

Bellabeat Fitness Data Analysis Project

Background + Scenario

- Bellabeat, a rapidly growing company specializing in health-focused products for women, is poised to expand its presence in the global smart device market.
- I am acting as a junior data analyst working on the marketing analyst team at Bellabeat
- Co-founders Urška Sršen and Sando Mur believe that leveraging insights from smart device fitness data will unlock significant growth opportunities.

Step One: Ask

- **Business Task:** I have the responsibility of analyzing consumer smart device usage data for a non-Bellabeat smart device to gain insights into user behavior. I will use this data to provide recommendations on how these trends can inform Bellabeat's marketing strategy and catapult them to be a global player in the smart device market.
- **Key Stakeholders:** Urška Sršen and Sando Mur (both co-founders), Bellabeat marketing analytics team

Step Two: Prepare

- **Data:** <https://www.kaggle.com/datasets/arashnic/fitbit>
- **What is in it?:** The dataset contains personal information from 30 individuals who consented to share their data through Fitbit which sells a similar type of product to Bellabeat. This includes detailed records of physical activity, heart rate, weight logs and sleep tracking on a minute-by-minute basis. Additionally, it provides insights into daily activities, such as step count and heart rate.
- **Is it reliable?:** We will use the ROCCC method to analyze this data source for credibility and bias
 - (R)eliability: This dataset was collected from 30 real FitBit users who voluntarily provided their personal tracking data. The data was gathered with a survey through Amazon Mechanical Turk.
 - (O)riginality: This dataset is directly from users who submitted their data via the survey.
 - (C)omprehensiveness: This dataset is rather small with only 30 users, and there is no indication of factors like race, sex, nationality, or place of residence. How old were they? Were they chosen randomly? These questions cannot be answered as the descriptive data is very limited.
 - (C)urrent: This dataset is from a survey conducted in 2016.
 - (C)ited: This dataset was created by Amazon Mechanical Turk.
- **Dataset Obstacles:** For the dataset I was given, I decided to focus on the merged daily activity table and the sleep data table. The daily activity table contains 33 entries of user data, while the sleep data table only contains 24 entries of user data, meaning there is sleep data missing for 9 users. We will go through and only use the 24 user entries that appear in both tables to have an accurate analysis.

Step Three: Process

- **Data Cleaning:** For the data cleaning process, I chose to use SQL and Excel. With SQL, I was able to filter out any duplicate entries with the DISTINCT function. Next, I was able to go ahead and clean the dailyActivity_merged.csv file by calculating the average values for each user, consolidating their data into a single row per user for each category. For the sleepDay_merged.csv file, I applied a similar approach, but instead of averaging, I summed the Total Sleep Records for each user and grouped them by their ID's.

```
1 SELECT DISTINCT Id,
2   AVG(Calories) AS AvgCalories,
3   AVG(TotalDistance) AS AvgTotalDistance,
4   AVG(TrackerDistance) AS AvgTrackerDistance,
5   AVG(SedentaryMinutes) AS AvgSedentaryMinutes,
6   AVG(VeryActiveMinutes) AS AvgVeryActiveMinutes,
7   AVG(FairlyActiveMinutes) AS AvgFairlyActiveMinutes,
8   AVG(LightlyActiveMinutes) AS AvgLightlyActiveMinutes,
9   AVG(TotalSteps) AS AvgTotalSteps
10
11 FROM `bellabeat-casestudy-433900.Fitbit.dailyActivity_merged`
```

```
1 SELECT DISTINCT Id,
2   SUM(TotalSleepRecords) AS TotalSleepRecords,
3   AVG(TotalMinutesAsleep) AS TotalMinutesAsleep,
4   AVG(TotalTimeInBed) AS TotalTimeInBed
5
6 FROM `bellabeat-casestudy-433900.Fitbit.sleepDay_merged`
7
8 GROUP BY Id;
```

- I followed this up by transferring both tables into Excel to get rid of the 9 user data entries that don't appear in both tables and changed the user ID numbers into a simpler user# naming system.

Id	AvgCalories	AvgTotalDistance	AvgTrackerDistance	AvgSedentaryMinutes	AvgVeryActiveMinutes	AvgFairlyActiveMinutes	AvgLightlyActiveMinutes	AvgTotalSteps
User1	1,816.42	7.81	7.81	848.16	38.71	19.16	219.94	12,116.74
User2	2,811.30	5.3	5.3	1,161.87	9.57	21.37	178.47	7,282.97
User3	1,573.48	1.71	1.71	1,206.61	0.13	1.29	115.45	2,580.06
User4	2,172.81	0.63	0.63	1,317.42	1.32	0.77	38.58	916.13
User5	1,540.65	3.45	3.45	689.42	0.1	0.26	256.65	5,566.87
User6	1,724.16	3.19	3.19	1,220.10	1.35	2.58	198.19	4,716.87
User7	2,043.44	6.36	6.36	687.17	13.5	20.56	252.5	9,519.67
User8	1,513.67	7.52	7.52	707.53	18.9	61.27	174.77	10,984.57
User9	2,385.81	1.63	1.63	1,237.26	5.19	5.35	76.94	2,267.23
User10	2,037.68	4.89	4.89	735.81	3.58	12.32	228.77	7,268.84
User11	3,093.87	8.39	8.39	836.68	23.16	20.35	229.35	10,813.94
User12	2,186.19	3.25	3.25	829.9	6.61	1.74	209.1	4,796.55
User13	2,033.26	5.08	5.08	1,093.61	10.39	13.71	284.97	7,685.13
User14	2,965.55	6.96	6.96	766.42	5.13	26.03	237.48	8,572.06
User15	1,875.68	5.64	5.64	668.35	23.42	13	206.19	8,612.58
User16	3,359.63	6.21	6.21	754.43	87.33	29.83	147.93	8,304.43
User17	2,261.14	5.34	5.34	796.29	1.57	2.04	288.36	7,046.71
User18	2,131.77	1.81	1.81	1,299.42	11	14.81	40.15	2,519.69
User19	1,982.03	6.59	6.52	662.32	22.81	18.52	245.81	9,794.81
User20	2,544.00	8.02	7.58	1,055.35	31.04	16.27	280.73	11,323.42
User21	2,566.35	6.39	6.39	850.45	42.58	25.35	143.84	9,371.77
User22	2,945.81	11.48	11.48	1,148.00	85.16	9.58	150.97	14,763.29
User23	3,436.58	6.91	6.91	716.13	58.68	10.26	156.1	8,717.71
User24	1,962.31	1.19	1.19	1,060.48	0.97	4.03	91.79	1,853.72

Id	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed
User1	27	360.28	383.2
User2	4	294	346
User3	3	652	961
User4	8	417	437.8
User5	28	506.18	537.64
User6	1	61	69
User7	15	446.8	491.33
User8	32	293.64	461.14
User9	8	349.38	379.75
User10	27	476.65	501.96
User11	31	403.13	426.21
User12	39	385.18	416.82
User13	5	127.6	140
User14	30	421.14	441.96
User15	38	463.48	505.87
User16	27	432	460.62
User17	22	478.78	510.17
User18	3	349.67	369
User19	34	448	466.13
User20	2	68.5	71.5
User21	24	453.13	466.42
User22	3	297	301.67
User23	36	443.34	483.31
User24	15	435.67	453.8

Step Four: Analyze

- **Finding Patterns & Trends:** Due to the fact that humans sleep for a third of their lives on average, I wanted to dive deeper into the sleep data to better understand how it impacts these users' daily activity. I can then use this user behavior data to provide valuable insights for Bellabeat.

- Looking at the sleep data for each user and then comparing it to their activity it is pretty easy to see that it is hard to be active if you don't get enough sleep or if you sleep for too long. On one end of the spectrum User3 slept the most (652 min), and only burned 1,573 calories. On the other hand, User6 only slept for 61 minutes and ended up burning only 1,724 calories. I would assume that the cause for this is being tired when you have underslept or having less hours of sunlight if you have overslept. I have come to a solid conclusion that it is most optimal to get around 375 to 450 minutes of sleep based on the data from these 24 users.. There also seems to be the same correlation between sleep and steps taken.
- The very active, fairly active, lightly active, and sedentary labels are descriptions through Fitbit based on the frequency, duration, intensity, and patterns of your movement. Provided with the user's data, you are clearly able to see how well this correlates with someone's overall activity. If someone is described as being very active, they tend to travel farther in a day, take more steps, etc. This is shown in viz 3 by the trend line.

Step Five: Share

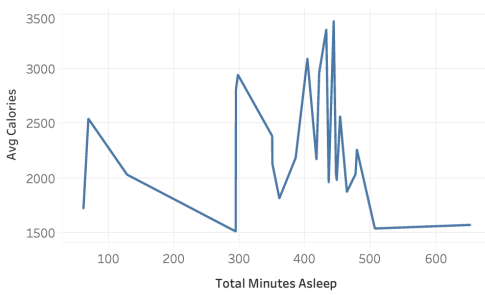
- Here is the dashboard I was able to create with Tableau containing four useful visualizations that illustrate these users' data through Fitbit.

- Viz 1: Compares the total minutes asleep to the average amount of calories burned by the users
- Viz 2: Compares the total minutes asleep to the average total amount of steps taken by the users
- Viz 3: Compares the amount of very active minutes to the average total distance traveled by the users
- Viz 4: Illustrates the division of the activity level labels that are given by Fitbit to the users based on the frequency, duration, intensity, and patterns of their movement

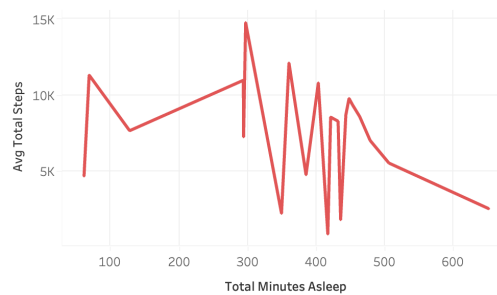
Fitbit User Data Dashboard



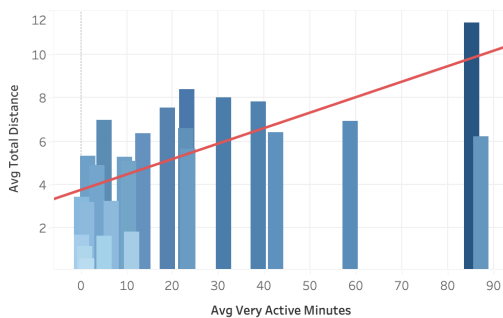
Avg Calories Burned vs Total Minutes Asleep



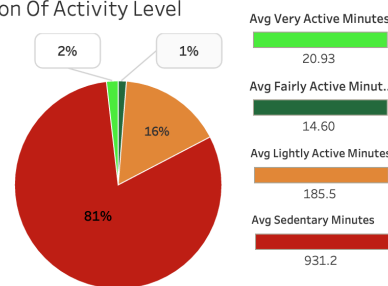
Avg Total Steps vs Total Minutes Asleep



Avg Total Distance Vs Avg Very Active Minutes



Division Of Activity Level



Avg Very Active Minutes	20.93
Avg Fairly Active Minutes	14.60
Avg Lightly Active Minutes	185.5
Avg Sedentary Minutes	931.2

Step 6: Act

- Key Takeaways:

- Eighty-one percent of the minutes that users spend each day are spent in sedentary activities. On average, users are sedentary for 12 hours a day, lightly active for 4 hours, and fairly active or very active for just 30 minutes.
- The users who go through their day with the highest levels of activity sleep between 375 to 450 minutes per night.
- Bellabeat's marketing team can utilize the patterns to better understand how customers from similar companies use their goods, which may have a direct effect on the company's marketing approach.

- Bellabeat Moving Forward:

- Enhance Bellabeat products to remind users of the best times to go to bed, with a target of around 415 minutes on average per night, based on these trends in order to be the most active.
- Enhance Bellabeat products to focus on very active minutes.
- Make gathering thorough user data, such as age, gender, body fat percentage, and demographics, a top priority in order to enhance the accuracy of analysis on devices.
- After a long stretch of sedentary minutes, the Bellabeat device can also vibrate to remind the user to get up and start moving!