Data Analysis Case Study - Bellabeat

Bellabeat is a manufacturer of the unique, discreetly-profiled Leaf which has the potential to disrupt the saturated market of more traditional wrist-worn smart-devices. Publicly available datasets provide information on how these devices are used to calculate fitness metrics, in addition to features which are common amongst the top-rated smart-devices. An analysis of this data provides an insight into how smart-device users are overall more fit than non-users, and how these devices hold great potential for enhancing sleep quality. Furthermore, these insights provide Bellabeat with a tailored marketing strategy to highlight the outstanding features of the Leaf, which go beyond the capabilities of traditional and more expensive competing smart-devices.

ASK

The company Bellabeats produces a range of wellness gadgets, such as the Leaf, designed to enhance the fitness of women through the integration with the Bellabeat app. An analysis of smart-device usage and fitness-tracking data, which is already publicly available, will allow for an identification of trends in the way smart-device customers use their devices and thereby will provide insights into how Bellabeat can increase profits. These insights will aim to answer the following three questions:

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

PREPARE

Two datasets were used for this analysis. The first is the FitBit Fitness Tracker Data set sourced from Kaggle [1] under the CC0: Public Domain licence. These data are the result of an Amazon survey with a population of thirty participants, which detail their fitness activity as logged by a smart device. The identity of participants is not included in the dataset. This dataset is in long format. The dataset follows ROCCC guidelines for credibility, however one concern is that the gender of the participants is not disclosed. Therefore, it will not be possible to identify differences in women's smart-device usage compared to that of men, and hence no specific recommendations can be made specifically for women's needs. The second data set on Kaggle contains approximately 500 records of smartwatches available on the Indian market with specifications such as selling price, product colour and average customer ratings [2]. Cleaning the second dataset was much more straightforward, and it involved filtering the dataset to just the highest-rated 100 smartwatches out of the full list.

PROCESS

For this case study, the tools Google Sheets and RStudio have been used. Since the largest dataset contains thirty entries, the small size is well-suited to the open-access Google

Sheets spreadsheet application, where data cleaning can take place. For analysis and detailed visualisations, RStudio will be used.

The first dataset was cleaned by formatting the date/time fields which were inconsistent throughout all the records, e.g. April was given as ‘04’ and ‘4’ which led to inconsistent formatting of the date cells. Whitespace was also trimmed. The simplest way to correct the formatting problem was to use a regular expression to extract the MM/DD/YYYY and disregard the time by using the split function, which was always set to midnight. Since the month was always before October ‘10’, only the first digit of MM was needed. A regular expression extracted M/DD/YYYY and a ‘0’ was concatenated to this as a string. Durations were also converted from HH:MM:SS to minutes. This allowed for simpler comparisons later on in the analysis. The second dataset required formatting the price column as a ‘price’. The data was then cleaned, where excess whitespace was trimmed on multiple entries. Both data sets were downloaded and stored locally.

ANALYSE/SHARE

Using Google Sheets, a refined list of 24 customers was created where the list contained people whose data entries were present in the activity sheet and sleep records sheet. This allowed for a more complete analysis as joining the two datasets ensured that the same people were considered during the analysis stage, i.e. any customer with no sleep records was discarded by using Google Sheets IF statements. The same customers had many fitness observations over a period of days, and the most relevant ones were returned using VLOOKUP. Averages were calculated across the complete date range for each fitness attribute to enable meaningful comparisons between customers by using Google Sheets functions. Performing an analysis on the sales price of the 100 most popular smartwatches involved using a current exchange rate of Indian Rupee (INR) to Australian Dollars (AUD) of 1 INR = 0.017 AUD.

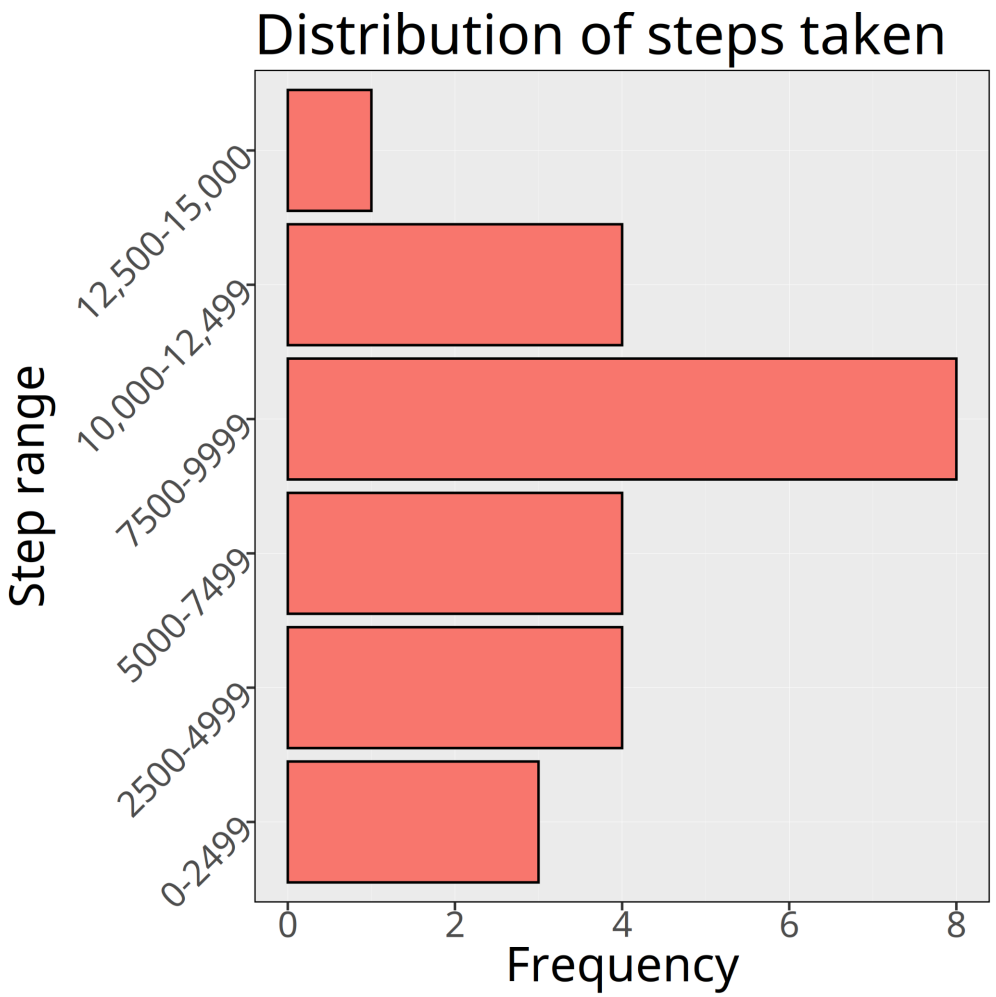

Figure 1: A frequency distribution of the amount of steps taken in a single day. Each bar represents the amount of customers who walked a number of steps falling into their respective step range category. The results indicate that most customers walked between 7500 and 9999 steps. This is in accordance with the amount of steps recommended by Harvard Health.

To compare the baseline of physical activity that each customer maintains, an analysis was made of the average amount of steps taken per day. The results indicate that most users took between 7500 and 9999 steps. According to Dr. Anthony Komaroff, the Editor in Chief of Harvard Health, the minimum recommended amount of steps for an average senior adult is 7500 steps [3]. The findings of the analysis indicate that approximately 60% of users exceed this target. A possible explanation is that owners of smart-devices are more conscious of their overall fitness compared to people who do not use them. Bellabeat could highlight this finding in its advertisements to boost the appeal of owning a Leaf device. They could also could improve their product by providing tailored reminders to its users to meet basic fitness targets, such as steps taken via the app.

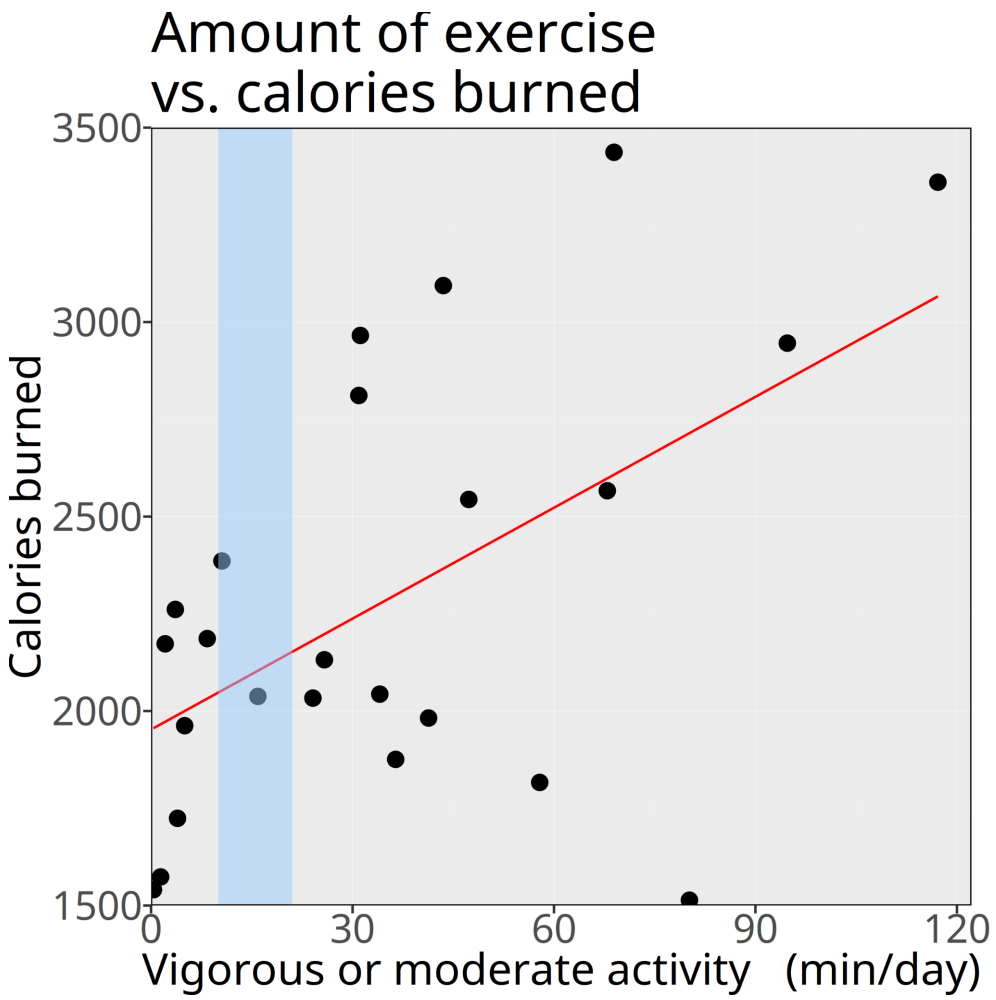

Figure 2: A scatterplot showing the correlation between the amount of vigorous or moderate activity done in a day versus the amount of calories burned. The solid black dots represent the data from a single customer. The blue shaded region identifies the recommended amount of (semi)strenuous exercise per day according to W.H.O.. The red line is a regression analysis, which outlines the positive correlation between amount of exercise and burned calories.

To follow up on the possible benefit of smart devices on the fitness of its users, further analysis was performed on the average amount of moderate to vigorous activity per day compared to average calories burned per day. To qualify the amount of exercise as sufficient, the World Health Organization recommendations of 75 to 150 minutes per week [4] serve as a benchmark (the blue shaded region in figure 2). The results indicate that 30 % of users perform below par, and 70 % of users meet the recommended levels. According to a government health survey, on average, only 43 % of adults meet the recommended activity requirements [5]. Therefore, we can conclude that smart device owners significantly outperform the average adult in terms of moderate to vigorous activity. Bellabeat should focus on this result in their marketing strategy as a clear benefit of owning one of their devices, especially for targeted calorie goals, e.g. weight loss.

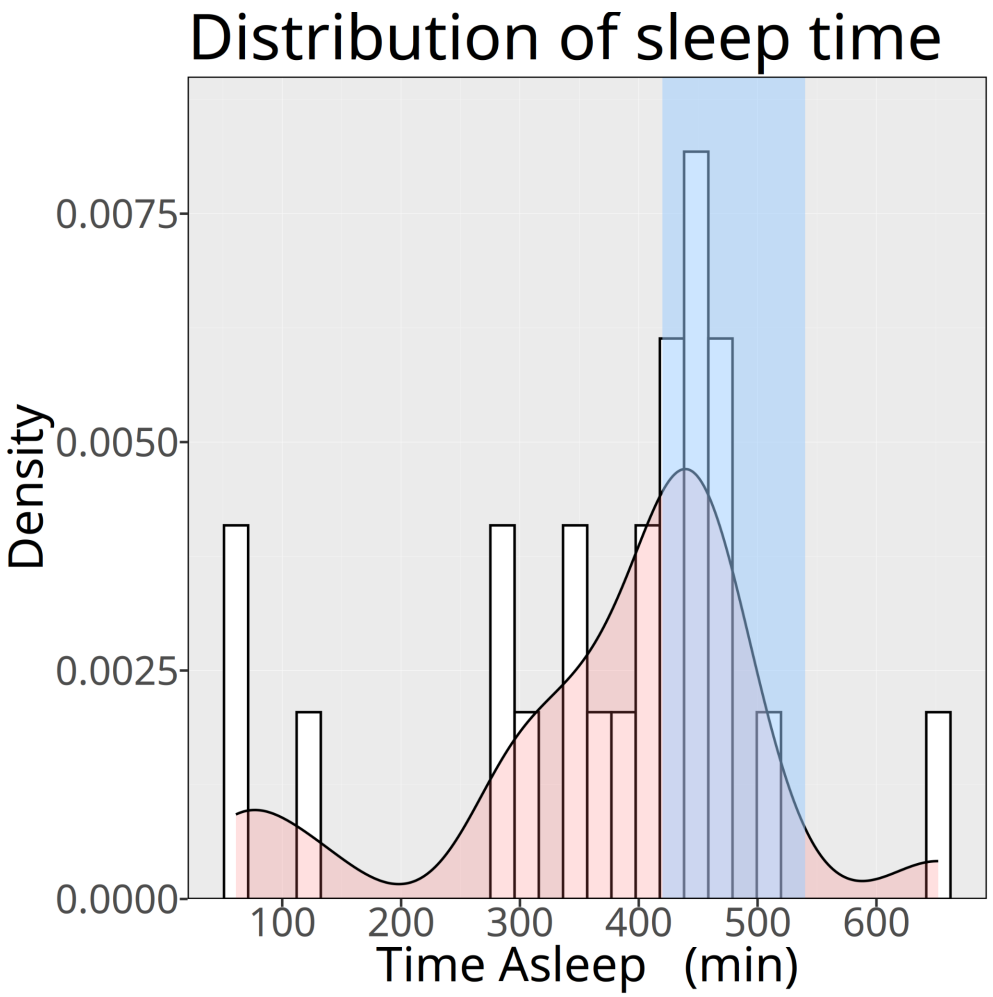

Figure 3: A frequency distribution of the amount of sleep across all customers. The white bars represent the frequency binned for total time asleep. The shaded red region shows the density across the frequency distribution, indicating the mode to be 440 minutes of sleep per night. The blue shaded region indicates the recommended amount of sleep per night as per health guidelines. These results indicate that 50 % users sleep less than the recommended amounts of sleep per night.

Aside from exercise, sleep is a major factor of overall health. One appealing feature of smart-devices is their ability to track sleep quantity and quality and giving the user feedback on how they can make improvements. To take a snapshot of this capability, the sleep duration of 24 participants was analysed and plotted as a frequency histogram as shown in figure 3. The white bars indicate the relative frequency of the respective amount of sleep, with each bar binned over 20 minute increments of sleep. The red shading represents the density of each sleep time, where the area under the curve adds to 100 % of the users. More common sleep times correspond to a taller density. The shaded blue region indicates the ideal amount of sleep, i.e 7-9 hours, as recommended by the Sleep Health Foundation [6]. These results indicate only 50 % of users meet these requirements, which is surprising given the importance of sleep. Bellabeat can exploit this by providing a well-marketed holistic ‘points based’ health score including sleep quality, where sleep is scored depending on whether it meets its target duration. Reminders can be issued throughout the day via the app.

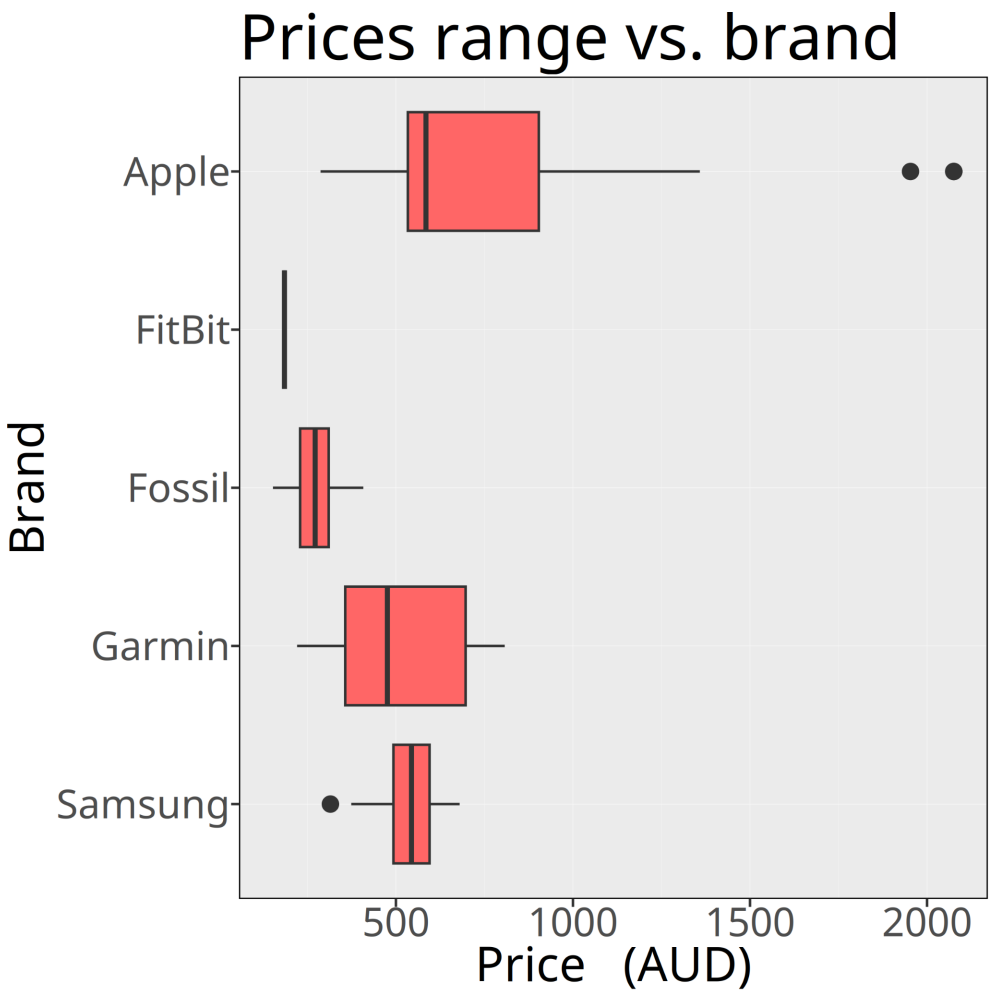

Figure 4: A box and whisker plot showing the price distribution for each brand of smart device in the list of top 100 devices. Each brand has a corresponding box and whisker plot, where the median price is represented by the black solid line within each box, the first and third interquartile range by the left and right edges of each box respectively, and the range of prices represented by the lines protruding (whiskers) from each box. Outliers are shown as a solid black dot. In the case of FitBit, only one data point is present and so the single price is represented by a solid line. These results indicate that Apple smart devices have the most expensive products, however on average, all brands have most of their devices around a common price point of 500 AUD.

A comparison was made of the distribution of prices per brand of smart device featured in the top 100 as per customer satisfaction. Apple was the most popular brand of device as well as having the greatest price variance. The median cost of a smart device regardless of brand is approximately 500 AUD. The Bellabeat Leaf is significantly cheaper than these, at 150 AUD. Taking into account the lack of a screen to keep a discrete profile, the device is a very economical investment as it performs all of the health tracking functions of other branded smart devices for 30 % of the price. Bellabeat can highlight the economy of their device in their marketing and present the lack of screen in a positive light, focusing on the unique discreet profile it offers.

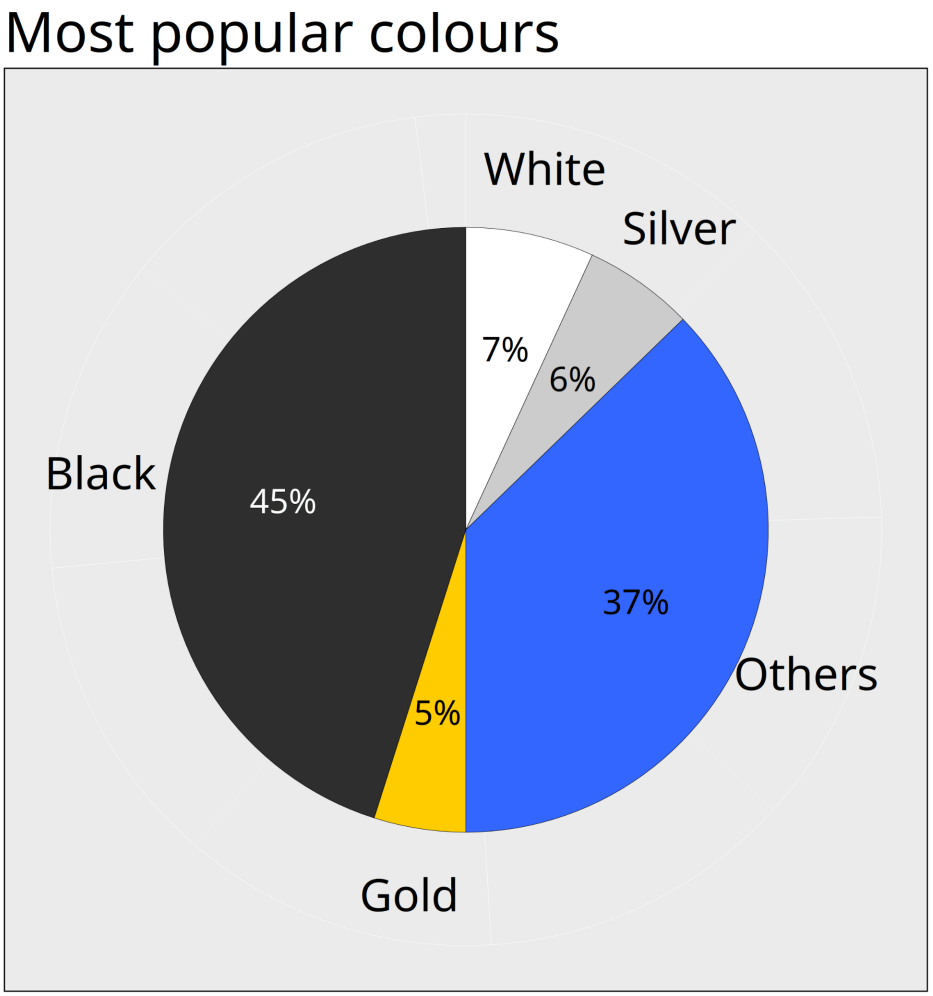

Figure 5: A pie chart showing the percentage of each colour amongst all most popular colours. The results indicate that by far, black is the most popular colour, occupying almost half of all devices. The other most popular colours are white, silver and gold. All other colours, of which there are twenty remaining, occupy less than 40 % of all remaining device colours.

The most popular colour of smart device is black, occupying 45 % of all top-rated devices. Since the Bellabeat device is black, it already occupies a safe position of favourability. The unique aspect of the Bellabeat Leaf is the diverse range of accessories and bracelets, many of which are white, silver and gold, i.e. the other most popular colours. The marketing team could focus on the different possibilities the device offers, as its ability to be customised is far beyond what other branded smart devices offer. Additionally, it can be worn as a bracelet, a necklace and a brooch. Bellabeat could showcase more design combinations featuring these most popular colours in their advertisements to highlight their large range of customisation, which sets them apart from competing brands.

CONCLUSION

This case study demonstrates a few trends in smart-device usage, namely the positive effect of owning a smart-device has on a user's overall level of fitness, where users are 70 % more likely to meet or exceed exercise recommendations compared to 43 % of the average population. Additional trends on sleep monitoring show that at least 50 % of users meet the recommended amounts of sleep, with lots of room for improvement which can be achieved by relaying tailored feedback to the user via the Bellabeat app. These findings all contribute to a bespoke, optimised marketing strategy whereby advertisements should combine their focus on traditional smart-device features, such as exercise and sleep tracking, in combination with the economical pricing of the Leaf device, the benefits of its screenless discrete profile, and the wide range of possibilities in how the Leaf can be worn and customised.

REFERENCES

[1] FitBit Fitness Tracker Data, *Pattern recognition with tracker data: Improve Your Overall Health,* provided by user 'Möbius'. URL: https://www.kaggle.com/datasets/arashnic/fitbit?resource=download

[2] Fitness Trackers Products Ecommerce, *Different products from various Brands with specifications, ratings & reviews,* provided by user 'Devashree Madhugiri'. URL: https://www.kaggle.com/datasets/devsubhash/fitness-trackers-products-ecommerce?resource=download&select=Fitness_trackers_updated.csv

[3] A. L. Komaroff, M.D., "How Many steps should I take each day?", *Harvard Health Letter,* (2019), URL: https://www.health.harvard.edu/staying-healthy/how-many-steps-should-i-take-each-day

[4] Fact Sheet: Physical activity, World Health Organization (2022), URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity

[5] Australian Health Survey: Physical Activity, Australian Bureau of Statistics, (2013), URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/australian-health-survey-physical-activity/latest-release

[5] "How much sleep do you really need?", *Sleep Health Foundation,* (2016), URL: https://www.sleephealthfoundation.org.au/how-much-sleep-do-you-really-need.html