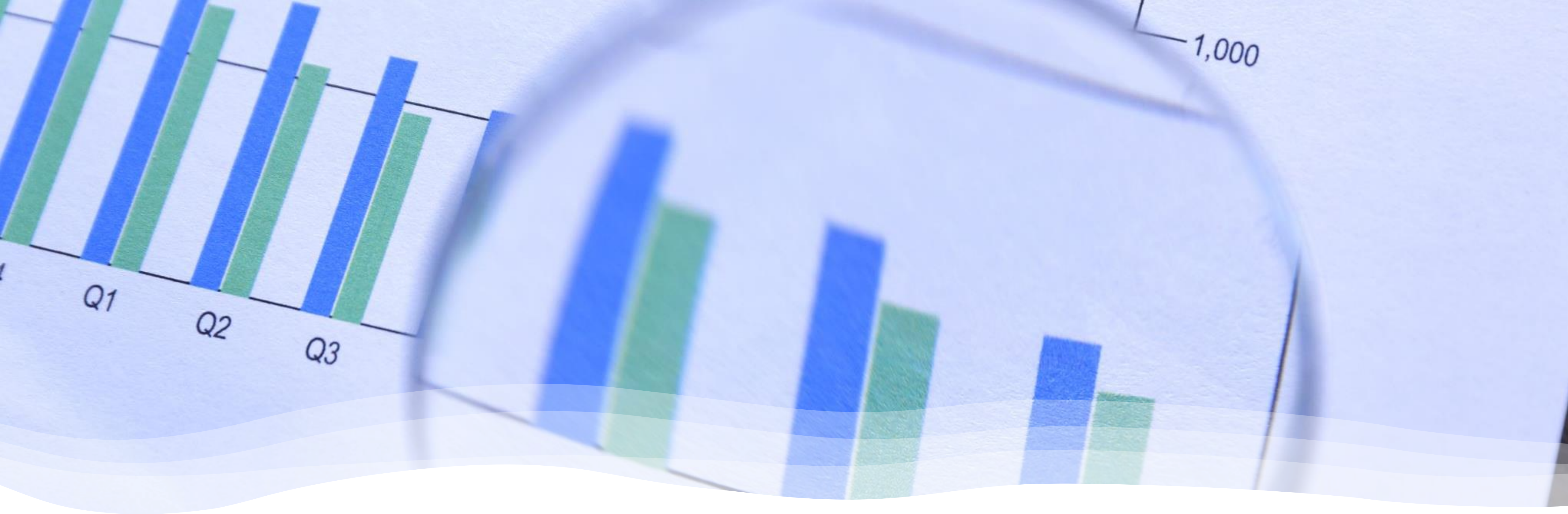


Data Processing & Data Analysis Part II for Case Study “How Can a Wellness Technology Company Play It Smart?”

By Kristin Lu

February 1, 2024

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.
- 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 part II of this analysis:
  - 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 – 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
```

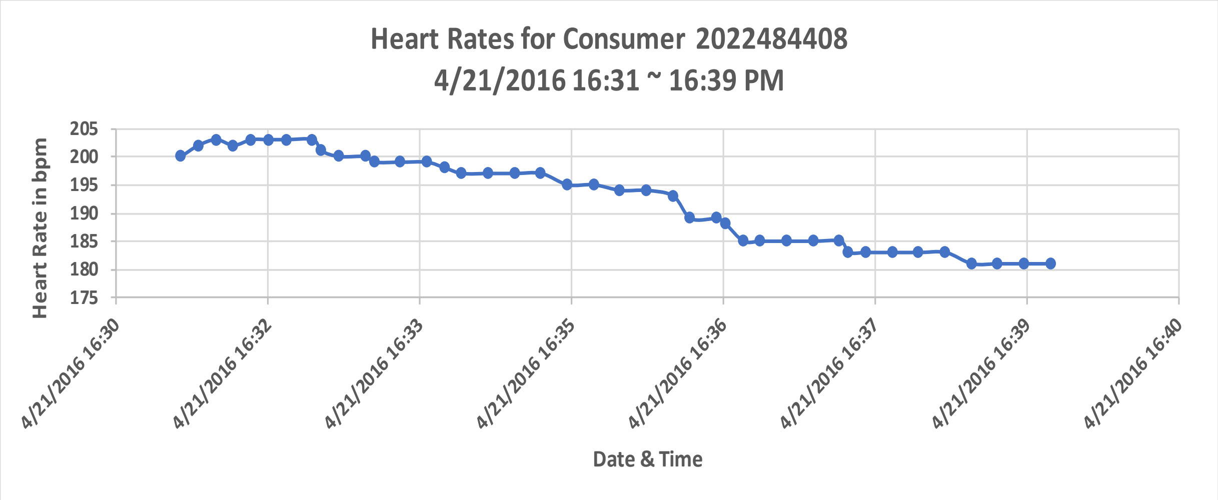
Data Processing – Sleep Duration & Sleep Quality

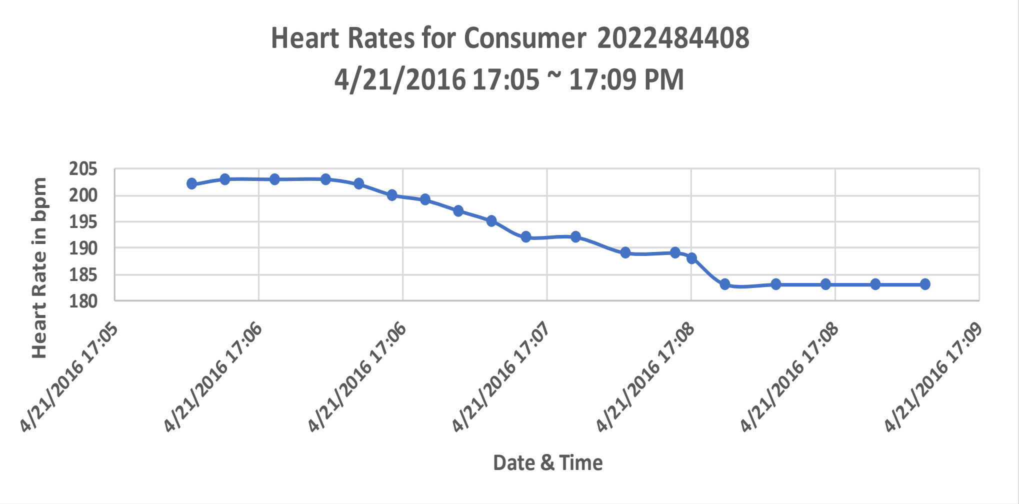
- Opened sleepDay_merged.csv and saved as Excel workbook.
- Removed duplicates rows.
- Removed the rows related to the following consumers who had less than 15 days (50% of the 31 days) of data: 1644430081, 1844505072, 1927972279, 2320127002, 4020332650, 558609924, 677588955, 70077441712, 8053475328.

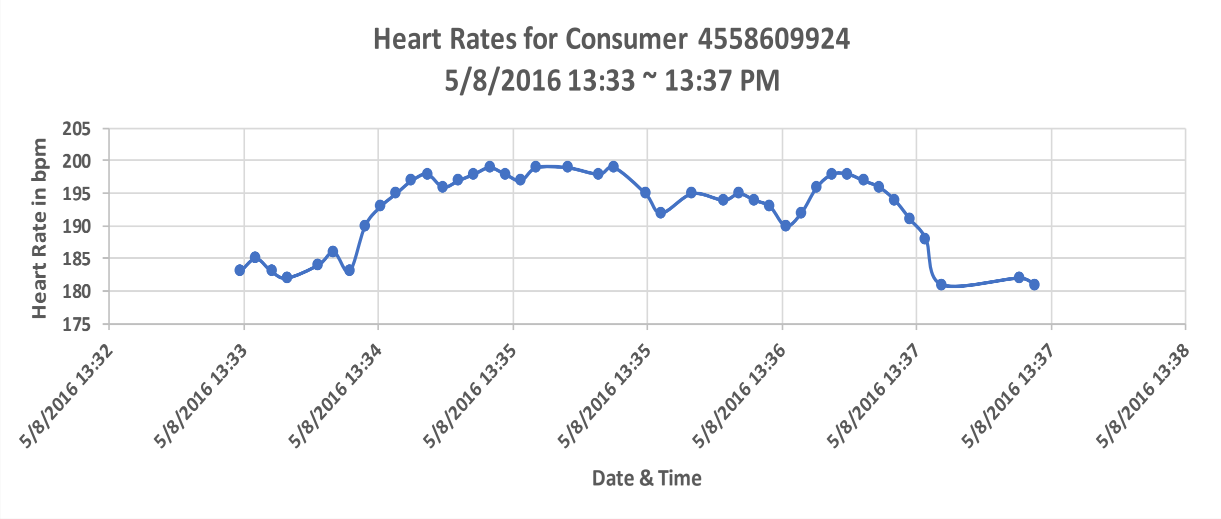
Data Analysis: Heart Rate Monitoring

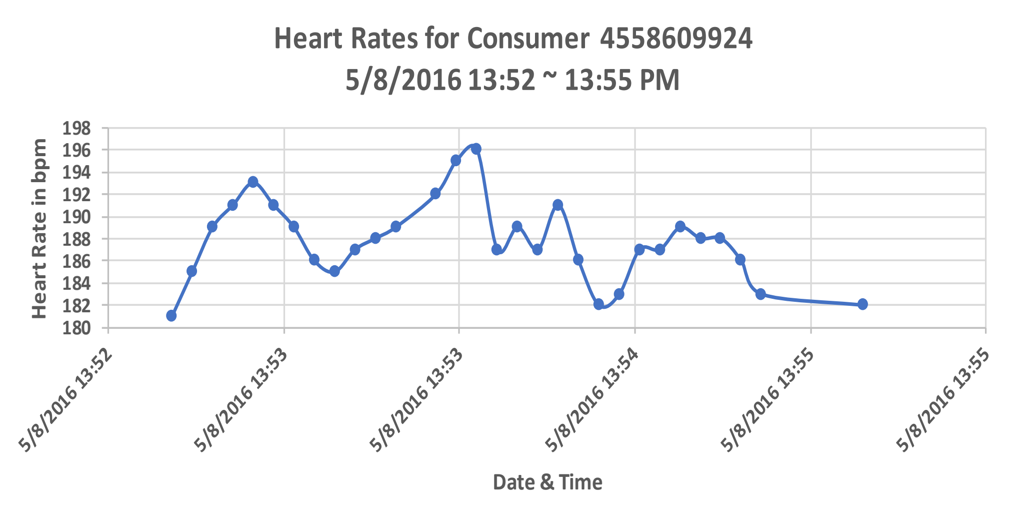
- In the worksheet heartrate_sec_high, created a PivotTable. Aggregated data by consumer – there were 4 consumers with heart rate over 180 bpm. Calculated average heart rate by consumer.
- In heartrate_sec_high, sorted data first by **Id** column, then by **Time** column.
- The above consumers were observed 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

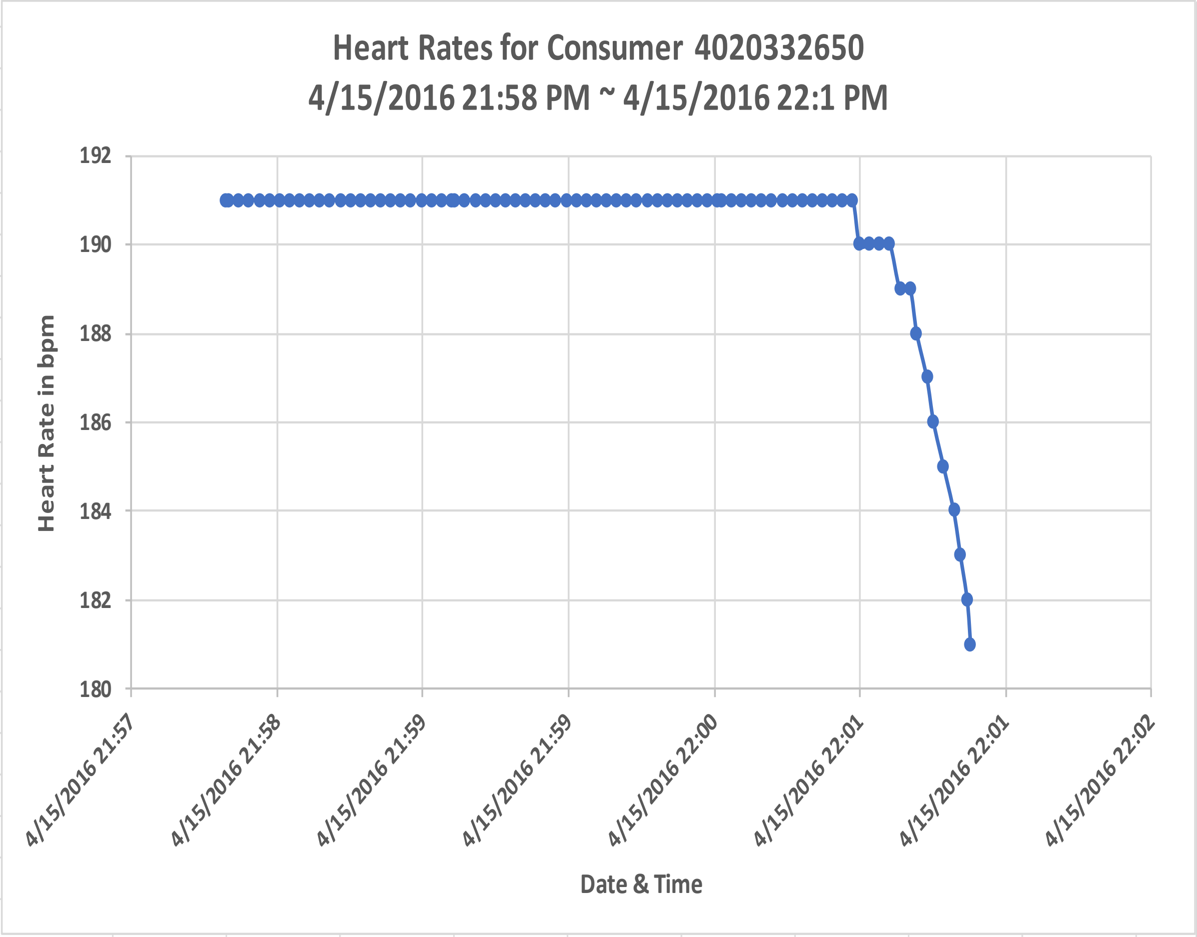

Heart Rates for Consumer 2022484408
4/21/2016 16:31 ~ 16:39 PM

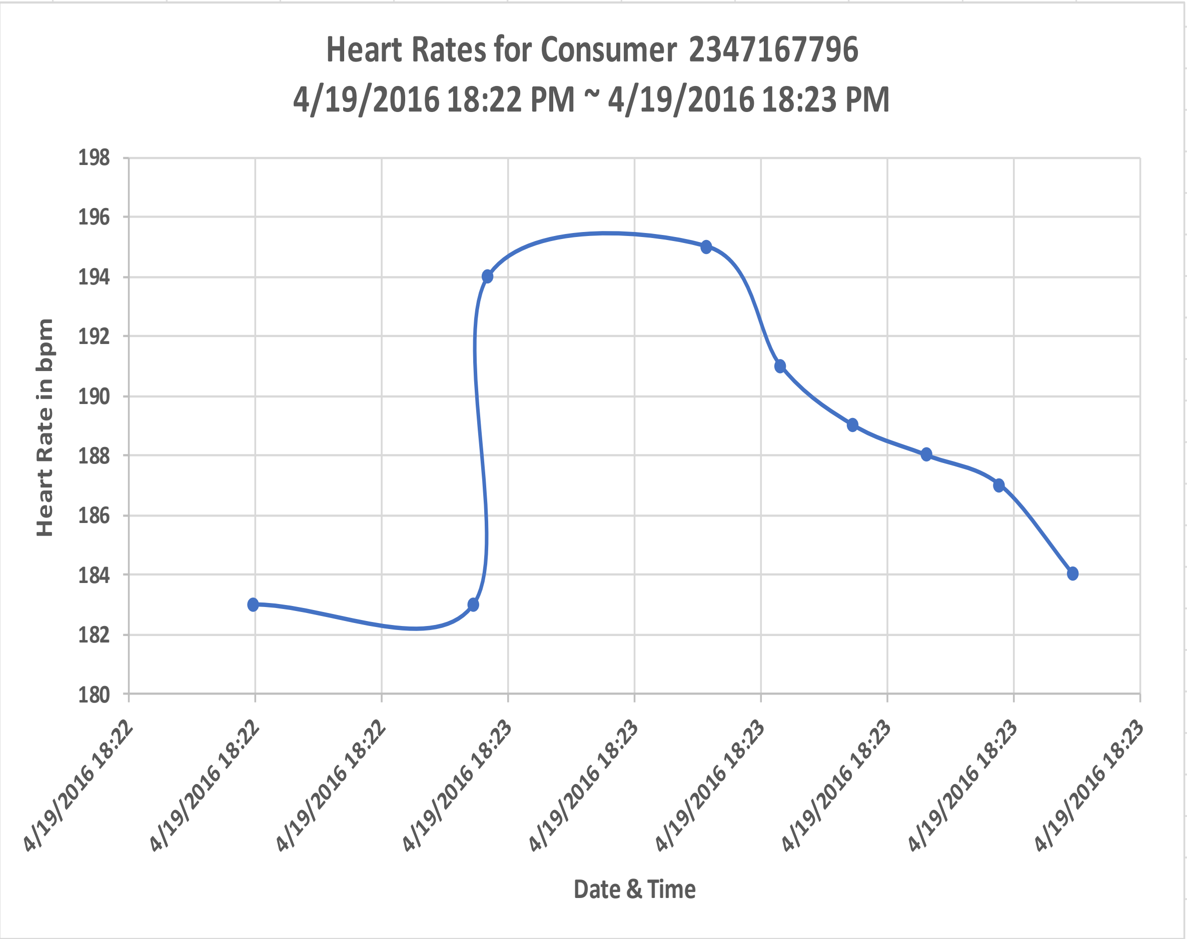

Heart Rates for Consumer 2022484408
4/21/2016 17:05 ~ 17:09 PM


Heart Rates for Consumer 4558609924
5/8/2016 13:33 ~ 13:37 PM


Heart Rates for Consumer 4558609924
5/8/2016 13:52 ~ 13:55 PM

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

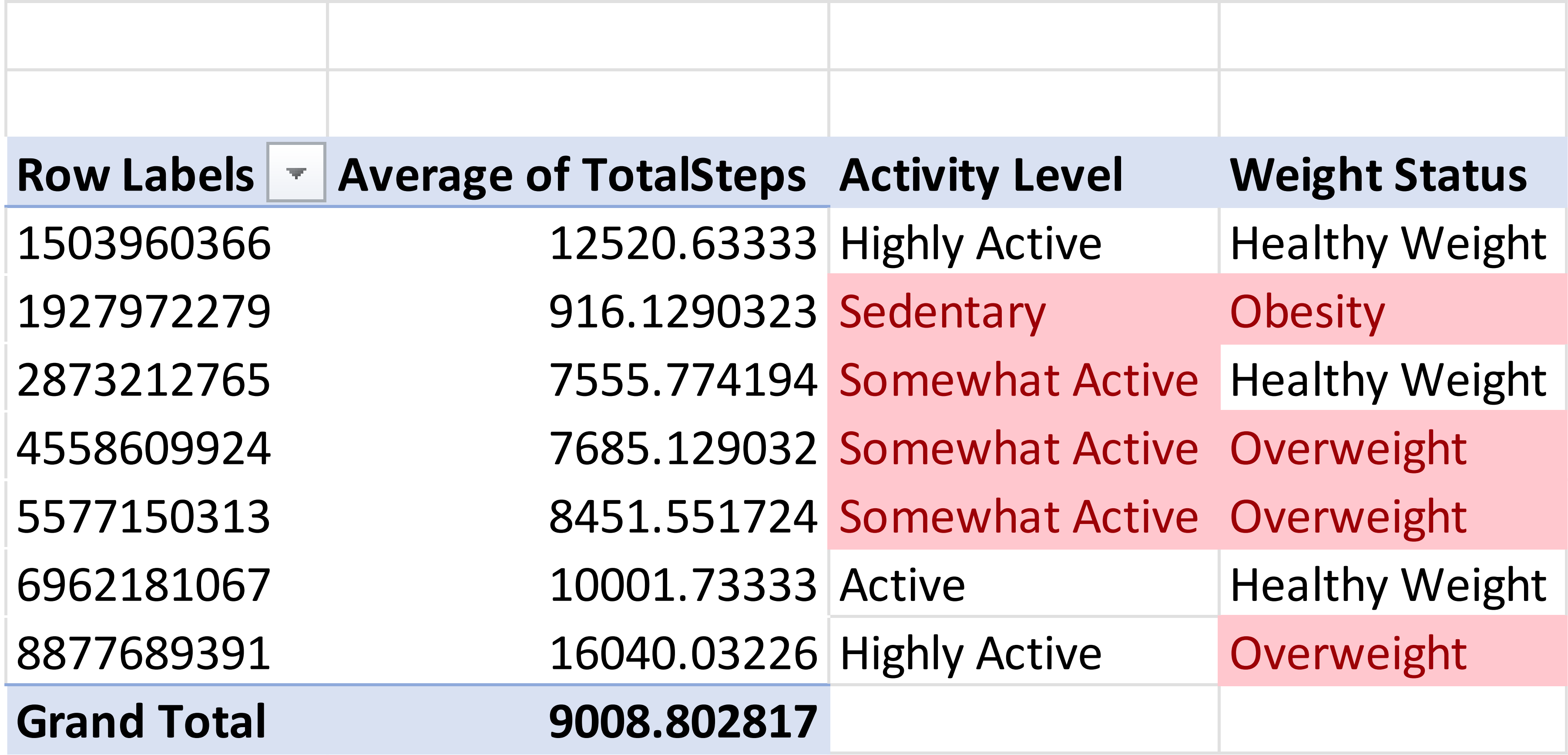
- Created a Pivot Table in dailyActivity_weightInfo_merged file. Aggregated data by consumer and got the average of daily total steps taken by each consumer. Added a column “**Activity Level**” next to the PivotTable. Categorized the activity level (as described earlier in this document) for each consumer.

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

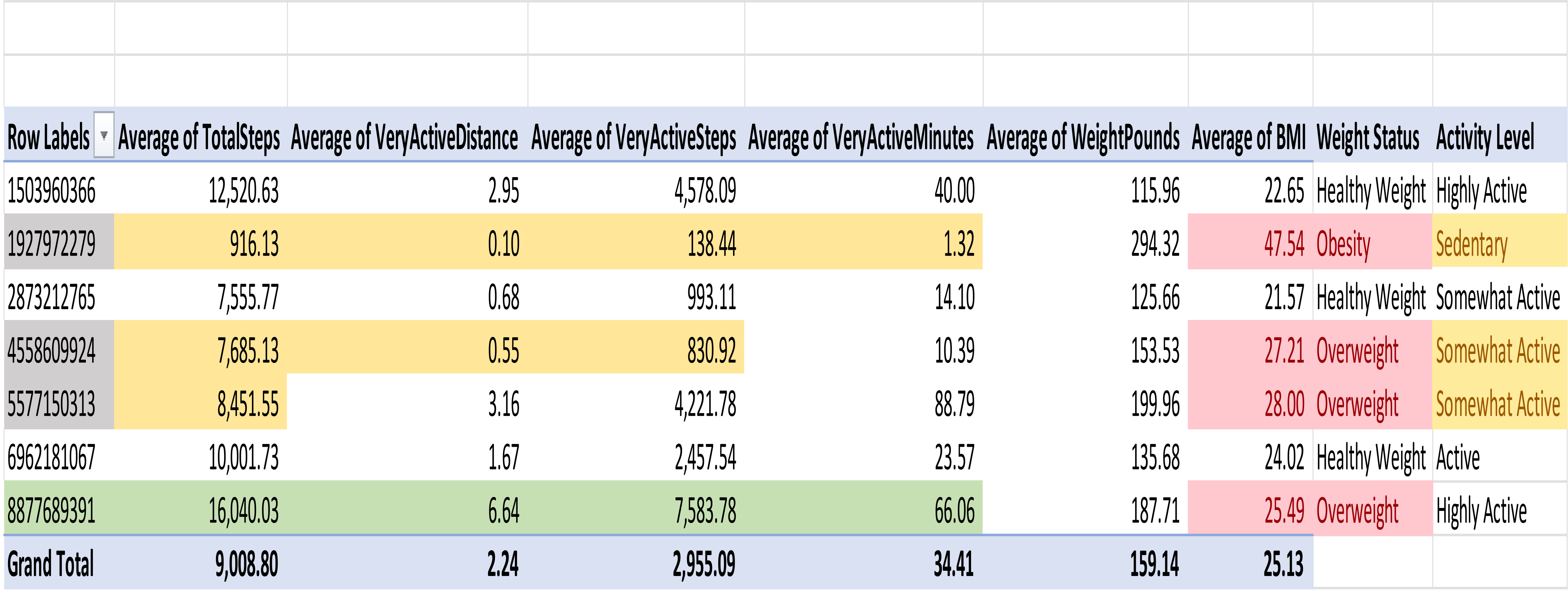


Row Labels	Average of TotalSteps	Activity Level	Weight Status
1503960366	12520.63333	Highly Active	Healthy Weight
1927972279	916.1290323	Sedentary	Obesity
2873212765	7555.774194	Somewhat Active	Healthy Weight
4558609924	7685.129032	Somewhat Active	Overweight
5577150313	8451.551724	Somewhat Active	Overweight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

Data Analysis: Weight Watching

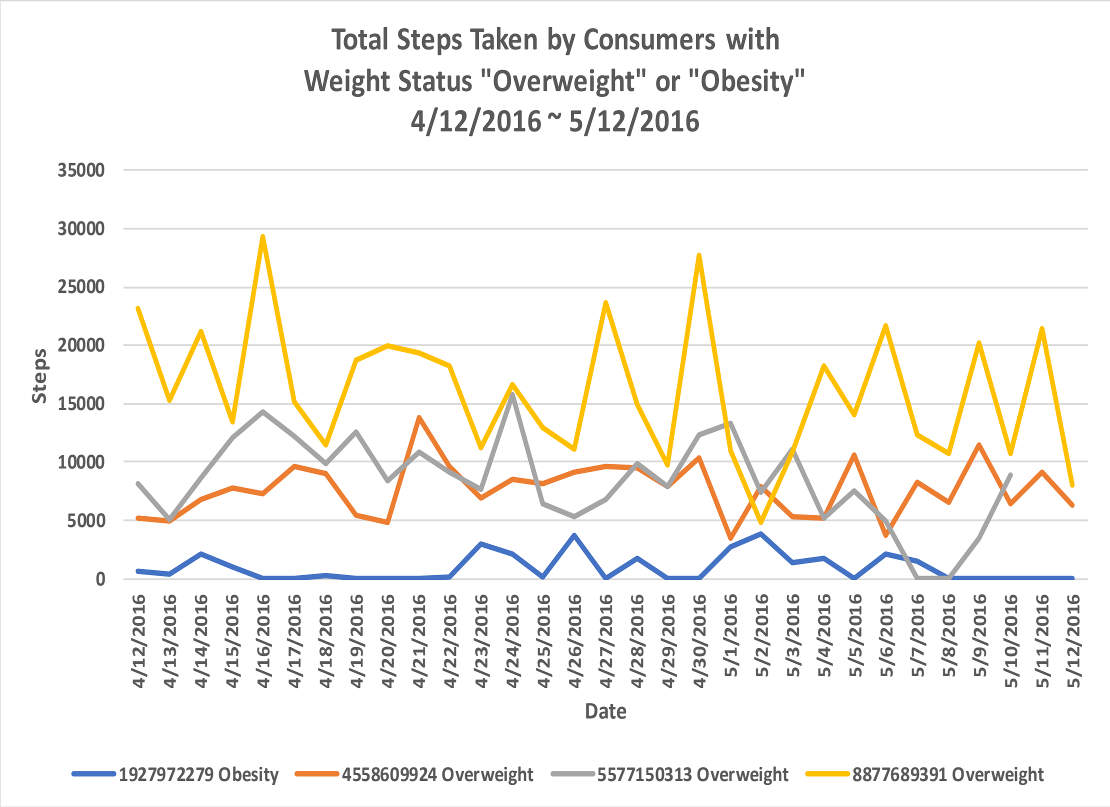
- According to this article "How many steps should people take per day?", a 2018 analysis of 363 people with obesity found that people who walked **10,000 steps a day**, including **at least 3,500 steps** engaging in **moderate-to-vigorous activity** lasting 10 minutes or longer, had increased weight 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

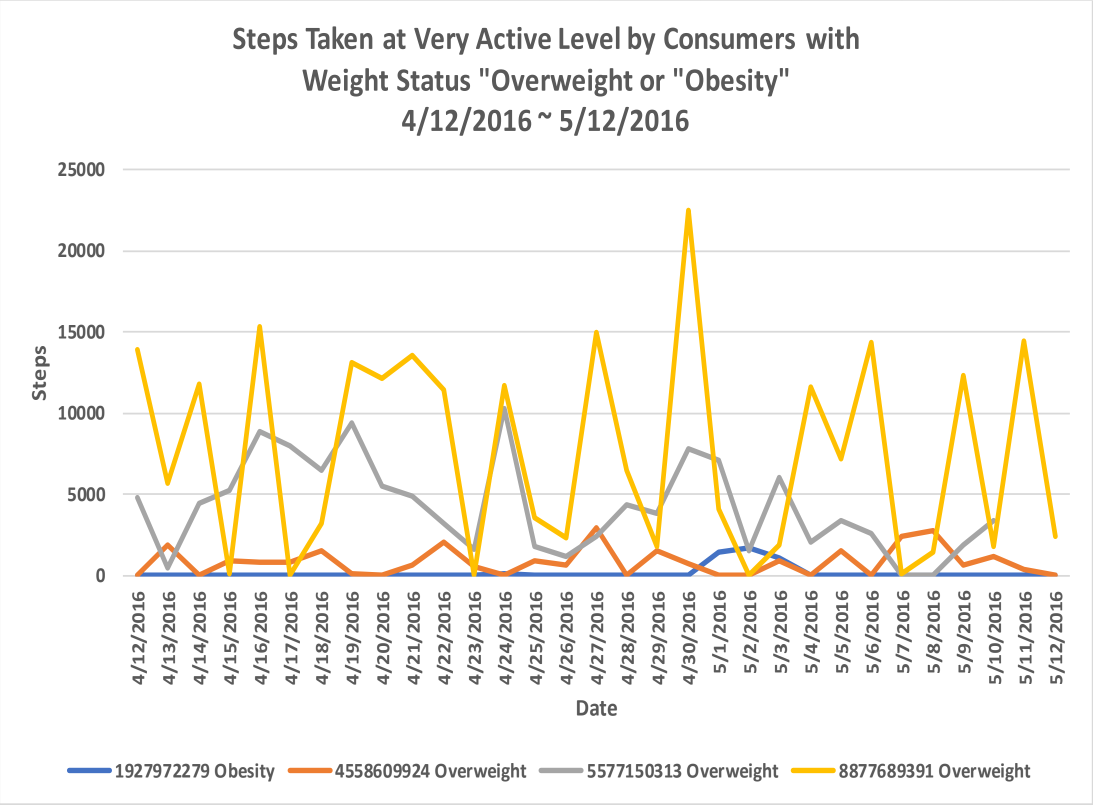


Row Labels	Average of TotalSteps	Average of VeryActiveDistance	Average of VeryActiveSteps	Average of VeryActiveMinutes	Average of WeightPounds	Average of BMI	Weight Status	Activity Level
1503960366	12,520.63	2.95	4,578.09	40.00	115.96	22.65	Healthy Weight	Highly Active
1927972279	916.13	0.10	138.44	1.32	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

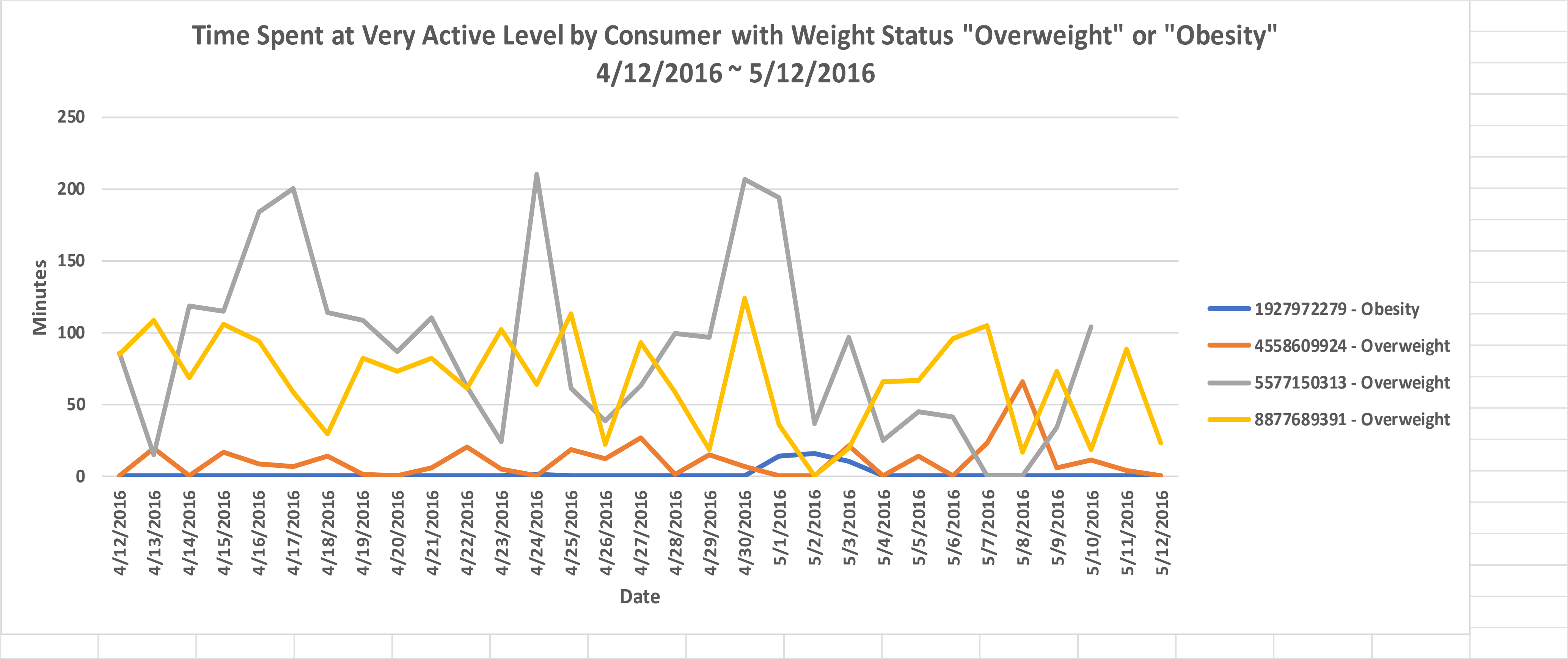


Total Steps Taken by Consumers with Weight Status "Overweight" or "Obesity" 4/12/2016 ~ 5/12/2016



Steps Taken at Very Active Level by Consumers with Weight Status "Overweight or "Obesity" 4/12/2016 ~ 5/12/2016

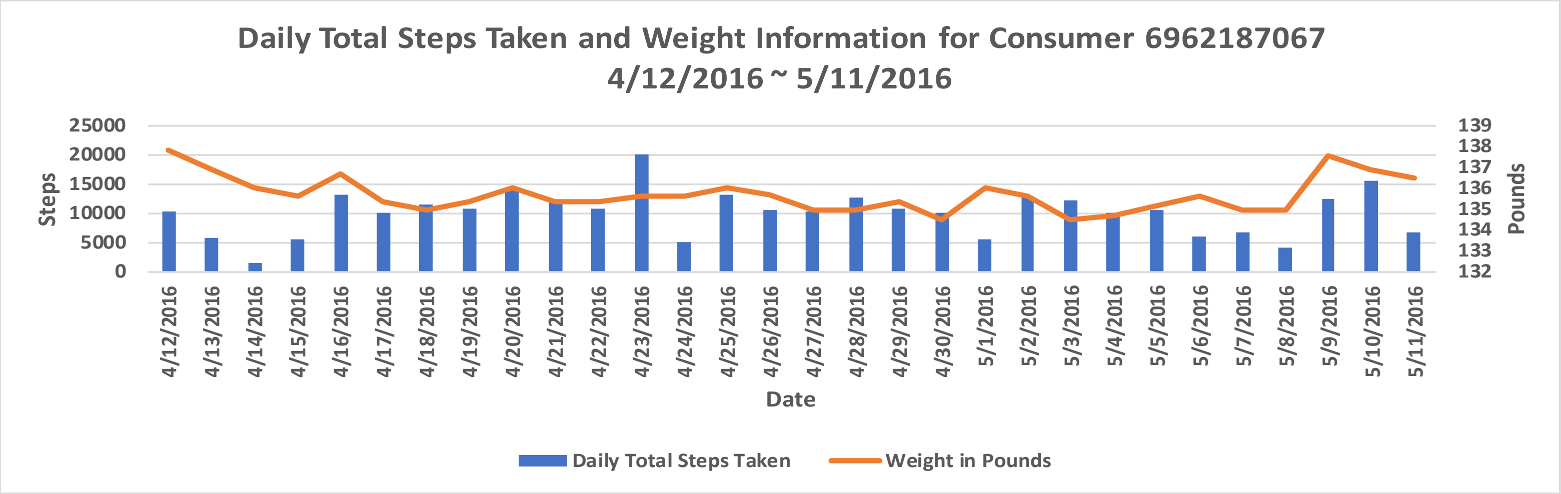
Data Analysis: Weight Watching

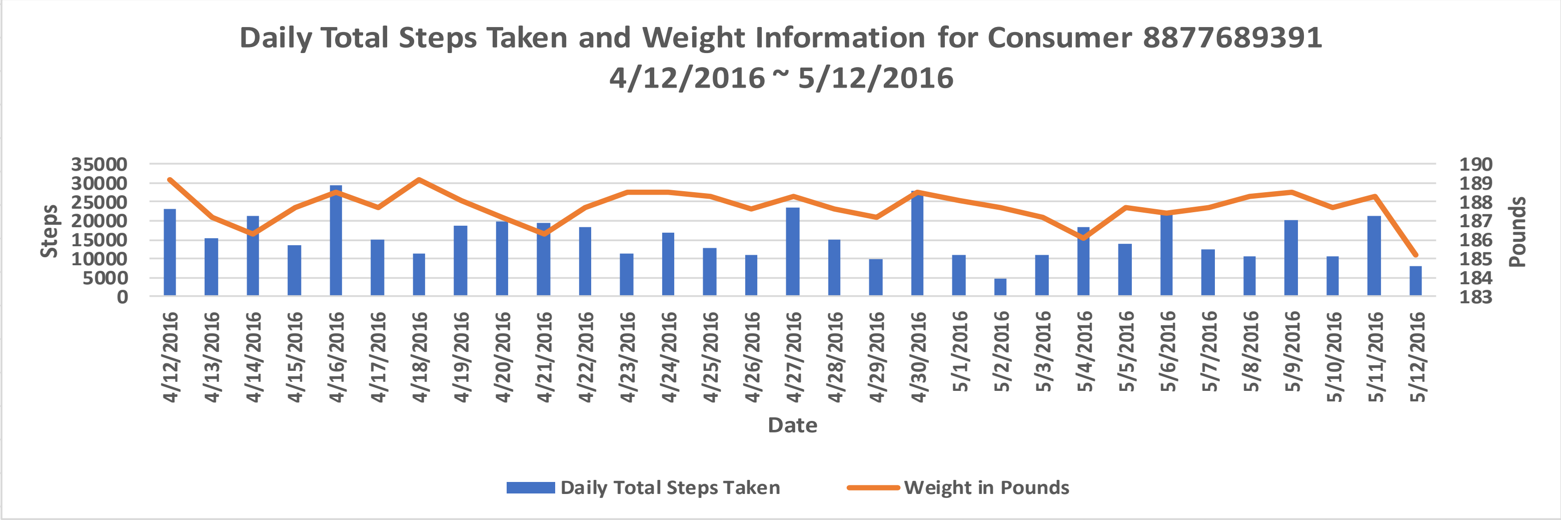


Data Analysis: Weight Watching

- 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


Daily Total Steps Taken and Weight Information for Consumer 6962187067
4/12/2016 ~ 5/11/2016


Daily Total Steps Taken and Weight Information for Consumer 8877689391
4/12/2016 ~ 5/12/2016

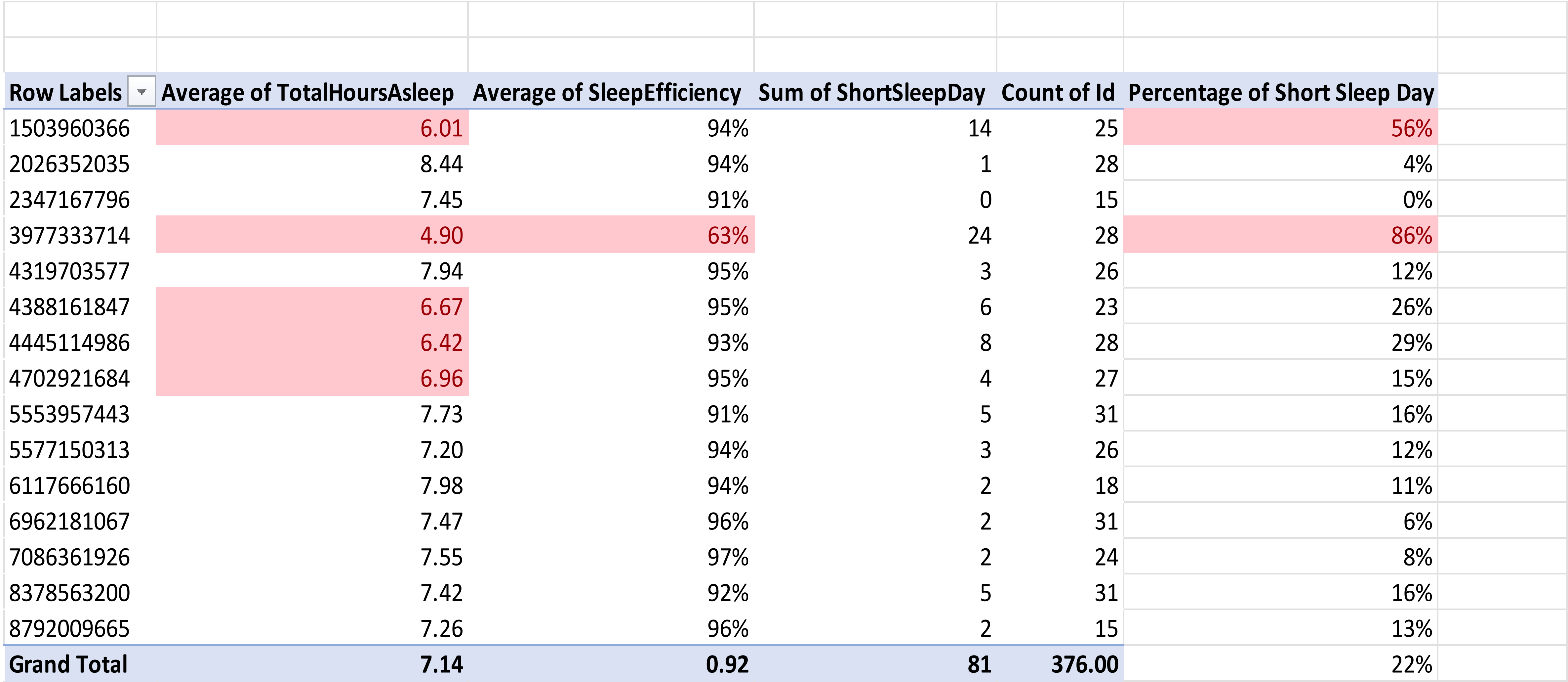
Data Analysis: Sleep Duration and Sleep Efficiency Tracking

- 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

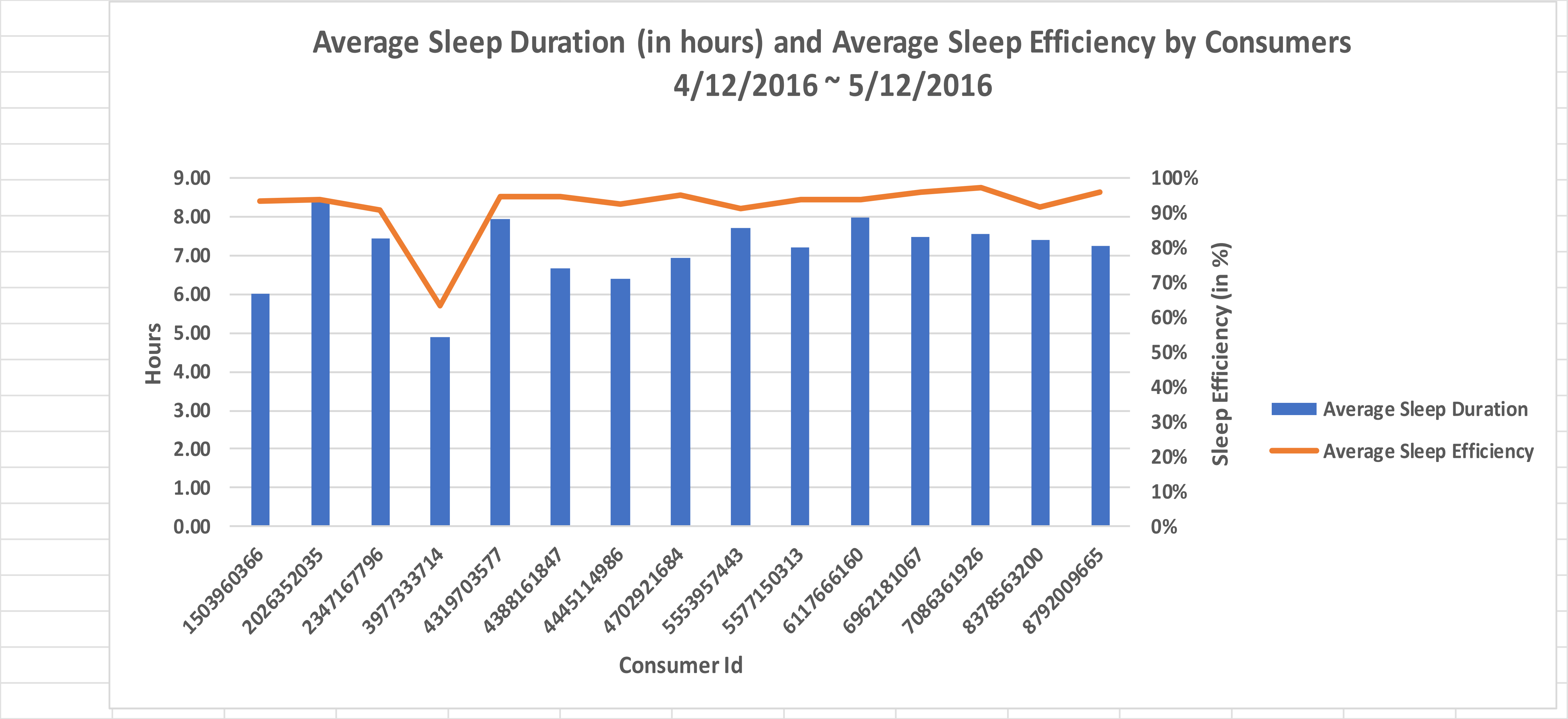
- Created a Pivot Table and calculated the average of TotalHoursAsleep and SleepEfficiency by consumers. Also got the sum of the ShortSleepDay by consumers. Divided the sum of ShortSleepDay by the count of Id (number of rows associated with the consumer) and got percentage of short sleep day. The Pivot 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.