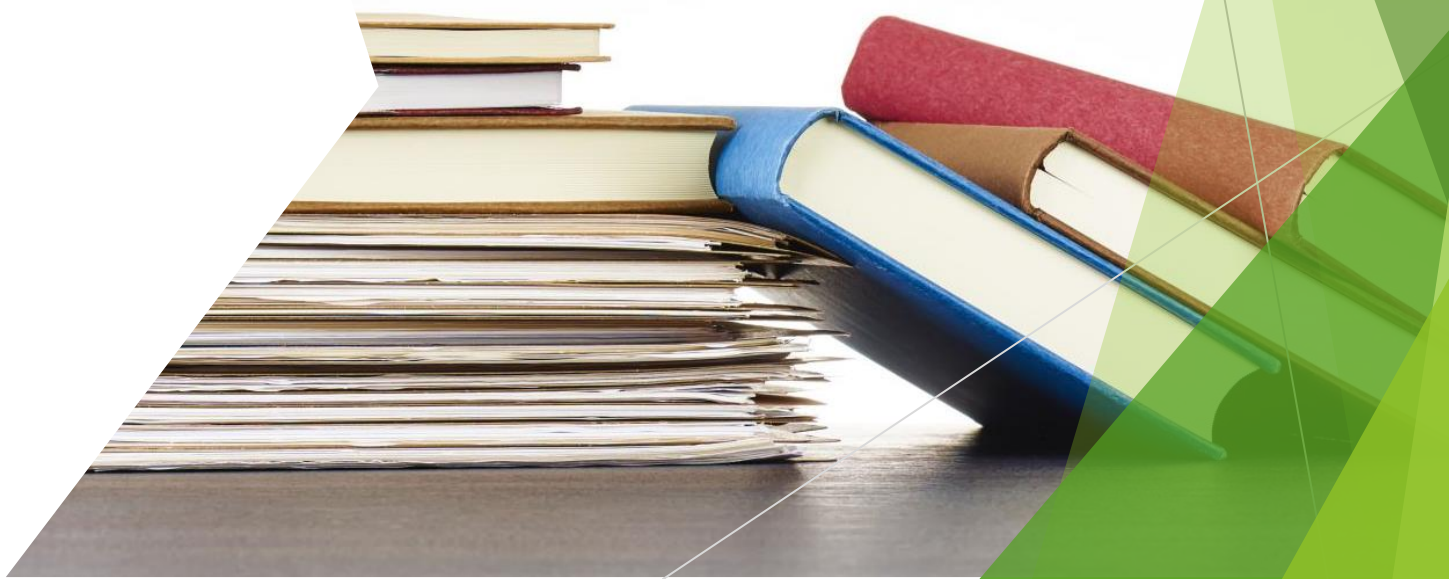
Case Study: How Can a Wellness Technology Company Play It Smart?

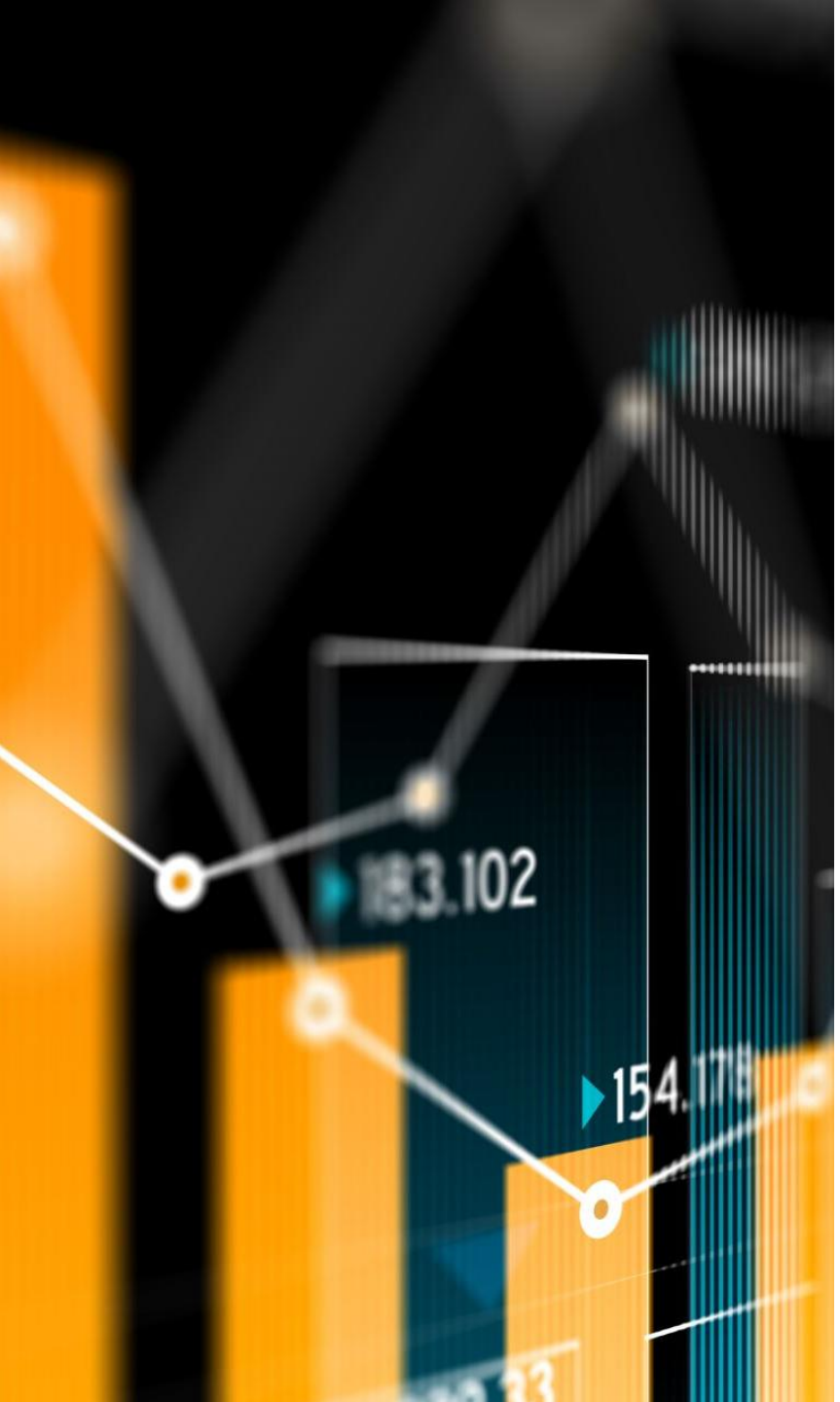
By Kristin Lu

February 1, 2024



Overview of the Case Study

In this case study, publicly available smart device fitness data was analyzed to gain insights into how consumers use their fitness tracker/smart devices. Insights discovered will help guide marketing strategy for Bellabeat, a high-tech manufacturer specializing in women's health products.



Statement of Business Task

Analyze publicly available fitness tracker (smart device) usage data to discover trends which may help influence the marketing strategy for a wellness technology company.

Data Preparation and Data Exploration

- The data used for this case study is the "**FitBit Fitness Tracker Data**" which was downloaded from Kaggle.
- This Kaggle data set 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 users' habits.

Data Preparation and Data Exploration

- There are 18 files in the dataset. Not all the files were used for the analysis. The following is a description of the files used in this case study:
  - dailyActivity_merged.csv - this file contains the following columns: Id, ActivityDate, (in Short Date format “m/d/yyyy”), TotalSteps, TotalDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance, SedentaryActiveDistance, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes, Calories, etc.
  - heartrate_seconds_merged.csv - this file contains the following columns: Id, Time (in “m/d/yyyy h:mm” format), Value (heartrate).
  - sleepDay_merged.csv - this file contains the following columns: Id, SleepDay (in “m/d/yyyy h:mm” format), TotalSleepRecords, TotalMinutesAsleep, and TotalTimeInBed.
  - weightLogInfo_merged.csv - this file contains the following columns: Id, Date (in “m/d/yyyy h:mm” format), WeightKg, WeightPounds, Fat, BMI, etc.

Data Processing – Daily Activity

- Open dailyActivity_merged.csv and save it as an Excel Workbook.
- Built a Pivot Table to count number of rows associated with each consumers. Removed all rows associated with consumers with less than 15 days of data (15 days is 50% of 31 days, that is, the number of days from 4/12/2016 to 5/12/2016). Otherwise, the data might be biased. One consumer (4057192912) were impacted.
- Built a pivot table to sum up the total calories consumed by each consumer. Do the same to sum up the total calories consumed by each consumer in hourlyCalories_merged file. Compared these two sets of values. Delete rows associated with consumers whose total calorie values calculated here are significantly different from that values calculated in another file (difference greater than 1000 calories). Four consumers (8583815059, 6117666160, 4388161847, and 4319703577) were impacted.
- It's not clear that if sleeping time is considered SedentaryMinutes. Let's added a new column TotalMinutes to sum up VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes and SedentaryMinutes. There are 400+ rows have TotalMinutes field equal to 1440. Looks like sleeping time is considered SedentaryMinutes for some consumers, but not all the consumers. Anyway, remove rows with TotalMinutes less than 720 (half day).

Data Processing – Daily Activity (Continue)

- Set a filter such that only rows with 0 TotalSteps value will be displayed. Noticed that there were quite a few rows with 0 value in the all the following fields: TotalSteps, TotalDistance, VeryActiveDistance, ModeratelyActiveDistance, LightlyActiveDistance, VeryActiveMinutes, FairlyActiveMinutes, and LightlyActiveMinutes – most of these rows have 1000+ in their Calories field, but few of these rows (3) just have 0 in their Calories filed. Looks like there is inconsistent way to calculate calorie consumed by each consumer. According to an article, people are still burn calories while they are in sedentary state or even sleep to maintain basic bodily functions. Remove all rows with 0 Calories field.
- Note: hence "calories burned" is not a reliable way to evaluate the effectiveness of one's walking/exercise. We will fill filter out the consumers who have quite a few rows with 0 TotalSteps while having 1000+ in Calories fields in some analysis.

Here is the article cited above: https://www.health.harvard.edu/staying-healthy/burning-calories-without-exercise?fbclid=IwAR1wfcE9be0GsymB5JHQDPv0HBoYn_LaRT3CBNmHUOpj13hwKBktGBdu6nw#:~:text=It's%20true%3A%20just%20sitting%20on,up%20watching%20TV%20or%20reading

Data Processing – Heartbeat Rate

- Opened heartrate_seconds_merged.csv and save it as an Excel Workbook.
- The file contains over 1,000,000 heart rate data from 7 consumers, with multiple heart rate values collected in one minute.
- According this article Target Heart Rate and Estimated Maximum Heart Rate published by CDC, for **vigorous-intensity physical activity**, your **target heart rate** should be between **77%** and **93% of your maximum heart rate**. And to estimate your **maximum age-related heart rate**, **subtract your age from 220**. Assumed that the consumers in this dataset are between **20-60 years old** and have their **maximum age-related heart rate** between **200-160 bpm**. Therefore, the target heart rate for high-intensity physical activity should be between **186-154 bpm** (for a young person such as 20 years old) and between **148.8-123.2 bpm** (for an older person such as 60 years old). For simplicity, we checked **if any consumer's heart rate had ever exceeded 180 bpm.**
- Sorted the data by the Value column (i.e., heart rate) and performed conditional formatting on heart rate values **above 180 or below 40**.
- Turned on filtering. In the drop-down menu next to the column label of the Value column, selected only values above 180 or below 40. Copied the corresponding rows and pasted them into two new worksheets. One sheet (heartrate_sec_high) stores rows with heart rates above 180, and another sheet (heartrate_sec_low) stores rows with heart rates below 40.
- In the worksheet heartrate_sec_high, created a PivotTable. Aggregated data by consumer – there were 4 consumers with heart rate over 180. Calculated average heart rate by consumer. Selected some consumers for further analysis.

Data Processing – Weight Watching

- Opened weightLogInfo_merged.csv and saved it as an Excel Workbook.
- This file contains the weight information (such as WeightKg, WeightPounds, BMI, etc.) of 8 consumers. Let's join this file with the dailyActivity file so we may look at variables like steps taken, activity level, calorie burned, and weight changes together.
- Formatted the Date field to "Date/Short Date" format which does not include hour and minute information. That's the format ActivityDate field (in dailyActivity file) use.
- Used the SQL script in the following slide to perform **LEFT JOIN** first to get all the rows in dailyActivity file with WeightPounds and BMI information merged in, then got the fields (including TotalSteps, TotalDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, Calories, WeightPounds and BMI) we need. Saved the result to a **table** in the database.
- Applied the SQL script in the slide after next slide to the table obtained in the previous step. The result only contains rows related to the 7 consumers who have data in the weightLogInfo file (note: **we deleted rows related to consumer 4319703577 in dailyActivity_merge file due to data inconsistency issue** mentioned earlier). Saved the result to a **.csv** (dailyActivity_weightInfo_merged) file.
- In weightLogInfo file, used a PivotTable to count how many rows (or days) of data each consumer have. Among the 7 consumers, only **6962181067** and **8877689391** have more than 20 days of data. Deleted rows related to other consumers and saved the result to a new worksheet (dailyActivityWeightReduced).
- Used the **AVERAGEIF** function to calculate the average (excluding NULL cells) weight and BMI of the two consumers above. **Fill in the blank cells** in the above 2 consumers' data **with calculated values**.

Data Processing – Weight Watching

SQL Code to Join dailyActivity_merged.csv and WeightLogInfo_merged.csv

```
SELECT
  `klu0629.fitabase_data.dailyActivity_merged`.Id, ActivityDate, TotalSteps,
  TotalDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance,
  VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, Calories,
  WeightPounds, BMI
FROM `klu0629.fitabase_data.dailyActivity_merged`
LEFT JOIN `klu0629.fitabase_data.weightLogInfo_merged_v2`
ON `klu0629.fitabase_data.dailyActivity_merged`.Id =
  `klu0629.fitabase_data.weightLogInfo_merged_v2`.Id AND
  `klu0629.fitabase_data.dailyActivity_merged`.ActivityDate =
  `klu0629.fitabase_data.weightLogInfo_merged_v2`.Date
ORDER BY `klu0629.fitabase_data.dailyActivity_merged`.Id, ActivityDate
```

Data Processing – Weight Watching

SQL Code to Retrieve the Rows Associated with the 8 Consumers Who have data in WeightLogInfo.csv

```
SELECT
  *
FROM
`klu0629.fitabase_data.dailyActivity_WeightInfo_SQL_LeftJoin`
WHERE Id IN
  (
    SELECT DISTINCT(Id)
    FROM
      `klu0629.fitabase_data.dailyActivity_WeightInfo_SQL_LeftJoin`
    WHERE BMI IS NOT NULL
  )
ORDER BY Id, ActivityDate
```

Data Processing – Sleep Duration & Sleep Quality

- Opened sleepDay_merged.csv and saved as Excel workbook.
- Removed duplicates rows.
- Removed the rows related to the following consumers who had less than 15 days (50% of the 31 days) of data: 1644430081, 1844505072, 1927972279, 2320127002, 4020332650, 558609924, 677588955, 70077441712, 8053475328.

Data Analysis: Daily Activity Tracking

Made some plots to see how TotalSteps and Total Distance data are distributed.

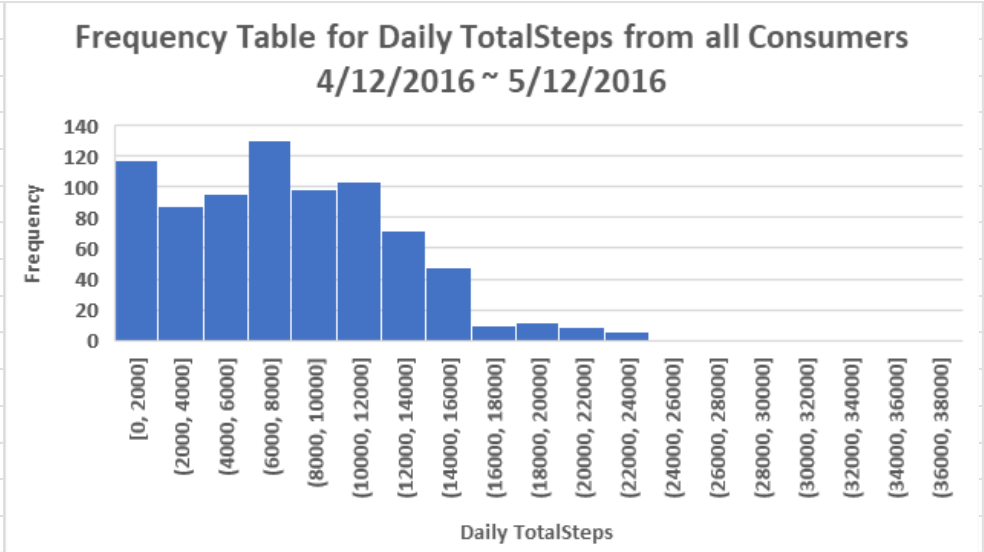
- The Daily TotalSteps histogram shows that the most common set of values for daily total steps are between **6,000 ~ 8,000** for consumers in this dataset. There is a long tail in the right side of the plot.
- The Daily TotalSteps boxplot shows that the **median value of daily total steps** is **7,396**. There are **some outliers above upper whisker** means there were some consumers who took much more steps than other consumers.
- The Daily TotalDistance histogram shows that the most common values for daily total distance are between **0 ~ 1.5 miles** in this dataset. The next common values are between **4.5** and **6 miles** for consumers in this dataset. There is a long tail in the right side of the plot.
- The Daily TotalDistance boxplot shows that the **median daily total distance** value is **5.19 miles**. There are **quite a few outliers above the upper whisker** means there were quite a few consumers who walked much longer in distance than other consumers.

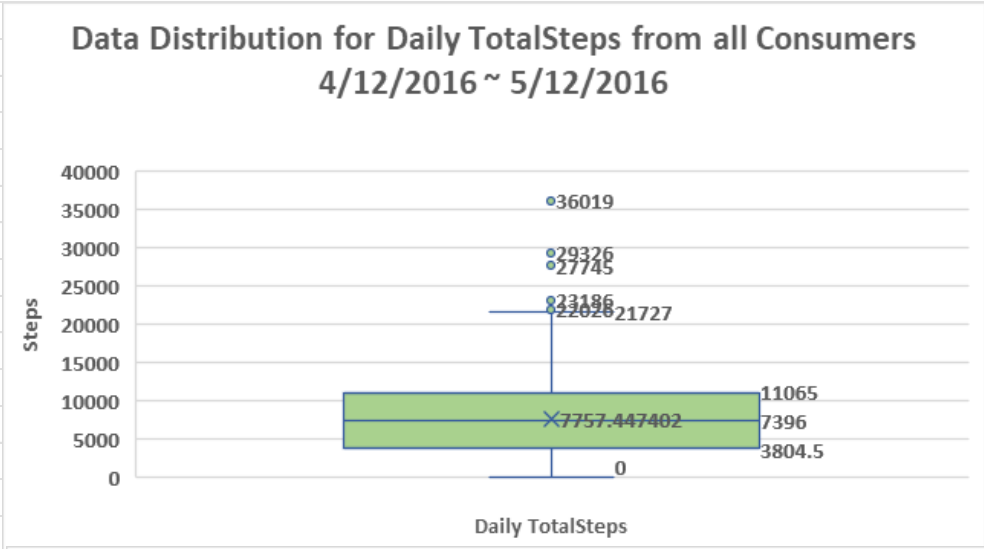
Data Analysis: Daily Activity Tracking

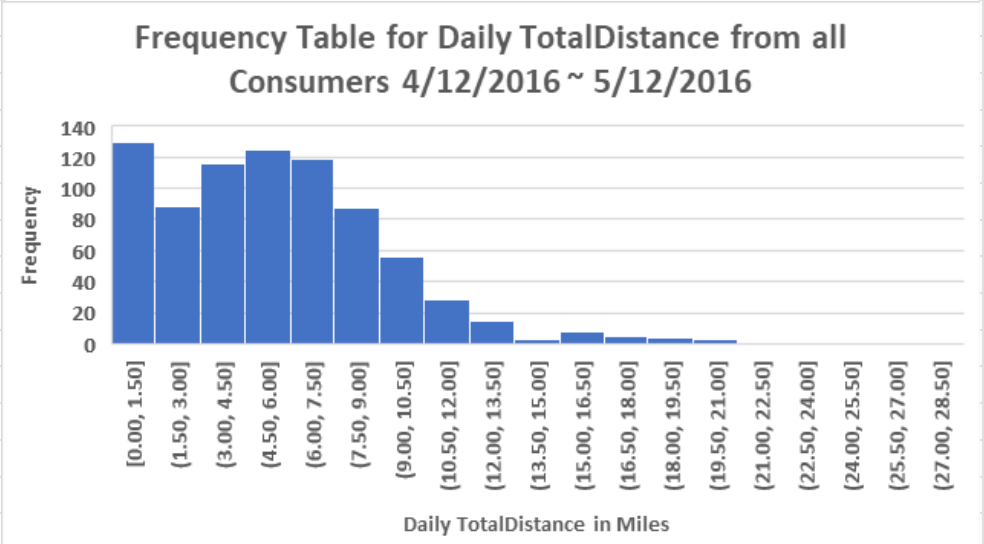
Made some plots to see how TotalSteps and Total Distance data are distributed.

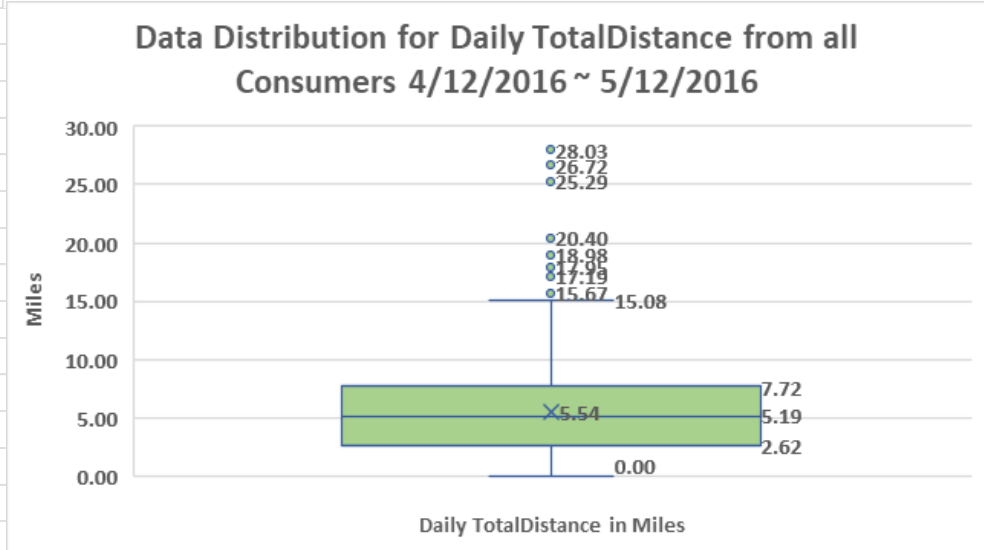
- The Daily TotalSteps histogram shows that the most common set of values for daily total steps are between **6,000 ~ 8,000** for consumers in this dataset. There is a long tail in the right side of the plot.
- The Daily TotalSteps boxplot shows that the **median value of daily total steps** is **7,396**. There are **some outliers above upper whisker** means there were some consumers who took much more steps than other consumers.
- The Daily TotalDistance histogram shows that the most common values for daily total distance are between **0 ~ 1.5 miles** in this dataset. The next common values are between **4.5** and **6 miles** for consumers in this dataset. There is a long tail in the right side of the plot.
- The Daily TotalDistance boxplot shows that the **median daily total distance** value is **5.19 miles**. There are **quite a few outliers above the upper whisker** means there were quite a few consumers who walked much longer in distance than other consumers.

Data Analysis: Daily Activity Tracking









Data Analysis: Daily Activity Tracking

Made some plots to see how VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes and SedentaryMinutes data are distributed.

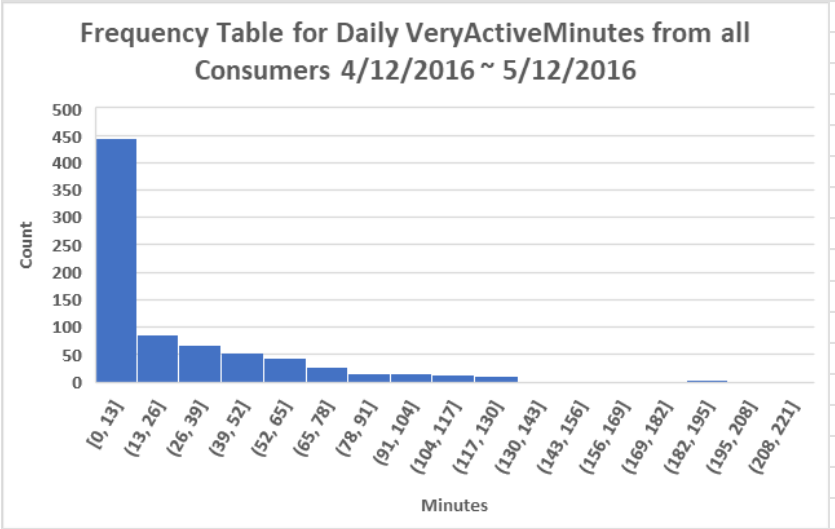
- The Daily VeryActiveMinutes histogram shows that the most common daily VeryActiveMinutes values are between **0 ~ 13** minutes for consumers in this dataset.

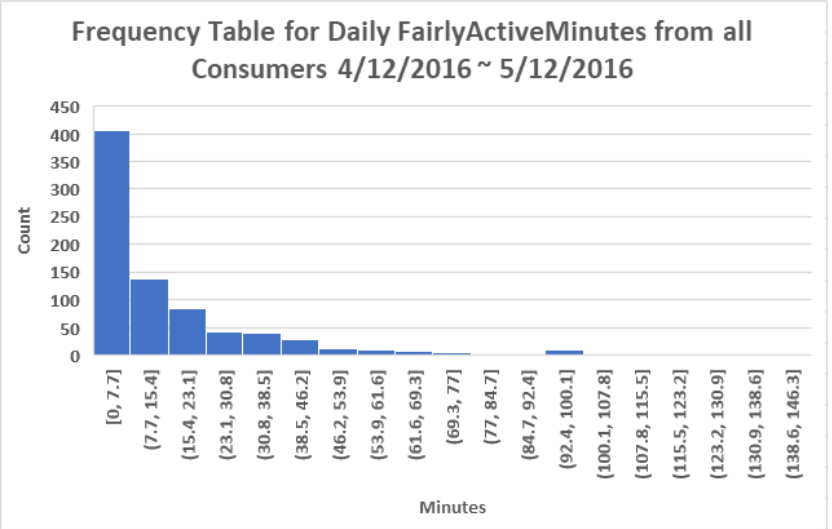
- The Daily VeryActiveMinutes boxplot shows that the **median daily VeryActiveMinutes** value is **6**. There are **some outliers above the upper whisker** means there were some consumers having much longer VeryActiveMinutes than other consumers.

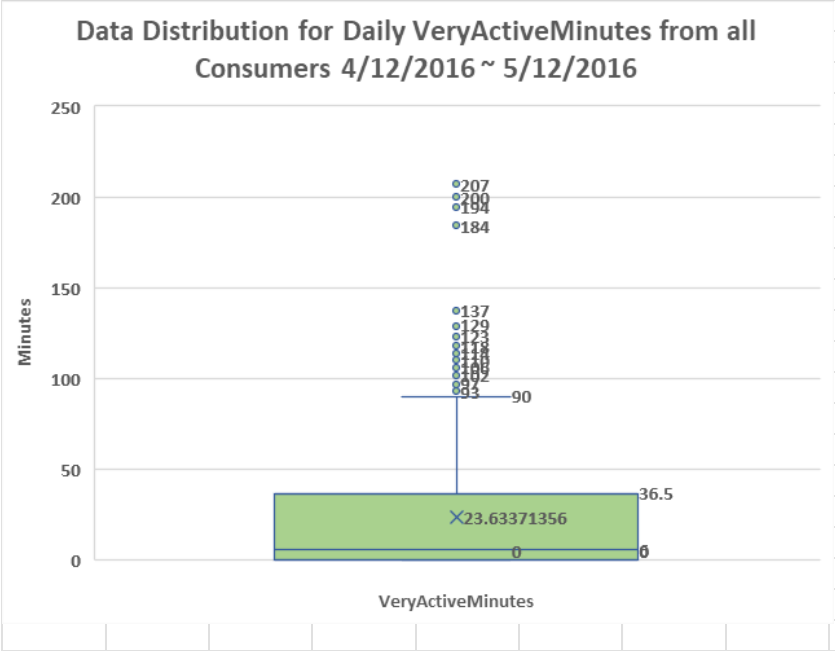
- The Daily FairlyActiveMinutes histogram shows that the most common values for daily FairlyActiveMinutes are between **0 ~ 7.7 minutes.**

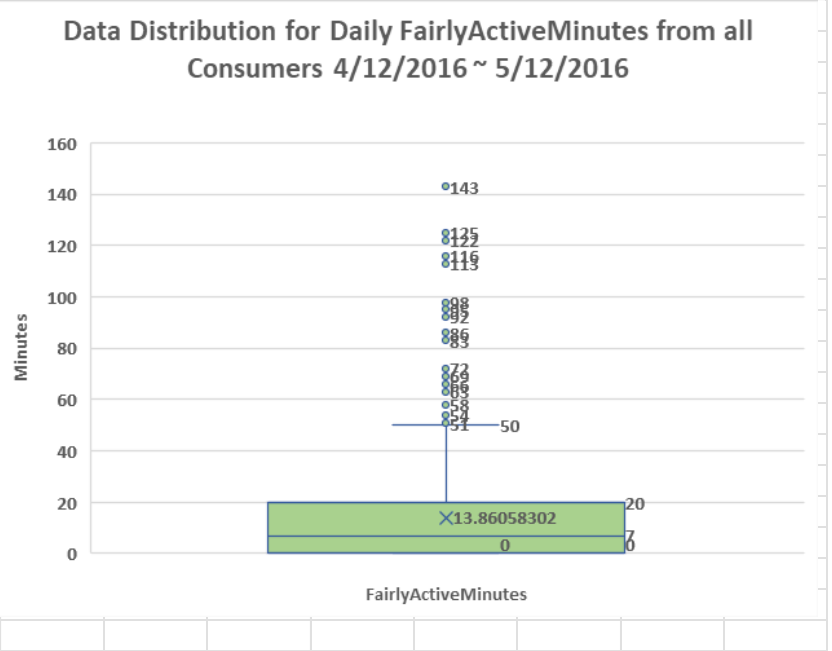
- The Daily FairlyActiveMinutes boxplot shows that the **median daily FairlyActiveMinutes** value is **7** minutes. There are **quite a few outliers above the upper whisker** means there were quite a few consumers having much longer FairlyActiveMinutes than other consumers.

Data Analysis: Daily Activity Tracking









Data Analysis: Daily Activity Tracking

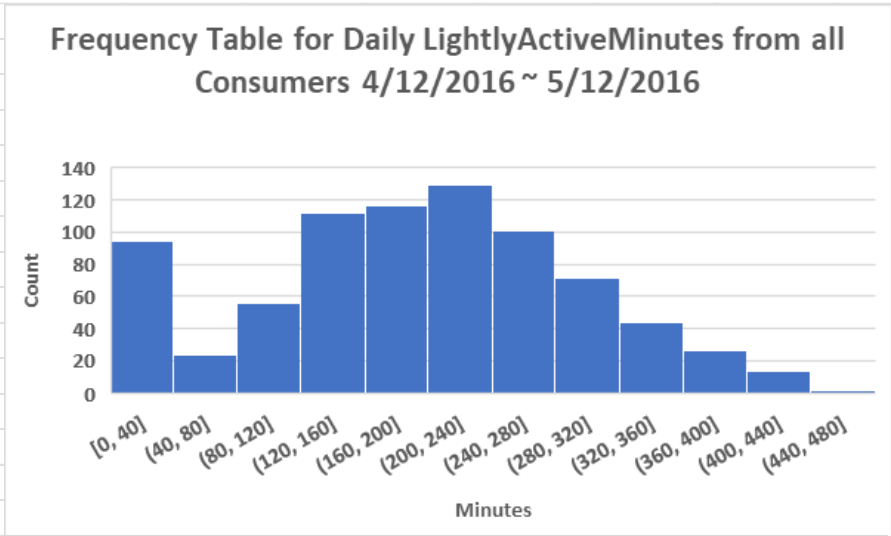
- The Daily LightlyActiveMinutes histogram shows that the most common daily LightlyActiveMinutes values are between **200 ~ 240** minutes for consumers in this dataset.

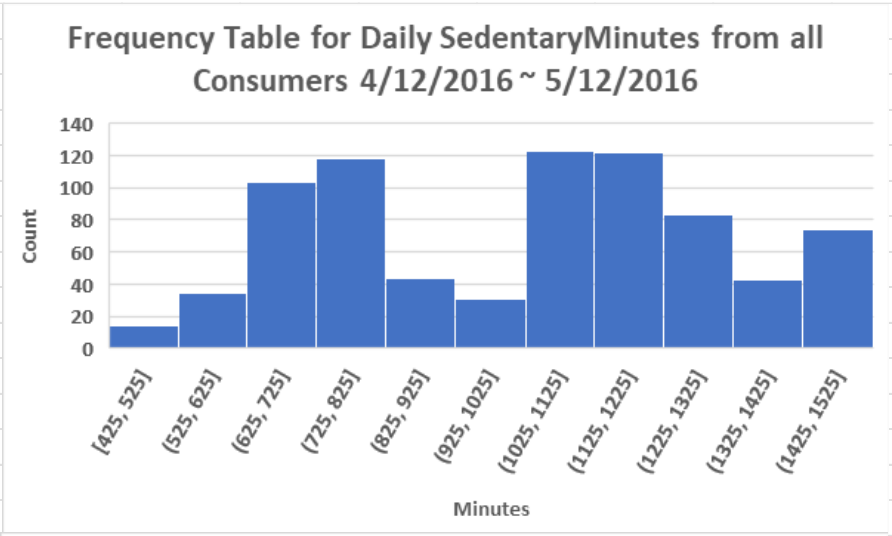
- The Daily LightlyActiveMinutes boxplot shows that the **median daily LightlyActiveMinutes** value is **199**.

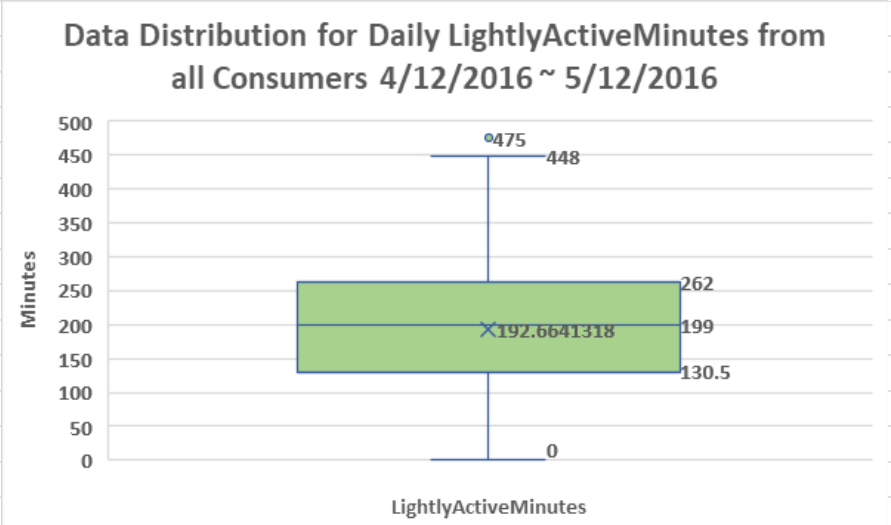
- The Daily SedentaryMinutes histogram shows that the most common values for daily SedentaryMinutes are between **1,025 ~ 1,125 minutes.**

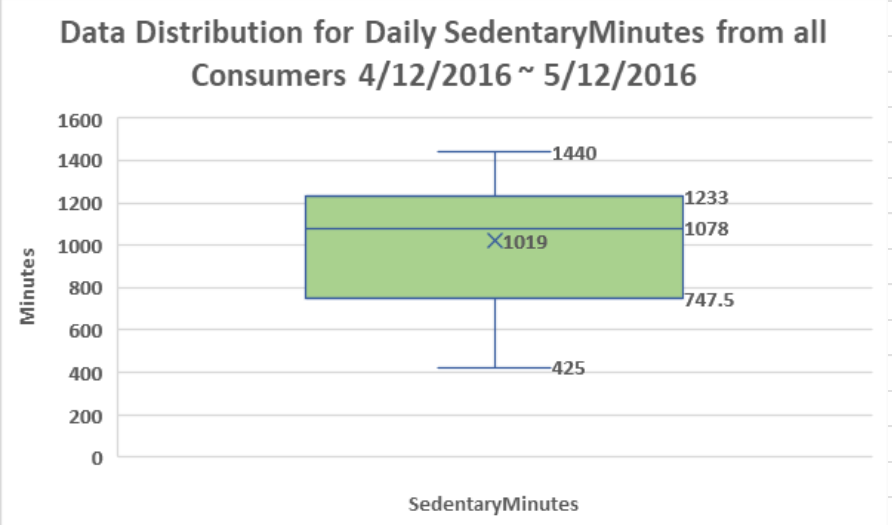
- The Daily SedentaryMinutes boxplot shows that the **median daily SedentaryMinutes** value is 1,078 minutes.

Data Analysis: Daily Activity Tracking





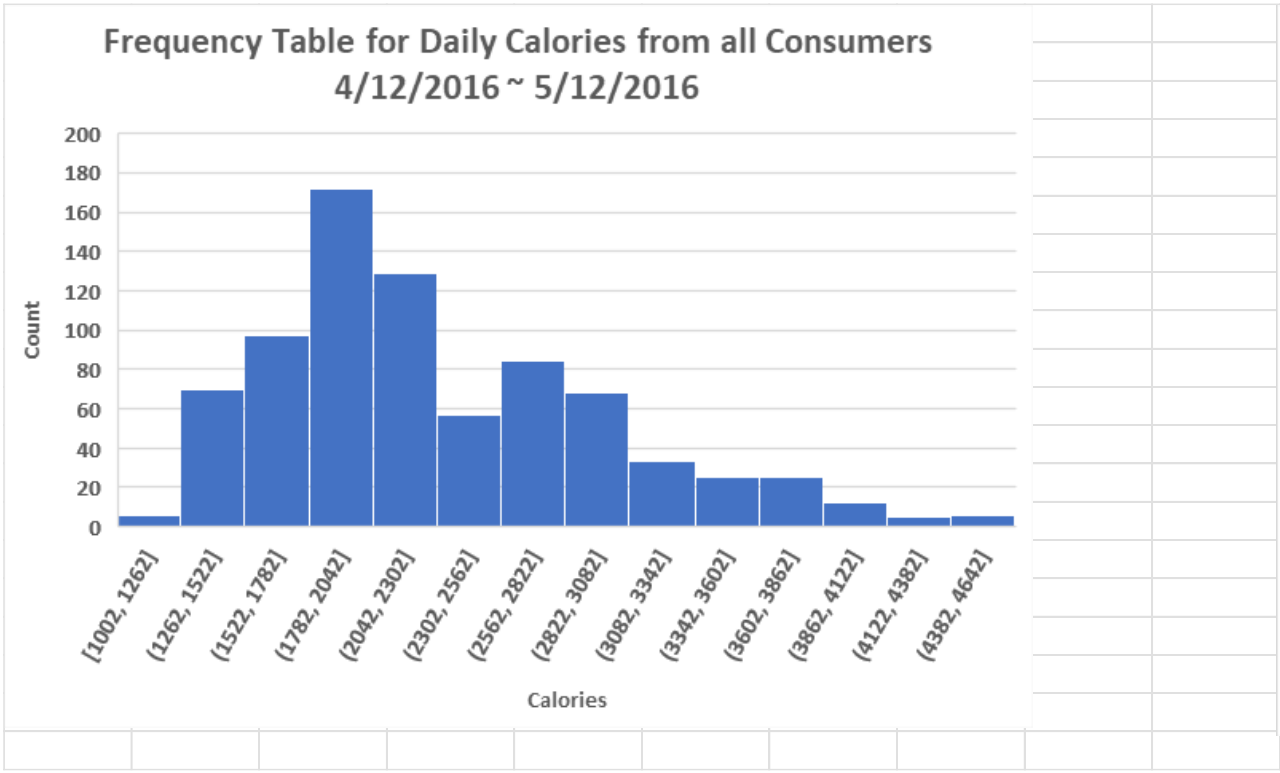


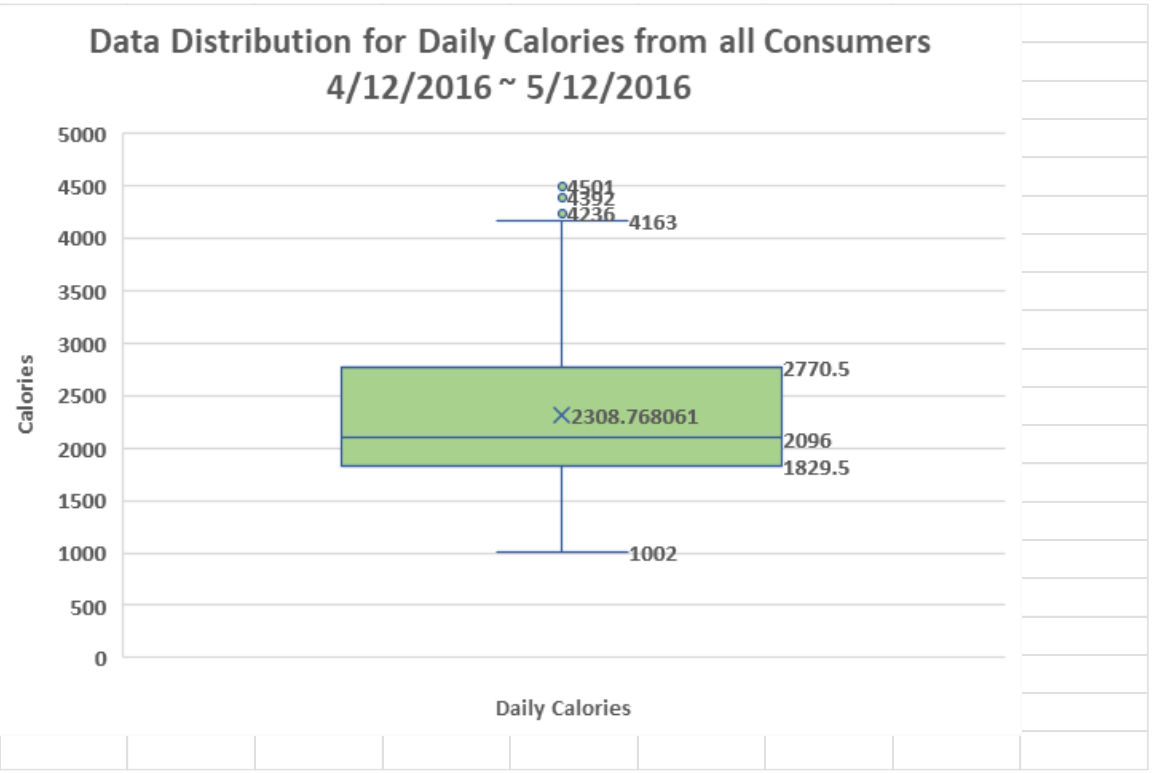


Data Analysis: Daily Activity Tracking

- Made some plots to see how Calories data are distributed.
  - The Daily Calories histogram shows that the most common set of daily Calories values are between **1,782 ~ 2,042 calories** for consumers in this dataset.
  - The Daily Calories boxplot shows that the **median daily calories** value is **2,096**. There are **some outliers above the upper whisker**, which means there were some consumers who burned much more calories than others.

Data Analysis: Daily Activity Tracking





Data Analysis: Daily Activity Tracking

- **Sorted** the worksheet by the **Calories** column and found the largest number in the column was **4,552**. This data related to consumer 5577150313.
- Input other numbers tracked for the day such as **TotalSteps**, **TotalDistance,** total walking time per activity level, and **weight** information (from another file weightLogInfo_merge.csv) into two **steps-to-calories converters** Walking Calorie Calculator & How to Count and Track Calories Burned Walking to get estimated values which were between **1,800 to 2,300.**
- Added **1,819** to those numbers. The calorie values got (**3,619 or 4,319**) are **still less than 4,552**. 1,819 is the calorie value found in the data tracking the activity of consumer 5577150313 on 5/7/2016 - the consumer did not walk that day (total steps and total distance were both 0). 1,819 could be the basal metabolic rate (or the energy/calories) for consumer 5577150313 (means the consumer needs that number of calories to perform basic body functions).
- Another consumer's daily activity data was input into the same calorie calculators mentioned above and the calculator's estimated calories were found to be lower than the calorie values found in this dataset.
- According to this article 6 Factors That Can Affect How Many Calories You Burn and this article What Affects How Many Calories You Burn? 6 Factors to Consider, factors like body weight, **muscle mass**, **age**, etc. can affect how much calories you burn. **No information on muscle mass or age was found in this dataset, so we were limited in performing the following tasks:**
  - **determining the accuracy of the calorie values in this dataset**
  - **discovering the relationship between calorie burned and other attributes tracked here** (like TotalSteps, TotalDistance, walking distance per activity level, time spent per activity level, etc.).

  Therefore, **we didn't do much analysis of calories in this report**.

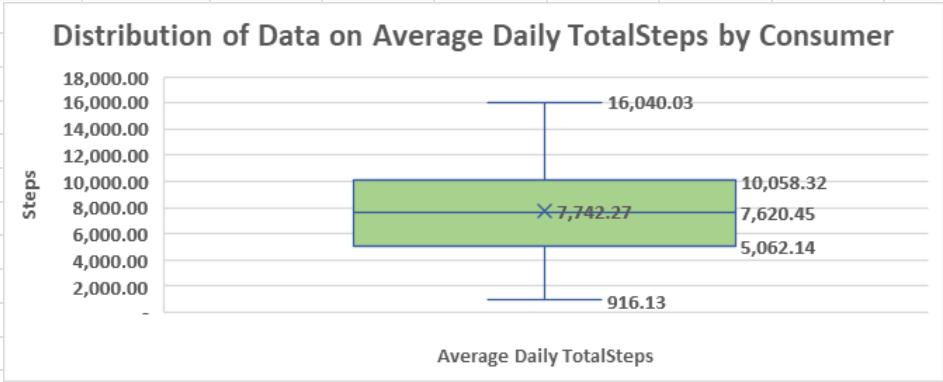
Data Analysis: Daily Activity Tracking

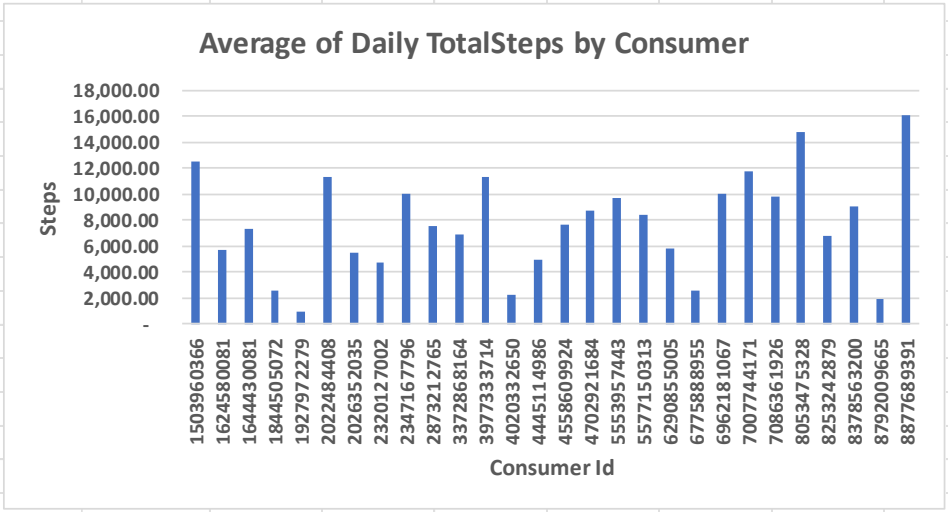
- Created a PivotTable. Aggregated data by consumer and calculated the average of daily TotalSteps for each consumer.
- Made a boxplot to check the distribution of data on the average of daily TotalSteps taken by each consumer and found that the **3rd quartile is around 10,058** which means nearly **75% of the consumers walked fewer than 10,000 steps a day**.
- Conditional formatting the Average of TotalSteps column to display the top 20% and bottom 20% values in different colors.

Data Analysis: Daily Activity Tracking

Row Labels	Average of TotalSteps	Average of TotalDistance
1503960366	12,520.63	8.07
1624580081	5,743.90	3.91
1644430081	7,282.97	5.30
1844505072	2,575.96	1.70
1927972279	916.13	0.63
2022484408	11,370.65	8.08
2026352035	5,456.07	3.39
2320127002	4,716.87	3.19
2347167796	10,077.18	6.73
2873212765	7,555.77	5.10
3372868164	6,861.65	4.71
3977333714	11,337.62	7.76
4020332650	2,267.23	1.63
4445114986	4,930.83	3.34
4558609924	7,685.13	5.08
4702921684	8,766.07	7.11
5553957443	9,676.31	6.34
5577150313	8,451.55	6.32
6290855005	5,851.32	4.43
6775888955	2,541.80	1.83
6962181067	10,001.73	6.73
7007744171	11,776.36	8.34
7086361926	9,766.07	6.67
8053475328	14,763.29	11.48
8253242879	6,842.28	4.93
8378563200	9,088.14	7.21
8792009665	1,919.93	1.23
8877689391	16,040.03	13.21
Grand Total	**7,757.45**	**5.54**

12,520.63
5,743.90
7,282.97
2,575.96
916.13
11,370.65
5,456.07
4,716.87
10,077.18
7,555.77
6,861.65
11,337.62
7,685.13
4,930.83
9,676.31
8,766.07
8,451.55
5,851.32
2,541.80
10,001.73
11,776.36
9,766.07
14,763.29
6,842.28
9,088.14
1,919.93
16,040.03





Data Analysis: Daily Activity Tracking

- Used a nested IF function to classify the “Activity Level” for each consumer based on the following guideline on steps and activity levels described in this article How Many Steps a Day Is Considered Active?
  - **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
  - **Highly active:** More than 12,500 steps daily
- Counted the number of consumers at each activity level and created a pie chart. **25%** of the consumers were at “**Somewhat Active**” level and **25%** of the consumers were at “**Sedentary**” level.
- According to this article How many steps should people take per day?, CDC recommends that most adults aim for **10,000** steps per day for health benefits. For most people, this is the equivalent of about 8 kilometers, or **5** miles.

Data Analysis: Daily Activity Tracking

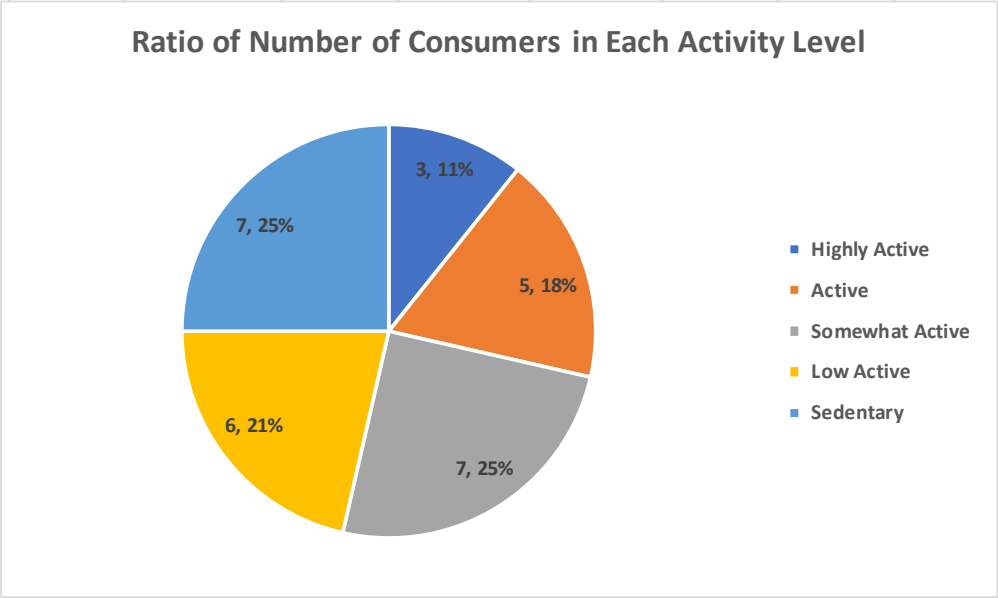
- Did some calculations and found that approximately **71%** of consumers in the dataset took **less than 10,000 steps** per day. According to this articles How many steps should people take per day?, most people in the United States only take 3,000–4,000 steps per day. Therefore, Bellabeat's marketing strategy should include **encouraging people to purchase and wear Bellabeat's smart devices to track/count the number of steps they take every day to ensure they achieve their goal of walking at least 10,000 steps a day.**

Data Analysis: Daily Activity Tracking

Row Labels	Average of TotalSteps	Activity Level
1503960366	12,520.63	Highly Active
1624580081	5,743.90	Low Active
1644430081	7,282.97	Low Active
1844505072	2,575.96	Sedentary
1927972279	916.13	Sedentary
2022484408	11,370.65	Active
2026352035	5,456.07	Low Active
2320127002	4,716.87	Sedentary
2347167796	10,077.18	Active
2873212765	7,555.77	Somewhat Active
3372868164	6,861.65	Low Active
3977333714	11,337.62	Active
4020332650	2,267.23	Sedentary
4445114986	4,930.83	Sedentary
4558609924	7,685.13	Somewhat Active
4702921684	8,766.07	Somewhat Active
5553957443	9,676.31	Somewhat Active
5577150313	8,451.55	Somewhat Active
6290855005	5,851.32	Low Active
6775888955	2,541.80	Sedentary
6962181067	10,001.73	Active
7007744171	11,776.36	Active
7086361926	9,766.07	Somewhat Active
8053475328	14,763.29	Highly Active
8253242879	6,842.28	Low Active
8378563200	9,088.14	Somewhat Active
8792009665	1,919.93	Sedentary
8877689391	16,040.03	Highly Active
Grand Total	**7,757.45**	

Activity Level	Count
Highly Active	3
Active	5
Somewhat Active	7
Low Active	6
Sedentary	7

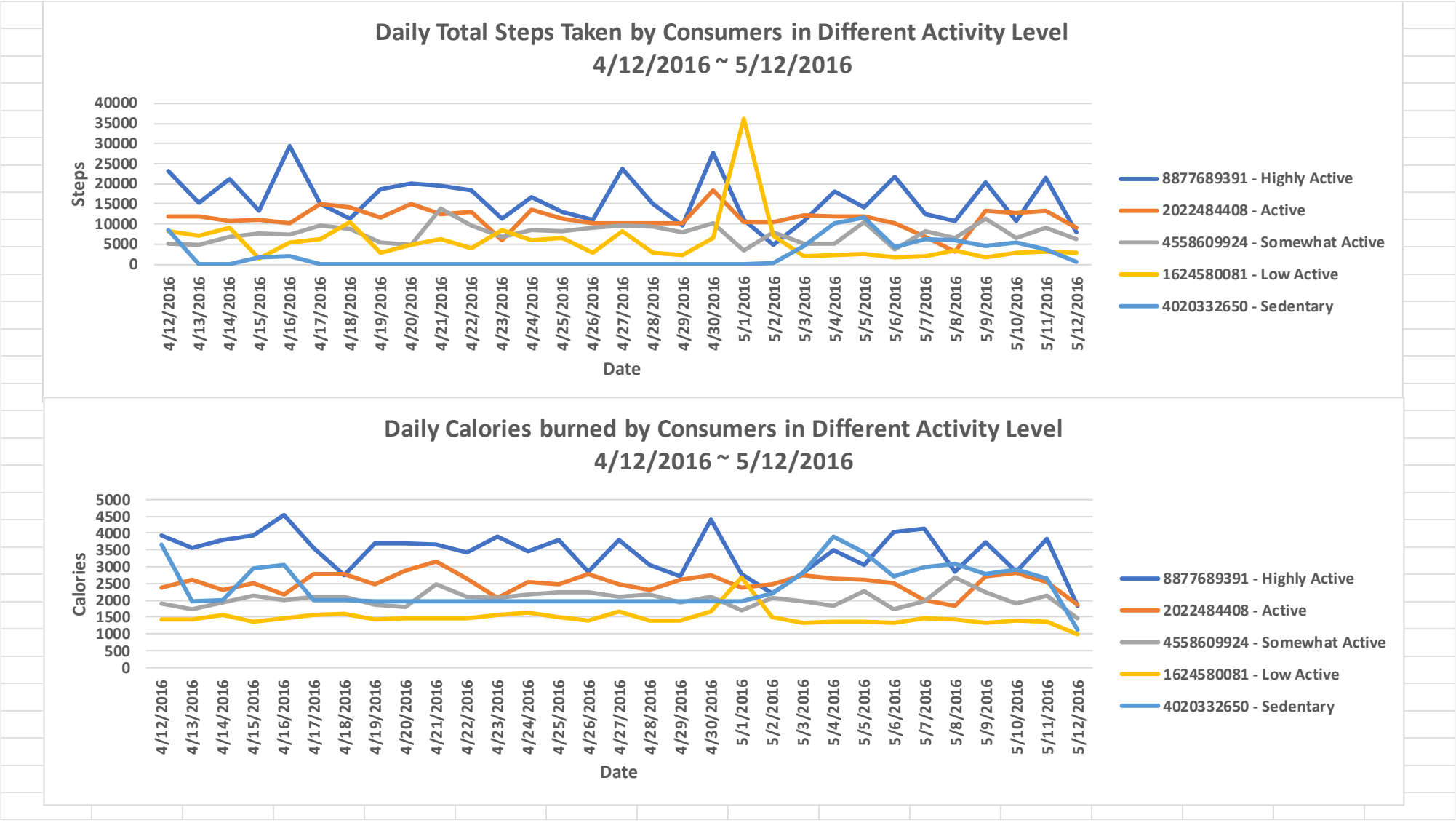
Ratio of Consumer with Average TotalSteps < 10,000 per day **0.714286**



Data Analysis: Daily Activity Tracking

- Selected one consumer for each activity level from those whose activity was tracked daily from 4/12/2016 to 5/12/2016.
- Created a line chart of the total daily steps taken by these consumers.
- Created a line chart of calories consumed by these consumers per day.
- From these line charts, we can see that there is a **positive correlation** between the number of steps taken and the calories burned.
- However, according to some articles like 6 Factors That Can Affect How Many Calories You Burn and What Affects How Many Calories You Burn? 6 Factors to Consider, factors such as **age**, **body mass**, **weight** etc. can affect the number of calories you burn. **A younger/heavier person may burn more calories than an older/lighter person while performing the same exercise**.
- What's interesting in the line chart is that a consumer with a "sedentary" activity level (ID 4020332650) consumed more calories than consumers with a "low active" or "somewhat active" level. Consumer 4020332650 might be younger or heavier. Therefore, Bellabeat's marketing strategy should include **encouraging people who need to burn more calories (such as those who are heavier or older) to purchase and wear Bellabeat's smart devices to track calorie burned for health or other benefits** (e.g., looking younger, slimmer or more energetic).

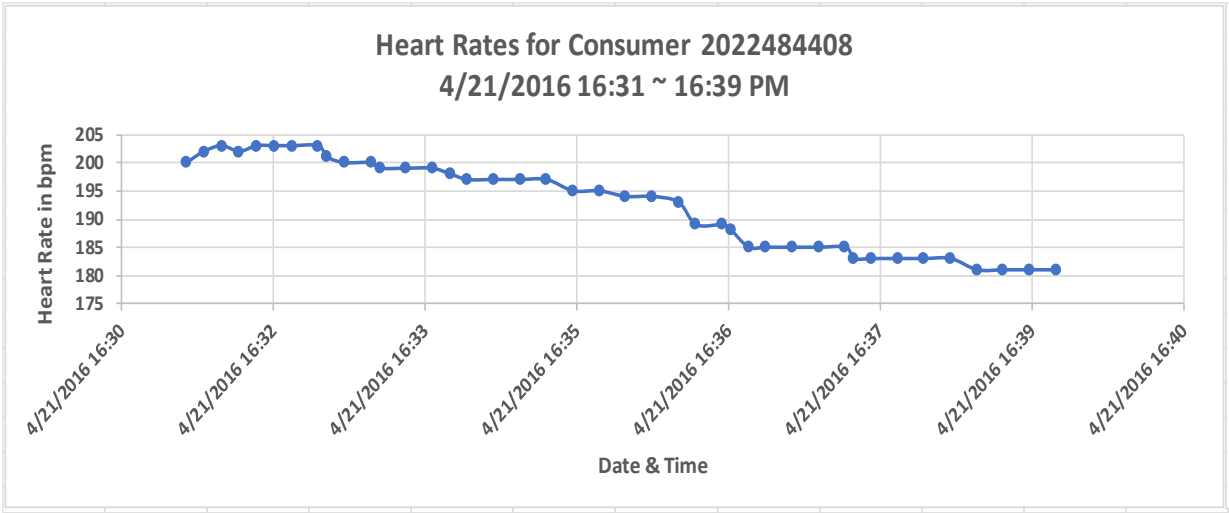
Data Analysis: Daily Activity Tracking

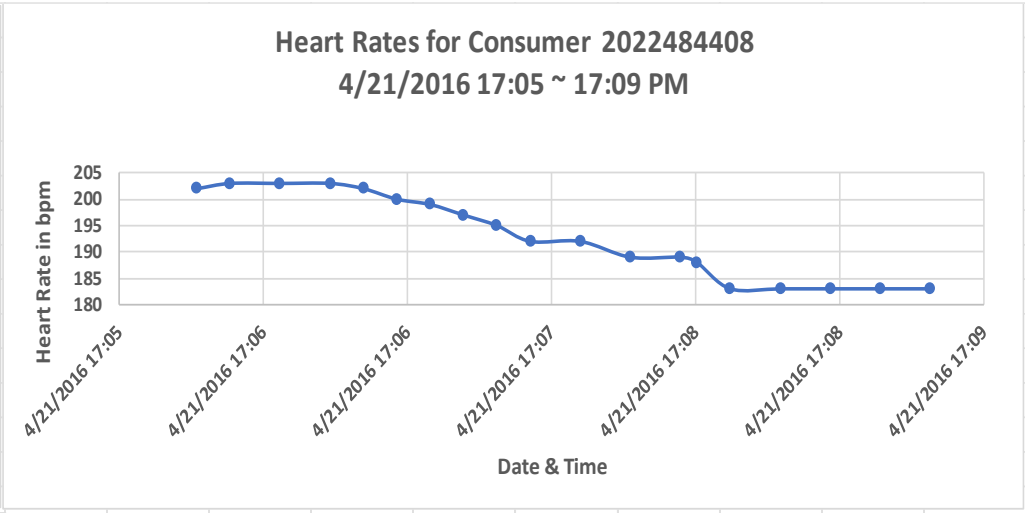


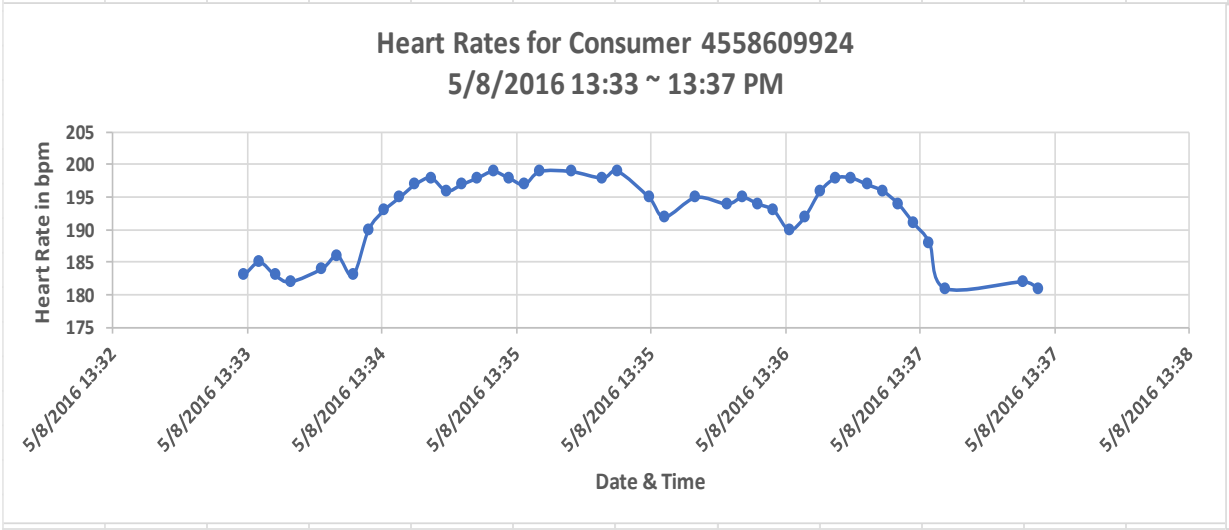
Data Analysis: Heart Rate Monitoring

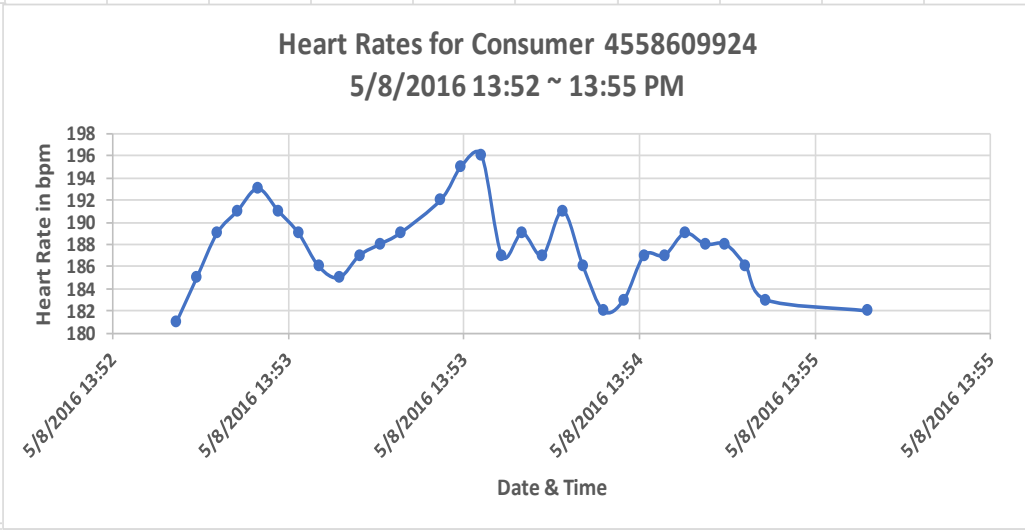
- In the worksheet heartrate_sec_high, created a PivotTable. Aggregated data by consumer – there were 4 consumers with heart rate over 180 bpm. Calculated average heart rate by consumer.
- In heartrate_sec_high, sorted data first by **Id** column, then by **Time** column.
- The above consumers were observed to have excessively high heart rates for several consecutive minutes on certain dates :
  - Consumer 2022484408 experienced high heart rates on two separate days – two minutes on 4/15/2016, and nine and five minutes on 4/21/2016. Select the date with the longest duration of high heart rate and made two charts with time and heart rates.
  - Consumer 4558609924 experienced high heart rates for five and four minutes on 5/8/2016. Created two charts with time and heart rates.
  - Consumer 4020332650 experienced high heart rates for four minutes on 4/15/2016. Created a chart with time and heart rates.
  - Consumer 2347167796 experienced high heart rates for two minutes on 4/19/2016. Created a chart with time and heart rates.
- Therefore, Bellabeat's marketing strategy should include **encouraging people who need to monitor their heart rate (such as people with tachycardia) to purchase and wear Bellabeat smart devices to monitor their heart rate during walking, exercise, etc.**

Data Analysis: Heart Rate Monitoring

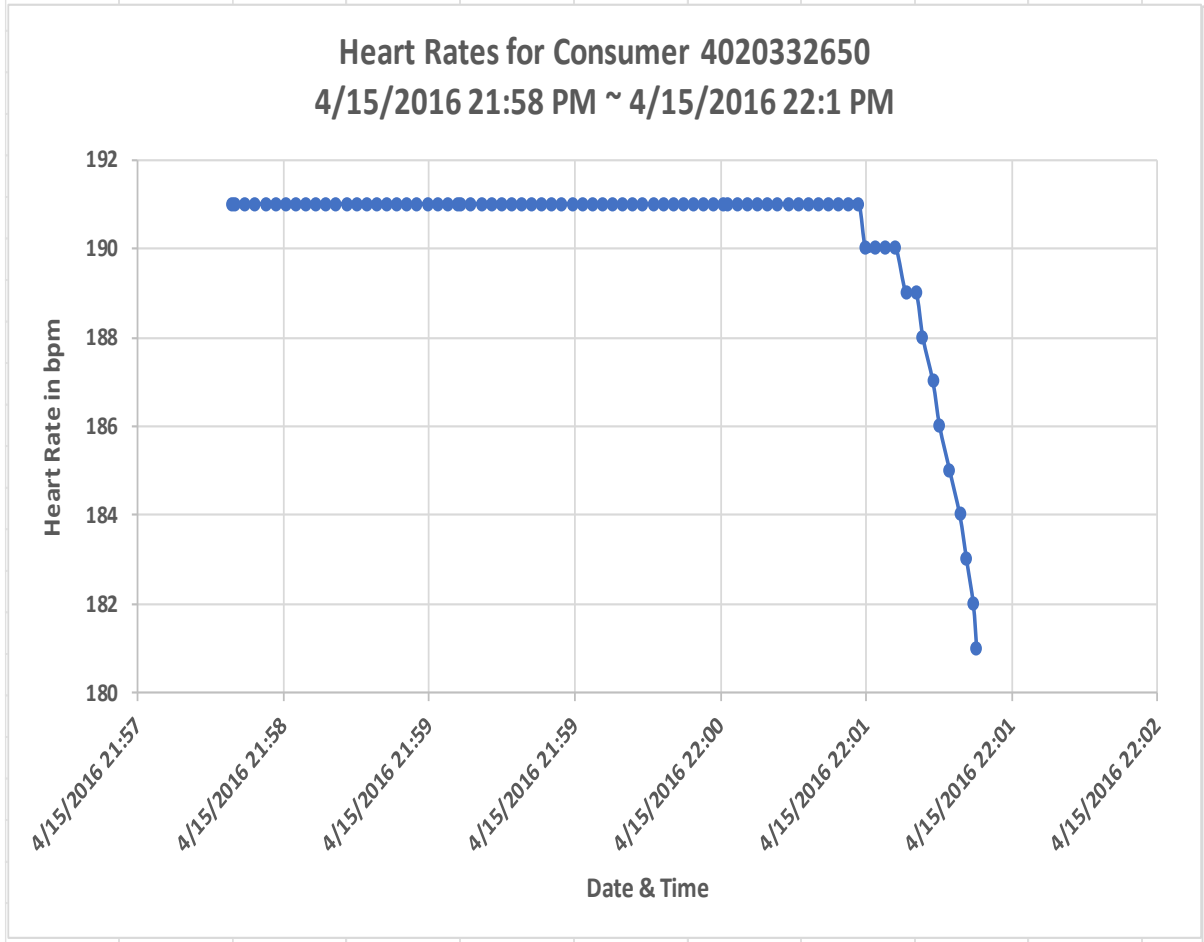


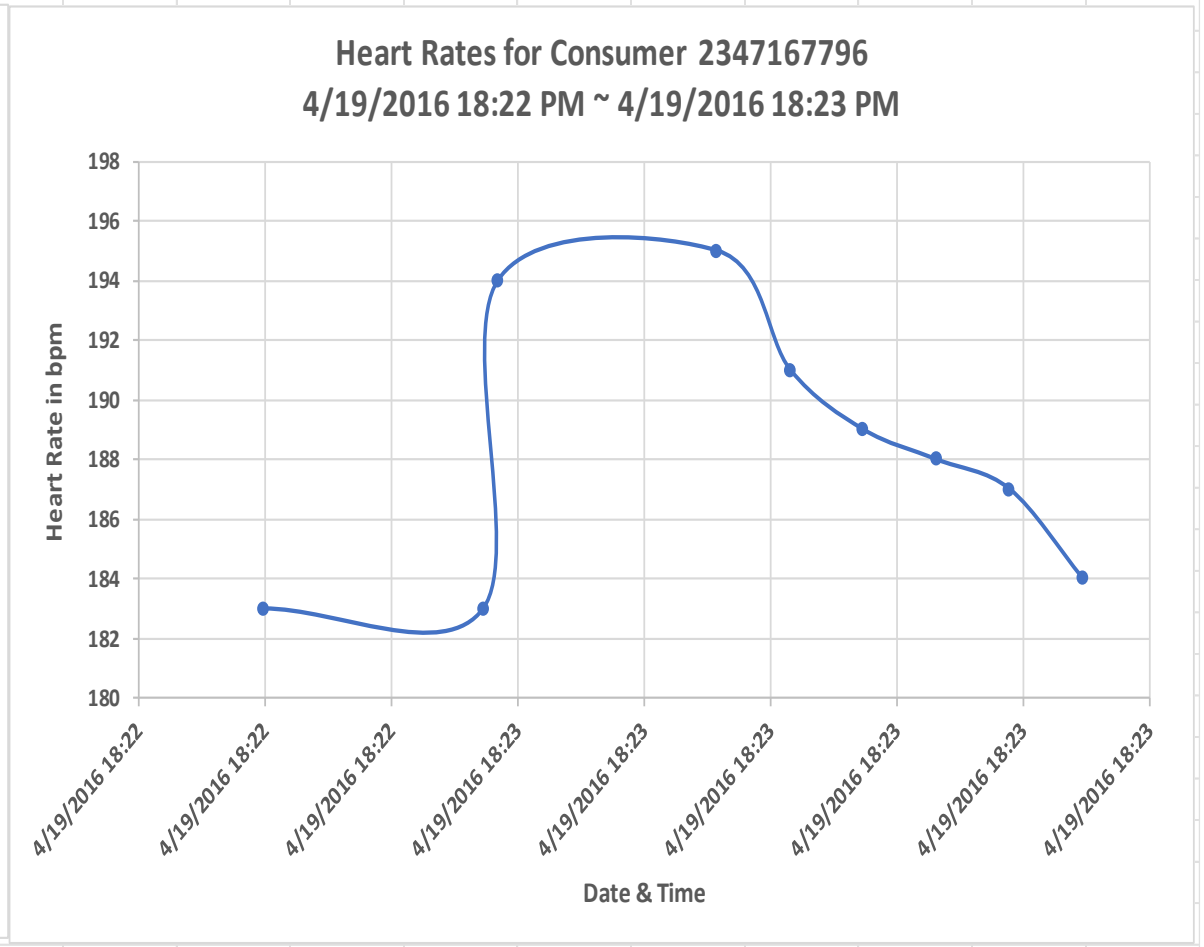






Data Analysis: Heart Rate Monitoring





Data Analysis: Weight Watching

- Created a PivotTable in weightLogInfo file. Aggregated data by consumer and got the average BMI values for all the consumers in this file. Named this worksheet **PivotBMIWeightStatus**. Categorized each consumer's **weight status** using the following guidelines described in the article About Adult BMI published by the CDC:

BMI	Weight Status
Below 18.5	Underweight
18.5 – 24.9	Healthy Weight
25.0 – 29.9	Overweight
30.0 and Above	Obesity

- Created a Pivot Table in dailyActivity_weightInfo_merged file. Aggregated data by consumer and got the average of daily total steps taken by each consumer. Added a column "**Activity Level**" next to the PivotTable. Categorized the activity level (as described earlier in this document) for each consumer.

Data Analysis: Weight Watching

- Copied the PivotBMIWeightStatus worksheet from weightLogInfo_merged file to dailyActivity_weightInfo_merged file. Added a new column "**Weight Status**" next to the "Activity Level" column created in the previous step. Used VLOOKUP function to get the weight status value from the lookup table in PivotBMIWeightStatus worksheet.

- Checked the activity level of the consumers whose weight status were marked as either "**Obesity**" or "**Overweight**". Consumer **1927972279** (whose activity level is "Sedentary" and weight status is "Obesity"), consumer **4558609924** (whose activity level is "Somewhat Active", and weight status is "Overweight") and consumer **5577150313** (whose activity level is "Somewhat Active", and weight status is "Overweight") **may need special reminders to take more steps** or **do more other exercises** each day so that **their weight status won't become an issue**.

Data Analysis: Weight Watching

Row Labels	Average of TotalSteps	Activity Level	Weight Status
1503960366	12520.63333	Highly Active	Healthy Weight
1927972279	916.1290323	Sedentary	Obesity
2873212765	7555.774194	Somewhat Active	Healthy Weight
4558609924	7685.129032	Somewhat Active	Overweight
5577150313	8451.551724	Somewhat Active	Overweight
6962181067	10001.73333	Active	Healthy Weight
8877689391	16040.03226	Highly Active	Overweight
Grand Total	**9008.802817**		

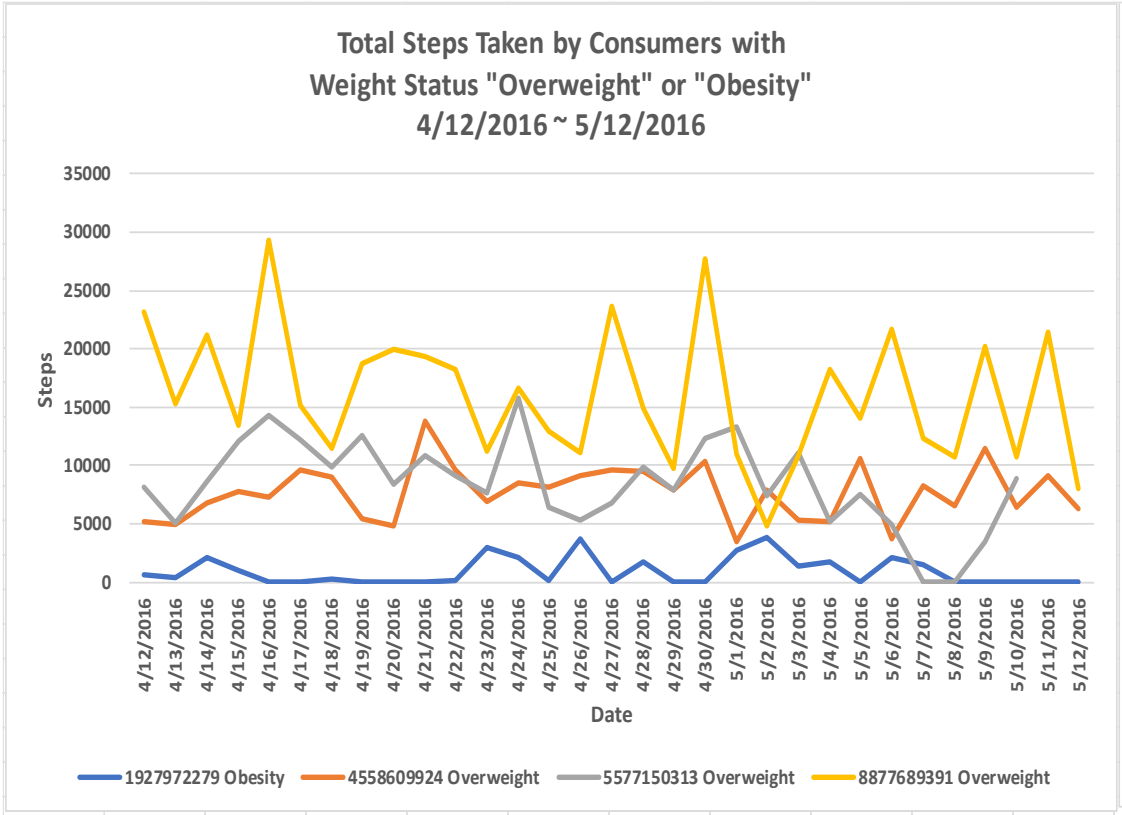
Data Analysis: Weight Watching

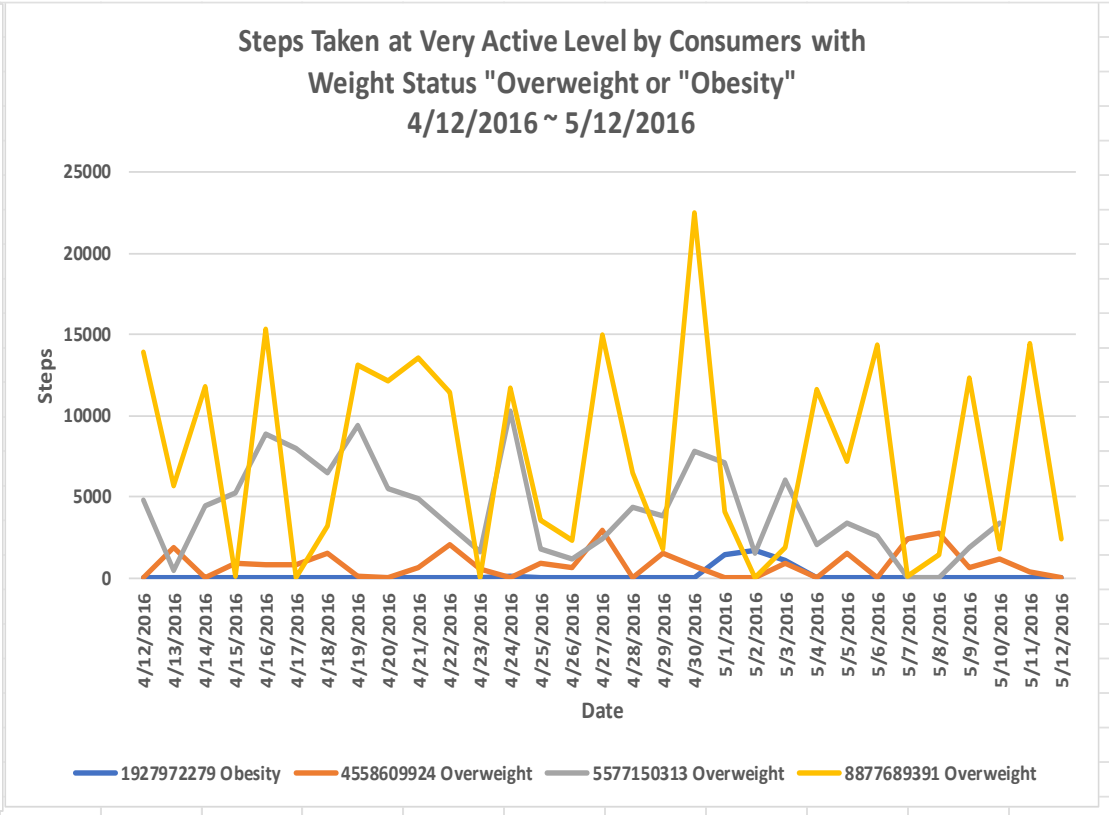
- According to this article "How many steps should people take per day?“, a 2018 analysis of 363 people with obesity found that people who walked **10,000 steps a day**, including **at least 3,500 steps** engaging in **moderate-to-vigorous activity** lasting 10 minutes or longer, had increased weight loss.
- Added a new column “**VeryActiveSteps**” in file dailyActivity_weightInfo_merged. Used an **IF function** and the following logic to calculate the value for VeryActiveSteps:
  - If TotalDistance is not 0: Divide VeryActiveDistance by TotalDistance, then multiply the result by TotalSteps.
  - If TotalDistance is 0: the value for VeryActiveSteps is 0.
- Researching deeper into the daily activity data of consumers with the weight status of either "**overweight**" or "**obesity**“ and found that **only one** consumer took serious action to lose weight by walking more than 10,000 steps a day, at least 3,500 of which were at a very active level for 10 minutes or longer.
- Therefore, Bellabeat's marketing strategy should include **identifying people who may have weight or BMI level concerns and** encouraging them to **purchase and wear Bellabeat’s smart devices to track their weight and BMI level.**

Data Analysis: Weight Watching

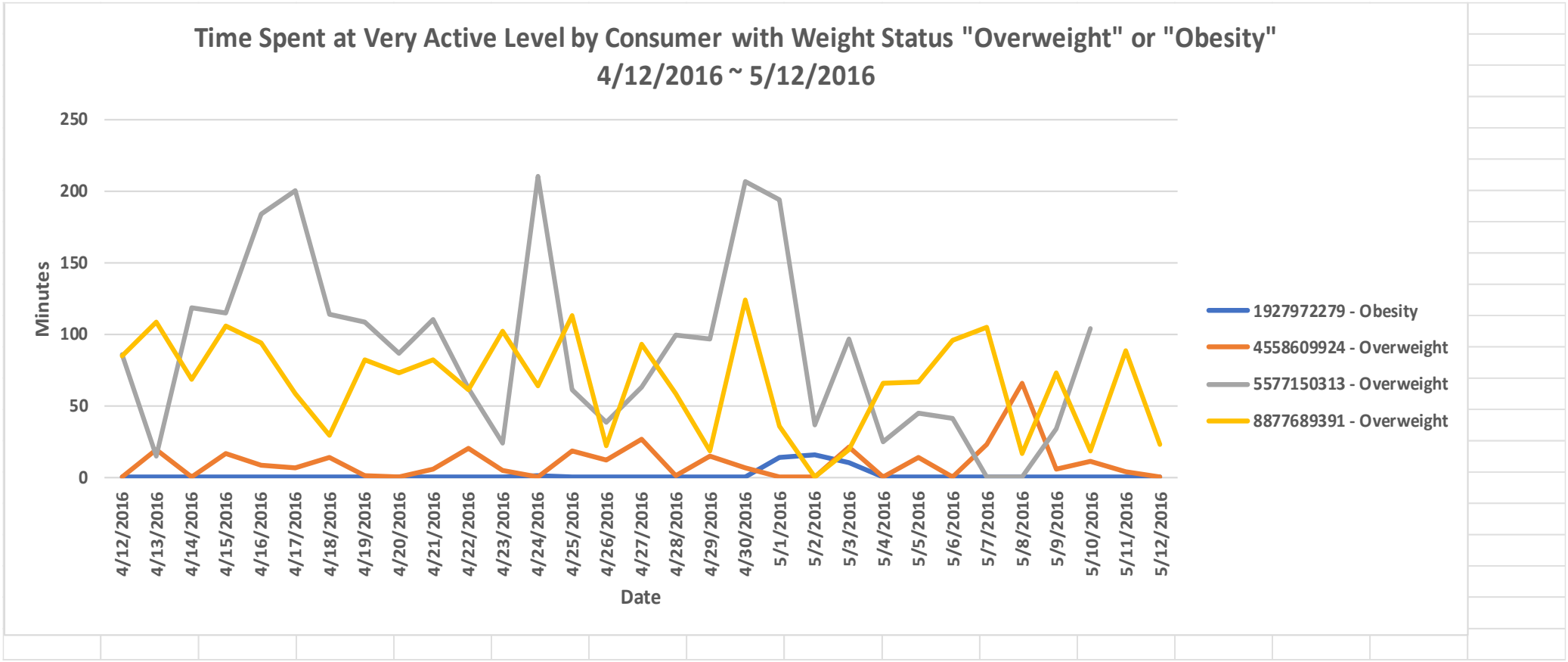
Row Labels	Average of TotalSteps	Average of VeryActiveDistance	Average of VeryActiveSteps	Average of VeryActiveMinutes	Average of WeightPounds	Average of BMI	Weight Status	Activity Level
1503960366	12,520.63	2.95	4,578.09	40.00	115.96	22.65	Healthy Weight	Highly Active
1927972279	916.13	0.10	138.44	1.32	294.32	47.54	Obesity	Sedentary
2873212765	7,555.77	0.68	993.11	14.10	125.66	21.57	Healthy Weight	Somewhat Active
4558609924	7,685.13	0.55	830.92	10.39	153.53	27.21	Overweight	Somewhat Active
5577150313	8,451.55	3.16	4,221.78	88.79	199.96	28.00	Overweight	Somewhat Active
6962181067	10,001.73	1.67	2,457.54	23.57	135.68	24.02	Healthy Weight	Active
8877689391	16,040.03	6.64	7,583.78	66.06	187.71	25.49	Overweight	Highly Active
Grand Total	**9,008.80**	**2.24**	**2,955.09**	**34.41**	**159.14**	**25.13**		

Data Analysis: Weight Watching





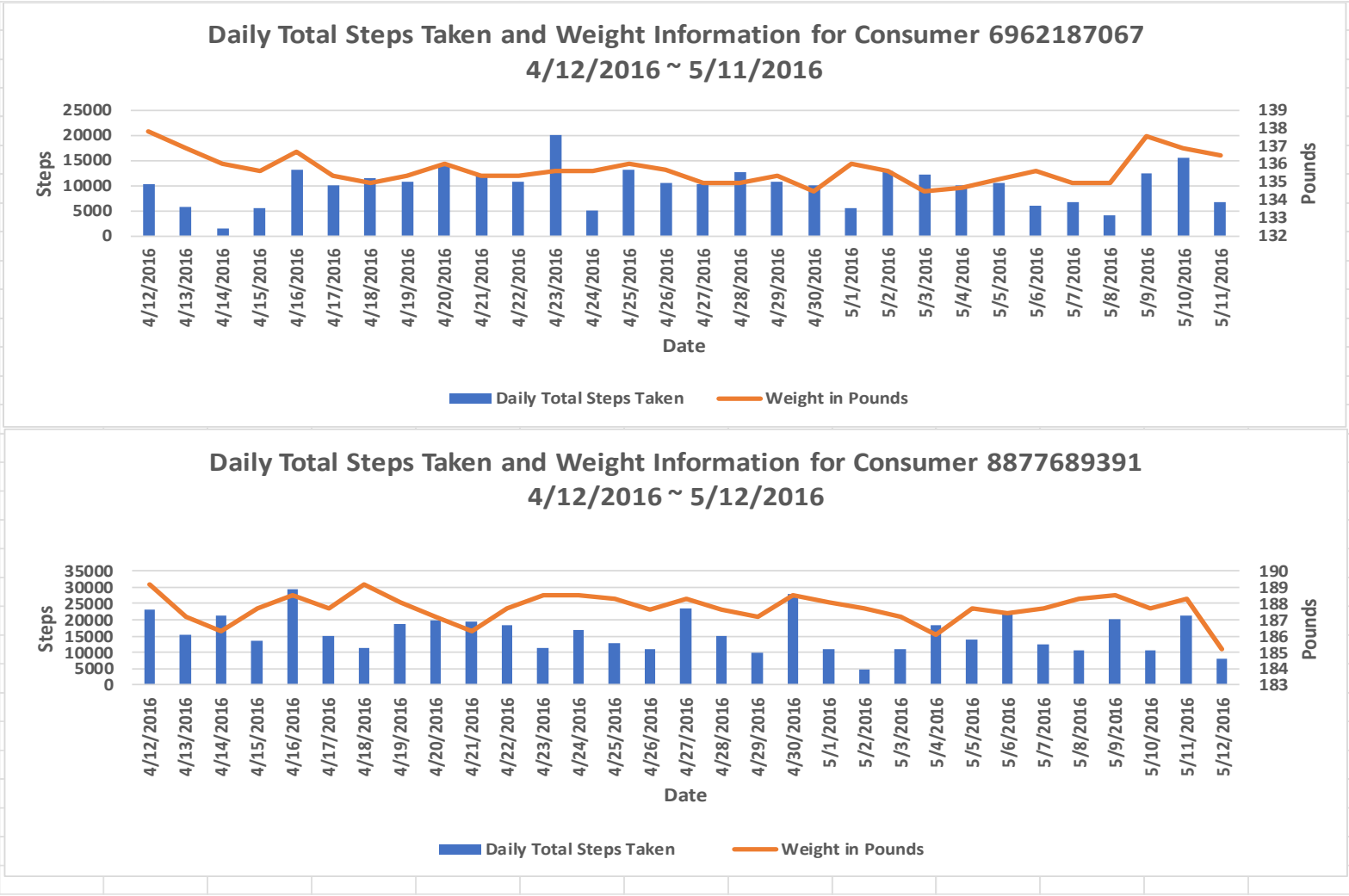
Data Analysis: Weight Watching



Data Analysis: Weight Watching

- Only 2 consumers (6962181067 and 8877689391) recorded their weight on more than 50% of the days between 4/12/2016 to 5/12/2016.
- Made **two** combo charts with total steps taken and weight information for the 2 consumers 6962181067 (whose activity level is “Active”, and weight status is “Healthy Weight”) and 8877689391 (whose activity level is “Highly Active”, and weight status is “Overweight”).
- Both consumers "seem" to be doing a good job controlling their weight. By the end of the data tracking period, both consumers had lost some weight. Roughly speaking, the more they walked, the lighter they weighed, and the less they walked, the heavier they weighed.
- However, other factors such as **food intake may affect the amount of weight a person loses**. **There were no data on caloric intake from food in this dataset**, so **we were limited in assessing the impact of consumers' walking on their weight control.**

Data Analysis: Weight Watching



Data Analysis: Sleep Duration and Sleep Efficiency Tracking

- Added 3 new columns: TotalHoursAsleep, ShortSleepDay, and SleepEfficiency. Here are explanations of how the values in these columns were derived:
  - **TotalHoursAsleep**: divide TotalMinutesAsleep by 60 to get TotalHoursAsleep.
  - **ShortSleepDay**: according to an article, experts recommend adults get at least **7** hours of sleep per night for better health. Consistently getting **less than 6 hours of sleep can have consequences for a person's health and quality of life**. Use **IF** function to determine whether the value in the TotalHoursAsleep column of the same row is less than 6 hours. If so, enter 1, otherwise enter 0.
  - **SleepEfficiency**: according to this web link sleep efficiency, sleep efficiency is the percentage of time spent asleep while in bed. It is calculated by dividing the amount of time spent asleep (in minutes) by the total amount of time in bed (in minutes). A normal sleep efficiency is **85%** or higher.

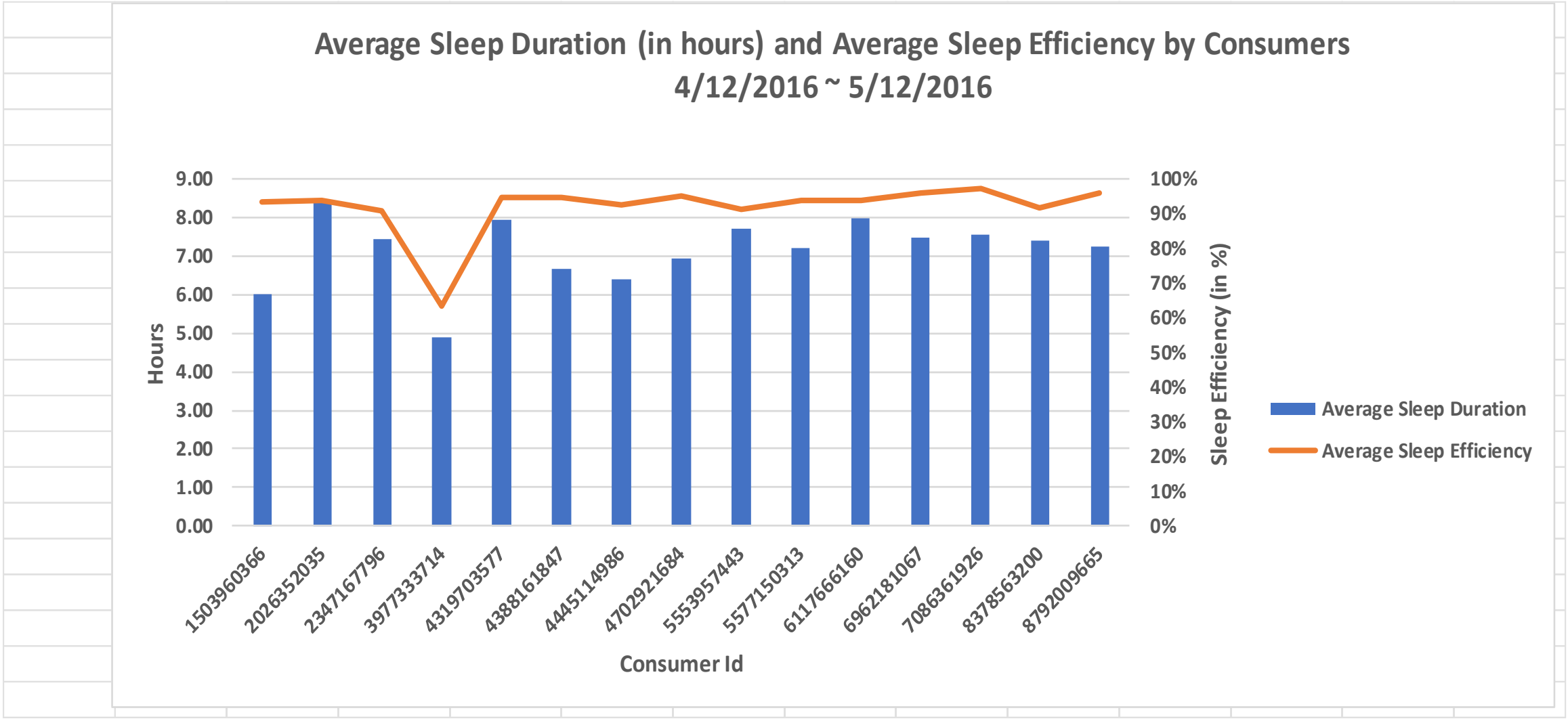
Data Analysis: Sleep Duration and Sleep Efficiency Tracking

- Created a Pivot Table and calculated the average of TotalHoursAsleep and SleepEfficiency by consumers. Also got the sum of the ShortSleepDay by consumers. Divided the sum of ShortSleepDay by the count of Id (number of rows associated with the consumer) and got percentage of short sleep day. The Pivot Table shows that 4 consumers suffered from short sleep duration (average sleep duration is less than 7 hours) - Consumer 3977333714 had an average sleep duration of 4 hours and her sleep efficiency is below 85%. Two consumers had short sleep duration (less than 6 hours of sleep per day) for more than 50% of the recording period – they consistently get less than 6 hours of sleep, which may have a negative impact on their health or quality of life.
- Created a combo chart which shows the average sleep hours and average sleep efficiency by consumers between 4/12/2016 and 5/12/2016. Consumer 3977333714 had both short sleep hours and sleep efficiency issues.

Data Analysis: Sleep Duration and Sleep Efficiency Tracking

Row Labels	Average of TotalHoursAsleep	Average of SleepEfficiency	Sum of ShortSleepDay	Count of Id	Percentage of Short Sleep Day
1503960366	6.01	94%	14	25	56%
2026352035	8.44	94%	1	28	4%
2347167796	7.45	91%	0	15	0%
3977333714	4.90	63%	24	28	86%
4319703577	7.94	95%	3	26	12%
4388161847	6.67	95%	6	23	26%
4445114986	6.42	93%	8	28	29%
4702921684	6.96	95%	4	27	15%
5553957443	7.73	91%	5	31	16%
5577150313	7.20	94%	3	26	12%
6117666160	7.98	94%	2	18	11%
6962181067	7.47	96%	2	31	6%
7086361926	7.55	97%	2	24	8%
8378563200	7.42	92%	5	31	16%
8792009665	7.26	96%	2	15	13%
Grand Total	**7.14**	**0.92**	**81**	**376.00**	22%

Data Analysis: Sleep Duration and Sleep Efficiency Tracking



Data Analysis: Sleep Duration and Sleep Efficiency Tracking

- The usage tread shows some consumers care about and tracking their **sleep duration** and **sleep efficiency**.
- Therefore, Bellabeat's marketing strategy should include
  a) **identifying groups who may have sleep deprivation or sleep efficiency issues**, such as older women, women who need to take care of their families and spend time working, women with chronic medical conditions, or women who frequently experience menstrual stress or pain.
  b) **encouraging the above-mentioned people to purchase and wear Bellabeat's smart devices** to track sleep duration and sleep efficiency.

Key Findings

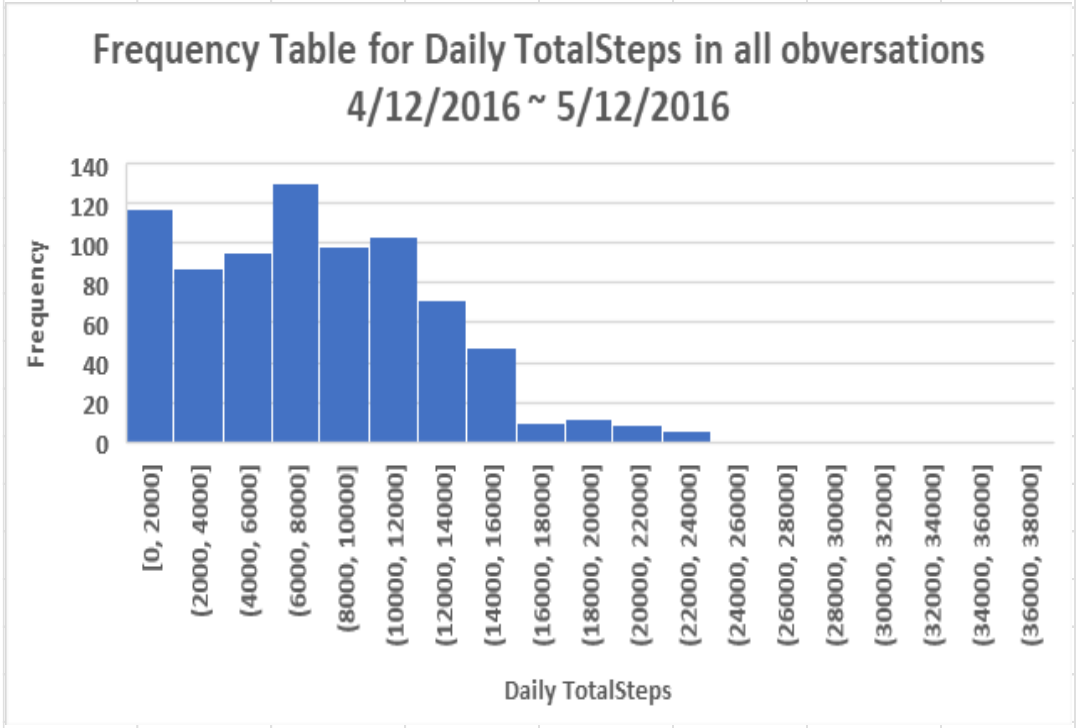
- Here are the trends found while analyzing smart device usage data from a public dataset:
  - The consumers in this dataset used smart devices to **track daily activity** such as how many steps they took, how far they walked at each intensity level, how long each intensity level lasted, and how many calories they burned.
  - The consumers used smart devices to **monitor** their **heart rate**.
  - The consumers used smart devices to **watch** their **weight** and **BMI level**.
  - The consumers used smart devices to track their **sleep duration** and **sleep efficiency**.
- These trends can apply to Bellabeat customers because:
  - Bellabeat's products like **Bellabbeat app**, **Leaf** and **Time** track user **activity**, **sleep**, and **stress**.
  - From the previous bullet item, we know that consumers use non-Bellabeat smart devices to track their **daily activity**, **sleep**, **heart rate**, **weight**, and **BMI** level.
  - Potentially biased data were removed prior to our analysis, meaning that our data represent a good sampling distribution. Therefore, trends found in our analysis regarding **activity** and **sleep** should apply to Bellabeat customers.
  - Stress may/may not cause abnormal heart rate or weight changes. However, the trends discovered in our analysis should help Bellabeat get the data its users want or need.

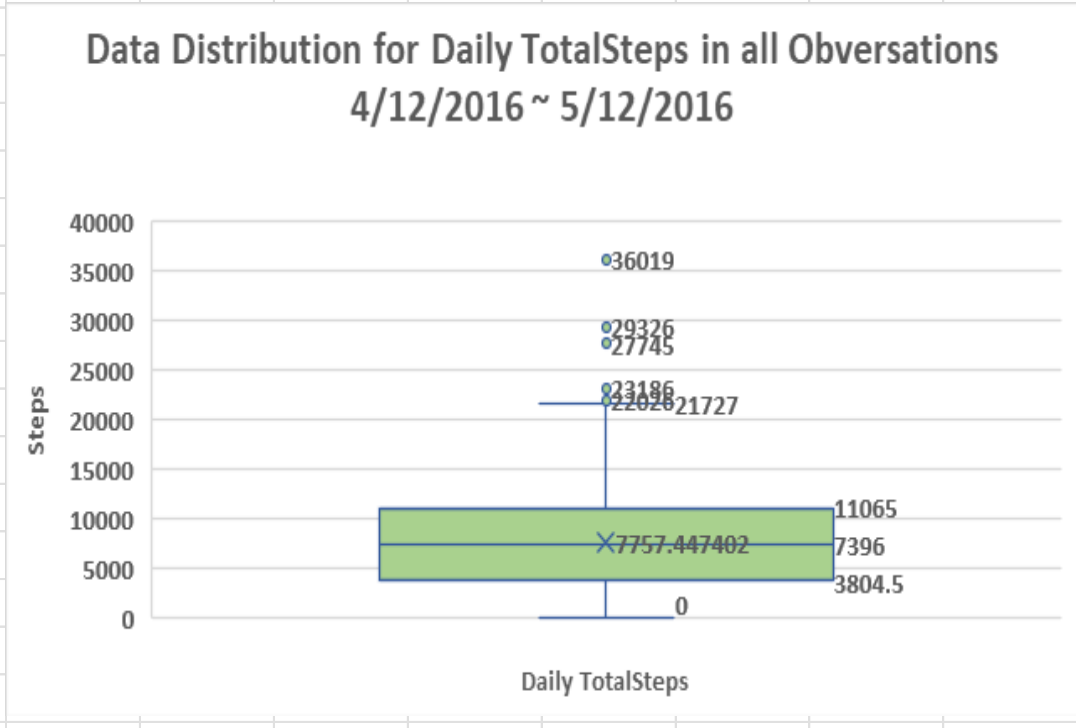
Key Findings: Daily Activity Tracking

Interesting findings:

- **Each** consumer in this dataset used a smart device to track their daily activities, including total steps taken, total distance walked, distance walked per activity level, total minutes spent per activity level, and total calories burned.
- By studying the frequency table (histogram chart) and data distribution chart (box plot) of the daily total steps for all observations in the dataset, the **following trends were found**:
  - The **most common set** of values for **daily total steps** were **between 6,000 and 8,000 steps.** The **next common set** of values for **daily total steps** were **between 0 and 2,000 steps.**
  - The **median** (the middle number in a sorted list of numbers) of the daily total steps was **7,396**.
  - The upper quartile (or third quartile, is the value under which 75% of data points are found when arranged in increasing order) of the box plot shows that **75%** of the values for daily total steps were **under 11,065**.

Key Findings: Daily Activity Tracking





Key Findings: Daily Activity Tracking

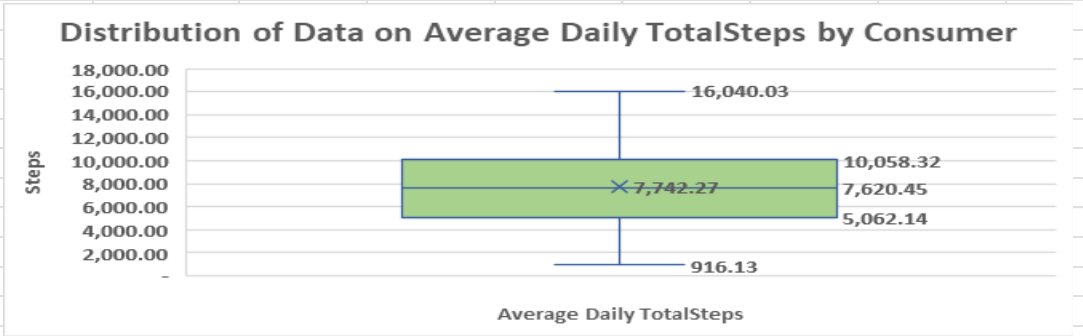
Interesting findings:

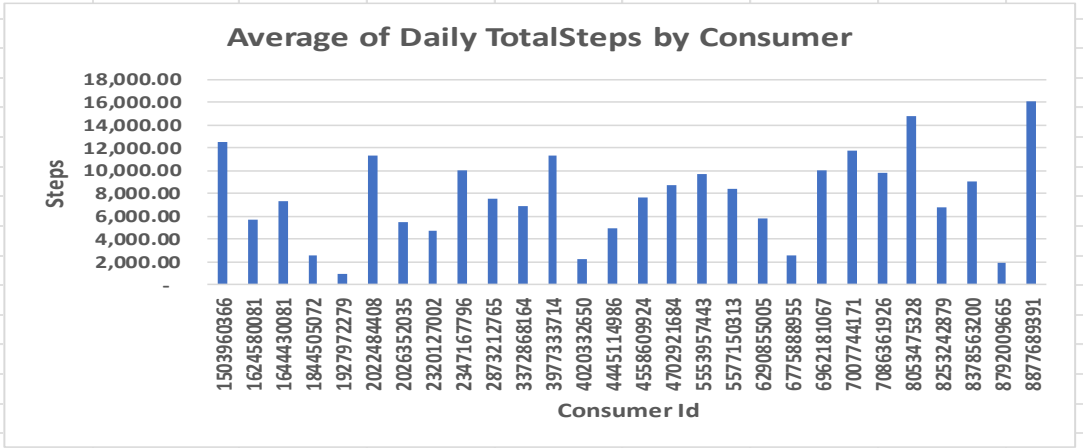
- **Aggregating** the data **by consumer Id**, **the following trends were discovered**:
  - Median average total daily steps per consumer is **7,620.45.**
  - The upper quartile of the boxplot shows that **75% of consumers** have an average daily total step count **below 10,058.32**.
  - According to this document, Walking Meeting - Preventing Chronic Disease, The Centers for Disease Control and Prevention (CDC) recommends walking **at least 10,000 steps per day**. It appears that **most consumers in this dataset were not meeting CDC recommendations.**

Key Findings: Daily Activity Tracking

Row Labels	Average of TotalSteps	Average of TotalDistance
1503960366	12,520.63	8.07
1624580081	5,743.90	3.91
1644430081	7,282.97	5.30
1844505072	2,575.96	1.70
1927972279	916.13	0.63
2022484408	11,370.65	8.08
2026352035	5,456.07	3.39
2320127002	4,716.87	3.19
2347167796	10,077.18	6.73
2873212765	7,555.77	5.10
3372868164	6,861.65	4.71
3977333714	11,337.62	7.76
4020332650	2,267.23	1.63
4445114986	4,930.83	3.34
4558609924	7,685.13	5.08
4702921684	8,766.07	7.11
5553957443	9,676.31	6.34
5577150313	8,451.55	6.32
6290855005	5,851.32	4.43
6775888955	2,541.80	1.83
6962181067	10,001.73	6.73
7007744171	11,776.36	8.34
7086361926	9,766.07	6.67
8053475328	14,763.29	11.48
8253242879	6,842.28	4.93
8378563200	9,088.14	7.21
8792009665	1,919.93	1.23
8877689391	16,040.03	13.21
Grand Total	**7,757.45**	**5.54**

12,520.63
5,743.90
7,282.97
2,575.96
916.13
11,370.65
4,716.87
10,077.18
7,555.77
6,861.65
11,337.62
2,267.23
4,930.83
7,685.13
8,766.07
9,676.31
8,451.55
5,851.32
2,541.80
10,001.73
11,776.36
9,766.07
14,763.29
6,842.28
9,088.14
16,040.03





Key Findings: Daily Activity Tracking

Interesting findings:

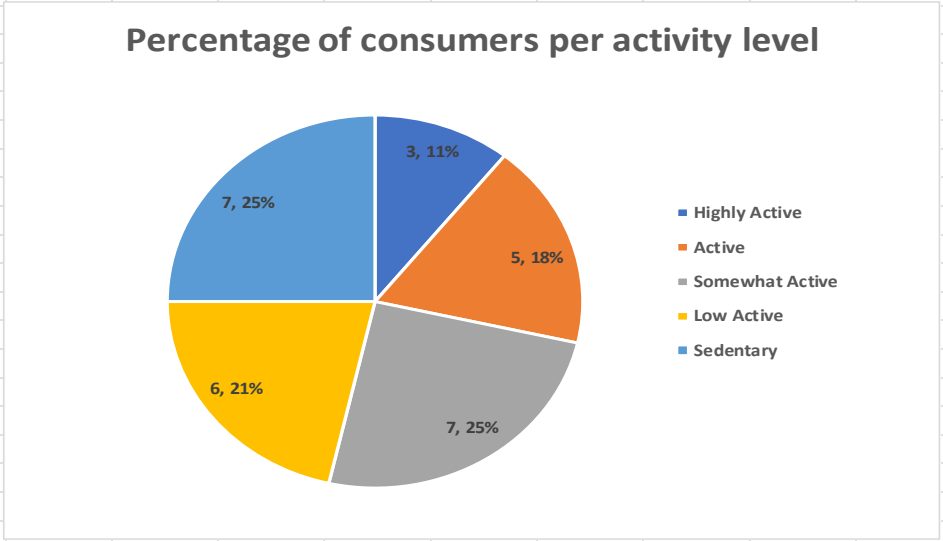
- Each consumer's "activity level“ was categorized based on their average number of steps taken per day and the following guidelines mentioned in the article How Many Steps a Day Is Considered Active?:
  - **Sedentary:** Less than 5,000 steps daily
  - **Low active:** About 5,000 to 7,499 steps daily
  - **Somewhat active:** About 7,500 to 9,999 steps daily
  - **Active:** More than 10,000 steps daily
  - **Highly active:** More than 12,500 steps daily
- The following trends where found:
  - **71%** of consumers took less than 10,000 steps per day on average (**25%** of consumers had an activity level of "**sedentary**", **21%** of consumers had an activity level of "**low active**", and **25%** of consumers have an activity level of "**somewhat active**“).

Key Findings: Daily Activity Tracking

Row Labels	Average of TotalSteps	Activity Level
1503960366	12,520.63	Highly Active
1624580081	5,743.90	Low Active
1644430081	7,282.97	Low Active
1844505072	2,575.96	Sedentary
1927972279	916.13	Sedentary
2022484408	11,370.65	Active
2026352035	5,456.07	Low Active
2320127002	4,716.87	Sedentary
2347167796	10,077.18	Active
2873212765	7,555.77	Somewhat Active
3372868164	6,861.65	Low Active
3977333714	11,337.62	Active
4020332650	2,267.23	Sedentary
4445114986	4,930.83	Sedentary
4558609924	7,685.13	Somewhat Active
4702921684	8,766.07	Somewhat Active
5553957443	9,676.31	Somewhat Active
5577150313	8,451.55	Somewhat Active
6290855005	5,851.32	Low Active
6775888955	2,541.80	Sedentary
6962181067	10,001.73	Active
7007744171	11,776.36	Active
7086361926	9,766.07	Somewhat Active
8053475328	14,763.29	Highly Active
8253242879	6,842.28	Low Active
8378563200	9,088.14	Somewhat Active
8792009665	1,919.93	Sedentary
8877689391	16,040.03	Highly Active
Grand Total	**7,757.45**	

Activity Level	Count
Highly Active	3
Active	5
Somewhat Active	7
Low Active	6
Sedentary	7

Percentage of Consumers with Average Daily TotalSteps less than 10,000 is: **0.714286**



Key Findings: Daily Activity Tracking

Interesting findings:

- According to this article How many steps should people take per day?, most people in the United States take less than 10,000 per day. Same trend was found in the dataset we analyzed. Therefore, Bellabeat's marketing strategy should include **encouraging people to purchase and wear Bellabeat's smart devices to track/count the number of steps they take every day to ensure they achieve their goal of walking at least 10,000 steps a day.**

Key Findings: Weight and BMI Level Tracking

Interesting findings:

- About **24%** of consumers in the dataset (8 out of 33 consumers) used smart devices to track their **weight** and **BMI level**. The weight status of a consumer was determined by his/her BMI value. Three consumers (or **9%** of all consumers) had the weight status of "overweight," and one consumer (or **3%** of all consumers) had the weight status of "obesity."
- Combining consumers' daily activity information and weight information and found that:
  - Of the three consumers with the "**overweight**" weight status, two were at the "**somewhat active**" activity level (an average of **5,000 to 7,499 steps per day**).
  - The consumer whose weight status was "**obesity**" had the activity level of "**sedentary**" (**less than 5,000 steps per day** on average).

Key Findings: Weight and BMI Level Tracking

Row Labels	Average of TotalSteps	Activity Level	Weight Status
1503960366	12520.63333	Highly Active	Healthy Weight
1927972279	916.1290323	Sedentary	Obesity
2873212765	7555.774194	Somewhat Active	Healthy Weight
4558609924	7685.129032	Somewhat Active	Overweight
5577150313	8451.551724	Somewhat Active	Overweight
6962181067	10001.73333	Active	Healthy Weight
8877689391	16040.03226	Highly Active	Overweight
Grand Total	**9008.802817**		

Key Findings: Weight and BMI Level Tracking

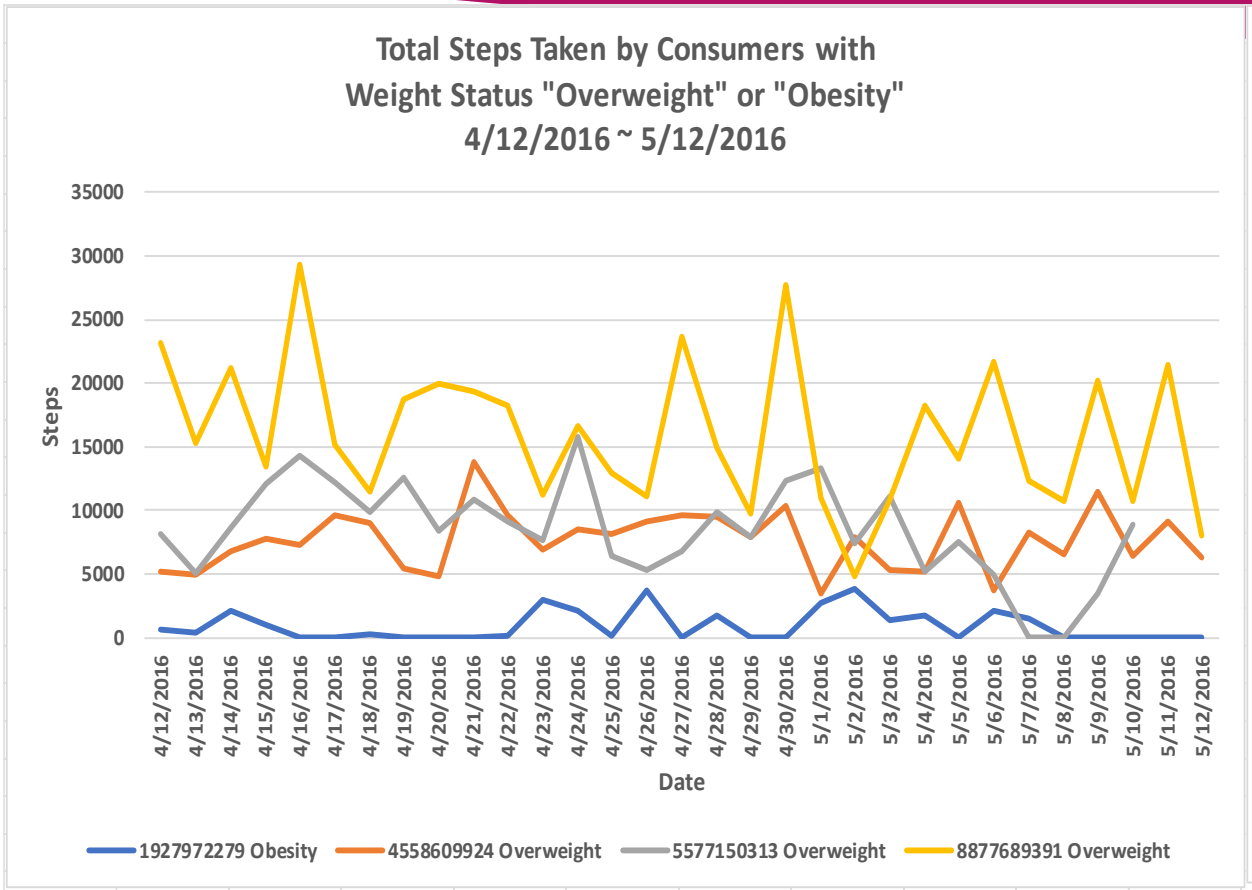
Interesting findings:

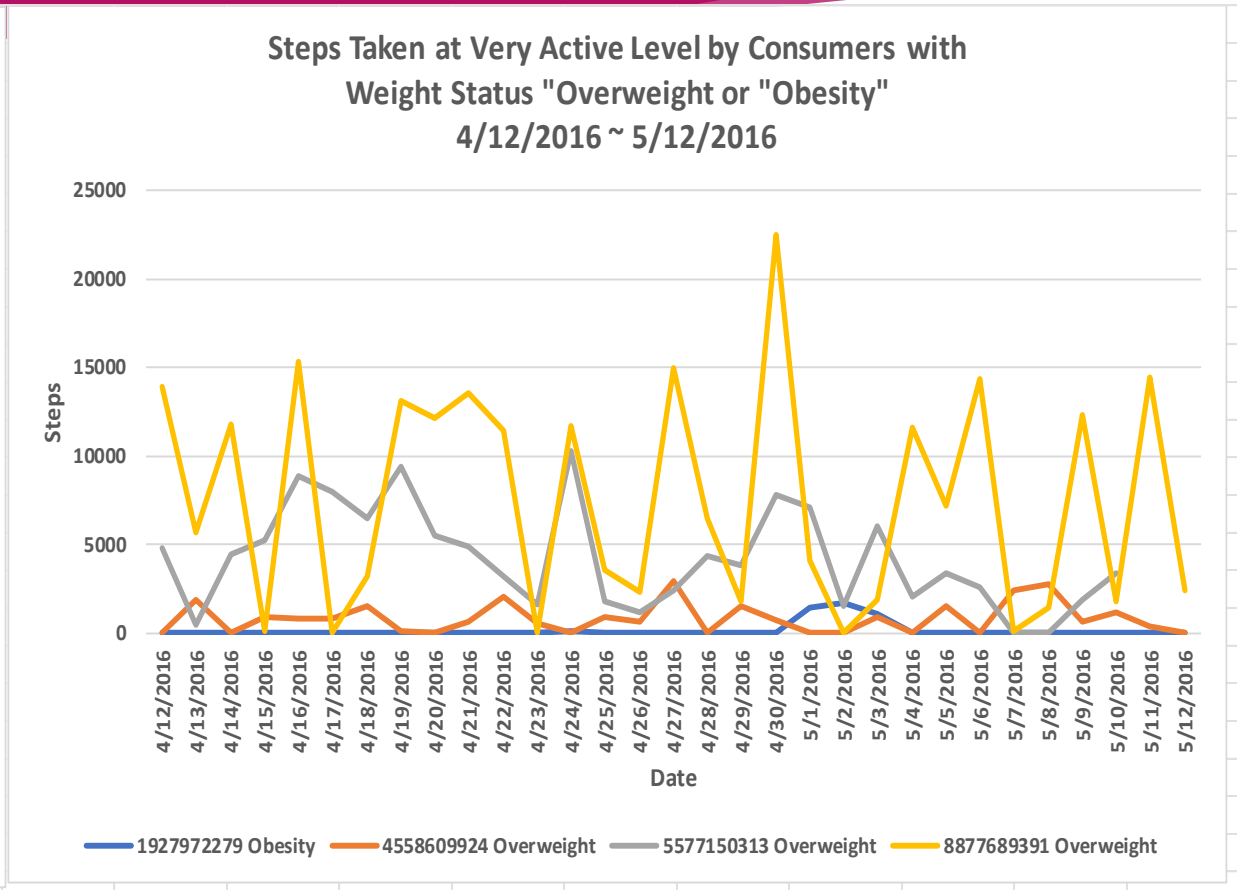
- According to this article "How many steps should people take per day?“, a 2018 analysis of 363 people with obesity found that people who walked **10,000 steps a day**, including **at least 3,500 steps** engaging in **moderate-to-vigorous activity** lasting 10 minutes or longer, had increased weight loss.
- Researching deeper into the daily activity data of consumers with the weight status of either "**overweight**" or "**obesity**“ and found that **only one** consumer took serious action to lose weight by walking more than 10,000 steps a day, at least 3,500 of which were at a very active level for 10 minutes or longer.
- Therefore, Bellabeat's marketing strategy should include **identifying people who may have weight or BMI level concerns and** encouraging them to **purchase and wear Bellabeat’s smart devices to track their weight and BMI level.**

Key Findings: Weight and BMI Level Tracking

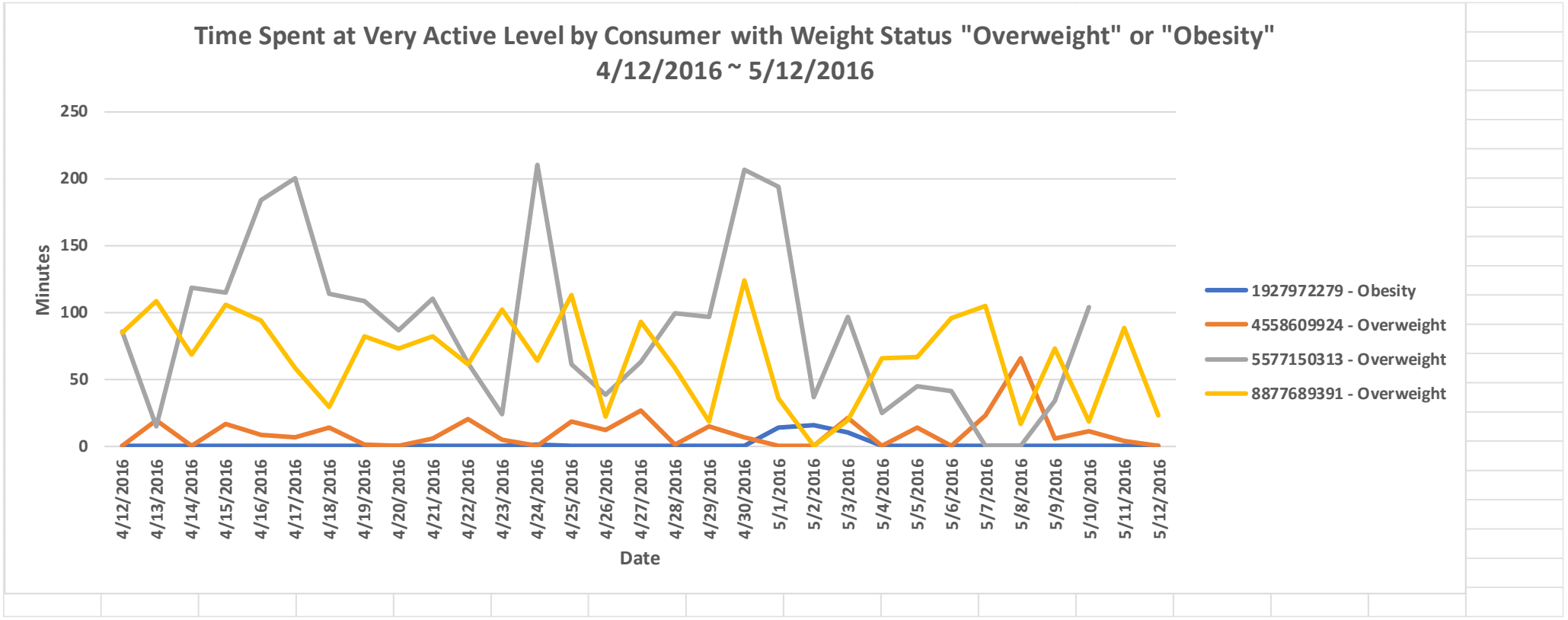
Row Labels	Average of TotalSteps	Average of VeryActiveDistance	Average of VeryActiveSteps	Average of VeryActiveMinutes	Average of WeightPounds	Average of BMI	Weight Status	Activity Level
1503960366	12,520.63	2.95	4,578.09	40.00	115.96	22.65	Healthy Weight	Highly Active
1927972279	916.13	0.10	138.44	1.32	294.32	47.54	Obesity	Sedentary
2873212765	7,555.77	0.68	993.11	14.10	125.66	21.57	Healthy Weight	Somewhat Active
4558609924	7,685.13	0.55	830.92	10.39	153.53	27.21	Overweight	Somewhat Active
5577150313	8,451.55	3.16	4,221.78	88.79	199.96	28.00	Overweight	Somewhat Active
6962181067	10,001.73	1.67	2,457.54	23.57	135.68	24.02	Healthy Weight	Active
8877689391	16,040.03	6.64	7,583.78	66.06	187.71	25.49	Overweight	Highly Active
Grand Total	**9,008.80**	**2.24**	**2,955.09**	**34.41**	**159.14**	**25.13**		

Key Findings: Weight and BMI Level Tracking





Key Findings: Weight and BMI Level Tracking



Key Findings: Sleep Duration and Sleep Efficiency Tracking

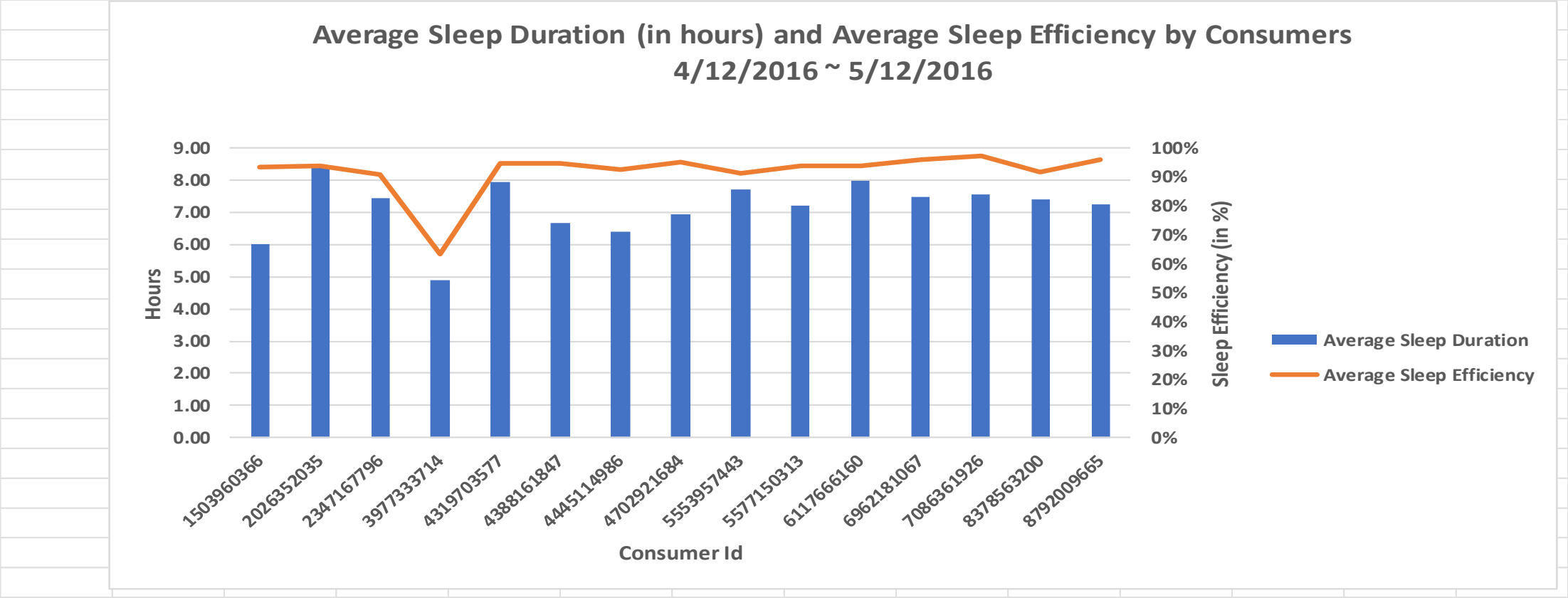
Interesting findings:

- Approximately **70%** of consumers in the dataset (24 out of 33 consumers) used smart device to track their sleep durations and sleep efficiencies (or total sleep minutes and total time in bed) . The following issues were found:
  - **Sleep deprivation issues:** according to this article Is 6 Hours of Sleep Enough?, **consistently getting less than 6 hours of sleep can have consequences for a person's health and quality of life.** 2 consumers (or **6%** of the consumers in the dataset) slept **less than 6 hours per day** on **more than 50% of the days** during the data collection date range**.**
  - **Sleep efficiency issue:** sleep efficiency is commonly defined as the ratio of **total sleep time** to **total time in bed**. A normal sleep efficiency is **85%** or higher. 1 consumer (or **3%** of the consumers in the dataset) had **poor sleep efficiency**, with an average sleep efficiency of about **63%**.
- For some consumers, it may be **important** to be able to **use a smart device to track their sleep durations and sleep efficiencies** and **receive alerts** when **the device detects** that the customer is **consistently getting less than 6 hours of sleep** or **has poor average sleep efficiency**.
- Therefore, Bellabeat's marketing strategy should include **identifying the potential buyers**, and encouraging them to **purchase and wear Bellabeat's smart devices to track their sleep durations and sleep efficiencies.**

Key Findings: Sleep Duration and Sleep Efficiency Tracking

Row Labels	Average of TotalHoursAsleep	Average of SleepEfficiency	Sum of ShortSleepDay	Count of Id	Percentage of Short Sleep Day
1503960366	6.01	94%	14	25	56%
2026352035	8.44	94%	1	28	4%
2347167796	7.45	91%	0	15	0%
3977333714	4.90	63%	24	28	86%
4319703577	7.94	95%	3	26	12%
4388161847	6.67	95%	6	23	26%
4445114986	6.42	93%	8	28	29%
4702921684	6.96	95%	4	27	15%
5553957443	7.73	91%	5	31	16%
5577150313	7.20	94%	3	26	12%
6117666160	7.98	94%	2	18	11%
6962181067	7.47	96%	2	31	6%
7086361926	7.55	97%	2	24	8%
8378563200	7.42	92%	5	31	16%
8792009665	7.26	96%	2	15	13%
Grand Total	**7.14**	**0.92**	**81**	**376.00**	22%

Key Findings: Sleep Duration and Sleep Efficiency Tracking

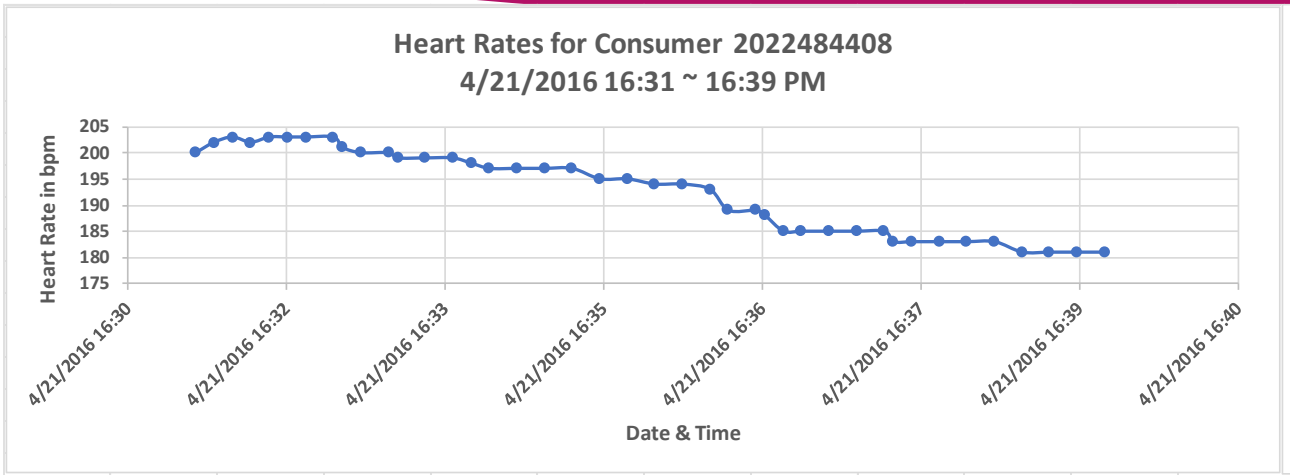


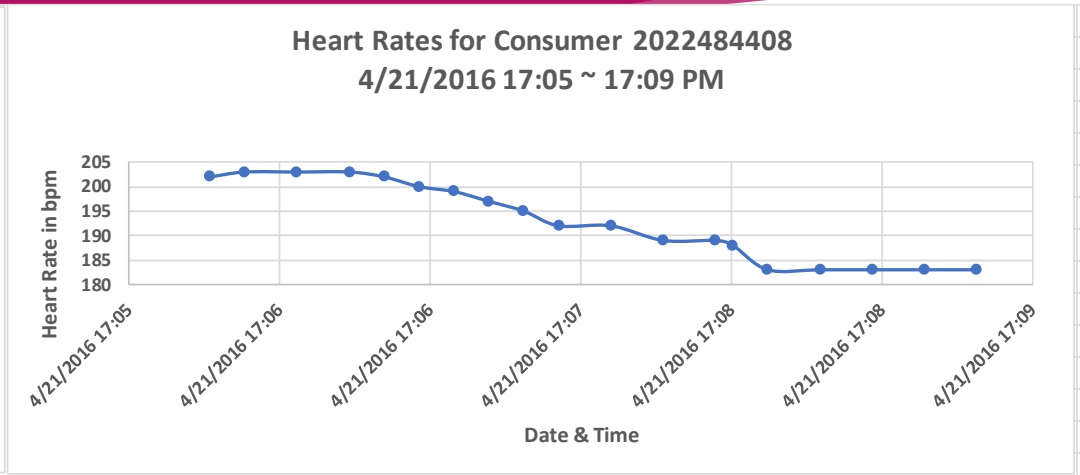
Key Findings: Heart Rate Monitoring

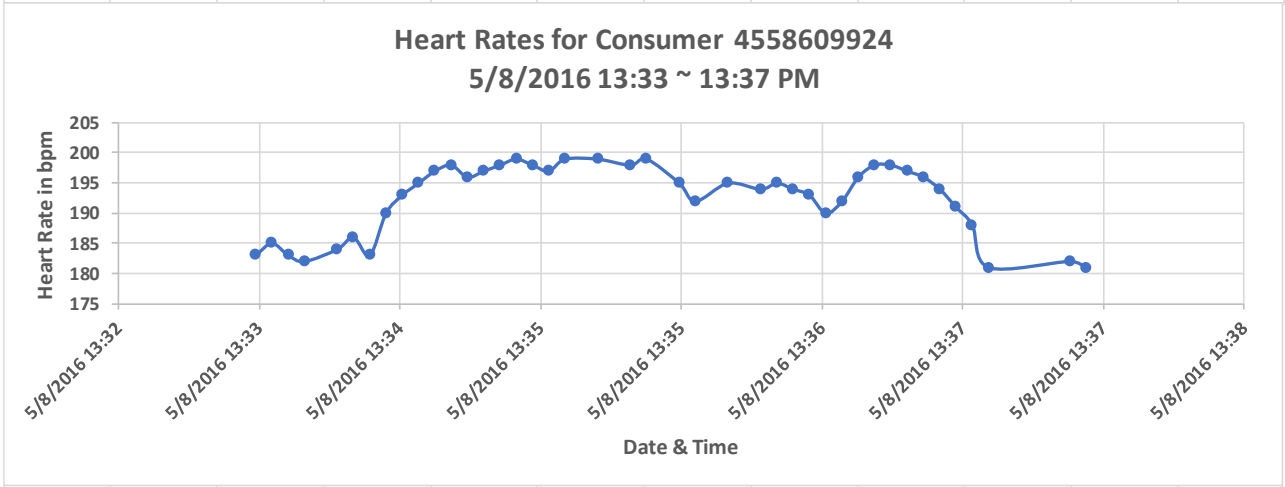
Interesting findings:

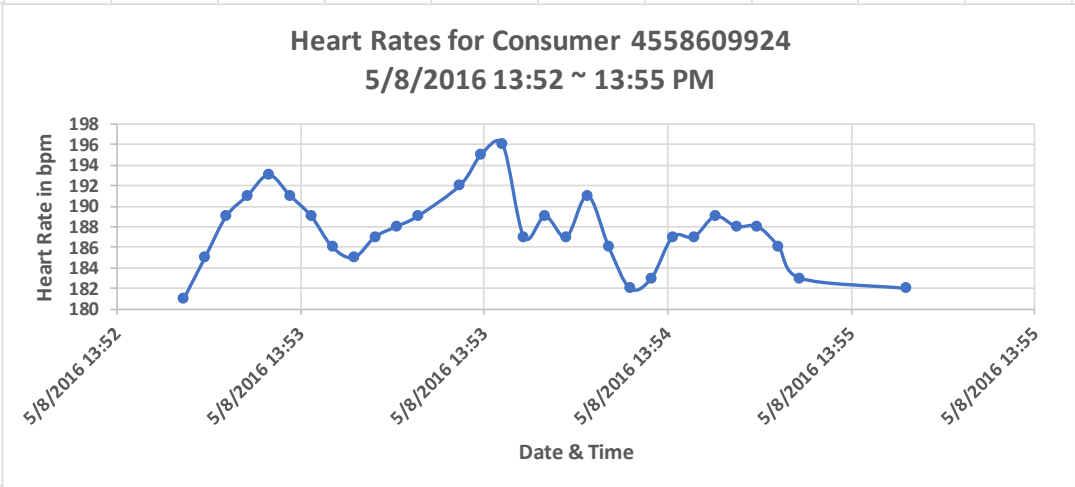
- Approximately **20%** of consumers in the dataset (7 out of 33 consumers) used smart device to monitor their heart rates. Among them, 4 consumers (or **12%** of all the consumers in the dataset) experienced high heart rates (which were **over 180 bpm**) that lasted for several minutes.
- For some consumers, it may be **important** to be able to **use a smart monitoring device** and **receive alerts** when **the device detects an abnormal heart rate**.
- Therefore, Bellabeat's marketing strategy should include **identifying the potential buyers**, and encouraging them to **purchase and wear Bellabeat's smart devices to monitor their heart rate during walking, exercise**, etc.

Key Findings: Heart Rate Monitoring

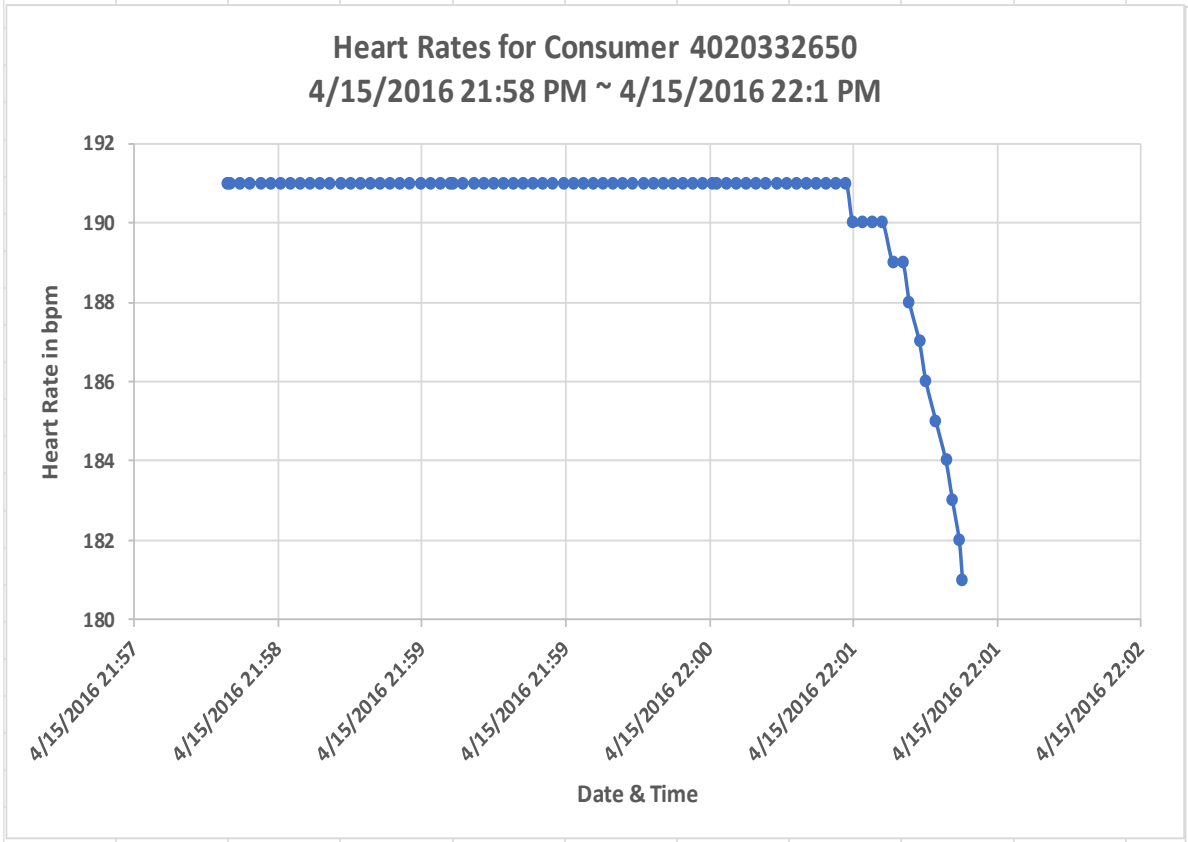


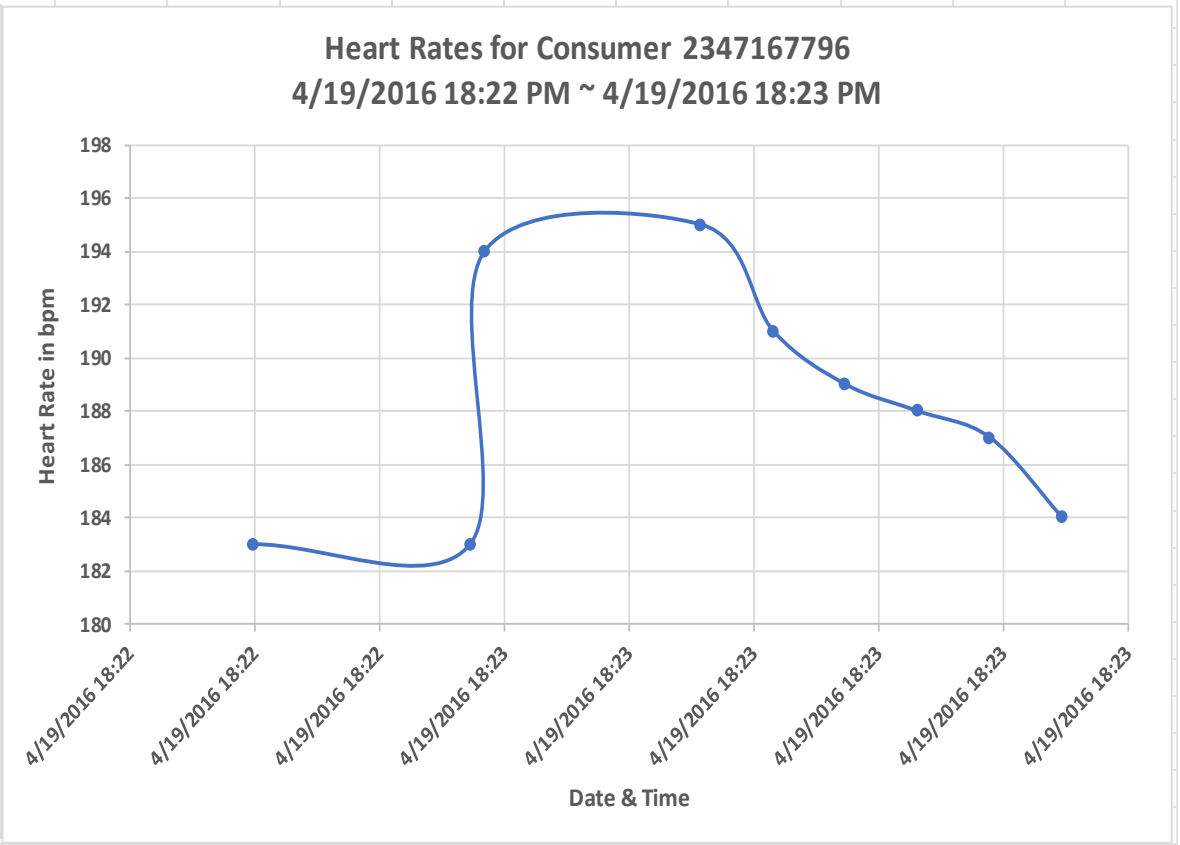






Key Findings: Heart Rate Monitoring





Recommendations

High-level Recommendations for Bellabeat's Marketing Strategy:

Regarding monitoring customers' heart rates:

- Request the technical team to **add** **the function of monitoring customers' heart rate** on Bellabeat's smart device **Time** (if it does not already have this function). This feature might include **sending an alert** to the customer if the device **detects abnormal heart rates** (too high or too low) for several minutes.
- **Identify people who may have heart rate issues** (tachycardia or bradycardia) and **promote this feature** to them. Approximately **20% of consumers** in the dataset were interested in and **used** this feature, and **12%** of consumers in the dataset **experienced** heart rate issues, so if Bellabeat's marketing team can find potential buyers, there might be **new growth opportunities** **in this area**. Potential customers may include:
  - People with heart disease or atrial fibrillation
  - People above middle age
  - People with thyroid disease or diabetes
  - Smokers

Recommendations

High-level Recommendations for Bellabeat's Marketing Strategy:

Regarding tracking customers' weights and BMI levels:

- Request technical team to **add** the function to **track customers' weights and BMI levels** on Bellabeat's smart device **Time** (if this feature does not already exist). This feature might include **sending an alert** to the customer if the device detects that the customer's **BMI level** indicates she is **overweight or obese**, and **sending reminders** to help the customer manage her weight (e.g., **walk 10,000 steps per day**, with at least **3,500** of those steps being **moderate to vigorous**).
- **Identify people who may have weight or BMI level concerns** and **promote this feature to them**. About **20%** of consumers in the dataset were interested in and used this feature, and **12%** of consumers in the dataset were either **overweight** or **obese**, so if Bellabeat's marketing team can find potential buyers, there could be **new growth opportunities in this area**. Potential customers may include:
  - Women of certain racial groups (e.g., African-American women, Hispanic women)
  - Women of a certain age (for example, women between 40 and 59 years old)
  - Women who weigh more than 150 pounds.

Recommendations

High-level Recommendations For Bellabeat's Marketing Strategy:

Regarding tracking customers' daily activities

- Request the technical team to **add** the function of **tracking customers' daily activities** (including total walking steps, total walking distance, walking distance for each activity level, total minutes spent for each activity level, and total calories burned) on Bellabeat's smart device **Time** (if this feature does not already exist). This feature might include **sending reminders** to customers who take less than 10,000 steps per day.
- **Encourage everyone to purchase** Bellabeat's smart device **Time**, follow CDC's recommendation to **set a goal of 10,000 steps each day**, and **track the daily activity using Time**. **Every** consumer in the dataset was interested in and used this feature, but at least **70%** of consumers in the dataset took less than 10,000 steps per day. If Bellabeat's marketing team can properly motivate buyers, there could **be a lot of growth** in this area.

Recommendations

High-level Recommendations For Bellabeat's Marketing Strategy:

Regarding tracking customers' sleep durations and sleep efficiencies:

- Request the technical team to **add the function to track customers' sleep durations and sleep efficiencies** (or total sleep minutes and total time in bed) on Bellabeat's smart device **Time** (if this feature does not already exist). This functionality may include **sending an alert** to the customer if the device detects that the customer is **consistently getting less than 6 hours of sleep** or **has poor average sleep efficiency**.
- **Identify people who may have sleep deprivation or sleep efficiency issues and promote this feature** to them. About **70%** of the consumers in the dataset were interested in and used this feature, at least **6%** of the consumers in the dataset had sleep deprivation issues, and at least **3%** of the consumers in the dataset had sleep efficiency issues. If Bellabeat's marketing team can find potential buyers, there might be **some growth in this area**. Potential customers may include:
  - Elderly women
  - Women who need to take care of their families and spend time working
  - Women with chronic illness
  - Women who frequently experience menstrual stress or pain