Bellabeat report
Author: Lingfeng Yang

Executive Summary

Bellabeat is a manufacturer of high-tech health-focused smart devices for women. Urška Sršen and Sando Mur, Chief Creative Officer and mathematician respectively, co-founded Bellabeat based on the idea of smart products catering to the needs of women. As a successful fast-growing company within the smart device market, Bellabeat has the potential to compete against larger market competitors by employing strategies that unlock growth opportunities within the market. Data analysis was conducted by the marketing analytics team to provide insights on new strategies and market opportunities.

Products

Bellabeat currently offers five products to consumers in the smart-device health market:

- **Bellabeat App**: Provides user health data related to activity, sleep, stress, menstrual cycle, and mindfulness habits. Connects to their line of smart products
- **Leaf**: wellness tracker that can be worn as a bracelet, necklace, or clip. Connects to the app to track activity, sleep, and stress.
- **Time**: a smartwatch to track activity, sleep, and stress
- **Spring**: A smart water bottle to track hydration levels
- **Bellabeat membership**: Membership program that gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health/beauty, and mindfulness

Marketing and Sales

At the current stage of Bellabeat's development as a company, the overarching strategy has a focus on company and user growth. The company has two primary sales channels; Online retailers and the company's own website. Bellabeat's marketing strategy employs both traditional advertising and current digital media, with a heavier focus on the latter.

Marketing Strategies:

- Traditional Media: radio, out-of-home billboards. print, and television
- Digital Marketing:
    - Google Search,
    - Facebook and Instagram pages,
    - active twitter engagement.
    - Youtube ads
    - Ads through Google Display Network to support marketing campaigns around key dates

Task

How can Bellabeat increase its smartwatch marketing strategy to increase consumer reach, competitiveness, and sales?

The objective of this report is to describe the process, detail the findings of the analysis, and provide high-level recommendations for growth. Focusing on Bellabeat's Time smartwatch; how can the company increase its smartwatch marketing strategy to increase consumer reach, competitiveness, and sales?

Insights from the data analysis can provide a view into how the smartwatch is primarily being used, gaining an understanding of the primary target base. This can answer questions such as; "*is there any room for expansion or downscale (more specific targets)*", and "*What are the primary functions that users use smartwatches for?*".

Process
Tools
Data Source: Kaggle
Original Format: CSV
Data Storage: On the cloud in Google Drive and Google BiqQuery
Data Manipulation: Google BigQuery SQL
Data Visualization: Tableau
Reports & Documentation: Google Doc

Data
The primary data source used in the analysis was public Fitbit user data acquired from Kaggle under CC0: Public Domain license. The data contains logged information about various health factors such as activity level and sleep from 33 unique users during the period of 2016-04-12 to 2016-05-12.

The data is organized by the date-time variable as well as user Id. The dataset is wide or long depending on the timeframe that the data is organized; a data table organized by minute is longer (narrow) than by the hour, which is wider

The data, originally stored in CSV formats were uploaded to Google BigQuery into separate tables. Data Integrity was verified through Google Big Query's SQL functions. The raw datasets were imported, reviewed, cleaned, and saved into new datasets so as to not damage the original source.*
*Detailed documentation can be found in the "Data Documentation" file

Verification was performed through manual checking on BigQuery with SQL, Some methods of verification include elementary queries on table size, unique users, and variable formats.

Complications
Many of the datasets have problems with timestamp formatting when uploading to BigQuery. This was discovered while in the process of uploading. The STRING format of the timestamp in CSV was not recognizable by BigQuery for automatic conversion to DATETIME. This was caused by the meridian symbol being interpreted as timezone. The solution was for the variable to be uploaded as STRING, reformatted using a custom SQL query, and then converted to DATETIME, allowing the use of built-in functions.

Analysis
Types of data:
1. *Daily_Activity*
   - Cumulation of Daily Calories/Intensity/Steps
   - Describes statistics for tracked days:
     - Steps
     - Distance; distance by activity level
     - Minutes by activity level
     - Calories
2. *Heartrate_Seconds*
   - Heart rate per second over a period of time
3. *Hourly_(Intensities/Steps/Calories)*
   - Cumulative steps were taken over an hour
   - Cumulative calories burned over an hour
   - Cumulative and average intensity level over an hour
4. *Calories/Intensity/METs/Steps_per_Minute (Narrow/Wide)*
   - Cumulative steps over a minute
   - Cumulative calories burned over a minute
   - The intensity level at a certain minute
   - MET rate at a certain minute
   - Available in narrow or wide format
5. *Minute_Sleep*
   - A binary value indicating whether the user is asleep at a time (minute intervals)
   - Time when the user is asleep
6. *Sleep_per_Day*

- A binary value indicating if the user logged a sleep record
- Minutes the user is asleep
- Minutes the user is in bed

7. *Weight_Log_Info*
   - Log of Weight(kg/lbs), fat percentage, and BMI at a certain date
   - Binary indicator whether the log was manually recorded or automatically recorded

Findings

Why are users using smartwatches (Fitbit)?
- *To track physical activity/sleep/weight?*
  Calculate the proportion of data logs that can be attributed to daytime activity (rested and active) to sleep. Distributions of the per-minute factors (Steps, Intensity, METs, Heartrate, Calories) help show the proportion.
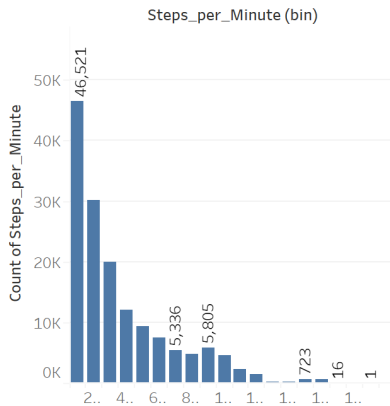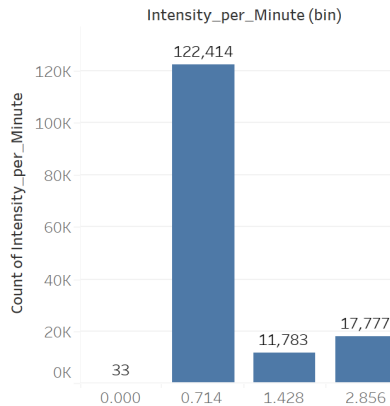
Per_Minute_Distributions
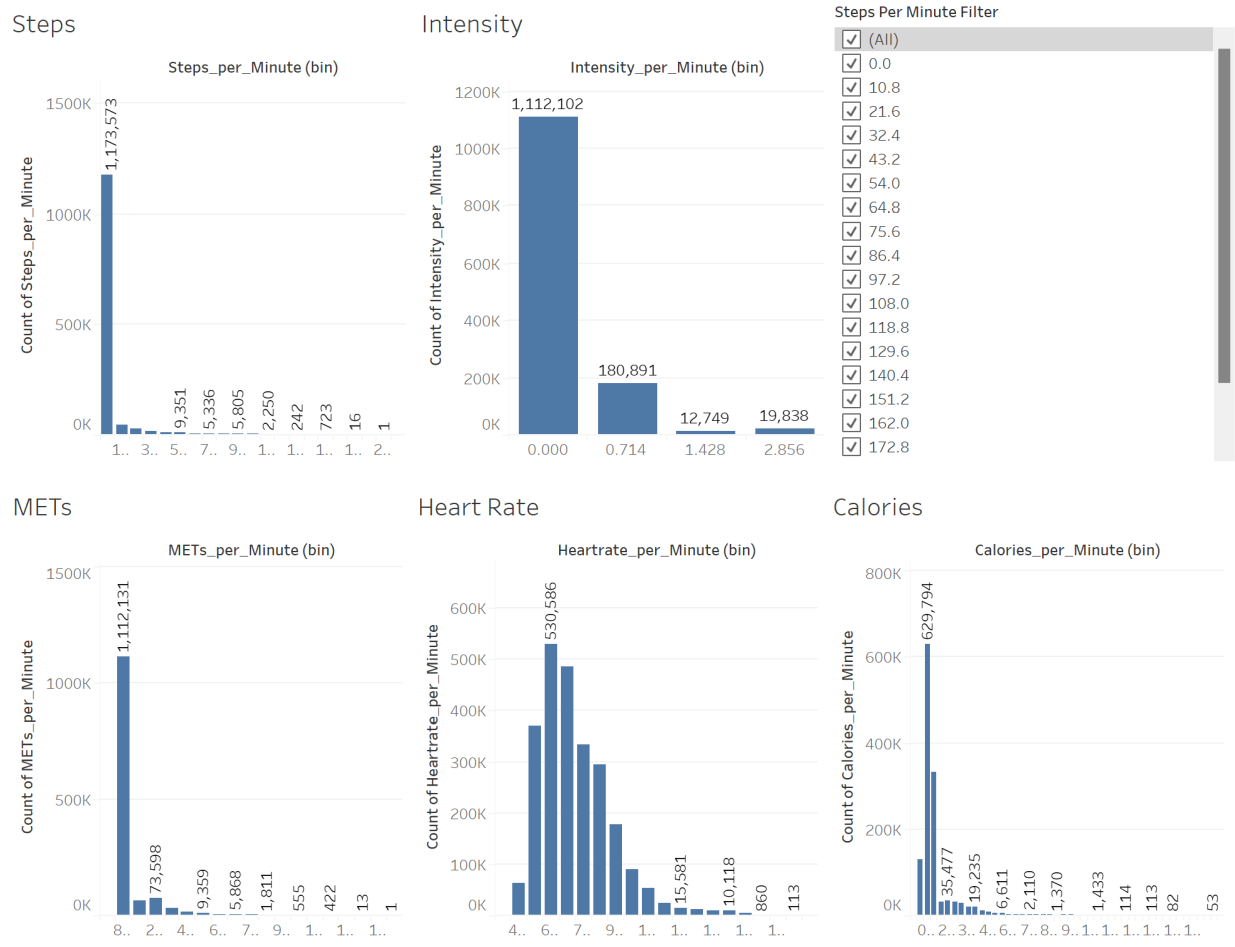
Distributions of all factors without filter
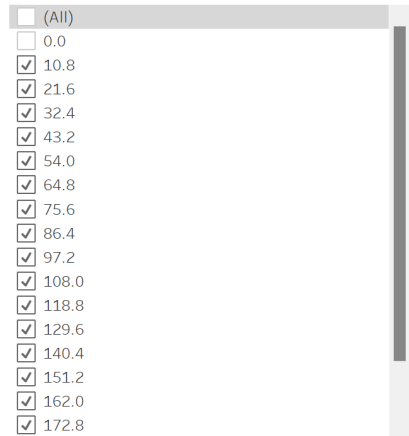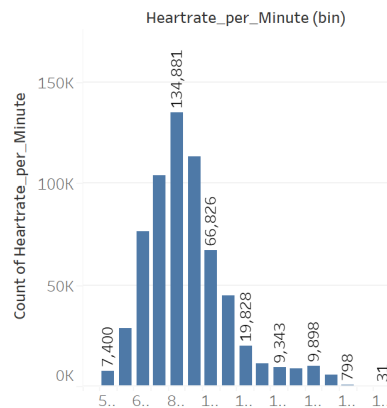
Per Minute Distributions

Steps

Steps_per_Minute (bin)

Intensity

Intensity_per_Minute (bin)

Steps Per Minute Filter

- [ ] (All)
- [ ] 0.0
- [x] 10.8
- [x] 21.6
- [x] 32.4
- [x] 43.2
- [x] 54.0
- [x] 64.8
- [x] 75.6
- [x] 86.4
- [x] 97.2
- [x] 108.0
- [x] 118.8
- [x] 129.6
- [x] 140.4
- [x] 151.2
- [x] 162.0
- [x] 172.8

METs

METs_per_Minute (bin)

Heart Rate

Heartrate_per_Minute (bin)
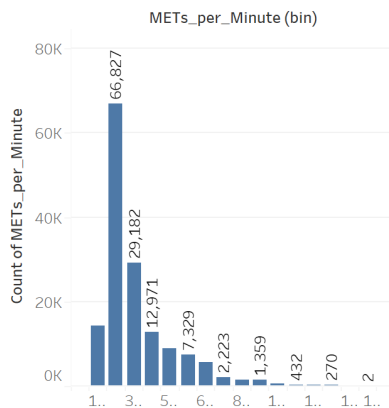
Calories

Calories_per_Minute (bin)
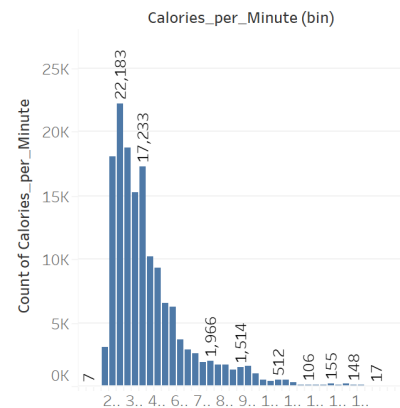
Distributions filter to remove when Steps per Minute = 0

Can see a majority of the logged measurement occur during rest (steps = 0, intensity=0, and resting heart rate/METs/calorie levels). Inferring that activity logs where steps_per_min = 0 is the user at rest (84.92% of steps_per_min logs are zeros) All distributions are heavily right-skewed. We can infer that majority of logs are used during rest/sleep.

Steps_Zeros

Non-Zeros
15.08%

Zeros
84.92%

84.92% of the per-minute data points are when steps per minute = 0.
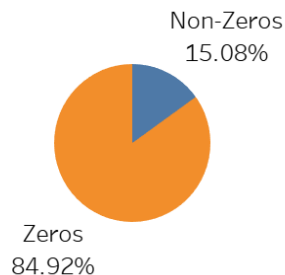
Improve physical performance?
It is hard to determine the type of physical activity based on given factors. The most probable type of activity would be walking/hiking/running can be determined by distance and activity intensity; high intensity and low distance may be other forms of exercise Primary users of the data to track running. The ratio of high intensity, high distance to high intensity, low distance

Improving sleep?
As a large majority of logs in the dataset occur when steps are 0, it is important to analyze the sleep patterns of the users. The dataset provides information on the sleep time and time in bed of the users.

Sleep_Over_Time
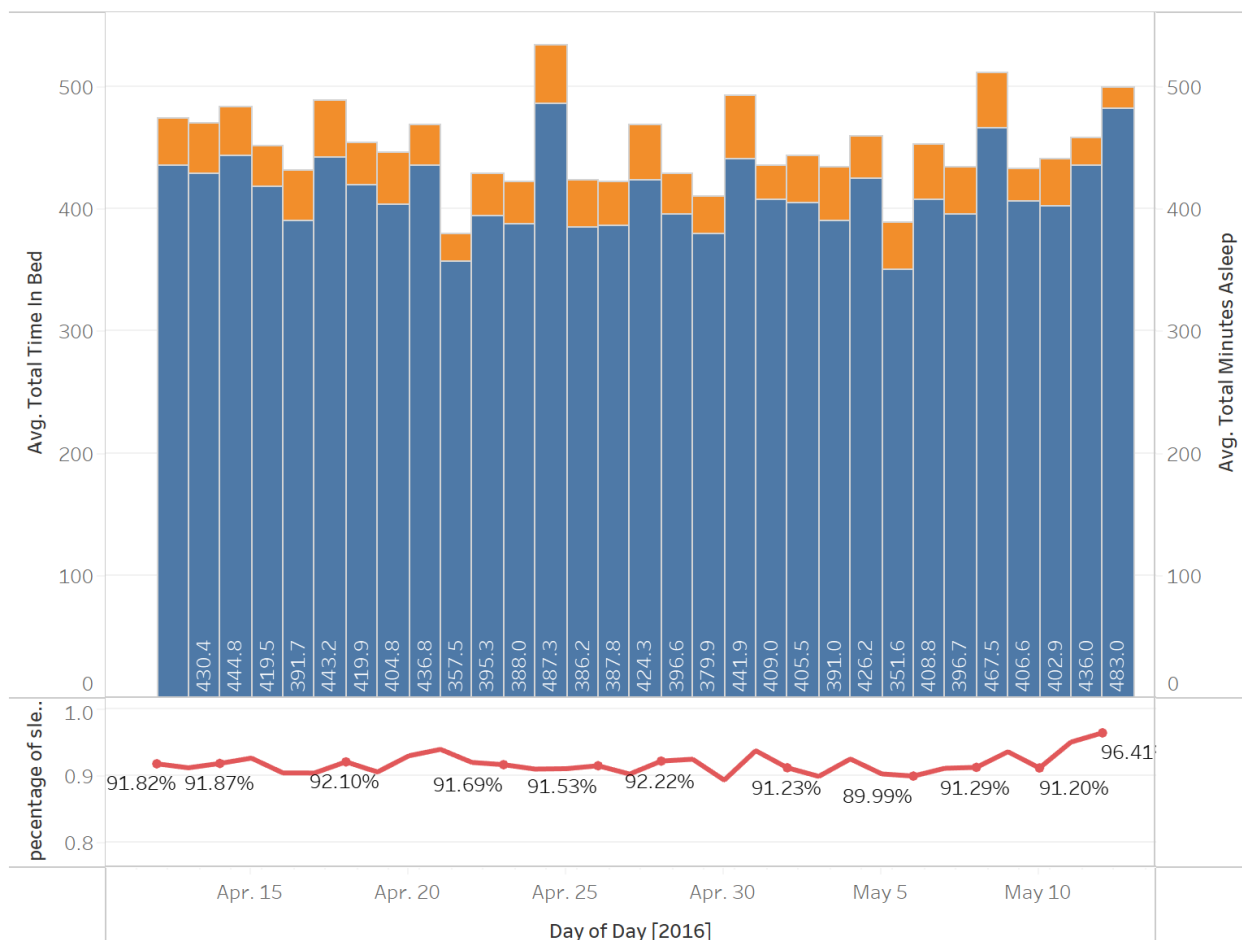
There is no noticeable improvement in both the number of minutes asleep and the percentage of minutes in sleep to minutes in bed. On average, users spend 452.45 minutes in bed and 413.80 minutes asleep. The percentage of time asleep to time in bed does not show a significant improvement over the course of the data time period.

While the analysis does not show improvement, a major reason for users to use smartwatches can be attributed to tracking their sleep patterns due to the proportion of data logs.

Improve general health (weight)?
Analyzing the improvement of weight/fat/BMI over the logged period using a time series of the factors.

Time series of weight/BMI/fat

There is no noticeable improvement in weight/BMI/fat. The data has sparse logs for weight and BMI, as well as containing many null values for fat. This is an unreliable analysis of the usage of smartwatches for tracking health measures. Although the data contains a variable indicating the log to be manually or automatically saved, it is not enough to provide a reliable analysis.

The time period may not be long enough for noticeable changes to occur in these factors. In general, improvement in these factors would take months of time, not available in the dataset

Limitations
Weight
A bias was discovered after the analysis was conducted. The data collect information on a relatively short time period, this would skew the data away from meaningful discoveries about user's usage to track weight. Biologically, a person's weight and health do not change significantly in a month, it would be hard to tell from the data if the user's intentions are to improve health/weight conditions.

Activity Type
Another limitation of the data is that it does not differentiate between the type of physical activity performed when tracking activity intensity. This can be a result of the technological limitations of a smartwatch as the most reliable measure to be collected are steps, heart rate, and location. Unless the user manually defines the type of activity, there is no concrete method to differentiate between, for example, running and biking. Rough estimates can be deduced based on a combination of activity intensity, distance, location, speed, heart rate, etc. While it is presumed that generally, most users will use a watch to track aerobic exercises, without more specific data, it is difficult to provide further recommendations based on customer segmentation of activity type.

Recommendation
Using this dataset, we can infer that a majority of smart-watch users purchase the product to track and maintain their fitness activity levels and progress, as well as sleeping patterns. As Bellabeat is looking to improve its operations and products through marketing, there is a large opportunity for the company to focus on the Bellabeat Time smartwatch. Future marketing campaigns should have an emphasis on the benefits of the product when used during those activities.