Yafeat Adugna

BELLABEAT CASE STUDY CLEAN VERSION / Google Analytics Course

Company: Bellabeat - a high-tech manufacturer of health-focused products for women.
- **Task** - analyze smart device data to gain insight into how consumers are using their smart devices.
  - Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.
- **Scenario** - junior data analyst working on the marketing analyst team at **Bellabeat**. Co founder Urška Sršen believes analyzing smart device fitness data could help unlock new growth opportunities for the company.
  - Present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat's marketing strategy.
- **Bellabeat marketing analytics team:** responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.

PHASE 1: ASK

- ★ **Problem:** Identifying usage of product (trends)
- ★ **Goal:** Bellabeat, wants to gain insight from their fitness device to understand how customers are using their product.
- ★ **Objective:** To Identify if there is a trend which leads to new growth opportunities.

Stakeholders are
- Urška Sršen: Bellabeat's co-founder and Chief Creative Officer.
- Sando Mur: Mathematician and Bellabeat cofounder
- Bellabeat marketing analytics team

Questions to help guide my analysis?
- ★ What are some trends in smart device usage?
- ★ How could these trends apply to Bellabeat customers?
  - Is there an increase or decrease in trends, what happens if there is a decrease trend or an increase trend?
- ★ How could these trends help influence Bellabeat marketing strategy?
- ★ Geography wise, Are we looking at whole users in a specific location? Does it matter?

BrainStorm Based Business Task and Questions.

Q - What are some trends in smart device usage?

- - **I ask** - How is trend measured, in a specific product? Is there a specific timeline?
- - **Answers I think -** FIND Trends for every product (because I am presenting one of the products) - so I can group the trends by products.

Q - How could these trends apply to Bellabeat customers?

- If there is a **weaker trend** or a decline in usage of a product - The data can be used as an insight to improve that aspect OR specific product.
  - **Solution I think -**
    - Drive discounts
    - Targeted ads to use that product
    - Sending surveys (used to see the personal reasons of why there is a decline) and trying to see if there is a trend.
- If there is a **Strong trend** or an increase in usage of a product - The data can be used as a statistic for new customers to show how existing customers are leading a healthier lifestyle.
  - Recommendation:
    - From Marketing Standpoint: Offer customers (goods discount) in exchange for review of their service, **TO BE USED FOR ADVERTISEMENTS TO ENCOURAGE NEW CUSTOMERS TO JOIN.**
    - OFFER A LOYALTY PROGRAM
      - FOR NON-SUBSCRIBERS - Send Appreciation Emails of members, **OFFER ONE YEAR DISCOUNT TOWRDS THEIR FIRST ANNUAL SUBSCRIPTIONS.** (this drives existing customers to be a life-time customers using a loyalty program)

Q - How could these trends help influence Bellabeat marketing strategy?
- The trends could help the marketing team understand where to focus their marketing budget on and what they're approach should be.

PHASE 2: PREPARE

- **What type of data is right for the question you're answering?**
  - It is a Making predictions problem so, we use existing data and historical data which would be a good fit for this problem.
- **Decide what data to use?**
  - Since we are requested to look into public data set we decided to use (a third party data)
    - **Who is managing the data**?
      - **Where is the data stored?** Stored at an online data repository website.
      - **How did you verify the data's integrity?** It was consensual
      - **How does it help you answer your question?**
        - Question was: gain insight for devices to understand how customers are using the product.
          - Datasets chosen for this Problem are (highlevel):
            - Daily Activity dataset
            - Sleep_day dataset

Sršen tells you that this data set might have some limitations, and encourages you to consider adding another data to help address those limitations.

DATA CREDIBILITY

- ☑ R - Reliable - accurate
  - It is accurate because the data was collected from real customers

- ☑ ~~O - Original~~
    - The data was original since it was collect directly from users
- ☑ ~~C - Comprehensive (everything you need in one place)~~
- ☐ **C** - Current
    - Not current, but in this case we are trying to understand customer behavior.
- ☑ ~~C - Cited~~
    - Creator of the data was cited.

<mark>DESCRIPTION OF DATA USED</mark> - This dataset contains a survey result via Amazon Mechanical Turk between the date 03.12.2016-05.12.2016 from thirty fitbit users with personal fitness tracker. These 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.

PHASE 3: PROCESS

- What tools are you choosing and why?
    - **SQL & EXCEL** because in my options it is easier to clean data with.
- Documenting the cleaning process?
    - Joined two datasets
        - Daily_Activity and sleep_day datasets were joined using **SQL**
    - Daily Activity data set
        - Removed duplicates was found - 0 found
        - Removed Empty space - 0 found
        - Convert to right data type - 0 found
    - Sleep_Day data set
        - 3 duplicates found and removed, Now there is 410 unique values
        - Removed Empty space - 0 found
        - Convert to right data type - 0 found

PHASE 4 & 5: ANALYSIS & SHARE

After Cleaning the data, To get a better understanding, the datasets - daily activity and sleep day were merged using their primary key called "ID".

```sql
 -- After Cleaning the data, To get a better understanding, the datasets - daily
activity and sleep day were merged using their primary key called "ID".


SELECT
  *
FROM
  `my-project-1234-391913.dailyActivity.dailyActivity` AS daily_activity
INNER JOIN
  `my-project-1234-391913.dailyActivity.sleepday` AS sleep
```

```
ON
  dailyactivity.id = sleep.id


 -- To show the distribution of customers and users, for the first analysis I used the
"steps" column to see if user were
    ● Beginners - < 40,000 steps in 2 month
    ● Fairly Active - between 40,000 and 90,000 steps in 2 month
    ● Very Active - > 90,000 step in 2 month
          ○  Data Points ranged between 2,366 - 156,880 steps in 2 month
 --This can help answer the business question by showing the distribution of
smartwatch users.

SELECT
  id,
  CASE
    WHEN total_steps_per_day < 40000 THEN "Beginner"
    WHEN total_steps_per_day BETWEEN 40000 AND 90000 THEN "fairly_active"
  ELSE
  "Very_active"
END
  AS activity_level_per_total_steps
FROM (
  SELECT
    ID,
    total_steps/60 AS total_steps_per_day
  FROM (
    SELECT
      ID,
      SUM(TotalSteps) AS total_steps
    FROM
      `my-project-1234-391913.dailyActivity.ac_sleep`
    GROUP BY
      Id ))
```
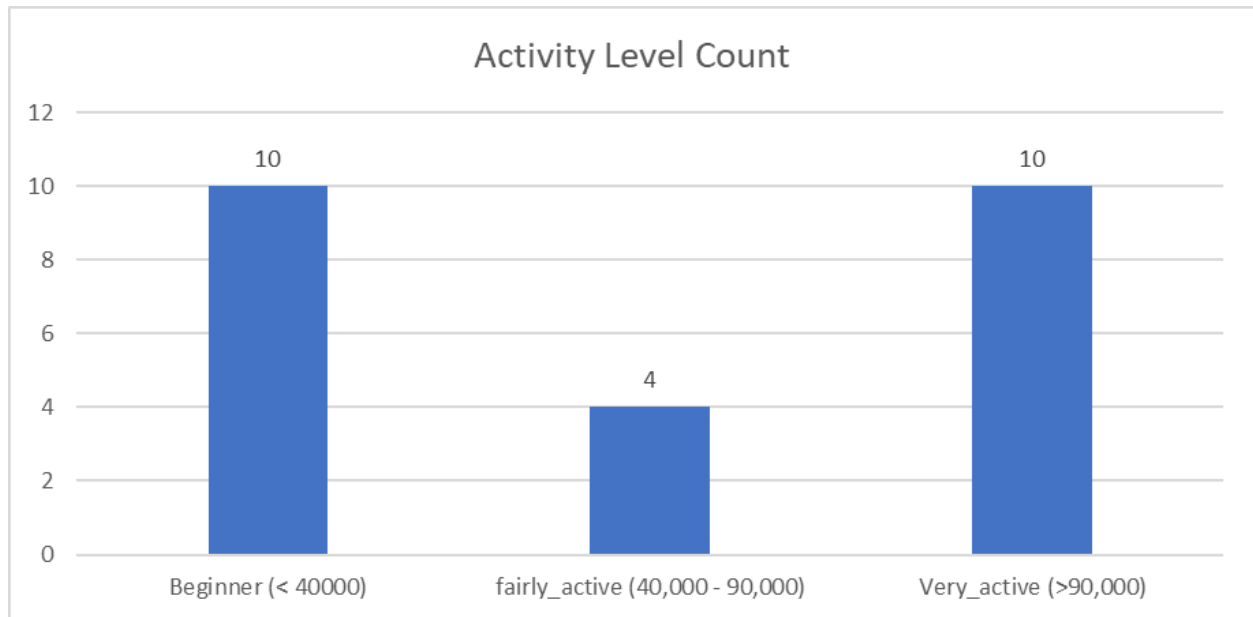
Activity Level Count

Analysis:
- Results show that there is an **even** distribution between Beginners and Ver_active users.

2.

-- Second Analysis was to Find the Average hours of sleep per user to show users sleeping distribution.

```sql
SELECT
  id,
  AVG(sleep_day) AS Average_sleep_per_day_in_hr
FROM (
  SELECT
    id,
    TotalMinutesAsleep/60 AS sleep_day
  FROM
    `my-project-1234-391913.dailyActivity.ac_sleep` )
GROUP BY
  Id
```

Average Hours of Sleep per day
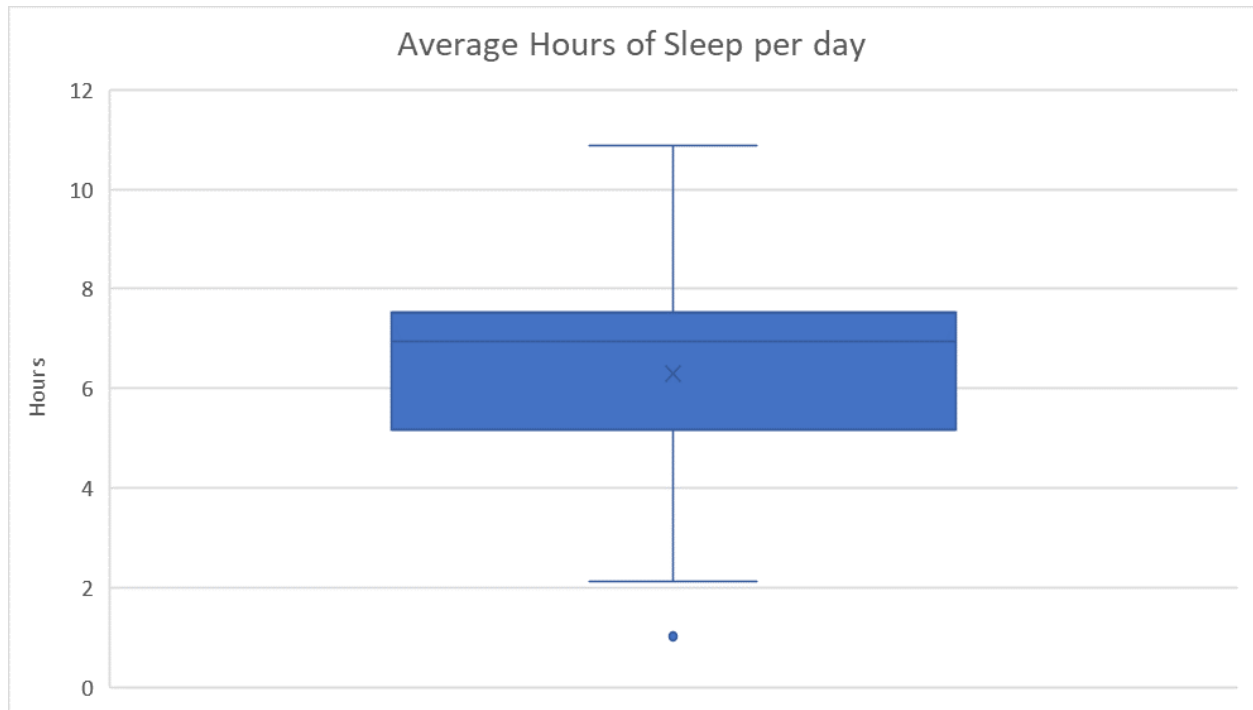
3.
-- third Analysis was to find Correlation between Sleep Hours and Total Calories
Burned.
  ● This shows the marketing team whether or NOT total calories burn is correlated
    with having a consistent and better sleep
      ○ In other words whether or not Carliers burn leads to having better sleep.
--This can help answer the business question by showing how users might use these two
different features on a relationship basis.

```sql
SELECT
  id,
  SUM(Calories) AS total_calories_burn,
  AVG(sleep_day) AS Average_sleep_per_day_in_hr
FROM (
  SELECT
    id,
    Calories,
    TotalMinutesAsleep/60 AS sleep_day
  FROM
```
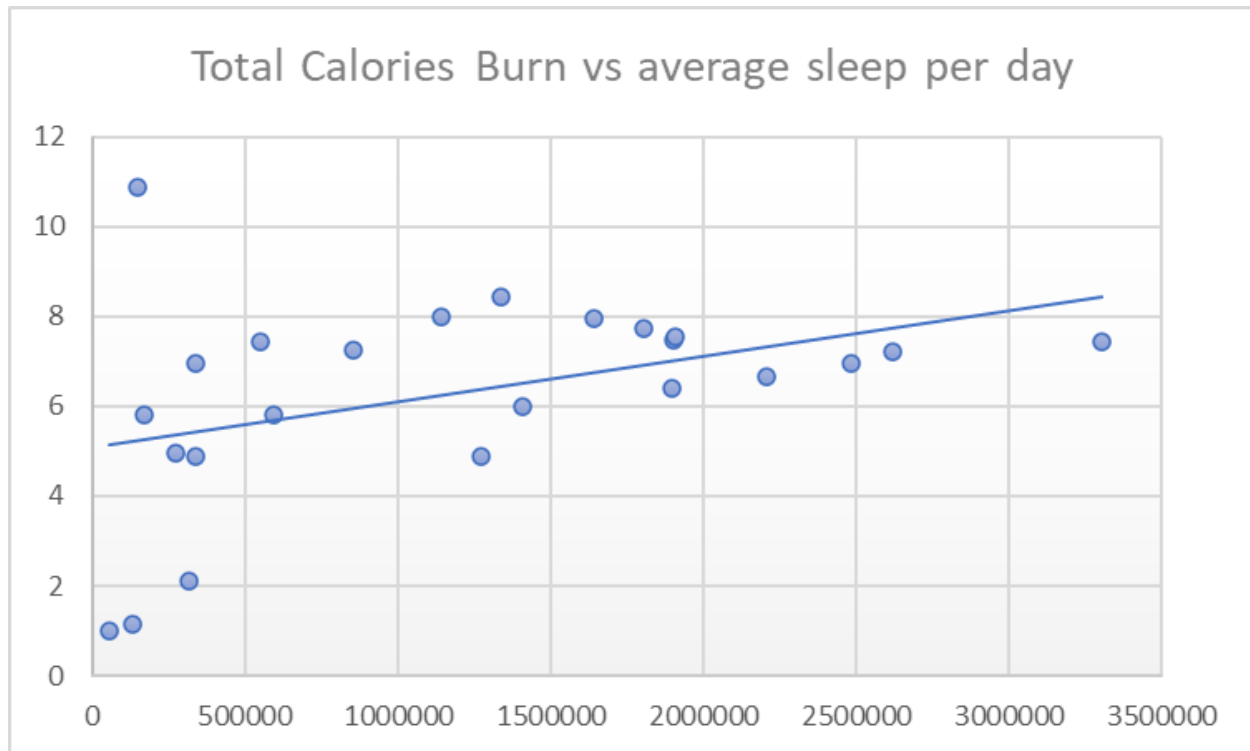
```
    `my-project-1234-391913.dailyActivity.ac_sleep` )
GROUP BY
  Id
```

Total Calories Burn vs average sleep per day

Analysis:
- Results show that Total Calories Burned and Average sleep per night is slightly correlated.
- This shows that these two feature are often used together
  - Also shows that consist calorie burn leads to consistent sleep schedule

4.

```
-- Forth Analysis was to find Correlations between Total Calories Burned and Total
Steps.
--This can help answer the business question by showing how users might use these two
different features on a relationship basis.

SELECT
    id,
    total_calories,
    total_step
  FROM (
```
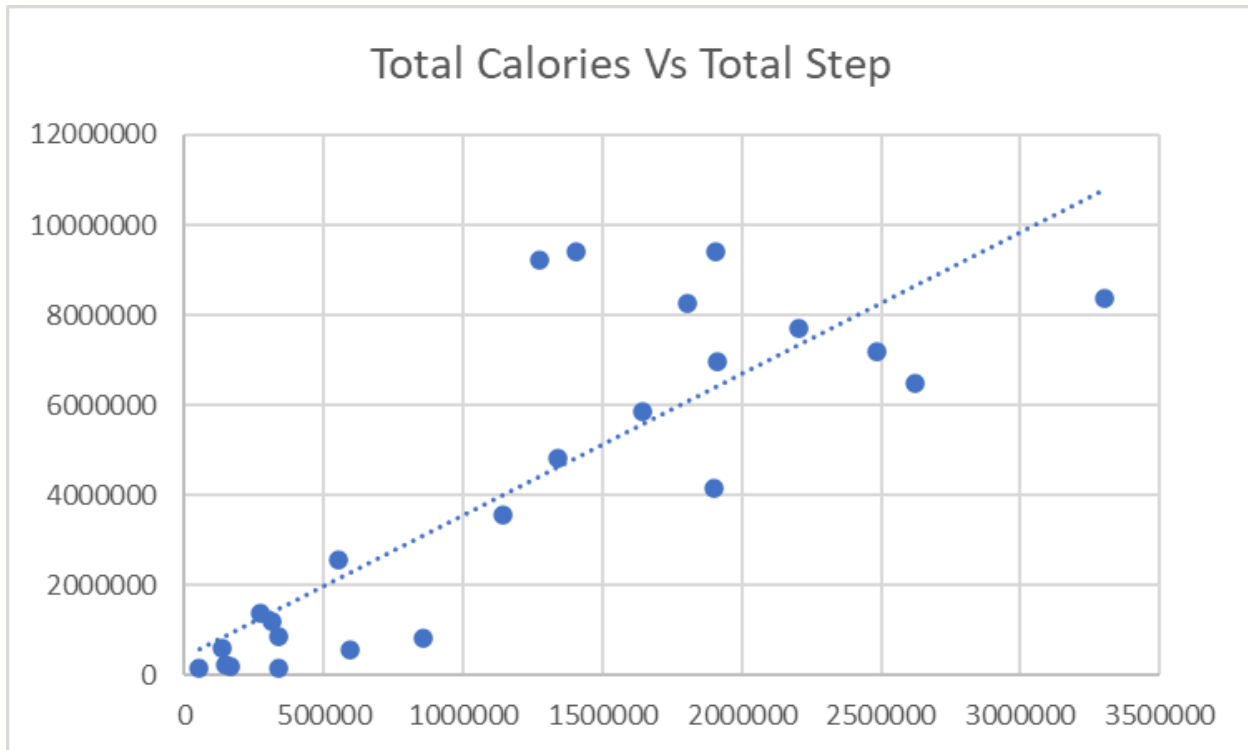
```
SELECT
    id,
    SUM(Calories) AS total_calories,
    SUM(TotalSteps) AS total_step
  FROM
    `my-project-1234-391913.dailyActivity.ac_sleep`
group by Id)
```


Total Calories Vs Total Step

Analysis
- Results showed there is a strong correlation between Total Calories Burned and Total Steps taken.
- This shows that these two feature are often used together because they go hand to hand
  - This shows the strongest correlation and reason people use smartwatch devices.

PHASE 5 & 6: SHARE AND ACT

In conclusion:

- There are few interesting insights found in FitBit Fitness Tracker Data that would be useful for Bellabeat and leveraging these insights could lead to enhanced process optimization and improved outcomes.

1) Main users of FitBit are early birds also known as beginners and people who take their well being very seriously,
    a) Given this insight, Bellabeat's marketing team can strategically focus on promoting their **membership services** to these specific user segments due to their heightened motivation for working out.

2) The average user maintains a consistent sleep duration of 8 hours of sleep across all groups. Additionally, there is a mild correlation between higher daily step counts and more consistent sleep patterns.
    a) Knowing this Bellabeat can utilize their time tracking feature to further study the behavior of their users. Furthermore, this findings shows the potential for consistent exercise leads to improved sleep quality, indicating reduction of stress and an enhanced overall quality of life for all users.

3) Lastly, there is a clear indication of a relationship between intense physical activity, as measured by steps taken, and total calories burned.
    a) Not only this will allow Bellabeats "Leaf" product to be utilized often with Bellabeat app, but also the marketing can team can target very active users with cross functional products Bellabeat offers such as: the spring water bottle and leaf bracelet. These items complement the user experience and also offer the ability to track heart rate and time efficiently.

Thank you for reading and I hope you enjoyed it!