

**Sade Fasan**  
**Google Data Analytics Case Study**



## Table of Contents

<b>Section 1 - Introduction .....</b>	<b>2</b>
About bellabeat .....	2
Company products .....	2
My assignment .....	3
Deliverables .....	3
The data analysis process .....	3
Deliverable 1 - A clear summary of the business task .....	3
<b>Section 2 - Ask questions to make data-driven decisions .....</b>	<b>4</b>
About the data.....	4
Storage .....	4
Format .....	4
Credibility and bias .....	5
Problems with the data .....	5
Assumptions made .....	5
Licensing, privacy, security, and accessibility .....	6
Questions that guided my analysis .....	6
<b>Section 3 - Prepare data for exploration .....</b>	<b>6</b>
<b>Section 4 - Process Data from dirty to clean .....</b>	<b>7</b>
Cleaning data activity merged sheet .....	8
Cleaning sleep day merged sheet.....	9
Cleaning weight log info merged sheet .....	10
<b>Section 5 - Analyse Data to Answer Questions .....</b>	<b>11</b>
Pivot tables and pivot charts .....	11
<b>Section 6 - Share Data Through the Art of Visualization .....</b>	<b>13</b>
Insights from daily activity.....	13
Insights from weight information .....	16
Insights from sleep information .....	17
<b>Section 7 - Scope for further study.....</b>	<b>18</b>
<b>Section 8 - Act to make data-driven decisions.....</b>	<b>19</b>
Products not selected to apply insights and unlock new growth opportunities .....	19
Product selected to apply insights and unlock new growth opportunities .....	19

<b>Section 9 - How can bellabeat take advantage of global growth opportunities with smart device data? Final recommendations to Bellabeat.....</b>	<b>19</b>
<b>Section 10 - Conclusion .....</b>	<b>21</b>

## Section 1 - Introduction

For my capstone project, I decided to do a case study about health and wellness.

For this project, **I am a data analyst within bellabeats marketing analytics Team.** Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Analysing smart device fitness data could help unlock new growth opportunities.

### About bellabeat

Bellabeat, is 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, cofounder and Chief Creative Officer of bellabeat, believes that analysing smart device fitness data could help unlock new growth opportunities for the company. Collecting data on **activity, sleep, stress, and reproductive health** has allowed bellabeat to empower women with knowledge about their own health and habits.

The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively.

### Company products

We have 5 different products which are:

1. **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. The bellabeat app connects to their line of smart wellness products.
2. **Leaf:** Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the bellabeat app to track activity, sleep, and stress.
3. **Time:** This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.
4. **Spring:** This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.
5. **Bellabeat membership:** Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully

personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

### My assignment

I have been asked to focus on one of bellabeat's products and analyse smart device data to gain insight into how consumers are using their smart devices. The insights I discover will then help guide marketing strategy for the company. I will present my analysis to the bellabeat executive team (key stakeholders) along with my high-level recommendations for bellabeat's marketing strategy. The key stakeholders are Urška Sršen: cofounder and Chief Creative Officer and Sando Mur: Mathematician and cofounder.

### Deliverables

I have produced a report (this document) 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 my analysis
5. Supporting visualizations and key findings
6. My top high-level content recommendations based on my analysis

### The data analysis process

I have followed the 6 steps of the data analysis process to structure this case study which is to:

1. Ask questions to make data-driven decisions,
2. Prepare data for exploration,
3. Process data from dirty to clean,
4. Analyse data to answer questions,
5. Share data through the art of visualisation, and
6. Act on recommendations.

### Deliverable 1 - A clear summary of the business task

One of the deliverables was a clear summary of the business task. I have stated these in Table 1.

Title of assignment	Bellabeat - How can a wellness technology company play it smart?
Industry focus	Health and fitness
Problem statement	How can Bellabeat take advantage of global growth opportunities with smart device data.
Business use case (What I'm solving for)	<ul style="list-style-type: none"><li>• Ways device data can help unlock new growth opportunities.</li><li>• How can insights be applied to one of the current products</li></ul>
Goals	<ul style="list-style-type: none"><li>• Analyse smart device usage data, to gain insight into how consumers use non-Bellabeat</li></ul>

	smart devices. <ul style="list-style-type: none"> <li>• select one bellabeat product to apply these insights to in my presentation.</li> <li>• Help guide marketing strategy</li> </ul>
Deliverables	<ul style="list-style-type: none"> <li>• A clear summary of the business task</li> <li>• A description of all data sources used</li> <li>• Documentation of any cleaning or manipulation of data</li> <li>• A summary of my analysis</li> <li>• Supporting visualizations and key findings</li> <li>• My top high-level content recommendations based on my analysis.</li> </ul>
Are datasets available?	Fitbit Fitness Tracker Data by Mobius on Kaggle.
Dataset list	Other - if required
Website to scrape data if required	N/A
Key stakeholders	Executive team: Urška Sršen: cofounder and Chief Creative Officer. Sando Mur: Mathematician and cofounder

Table 1: Statement of business task

## Section 2 - Ask questions to make data-driven decisions

This section involves understanding what questions I'm trying to answer with the data and ensures I fully understand stakeholder expectations. I also provide information about the dataset, comment on its reliability and state the assumptions I've made.

### About the data

#### Fitbit Fitness Tracker Data

- This data set is in the Public Domain and is made available through Mobius who is a data scientist and Kaggle competition master.
- It contains personal fitness tracker from thirty Fitbit users.
- Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring.
- It includes information about daily activity, steps, and heart rate that can be used to explore smart device users' daily habits.

### Storage

The initial dataset was stored on my local environment; however, remote storage options include Google cloud and GitHub. It comprises of 18 separate excel spreadsheets.

### Format

The data is organised in the long data format - long data subjects can have multiple rows that hold the values of subject attributes.

### Credibility and bias

The participants gave consent and their data was anonymised. Good data sources are reliable, original, comprehensive, current and cited. This data source is reliable because it has a usability score of 10; the highest on Kaggle. It also has a gold dataset medal with over 50 votes.

To validate the data's integrity, I traced the data to its [source](#) (authors who published it) and their names are accessible just in case I had questions. I therefore assume its truthful and reliable.

### Problems with the data

The data was generally complete; however, a column had a significant number of missing values, making it unusable. The data was gathered over a two-day period; however, the dates show that participants had in fact submitted data worth a few months. The two-day period might not have given more participants to respond, leading to a smaller than ideal sample size.

Finally, only thirty participants contributed to the data which is far from ideal. According to the sample size calculator shown in Figure 1, nearly 400 participants were needed to achieve the shown parameters. Therefore, the results are not truly representative of the greater population.

### Sample Size Calculator

#### Find Out The Sample Size

This calculator computes the minimum number of necessary samples to meet the desired statistical constraints.

**Result**

Sample size: **385**

This means 385 or more measurements/surveys are needed to have a confidence level of 95% that the real value is within  $\pm 5\%$  of the measured/surveyed value.

Confidence Level: ?	95% ▾	
Margin of Error: ?	5%	
Population Proportion: ?	50%	Use 50% if not sure
Population Size: ?		Leave blank if unlimited population size.

**Calculate** ▶ **Clear**

Figure 1 - Sample size calculator

### Assumptions made

The data came from thirty participants globally. Given the similarities between the participant lifestyles to people in the USA (assuming that's where most of the participants live) and the UK, it is assumed that my conclusions can also be applied to the UK population.

Finally, the Fitbit data could have been gathered from both men and women. However, bellabeat only caters to women. It is therefore assumed that the data can be used to draw conclusions about women's health; even though data specifically from women would have been ideal.

### Licensing, privacy, security, and accessibility

It's a crowd-sourced Fitbit dataset taken from 03.12.2016-05.12.2016, by Furberg, Robert; Brinton, Julia; Keating, Michael; Ortiz, Alexa. These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016 - 05.12.2016. The dataset is free to analyse and used to gain insights for bellabeat as it is in the public domain. Amazon Mechanical Turk is a crowdsourcing marketplace that makes it easier for individuals and businesses to outsource their processes and jobs to a distributed workforce who can perform tasks virtually.

### Questions that guided my analysis

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

## Section 3 - Prepare data for exploration

This phase of the data analysis process involves looking at the data set and preparing it for exploration.

The three files listed below were chosen for analysis because they contained **merged and therefore comprehensive data** from the participants results. They were also selected because they focus on data that the stakeholders wanted to know and bellabeat devices specifically tracked this data with almost all their products. It was important for me to **relate my findings** back to their products.

The features of the three sheets used for analysis are described below.

### 1. Sleep day merged

This sheet contains information about the participants sleep activity. It has five columns. The column headers are:

- Id: participant Identifier
- Sleep day: the date the record was made
- Total sleep records: the number of sleep records made on the day
- Total minutes asleep: the amount of time the participants spent sleeping
- Total time in bed: amount of time awake and asleep in bed.

### 2. Daily activity merged

This sheet tracked the participants activity levels. It contains information about the participants activity. It has nine columns. The column headers are:

- Id: participant Id
- Activity Date: the date the record was made
- Total Steps: number of steps the participant took
- Total Distance: total distance each participant went. The units for this is unknown.
- Very Active Minutes: amount of very active minutes each participant had on the day

- Fairly Active Minutes: amount of fairly active minutes each participant had on the day
- Lightly Active Minutes: amount of lightly active minutes each participant had on the day
- Sedentary Minutes: number of sedentary minutes each participant had on the day
- Calories: for the participants daily activity, the calories column is assumed to be calories consumed as Fitbit devices track calories consumed as standard.

### 3. **Weight log info**

This sheet tracked the participants weight information and how they recorded the data. It has seven columns. The column headers are:

- Id: participant Id
- Date: the date the record was made
- Time (24 hr): time the record was made in the 24-hour format.
- Weight (Kg): participants weight in Kg.
- BMI: participants Body Mass Index.
- Weight (Pounds): participants weight in pounds
- Is Manual Report: this states if their weight information was recorded manually or automatically and is True or False; a Boolean data type.

## Section 4 - Process Data from dirty to clean

This phase of the data analysis process involved checking the data is valid, complete, and clean before I began any analysis. I did lots of data manipulation (which is changing or altering data in order to make it more readable and organised).

I have outlined the changes I made along with screenshots of before and after the changes. For all the documents, I made a copy of the original sheet so that original can always be referred to. I also made the column headers bold and expanded them for a visually appealing sheet.

## Cleaning data activity merged sheet

Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActive	VeryActive	Moderate	LightActive	Sedentary	VeryActive	FairlyActive	LightlyActive	Sedentary	Calories
1	1.5E+09 04/12/2016	13162	8.5	8.5	0	1.88	0.55	6.06	0	25	13	328	728	1985
2	1.5E+09 4/13/2016	10735	6.97	6.97	0	1.57	0.69	4.71	0	21	19	217	776	1797
3	1.5E+09 4/14/2016	10460	6.74	6.74	0	2.44	0.4	3.91	0	30	11	181	1218	1776
4	1.5E+09 4/15/2016	9762	6.28	6.28	0	2.14	1.26	2.83	0	29	34	209	726	1745
5	1.5E+09 4/16/2016	12669	8.16	8.16	0	2.71	0.41	5.04	0	36	10	221	773	1863
6	1.5E+09 4/17/2016	9705	6.48	6.48	0	3.19	0.78	2.51	0	38	20	164	539	1728
7	1.5E+09 4/18/2016	13019	8.59	8.59	0	3.25	0.64	4.71	0	42	16	233	1149	1921
8	1.5E+09 4/19/2016	15506	9.88	9.88	0	3.53	1.32	5.03	0	50	31	264	775	2035
9	1.5E+09 4/20/2016	10544	6.68	6.68	0	1.96	0.48	4.24	0	28	12	205	818	1786
10	1.5E+09 4/21/2016	9819	6.34	6.34	0	1.34	0.35	4.65	0	19	8	211	838	1775
11	1.5E+09 4/22/2016	12764	8.13	8.13	0	4.76	1.12	2.24	0	66	27	130	1217	1827
12	1.5E+09 4/23/2016	14371	9.04	9.04	0	2.81	0.87	5.36	0	41	21	262	732	1949
13	1.5E+09 4/24/2016	10039	6.41	6.41	0	2.92	0.21	3.28	0	39	5	238	709	1788
14	1.5E+09 4/25/2016	15355	9.8	9.8	0	5.29	0.57	3.94	0	73	14	216	814	2013
15	1.5E+09 4/26/2016	13755	8.79	8.79	0	2.33	0.92	5.54	0	31	23	279	833	1970
16	1.5E+09 4/27/2016	18134	12.21	12.21	0	6.4	0.41	5.41	0	78	11	243	1108	2159
17	1.5E+09 4/28/2016	13154	8.53	8.53	0	3.54	1.16	3.79	0	48	28	189	782	1898
18	1.5E+09 4/29/2016	11181	7.15	7.15	0	1.06	0.5	5.58	0	16	12	243	815	1837
19	1.5E+09 4/30/2016	14673	9.25	9.25	0	3.56	1.42	4.27	0	52	34	217	712	1947
20	1.5E+09 05/01/2016	10602	6.81	6.81	0	2.29	1.6	2.92	0	33	35	246	730	1820

Figure 2 - Screenshot of original data daily activity merged

- For Figure 2, I deleted column E because it had the same data as column D.
- I formatted the dates to a mmm-dd-yyyy format. A lot of the cells remained unchanged despite this, so I used the “text-to-columns” feature to change the dates and remove the delimiter.
- The relevant values were converted to numbers.
- Calories is assumed to be calories consumed as Fitbit devices track calories consumed as standard.

Figure 3 is a screenshot of the cleaned version of daily activity merged sheet

Id	ActivityDate	TotalSteps	TotalDistance	VeryActiveMinutes	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes	Calories
1	1503960366 Apr-12-2016	13162	8.5	25	13	328	728	1985
2	1503960366 Apr-13-2016	10735	6.96999979	21	19	217	776	1797
3	1503960366 Apr-14-2016	10460	6.739999771	30	11	181	1218	1776
4	1503960366 Apr-15-2016	9762	6.28000021	29	34	209	726	1745
5	1503960366 Apr-16-2016	12669	8.159999847	36	10	221	773	1863
6	1503960366 Apr-17-2016	9705	6.480000019	38	20	164	539	1728
7	1503960366 Apr-18-2016	13019	8.590000153	42	16	233	1149	1921
8	1503960366 Apr-19-2016	15506	9.880000114	50	31	264	775	2035
9	1503960366 Apr-20-2016	10544	6.679999828	28	12	205	818	1786
10	1503960366 Apr-21-2016	9819	6.340000153	19	8	211	838	1775
11	1503960366 Apr-22-2016	12764	8.130000114	66	27	130	1217	1827
12	1503960366 Apr-23-2016	14371	9.039999962	41	21	262	732	1949
13	1503960366 Apr-24-2016	10039	6.409999847	39	5	238	709	1788
14	1503960366 Apr-25-2016	15355	9.800000191	73	14	216	814	2013
15	1503960366 Apr-26-2016	13755	8.789999962	31	23	279	833	1970
16	1503960366 Apr-27-2016	18134	12.21000004	78	11	243	1108	2159
17	1503960366 Apr-28-2016	13154	8.529999733	48	28	189	782	1898
18	1503960366 Apr-29-2016	11181	7.150000095	16	12	243	815	1837
19	1503960366 Apr-30-2016	14673	9.25	52	34	217	712	1947
20	1503960366 May-01-2016	10602	6.809999943	33	35	246	730	1820

Figure 3: Cleaned version of daily activity merged sheet



## Cleaning sleep day merged sheet

	A	B	C	D	E	F	G
1	Id	SleepDay	TotalSleep	TotalMinu	TotalTimeInBed		
2	1.5E+09	#####	1	327	346		
3	1.5E+09	4/13/2016	2	384	407		
4	1.5E+09	4/15/2016	1	412	442		
5	1.5E+09	4/16/2016	2	340	367		
6	1.5E+09	4/17/2016	1	700	712		
7	1.5E+09	4/19/2016	1	304	320		
8	1.5E+09	4/20/2016	1	360	377		
9	1.5E+09	4/21/2016	1	325	364		
10	1.5E+09	4/23/2016	1	361	384		
11	1.5E+09	4/24/2016	1	430	449		
12	1.5E+09	4/25/2016	1	277	323		
13	1.5E+09	4/26/2016	1	245	274		
14	1.5E+09	4/28/2016	1	366	393		
15	1.5E+09	4/29/2016	1	341	354		
16	1.5E+09	4/30/2016	1	404	425		
17	1.5E+09	#####	1	369	396		
18	1.5E+09	#####	1	277	309		
19	1.5E+09	#####	1	273	296		
20	1.5E+09	#####	1	247	264		
21	1.5E+09	#####	1	334	367		

Figure 4: Screenshot of original sleep data merged

- The most important action here was to format the date correctly. Again, this was done using Excel's text-to-columns feature.
- The time was deleted because the records were all at 00:00AM
- The minutes were converted to number formats

Figure 5 is a screenshot of the cleaned version of the sleep data

	A	B	C	D	E	F	G
1	Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed		
2	1503960366	Dec-04-2016	1	327	346		
3	1503960366	Apr-13-2016	2	384	407		
4	1503960366	Apr-15-2016	1	412	442		
5	1503960366	Apr-16-2016	2	340	367		
6	1503960366	Apr-17-2016	1	700	712		
7	1503960366	Apr-19-2016	1	304	320		
8	1503960366	Apr-20-2016	1	360	377		
9	1503960366	Apr-21-2016	1	325	364		
10	1503960366	Apr-23-2016	1	361	384		
11	1503960366	Apr-24-2016	1	430	449		
12	1503960366	Apr-25-2016	1	277	323		
13	1503960366	Apr-26-2016	1	245	274		
14	1503960366	Apr-28-2016	1	366	393		
15	1503960366	Apr-29-2016	1	341	354		
16	1503960366	Apr-30-2016	1	404	425		
17	1503960366	Jan-05-2016	1	369	396		
18	1503960366	Feb-05-2016	1	277	309		
19	1503960366	Mar-05-2016	1	273	296		
20	1503960366	May-05-2016	1	247	264		
21	1503960366	Jun-05-2016	1	334	367		

Figure 5: Screenshot of the cleaned version of the sleep data

### Cleaning weight log info merged sheet

	A	B	C	D	E	F	G	H	I	J
1	Id	Date	WeightKg	WeightPoi	Fat	BMI	IsManualF	LogId		
2	1.5E+09	#####	52.6	115.9631	22	22.65	TRUE	1.46E+12		
3	1.5E+09	#####	52.6	115.9631		22.65	TRUE	1.46E+12		
4	1.93E+09	4/13/2016	133.5	294.3171		47.54	FALSE	1.46E+12		
5	2.87E+09	4/21/2016	56.7	125.0021		21.45	TRUE	1.46E+12		
6	2.87E+09	#####	57.3	126.3249		21.69	TRUE	1.46E+12		
7	4.32E+09	4/17/2016	72.4	159.6147	25	27.45	TRUE	1.46E+12		
8	4.32E+09	#####	72.3	159.3942		27.38	TRUE	1.46E+12		
9	4.56E+09	4/18/2016	69.7	153.6622		27.25	TRUE	1.46E+12		
10	4.56E+09	4/25/2016	70.3	154.985		27.46	TRUE	1.46E+12		
11	4.56E+09	#####	69.9	154.1031		27.32	TRUE	1.46E+12		
12	4.56E+09	#####	69.2	152.5599		27.04	TRUE	1.46E+12		
13	4.56E+09	#####	69.1	152.3394		27	TRUE	1.46E+12		
14	5.58E+09	4/17/2016	90.7	199.9593		28	FALSE	1.46E+12		
15	6.96E+09	#####	62.5	137.7889		24.39	TRUE	1.46E+12		
16	6.96E+09	4/13/2016	62.1	136.9071		24.24	TRUE	1.46E+12		
17	6.96E+09	4/14/2016	61.7	136.0252		24.1	TRUE	1.46E+12		
18	6.96E+09	4/15/2016	61.5	135.5843		24	TRUE	1.46E+12		
19	6.96E+09	4/16/2016	62	136.6866		24.21	TRUE	1.46E+12		
20	6.96E+09	4/17/2016	61.4	135.3638		23.96	TRUE	1.46E+12		
21	6.96E+09	4/18/2016	61.2	134.9229		23.89	TRUE	1.46E+12		

Figure 6: Screenshot of original weight log info merged

- For Figure 6, I made the relevant columns numbers and adjusted the decimal places. E.g., reduced weight, and BMI to two decimal places as that's how they're typically reported.

- The date was formatted using excels text-to-columns feature and then formatted into a date using the mmm-dd-yyyy format.
- I formatted the time to a 24-hour format.
- The LogId column was deleted as each entry was already a unique record.
- The weight in Kg and pounds were kept in case the audience has a preference during more interactive visualisation.
- The fat column has also been deleted as there are a significant number of missing values (just 2 out of 67 values were present). The values for fat couldn't be calculated either as the gender of the participant was missing.

Figure 7 is the screenshot of the final cleaned version.

	A	B	C	D	E	F	G	H
	<b>Id</b>	<b>Date</b>	<b>Time (24 hr)</b>	<b>WeightKg</b>	<b>WeightPounds</b>	<b>BMI</b>	<b>IsManualReport</b>	
2	1503960366	May-02-2016	11:59:59 PM	52.60	115.96	22.65	TRUE	
3	1503960366	May-03-2016	11:59:59 PM	52.60	115.96	22.65	TRUE	
4	1927972279	Apr-13-2016	01:08:52 AM	133.50	294.32	47.54	FALSE	
5	2873212765	Apr-21-2016	11:59:59 AM	56.70	125.00	21.45	TRUE	
6	2873212765	May-12-2016	11:59:59 PM	57.30	126.32	21.69	TRUE	
7	4319703577	Apr-17-2016	11:59:59 AM	72.40	159.61	27.45	TRUE	
8	4319703577	May-04-2016	11:59:59 PM	72.30	159.39	27.38	TRUE	
9	4558609924	Apr-18-2016	11:59:59 AM	69.70	153.66	27.25	TRUE	
10	4558609924	Apr-25-2016	11:59:59 AM	70.30	154.98	27.46	TRUE	
11	4558609924	May-01-2016	11:59:59 PM	69.90	154.10	27.32	TRUE	
12	4558609924	May-02-2016	11:59:59 PM	69.20	152.56	27.04	TRUE	
13	4558609924	May-09-2016	11:59:59 PM	69.10	152.34	27.00	TRUE	
14	5577150313	Apr-17-2016	09:17:55 AM	90.70	199.96	28.00	FALSE	
15	6962181067	Apr-12-2016	11:59:59 PM	62.50	137.79	24.39	TRUE	
16	6962181067	Apr-13-2016	11:59:59 AM	62.10	136.91	24.24	TRUE	
17	6962181067	Apr-14-2016	11:59:59 AM	61.70	136.03	24.10	TRUE	
18	6962181067	Apr-15-2016	11:59:59 AM	61.50	135.58	24.00	TRUE	
19	6962181067	Apr-16-2016	11:59:59 AM	62.00	136.69	24.21	TRUE	
20	6962181067	Apr-17-2016	11:59:59 AM	61.40	135.36	23.96	TRUE	
21	6962181067	Apr-18-2016	11:59:59 AM	61.20	134.92	23.89	TRUE	

Figure 7: Screenshot of the cleaned version of weight log info

## Section 5 - Analyse Data to Answer Questions

The goal of this phase was to identify trends and relationships within the data so that I could accurately answer the questions that stakeholders were asking.

### Pivot tables and pivot charts

I started by inserting pivot tables and pivot charts to summarise the cleaned data. This enabled me to spot trends and patterns much quicker than manually inspecting the data sets.

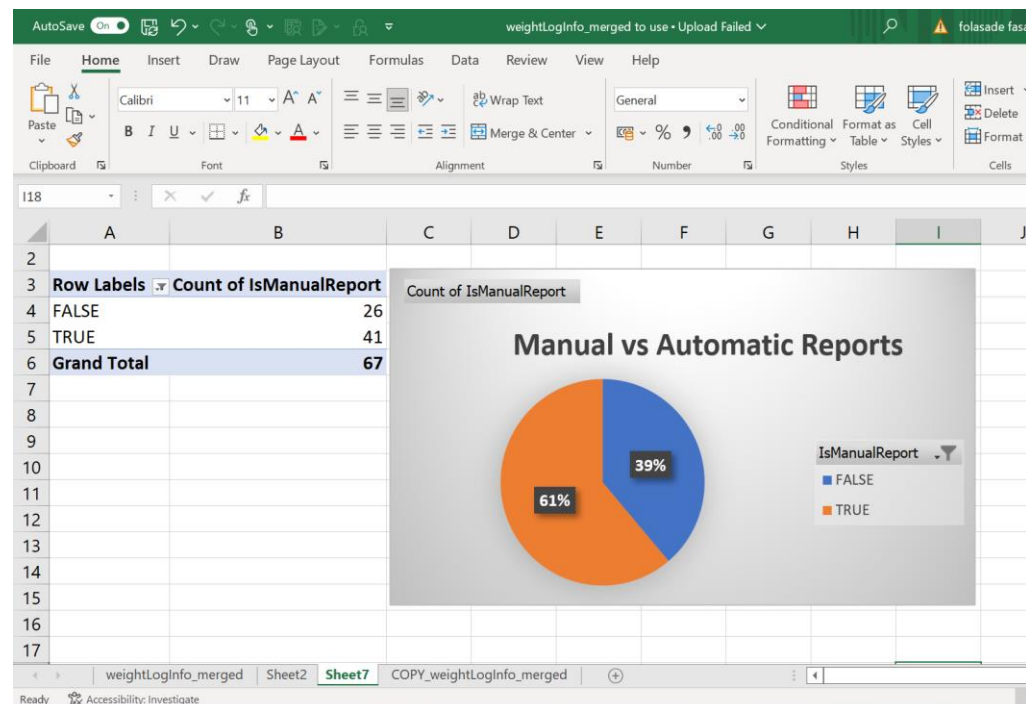


Figure 8: Pivot table & chart showing how many people recorded weight information manually vs automatically.

Figure 8 shows a pivot table I inserted to visualise how many people recorded weight information themselves vs automatically. The results from the pivot pie chart I inserted show that 61% of participants recorded their readings manually. This is a potential marketing angle to make fitness tracking easier for bellabeat customers.

Table 2 shows the pivot table I inserted to show the average, minimum and maximum weights as well as the participants BMI.

Average of WeightKg	Max of WeightKg	Min of WeightKg
72.03582137	133.5	52.59999847
Average of WeightPounds	Max of WeightPounds	Min of WeightPounds
158.8118014	294.31712	115.9631465
Average of BMI	Max of BMI	Min of BMI
25.18522379	47.54000092	21.45000076

Table 2: Pivot table showing the average, minimum and maximum weights as well as BMI.

I made a pivot table to get the average values for total time in bed and total minutes asleep, as shown in Table 3.

Row Labels	Sum of TotalSleepRecords	Average of TotalTimeInBed	Average of TotalMinutesAsleep
1	367	451.9482289	413.6866485
2	86	500.0930233	453.1395349
3	9	683	644
<b>Grand Total</b>	<b>462</b>	<b>458.6392252</b>	<b>419.4673123</b>

Table 3: Pivot table showing the average values for total time in bed and total minutes asleep

Figure 9 shows the bar chart I made, comparing how long participants spent in bed vs how many minutes of sleep they got.

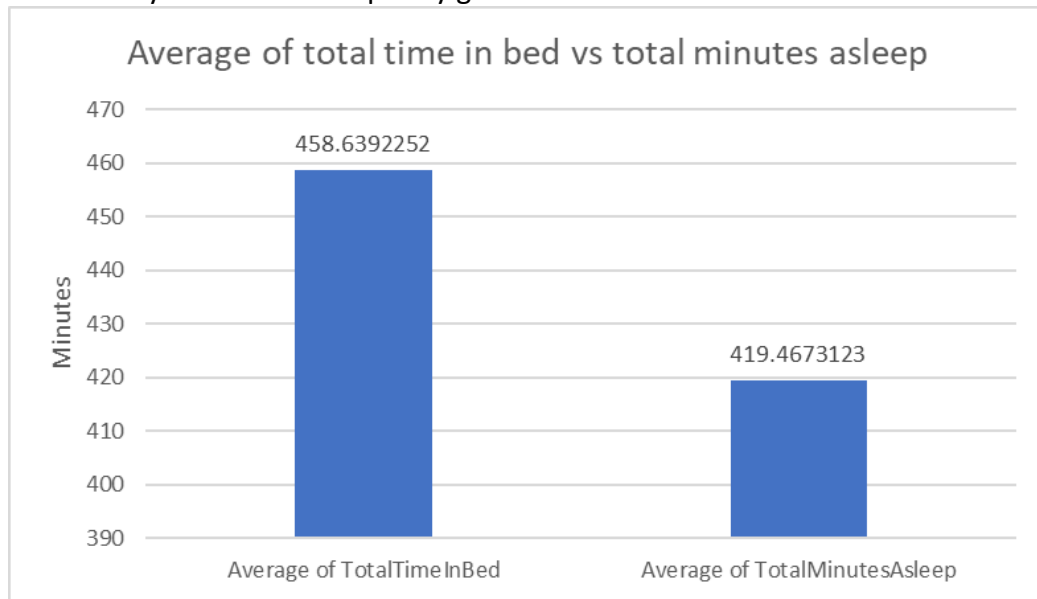


Figure 9: Bar chart comparing how long participants spent in bed vs how many minutes of sleep they got.

## Section 6 - Share Data Through the Art of Visualization

This phase involves sharing results from the data analysis I did. I aimed for the stakeholders to know exactly what they were looking at within the **first five seconds** of seeing it. In the five seconds after that, they should **understand the conclusion** my visualisation is making.

### Insights from daily activity

I analysed the data from the users' daily activities and wanted to investigate the points below from my initial look at the data:

- Total steps vs distance,
- Calories consumed vs total distance,
- Calories consumed vs sedentary mins
- Average numbers for variables.

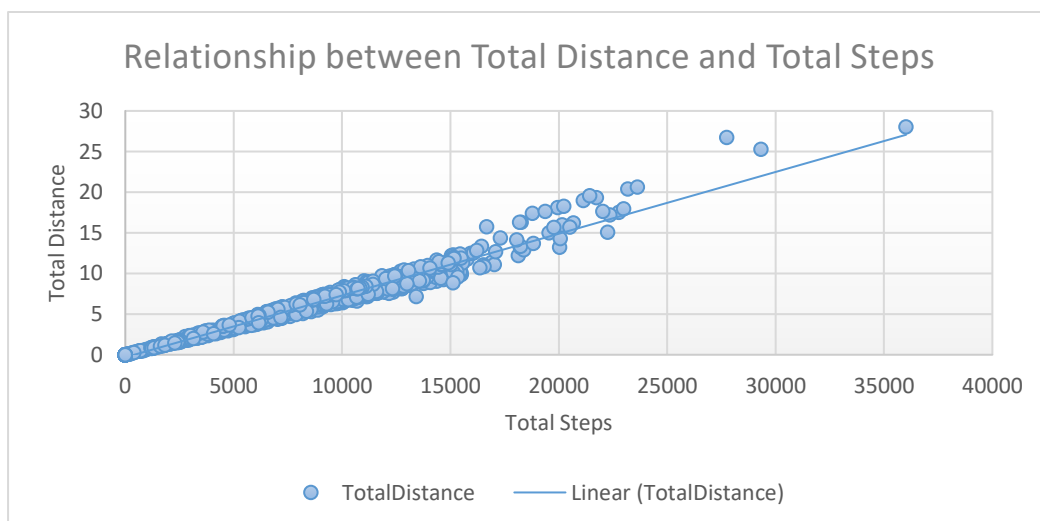


Figure 10: Scatterplot showing the relationship between total distance and total steps.

Figure 10 shows the relationship between total distance and total steps. The correlation coefficient between total distance and total steps taken by the participants is 0.99. This indicates a very strong positive relationship, meaning the further the distance of activity, the more steps a participant took. A correlation coefficient is a number between -1 and 1 that tells you the **strength** and **direction** of a relationship between variables.

Figure 11 shows the minutes and percentage of each type of activity by the participants. 81% of the participants activity levels was sedentary out of 100%, meaning there was a significant proportion of sedentary activity. This was followed by lightly active minutes at 16%. Just 3% of activity was fairly or very active.

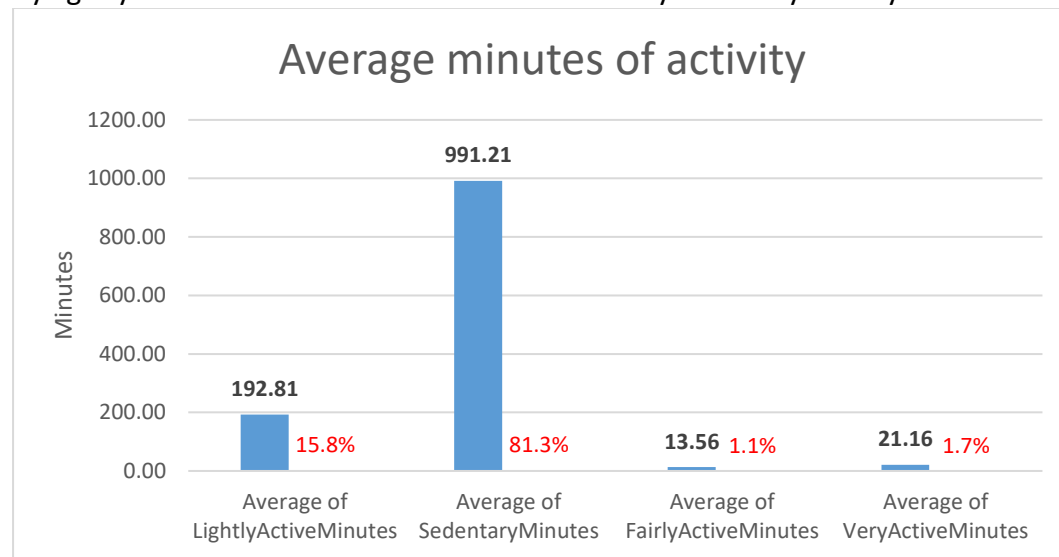


Figure 11: Bar chart showing average minutes of activity

Average of Calories	Average of Lightly Active Minutes	Average of Sedentary Minutes	Average of Total Steps	Average of Fairly Active Minutes	Average of Very Active Minutes	Average of Total Distance
2303.60	192.81	991.21	7637.91	13.56	21.16	5.49

Table 4: The average from the participants calorie intake, active minutes and total distance of activity

Table 4 highlights the average statistics from the participants calorie intake, active minutes and total distance of activity. According to the NHS, generally, the recommended daily calorie intake is 2,000 calories a day for women. The average number of steps taken is just 7,637. This is far from the recommended 10,000 steps daily.

The total amount of average minutes of activity is 227.53 which equals to 3 hours, 47 minutes and 24 seconds. On average, 34.72 minutes daily is used for fairly or very active minutes of exercise which is inline with the NHS recommendations of at least 150 minutes of moderate intensity activity a week or 75 minutes of vigorous intensity activity a week spread evenly over 4 to 5 days a week, or every day. The average sedentary minutes daily is 991.21 which equals to approximately 16.5 hours daily.

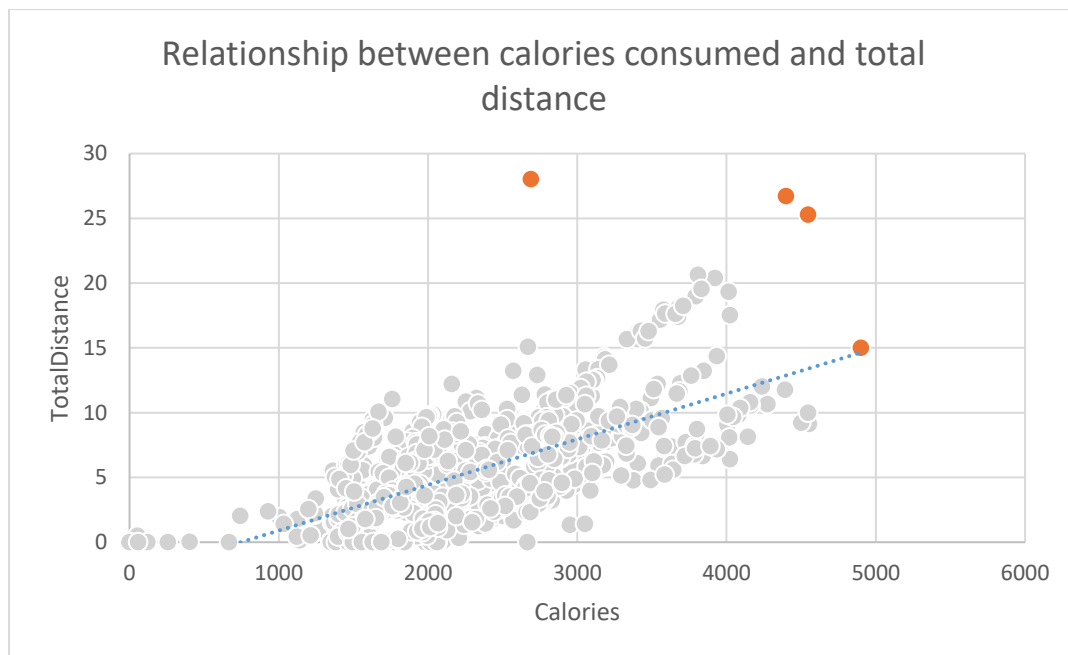


Figure 12: Scatterplot showing the relationship between calories consumed and total distance

Figure 12 shows the relationship between calories consumed and total distance. The correlation coefficient is 0.65 indicating a strong positive relationship. The more calories the participants consumed, the further distance of activity they went. The orange dots indicate the 4 outliers from the dataset.

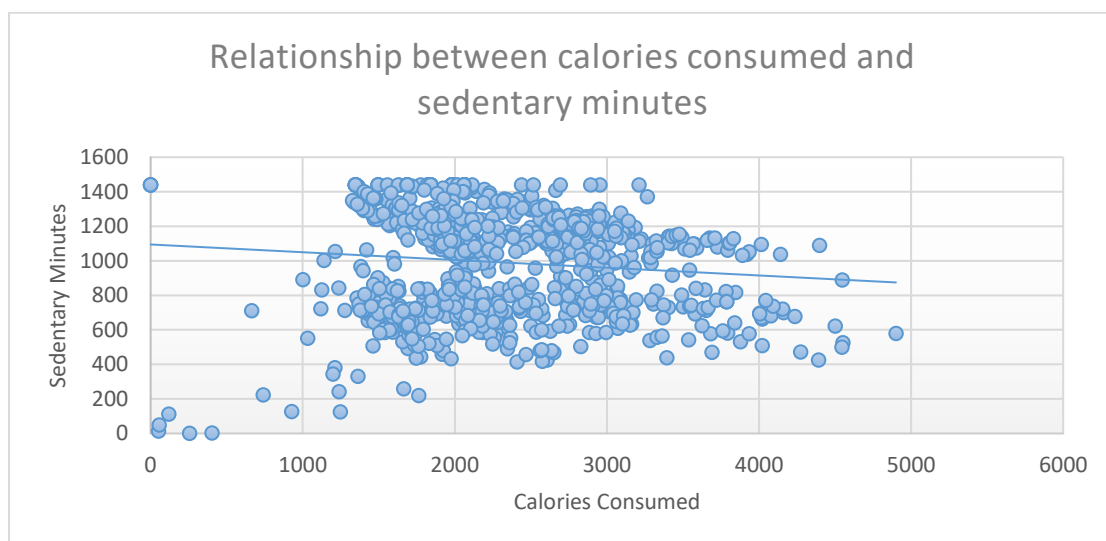


Figure 13: Scatterplot showing the relationship between calories consumed and sedentary minutes

Figure 13 shows the relationship between calories consumed and sedentary minutes. The correlation coefficient is -0.11 indicating a weak negative relationship, as supported by the negative trendline. The more sedentary minutes of activity a participant had, the less calories they consumed; however, as the relationship is a weak one, the data isn't strong enough to make data-driven business decisions. Further study will be needed to establish the relationship between calories consumed and sedentary activity. Alternatively, a well-established study could be referenced.

### Insights from weight information

I analysed the data from the users' weight information and wanted to investigate the points below from my initial look at the data.

- How many people recorded manually? Do they like being in control or want it automated? More people recorded manually. Perhaps provide manual recordings as an option or make accurate auto recordings a key selling point.
- Relationship between weight and BMI
- Participants BMI on average

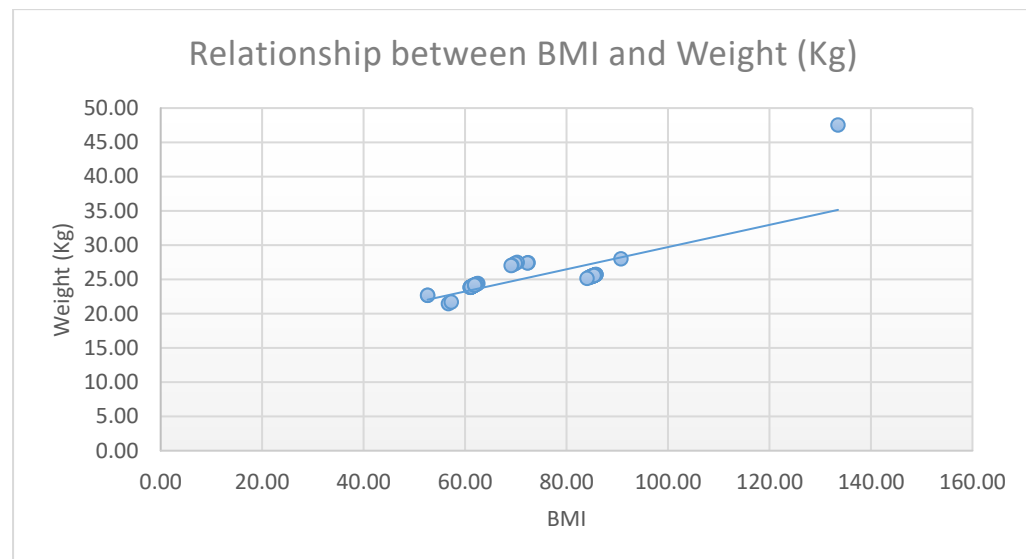


Figure 14: Scatterplot of the relationship between BMI and Weight (Kg)

Figure 14 shows a scatterplot of the relationship between BMI and weight. The correlation coefficient is 0.74 indicating a strong positive relationship. The more a participant weighed, the higher their BMI. However, this correlation is unsurprising because the **formula for BMI is  $\text{kg/m}^2$** .

The NHS states that the body mass index (BMI) is a measure that uses height and weight to work out if a weight is healthy. For most adults, an ideal BMI is between 18.5 to 24.9.

Average of WeightKg	Average of WeightPounds	Average of BMI
72.03582137	158.8118014	25.18522379

Table 5: Participants average weight and BMI

Table 5 shows the participants average weight in Kg and pounds, as well as the average BMI. The average BMI of the participants was 25.18 indicating the average participant was overweight, despite getting the recommended amounts of exercise on average.



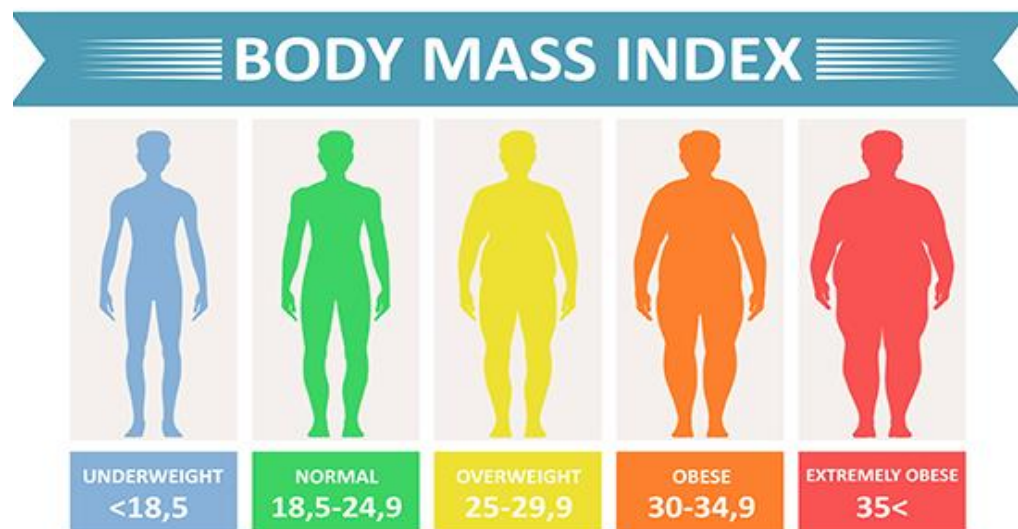


Figure 15: BMI index. Creator: VikiVector | Credit: Getty Images/iStockphoto

Row Labels	Count of IsManualReport
FALSE	26
TRUE	41
<b>Grand Total</b>	<b>67</b>

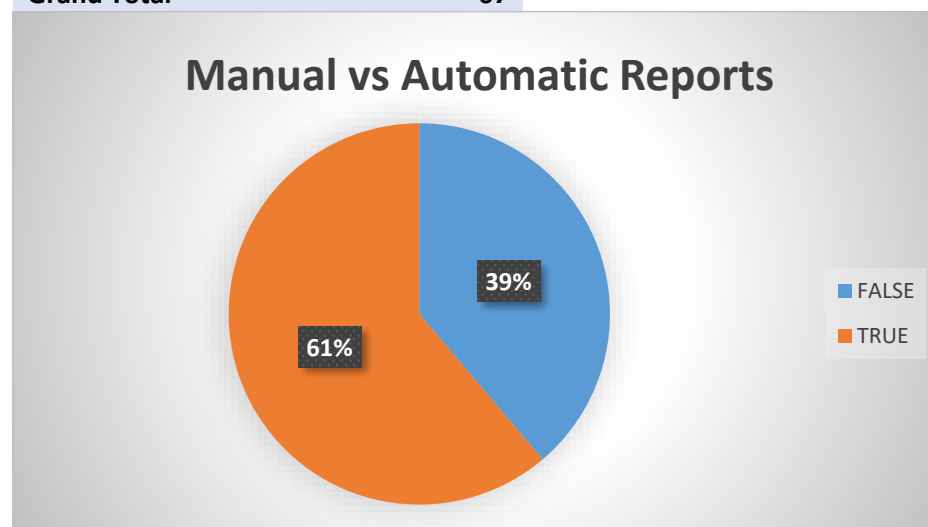


Figure 16: Image of pie chart and table showing manual vs automatic reports of weight

Figure 16 shows what proportion of weight information was manually and automatically recorded. 61% of the participants reported information about their weight manually. This is a significant amount and there could be an angle for marketing bellabeat vs its competitors.

### Insights from sleep information

I analysed the data from the users' sleep information and wanted to investigate the points below from my initial look at the data.

- How many sleep records by group?
- Average of total time in bed vs total mins asleep.
- Total mins asleep vs calories consumed?

The pivot table below shows the total number of sleep records, the average total time in bed and the average total minutes asleep.

Row Labels	Sum of TotalSleepRecords	Average of TotalTimeInBed	Average of TotalMinutesAsleep
1	367	451.9482289	413.6866485
2	86	500.0930233	453.1395349
3	9	683	644
<b>Grand Total</b>	<b>462</b>	<b>458.6392252</b>	<b>419.4673123</b>

Table 6: Total number of sleep records, the average total time in bed and the average total minutes asleep

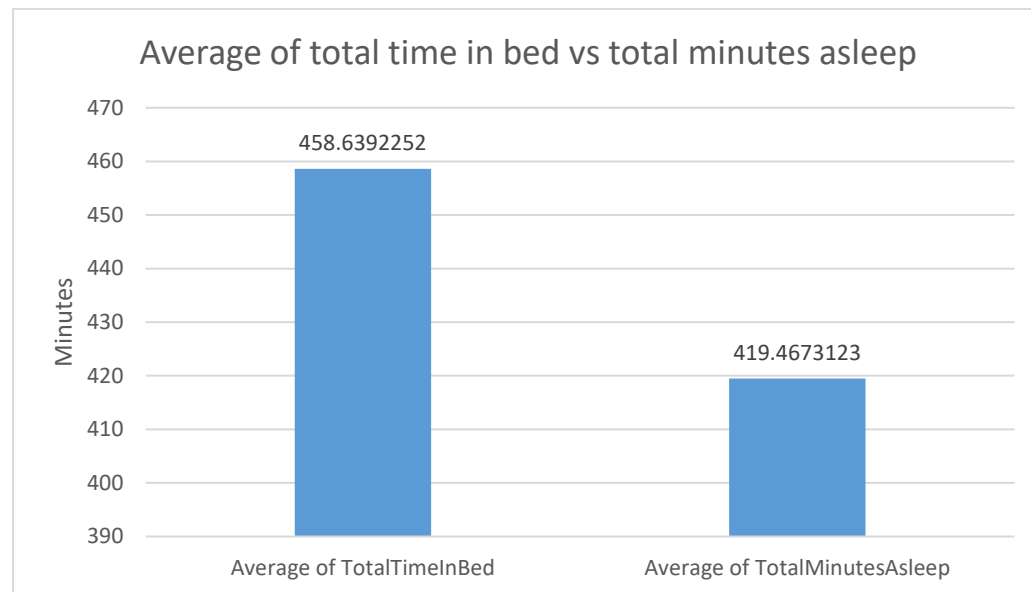


Figure 17: Bar chart showing average of total time in bed vs total minutes asleep

Figure 17 shows that on average, the participants spent 39 minutes in bed before finally falling asleep. There is a growth opportunity to help people fall asleep quicker, therefore leading to a marketing angle. The Y axis of the original chart didn't start at 0 however, this has been done to emphasise the difference between the 2 variables. In a case where it was imperative to avoid misrepresentation, the Y axis would start at 0.

## Section 7 - Scope for further study

This section talks briefly about what could be done in future for better results and further insight.

- The average total minutes asleep could be compared against the total number of calories consumed. Do people who consume more calories spend more total time asleep?
- Stress levels can be analysed, as its also a fundamental part to holistic wellness.
- More insights can be uncovered if the data was merged with other tools like SQL and Python. I would also develop interactive visualisations to filter specific data (e.g., weight in Kg vs Lbs for those who prefer different units). This would be achieved with Tableau or Power BI. These would be my next steps, given more time with the data.

- Finally, a larger sample size with solely women, would yield more accurate and representative results.

## Section 8 - Act to make data-driven decisions

The final stage of the data analysis process is to act to make data driven decisions. This involves me sharing my results to stakeholders with visualisations in Microsoft PowerPoint for example.

### Products not selected to apply insights and unlock new growth opportunities

- The Bellabeat app wasn't selected as the focus as it's the central hub for bellabeats products to function. It's assumed that whichever product the user gets, they'll have to use the app.
- The Time and the Leaf weren't selected as they're Jewellery which people may not always have on. Also, they don't track nutrition.
- The Spring wasn't selected as it only caters to hydration. The data shows that fitness device wearers need more of a lifestyle tracker than a simple tracker.

### Product selected to apply insights and unlock new growth opportunities

**Bellabeat membership:** Bellabeat offers a subscription-based membership program for users. Membership gives users 24/7 access to fully **personalised guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.**

The Bellabeat membership has been selected to apply the data insights to because it's fully personalised. Most importantly, it provides guidance on nutrition, sleep and activity levels, as well as other factors that hasn't been analysed but are fundamentally important to a healthy lifestyle.

## Section 9 - How can bellabeat take advantage of global growth opportunities with smart device data? Final recommendations to Bellabeat

1. The Fitbits these participants used didn't measure stress, or stress results were not provided in the data set. This is a key selling point for the membership as membership involves promoting mindfulness which prevents stress, therefore reducing conditions like high blood pressure, heart attacks and strokes. The marketing team should emphasise that **having a Bellabeat membership will actively reduce their stress levels.**
2. The further the distance of activity, the more steps a participant took. As Bellabeat membership is fully personalised, rewards and incentives could be offered to the user. For example, after X distance, they unlock a personal reward they like for example, heavy discounts on Graze products. This therefore incentivises the user to remain subscribed to the bellabeat membership. **The marketing team should emphasize the user being rewarded for doing what they love.**

3. 81% of the participants average activity throughout the day was sedentary activity and 3% of activity was fairly or very active. Like the above point, rewards could be given if these percentages changed. For example, if fair or very active minutes increased to 5% for the user. The membership would be capable of tracking this measure through the Leaf or Time. This would inevitably lead to **reduced chances of diseases, and an increase in feeling good which the marketing team should promote.**
4. Key metrics like calorie intake, BMI, amount of moderate and vigorous activity and step count have national recommended levels. The data uncovered that while participants got recommended levels of exercise, they were still overweight on average. A Bellabeat membership will enable these to be tracked and used to guide the customer towards the best goal for them. **The marketing team should promote weight loss through individualised fitness plans.**
5. The more calories the participants consumed, the further distance of activity they went. The membership will be the best way for users to track how far they should go and therefore how many calories are best to consume. **Again, reinforcing individualized fitness with personalized goals and overall wellbeing.** This is what the marketing team should promote. Rewards and discounts can again, be offered to users who surpass their goals.
6. 61% of the participants reported information about their weight manually. This is a significant amount. Depending on the reason for this, the marketing team could promote the option for users to record their weight information manually (if possible) or **promote the fact that their weight information is automatically tracked with accuracy, vs its competitors.**
7. On average, people spent 39 minutes in bed before finally falling asleep. There is a growth opportunity to help people fall asleep quicker, therefore **the marketing team should promote better sleep and a longer night of sleep towards the recommended 7 – 8 hours** (or whatever works best for the user). The app could provide educative content about sleep hygiene and actively guide the user through this. The membership could also provide a space for meditation and calming sounds (or provide discounts to an app like Headspace) to help users fall asleep faster.
8. Bellabeat should continue to focus on **digital marketing.** This provides a more targeted reach with lower costs than traditional media. Marketing strategies like **content writing, copywriting and video marketing** should also be used for the best results and greater profits.
9. Finally, a drive to the membership will provide recurring income for bellabeat which will increase revenues and net profits, as opposed to one-off payments from the Leaf, Time and Spring. **The marketing team as well as the wider**

**teams, should therefore do all they can to promote the membership to as many as possible.**

## Section 10 - Conclusion

In conclusion, I have combined Microsoft Excel, with my domain knowledge in health to explore, visualise and present fitness data in order to create revenue for bellabeat and help women improve their health. There are ample ways for bellabeat to take advantage of global growth opportunities with their smart device data. The marketing team should focus on promoting the bellabeat membership and emphasize the points below:

1. Having a Bellabeat membership will actively reduce their stress levels.
2. The marketing team should emphasize the user being rewarded for doing what they love.
3. The marketing should promote that having a membership will reduce chances of diseases, and lead to an increase in feeling good.
4. The marketing team should promote weight loss through individualised fitness plans.
5. The marketing team should promote the fact that their weight information is automatically tracked with accuracy, vs its competitors.
6. The marketing team should promote better sleep and a longer night of sleep towards the recommended number of hours.
7. Bellabeat should continue to focus on digital marketing as the main advertising channel.