Case Studay Report
Chuan (Jasper) Liu
2025-04-17

Case Study 2: How Can a Wellness Technology Company Play It Smart ?

In this case study, we will perform data analysis for a high-tech manufacturing company in order to gain insight into how consumers use non-Bellabeat smart devices. The company provided their 2-year shared bike usage data for both their members and non-member users.

Ask：Questions to be addressed

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?

Prepare

1. Access data: This is public data from the website: https://www.kaggle.com/datasets/arashnic/fitbit, given by the class.

2. The data consists of 18 csv files: the files for recorded the activities time, calories consumption, sleep time and status and etc. from 2016/4 to 2016/5

3. The data is first-party data collected by Fitbit , so there is a low chance of bias, but due to it being the company's own data the credibility is very high.

4. The data is open source and covered by licence : https://creativecommons.org/publicdomain/zero/1.0/

Process

After downloading all files, we found that the date format in all files is not in the right format. I use google sheets to convert all dates from D/M/Y to Y-M-D, so BigQuery can read the date correctly. I also checked the length of each cell in column “Id” use “LEN()” function, and use filter to remove all rows with null values.

Uploaded all files to BigQuery, I checked the schema of all 10 files to make sure all columns in each file are complete and named correctly. Then, checked the type of cell in each column, make sure all data type are correct. And each table has a column called “Id”, we can use this column as our primary key.

Next, we will use SQL code to look into details about our files:

Let’s check the classification of time in our data, we are using the key word : **date, minute, daily, hourly, day, seconds**

```
SELECT
table_name,
column_name
FROM
`my-first-project-452000.fitdata.INFORMATION_SCHEMA.COLUMNS`
WHERE
REGEXP_CONTAINS(LOWER(column_name), "date|minute|daily|hourly|day|seconds");
```

We have the results below:

Row	table_name	column_name
1	minuteCaloriesNarrow	ActivityMinute
2	minute_intensitiesNarrow	ActivityMinute
3	minuteMETsNarrow	ActivityMinute
4	minuteStepNarrow	ActivityMinute
5	minuteSleep	date
6	dailyActivity	ActivityDate
7	dailyActivity	VeryActiveMinutes
8	dailyActivity	FairlyActiveMinutes
9	dailyActivity	LightlyActiveMinutes
10	dailyActivity	SedentaryMinutes

From our results, we have 5 tables recorded by minute, so we will start our analyze based on it. As we can see in the table “minuteSleep”, the date column name is different from others, let change it first:

```
ALTER TABLE my-first-project-452000.fitdata.minuteSleep_cleaned
RENAME COLUMN date to ActivityMinute;
```

Row	table_name	column_name
1	minuteCaloriesNarrow	ActivityMinute
2	minute_intensitiesNarrow	ActivityMinute
3	minuteMETsNarrow	ActivityMinute
4	minuteSleep_cleaned	ActivityMinute
5	minuteStepNarrow	ActivityMinute
6	dailyActivity	ActivityDate
7	dailyActivity	VeryActiveMinutes
8	dailyActivity	FairlyActiveMinutes
9	dailyActivity	LightlyActiveMinutes
10	dailyActivity	SedentaryMinutes

Now that we should look at the columns that are shared among the minute tables:

```
SELECT
column_name,
data_type,
COUNT(table_name) AS table_count
FROM
`my-first-project-452000.fitdata.INFORMATION_SCHEMA.COLUMNS`
WHERE
REGEXP_CONTAINS(LOWER(table_name),"minute")
GROUP BY
column_name,
data_type;
```

Row	column_name	data_type	table_count
1	Id	INT64	5
2	ActivityMinute	TIMESTAMP	5
3	Calories	FLOAT64	1
4	Intensity	INT64	1
5	METs	INT64	1
6	value	INT64	1
7	logId	INT64	1
8	Steps	INT64	1

Say we are considering sleep related products as a possibility, let's take a moment to see if/ how people nap during the day. To do this we are assuming that a nap is any time someone sleeps but goes to sleep and wakes up on the same day

```
SELECT
  Id,
  sleep_start AS sleep_date,
  COUNT(logId) AS number_naps,
  ROUND(SUM(TIMESTAMP_DIFF(end_time, start_time, SECOND)) / 3600.0, 2) AS
total_hours_sleeping
FROM (
  SELECT
    Id,
    logId,
    MIN(DATE(ActivityMinute)) AS sleep_start,
    MAX(DATE(ActivityMinute)) AS sleep_end,
    MIN(ActivityMinute) AS start_time,
    MAX(ActivityMinute) AS end_time
  FROM
    `my-first-project-452000.fitdata.minuteSleep`
  WHERE
    value = 1
  GROUP BY
    Id, LogId
)
WHERE
  sleep_start = sleep_end
GROUP BY
  1, 2
```

Row	Id	sleep_date	number_naps	total_hours_slee...
1	1503960366	2016-03-13	1	7.08
2	1503960366	2016-03-14	1	6.38
3	1503960366	2016-03-15	1	5.42
4	1503960366	2016-03-16	2	6.05
5	1503960366	2016-03-17	1	7.2
6	1503960366	2016-03-18	1	6.82

Join tables

When we check the new table, we found that some of the data from table "`minuteSleep_cleaned`" are not joined in the new table, because their time sitting in "`minuteSleep_cleaned`" is with microsecond accuracy

	Id	ActivityMinute	value	logId
1	1503960366	2016-03-13 03:20:30 UTC	1	11114919637
2	1503960366	2016-03-13 03:52:30 UTC	1	11114919637
3	1503960366	2016-03-13 03:54:30 UTC	1	11114919637
4	1503960366	2016-03-13 04:12:30 UTC	1	11114919637
5	1503960366	2016-03-13 04:44:30 UTC	1	11114919637
6	1503960366	2016-03-13 06:35:30 UTC	1	11114919637
7	1503960366	2016-03-13 07:33:30 UTC	1	11114919637
8	1503960366	2016-03-13 07:38:30 UTC	2	11114919637
9	1503960366	2016-03-13 07:51:30 UTC	1	11114919637

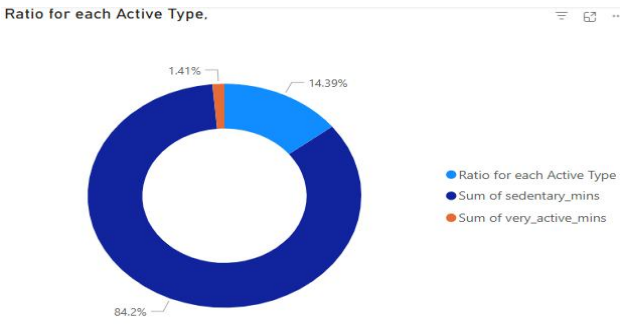
Let's round it:

```
UPDATE `my-first-project-452000.fitdata.minuteSleep`
SET ActivityMinute = TIMESTAMP_TRUNC(ActivityMinute, MINUTE)
WHERE TRUE
```

Analyze

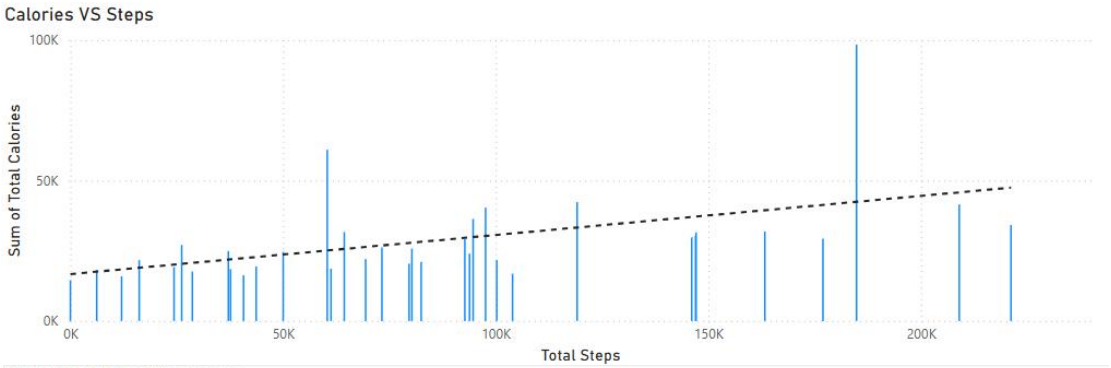
First let's check the time spend on activity per day

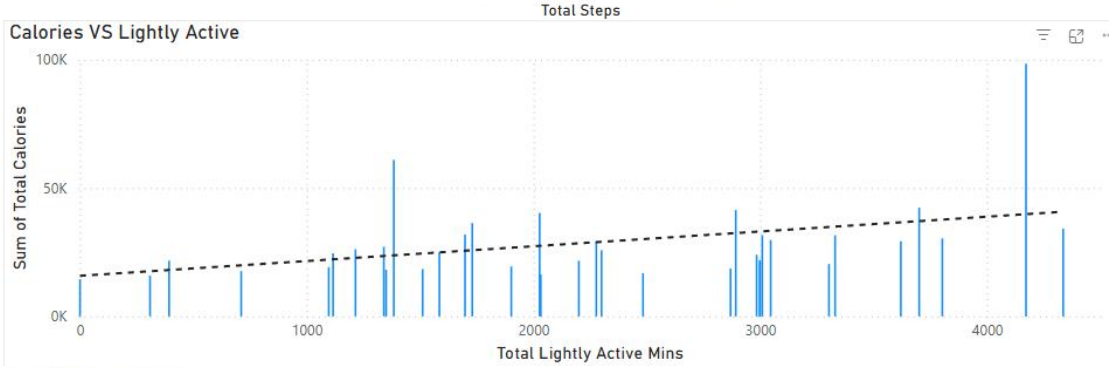
```
Select
 Distinct Id,
 SUM(SedentaryMinutes) as sedentary_mins,
 SUM(LightlyActiveMinutes) as lightly_active_mins,
 SUM(FairlyActiveMinutes) as fairly_active_mins,
 SUM(VeryActiveMinutes) as very_active_mins
From `my-first-project-452000.fitdata.dailyActivity_merged`
Group by Id
```

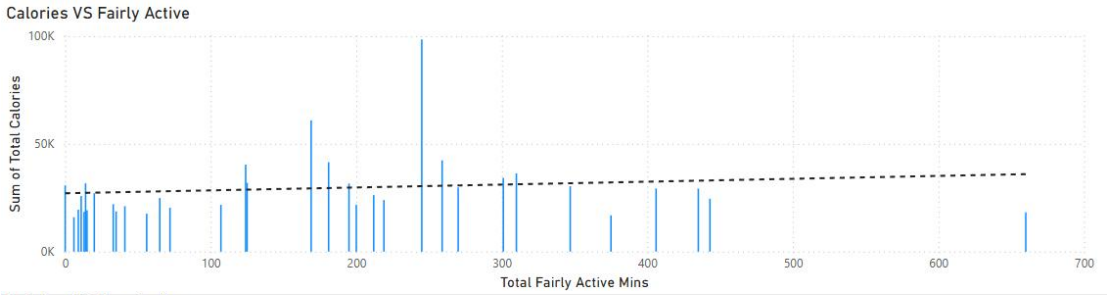


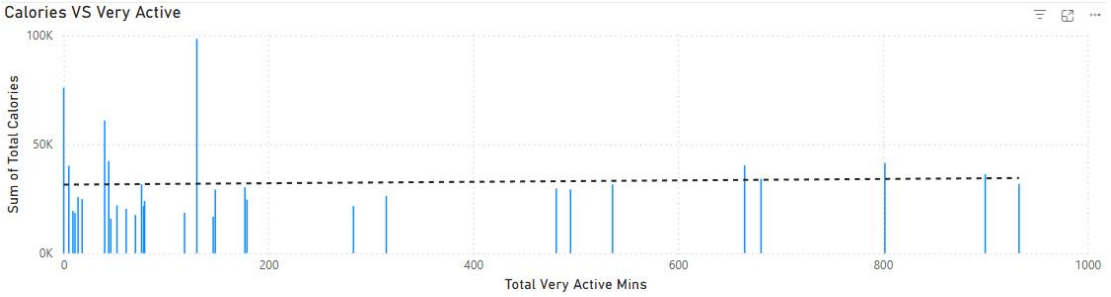
Activities and calories comparison

```
Select
 Id,
 SUM(TotalSteps) as total_steps,
 SUM(VeryActiveMinutes) as total_very_active_mins,
 Sum(FairlyActiveMinutes) as total_fairly_active_mins,
 SUM(LightlyActiveMinutes) as total_lightly_active_mins,
 SUM(Calories) as total_calories
From `my-first-project-452000.fitdata.dailyActivity_merged`
Group By Id
```







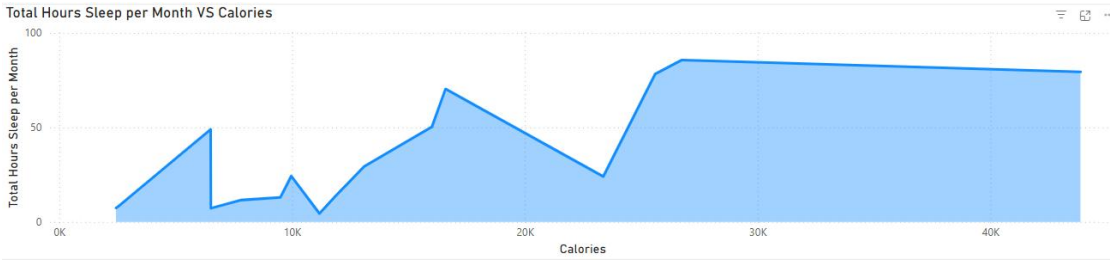


Say we are considering sleep related products as a possibility, let's take a moment to see if/ how people nap during the day. To do this we are assuming that a nap is any time someone sleeps but goes to sleep and wakes up on the same day

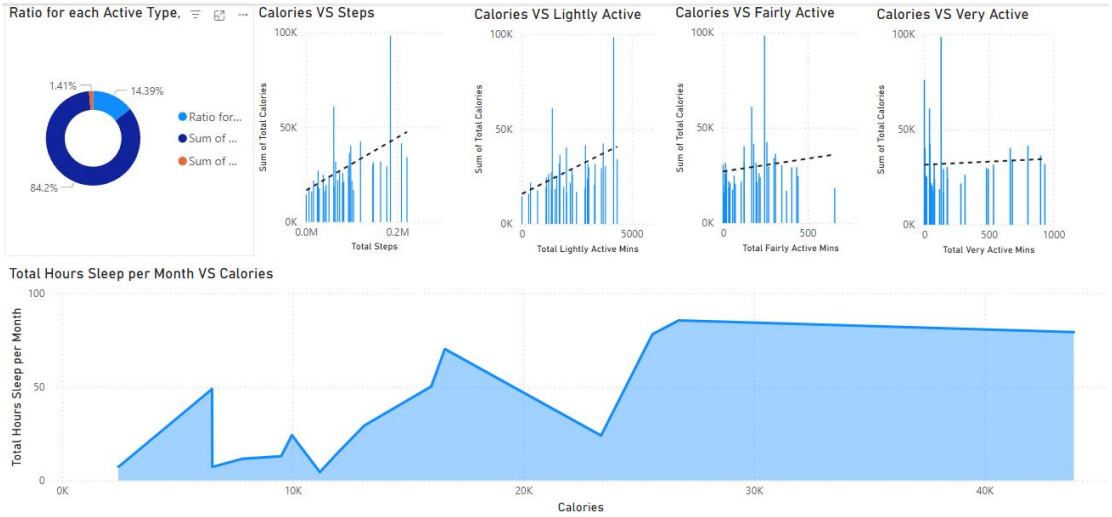
```
CREATE OR REPLACE TABLE `my-first-project-452000.fitdata.sleep_total` AS
SELECT
  Id,
  sleep_start AS sleep_date,
  COUNT(logId) AS number_naps,
  ROUND(SUM(TIMESTAMP_DIFF(end_time, start_time, SECOND)) / 3600.0, 2) AS
total_hours_sleeping
FROM (
  SELECT
    Id,
    logId,
    MIN(DATE(ActivityMinute)) AS sleep_start,
    MAX(DATE(ActivityMinute)) AS sleep_end,
    MIN(ActivityMinute) AS start_time,
    MAX(ActivityMinute) AS end_time
  FROM
    `my-first-project-452000.fitdata.minuteSleep`
  WHERE
    value = 1
  GROUP BY
    Id, LogId
)
WHERE
  sleep_start = sleep_end
GROUP BY
  1, 2
```

Then we join the Calories column in “dailyActivity” table with our sleep summary

```
Select
 A.Id,
 SUM(B.number_naps) as total_naps,
 ROUND(SUM(B.total_hours_sleeping),2) as total_hours_sleep,
 SUM(Calories) as calories
From `my-first-project-452000.fitdata.dailyActivity_merged` A
Inner Join `my-first-project-452000.fitdata.sleep_total` B
ON A.Id = B.Id AND A.ActivityDate = B.sleep_date
Group By A.Id
```



POWER BI DASHBOARD



Share

1. What are some trends in smart device usage?

From the charts above we can see:

- People who do not exercise regularly are still the majority；
- The positive relationship between exercise time and calorie expenditure was more pronounced among people who did not exercise regularly；
- The consumption of calories is positively correlated with sleep duration and quality；

2. How could these trends apply to Bellabeat customers?

According to the above data, we can give different kinds of feedback for different groups of people, so as to increase the satisfaction and enthusiasm of users.

3. How could these trends help influence Bellabeat marketing strategy?

The fact that people who do not exercise frequently are the main group indicates that there is still much room for improvement in people's enthusiasm for exercise and their physical health. It also shows that the current sports-wearing products on the market have not played a very effective role. We can improve our products through data to capture market share.

ACT

In our system, users are classified through their data. For users with insufficient physical activity, we provide more feedback on calorie consumption. For users who exercise frequently, we provide more feedback on the improvement of sleep quality and the benefits it brings. In this way, users can all feel the benefits brought by exercise. This can better enhance users' satisfaction and dependence on our products.