

Google Data Analytics Capstone Project

(Bellabeat Case Study)

Title— Bellabeat Case Study

How Can a Wellness Technology Company Play It Smart?

Author— Amit Baviskar

Date- 5 Dec 2023

09/12/2023

For More Information [Bellabeat](#)

Author— Amit Baviskar



LoGo

Introduction

This project serves as the culmination of our efforts toward earning the Google Data Analytics Certification.

In this scenario, our role involves functioning as Data Analysts within Bella Beat dynamic marketing team. Founded by Urška Sršen and Sando Mur, Bellabeat stands out as a high-tech company specializing in the production of health-focused smart products. These innovative devices gather data on various aspects such as activity, sleep, stress, and reproductive health.

The overarching mission of Bella beat is to Empower Women by providing them with insightful knowledge about their health and daily habits. To uncover new avenues for growth, we are tasked with delving into the analysis of data sourced from smart devices. This analysis holds the potential to significantly contribute to the development of the company's marketing strategy.

ASK PHASE:

In this phase, we need to identify the stakeholders

STAKEHOLDERS:

- 1) Urška Sršen: Bella Beat Co-founder and Chief Creative Officer.
- 2) Sando Mur: Mathematician and Bella Beat 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 Bella Beat 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 Bella Beat achieve them.

Business Task:

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

Business Objectives:

- What are the trends identified?
- How could these trends apply to Bellabeat customers?
- How could these trends help influence Bellabeat marketing strategy?

PREPARE:

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 trackers 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 integrity:

ROCCC stands for Reliability, Originality, Comprehensiveness, Currentness, and Citations. Let's break down the analysis:

ROCCC analysis

Reliability: LOW (1) – The data set was collected from 30 individuals whose gender is unknown.

Originality: LOW (1) – Third-party data was collected using Amazon Mechanical Turk.

Comprehensiveness: MEDIUM (2) – The dataset contains multiple fields on daily activity intensity, calories used, daily steps taken, daily sleep time, and weight records.

Currentness: MEDIUM (2) – The data is 7 years old, but the habits of how people live do not change significantly over a few years.

Cited: HIGH (3) – The data collector and source are well-documented.

In this representation, the numbers in parentheses (**1, 2, 3**) indicate the different levels, with higher numbers representing higher levels of reliability, originality, comprehensiveness, currency, and citation.

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 enhances 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.

What tools are you choosing and why? I employed a combination of Excel and SQL for data analysis. Each table, imported from the original dataset, was named in accordance with its counterpart in the source dataset. During the cleaning phase, I utilized both Excel and SQL to adjust data types, ensuring that datetime fields contained only date information or were refined to exclusively represent time (or activity hours).

To find out how many users are in each dataset, I first changed the data type to text before putting the dataset into SQL. The datasets chosen will be the following:

CSV File Name

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

As above analysis, we can observe more users daily_activity, daily_Calories , daily_Intensities , daily_steps, are preferred than daily_Steps, sleep_Day, weight_Log_Info has users are less in this scenario

ANALYSE:

To find out how many users are in each dataset, I first changed the data type to text before putting the dataset into SQL.

Data in Data Sample Set: Data present in Sample file Cannot be Import in Mysql. We Can Import Data in Mysql Database by Converting all Data into Text.

1) Determining the number of users in each dataset:

```
1  -- Count Number Of Unquie User In FitBand According To Following Activites
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
```

Table Formed	Distinct Count
• dailyActivity_merged.csv	33
• dailyCalories_merged.csv	33
• dailyIntensities_merged.csv	33
• dailySteps_merged.csv	33
• sleepDay_merged.csv	24
• weightLogInfo_merged.csv	8

Based on the analysis, it's clear that more users are actively engaged with features such as daily activity, daily calories, daily intensities, and daily steps, while there is comparatively less interest in daily steps, sleep data, and weight log information.

But the User Data sample Size is Small.

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

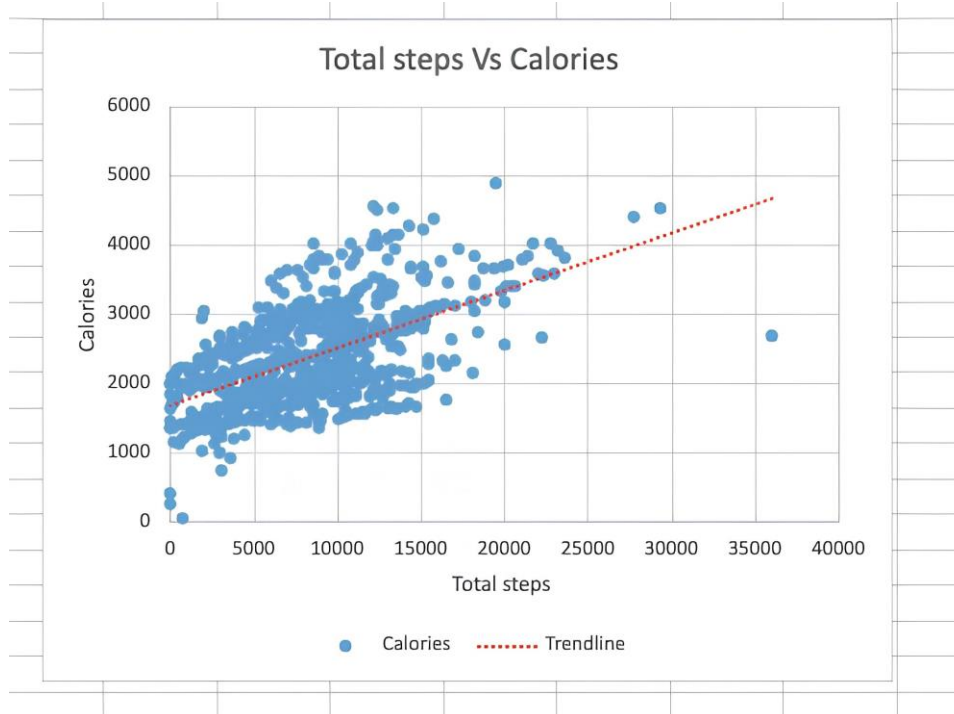
```

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

```

Result Grid											
Filter Rows: Export Wrap Cell Contents											
		TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActiveDistance	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes
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	0.56999992847443	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	0.019999995529652	8	95
7086361926	4/18/2016	9827	6.71000003814697	6.71000003814697	0	3.17000007629395	1.22000002861023	2.30999994277954	0	61	51
15210661566	4/17/2016	9810	6.34000015758780	6.34000015758780	0	1.34000001117786	0.340000004010656	4.65000000653674	0	10	8

In the SQL script mentioned earlier, data cleaning involves excluding records of the least active members by filtering out entries where certain columns—such as "Very Active Distance," "Moderately Active Distance," "Light Active Distance," "Very Active Minutes," "Fairly Active Minutes," and "Lightly Active Minutes"—are non-zero. Following this cleaning process, the refined data is then analyzed in Excel to investigate the correlation between total steps and calorie consumption



Scatter Plot Graph Showing Relationship Of Calories and Total Step

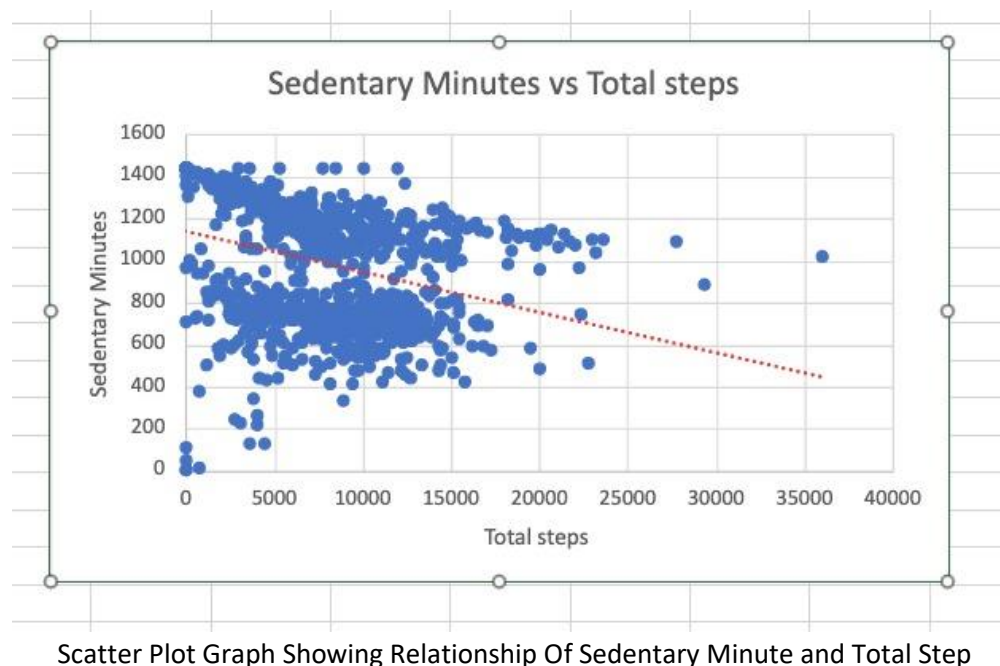
Visualizing the data through a scatter plot unveils a positive relationship. The plot indicates that users with higher step counts also tend to have a higher calorie intake, a correlation underscored by the trend line.

3) To find and analyze relation between total step And Sedentary Time?

```
25  ## For finding activity level and calories burnt
26
27 • SELECT Id, ActivityDate, Calories, TotalSteps,
28       TotalDistance,
29       (VeryActiveDistance+ModeratelyActiveDistance+LightActiveDistance) AS Total_Active_Distance,
30       SedentaryActiveDistance,
31       (VeryActiveMinutes+FairlyActiveMinutes+LightlyActiveMinutes) AS Total_Active_Minutes,
32       SedentaryMinutes
33 FROM daily_activity
34 order by Total_Active_Distance DESC;
```

Result Grid									
Filter Rows:									
Export: Wrap Cell Content:									
Id	ActivityDate	Calories	TotalSteps	TotalDistance	Total_Active_Distance	SedentaryActiveDistance	Total_Active_Minutes	SedentaryMinutes	
1624580081	5/1/2016	2690	36019	28.0300006866455	28.020000100135757	0.0199999995529652	420	1020	
8877689391	4/30/2016	4398	27745	26.7199993133545	26.66999967396258	0	351	1089	
8877689391	4/16/2016	4547	29326	25.2900009155273	25.159999847412166	0	552	888	
8877689391	4/27/2016	3808	23629	20.6499996185303	20.609999997072584	0	336	1104	
8877689391	4/12/2016	3921	23186	20.3999996185303	20.380000442266457	0	404	1036	
8877689391	5/11/2016	3832	21420	19.5599994659424	19.520000129938122	0	313	1127	
8877689391	5/6/2016	4015	21727	19.3400001525879	19.23999980092046	0	345	1095	
8877689391	4/14/2016	3793	21129	18.9799995422363	18.890000164508855	0.0199999995529652	379	1061	
8877689391	5/10/2016	3710	20736	18.75	18.140000164508855	0.0199999995529652	390	1131	

To uncover insights, we combined data on the least active and less sedentary minutes, merging them with sedentary minutes for a direct comparison with total steps. Our analysis indicates a negative relationship: an uptick in sedentary minutes coincides with a decline in total steps (active minutes). This correlation is graphically depicted by a downward-trending line, emphasizing the consistent reduction in total steps as sedentary minutes increase.



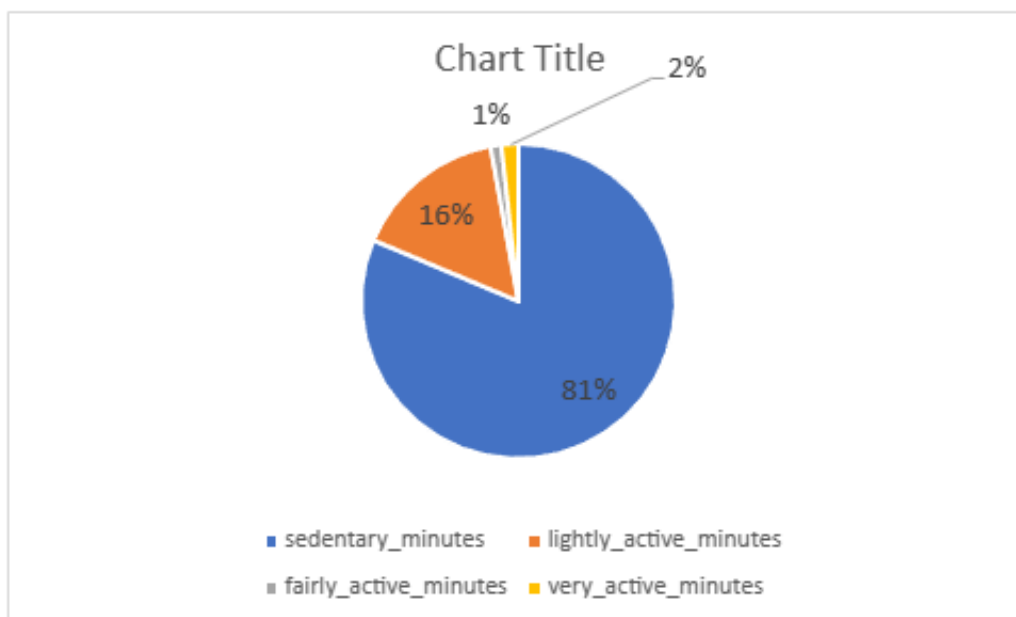
To obtain relevant insights, we included both the least active and less sedentary minutes, combining them with sedentary minutes for comparison with the total step count. Upon analysis of the dataset, it becomes apparent that there is a negative relationship between sedentary minutes and total steps (active minutes). The trend line further supports this observation by displaying a downward trajectory, indicating that an increase in sedentary minutes is associated with a decrease in total steps.

4) Examine Active And sedentary Time In Daily Life Cycle

```
39
40
41 -- Calculate Time Spent On Activity Per Day
42 • SELECT Distinct Id, SUM(SedentaryMinutes) as sedentary_minutes,
43    SUM(LightlyActiveMinutes) as lightly_active_minutes,
44    SUM(FairlyActiveMinutes) as fairly_active_minutes,
45    SUM(VeryActiveMinutes) as very_active_minutes
46 FROM daily_activity
47 GROUP BY Id;
48
```

Id	sedentary_minutes	lightly_active_minutes	fairly_active_minutes	very_active_minutes
1503960366	26293	6818	594	1200
1624580081	38990	4758	180	269
1644430081	34856	5354	641	287
1844505072	37405	3579	40	4
1927972279	40840	1196	24	41
2022484408	34490	7981	600	1125
2026352035	21372	7956	8	3
2320127002	37823	6144	80	42
2347167706	12369	4545	370	243

Result 9



Pie Chart Showing Minute Distribution in Activity in A Daily

Upon examination, it becomes evident that a majority of Bellabeat product users allocate more time to sedentary minutes than the combined duration of lightly active, active, and very active minutes. This revelation underscores a crucial observation — a considerable segment of the user base tends to

have a higher proportion of inactive time compared to active time. This trend may have implications for their overall health.

5) Calculation Of Most Active Weekday (On Average):

```
--  
49 -- Calculation Of Most Active Weekday (On Average)  
50 • use fitband;  
51 • SET SQL_SAFE_UPDATES = 0; -- In MySQL, the SQL_SAFE_UPDATES mode can be disabled by setting it back to 1.  
52 -- It's important to note that setting SQL_SAFE_UPDATES to 0  
53 -- can potentially lead to unintended updates and deletions, so use it with caution  
54 • Update daily_activity  
55 Set ActivityDate = STR_TO_DATE(ActivityDate, '%m/%d/%Y %h:%i'); --  
56  
57 -- Add day_of_week column on daily_activities  
58 • Alter Table daily_activity  
59 ADD day_of_week nvarchar(20)  
60  
61 -- To Extract Day Of Week From Date
```



The screenshot shows a database interface with a 'Result Grid' tab selected. The grid displays data for the 'daily_activity' table, with columns: avg_steps, avg_distance, avg_calories, and Day_of_Week. The data is sorted by avg_steps in descending order. The interface also includes a 'Filter Rows' button, an 'Exports' button, and a 'Wrap Cell Contents' checkbox.

avg_steps	avg_distance	avg_calories	Day_of_Week
8125.006578947368	5.832236808050717	2356.0131578947367	Tuesday
7559.373333333333	5.488333327832321	2302.62	Wednesday
7405.836724693878	5.312244923506565	2199.5714285714284	Thursday
7448.230158730159	5.309920621326282	2331.785714285714	Friday
8152.975806451613	5.854677422708202	2354.967741935484	Saturday

```

60
61 -- To Extract Day Of Week From Date
62 UPDATE daily_activity
63 SET Day_of_Week = dayname(ActivityDate);
64
65 • Select AVG(TotalSteps) as avg_steps,
66     AVG(TotalDistance) as avg_distance,
67     AVG(Calories) as avg_calories,
68     Day_of_Week
69 From daily_activity
70 Group By Day_of_Week;
71
72 -- To reverse the setting, you can use the following command:
73 • SET SQL_SAFE_UPDATES = 1;
74
75
76

```

Result Grid

avg_steps	avg_distance	avg_calories	Day_of_Week
8125.006578947368	5.832236808050717	2356.0131578947367	Tuesday
7559.373333333333	5.488333327832321	2302.62	Wednesday
7405.836734693878	5.312244923506565	2199.5714285714284	Thursday
7448.230158730159	5.309920621326282	2331.785714285714	Friday
8152.975806451613	5.854677422708202	2354.967741935484	Saturday

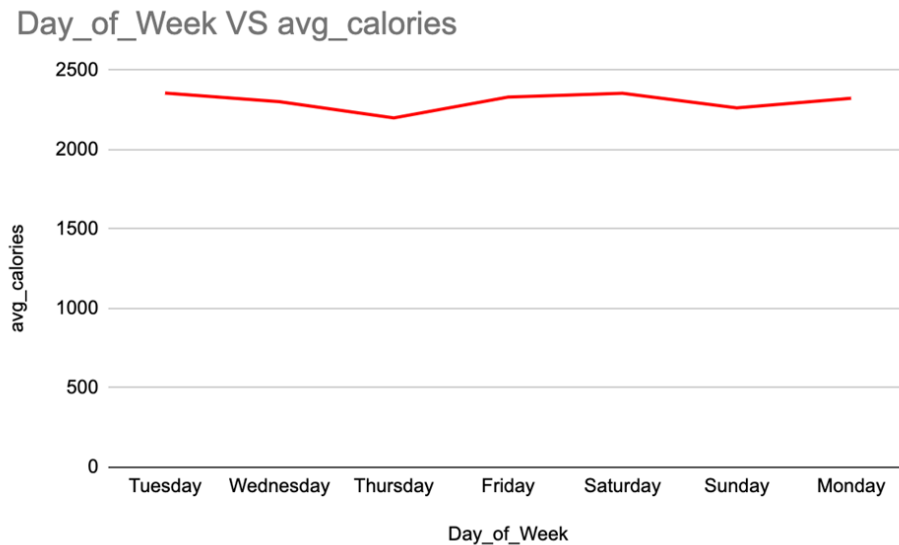
avg_steps	avg_distance	avg_calories	Day_of_Week
8125.006578947370	5.832236808050720	2356.0131578947400	Tuesday
7559.373333333330	5.488333327832320	2302.62	Wednesday
7405.836734693880	5.312244923506570	2199.5714285714300	Thursday
7448.230158730160	5.309920621326280	2331.785714285710	Friday
8152.975806451610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

Observation BY Plotting Graph

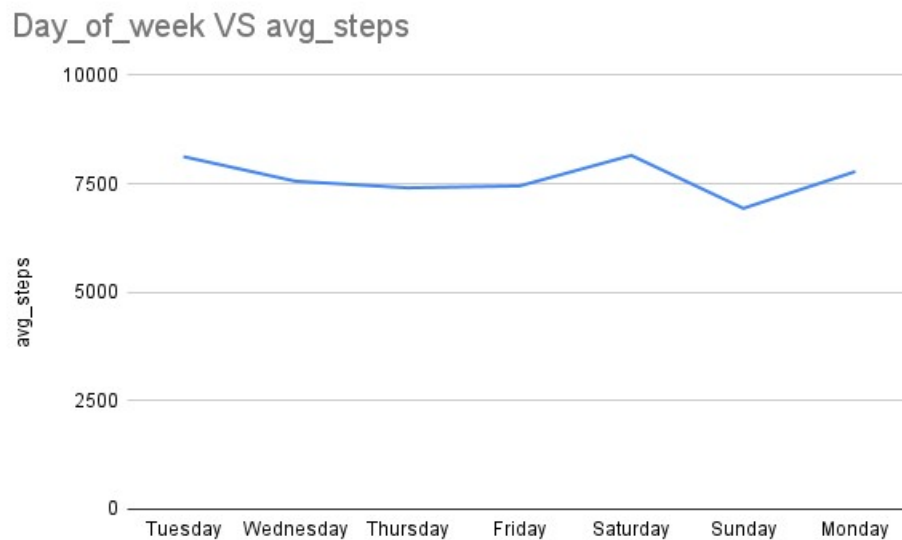
Upon scrutinizing the data, a clear trend comes to light concerning the average total steps, total distance, and calories. The line chart distinctly portrays that Saturday and Tuesday, closely followed by Monday, emerge as the peak days for physical activity. In contrast, Sunday stands out as the least preferred day for engaging in exercise or any form of physical exertion.

The data suggests a consistent cycle in people's activity levels throughout the week. Saturdays and Tuesdays exhibit heightened levels of activity, potentially influenced by factors such as the weekend's

relaxed atmosphere, the commencement of the workweek, and the availability of more leisure time for exercise. Mondays also show increased activity levels, indicating a motivated start to the week

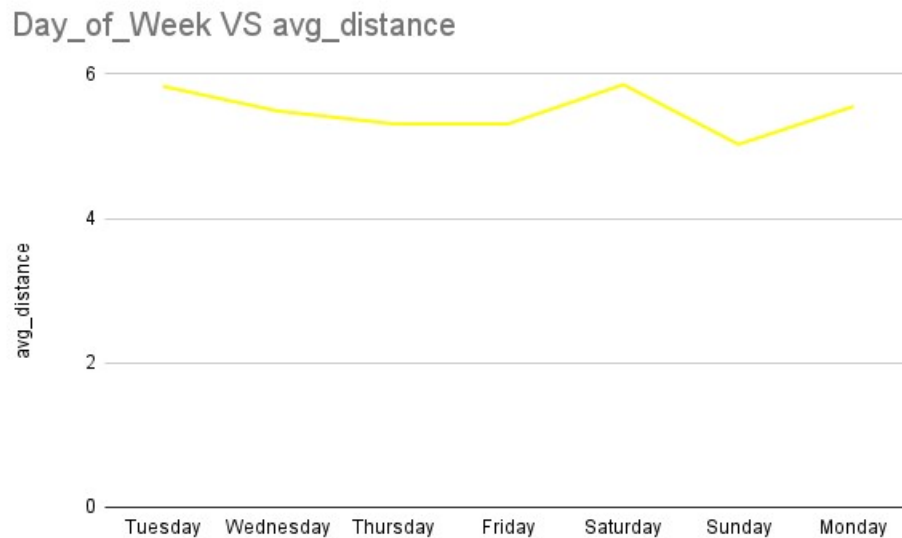


Line Graph Showing Relationship of Day of Week and Calories



Line Graph Showing Relationship of Day of Week and Average Step

From Above graph (Line Graph Showing Relationship of Day of Week and Average Step), We Can conclude That the Monday, Tuesday and Saturday is Most Active Day in a Week. Sunday is the Least Active Day of the Week.



Line Graph Showing Relationship of Day of Week and Average Distance Covered

From Above graph (Line Graph Showing Relationship of Day of Week and Average Distance Covered), We Can conclude That the Monday, Tuesday and Saturday is Most Active Day in a Week By Considering Distance Covered. Sunday is the Least Active Day of the Week.

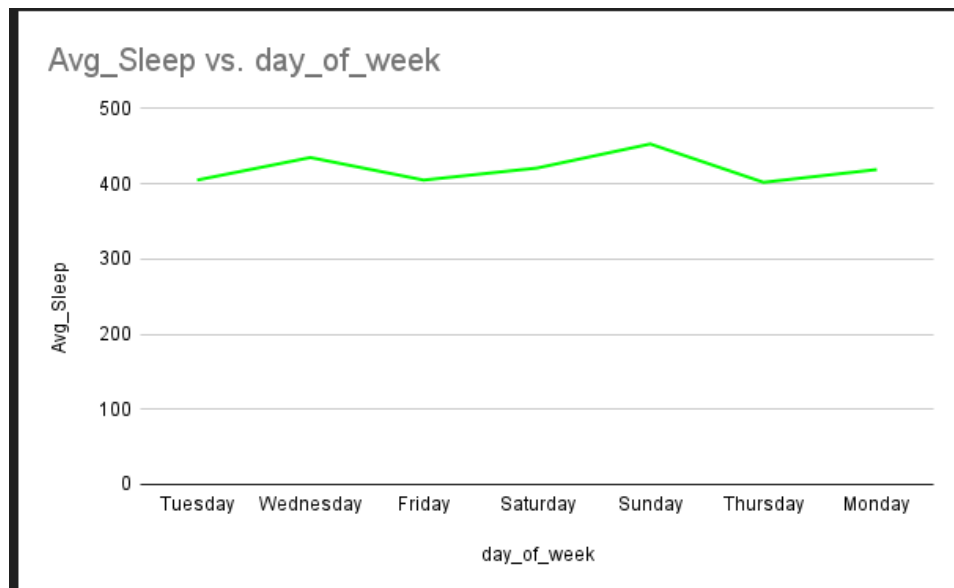
6) Find the relation between average sleep on weekdays.

```

77 • use fitband;
78 • Update sleep_day
79   Set SleepDay = STR_TO_DATE(SleepDay, '%m/%d/%Y %h:%i:%s %p');
80
81   -- Add Day Of Week Column On sleep_date
82 • Alter Table sleep_day
83   ADD day_of_week varchar(20);
84
85 • UPDATE sleep_day
86   SET Day_of_Week = dayname(SleepDay);
87
88 • Select day_of_week,
89   round(avg(TotalMinutesAsleep), 0) AS Avg_Sleep FROM sleep_day
90   Group By day_of_week;
91
92 • SET SQL_SAFE_UPDATES = 1;

```

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



Line Graph Showing Relationship of Day of Week and Average Sleep

Following data analysis, it's apparent that Sunday, Thursday, and Monday rank as the least active days of the week. This pattern correlates with extended periods of sedentary behavior and reduced active time, posing a potential challenge for individuals striving to achieve their fitness goals within a designated timeframe. Implementing targeted strategies to enhance activity levels on these specific days could play a pivotal role in advancing overall goal attainment.

SHARE:

Emerging Trends:

Subdued Sundays :Users exhibit lower activity levels on Sundays.

Sleep Harmony :A potential link between sleep quality and daily activity.

Application to Bellabeat Customers

Sunday Boost:

Motivate users to elevate Sunday activity via targeted reminders and incentives.

Introduce custom challenges and rewards for amplifying step counts on Sundays.

Revamped Sleep Features:

Enhance sleep tracking to spotlight the interplay between sleep and daily activity.

Educate users on the merits of increased daytime activity for enhanced sleep quality.

Strategic Tweaks in Bellabeat Marketing:

Sunday-Centric Campaigns:

Craft campaigns homing in on Sundays, underlining the importance of staying active.

Leverage social media, in-app notifications, and emails for impactful messaging.

Harmonized Sleep-Activity Messaging:
Integrate in-app messages linking daily activity with improved sleep

ACT:

The analysis of user behavior reveals a predominant inclination towards a sedentary lifestyle, emphasizing the need for proactive measures to encourage healthier habits. Here are some recommendations based on the findings:

1. Smart Device Notifications for Activity:

- Send timely notifications through smart devices during peak activity hours to prompt users to engage in physical activities, combating the prevalent sedentary lifestyle.

2. Sleep Reminders and Enhanced Features:

- Implement app notifications to remind users to prioritize sufficient sleep daily. Additionally, consider introducing new sleep measurement features or products to enhance users' sleep tracking experience.

3. Calorie Challenges and Rewards System:

- Set up daily or weekly calorie challenges for users, rewarding points to top performers. Accumulated points can be redeemed as discounts for their future product purchases, encouraging a healthier lifestyle.

4. Sunday Activity Boost:

- Recognize the lower step count on Sundays and leverage targeted reminders to motivate users to achieve at least 8000 steps on that day. Create engaging app content, such as posts and videos, emphasizing the importance of increased daily steps.

5. Addressing Sedentary Behavior:

- Highlight the sedentary lifestyle trend among users (79% to 81% of their time) and provide motivational content to encourage more physical activity.

6. Promoting the Steps-Calories Relationship:

- Showcase the positive correlation between walking more steps and burning calories. Motivate users to adopt a more active lifestyle by illustrating the benefits of increasing daily step counts.

7. Sleep Health Content and Features:

- Emphasize the importance of adequate sleep (at least 8 hours a day) through health-conscious content within the app. Introduce features like creating personalized sleep schedules to help users prioritize and manage their sleep.

8. Sunday Activity Push:

- Recognize Sunday as a day of low activity across various metrics. Encourage users to boost their activity levels on Sundays by delivering targeted reminders and creating content that highlights the benefits of increased activity during the weekend.

By implementing these recommendations, the company can not only address existing patterns but also actively engage users in adopting healthier lifestyles, contributing to their overall well-being