BELLABEAT CASE STUDY

How Can a Wellness Technology Company Play It Smart?

Ayomide Onayemi

January, 2024

Contents

PROJECT OVERVIEW 2
STEP 1: ASK 2
1.0 INTRODUCTION 2
1.1 BUSINESS TASK 2
STEP 2: PREPARE 2
2.1 Data Source Used 2
2.2 Is the data ROCCC? 2
STEP 3: PROCESS 3
STEP 4: ANALYZE 4
4.1 Findings from the Analysis 5
STEP 5: SHARE 5
STEP 6: ACT 8

PROJECT OVERVIEW

This case study is the Capstone project for the Google Data Analytics Professional Certificate. The analysis is structured around the utilization of the 6 steps of Data Analysis.

STEP 1: ASK

1.0 INTRODUCTION

Bellabeat, a high-tech manufacturer of health-focused products for women. has significantly impacted women's well-being by gathering data on activity, sleep, stress, and reproductive health. Since its inception in 2013, Bellabeat has undergone remarkable growth, solidifying its position as a technology-driven wellness entity dedicated to women's health.

Key Stakeholders:

- Urška Sršen (Co-founder and Chief Creative Officer)
- Sando Mur (Co-founder and Mathematician)
- Bellabeat Marketing Analytics Team
- Consumers (Bellabeat Customers)

1.1 BUSINESS TASK

The objective is to analyze smart device data to gain insights into customer engagement with the Fitbit app. Leveraging trends identified through this analysis, high-level recommendations will be provided to inform Bellabeat's marketing strategy.

STEP 2: PREPARE

2.1 Data Source Used

- FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius)
- Generated by respondents from a distributed survey via Amazon Mechanical Turk between 12 March 2016 to 12 May 2016.
- Content: Personal fitness tracker data from thirty Fitbit users, including minute-level output for physical activity, heart rate, and sleep monitoring. Information covers daily activity, steps, and heart rate.

2.2 Is the data ROCCC?

ROCCC stands for Reliable, Original, Comprehensive, Current, and Cited.

Reliable - LOW – selection bias, this data was collected 7 years ago and probably does not effectively represent the current socio-economical factors.

Original - LOW – data is sourced through Amazon Mechanical Turk, a third-party platform, introducing potential biases and limitations.

Comprehensive - MEDIUM – while the dataset includes relevant parameters, the sample size of 30 respondents limits its comprehensiveness.

Current - LOW – The data is over 7 years old, potentially rendering the insights outdated and less applicable to the current landscape.

Cited - LOW - The data is collected from a third party, making the exact sources and methodology unclear and potentially introducing uncertainty.

STEP 3: PROCESS

The file below was selected and used for analysis.

dailyActivity_merged.csv

Data cleaning and manipulation with Excel.

1. Download data and familiarize: The screenshot of the table below

	ActivityDate	TotalSteps	TotalDistanc	TrackerDista	LoggedActiv	VeryActiveD	Moderately/	LightActiveD	SedentaryAc	VeryActiveM	FairlyActive	LightlyActive	SedentaryM	Calories
2	4/12/2016	13162	8.5	8.5	0	1.879999995	0.550000012	6.059999943	0	25	13	328	728	198
3	4/12/2016	8163	5.309999943	5.309999943	0	0	0	5.309999943	0	0	0	146	1294	143
4	4/12/2016	10694	7.769999981	7.769999981	0	0.140000001	2.299999952	5.329999924	0	2	51	256	1131	319
5	4/12/2016	6697	4.429999828	4.429999828	0	0	0	4.429999828	0	0	0	339	1101	203
6	4/12/2016	678	0.469999999	0.469999999	0	0	0	0.469999999	0	0	0	55	734	222
7	4/12/2016	11875	8.340000153	8.340000153	0	3.309999943	0.769999981	4.260000229	0	42	14	227	1157	239
8	4/12/2016	4414	2.74000001	2.74000001	0	0.189999998	0.349999994	2.200000048	0	3	8	181	706	145
9	4/12/2016	10725	7.489999771	7.489999771	0	1.169999957	0.310000002	6.010000229	0	13	9	306	1112	212
10	4/12/2016	10113	6.829999924	6.829999924	0	2	0.620000005	4.199999809	0	28	13	320	964	234
11	4/12/2016	8796	5.909999847	5.909999847	0	0.109999999	0.930000007	4.880000114	0	2	21	356	1061	198
12	4/12/2016	4747	3.24000001	3.24000001	0	0	0	3.230000019	0.01	0	0	280	1160	178
13	4/12/2016	8856	5.980000019	5.980000019	0	3.059999943	0.910000026	2.00999999	0	44	19	131	777	145
14	4/12/2016	8539	6.119999886	6.119999886	0	0.150000006	0.239999995	5.679999828	0	4	15	331	712	365
15	4/12/2016	5394	4.03000021	4.03000021	0	0	0	3.940000057	0	0	0	164	1276	228
16	4/12/2016	7753	5.199999809	5.199999809	0	0	0	0	0	0	0	0	1440	211
17	4/12/2016	10122	7.78000021	7.78000021	0	0	0	0	0	0	0	0	1440	295
18	4/12/2016	3276	2.200000048	2.200000048	0	0	0	2.200000048	0	0	0	196	787	211
19	4/12/2016	5135	3.390000105	3.390000105	0	0	0	3.390000105	0	0	0	318	1122	190
20	4/12/2016	7213	5.880000114	5.880000114	0	0	0	5.849999905	0	0	0	263	718	294
21	4/12/2016	11596	7.570000172	7.570000172	0	1.370000005	0.790000021	5.409999847	0	19	13	277	767	202
22	4/12/2016	8135	6.079999924	6.079999924	0	3.599999905	0.379999995	2.099999905	0	86	16	140	728	340
23	4/12/2016	0	0	0	0	0	0	0	0	0	0	0	1440	149

2. Sort Data by "ActivityDate", from oldest to newest

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Id	ActivityDate	TotalSteps	TotalDistanc	TrackerDista	LoggedActiv	VeryActiveD	Moderately/	LightActiveD	SedentaryAc	VeryActiveM	FairlyActive	LightlyActive
2	1503960366	4/12/2016	13162	8.5	8.5	0	1.879999995	0.550000012	6.059999943	0	25	13	32
3	1624580081	4/12/2016	8163	5.309999943	5.309999943	0	0	0	5.309999943	0	0	0	14
4	1644430081	4/12/2016	10694	7.769999981	7.769999981	0	0.140000001	2.299999952	5.329999924	0	2	51	25
5	1844505072	4/12/2016	6697	4.429999828	4.429999828	0	0	0	4.429999828	0	0	0	33
6	1927972279	4/12/2016	678	0.469999999	0.469999999	0	0	0	0.469999999	0	0	0	5
7	2022484408	4/12/2016	11875	8.340000153	8.340000153	0	3.309999943	0.769999981	4.260000229	0	42	14	22
8	2026352035	4/12/2016	4414	2.74000001	2.74000001	0	0.189999998	0.349999994	2.200000048	0	3	8	18
9	2320127002	4/12/2016	10725	7.489999771	7.489999771	0	1.169999957	0.310000002	6.010000229	0	13	9	30
10	2347167796	4/12/2016	10113	6.829999924	6.829999924	0	2	0.620000005	4.199999809	0	28	13	32
11	2873212765	4/12/2016	8796	5.909999847	5.909999847	0	0.109999999	0.930000007	4.880000114	0	2	21	35
12	3372868164	4/12/2016	4747	3.24000001	3.24000001	0	0	0	3.230000019	0.01	0	0	28
13	3977333714	4/12/2016	8856	5.980000019	5.980000019	0	3.059999943	0.910000026	2.00999999	0	44	19	13
14	4020332650	4/12/2016	8539	6.119999886	6.119999886	0	0.150000006	0.239999995	5.679999828	0	4	15	33
15	4057192912	4/12/2016	5394	4.03000021	4.03000021	0	0	0	3.940000057	0	0	0	16
16	4319703577	4/12/2016	7753	5.199999809	5.199999809	0	0	0	0	0	0	0	
17	4388161847	4/12/2016	10122	7.78000021	7.78000021	0	0	0	0	0	0	0	
18	4445114986	4/12/2016	3276	2.200000048	2.200000048	0	0	0	2.200000048	0	0	0	19
19	4558609924	4/12/2016	5135	3.390000105	3.390000105	0	0	0	3.390000105	0	0	0	31
20	4702921684	4/12/2016	7213	5.880000114	5.880000114	0	0	0	5.849999905	0	0	0	26
21	5553957443	4/12/2016	11596	7.570000172	7.570000172	0	1.370000005	0.790000021	5.409999847	0	19	13	27

3. Put data in a table and name "Bellafit"
4. Check for blanks: =IF(SUM(COUNTBLANK(Bellafit))>0, "Blanks Found", "No Blanks")

5. Distinct count of Id: 33 – contrary to the initially stated number of 30.
6. Changed date format: mm/dd/yyyy to dd/mm/yyyy
7. Extract "DayOfTheWeek" from ActivityDate: =TEXT(WEEKDAY([@ActivityDate]),"dddd")

Id	ActivityDate	DayOfTheWeek	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActiveDistance
1503960366	12/04/2016	Tuesday	13162	8.5	8.5	0	1.879999995	0.550000012
1624580081	12/04/2016	Tuesday	8163	5.309999943	5.309999943	0	0	0
1644430081	12/04/2016	Tuesday	10694	7.769999981	7.769999981	0	0.140000001	2.299999952
1844505072	12/04/2016	Tuesday	6697	4.429999828	4.429999828	0	0	0
1927972279	12/04/2016	Tuesday	678	0.469999999	0.469999999	0	0	0
2022484408	12/04/2016	Tuesday	11875	8.340000153	8.340000153	0	3.309999943	0.769999981
2026352035	12/04/2016	Tuesday	4414	2.74000001	2.74000001	0	0.189999998	0.349999994
2320127002	12/04/2016	Tuesday	10725	7.489999771	7.489999771	0	1.169999957	0.310000002
2347167796	12/04/2016	Tuesday	10113	6.829999924	6.829999924	0	2	0.620000005
2873212765	12/04/2016	Tuesday	8796	5.909999847	5.909999847	0	0.109999999	0.930000007
3372868164	12/04/2016	Tuesday	4747	3.24000001	3.24000001	0	0	0
3977333714	12/04/2016	Tuesday	8856	5.980000019	5.980000019	0	3.059999943	0.910000026
4020332650	12/04/2016	Tuesday	8539	6.119999886	6.119999886	0	0.150000006	0.239999995
4057192912	12/04/2016	Tuesday	5394	4.03000021	4.03000021	0	0	0
4319703577	12/04/2016	Tuesday	7753	5.199999809	5.199999809	0	0	0
4388161847	12/04/2016	Tuesday	10122	7.78000021	7.78000021	0	0	0
4445114986	12/04/2016	Tuesday	3276	2.200000048	2.200000048	0	0	0
4558609924	12/04/2016	Tuesday	5135	3.390000105	3.390000105	0	0	0
4702921684	12/04/2016	Tuesday	7213	5.880000114	5.880000114	0	0	0
5553957443	12/04/2016	Tuesday	11596	7.570000172	7.570000172	0	1.370000005	0.790000021

8. Calculate the TotalMinutes: =SUM([@VeryActiveMinutes],[@FairlyActiveMinutes],[@LightlyActiveMinutes],[@SedentaryMinutes])
9. Calculate the TotalHours: =[@TotalMinutes]/60

LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes	TotalMinutes	TotalHours
6.059999943	0	25	13	328	728	1094	18.23333333
5.309999943	0	0	0	146	1294	1440	24
5.329999924	0	2	51	256	1131	1440	24
4.429999828	0	0	0	339	1101	1440	24
0.469999999	0	0	0	55	734	789	13.15
4.260000229	0	42	14	227	1157	1440	24
2.200000048	0	3	8	181	706	898	14.96666667
6.010000229	0	13	9	306	1112	1440	24
4.199999809	0	28	13	320	964	1325	22.08333333
4.880000114	0	2	21	356	1061	1440	24
3.230000019	0.01	0	0	280	1160	1440	24
2.00999999	0	44	19	131	777	971	16.18333333
5.679999828	0	4	15	331	712	1062	17.7
3.940000057	0	0	0	164	1276	1440	24
0	0	0	0	0	1440	1440	24
0	0	0	0	0	1440	1440	24
2.200000048	0	0	0	196	787	983	16.38333333
3.390000105	0	0	0	318	1122	1440	24
5.849999905	0	0	0	263	718	981	16.35
5.409999847	0	19	13	277	767	1076	17.93333333

STEP 4: ANALYZE

The data cleaning and manipulation are done, next is the “analyze” phase. Data aggregation using pivot table.

	A	B	C	D	E	F	G
1	Day of the Week	Frequency	Average of Calories	Average of TotalSteps	Average of TotalDistance	Average of TotalMinutes	Average of TotalHours
2	Sunday	121	2263	6933.231405	5.027190074	1198.743802	19.97906336
3	Monday	120	2324.208333	7780.866667	5.55291666	1257.108333	20.95180556
4	Tuesday	152	2356.013158	8125.006579	5.832236808	1241.993421	20.69989035
5	Wednesday	150	2302.62	7559.373333	5.488333328	1213.213333	20.22022222
6	Thursday	147	2199.571429	7405.836735	5.312244924	1178.782313	19.64637188
7	Friday	126	2331.785714	7448.230159	5.309920621	1236.674603	20.61124339
8	Saturday	124	2354.967742	8152.975806	5.854677423	1208.548387	20.14247312
9	**Grand Total**	**940**	**2303.609574**	**7637.910638**	**5.489702122**	**1218.753191**	**20.31255319**

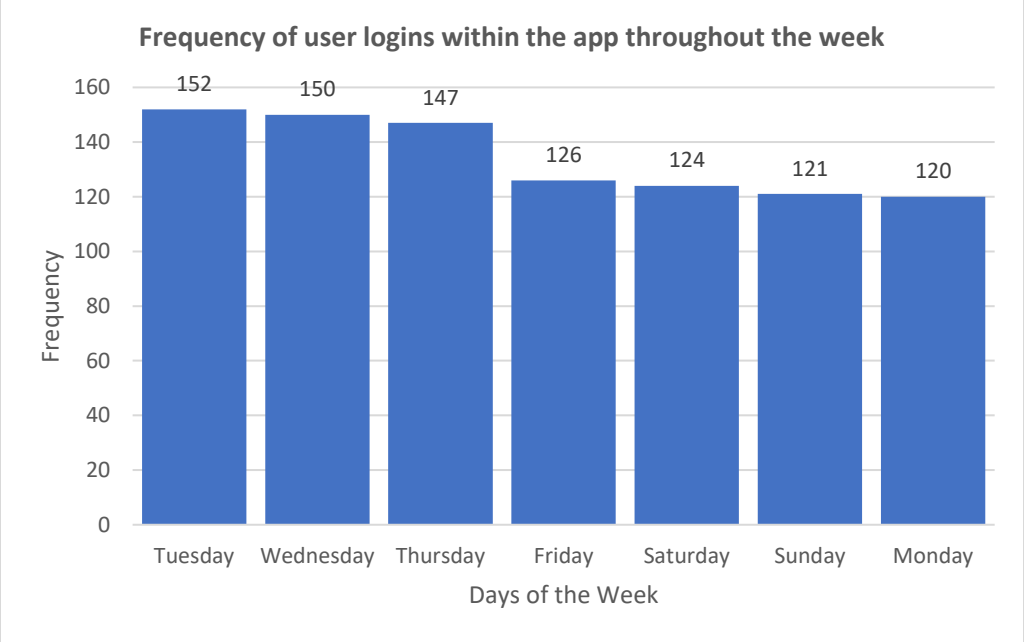
4.1 Findings from the analysis

1. Tuesday is the most active day with the highest recorded number of steps and distance. Additionally, user engagement, indicated by the frequency of logins, reaches its peak on Tuesday.

2. Fitbit users, on average, recorded 7,638 steps and covered a distance of 5.49 km per week. Moreover, the average caloric expenditure amounted to 2,304 calories, with users spending an average of 20.3 hours engaging in activities throughout the week.

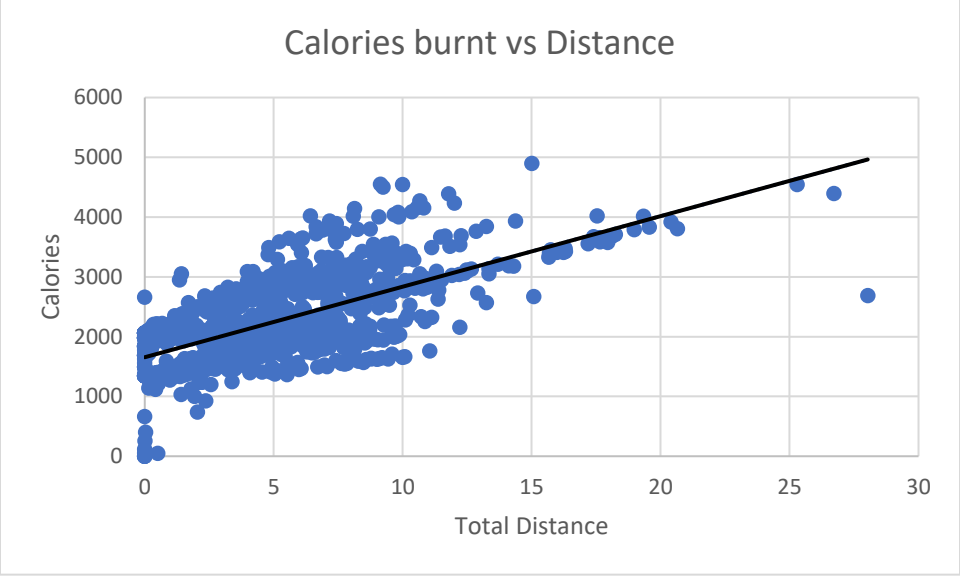
STEP 5: SHARE

The column chart below shows peak user logins on Tuesdays, reflecting heightened engagement, while a significant drop from Friday to Monday suggests reduced app interaction over the weekend and the start of the workweek.



The scatter plot below shows the positive relationship between calories burnt and steps taken, with a correlation coefficient of 0.59. This means that as the number of steps increases, there is a corresponding rise in calories burned, highlighting the direct connection between physical activity and energy expenditure.

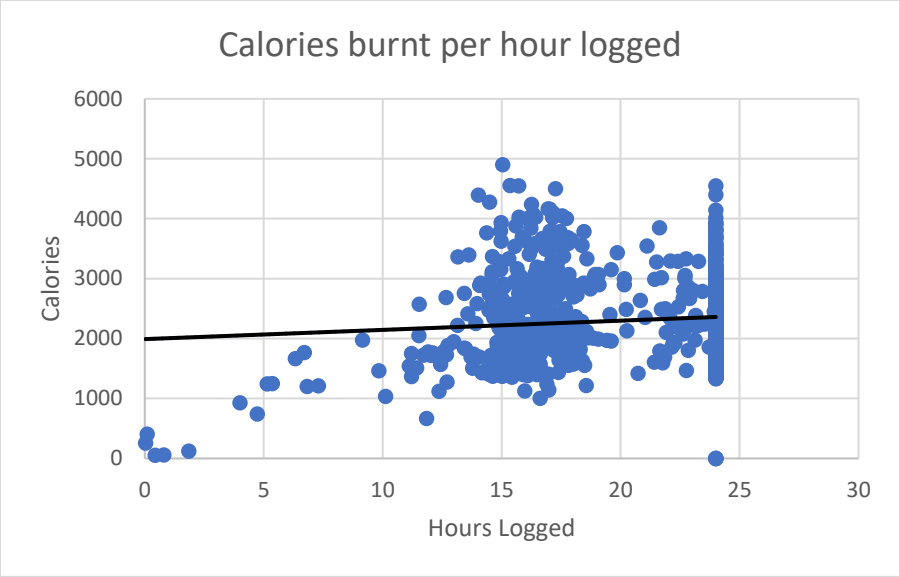
The plot below reveals a positive correlation (correlation coefficient: 0.644962) between calories burned and distance covered. This signifies that users covering greater distances exhibit higher calorie expenditure, emphasizing the direct association between physical activity distance and energy burnt.



The plot below reveals a positive relationship between calories burned and hours logged (correlation coefficient of 0.094951). This signifies a weak positive correlation, implying that there is a slight tendency for users who log more hours to burn more calories.

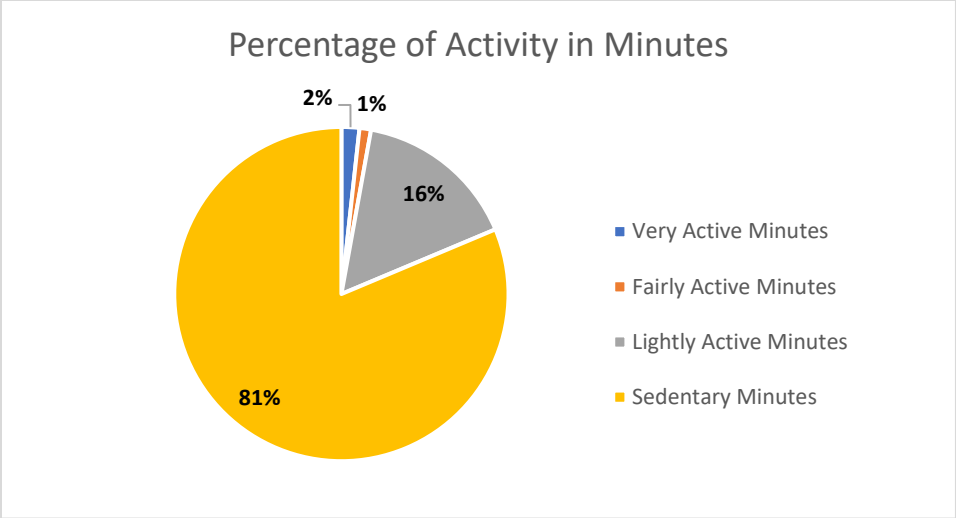
A few outliers are also observed in the plot:

1. Zero value outliers indicate an occurrence where neither calories burnt nor hours logged are recorded.
2. Another outlier is noted at the 24-hour mark with a corresponding zero-calorie burnt. This outlier suggests a prolonged activity duration without any associated calorie expenditure.



The pie chart below provides valuable insights into Fitbits user activity:

- Fitbit users predominantly log sedentary activities, indicating that a significant portion of their time involves low physical activity levels.
- Users dedicate only a small portion of their time to very active and fairly active activities, or they may not participate in physical activities like running or workouts.
- Approximately 16% of users' time is dedicated to light physical activities, showcasing a moderate engagement in activities involving mild exertion.



STEP 6: ACT

The trends identified:

- Notable day-specific peak indicates a recurring pattern of increased physical activity.
- A higher percentage of Fitbit users engage in sedentary activities rather than utilizing the app for fitness tracking.
- Users exhibit higher app engagement on weekdays compared to weekends.

How can these inform Bellabeat marketing strategy?

- Bellabeat can capitalize on the observed weekday peaks by strategically deploying promotions, challenges, or content releases, particularly on Tuesdays.
- Implement notification prompts to encourage users to engage in physical activity and track their activities using the app. Timely reminders can boost user adherence to fitness goals.
- Develop educational content highlighting the importance of active living and the health benefits of regular exercise.