Google Data Analytics Capstone Project (Bellabeat Case Study)

—————————————————————————————

Title— Bellabeat Case Study
How Can a Wellness Technology Company Play it Smart ?
Author— Sumit Baviskar
Date- 10 Dec 2023

Introduction

This is a project for a Google Data Analytics Certification, where we're helping Bellabeat, a company that makes smart health products. The founders, Urška Sršen and Sando Mur, want to use data from these products to better understand how women take care of their health. Our goal is to analyze this data and come up with ideas for how Bellabeat can grow even more.

First, we need to get the data that shows how people use Bellabeat's smart devices. We'll look at things like how much they move, their sleep, stress, and reproductive health. It's important to make sure the data is accurate and makes sense.

Next, we need to decide what we want to find out. Do we want to know how people use the products, what makes them stop using them, or maybe find groups of people who use them in similar ways?

We'll look at the data in different ways, like using graphs and charts, to see if we can find any patterns or interesting things. We might group people based on things like age or how often they use the products.

We'll also check if certain things are connected, like if people who are stressed sleep less. We'll look at how things change over time, like if more people use the products during certain seasons.As we go along, we'll see if there are ways to make the products better or if there are new things Bellabeat could add. We might also suggest partnerships that would be good for Bellabeat.When we find interesting things, we'll use pictures and charts to show the team. Finally, we'll put everything together and suggest things the team can do based on what we found. We'll talk about all of this in a simple and clear way so that everyone understands.

For more information [Bellabeat](#)

ASK PHASE ────────────────────────────────────

In this phase ,we need to identity the stakeholders

STAKEHOLDERS:

1) **Urška Sršen:** Bellabeat's Co-founder and Chief Creative Officer.
2) **Sando Mur:** Mathematician and Bellabeat's Co-founder , key member of the Bellabeat executive team
 3) Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and
reporting data that helps guide Bellabeat's marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat''s mission and business goals — as well as how you, as a junior data analyst, can help Bellabeat achieve them.

Business Task:

Analyze fitness data from Fitbit users to uncover consumer usage insights and apply high-level marketing recommendations to one of Bellabeat's products to drive growth and enhance digital marketing strategies for the company.

Business Objectives:

1. What are the trends identified?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

Sršen encourages me to use public data that explores smart device users' daily habits. She recommends a specific data set for us to view:

- [FitBit Fitness Tracker Data](#) Available on Kaggle . This Kaggle data set 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.

This data is contained in 18 CSV files and gonna used few CSV files for analysis

Data integrity:
Providing a ROCCC analysis for a dataset. ROCCC stands for Reliability, Originality, Comprehensiveness, Currentness, and Citations. Let's break down the analysis:

- Reliability (LOW): The dataset was collected from 30 individuals, and their gender is unknown. The small sample size and lack of gender information might raise concerns about the reliability of the dataset. Larger and more diverse samples generally contribute to higher reliability.
- Originality (LOW): The data was collected using Amazon Mechanical Turk, indicating that it's not original to the researchers but rather sourced from a third party. While Mechanical Turk is a common platform for data collection, it might affect the originality of the dataset.
- Comprehensiveness (MEDIUM): The dataset covers multiple fields, including daily activity intensity, calories used, daily steps taken, daily sleep time, and weight records. This suggests a moderate level of comprehensiveness, addressing various aspects of daily life. However, the assessment depends on the specific research goals and whether these fields are relevant.

- Currentness (MEDIUM): The data is 7 years old, and mention in dataset has habits may not change significantly over a few years. However, depending on the context of the study, changes in lifestyle patterns, technology, or societal trends could impact the relevance of the data. It's essential to consider the specific domain and how quickly things may evolve.
- Citations (HIGH): The data collector and source are well documented, which is a positive aspect. High citation and documentation contribute to transparency and enable others to assess the validity of the data.

In summary, the dataset has some limitations in terms of reliability due to a small sample size and a lack of gender information. The use of Mechanical Turk affects the originality, but the dataset's comprehensiveness is moderate. The data's age may impact its current relevance, but the high level of citation and documentation enhance its credibility. Researchers should carefully consider these factors when using or interpreting the dataset for their studies.

Tools:

I used Excel and SQL to load all files for my initial review of the data provided to see if there were any initial errors, cleaned data cleaning, transformation, analysis, and visualization.

PROCESS————————————————————————————————————

In this phase, I viewed all 18 files and decided to use 6 of the 18 datasets that are available to help with my analysis. The datasets chosen will be the following:

- **dailyActivity_merged.csv**
- **dailyCalories_merged.csv**
- **dailyIntensities_merged.csv**
- **dailySteps_merged.csv**
- **weightLogInfo_merged.csv**

I applied a dual approach, leveraging both Excel and SQL, to perform comprehensive data analysis. Each table, seamlessly imported from the original dataset, bore nomenclature consistent with its counterpart in the source dataset. In the cleaning phase, I adeptly employed both Excel and SQL to refine data types, meticulously configuring datetime fields to exclusively house date information or be tailored specifically to capture time (or activity hours).

1) Finding number of users in Bellabeat using different datasets ?

To find out how many users are in each dataset, I first changed the data type to text before putting the dataset into SQL.As the heart_rate file has been not imported file data fully , due to Data Type Mismatch or Encoding mismatches . I have imported the data in STRING datatype.

SQL CODE—

```
1    -- Number Of Users
2    |
3 •  use fitband;
4 •  select Count(distinct ID)  from daily_activity;
5 •  select Count(distinct ID)  from daily_calories;
6 •  select Count(distinct ID)  from daily_intensities;
7 •  select Count(distinct ID)  from daily_steps;
8 •  select Count(distinct ID)  from sleep_day;
9 •  select Count(distinct ID)  from weight_log;
10 • select Count(distinct ID)  from heart_rate;
11
```

Result: Number of users in following;

33

33

33

33

24

8

The analysis highlights a noticeable trend: daily activity, daily calories, daily intensities, and daily steps gain more user attention, whereas daily steps, sleep data, and weight log information have fewer users engaged.

Recommendation:

In response to user preferences, it is recommended to focus product efforts on refining and expanding features related to daily activities, calorie tracking, intensity monitoring, and step counts. Strengthening these aspects could lead to increased sales and heightened user satisfaction.

2) To find the relationship between calories and total steps ?

SQL CODE—

#clean the data and find activity level and calories burnt

Select * from daily_activity

Where
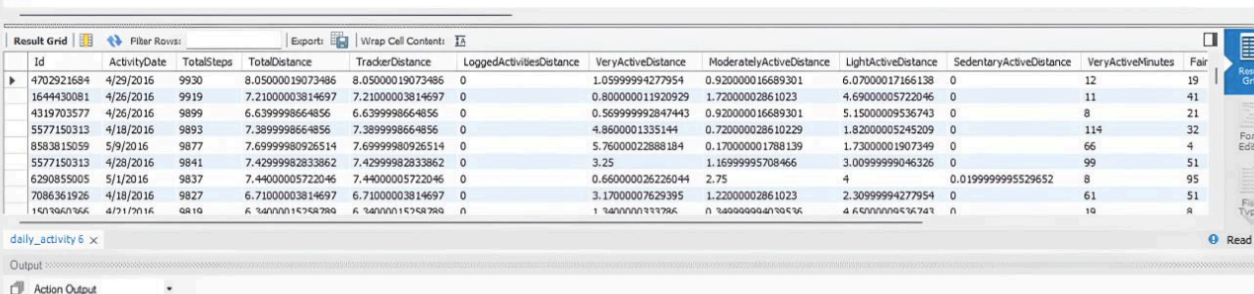
VeryActiveDistance+ModeratelyActiveDistance+LightActiveDistance != 0

And VeryActiveMinutes+FairlyActiveMinutes+LightlyActiveMinutes != 0

Order BY TotalSteps Desc;

<u>SQL OutPut—</u>

```
12
13
14    ## For finding activity level and calories burnt and also clean the data
15
16 •  SELECT *
17    FROM daily_activity
18    WHERE VeryActiveDistance+ModeratelyActiveDistance+LightActiveDistance != 0
19    AND VeryActiveMinutes+FairlyActiveMinutes+LightlyActiveMinutes != 0
20    ORDER BY TotalSteps DESC;
21
```

Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActiveDistance	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	Fair
4702921684	4/29/2016	9930	8.05000019073486	8.05000019073486	0	1.05999994277954	0.920000016689301	6.07000017166138	0	12	19
1644430081	4/26/2016	9919	7.21000003814697	7.21000003814697	0	0.800000011920929	1.72000002861023	4.69000005722046	0	11	41
4319703577	4/26/2016	9899	6.6399998664856	6.6399998664856	0	0.569999992847443	0.920000016689301	5.15000009536743	0	8	21
5577150313	4/18/2016	9893	7.3899998664856	7.3899998664856	0	4.8600001335144	0.720000028610229	1.82000005245209	0	114	32
8583815059	5/9/2016	9877	7.69999980926514	7.69999980926514	0	5.76000022888184	0.170000001788139	1.73000001907349	0	66	4
5577150313	4/28/2016	9841	7.42999982833862	7.42999982833862	0	3.25	1.16999995708466	3.00999999046326	0	99	51
6290855005	5/1/2016	9837	7.44000005722046	7.44000005722046	0	0.660000026226044	2.75	4	0.0199999995529652	8	95
7086361926	4/18/2016	9827	6.71000003814697	6.71000003814697	0	3.17000007629395	1.22000002861023	2.30999994277954	0	61	51
1503960366	4/21/2016	9819	6.34000015258789	6.34000015258789	0	1.34000003337385	0.349999994039535	4.65000009536743	0	19	8

In the provided SQL code, data for the least active members are cleaned by removing entries of individuals with columns such as Very Active Distance, Moderately Active Distance, Light Active Distance, Very Active Minutes, Fairly Active Minutes, and Lightly Active Minutes having values of zero.Taking the dataset into excel for further data visualization and forming scatter chart from the dataset.
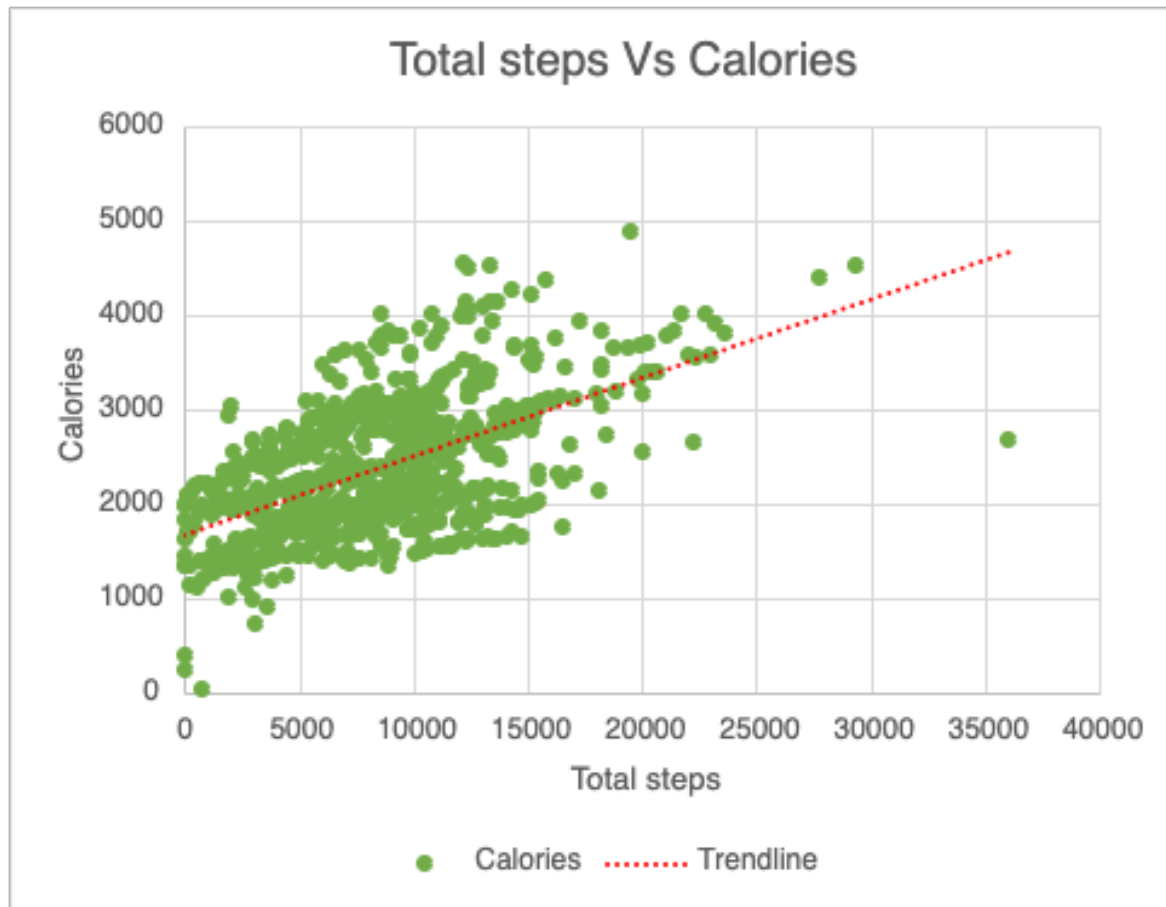
Fig 1: **Scatter Chart For Relationship Between Total steps and Calories**

Recommendation:

To engage with users whose calorie intake is higher, suggestions for active and calorie-burning activities can be made. Additionally, providing users with information on adjusting their calorie intake according to their fitness goals can contribute to their overall health and well-being
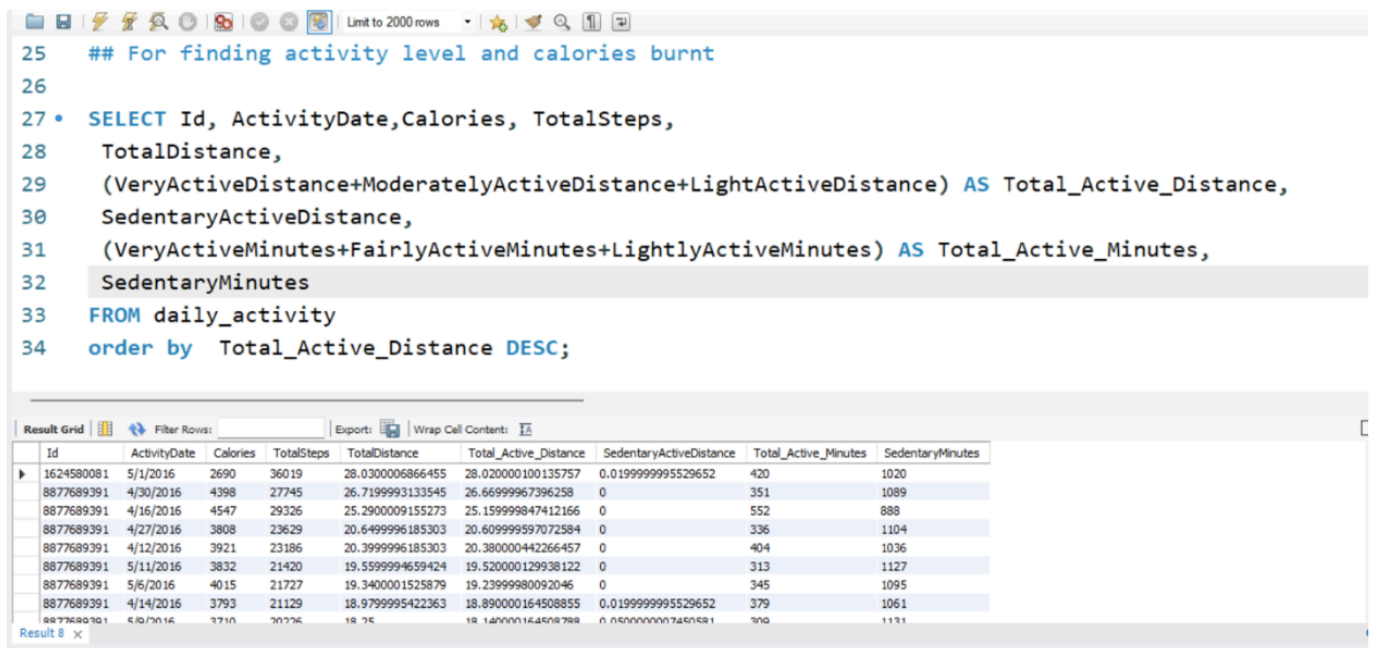
3) To find and analyze relation between total step and sedentary minutes?

SQL CODE—

to retrieve for finding activity level and calories burnt

Select Id, ActivityDate,Calories, TotalSteps,
TotalDistance,
(VeryActiveDistance+ModeratelyActiveDistance+LightActiveDistance) AS
Total_Active_Distance,
SedentaryActiveDistance,
VeryActiveMinutes+FairlyActiveMinutes+LightlyActiveMinutes) AS
Total_Active_Minutes,
SedentaryMinutes
From daily_activity
order by Total_Active_Distance Desc;

SQL Output——

To derive insights, we aggregated data on the least active and less sedentary minutes, combining them with sedentary minutes for a direct comparison with total steps. Our analysis reveals a negative relationship: an increase in sedentary minutes corresponds to a decrease in total steps (active minutes). This correlation is visually represented by a downward-trending line, underscoring the consistent decline in total steps as sedentary minutes rise.
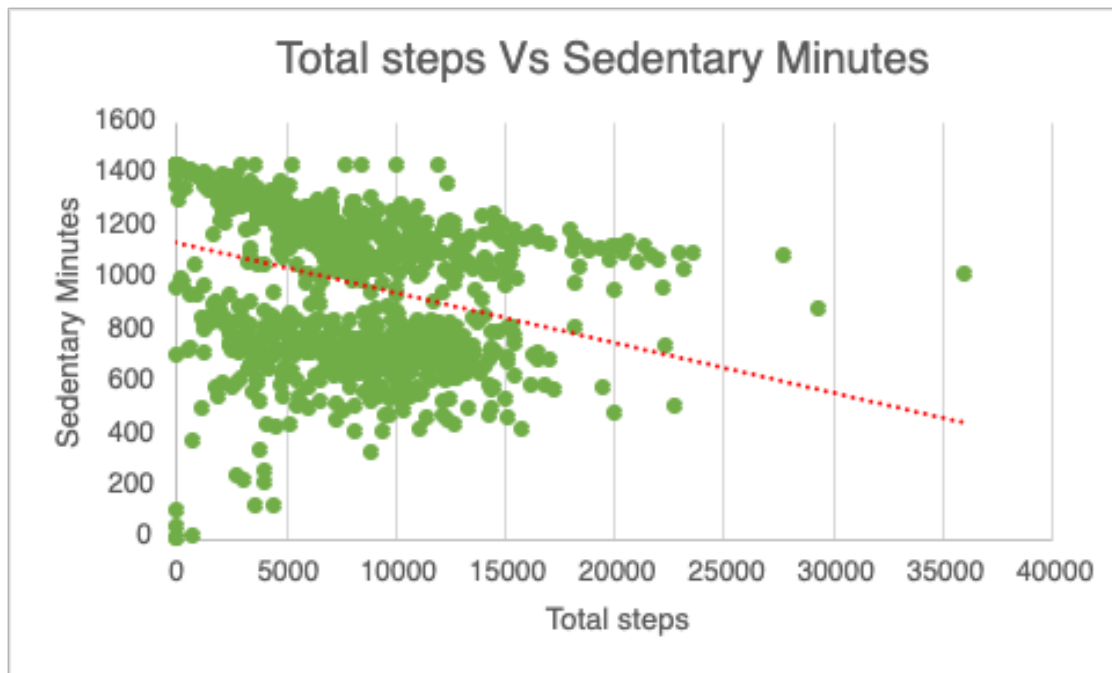

Fig 2: **Scatter Chart For Relationship Between Total steps and Sedentary Minutes**

Recommendation:
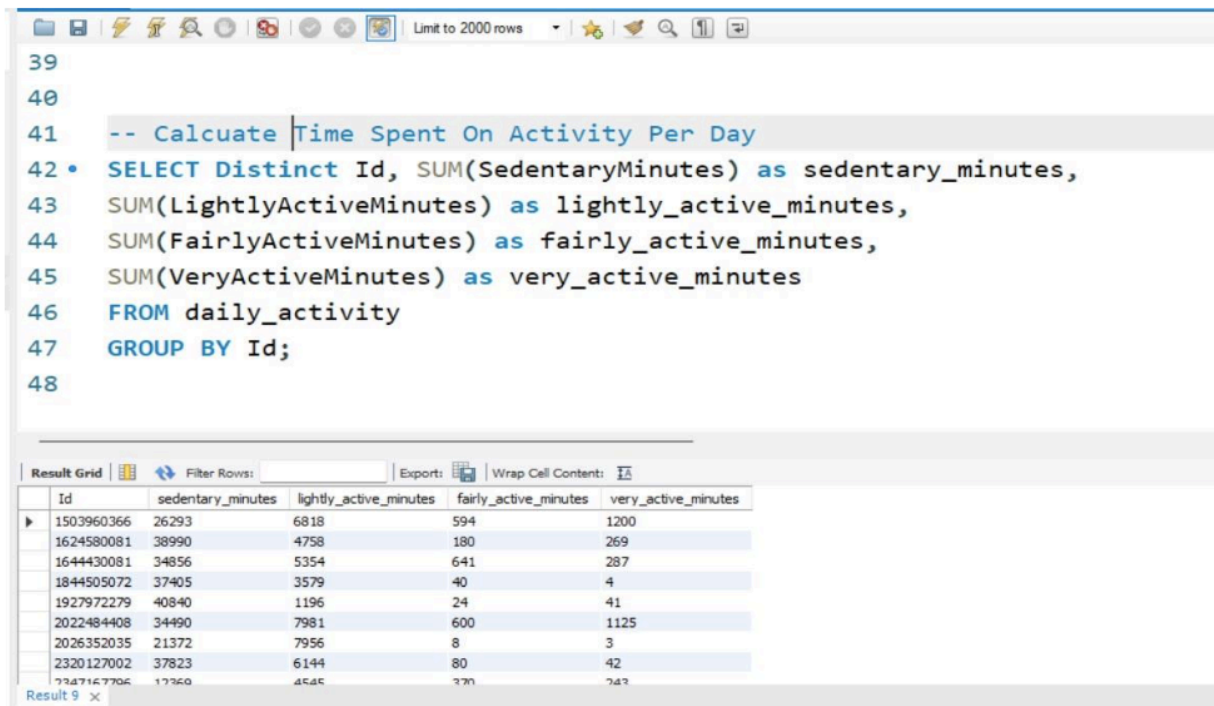User should be notified for Taking Regular Breaks while siting for that send notification and reminders, recommendation for active hobbies ,suggest walking ,reduce screen time and use practice Desk Exercises and standing desk.

4) Examine the active and sedentary minutes in daily part of life of users?

SQL CODE—

Calculation for active time per day

Select Distinct Id, SUM(SedentaryMinutes) as sedentary_minutes, SUM(LightlyActiveMinutes) as lightly_active_minutes, SUM(FairlyActiveMinutes) as fairly_active_minutes, SUM(VeryActiveMinutes) as very_active_minutes From daily_activity Group by Id;

Fig 3: **Pie Chart showing the relationship between various minutes in lifestyles**

As seen here, the majority of users using the Bellabeat product spend more time in sedentary minutes than in the combined durations of lightly active, fairly active, and very active minutes. This insight serves as a significant eye-opener, revealing that a substantial portion of the user base tends to have more inactive time than active time, which may impact their overall health.

5) Calculation of the most active week day on average?

SQL CODE—
```
# updating the column to string to date
use fitband;
SET SQL_SAFE_UPDATES = 0;
Update daily_activity
Set ActivityDate = STR_TO_DATE(ActivityDate, '%m/%d/%Y %h:%i');

#Add day_of_week column on daily_activities
Alter Table daily_activity
ADD day_of_week varchar(20)

# To Extract Day Of Week From Date
UPDATE daily_activity
SET Day_of_Week = dayname(ActivityDate);

# to calculate average of total steps, total distance, calories
Select AVG(TotalSteps) as avg_steps,
AVG(TotalDistance) as avg_distance,
AVG(Calories) as avg_calories,
Day_of_Week
From daily_activity
Group By  Day_of_Week;
```

SQL output

avg_steps	avg_distance	avg_calories	Day_of_Week
8125.006578947370	5.832236808050720	2356.0131578947400	Tuesday
7559.373333333330	5.488333327832320	2302.62	Wednesday
7405.836734693880	5.312244923506570	2199.5714285714300	Thursday
7448.230158730160	5.309920621326280	2331.785714285710	Friday
8152.975806451610	5.854677422708200	2354.967741935480	Saturday
6933.231404958680	5.027190073693470	2263	Sunday
7780.866666666670	5.552916660221920	2324.2083333333300	Monday

This is imported in Excel and line chart is obtain for observation.

After examining the data, a noticeable pattern emerges regarding the average total steps, total distance, and calories. The line chart indicates that Saturday and Tuesday, with Monday in close succession, are the days when people are most active. Conversely, Sunday and Monday appears to be the least favored day for exercise or physical activity.
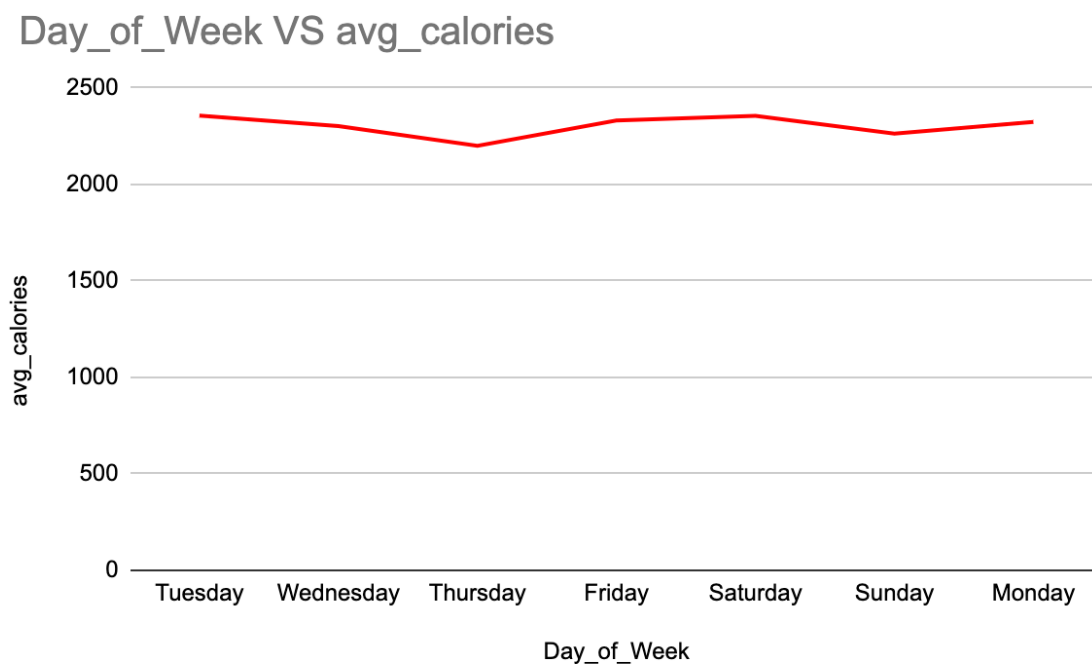

Fig 4:**Line Chart of Day Of Week And Average Calories**
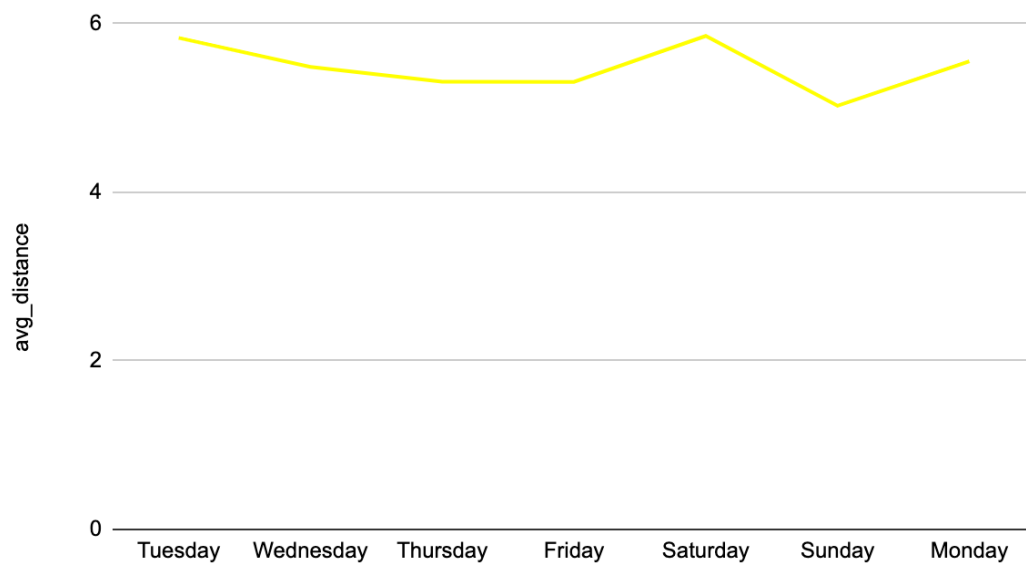
Day_of_Week VS avg_distance

Fig 5:Line Chart of Day Of Week And Average Distance
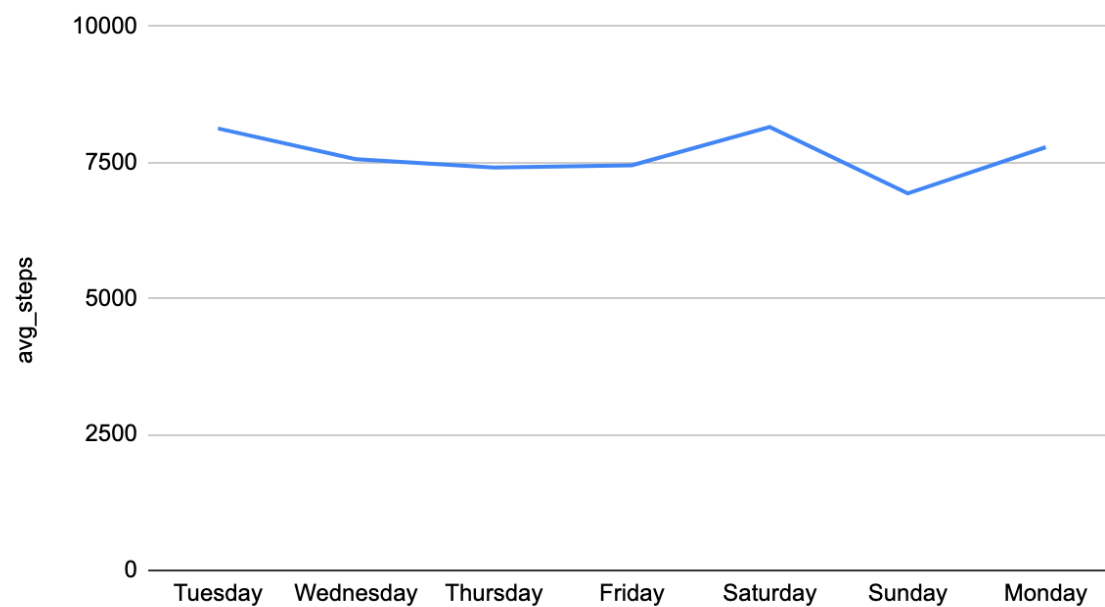
Day_of_week VS avg_steps

Fig 6:Line Chart of Day Of Week And Average Steps

<u>Recommendation</u>:
To boost activity on less active days, consider increasing notifications to motivate users. Furthermore, on days with higher activity levels, introducing notifications related to nutritious eating habits and reminders for adequate hydration could be valuable additional features.

6) Find the relation between average sleep on weekdays.

<u>SQL CODE-</u>

```
# updating the string to date
SET SQL_SAFE_UPDATES = 0;
use fitband;
Update sleep_day
Set SleepDay = STR_TO_DATE(SleepDay, '%m/%d/%Y %h:%i:%s %p');

# to adding column to add week of the day
Alter Table sleep_day
ADD day_of_week varchar(20);

# to extract the day of the week
UPDATE sleep_day
SET Day_of_Week = dayname(SleepDay);

#extracting the data
Select day_of_week,round(avg(TotalMinutesAsleep), 0) AS Avg_Sleep FROM sleep_day
Group By day_of_week;
```

Output of SQL code——

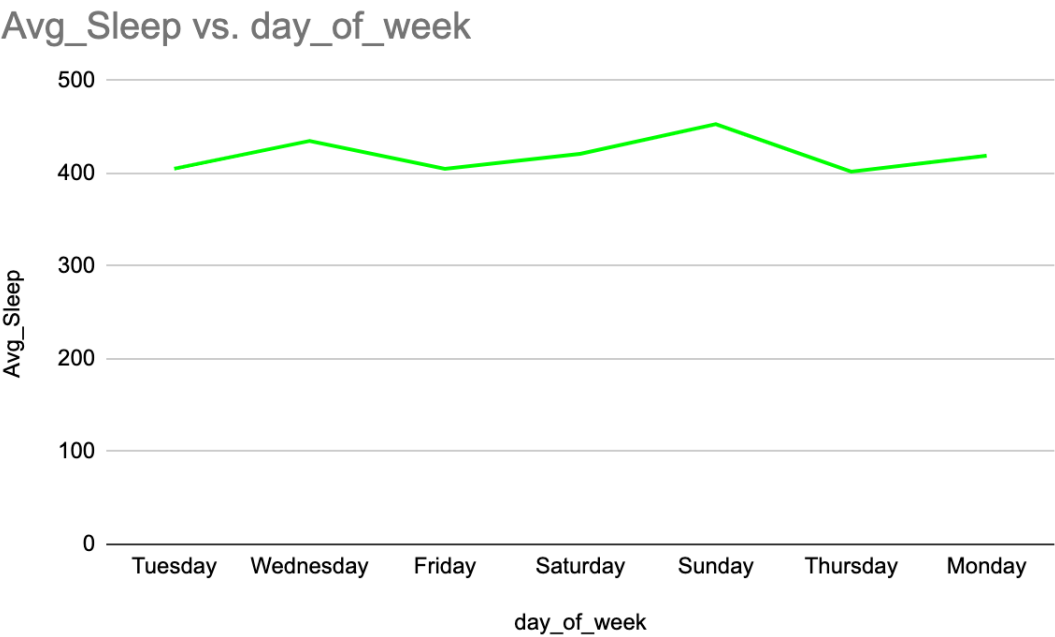| day_of_week	Avg_Sleep
Tuesday	405
Wednesday	435
Friday	405
Saturday	421
Sunday	453
Thursday	402
Monday	419

Avg_Sleep vs. day_of_week

Fig 7:Line Chart of Day Of Week And Average Calories

Recommendation:
After analyzing the data, it's evident that Sunday, Thursday, and Monday emerge as the least active days of the week. This trend translates to increased sedentary time and decreased active time, potentially hindering individuals from reaching their set fitness goals in a timely manner. Addressing strategies to boost activity on these specific days may contribute significantly to overall goal achievement.

Trends Identified:

1)Low Activity on Sundays:
2)Users walk the least on Sundays.
3)Sleep-Activity Correlation:
4)There is a potential correlation between sleep quality and daily activity levels.

Application to Bellabeat Customers

1) Sunday Activity Targeting: Encourage Bellabeat users to boost their activity on Sundays through targeted reminders and incentives.Provide personalized challenges or rewards to motivate increased step count on Sundays.

2) Sleep Enhancement Features:Enhance sleep tracking features to highlight the connection between sleep quality and daily activity.Educate users on the benefits of being more active during the day for improved sleep.

Influence on Bellabeat Marketing Strategy

1)Sunday-Specific Campaigns:Develop marketing campaigns that specifically target Sundays, emphasizing the importance of staying active on this day.Utilize social media, in-app notifications, and email campaigns to deliver these messages.

2)Integrated Sleep-Activity Messaging:Incorporate messages and content within the app that link daily activity with better sleep.Collaborate with sleep experts or influencers to create engaging content that highlights the benefits.

3)Promotion of Personalized Challenges: Introduce and promote personalized challenges that align with identified trends.Showcase success stories or testimonials related to improved activity levels on Sundays and enhanced sleep.

ACT

 (Suggestion for marketing team)

1) **Promote Physical Activity on Inactive Days**:-Encourage higher levels of physical activity, especially on less active days such as Sundays.

2)**Highlight the Link Between Physical Activity and Improved Sleep**:-Increase awareness about the positive correlation between regular physical activity and enhanced sleep quality.

3) **Adding new features** :-Bellabeat has the opportunity to amplify its sleep tracking feature by emphasizing the correlation between sleep quality and daily activity levels. This enhancement can serve as a compelling incentive for individuals to invest in Bellabeat products, promoting the establishment of healthier sleep habits through increased everyday physical activity.

4)**Adding minimum workout/calories burn goals**:-As users tend to walk the fewest steps on Sundays, a strategic approach would be to target them with an increased number of reminders encouraging a minimum of 8000 steps on this particular day. The company can further enhance user engagement by creating posts and videos within the app, emphasizing the significance of daily walking and the positive impact it has on overall well-being.