally, no significant correlation was found between sleep efficiency and user activity levels. This suggests that sleep quality alone does not necessarily determine how active users are during the day, although overall sleep duration may still play a role in their recovery and energy levels.

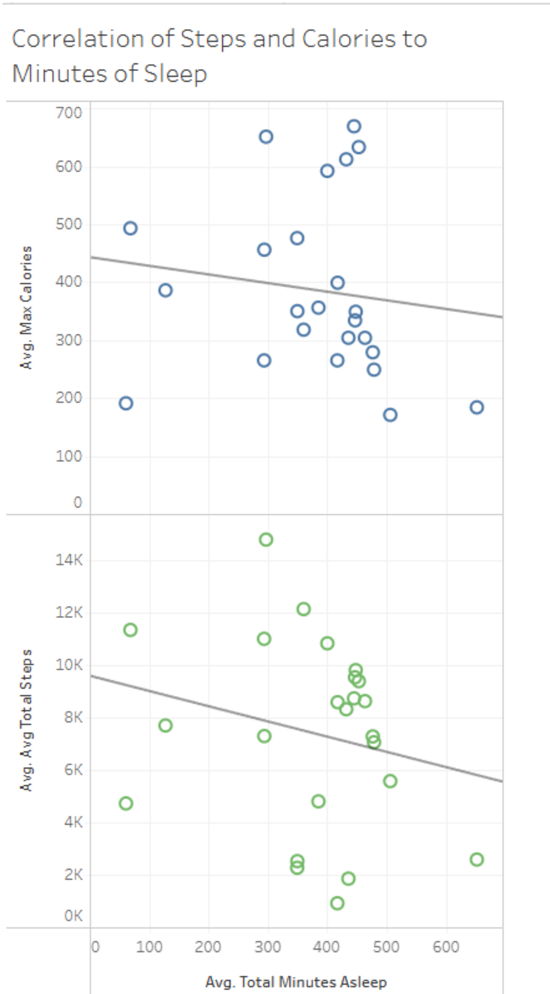
Supporting Visualizations and Key Findings



The figure above displays the correlation between the average calories burned to the average steps of the users. Based on the figure above, there is a positive correlation between the calories burned to the total steps taken which means the more steps the user take, the more calories are burned.



The figure above shows the average calories burned, average total steps taken , and average sedentary minutes per category of user. Based on the figure above, users who are highly active tend to have the most amount sedentary minutes which means that these users tend to rest a lot more compared to the other users. The figure also shows that those users who are not so active tend to burn a lot more calories compared to the other users. This is because they do low-intensity activities for a long duration.



The figure above shows the correlation of the minutes of sleep to the average total steps taken and calories burned. Based on the data above, there is a negative correlation in both graphs which means that the amount of sleep does not affect the total steps take and calories burned.

Recommendations

What Bellabeat can do is to create a personalized workout plan for the users that would bring balance between high and low intensity workouts to avoid burnout while still maximizing calorie burn. When the Bellabeat reads that the user is burning few calories, the smartwatch can notify the user and can recommend an increase in the duration of lighter exercises. Another recommendation is to incorporate more active minutes. This can be done by having the Bellabeat smartwatch notify the user to do light activities such as standing up, stretching, or even walking every hour when users are not doing any activity. Last but not the least, my other recommendation is to notify the user when they are not getting enough sleep. It is important for the users to sleep for at least 7-8 hours for better recovery and energy levels.