Case Study: How Can a Wellness Technology Company Play It Smart?

Background

Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, co-founder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company.

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Sršen knows that an analysis of Bellabeat's available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

You have been asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights you discover will then help guide marketing strategy for the company. You will present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat's marketing strategy.

Step 1: Defining Business Task

Business Task

- **How consumers are using their smart devices.**

Deliverables

A report with the following deliverables:

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of analysis
5. Supporting visualizations and key findings
6. Top high-level content recommendations based on your analysis

Stakeholders

Primary

- ***Urška Sršen:*** Bellabeat's co-founder and Chief Creative Officer
- ***Sando Mur:*** Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team

Secondary

- ***Bellabeat marketing analytics team:*** A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat"s mission and business goals — as well as how you, as a junior data analyst, can help Bellabeat achieve them.

Step 2: Data-set Integrity Check and Cleaning Preparation

Tools Used

- DB Browser for SQLite (SQL) - For exploring, processing and cleaning of data (Some of the files contained rows more than Excel can handle)
- Tableau - Analysis and visualization

Data & Source Integrity

Reliability

Data used for this project is from a fitness tracking device under brand Fitbit open/public data which can be found in this link. The data is about fitness data from 30 individuals wearing Fitbit fitness tracker. Since there is no information regarding participants' age, sex and profession, accuracy and unbiased stand can not be vetted.

Originality

Data set contains 30 user fitness tracking data down to the minute. This data was collected by Amazon Mechanical Turk between 03.12.2016 - 05.12.2016. This data has been preprocessed and uploaded by Kaggle user Mobius. Data set is under Creative Commons license which claims no copyright of the data and allows copy, modification and performing work, even for commercial purposes, all without asking permission. Details of the license can be found here

Since data has been collected from third party and was preprocessed as well, originality can not be vetted as well.

Comprehensiveness

30 eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. However, there are some discrepancies in the data set:

- There is no information about the age, sex, height and profession of the personnel providing fitness logs. This makes identifying pattern of activity of individual
- dailyActivities_merged.csv contains distance data but there is no information regarding distance measuring unit.

- There are inconsistencies in logged data. Not all participants have full 30 days of data. Moreover, Since this data spreads for both May and April months, there are participants with 31 days of data. This makes the time-frame for data analysis inconsistent.
- Some of the data contains 1440 mins of Sedentary minutes which is equivalent to hour. This indicates not wearing the tracker throughout the day.
- There are no information regarding what intensity is and how to measure and unit of intensity in the hourlyIntensities_merged.csv

Current

This data was collected by Amazon Mechanical Turk between 03.12.2016 - 05.12.2016. This indicates that data is quite outdated. This decreases usability of the data.

Cited

Digital Object Identifier (DOI) is 10.5281/zenodo.53894 which resolves to this

Data-set Observations

- The Fitbit data set contains 18 files containing different tracking data. These are:
    1. dailyActivity_merged.csv
    2. dailyCalories_merged.csv
    3. dailyIntensities_merged.csv
    4. dailySteps_merged.csv
    5. heartrate_seconds_merged.csv
    6. hourlyCalories_merged.csv
    7. hourlyIntensities_merged.csv
    8. hourlySteps_merged.csv
    9. minuteCaloriesNarrow_merged.csv
    10. minuteCaloriesWide_merged.csv
    11. minuteIntensitiesNarrow_merged.csv
    12. minuteIntensitiesWide_merged.csv
    13. minuteMETsNarrow_merged.csv
    14. minuteSleep_merged.csv
    15. minuteStepsNarrow_merged.csv
    16. minuteStepsWide_merged.csv
    17. sleepDay_merged.csv
    18. weightLogInfo_merged.csv
- dailyActivity_merged.csv contains all the column data from dailyCalories.csv, dailyIntensities.csv, dailySteps.csv.
- Some of the data are in both long and wide format. These are calorie, heart rate and steps data.
- heartrate_seconds_merged.csv contains heart rate of an individual for 24 hrs at 15 seconds interval.
- ***Summary of logged data:***

Logged Data	Second-wise	Minute-wise	Hour-wise	Day-wise
Calories	❌	✔	✔	✔
Intensity	✔	✔	✔	✔
Heart rate	✔	✔	✔	✔
Sleep	❌	❌	❌	✔
Weight	❌	❌	❌	Inadequate

Step 3: Cleaning and Processing Data

The data set has been marked as pre-processed. There is less chance of blanks or dirty data. However, in order to clean obtained data following cleaning activities have been performed:

- The date format of the daily and hourly data logs were modified with excel to 'YYYY-MM-DD HH:MM:SS' format for ease of future calculations.
- Checking for blank values - none found

```
-- This has been repeated for all of the field of the table
SELECT id
FROM dailyActivity_merged_date_mod
WHERE id is NULL
```

- Checking for negative values using following query - none found

```
SELECT
    min(TotalDistance), max(TotalDistance),
    min(TotalSteps), max(TotalSteps),
    min(LoggedActivitiesDistance), max(LoggedActivitiesDistance),
    min(SedentaryActiveDistance), max(SedentaryActiveDistance),
    min(SedentaryMinutes), max(SedentaryMinutes),
    min(Calories), max(Calories)
FROM
    dailyActivity_merged_date_mod
```

- Checking for full day sedentary activities in the dailyActivities_merged.csv as this can provide overview of each of the individuals daily activity patterns.

```
SELECT
    id,
    SedentaryMinutes,
    count() AS 'Sedentary Days (=> 1440 mins)'
FROM
    dailyActivity_merged_date_mod
WHERE
    SedentaryMinutes >= 1440
GROUP BY
    id, SedentaryMinutes
```

Result:

ID	Sedentary Days (=> 1440 mins)
1503960366	1
1844505072	9
1927972279	13
4020332650	14
4057192912	1
4319703577	1
4388161847	1
4702921684	1
5577150313	2
6117666160	5
6290855005	4
6775888955	9
7007744171	1
7086361926	1
8253242879	1
8583815059	6
8792009665	9

- Since this data spreads for both May and April months, there are participants with 31 days of data. There are inconsistencies in logged data as not all participants have full 30 days of data as well As a result, weekly cycle has been considered and 4 weeks (28 days) have been taken in consideration for time-frame of analysis since most of the participants' logged data can be taken into analysis.

```
SELECT
    id, count() as days_logged
FROM
    dailyActivity_merged_date_mod
GROUP BY
    id
ORDER BY
    days_logged
```

!s[./screenshots/days_logged.jpg]

```
-- Extracting 28 days of user acitvity data with no full day sedentary.
SELECT
```

```
    d_active.Id,
    d_active.ActivityDate,
    d_active.TotalSteps,
    d_active.TotalDistance,
    d_active.TrackerDistance,
    d_active.LoggedActivitiesDistance,
    d_active.VeryActiveDistance,
    d_active.ModeratelyActiveDistance,
    d_active.LightActiveDistance,
    d_active.SedentaryActiveDistance,
    d_active.VeryActiveMinutes,
    d_active.FairlyActiveMinutes,
    d_active.LightlyActiveMinutes,
    d_active.SedentaryMinutes,
    d_active.Calories
FROM
    dailyActivity_merged as d_active
INNER JOIN
    (SELECT Id
    FROM dailyActivity_merged
    GROUP BY Id
    HAVING count(Id) >= 28) as d4w
ON
    d_active.Id = d4w.Id
WHERE
    d_active.SedentaryMinutes < 1440
```

This query resulted 760 rows of user activity data which was analyzed in Tableau.

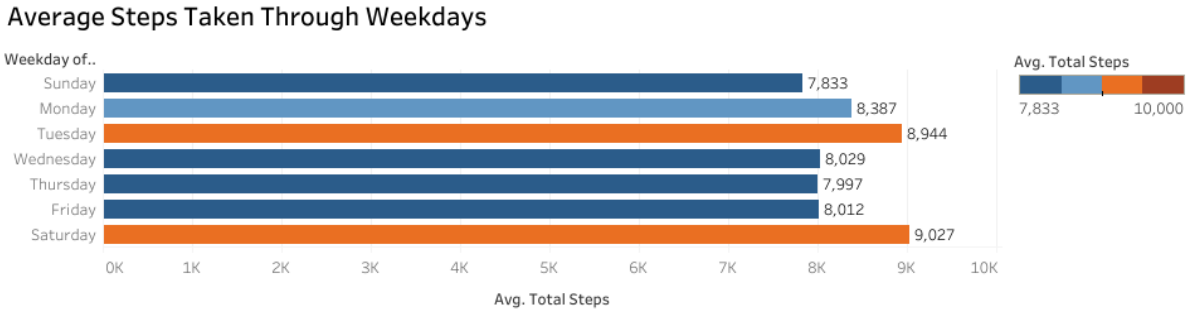
- Along with the above mentioned extracted table, sleepDay_merged.csv has been used for analysis.

Step 4: Analyzing and Visualizing Data

As per available data, following correlations have been analyzed and visualized:

Avg. Steps Taken Weekly

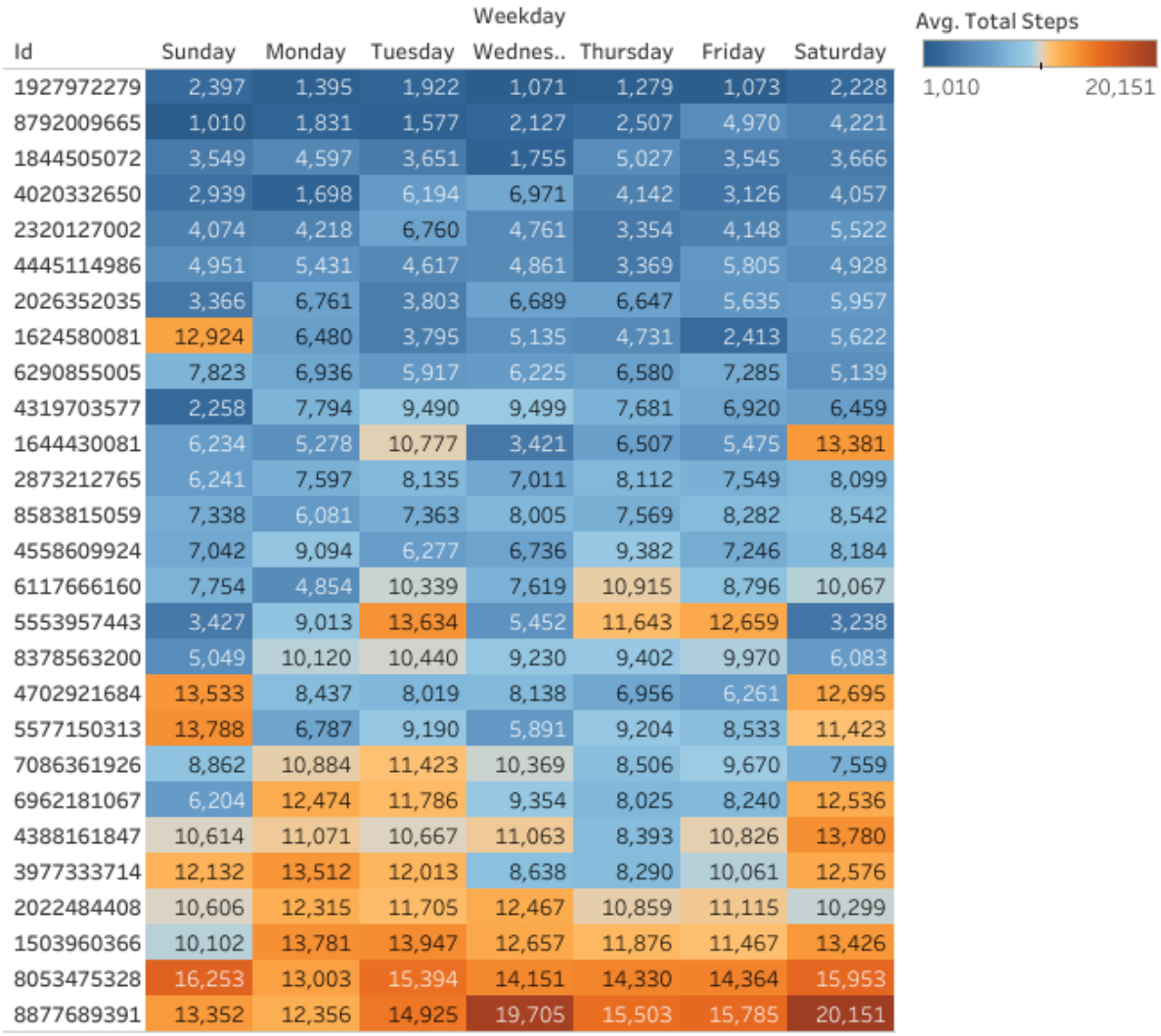
In the first analysis average steps taken by the participants weekly has been observed and found that Sunday being the lowest, 6991 steps, Tuesday and Saturday shares higher number which is above 8000 and other days are roughly around 7500 steps.



Average of Total Steps for each weekday for the month of April and May, 2016. Color shows average of Total Steps. Increased color intensity indicates higher number of steps.

Moreover, breakdown of each participants' average step counts throughout the weekdays have been plotted for better understanding of the movement pattern of the participants.

Avg. Steps Taken by Participants on Weekdays



Id	Sunday	Monday	Tuesday	Wednes..	Thursday	Friday	Saturday
1927972279	2,397	1,395	1,922	1,071	1,279	1,073	2,228
8792009665	1,010	1,831	1,577	2,127	2,507	4,970	4,221
1844505072	3,549	4,597	3,651	1,755	5,027	3,545	3,666
4020332650	2,939	1,698	6,194	6,971	4,142	3,126	4,057
2320127002	4,074	4,218	6,760	4,761	3,354	4,148	5,522
4445114986	4,951	5,431	4,617	4,861	3,369	5,805	4,928
2026352035	3,366	6,761	3,803	6,689	6,647	5,635	5,957
1624580081	12,924	6,480	3,795	5,135	4,731	2,413	5,622
6290855005	7,823	6,936	5,917	6,225	6,580	7,285	5,139
4319703577	2,258	7,794	9,490	9,499	7,681	6,920	6,459
1644430081	6,234	5,278	10,777	3,421	6,507	5,475	13,381
2873212765	6,241	7,597	8,135	7,011	8,112	7,549	8,099
8583815059	7,338	6,081	7,363	8,005	7,569	8,282	8,542
4558609924	7,042	9,094	6,277	6,736	9,382	7,246	8,184
6117666160	7,754	4,854	10,339	7,619	10,915	8,796	10,067
5553957443	3,427	9,013	13,634	5,452	11,643	12,659	3,238
8378563200	5,049	10,120	10,440	9,230	9,402	9,970	6,083
4702921684	13,533	8,437	8,019	8,138	6,956	6,261	12,695
5577150313	13,788	6,787	9,190	5,891	9,204	8,533	11,423
7086361926	8,862	10,884	11,423	10,369	8,506	9,670	7,559
6962181067	6,204	12,474	11,786	9,354	8,025	8,240	12,536
4388161847	10,614	11,071	10,667	11,063	8,393	10,826	13,780
3977333714	12,132	13,512	12,013	8,638	8,290	10,061	12,576
2022484408	10,606	12,315	11,705	12,467	10,859	11,115	10,299
1503960366	10,102	13,781	13,947	12,657	11,876	11,467	13,426
8053475328	16,253	13,003	15,394	14,151	14,330	14,364	15,953
8877689391	13,352	12,356	14,925	19,705	15,503	15,785	20,151

Average of Total Steps (color) taken by the participants for the month of April and May, 2016, broken down by Weekday vs. Id.

Speculation

Participants are not consistent with maintaining their movement. On Sunday (holiday) participants tend to be more sedentary than the other days of the week. Moreover, it has been observed that most of the participants do not meet the 10,000 steps recommendation by CDC (Centers for Disease Control and Prevention) [1].

Participants' Pattern of Being Sedentary

The below mentioned table portrays each participants' sedentary days. This helps to observe which participant is more careful and serious about wearing the fitness tracker and measure sedentary and activity minutes.

Moreover, average sedentary minutes by participants have been plotted in color mapped table in the following picture and it can be observed that almost half of the participants seem to have higher sedentary minutes.

Avg. Sedentary Minutes of Participants

Id	Activity Date: Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
6117666160	613.8	696.0	751.0	761.0	605.3	621.5	616.5	656.3
6962181067	696.3	647.0	683.0	655.2	581.8	714.5	675.3	662.3
5553957443	616.5	686.3	758.4	653.8	627.0	735.0	593.0	668.4
2026352035	770.8	785.5	898.4	594.2	567.4	600.5	611.3	689.4
5577150313	483.0	799.3	687.8	642.0	794.3	890.3	573.3	705.5
3977333714	659.8	814.5	716.4	635.6	749.3	698.8	694.3	707.5
4319703577	873.5	610.8	619.3	853.0	672.6	694.3	637.8	712.3
8378563200	655.8	746.5	743.0	712.4	656.8	774.0	733.5	716.1
4702921684	666.0	874.3	734.0	753.6	679.4	914.8	582.5	744.0
4388161847	615.0	1,006.5	752.3	829.0	879.4	1,014.3	600.8	816.6
1503960366	638.0	899.0	779.6	910.0	923.8	878.0	762.5	828.4
4445114986	972.3	867.8	805.8	807.8	669.8	833.0	904.5	829.9
7086361926	648.7	830.5	817.0	794.0	790.2	914.5	998.0	830.8
8792009665	1,122.0	916.0	744.5	824.5	999.3	796.0	990.7	889.7
4020332650	1,106.5	1,337.7	812.0	913.5	1,024.7	1,077.0	1,239.0	1,070.3
4558609924	1,073.8	1,048.8	1,145.8	1,172.2	954.8	1,146.8	1,115.3	1,093.6
2873212765	1,190.5	1,099.3	1,074.6	1,080.2	1,034.2	1,069.0	1,158.3	1,097.2
1844505072	908.0	1,211.0	1,262.3	1,355.8	1,025.8	1,066.8	937.3	1,111.1
2022484408	1,165.0	1,107.3	1,116.6	1,102.0	1,027.8	1,107.3	1,185.0	1,112.6
8877689391	1,152.0	1,195.5	1,140.0	1,121.2	1,052.6	1,123.8	1,011.3	1,112.9
8053475328	1,167.8	1,200.0	1,182.4	1,093.6	1,104.4	1,215.5	1,088.3	1,148.0
6290855005	1,149.5	1,080.3	1,186.7	1,148.3	1,154.3	1,128.3	1,229.5	1,153.5
1644430081	1,115.5	1,097.3	1,152.8	1,199.4	1,256.3	1,216.0	1,088.8	1,161.9
2320127002	1,255.0	1,263.5	1,192.0	1,227.4	1,190.2	1,250.0	1,175.3	1,220.1
8583815059	1,227.3	1,287.7	1,213.5	1,208.3	1,278.5	1,216.0	1,144.7	1,225.8
1927972279	1,356.5	1,349.3	1,068.3	1,169.5	1,141.7	1,234.3	1,343.5	1,228.9
1624580081	1,201.3	1,265.5	1,307.6	1,292.8	1,198.2	1,303.3	1,229.3	1,257.7

Avg. Sedentary Minutes: 480.0 – 1,356.5

Average of Sedentary Minutes (color) broken down by Weekday vs. Participant ID. Considering 8 hrs (480 mins) of sleep as common lower threshold sedentary minutes, half of the participants spend most of the days without mentionable activity throughout the week.

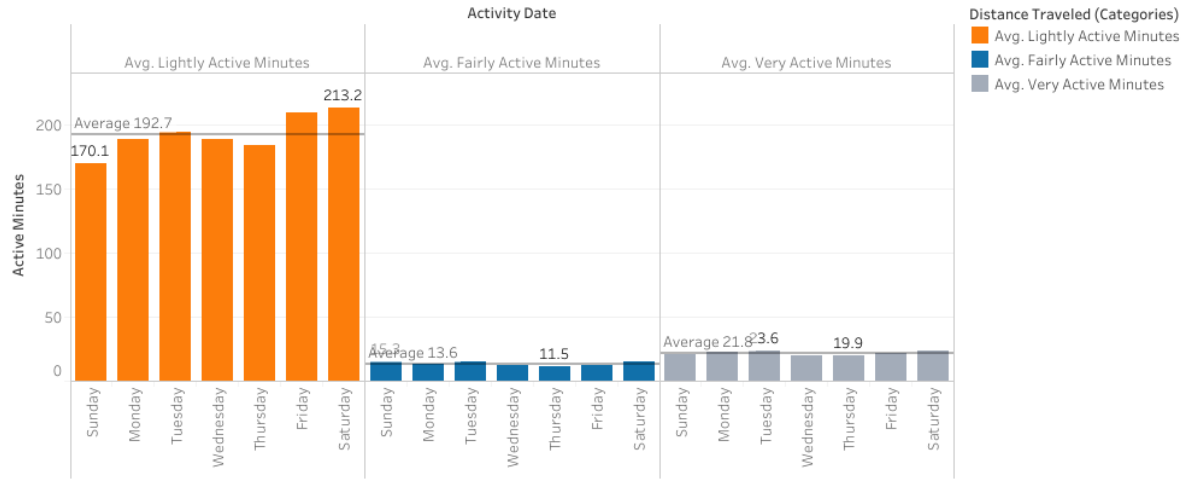
Speculation

1. Half of the participants spend most of the day without any significant movement or exercise which has been marked in Red colored shades.
2. Some of the participants have high number of sedentary days, indicating fitness tracker not being worn. Similar incident might have happened for the dark red shades of the color mapped table above which might be the result of tracker not being worn for some part of the day.

Time Spent in Different Activity Categories

Time spent in different activity types has been visualized by weekdays. This shows that most of the time spent by the participants is in the lightly active category averaging 192.7 minutes. Although average very active time (21.8 minutes) is slightly higher than the fairly active time (13.6 minutes) spent throughout the weekdays, these values are far behind the lightly active minutes.

Active Time



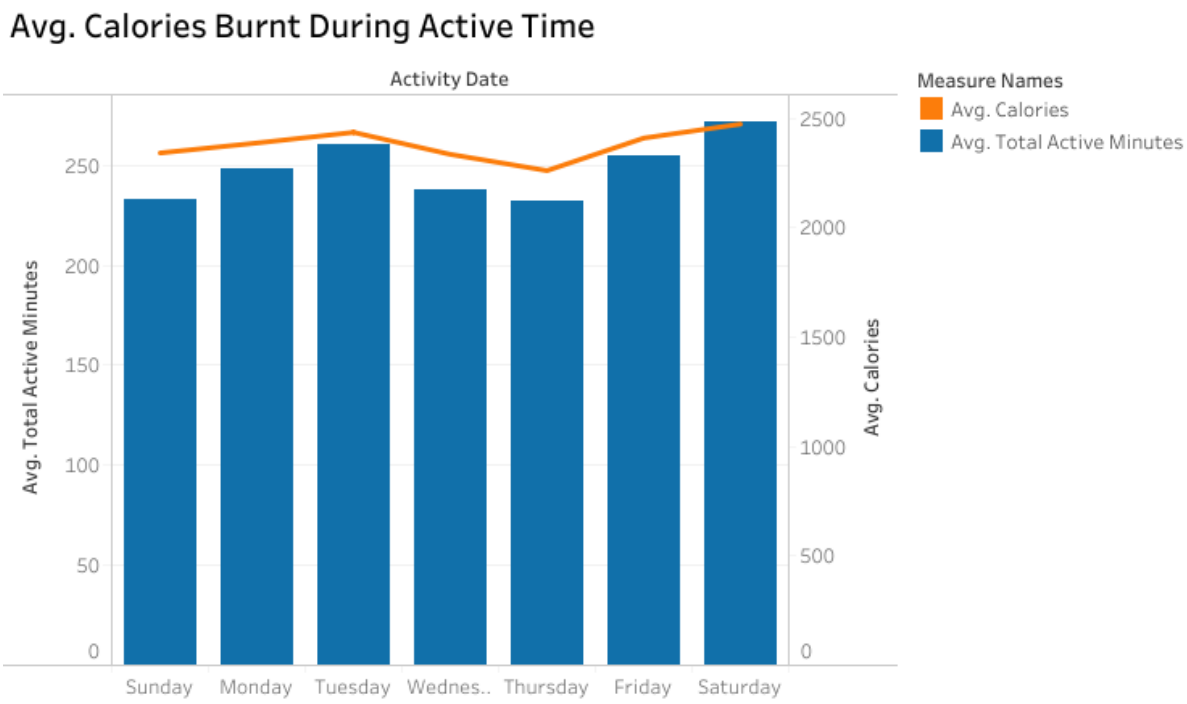
Types of activity minutes throughout the week by the Fitbit users, measured by type of activities for the month of April and May, 2016.

Speculation

Light activity includes regular daily activities such as walking slowly, sitting and using computer, light standing work i.e. cooking, washing dishes etc., fishing sitting, playing instruments etc[2]. Above visualization means that most of the participants are not very active and conscious about being active and keeping healthy lifestyle.

Calories Burnt During Active Time

Average calories burnt during active time on weekdays has been measured and visualized. Here, summation of lightly active, moderately active and very active time has been considered as "Total Active Minutes". Average of Total Active Minutes and Average of Calories Burnt throughout the weekdays have been plotted.



The trends of Avg. of Total Active Minutes and Avg. Calories for weekdays have been shown. Blue colored bars indicate avg. of total active minutes and orange line indicates avg. calories burnt on that specific weekday.

Moreover, the average calories burnt by the individual participants has been analyzed and it has been observed that half of the participants have calories burnt below 2000 mark.

Avg. Calories Burnt by Participants

	Weekday							
Id	Sunday	Monday	Tuesday	Wednes..	Thursday	Friday	Saturday	Grand To..
1624580081	1,822	1,480	1,399	1,451	1,362	1,390	1,539	**1,483**
3977333714	1,580	1,645	1,596	1,221	1,464	1,521	1,622	**1,514**
2026352035	1,404	1,639	1,431	1,627	1,548	1,568	1,572	**1,541**
1844505072	1,667	1,814	1,713	1,527	1,688	1,690	1,688	**1,666**
2320127002	1,700	1,701	1,869	1,747	1,537	1,701	1,820	**1,724**
5553957443	1,603	1,996	2,171	1,720	1,857	2,111	1,641	**1,876**
1503960366	1,769	1,939	1,968	1,869	1,852	1,826	1,895	**1,877**
2873212765	1,816	1,895	1,938	1,981	1,820	1,999	1,974	**1,917**
6962181067	1,783	2,125	2,097	1,973	1,764	1,987	2,173	**1,982**
4558609924	2,162	2,158	1,970	1,916	2,061	1,964	2,040	**2,033**
4319703577	1,682	2,140	2,233	2,239	1,850	2,078	2,020	**2,035**
8792009665	1,881	2,015	1,582	2,064	2,154	2,510	2,455	**2,086**
4445114986	2,212	2,234	2,192	2,200	1,976	2,334	2,204	**2,186**
1927972279	2,368	2,239	2,351	2,245	2,039	2,232	2,364	**2,252**
6117666160	2,285	1,770	2,528	2,338	3,138	2,418	2,556	**2,427**
2022484408	2,385	2,606	2,644	2,631	2,446	2,568	2,243	**2,510**
7086361926	2,438	2,802	2,810	2,719	2,329	2,695	2,334	**2,598**
4020332650	2,540	2,329	3,136	3,262	2,187	2,828	3,013	**2,720**
6290855005	2,978	2,609	2,690	2,767	2,774	2,833	2,726	**2,768**
1644430081	2,777	2,723	3,149	2,264	2,734	2,730	3,355	**2,811**
8583815059	2,808	2,677	2,814	2,812	2,837	2,911	2,995	**2,838**
8053475328	3,066	2,822	3,019	2,952	2,806	2,930	3,040	**2,946**
4702921684	3,406	2,982	2,972	2,988	2,598	2,790	3,454	**2,997**
4388161847	3,079	3,181	3,074	3,073	2,712	3,254	3,421	**3,099**
8877689391	3,158	3,116	3,221	3,672	3,083	3,519	4,245	**3,420**
8378563200	2,842	3,844	3,772	3,565	3,408	3,584	2,933	**3,437**
5577150313	4,497	3,126	3,526	2,691	3,656	3,380	3,976	**3,470**

Avg. Calories: 1,221 – 5,894

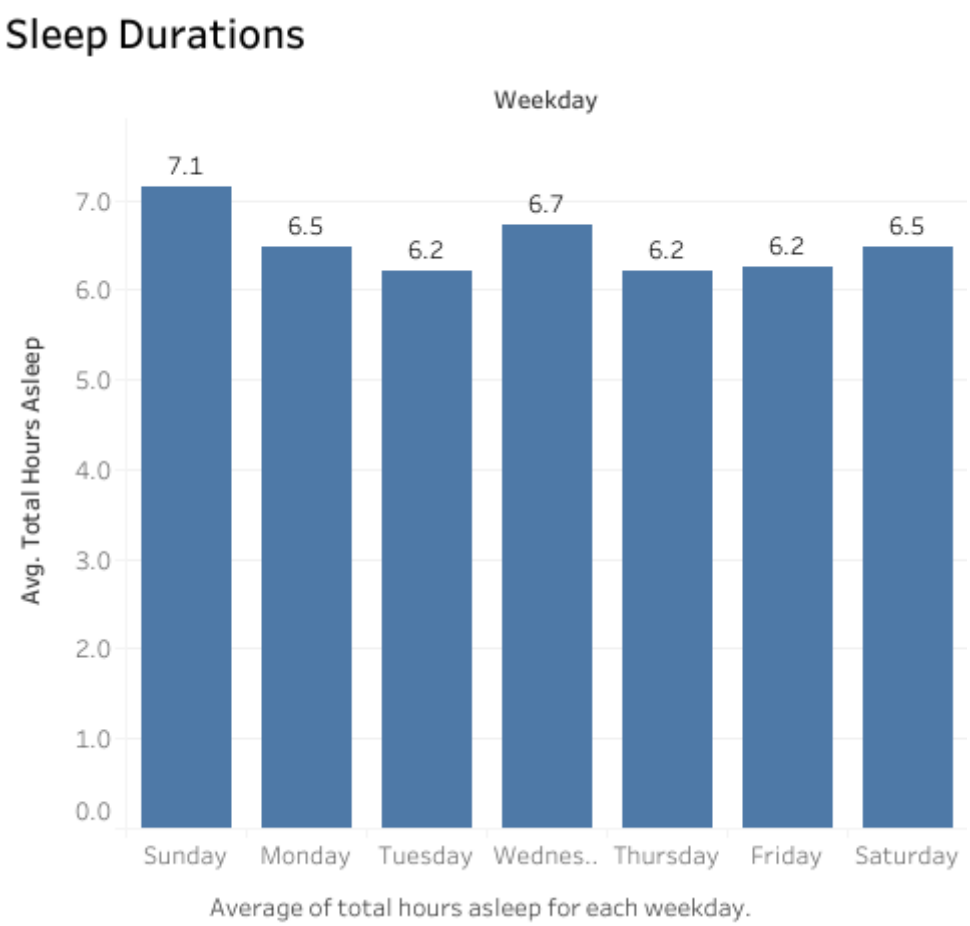
Average of Calories (color) broken down by Weekday vs. Id.

Speculation

Calorie burning requirement differs according to age, body weight, sex and calorie intake. These information are missing in this data set. Only weight information is available but for only 8 IDs and none of these weight logs represent any changes overtime. However, despite all of these missing information, as per general guideline women typically burn about 2000 calories per day and men burn about 2500 calories per day[3]. As per this guideline, the calories burnt throughout the weekdays are within this range.

Sleeping Durations of Participants

Average sleep duration of the participants throughout the weekdays have been measured and visualized in order to get information about average sleeping hours and sleeping patterns of participants.



Speculation

Although there were not age information of the participants, it is recommended that an adult (18-60 years) gets 7 hours and more sleep everyday [4]. In contrast to this information, most of the participants sleep duration is less than 7 hours on weekdays apart from Sunday, which is holiday.

Summary

Although the data set lacked reliability and comprehensiveness due to lack of proper data collection time-line; participants age, gender, sex and profession the data set yielded some usage pattern of the Fitbit fitness tracker users. Since healthy life-style gets down to participants' calorie usage and resting pattern, the activity time, steps taken, calories burnt, sedentary habits and sleep durations have been analyzed and visualized. Firstly, in case of activity it has been observed that most of the participants do not meet the 10,000 steps requirement by CDC. Moreover, half of the participants spend most of the day without any significant movement. However, the calorie burning requirement meets the desired value (men 2500 cal, women: 2000 cal) for half of the participants. This might be indicative of light exercise being done by the participants. On the other hand, considering sedentary condition, 1/3 of the participants have very high value of sedentary minutes. Some of the sedentary values might be indicative of tracker not being worn all the time throughout the day. On account of sleep, average sleeping duration is below 8 hour mark even Sunday being the highest with 7.1 hours.

Conclusion and Recommendation

As per Bellabeat stakeholder recommendation, Fitbit user open data has been analyzed to identify patterns of fitness device usage. Although the data set lacked some information, as per analysis following recommendations can be provided:

1. There has been high number of sedentary minutes observed in the data set. Potential cause of it might be not wearing the fitness tracker for long time in a day. The tracker needs to have feature to identify when it's on the users body or not. Otherwise this can produce wrong information about the user.
2. Users need to be notified about the lacking activity periodically. Users can be emailed most appropriately show as notification on the cellphone screen. In addition, calorie intake tracking can be introduced and thus curated activity plan can be suggested by the application.

References

1. Lifestyle Coach Facilitation Guide: Post-Core: https://www.cdc.gov/diabetes/prevention/pdf/postcurriculum_session8.pdf
2. Examples of Moderate and Vigorous Physical Activity: https://www.hsph.harvard.edu/obesity-prevention-source/moderate-and-vigorous-physical-activity/
3. How Many Calories Should You Burn Daily: https://www.nike.com/a/how-many-calories-should-you-burn-daily
4. How Much Sleep Do I Need?: https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html