BELLABEAT CASE STUDY

Kishren Jay

Introduction

Bellabeat creates smart health products including the Bellabeat app, Leaf, Time, and Spring. Additionally, it provides users with access to a membership programme that is based on subscriptions and provides them with curated guidance on leading healthy lives. Bellabeat has established itself as a female-focused, tech-driven wellness brand.

Bellabeat has made significant investments in digital marketing, making use of Google search, and participating in social media. The co-founder Sršen is aware that a review and analysis of Bellabeat's customer data would highlight further growth prospects.

Business tasks

Smart gadgets are utilized extensively in their daily lives. Bellabeat may profit from understanding the trends of smart device usage from the data collected and developing data-driven business strategies to investigate growth potential in the sector.

- Analyse data from non-Bellabeat smart devices to understand consumer needs.
- Study how these trends could apply to Bellabeat customers.
- Investigate how these trends can be implemented into Bellabeat marketing strategy.

Preparing the Data

- The **Fitbit Fitness Tracker Data** that Mobius made public and was saved on Kaggle is the data that was used in this study.
- Tracking data from **33** unique Fitbit users. Data from consented users include minute-level output for physical activity, heart rate, sleep monitoring, daily activity, calorie intake and sleep duration that can be analysed to investigate user trends.
- The dataset has in total **18 files** in **.csv** format organized in long format.
- Files that were selected for the analysis:
    - dailyActivity_merged.csv
    - dailyIntensities_merged.csv
    - dailySteps_merged.csv
    - dailyCalories_merged.csv
    - sleepDay_merged.csv
    - weightLogInfo_merged.csv

Processing the Data

- **Tool** : **Google Spreadsheet**
  Reason : Google Spreadsheet was sufficient to handle the data amount and number of rows while providing easy data cleaning and visualisation for the analysis.
- **Data Cleaning**
  - Date formats were mismatched with string and integer formats but were aligned using spreadsheet function.
  - Day of the week data was obtained using the **TEXT(column,"dddd")** function.
  - Zero value columns, TotalDistance, Tracker Distance, and LoggedAcitvies were deleted from the dailyActivity_merged.csv dataset.
  - TotalSteps column data that had values <100 steps were deleted due to improper use of Fitbit tracker.
- **Data Processing**
  - Used the **UNIQUE()** function to obtain the distinct user Ids and found inconsistent unique Ids in each dataset:

Dataset	Number of Unique Id
dailyActivity_merged.csv	33
dailyIntensities_merged.csv	33
dailySteps_merged.csv	33
dailyCalories_merged.csv	33
sleepDay_merged.csv	24
weightLogInfo_merged.csv	8

  - BMI column was extracted from weightLogInfo_merged.csv and calculated to be categorized into **BMI_Class** column using the following formula:

```
fx | = IF(H2 <= 25, "Healthy", IF(25>=H2>30,"Overweight", IF(H2>=30,"Obese")))
```

  - Utilized the **AVERAGE()** function to obtain averages for all the multiple tracker data for a unique type of Id.
  - TotalMinutesAsleep column was extracted from sleepDay_merged.csv file and calculated to be categorized into **Sleep_Quality** column using the following formula:

```
fx | =IF(F2<420,"Lack of Sleep",IF(F2>480,"Oversleep", "Optimum Sleep"))
```

  - Created new columns called **AvgTotalActiveMins/Day** & **AvgTotalNonActiveMins /Day** in dailyActivity_merged.csv dataset:

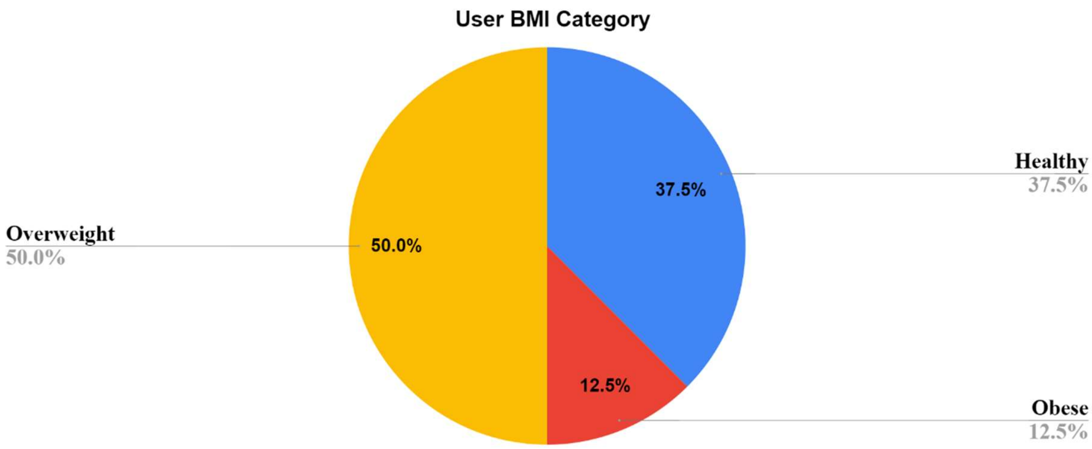
AvgTotalActiveMins/Day = AvgVeryActiveMins/Day + AvgFairlyActiveMinutes/Day + AvgLightlyActiveMinutes/Day

AvgTotalNonActiveMins /Day = AvgSedentaryMinutes/Day + AvgTotalMinutesAsleep

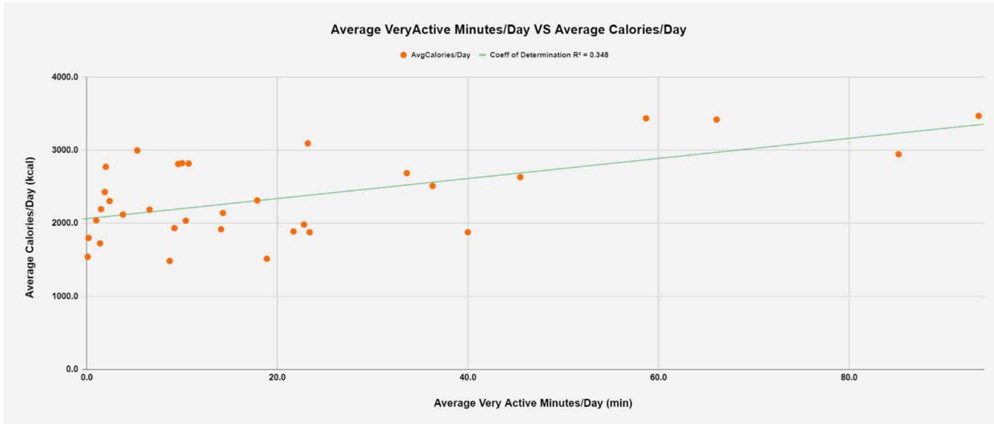
Data Insights & Visualisation

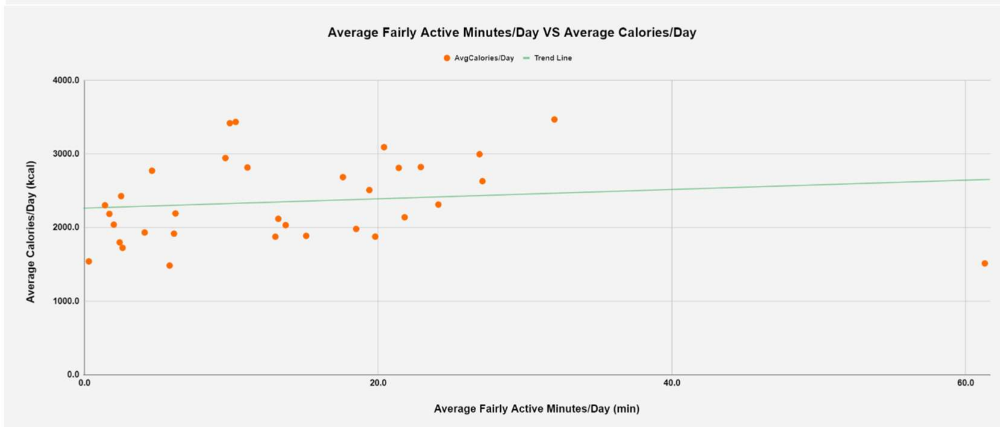
From the datasets, several hypotheses were initially made whereby there are relationships between:

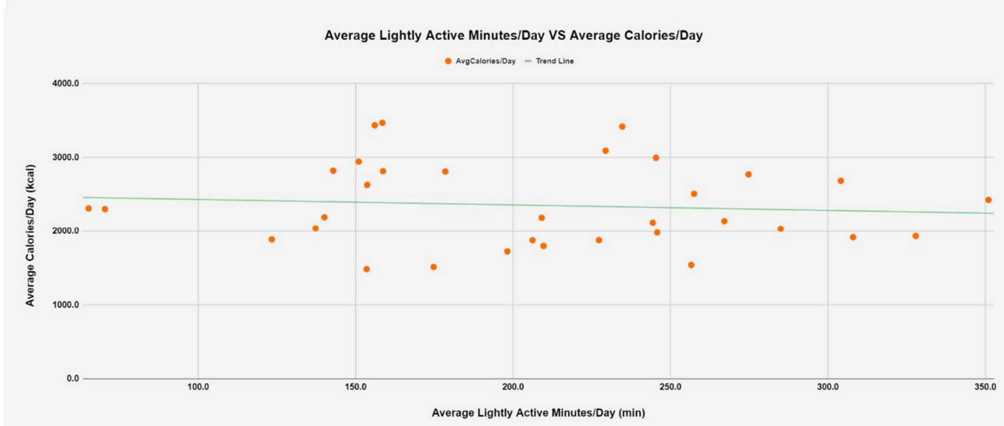
- User daily activity levels and calories per day
- User daily activity levels and amount of sleep per day
- User daily activity levels and user BMI

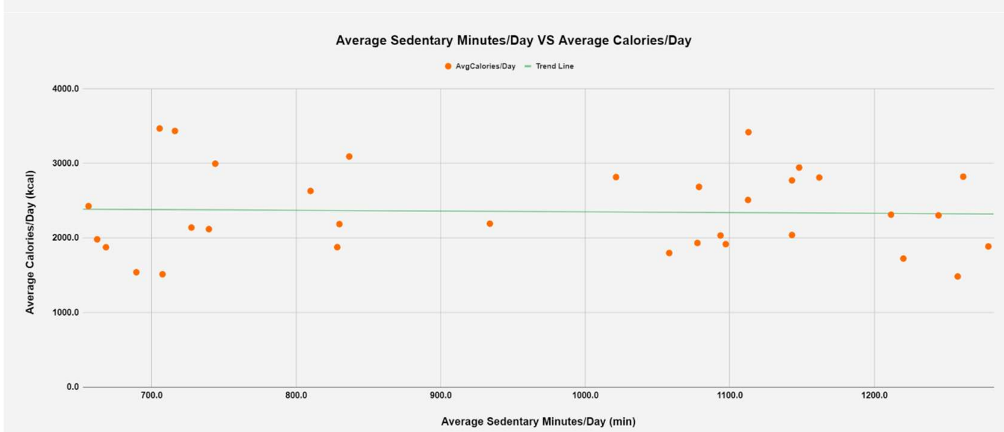


From the users that keyed in their BMI information, it can be shown that within the small sample size, 50% of the users are classified as overweight, while 37.5% are healthy and 12.5% of the users are obese.


Average VeryActive Minutes/Day VS Average Calories/Day
AvgCalories/Day
Coeff of Determination R² = 0.348
Average Calories/Day (kcal)
4000.0
3000.0
2000.0
1000.0
0.0
0.0
20.0
40.0
60.0
80.0
Average Very Active Minutes/Day (min)

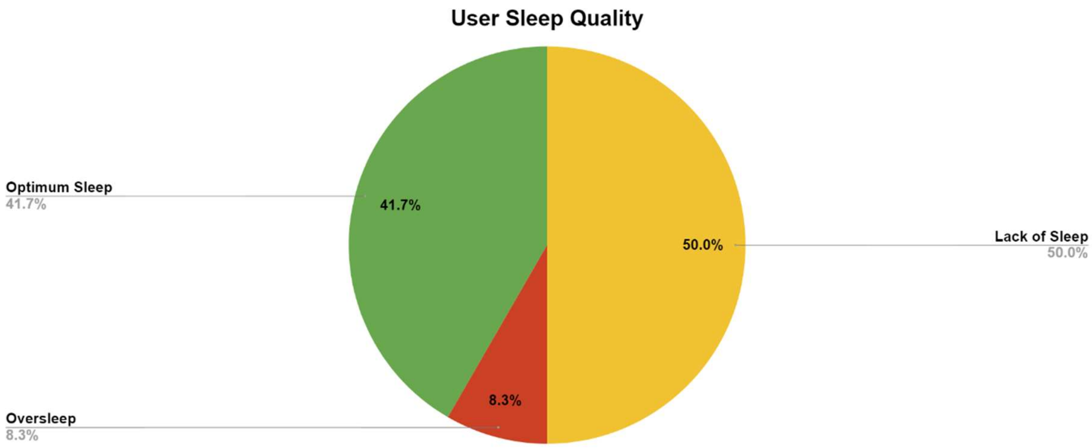

Average Fairly Active Minutes/Day VS Average Calories/Day
AvgCalories/Day
Trend Line
Average Calories/Day (kcal)
4000.0
3000.0
2000.0
1000.0
0.0
0.0
20.0
40.0
60.0
Average Fairly Active Minutes/Day (min)


Average Lightly Active Minutes/Day VS Average Calories/Day
AvgCalories/Day
Trend Line
Average Calories/Day (kcal)
4000.0
3000.0
2000.0
1000.0
0.0
100.0
150.0
200.0
250.0
300.0
350.0
Average Lightly Active Minutes/Day (min)

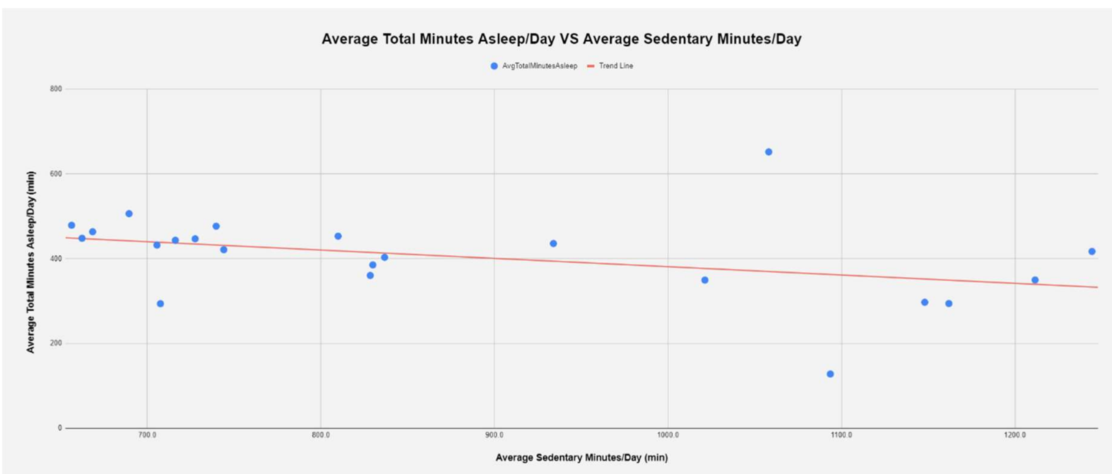

Average Sedentary Minutes/Day VS Average Calories/Day
AvgCalories/Day
Trend Line
Average Calories/Day (kcal)
4000.0
3000.0
2000.0
1000.0
0.0
700.0
800.0
900.0
1000.0
1100.0
1200.0
Average Sedentary Minutes/Day (min)

From the scatter plots above when comparing user activity level to the calories, the users that engage in more very active or fairly active minutes show a positive correlation to more calories burned per day.

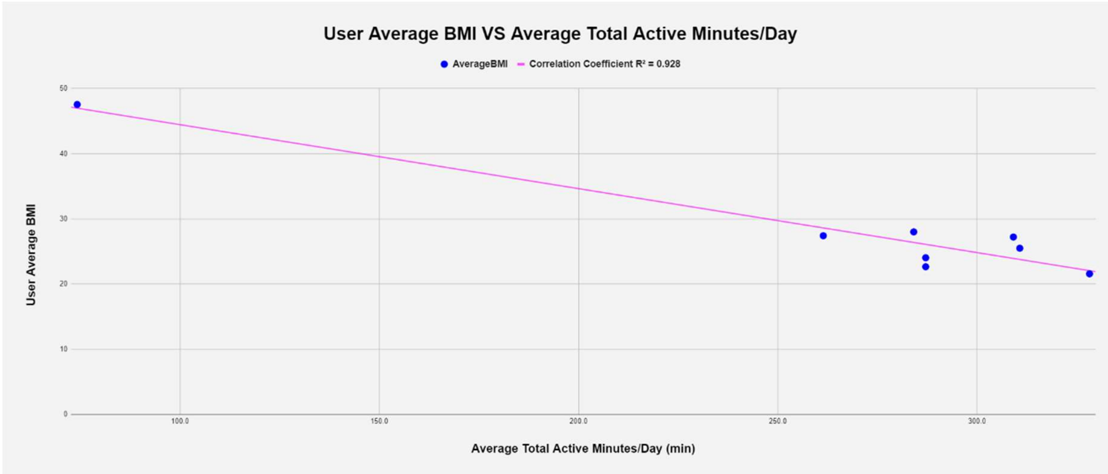
Meanwhile the scatterplots for users that engage in a sedentary or lightly active lifestyle have a correlation coefficient <0.1 which indicates there is no confident relativity to calories burned.

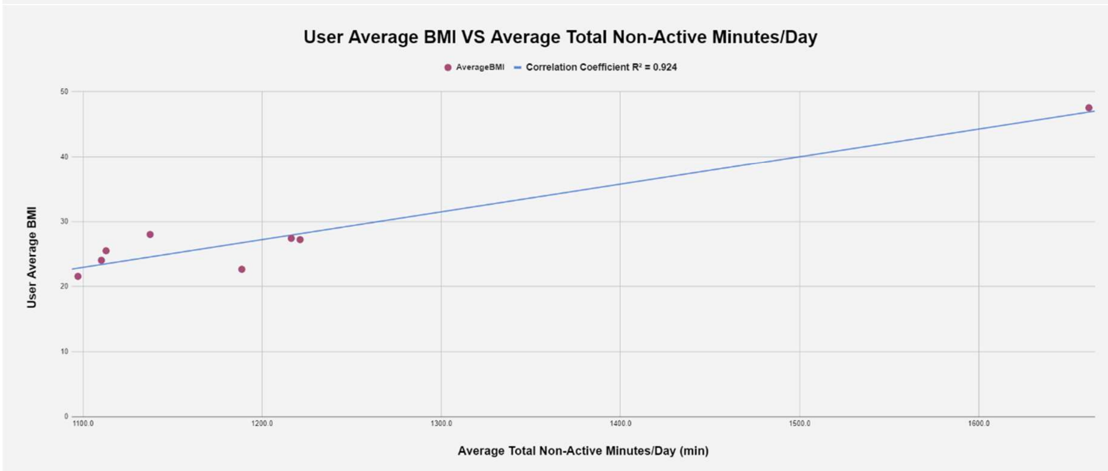


From the sample group of users that keyed in their total sleeping minutes, the categorization showed that 50% of users suffer from lack of sleep (< 7 hours), while 43.7% have optimum sleep (7 to 8 hours) and 8.3% tend to oversleep (> 8 hours).



This scatterplot visualisation shows the negative relationship of a user's average total minutes asleep each day and the user's average sedentary minutes each day. More user data is required to have a more confident correlation of the 2 variables.





The above visualisations indicate a strong relationship with a **correlation coefficient of 0.92** whereby users that are not active throughtout the days tend to have a higher average BMI that indicates them to either be in the overweight or obese category. The small sample size needs to be noted as only a handful of users keyed in their respective BMI.

Conclusion

From the analysis on the available tracker data, the main relationships discovered were:

- Users that spend more time being very active or fairly active have a positive relationship to the average amount of calories burned daily. As for minutes spent being lightly active or sedentary, a confident correlation cannot be said to exist within this sample population.
- Users who are more sedentary tend to have less optimum amount of daily sleep.
- Users who spend most of their being non active tend to have a higher/unhealthier BMI.

Recommendations

In order for Bellabeat to strive within the smart device industry, the results from the analysis of the tracked user data can be taken note to make data driven decisions that can assist in their marketing strategies that would lead Bellabeat to profit and grow within the sector.

1. Bellabeat should focus on ensuring data recording of users is and easier or automated task to ensure users don't neglect keying in required information. A connectivity function to a smart weighing scale should be implemented to track their user's weight for the ease of BMI calculation.
2. Bellabeat can implement a haptic feedback feature to the Time or Leaf product to alert users when they are too sedentary after a certain amount of time or spend too much time in bed without sleep.
3. Bellabeat's subscription plan that comes with a customized user experience should also be added with a point system whereby users get points for completing healthy daily routine objectives and are able to exchange the points for rewards.
4. Bellabeat can share data insights such as trends regarding the relationship of sedentary minutes and total minutes asleep or daily active minutes and BMI to the members via campaigns or built-in notifications in the Time product to educate the users on the benefits of keeping healthy.
5. Bellabeat should take note that the data collection is a priority as the more users routinely log in their data, the more trends can be analysed and discovered and implemented to conduct business profiting decisions.