Bellabeat Data Analytic Report

Bellabeat Smart Device Usage Analysis

Author : Sana Javed

Date : 24-09-2025

Version: 1.0

ASK :

Problem Statement:

As Bellabeat aims to expand its presence in the global smart device market, it needs a deeper understanding of how users interact with other non Bellabeat smart devices. The objective is to analyse trends in smart devices usage to identify behavioural trends, patterns of engagement, and popular features. These insights will be applied to a specific Bellabeat product, with the goal of improving market reach, enhancing brand positioning, and increasing customer engagement.

Stakeholders:

- UrškaSršen and other executive team – to get insights from data to improve product and it's marketing strategies.
- Marketing Analytics Team - Responsible for conducting the analysis and delivering actionable insights.
- Bellabeat Marketing Team - Will use the insights to shape targeted marketing campaigns.
- Product Development Team - May use findings to refine product features or future product development

Key questions:

1. What daily activity, sleep, calories and heart rate trends can be identified from smart device usage?
2. What percentage of user track their daily activities?
3. Does the insight from available dataset is sufficient to identify trends?
4. How can Bellabeat apply these insight to improve product and it's marketing strategies?

5. How could Bellabeat use these insights in its next marketing campaign?

PREPARE:

For this analysis Kaggle data set *FitBit Fitness Tracker Data* (CC0: Public Domain, dataset made available through Mobius) is being used. It contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits.

Data was downloaded locally and imported to Excel for initial quick review and cleaning.

Evaluating data quality using ROCCC framework:

Reliable: Fitbit data is from real users.

Original: Third party data, not from Bellabeat users

Comprehensive: 2 months data from only 30 users

Current: Data is from 2016

Cited: It's public Kaggle dataset

Because the dataset includes only 30 users and lacks demographic details, there is a risk of sampling bias. Still useful for identifying trends, can be improved by combining with Bellabeat own data.

Data is in long format, each user is identified with unique id and each row represent data of each user for day and time.

Files used : Daily Activty and SleepDay

PROCESS

Tools used: *Excel* (data preparation), *BigQuery* (SQL Analysis), *Tableau* (visualisation)

Data Preparation in Excel:

Daily Activity Sheet

- Adjust column widths so that all data is fully visible.
- Apply bold formatting to the column headers for clarity.
- Compare the column names in both files (the daily activity files) to ensure they match exactly before merging.

- Merge the two files (with the same column names) from both folders to create a single dataset containing all available dates.
- Some of the values in the Activity Date column are stored as text, convert them to a proper Date format (yyyy/mm/dd) that will be recognized by SQL.
- Use conditional formatting to highlight any blank cells and review them for completeness.
- Remove duplicates based on the combination of ID + Date.
- Sum all time-related columns and verify that the maximum and minimum values fall within a valid range.

SleepDay Sheet

- Adjust column widths so that all data is clearly visible.
- Apply bold formatting to the column headers for better readability.
- Convert the Date/Time column to a proper Date format that SQL will recognize:
  - Use Text to Columns with a fixed delimiter,
  - then apply the appropriate Date format (e.g., yyyy/mm/dd).
- Use conditional formatting to highlight and check for any blank cells.
- Remove duplicates based on the combination of ID + Date.
- Create derived columns:
  - AwakeInBed
  - TotalHoursAsleep (rounded to two decimal places)
- Validate values:
  - Ensure that all AwakeInBed values are positive,
  - and confirm that the resulting values appear reasonable.

SQL Analysis in BigQuery:

- Create a new dataset in BigQuery.
- Upload the files `daily_activity_clean` and `sleepday_clean` into the dataset as separate tables.
- Review the table schemas to ensure that all columns, data types, and formats have been imported correctly.

Verify Dataset Integrity

Ensure that both the daily_activity_clean and sleepday_clean tables were complete and free of duplicate records.

```
-- total rows and distinct ids
SELECT
 COUNT(*) AS total_rows,
```

```
 COUNT(distinct Id) AS distinct_id
FROM `sana-project-469323.bellabeat.daily_activity_clean`;
```

Result: 1,397 rows representing 35 distinct users.

```
--duplicate rows
SELECT
 Id, Date,COUNT(*) AS records
FROM `sana-project-469323.bellabeat.daily_activity_clean`
GROUP BY Id, Date
Having COUNT (*) > 1;
```

Result: No duplicates found.

Similar checks on sleepday clean confirmed 413 rows, 24 distinct users, and no duplicates.

Confirm Date Ranges

Checked the date span to ensure data covered the expected analysis window:

```
--checking date range
SELECT
 MIN (Date) AS start_date,
 MAX (Date) AS end_date
FROM `sana-project-469323.bellabeat.daily_activity_clean`;
```

Result: Activity data ranged from 2016-03-12 to 2016-05-12;
sleep data ranged from 2016-04-12 to 2016-05-12.

Combine Activity and Sleep Data

Created a consolidated table for user-level daily analysis:

```
--combining activity and sleep data
CREATE TABLE `sana-project-469323.bellabeat.user_dailys` AS
SELECT
 a.Id,
 a.Date,
 a.TotalSteps,
 a.Calories,
 s.TotalHoursAsleep
FROM `sana-project-469323.bellabeat.daily_activity_clean` a
 LEFT JOIN
 `sana-project-469323.bellabeat.sleepday_clean` s
 ON a.Id =s.Id AND a.Date = s.SleepDay;
```

Generate Summary Metrics

Calculated key descriptive statistics to support later analysis:

```sql
--Average steps per user
SELECT
 Id,
 ROUND(AVG(TotalSteps),2) AS AverageSteps
FROM `sana-project-469323.bellabeat.daily_activity_clean`
GROUP BY Id;

-- Daily trends across users
SELECT
 Date,
 ROUND(AVG (TotalSteps),2) AS AverageSteps,
 ROUND(AVG (Calories),2) AS AverageCalories
FROM `sana-project-469323.bellabeat.daily_activity_clean`
GROUP BY Date
ORDER BY Date;

--Weekly summary
SELECT
 FORMAT_DATE('%Y-%V', Date) AS WeekYear,
 ROUND(AVG (TotalSteps),2) AS AverageWeeklySteps,
 ROUND(AVG (Calories),2) AS AverageWeelyCalories,
 ROUND (AVG (TotalHoursAsleep),2) AS AverageWeeklySleep
FROM `sana-project-469323.bellabeat.user_dailys`
GROUP BY WeekYear
ORDER BY WeekYear;
```

ANALYZE AND SHARE:

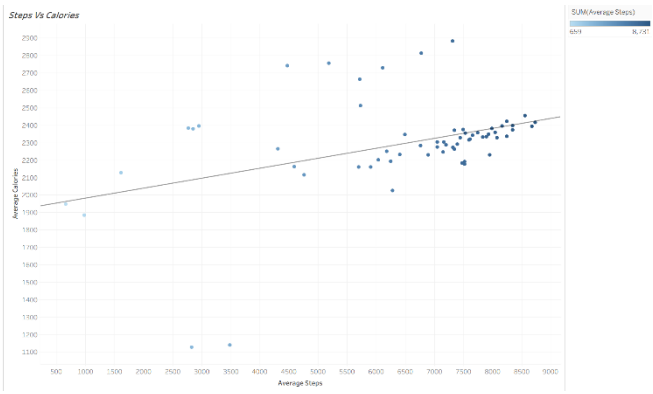
Key findings from Dashboard

Steps and Calories Over Time:

- Calorie burn increases with higher step counts (positive correlation)
- From early April to early May, users maintained high and stable step counts, leading to consistent calorie expenditure.

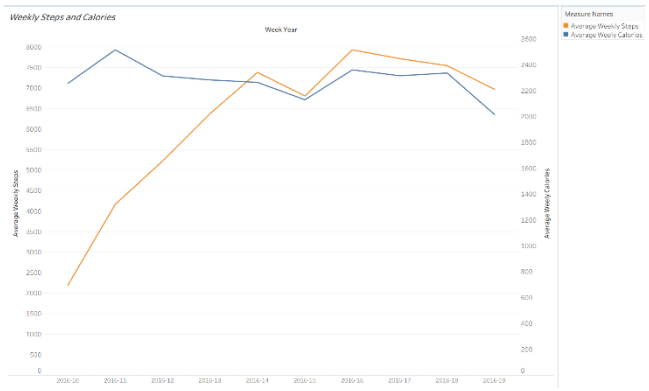
Steps and Calories

- The scatterplot confirms the linear relationship between steps and calories.
- Outliers: Some users logged high calories with relatively fewer steps, which may indicate other physical activities not tracked by steps alone.



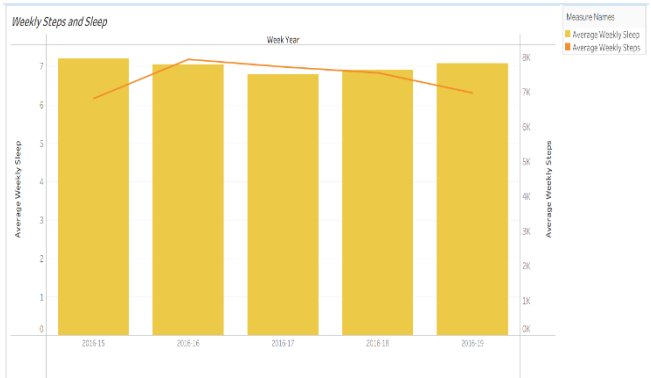
Weekly Steps and Calories

- Weekly averages show a steady increase in steps and calories during mid-weeks, followed by a decline at the end.
- Indicates fluctuations in user engagement, possibly influenced by motivation or external factors (e.g., weekends, weather).



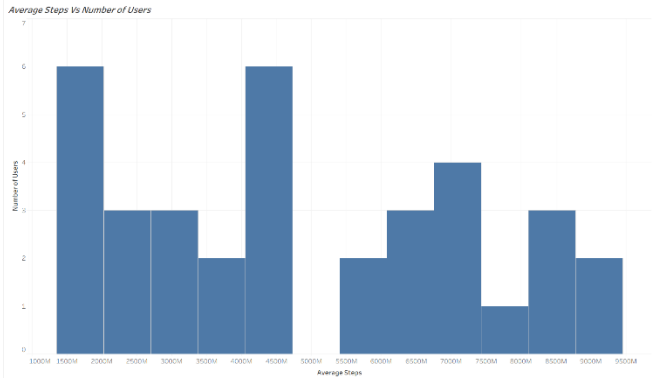
Weekly Steps and Sleep

- No strong correlation between steps and sleep duration.
- Even when activity rises, sleep hours remain stable, suggesting that physical activity doesn't directly affect sleep quality here.



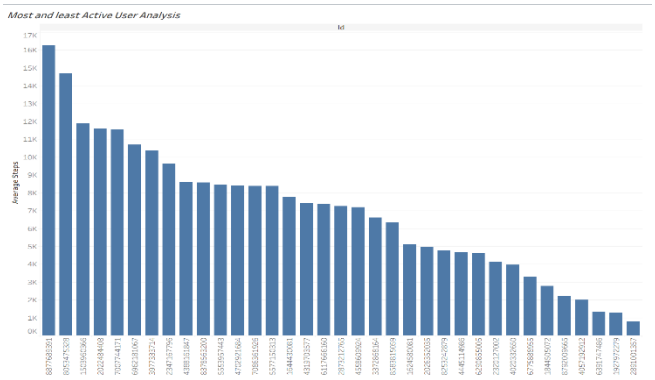
Average Steps Vs Number of Users:

- Most users are in the 3K–8K daily step range.
- Few users consistently hit the 10K+ recommended daily steps, indicating an opportunity to encourage higher activity.



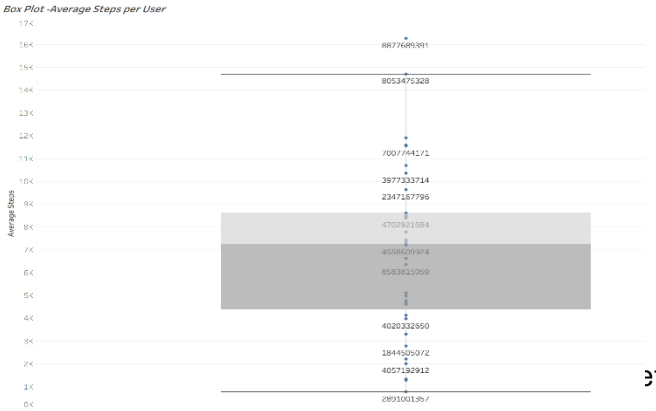
Most and Least Active Users:

- Large variation: some users average 15K+ steps, while others remain under 4K.



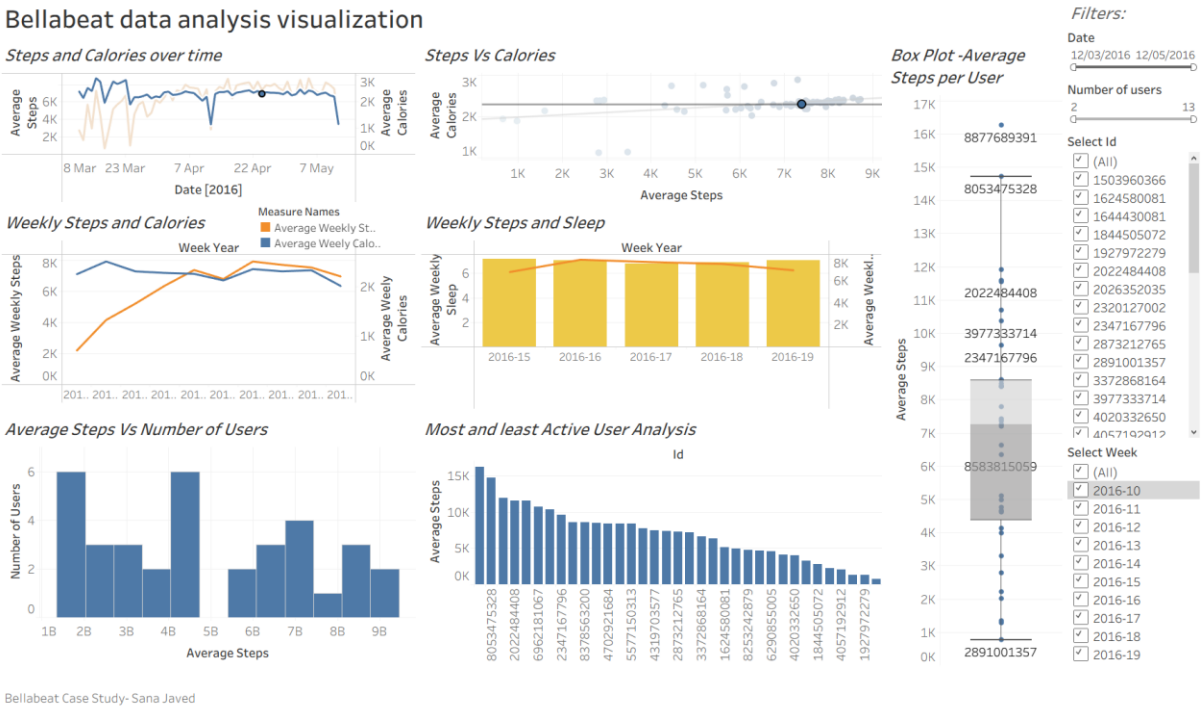
Box Plot- Average Step per User

- Median activity is about 7K steps/day.
- A few highly active users skew the average, masking the fact that many users fall below the 10K benchmark.



Key Insights:

- Positive link[illegible]eting around weight management and fitness tracking.
- Uneven activity levels across users → potential for personalized engagement strategies (push notifications, challenges, rewards).
- Sleep and activity not strongly related → promote sleep tracking features separately (wellness, recovery focus).
- Most users below 10K steps/day → campaigns and rewards to increase their physical activities (step challenges, suggest activities, reminders).



Bellabeat Case Study- Sana Javed

ACT

Objective:

Analyse trends in smart devices usage and to identify behavioural trends, patterns of engagement, and popular features. These insights will be applied to a specific Bellabeat

product, with the goal of improving market reach, enhancing brand positioning, and increasing customer engagement

Key insight to work on:

- Calories burned are strongly tied to daily steps
- Uneven activity levels across users
- Sleep tracking is underutilized compared to activity tracking.
- Only a small number of users achieve 10,000 steps/day.

Business Ideas:

- Step Challenges: Launch challenges to push users closer to 10K daily steps.
- Introduce dashboard in app to show users how activity impacts calorie burn directly
- Personalized Notifications: Send reminders to inactive users and celebrate milestones.
- Promote Bellabeat's sleep-tracking features

APPENDIX:

```
-- total rows and distinct ids
SELECT
 COUNT(*) AS total_rows,
 COUNT(distinct Id) AS distinct_id
FROM `sana-project-469323.bellabeat.daily_activity_clean`;
--total_rows  distinct_id
--1397        35

--duplicate rows
SELECT
 Id, Date,COUNT(*) AS records
FROM `sana-project-469323.bellabeat.daily_activity_clean`
GROUP BY Id, Date
Having COUNT (*) > 1;
--no duplicates

--checking date range
SELECT
 MIN (Date) AS start_date,
 MAX (Date) AS end_date
FROM `sana-project-469323.bellabeat.daily_activity_clean`;
--start_date  end_date
--2016-03-12  2016-05-12

-- total rows and distinct ids Sleep data
SELECT
 COUNT(*) AS total_rows,
 COUNT(distinct Id) AS distinct_id
```

```sql
FROM `sana-project-469323.bellabeat.sleepday_clean`
--total_rows  distinct_id
--413         24


--duplicate rows Sleep Data
SELECT
 Id, SleepDay,COUNT(*) AS records
FROM `sana-project-469323.bellabeat.sleepday_clean`
GROUP BY Id, SleepDay
Having COUNT (*) > 1;
--no duplicates

--checking date range Sleep data
SELECT
 MIN (SleepDay) AS start_date,
 MAX (SleepDay) AS end_date
 FROM `sana-project-469323.bellabeat.sleepday_clean`;
 --start_date end_date
--2016-04-12  2016-05-12


--combining activity and sleep
CREATE TABLE `sana-project-469323.bellabeat.user_dailys` AS
SELECT
 a.Id,
 a.Date,
 a.TotalSteps,
 a.Calories,
 s.TotalHoursAsleep
FROM `sana-project-469323.bellabeat.daily_activity_clean` a
 LEFT JOIN
 `sana-project-469323.bellabeat.sleepday_clean` s
 ON a.Id =s.Id AND a.Date = s.SleepDay;

--Average steps per user
SELECT
 Id,
 ROUND(AVG(TotalSteps),2) AS AverageSteps
FROM `sana-project-469323.bellabeat.daily_activity_clean`
GROUP BY Id;

-- Daily trends across users
SELECT
 Date,
 ROUND(AVG (TotalSteps),2) AS AverageSteps,
 ROUND(AVG (Calories),2) AS AverageCalories
FROM `sana-project-469323.bellabeat.daily_activity_clean`
GROUP BY Date
ORDER BY Date;

--Weekly summary
SELECT
 FORMAT_DATE('%Y-%V', Date) AS WeekYear,
 ROUND(AVG (TotalSteps),2) AS AverageWeeklySteps,
 ROUND(AVG (Calories),2) AS AverageWeelyCalories,
 ROUND (AVG (TotalHoursAsleep),2) AS AverageWeeklySleep
FROM `sana-project-469323.bellabeat.user_dailys`
```

```
GROUP BY WeekYear
ORDER BY WeekYear;
```