Case Study:

Bellabeat Business and Fitbit Data

Kolton Ascheman

7/31/2023

Business Task

Bellabeat is a company that creates smart fitness technology for women. These technology products are created to give users a better understanding about their overall fitness and wellness. This study is being done to analyze how current smart devices are being used for health and wellness. This study will focus specifically on the Bellabeat Leaf, a tracking smart device similar to that of the Fitbit, which its use will be looked at in this study.

About the Data

The dataset is FitBit Fitness Tracker Data, uploaded by Mobius on Kaggle. The data comes from a survey that was done by Amazon Mechanical Turk consisting of thirty users' data of fitness activity. The Dataset is CC0: Public Domain. It was downloaded onto a folder, then uploaded to Google Sheets for cleaning and analysis.

The dataset does have some limitations. Only 30 users were recorded, and their overall activity is recorded over a relatively short period of time. While a lot of data is available on the micro level, not as much is available on the macro level.

Limitations of this Analysis

Because this analysis is being done in Google Sheets, spreadsheets from the dataset with a large data size will not be examined in this study. A lot of the spreadsheets for these fitness features have spreadsheets with smaller file sizes that span a longer period of time. For example, sleepDay_merged vs. minuteSleep_merged.

Data Cleaning

The data was cleaned by checking the filters of each column in the following tables: dailyActivity_merged, WeightLogInfo_merged, and SleepDay_merged.

My Approach to Analyzing the Data

When I looked at this data, the question I wanted to answer was what specific features people were using on FitBits. To find this out, I transformed the three following tables into pivot tables, aggregating certain variables to see how many people were using each feature and how often they were using it. Almost all of the chosen variables were aggregated by COUNT for this reason.

Here is how each of the listed tables were transformed for analysis:

- dailyActivity_merged was transformed into a pivot table with unique IDs for every row.
    - TotalDistance was aggregated by COUNT for every unique ID.

        - Each COUNT that was aggregated for each unique ID was aggregated into a COUNTIF that measures that amount of users that recorded a particular number of days. This can be seen in the third bar chart.

    - LoggedActivitiesDistance was aggregated by SUM for every unique ID.

- WeightLogInfo_merged was transformed into a pivot table with unique IDs for every row.

    - WeightPounds was aggregated by COUNT for every unique ID.
    - Fat was aggregated by COUNT for every unique ID.
    - In the weightLogInfo dataset, the column “IsManualReport” had values containing TRUE or FALSE. Conditional formatting was used to sort which users had these values. The values were manually added as a 0 or 1 to each user in the pivot table.

- sleepDay_merged was transformed into a pivot table with unique IDs for every row.

    - SleepDay was aggregated by COUNT for every unique ID.

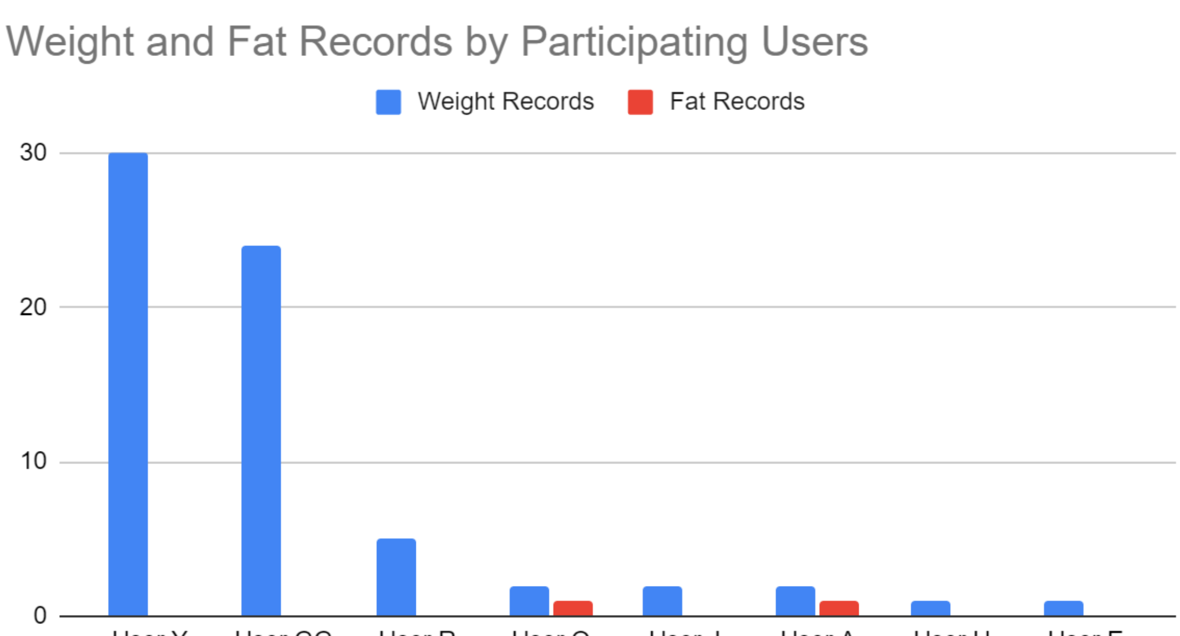
For the purposes of simplifying the tables, the user IDs, which were 10-digit numbers, were changed to letters. This process was done in a separate sheet using VLOOKUP.

Analysis

1. Weight

Of the 30 people in the dataset, only 8 recorded their weight during the timeframe. Of those 8 people:

- 5 people recorded their weight only once or twice
- 1 person recorded their weight 5 times
- 2 people recorded their weight more; 24 times and 30 times



Overall, only 2 to 3 people recorded their weight a substantial amount of times over the given timeframe.

Manual Weight Entry and Fat

41 of the 67 weight records were manually entered. People either entered all of their weight manually, or all of it automatically. No person used both methods. The method of recording weight data was not related to how many times people recorded data.

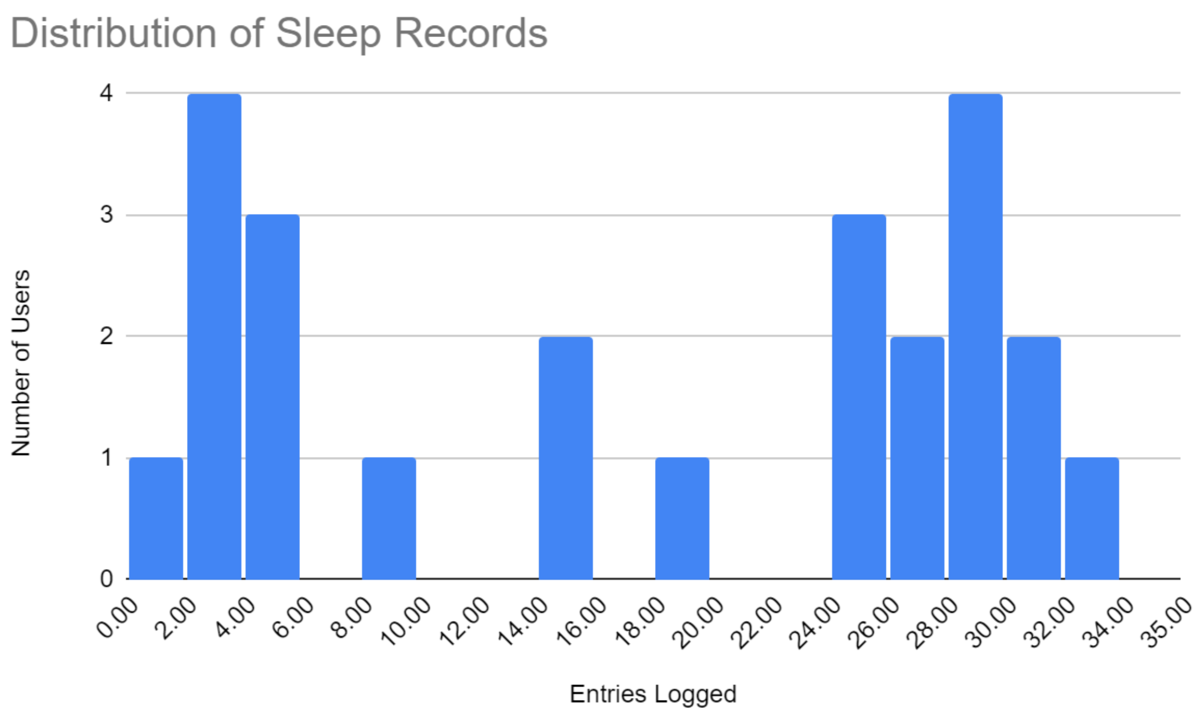
Only two people recorded their fat. Each person recorded their fat one time.

2. Sleep

Of the 30 people in the dataset, 24 people recorded their sleep. 12 people recorded at least 24 days of sleep. The other 12 people recorded their sleep for 1-18 days.
The pivot table below shows the number of sleep records for each user. A histogram following shows the distribution.

Number of Sleep Records by User

User	Count	User	Count	User	Count
User DD	32	User O	26	User R	5
User Y	31	User A	25	User E	5
User T	31	User AA	24	User C	4
User S	28	User P	24	User BB	3
User Q	28	User V	18	User X	3
User L	28	User FF	15	User D	3
User G	28	User I	15	User Z	2
User U	26	User M	8	User H	1



3. Distance and Steps

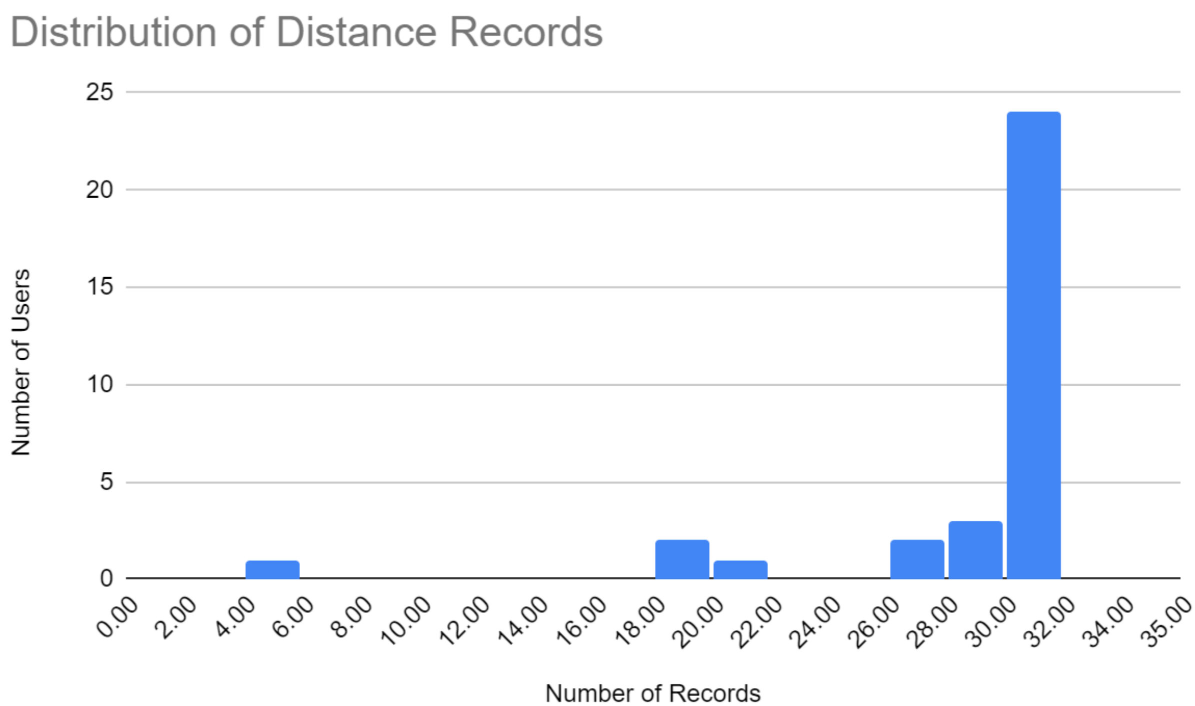
The distance and step features were used by all of the people in the dataset. 26 of the people recorded their distance and steps at least 29 days.

The following pivot table shows the count of distance entries for every user, and a histogram showing the distribution of these counts.

Number of Distance Records by User

User	Distance Entries	User	Distance Entries	User	Distance Entries
User GG	31	User O	31	User L	30
User EE	31	User M	31	User C	30
User DD	31	User J	31	User FF	29
User BB	31	User H	31	User W	29
User AA	31	User G	31	User V	28
User Y	31	User F	31	User Z	26
User T	31	User E	31	User X	26
User S	31	User D	31	User K	20
User R	31	User B	31	User CC	19
User Q	31	User A	31	User I	18

User P	31	User U	30	User N	4



Logged Activities

Only 4 people logged activities.

- 3 people logged activities 1-3 times.
- 1 person logged activities 12 times.

Conclusion

By far, distance was used the most by the people in this dataset. Sleep was recorded by most people, although only one third of people recorded a substantial amount of days. Weight was shown to have little to no use by people. Only two people extensively used this feature. Virtually no one recorded their fat. Logged activities were also rarely used by people, with only one person recording lots of activities.

My recommendation would be to focus on the features involving tracking distance. These are the features that the vast majority of the people in this dataset used. Recording sleep is something that should be looked into more. Depending on demand, it would make sense to focus on this use of the device as well.