Bellabeat Case Study Project

Introduction

Bellabeat is a high-tech manufacturer of health-focused products for women. It offers various smart devices that track activity, sleep, stress, and reproductive health data to empower women with knowledge about their health and habits.

Bellabeat is currently a successful small company, but it is looking for opportunities to expand its market share. This project embarks on a comprehensive market analysis, employing the data analysis process—Ask, Prepare, Process, Analyze, Share, and Act—to unravel consumer behaviors.

The main objective of this project is to gain insights into consumer usage patterns of smart devices and facilitate informed strategies for market expansion.

Ask Phase

Business Task

For this project, the stakeholders asked us to analyze customer behavior for non-Bellabeat devices and select one Bellabeat product to apply these insights to.

Stakeholders

- Urška Sršen - Bellabeat cofounder and Chief Creative Officer
- Sando Mur - Bellabeat cofounder and key member of Bellabeat executive team
- Bellabeat Marketing Analytics team

Bellabeat product I chose:

- Bellabeat app: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions.

Prepare Phase

We will analyze consumer behaviors using FitBit Fitness Tracker Data, a public dataset accessible on Kaggle through MÖBIUS. This dataset was generated by 30 consented and eligible Fitbit users, tracking their physical activity, heart rate, and sleep monitoring to explore users' habits. The information was collected by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016 and 05.12.2016.

By verifying the metadata, we can confirm it is open-source. The owner has dedicated the work to the public domain by waiving all their rights to the work worldwide under copyright law.

The data presents a potential challenge, arising from its limited sample size of 30 users and the absence of demographic and gender information, which could lead to sampling bias.

Process Phase

We will focus the analysis on SQL and later create visualizations in Tableau to share the results with stakeholders.

Data Selection

For the purpose of analyzing consumer behaviour with the health app, we will focus on the following tables:

- Daily Activity– Records IDs, dates, steps, distances, active time and calories
- Daily Sleep– Records IDs, datetime, total minutes asleep and total time in bed
- Hourly Steps– Records IDs, datetime, hourly step number

We select these tables because they offer relatively comprehensive information for our analysis.

Data Preview

Id	Activity Date	Total Steps	TotalDistance	Tracker Distance	LoggedActivities Distance	VeryActive Distance	ModeratelyActive Distance	LightActive Distance	SedentaryActive Distance	VeryActive Minutes	FairlyActive Minutes	LightlyActive Minutes	Sedentary Minutes	Calories
1503960366	4/12/201	13162	8.5	8.5	0	1.88	0.550000012	6.05999994	0	25	13	328	728	1985
1503960366	4/13/201	10735	6.97	6.97	0	1.5700001	0.689999998	4.71000004	0	21	19	217	776	1797
1503960366	4/14/201	10460	6.74	6.74	0	2.4400001	0.400000006	3.91000009	0	30	11	181	1218	1776
1503960366	4/15/201	9762	6.28	6.28	0	2.1400001	1.25999999	2.82999992	0	29	34	209	726	1745
1503960366	4/16/201	12669	8.16	8.16	0	2.71	0.409999996	5.03999996	0	36	10	221	773	1863
1503960366	4/17/201	9705	6.48	6.48	0	3.1900001	0.779999971	2.50999999	0	38	20	164	539	1728
1503960366	4/18/201	13019	8.59	8.59	0	3.25	0.639999986	4.71000004	0	42	16	233	1149	1921
1503960366	4/19/201	15506	9.88	9.88	0	3.53	1.320000052	5.03000021	0	50	31	264	775	2035
1503960366	4/20/201	10544	6.68	6.68	0	1.96	0.479999989	4.23999977	0	28	12	205	818	1786

Figure 1: Daily Activity Preview

Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed
1503960366	4/12/2016 12:00:00 AM	1	327	346
1503960366	4/13/2016 12:00:00 AM	2	384	407
1503960366	4/15/2016 12:00:00 AM	1	412	442
1503960366	4/16/2016 12:00:00 AM	2	340	367
1503960366	4/17/2016 12:00:00 AM	1	700	712
1503960366	4/19/2016 12:00:00 AM	1	304	320
1503960366	4/20/2016 12:00:00 AM	1	360	377
1503960366	4/21/2016 12:00:00 AM	1	325	364
1503960366	4/23/2016 12:00:00 AM	1	361	384

Figure 2: Daily Sleep Preview

Id	ActivityHour	StepTotal
1503960366	4/12/2016 12:00:00 AM	373
1503960366	4/12/2016 1:00:00 AM	160
1503960366	4/12/2016 2:00:00 AM	151
1503960366	4/12/2016 3:00:00 AM	0
1503960366	4/12/2016 4:00:00 AM	0
1503960366	4/12/2016 5:00:00 AM	0
1503960366	4/12/2016 6:00:00 AM	0
1503960366	4/12/2016 7:00:00 AM	0
1503960366	4/12/2016 8:00:00 AM	250

Figure 3: Hourly Step Preview

Cleaning and Formatting

Before we start to analyze the data, we will process it to look for any errors and inconsistencies.

To start with, we identify number of distinct users:

- 33 for Daily Activity dataset
- 24 for Daily Sleep dataset
- 33 for Hourly Steps dataset

Next, we check if there are duplicate entries with the same IDs and dates:

- 0 duplicate for Daily Activity dataset
- 3 duplicates for Daily Sleep dataset – Remove these entries
- 0 duplicate for Hourly Steps dataset

We notice that Daily Activity table records dates in *date* format while Daily Sleep table records dates in *datetime* format. In order to successfully join the 2 tables later, we extract the date information from 'SleepDay' column and store this in a new column as 'SleepDayConverted'.

Now that Daily Activity and Daily Sleep share consistent ID and date information, we perform a full join based on these common identifiers. However, for a more straightforward and conclusive dataset, we extract the key metrics for our analysis – daily steps, calorie consumption and sleep data – and calculate their average values over a 31-day period. These values are then combined and stored in a single table for easier analysis.

	Id	AverageSteps	AverageCalories	AverageSleep
1	1644430081	7282	2811	294
2	7086361926	9371	2566	453
3	4702921684	8572	2965	417
4	3977333714	10984	1513	293
5	6290855005	5649	2599	NULL

Figure 4: Average Value Table Preview

Analyze Phase

Now that the data is structured, we start to analyze trends of the Fitbit users and determine if it can help us on Bellabeat's strategy.

Types of users based on activity level

First, we are going to analyze the average steps per day. Upon review, it's evident that the average step count varies significantly, ranging from 916 steps per day to 16,040 steps per day.

In accordance with the guidelines by MedicineNet, we are going to categorize participants by their daily steps:

- **Sedentary:** Less than 5,000 steps daily
- **Low active:** About 5,000 to 7,499 steps daily
- **Somewhat active:** About 7,500 to 9,999 steps daily
- **Active:** More than 10,000 steps daily

We store the resulting table in Excel and create a pivot table:

Activity_Type	Number
Sedentary	8
Low Active	9
Somewhat Active	9
Active	7
Grand Total	**33**

Figure 5: User Categorization Based On Activity Level

The distribution is quite even, with most people being ‘low active’ and ‘somewhat active’.

Calorie consumption

Average calorie consumption spans from 1483 to 3436 per day. Due to the absence of standardized calorie consumption guideline and gender information, we won’t classify participants based on calorie consumptions.

Types of users based on sleep time

Following the guidance provided by the National Institutes of Health (NIH), we'll categorize users based on their sleep duration:

- **Insufficient sleeper:** Sleeps less than 7 hours a night
- **Sufficient sleeper:** Sleeps more than 7 hours a night

Again, we store the resulting table in Excel and create a pivot table:

Sleep_Type	Number
Insufficient Sleeper	13
Sufficient Sleeper	11
Grand Total	**24**

Figure 6: User Categorization Based On Sleep Duration

The distribution remains relatively even, with 2 more people experiencing insufficient sleep.

Time people spend on bed before falling asleep

We will extract this information by deducting ‘Total Minutes Asleep’ from ‘Total Time In Bed’. Then, we will calculate the average over a 31-day period.

	Id	AverageInBed
1	1503960366	22
2	1644430081	52
3	1844505072	309
4	1927972279	20
5	2026352035	31

Figure 6: Average Time To Fall Asleep Preview

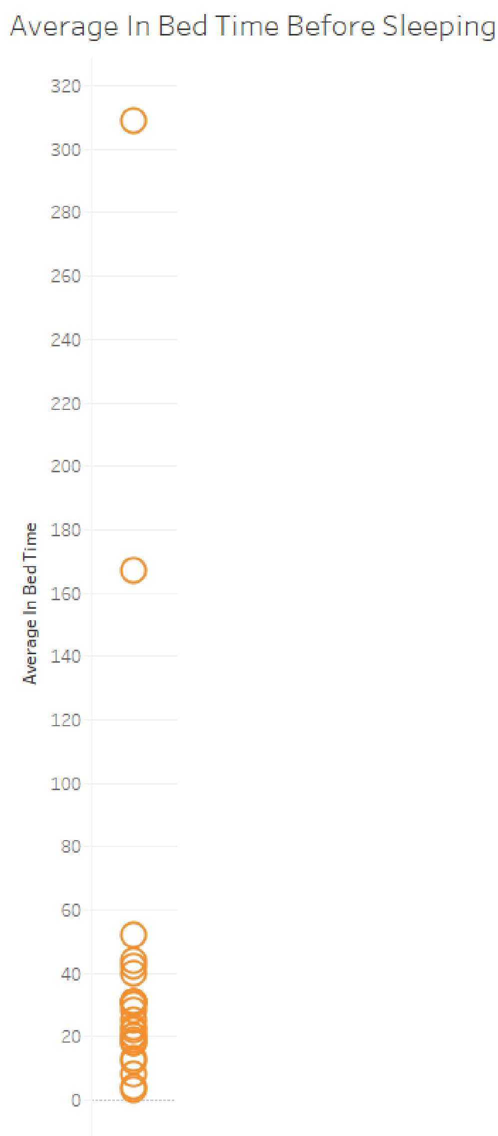


Figure 7: Average Time To Fall Asleep Visualization

Through visualization, it's evident that the majority of users typically fall asleep within 0-50 minutes after getting into bed.

Compare user types by joining activity type and sleep type

Having examined both activity levels and sleep patterns separately, our next step is to merge the resulting tables to analyze these two metrics collectively.

	Id	Active_type	Sleep_type
1	1644430081	LowActive	InsufficientSleep
2	7086361926	SomewhatActive	SufficientSleep
3	4702921684	SomewhatActive	InsufficientSleep
4	3977333714	Active	InsufficientSleep
5	2026352035	LowActive	SufficientSleep

Figure 8: Combined Activity and Sleep Analysis Preview

Upon consideration, we perform an inner join, which means only ID appears on both tables will be included. By doing this, we remove any participants with a null value in sleep type, ensuring a more concise and manageable table for analysis.

We then delve into exploring this merged table, starting by filtering active users to observe their sleep patterns.

	Id	Active_type	Sleep_Type
1	3977333714	Active	InsufficientSleep
2	8053475328	Active	InsufficientSleep
3	7007744171	Active	InsufficientSleep
4	1503960366	Active	InsufficientSleep
5	4388161847	Active	InsufficientSleep

Figure 9: Sleep Pattern For Active Users

Surprisingly, all active users exhibit insufficient sleep durations.

Moving forward, we are now curious what are the main active levels among participants with insufficient and sufficient sleep periods.

	Id	Active_type	Sleep_type
1	3977333714	Active	InsufficientSleep
2	8053475328	Active	InsufficientSleep
3	7007744171	Active	InsufficientSleep
4	1503960366	Active	InsufficientSleep
5	4388161847	Active	InsufficientSleep
6	1644430081	LowActive	InsufficientSleep
7	4020332650	Sedentary	InsufficientSleep
8	4445114986	Sedentary	InsufficientSleep
9	6775888955	Sedentary	InsufficientSleep
10	1927972279	Sedentary	InsufficientSleep
11	2320127002	Sedentary	InsufficientSleep
12	4558609924	SomewhatActive	InsufficientSleep
13	4702921684	SomewhatActive	InsufficientSleep

	Id	Active_type	Sleep_type
1	6117666160	LowActive	SufficientSleep
2	2026352035	LowActive	SufficientSleep
3	4319703577	LowActive	SufficientSleep
4	8792009665	Sedentary	SufficientSleep
5	1844505072	Sedentary	SufficientSleep
6	5553957443	SomewhatActive	SufficientSleep
7	2347167796	SomewhatActive	SufficientSleep
8	6962181067	SomewhatActive	SufficientSleep
9	8378563200	SomewhatActive	SufficientSleep
10	5577150313	SomewhatActive	SufficientSleep
11	7086361926	SomewhatActive	SufficientSleep

Figure 10: Active Levels for Different Sleep Time

The analysis indicates that insufficient sleep is predominantly associated with 'active' and 'sedentary' user categories, while sufficient sleep is notably linked to 'somewhat active' users.

Activity Peaks in a Day

Now, we want to identify the peak hours of user activity by analyzing the hourly step table. To start with, we extract the hour from 'ActivityHour' column and store this value as 'StepHour. We then aggregate the step count based on 'StepHour' and calculate the average steps over a 31-day period.

	StepHour	AverageStepPerHour
1	00:00:00.0000000	42
2	01:00:00.0000000	23
3	02:00:00.0000000	17
4	03:00:00.0000000	6
5	04:00:00.0000000	12

Figure 11: Average Step By Hour Preview

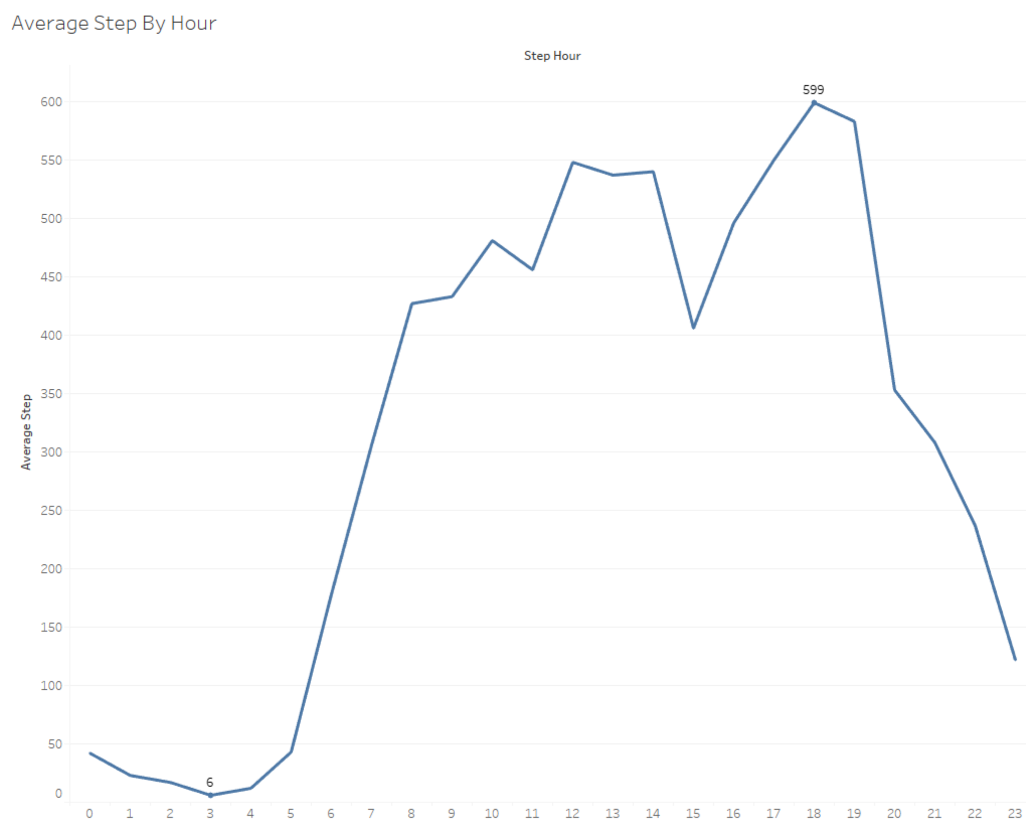


Figure 12: Average Step by Hour Visualization

From this graph we can see that average step per hour ranges from 0 to 600 in a day, with the peak observed at 6 p.m. This pattern aligns with typical daily routines observed among individuals.

Analyze using pattern – Monthly Usage

After examing customer activity level and sleep time, we now focus on understanding the Fitbit usage patterns to gain insights for Bellabeat's app. To achieve this, we aim to assess the user frequency within a month. I plan to classify users monthly utilization with the following criteria:

- **Infrequent user:** Use Fitbit less than 10 days a month
- **Moderate user:** Use Fitbit 10-20 days a month
- **Frequent user:** Use Fitbit more than 20 days a month

Monthly_Frequency	Number of users
Infrequent User	1
Moderate User	3
Frequent User	29
Grand Total	**33**

Figure 12: User Monthly Frequency Summary

The pivot table clearly shows that the majority of participants fall into the 'Frequent user' category.

Analyze using pattern – Daily Usage

After gaining an overview of monthly usage frequency, we are now interested in daily usage frequency. We first compute daily used minutes by aggregating 'Very Active Minutes', 'Fairly Active Minutes', 'Lightly Active Minutes' and 'Sedentary Minutes' from the Daily Activity table. Subsequently, I classify daily user utilization using the following criteria:

- **Less than half day:** Use Fitbit less than 12 hours a day
- **More than half day:** Use Fitbit 12-24 hours a day
- **All day:** Use Fitbit 24 hours a day

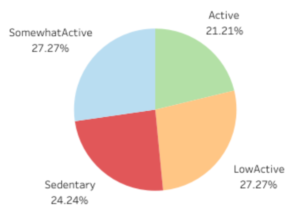
Daily_Frequency	Total Number
Less Than Half Day	25
More Than Half Day	437
All Day	478
Grand Total	**940**

Figure 12: User Daily Frequency Summary

After analyzing the recorded user data, it's evident that approximately half of the participants utilize Fitbit for the entire day, while the other half use it for more than half a day.

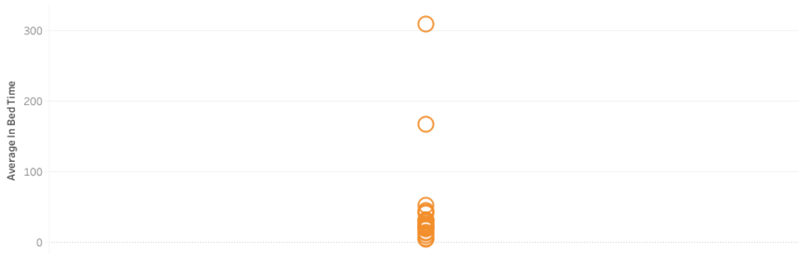
Share Phase

Active Type



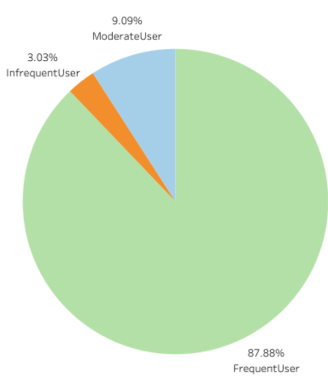
Sleeper Type

Average In Bed Time Before Sleeping

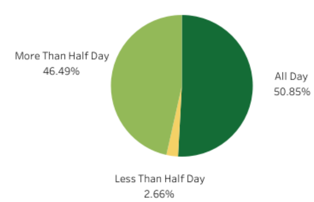


Average Step By Hour

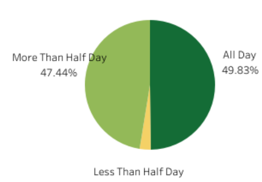
Monthly Usage Frequency



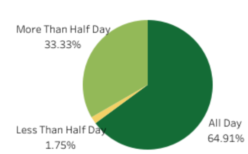
Daily Usage Overall



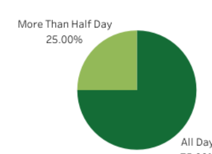
Daily Usage
Among Frequent Users



Daily Usage
Among Moderate Users



Daily Usage
Among Infrequent Users



Act Phase

Final Conclusion

In this project, we conducted a thorough analysis regarding Fitbit user behavior. We started by categorizing customers by their total steps per day and noticed an even distribution among different active levels, where 21% participants were classified as 'active'. Then, we focused our analysis on sleep time and divided customers based on whether they achieved 7 hours of sleep per day. The result showed the average sleep time among users was 6 hours per day, and over half of the users experienced insufficient sleep time. Additionally, the majority of participants needed 0-50 minutes to fall asleep.

To examine the interaction between active level and sleep time, we combined the dataset and observed the merged table. To our surprise, all of the 'active' participants slept less than 7 hours a day. Moreover, we noticed that most people with inadequate sleep time were 'active' and 'sedentary' users, while most people with adequate sleep time were 'low active' and 'somewhat active' users.

Our next step was to analyze monthly and daily using patterns. We were glad to see that 88% participants used Fitbit more than 20 days a month and 51% participants used Fitbit all day, indicating high utilizing frequency.

Finally, a dashboard was created for visual representation of the data.

Consideration

Considering the small sample size (30 users), I would advise Bellabeat to use internal data with a larger user base and additional information, such as heart rate and stress level, to construct a more comprehensive and accurate analysis. In addition, since Bellabeat is a female-focused organization, I suggest collecting data exclusively from women.

Apply insights to the Bellabeat app

Based on our findings, I would suggest Bellabeat do the following:

1. Design a daily dashboard interface
   It would be distracting for users to keep up with all the data tracked by the app; thus creating a dashboard interface that highlights where users aren't meeting the desired standards would be convenient.

2. AI-powered health coaching
   After highlighting where users' lifestyles could improve, Bellabeat could implement a coaching system that offers specialized health and fitness recommendations based on user behaviors and preferences.

3. Gamification
   Gaming and rewards systems could make it entertaining to start and maintain a healthy lifestyle. Bellabeat could offer games where users level up as their exercise level increases or provide rewards for completing challenges, such as step challenges, or maintaining healthy habits. These rewards can be redeemed for exclusive events or customized user interface.

4. Add sleep-related functions
   One way to promote a healthier sleep routine is by setting up reminders through the app to encourage the users to head to bed at a specific time. Moreover, Bellabeat can provide sleep coaching or ambient sounds that improve users' sleeping experience.

References

1. Pallavi Suyog Uttekar, M. (2021, July 15). How many steps a day is considered active?. MedicineNet. https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm

2. U.S. Department of Health and Human Services. (n.d.). How much sleep is enough?. National Heart Lung and Blood Institute. https://www.nhlbi.nih.gov/health/sleep/how-much sleep#:~:text=Experts%20recommend%20that%20adults%20sleep,or%20more%20hours%20a%20night.