Case Study:

bellabeat

Exploring Data Driven Decisions through Data Analytics

By

Ruben-Stuart Kantner

Student, B.Sc Business Administration

Hochschule Hannover, Germany

Contents

Introduction

In this case study I will apply my Data Analytics skills to analyze smart device fitness data to answer a business task. After introducing the company for whom this analysis is made, I will proceed through the phases Ask, Prepare, Process, Analyze, Share and Act of the Data Analysis process as taught in the Google Data Analytics Professional Certificate. Along the way I will identify the Business Task, the Stakeholders, relevant Data, its integrity and validity, proceed to process and analyze the data to gain actionable insights.

The Company

For this case study I will be working with the fictitious high-tech manufacturer of health-focused products for women - Bellabeat. Right now, the company is successful yet small but has the potential to become a large player in the global smart device market. They believe that analyzing and gaining insights from smart device data will enable them to find opportunities for growth.

Products:

Bellabeat offers a variety of health-focused tech devices. These include:

Bellabeat App : The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.

Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.

Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.

Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.

Bellabeat Membership: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

The Task

Bellabeat asks to analyze smart device usage data to gain insight into how consumers use non-Bellabeat smart devices. They then want to have these insights applied to one of their products.

Guiding questions are:

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?

Deliverables:

1. A clear summary of the business task
2. A description of all data sources used
3. Documentation of any cleaning or manipulation of data
4. A summary of the analysis
5. Supporting visuals and key findings
6. Top high-level content recommendations based on the analysis

1 Ask

1.1 The Business Task

Bellabeat produces high-tech smart devices for women's wellness. The company invests heavily in marketing, especially digital marketing. But wants to identify opportunities for growth to inform its marketing strategy.

As of now the company possesses data but lacks insight into how consumers use smart devices which could lead to more efficient use and targeting of marketing channels and product development. To gain comprehensive insight not only does the usage of Bellabeat products need to be analyzed but also that of non-Bellabeat products or smart devices in general.

To this end trends in the usage of non-Bellabeat smart device usage need to be identified and analyzed in such a way that marketing strategy can be informed by the gained insight.

Key Stakeholders for this project are the executives Urška Sršen and Sando Mur as well as the Bellabeat Marketing Analytics Team as part of which I am performing this analysis.

2 Prepare

To solve the business task Data on the usage of non-Bellabeat products needs to be acquired. Sršen provided directions to a dataset. The dataset is public under a CC0: Public Domain License and updated annually. It is titled "FitBit Fitness Tracker Data" and is hosted on the platform kaggle.com by user Möbius.

2.1 The Data

The dataset contains responses from a distributed survey via Amazon Mechanical Turk from 03.12.2016-05.12.2016. In this survey thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring over the course of 31 days.

The data is organized into 18 CSV files detailing daily, hourly and by the minute information on activity, calories, steps, heart rate, sleep and weight. The data is recorded in long format and hosted on kaggle.com.

2.2 Data Verification

Due to the small sample size and limited number of rows in each file the data could be verified as to its integrity using Excel. And the individual table's relationships through the user ID could be confirmed.

2.3 Data Limitations

The data's validity and integrity are limited. Limiting factors are:

- Completeness: Certain categories such as weight and sleep only have data from a fraction of the participants.
- Age: The dataset is almost 9 years old (as of March 2025). User behavior may have changed since the time of the survey in which the data was collected.
- Context: For some columns in this dataset key context is missing. i.e. distances are not labeled miles or kilometers and calories are not labeled consumed or burned and thus cannot be taken into consideration for this analysis.
- Sample Size: With only 30 participants the sample size is very small and opens the data up to concerns of sampling bias. This needs to be considered when applying the insight of this analysis.
- Duration: The timeframe from which the data was collected was only two months (March and April 2016). Such a short period leaves the results vulnerable to outside factors such as weather or other events that may have occurred during that time.
- Additionally, the purpose of the collection of the data is not available.

Relevant data to solve the business task are:

Daily Activity - Can be used to determine how often users use smart devices

Daily Steps - Can be used to determine how many steps the average user takes and whether that makes them active.

Daily Sleep - Can be used to determine how much users sleep and whether more active users sleep better

Hourly Steps - Can be used to determine at what times users are most active

Relevant but excluded due to data limitations are:

Calories (Context), Weight (Sample Size)

3 Process

To process the data for analysis Excel and its PowerQuery functionality is sufficient for data cleaning due to the small size and complexity of the selected data. The Excel extension PowerPivot can be used to create the relationships between the tables using an Entity Relationship Diagram (ERD) within a data model.

3.1 Data Consolidation

In the first step of processing all relevant tables are imported into a single workbook in Excel using PowerQuery to ensure that the original data remains intact.

3.2 Data Cleaning

In the next step each table is checked for duplicates and blank values.

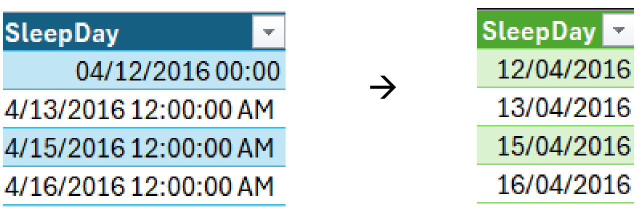
Table Name	Duplicates	Blanks
daily_activity	0	0
daily_sleep	3	0
daily_steps	0	0
hourly_steps	0	0

In this case only the table daily_sleep contained duplicates which are then removed.

3.3 Data Consistency

Following this, all tables are checked to ensure that the values of every column are in an appropriate format and that this format is uniformly applied to all values within the column.

Noticeable here is that in the daily_sleep table the sleepDay column, which details the day on which the sleep was logged, contains the time. This is unnecessary and therefore can be reduced to just the date.



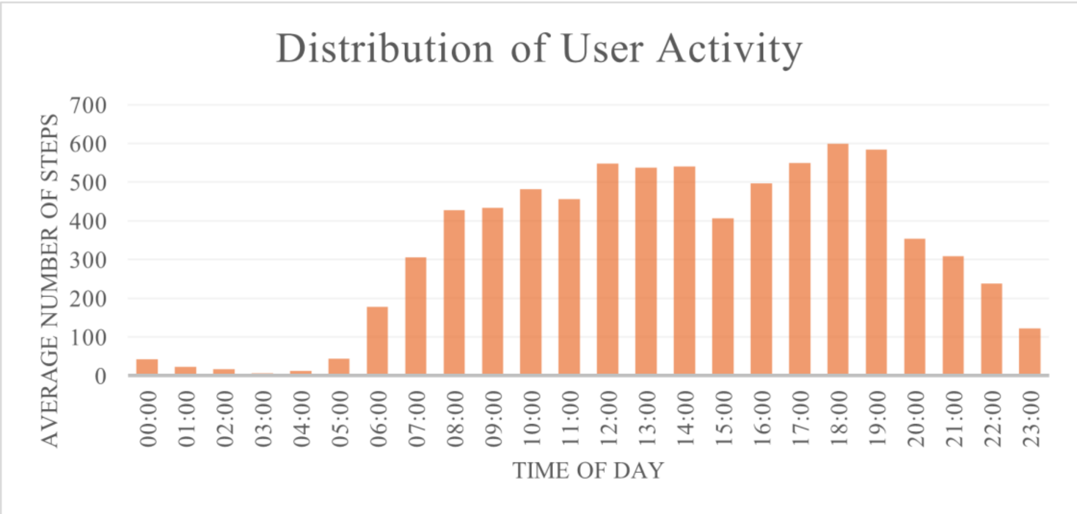
As evident in the above picture the date formats in all tables are generally different due to the outputs of different devices. This is corrected in the processing step to ensure data consistency. After ensuring no further issues remain with the data the cleaning step is complete.

4 Analyze

4.1 Overview

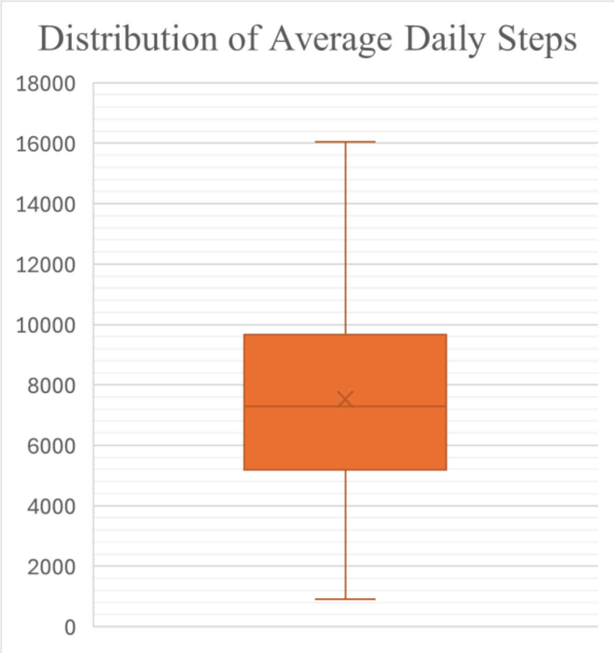
At the start of the analysis, it is important to get an overview of how often users use their smart devices. To this end the daily_activity table is checked for the number of days each unique user has logged. The average usage days of the sample are 28.48 out of 31. Since there are some outliers which could affect the average considerably, the median is also calculated, and the result is 31 out of 31 days.

Next the exact time of usage is analyzed to find trends or patterns. To do this a pivot table with the average number of steps logged each hour was created and evaluated through a chart. As a result, it was identified that activity starts picking up around 6:00 and falls quickly after 19:00. The times between 12:00 and 14:00 as well as between 17:00 and 19:00 were identified as the most active times.



4.2 Activity

After getting this broad overview of the usage of smart devices the next thing of interest is the activity of individual users. The data in daily_steps reveals a broad spectrum of activity among the users. To illustrate this a boxplot of the average daily steps of the users is created.



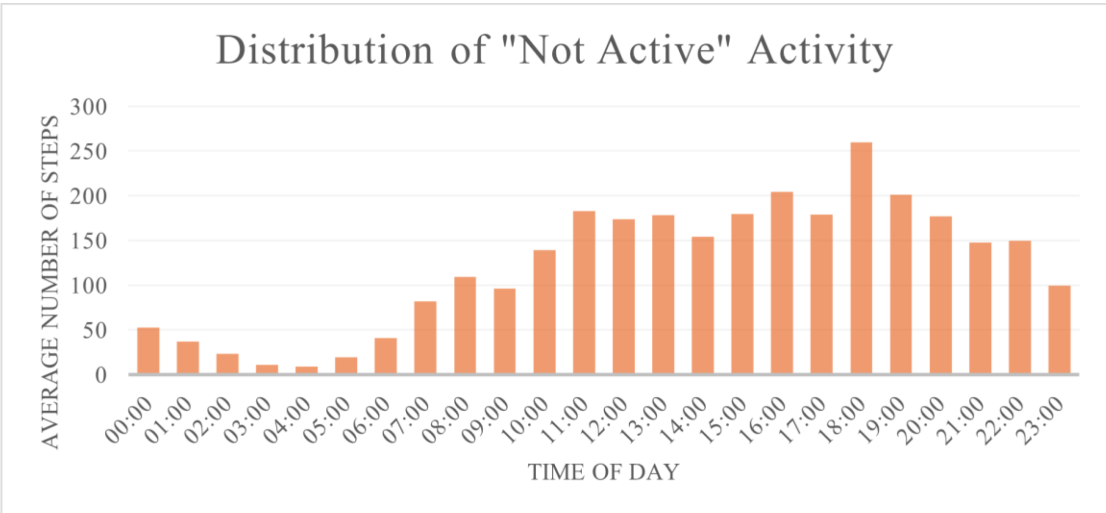
In the resulting visualization the distribution of average daily steps among the users can be seen. The distribution has a slight right skew as can be seen by the position of the median line and the length of the upper whisker compared to the lower whisker. Furthermore, the average (x) is slightly higher than the median (line through the box) which could indicate outliers on the upper end of the distribution.

Overall, it appears most users take a moderate number of steps every day with the middle quartiles ranging from just over 5000 to just under 10000 steps.

To now dig deeper and uncover potential relationships between activity, activity time and sleep the users are categorized according to their average daily steps with the following criteria:

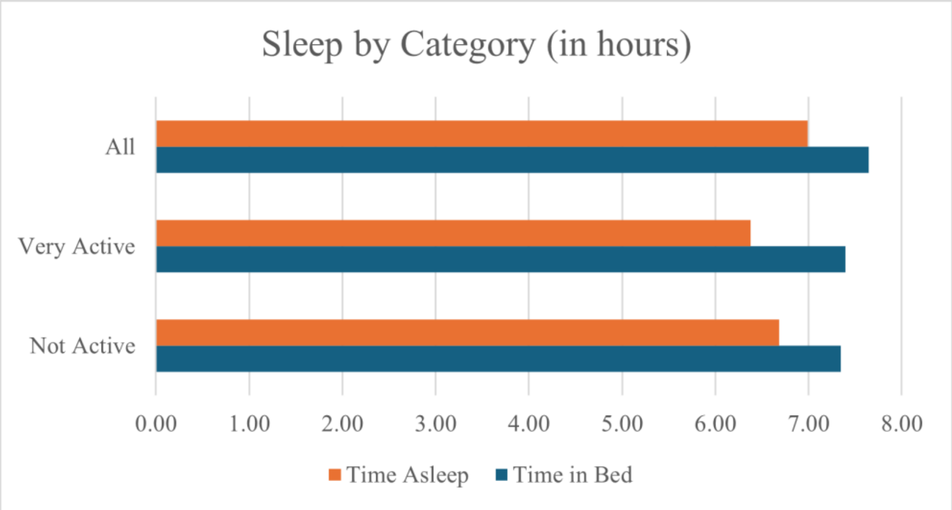
Category	Number of Users	Percentage
Very Active (> 10 000 steps)	7	21.21%
Active (5 000 – 10 000 steps)	18	54.55%
Not Active (< 5 000 steps)	8	24.24%

As expected, most users fall into the "Active" category of between 5000 and 10000 steps. Now the daily activity of the "Not Active" users is analyzed to determine whether they are active at different times compared to the other users. Once again, the pivot table for hourly steps was consulted and using slicers filtered for those users in the "Not Active" category. The resulting distribution shows no clear deviation to that from all users and indicates that there is no correlation between level of activity and time of activity.



4.3 Sleep

Lastly the relationship between activity and sleep is explored. The table sleepDay provides information about the time in bed and time slept for each user. The data is logged daily. By comparing the sleep data between those with high activity and those with low activity, a relationship between activity and sleep may be uncovered. To do this the sleep data is filtered by category. To investigate a potential relationship special interest is placed in the comparison of the "Very Active" and "Not Active" categories. In the following visualization they are compared to each other and the values for all users in the dataset.



Overall users appear to reach their recommended 7 hours of sleep. However, when looking at the "Very Active" and "Not Active" categories it becomes apparent that neither group reach the recommended amount of sleep. "Very Active" users also have the largest difference between the amount of time they spend in bed and the amount of time they sleep. Why these discrepancies exist cannot be discerned from the available data.

4.4 Conclusion and Summary of Analysis

After analyzing the available data using Excel and its integrated tools of PowerQuery and Pivot Tables, it appears that most smart device users use them consistently. They are most active between the times of 12:00 and 19:00. Most users are moderately active (54%) but fall short of the recommended 10 000 steps per day. The data indicates that there is a roughly equal number of very active and not active users. Activity level does not appear to have an impact on the time of activity or vice versa. While users generally reach the recommended 7 hours of sleep, very active and not active users do not, with very active users showing the largest shortcoming.

5 Insights & Recommendations

The analysis of the FitBit data revealed the following major insights:

1. Users utilize their devices consistently
2. The majority of users are moderately active
3. 75% of users do not reach the recommended 10 000 steps per day
4. 50% of users do not reach the recommended hours of sleep
5. Very Active and Not Active users are most likely to not sleep enough
6. No correlations between level of activity and time of activity could be found

Using this insight the following recommendations for Bellabeat's marketing strategy are made:

1. Bind consistent users by potentially offering premium services and loyalty rewards
2. Encourage users to reach the daily target steps by:
    - Incentivizing activity using challenges, push notifications, or recommending small activities to increase daily steps.
    - Specifically targeting moderately active and less active users.
3. Addressing the lack of sleep among very active and not active users. This can be done through educating on the benefits of sleep and how to achieve it or by recommending products, services and other ways to improve sleep time and quality.
4. Apply different strategies for Very Active and Not Active users as each group makes up a significant amount of the user base.
    - Very Active users: These users can be offered advanced challenges, products and insights to further improve their activity. They should also be particularly targeted for sleep related products and services.
    - Not Active users: These users should be encouraged to start getting active. Gentle reminders to move, offering ideas for small activities or providing pointers to entry-level fitness content, products and services could prove effective.
5. Plan notifications around the most active times of users.