Bellabeat Case Study

Introduction:

Welcome to the Bellabeat data analysis case study!.

Bellabeat, a high-tech manufacturer of health-focused products for women, and meet different characters and team members.Bellabeat primarily focused on creating health and wellness products for women, such as the Bellabeat Leaf, a smart jewelry piece that tracks various aspects of women's health, including activity, sleep, stress, and menstrual cycles.

At that time, Bellabeat did not have a traditional smartwatch in its product lineup. However, companies frequently update their product offerings, so it's possible that Bellabeat may have introduced new products, including smartwatches, since then.

The main objective of this case study is to analyze smart devices fitness data and determine how it could help unlock new growth opportunities for Bellabeat. We will focus on one of Bellabeat's products: **Bellabeat app**.

https://bellabeat.com/

Data Analysis Method

In this case study , followed these six step for data analysis:

1 .Ask - Business task
2. Prepare - A description of all data sources used.
3. Process -Documentation of any cleaning or manipulation of data
4. Analyze -A summary of your analysis
5. Share -Supporting visualizations and key findings
6. Act - Your top high-level content recommendations based on your analysis

..

ASK:

Business Task

Identify some trends in how consumers use the Bellabeat devices and how these trends can help improve new opportunities growth for Bellabeat as well as marketing strategy.

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?

Stakeholders

Urška Sršen – Bellabeat's cofounder and Chief Creative Officer
Sando Mur – Mathematician and Bellabeat's cofounder
Bellabeat's marketing analytics team – a team of data analytics..

Prepare:

Dataset Used

In this project, I will be using datasets from FitBit Fitness Tracker Data, which has been published by Möbius in Kaggle under the CC0: Public Domain Creative Common License.

Dataset Summary

I downloaded zip files provided, then extracted to csv files. There are 18 datasets, but I only used 6 datasets for this study. Each document represents different quantitative data tracked by Fitbit.

Dataset Limitations and Integrity

I used excel to take a glimpse of the data. The data collected was only during *2016-04-12* and end *2016-05-12* , so it was quite outdated as fitness trackers matured a lot since then. No demographic information (gender, location, age) collected makes the data bias even higher...

Bellabeat Case Study

Process

Tool used: Sql

To begin processing the data, I used SQL in BigQuery as one of the data analytics tools, to import the dataset, do the process of cleaning and organizing. The cleaning process included adjusting data type formats, removing duplicates and null data. I extracted the clean data to new csv and stored it. I documented the whole process of cleaning.

Using Spreadsheet to Convert Data Type

However, using BigQuery to convert the data type is quite challenging, as you need to convert it using *CAST()* function everytime in each query statement. As I am still learning and finding ways up until now, I decided to use Spreadsheet to change the data type of all date column into "Date" format...

Comparing Datasets

There are 3 datasets that seem to have some similarities: dailycalories, dailyintensities, dailysteps with dailyactivity. I checked if the id and date of those 3 tables are the same with the one in daily_activity.

It showed that data in calories, intensities, and steps have the same number of rows and described the exact data. We could conclude that the daily data of calories, intensities, and steps are the same with daily_activity, so we only focus on daily_activity data.

Checking Start-End Date and Id

I checked the start and end date in each dataset, as well as the length of id.

It showed that all datasets have the same start and end date: start *2016-04-12* and end *2016-05-12*. In term of id's length, all datasets also showed the same length: *10 characters*.

Cleaning Data

Find Duplicates

Duplicates might decrease the quality of data, hereby I had to find and remove them.

According to the result of finding duplicates, it showed that there are 3 duplicate rows in sleep_day dataset. We need to create a new sleep_day table, and remove the duplicates in the new table. In this case, I named the new table: *sleep_day_new*.

Find Unsuitable Data

During the checking and cleaning process, I found that there were some zero data in TotalSteps column inside the daily_activity dataset. Therefore, I decided to check and remove those zero value. I created new table and named it daily_activity_new, so that the previous dataset still remained.

Analyze

After cleaning the data, 3-clean dataset will be used to do the analysis process. In this process, I organized and formatted the data, performed some calculations, and identified trends as well as relationships between each variable.

..

Importing Datasets

After converting some date-related data into "Date" format, I opened [Bigquery Console](), then select "Create Project". Typed down the name of the project you are going to explore, in this case I used *data-analytics-00*. I created a new dataset for Bellabeat and named it *bellabeat*. Inside bellabeat dataset, I imported the .csv datasets I previously downloaded from [FitBit Fitness Tracker Data]() (Note: Don't forget to unzip the file).
After that, I started my work by finding the total number of users' id.

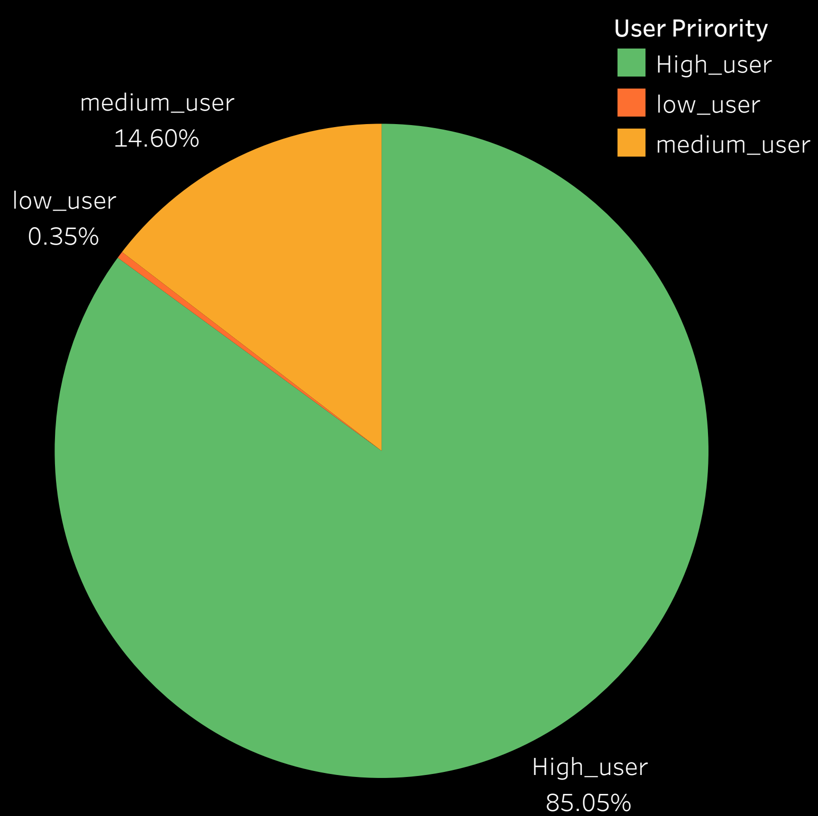
..

<--- page 1

page 3 --->

Bellabeat Case Study

User Type Distribution by Tracker-Wear

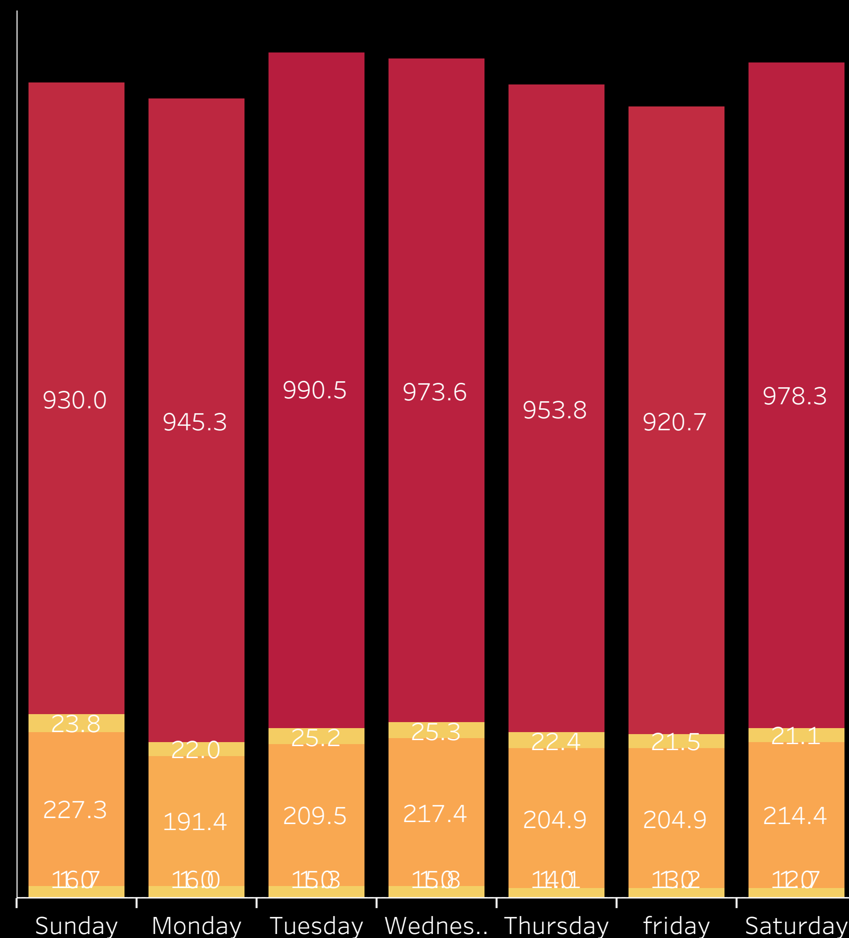


Firstly, I wanted to breakdown the users by how much they wore their FitBit Fitness Tracker. I created three user types:

High_user → wear their tracker for 21 – 31 days
Medium_user → wear their tracker for 11 – 20 days
Low_user → wear their tracker for 1 -10 days

Key Findings: More than 85% of the users used Fitbit Fitness Tracker frequently – between 21 to 31 days (high_user).

Average activity minutes by day of week



In this daily_activity dataset, I wanted to analyse some parameters. The first one was the average active minutes by day of week.

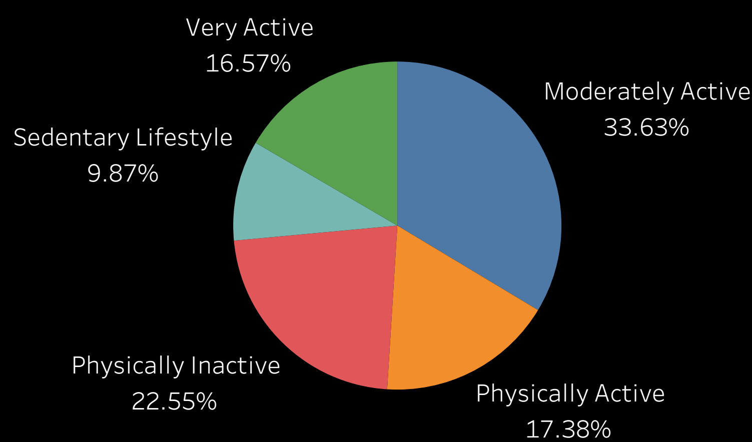
Key findings:
Total average active minutes shows slightly difference throughout the week
Sedentary Minutes are the highest type of active minutes

<--- page 2

Page 4 --->

Bellabeat Case Study

User type distribution by total steps



Tyron (2003) pointed out in his journal that having steps everyday might be a perfect parameter to acquire data about our physical activity. Therefore, we will use total steps data as part of our analysis.

Tudor-Locke and Bassett (2004) proposed a steps-per-day classification, such as:

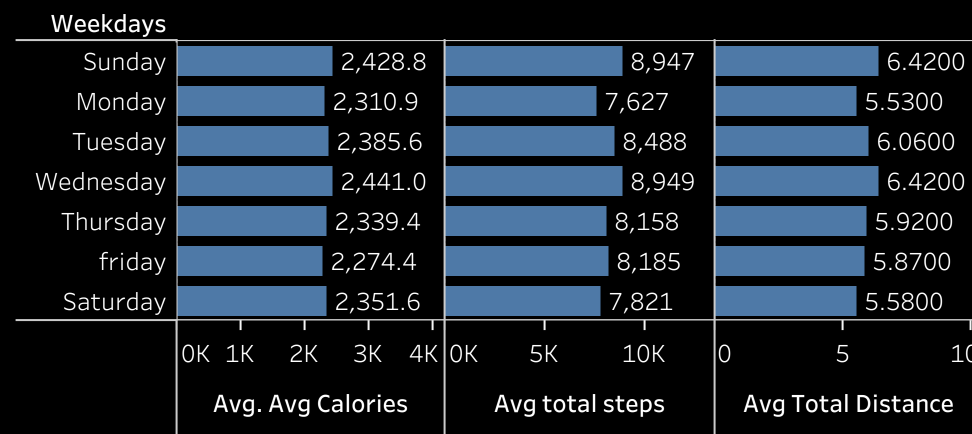
Sedentary Lifestyle → total steps per day <5000
Physically Inactive → total steps per day 5000 - 7499
Moderately Active → total steps per day 7500 - 9999
Physically Active → total steps per day 10000 – 12499
Vey Active → total steps per day >= 12500

I made the user type distribution based on the above categories.

Key Finding:

More than a third of the users are considered moderately active..

Average Steps, distances, calories by day of week



Next, I wanted to know the average of steps, distances, and calories spent per week.

Key Finding:

The highest average step and distance per day was on **Sunday** and **Wednesday** with **almost 9 thousand steps** and **6 km distance**. Average calories shows little difference throughout the week.

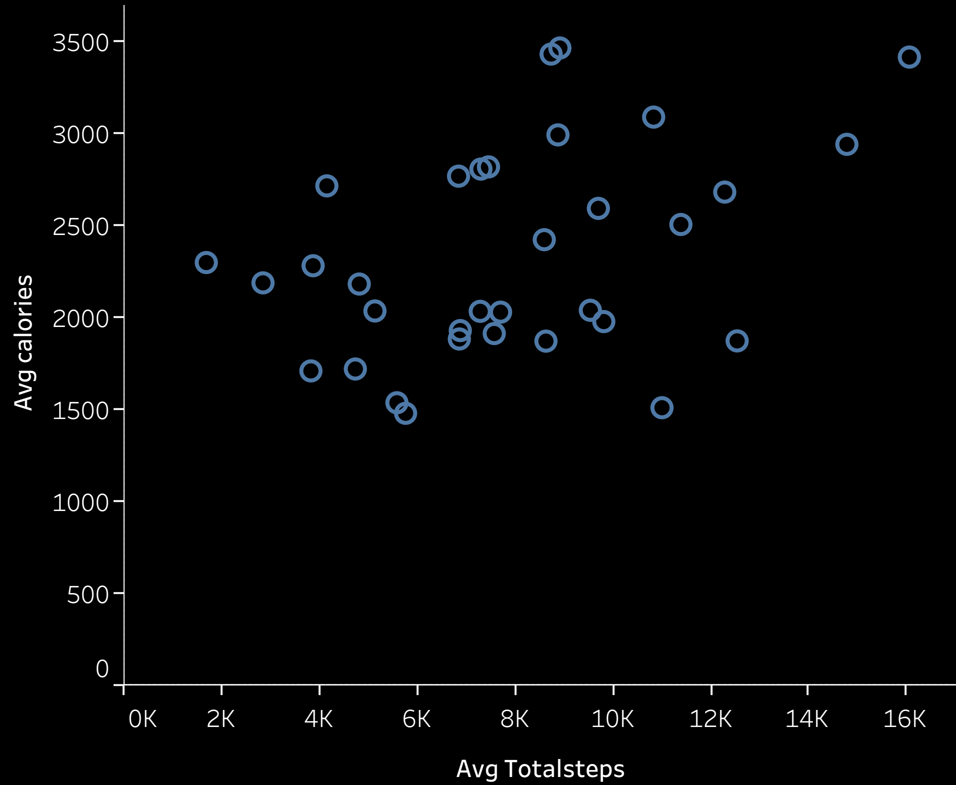
..

<---- Page3

Page 5 ---->

Bellabeat Case Study

Average total steps vs calories



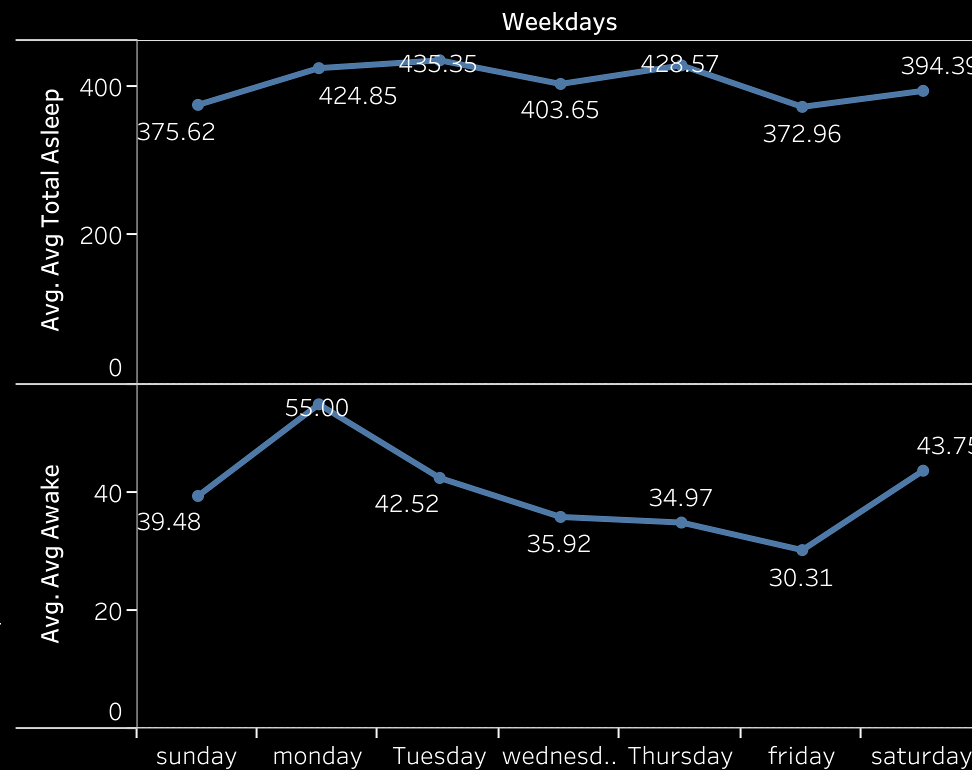
I wanted to look if there was a correlation between total steps and calories.

Key Takeaway:

There is high correlation between average total step and calories.

..

Average sleep time and awake time by day of week



In this sleep_day_new dataset, I wanted to analyze 2 parameters. The first one was the average sleep time and awake per week.

Key Takeaway:

There isn't a whole lot of difference between each day in terms of average time in bed and time awake. The highest average time in bed within a week falls on **Tuesday**.

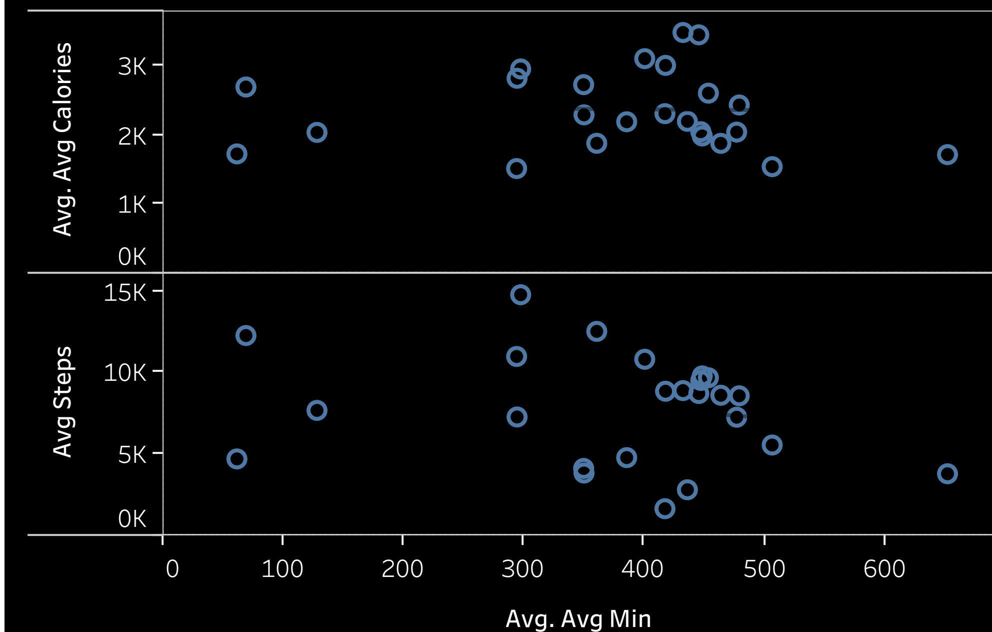
..

<--- Page 4

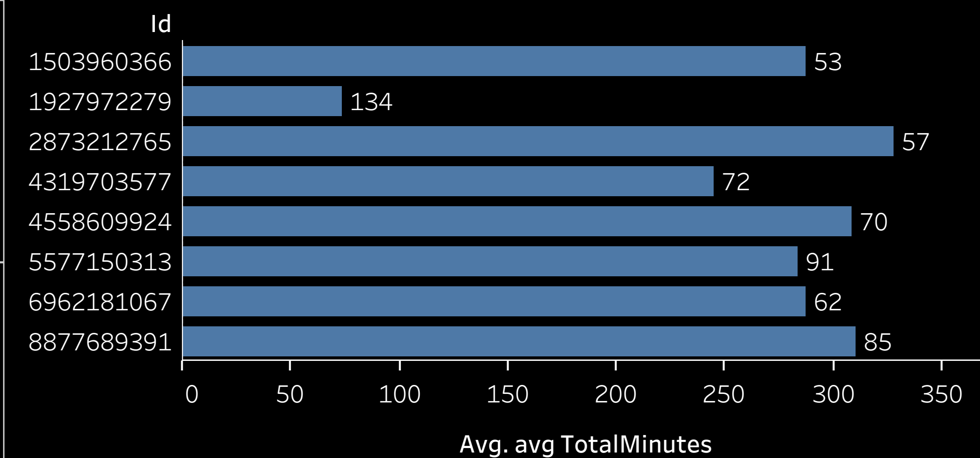
page 6 --->

Bellabeat Case Study

Average total minutes asleep, steps, calories



Average weight vs non-sedentary minutes



Next, I combined daily_activity_new and sleep_day_new to compare the total minutes asleep, steps, and calories. I used *inner join* formula because not all the sleep data are in the daily activity table, so I only needed to calculate the overlayed data.

In this weight_log_new dataset, I wanted to see the trend of individuals' weight by their activity minutes.

Key Takeaway:

There is **no correlation** between average total steps and average amount of minutes users sleep. **No correlation** is found between average total minutes of sleep and average calories.

Key Takeaway:

User with **low exercise** shows **overweighted** (134 kg). Users with weight of <70 kg had been more active.

..

..

<--- Page 5

Page 7-->

Bellabeat Case Study

5: Share

The share phase is done by presenting this report. Some conclusions I can draw from the data:

1. More than 85% of the users used Fitbit Fitness Tracker frequently – between 21 to 31 days (high_user).

2. Total average active minutes shows slightly difference throughout the week Sedentary Minutes are the highest type of active minutes

3. More than a third of the users are considered moderately active

4. The highest average step and distance per day was on Sunday and Wednesday with almost 9 thousand steps and 6 km distance. Average calories shows little difference throughout the week.

5. There is high correlation between average total step and calories.

6. There isn't a whole lot of difference between each day in terms of average time in bed and time awake. The highest average time in bed within a week falls on Tuesday.

7.There is no correlation between average total steps and average amount of minutes users sleep. No correlation is found between average total minutes of sleep and average calories.

8. User with low exercise shows overweighted (134 kg). Users with weight of <70 kg had been more active.

6: Act

The act phase would be done by the production and marketing team of the company. The main takeaway will be the top three recommendations:

1. Marketing team can showcase about **the importance of doing sport activities daily**, by highlighting the strong correlation between total active minutes with healthy weight, so that users might build a self-conscious about using the product regularly.

2. Production team can improve the **notification feature** of the tracker app as reminders for the users to achieve their goal and increase total steps each day.

3. Production team can provide **reward system**, based on total amount of steps reaching daily, weekly, monthly, to boost the energy and motivation of workout.