BELLABEAT CASE STUDY PORTFOLIO

Introduction

Bellabeat is a cutting-edge company that produces health-focused smart products designed for women.

The project's objective was to provide data-driven recommendations to improve the marketing strategy for the Bellabeat app. These recommendations were derived from an in-depth analysis of smart device data, offering insights into how consumers utilize their devices.

The report adheres to the six steps of the data analysis process: ask, prepare, process, analyze, share, and act.

ASK :

Identify Business Task

The primary objective was to analyze data from Bellabeat smart devices to understand how people are currently using their devices. The insights gained from this analysis aimed to provide high-level recommendations to inform and enhance Bellabeat's marketing strategy.

Consider Key Stakeholders

The key stakeholders for this project were:

- **Urška Sršen**: Co-founder and Chief Creative Officer of Bellabeat.
- **Sando Mur**: Co-founder, mathematician, and key executive team member at Bellabeat.
- **Bellabeat Marketing Analytics Team**: Tasked with implementing and optimizing marketing strategies based on the analysis and recommendations derived from this study.

PREPARE

Data Source

- The data for this analysis comes from the Fitbit Fitness Tracker Data, available on Kaggle and provided by Mobius.

Data Accessibility & Privacy

- The dataset is licensed under CC0: Public Domain. This means the owner has relinquished all rights to the work worldwide under copyright law, allowing it to be freely copied, modified, and distributed, including for commercial purposes, without the need for permission.

DATA ORGANIZATION

The data was available as 18 different CSV files. Each file contained various quantitative data generated from Fitbit health trackers, presented in either a long or wide format.

About the Dataset

Respondents generated this dataset to a distributed survey via Amazon Mechanical Turk between April 12, 2016, and May 12, 2016. Thirty-three eligible Fitbit users consented to share their personal tracker data, which includes minute-level output for physical activity, heart rate, and sleep monitoring.

I will use the daily and hourly data for this analysis rather than delving into detailed minute-based user performance. To conduct this analysis, I will need to combine several tables.

PROCESS

I have used MY SQL Workbench for this project to help process and analyze and for visualization, I have used Power BI.

The following tables were imported in My SQL Workbench:
Daily Activity_merged, Dailyactivity_sleep, Dailycalories_merged, Dailyintensities_merged, Dailysteps_merged, sleepday_merged,weightloginfo_merged.

Data Exploration and Quality Checks

Checking Unique IDs:

```
16 • select count(distinct(id))
17   from bellabeat.dailyactivity_merged;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(distinct(id))
33

The tables were checked for number of unique users.
All the tables had 33 unique users except sleepday_merged which had 24 users and weightloginfo_merged which had 8 users.

```
33   order by id_sleep;
34 • select count(distinct(id))
35   from sleepday_merged;
36
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(distinct(id))
24

```
34 • select count(distinct(id))
35   from weightloginfo_merged;
36
```

Result Grid | Filter Rows:

count(distinct(id))
8

Data Cleaning and De-duplication

Identifying Duplicates:

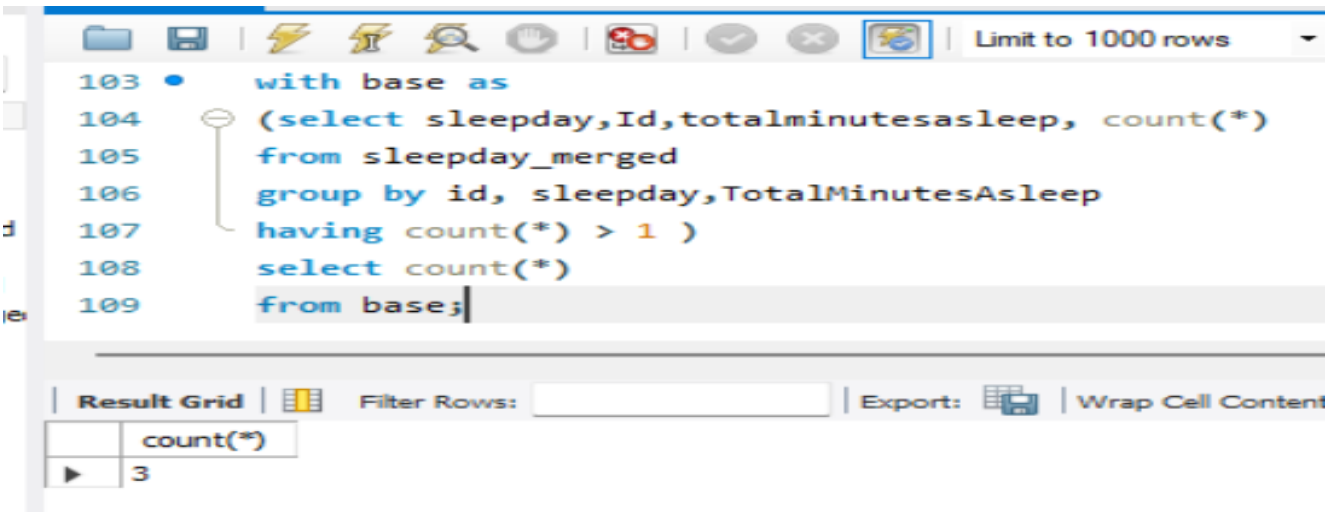
All the tables were then checked for the duplicate values grouped by id and date column.

```
38 •    with base as
39      (select activityday,Id, count(*)
40      from dailycalories_merged
41      group by id, activityday
42      having count(*) > 1 )
43      select count(*)
44      from base;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

count(*)
0

Three duplicates were found in sleepday_merged table.



```
110 •  select *
111    from sleepday_merged;
112 •  select *,
113    row_number() over (partition by id,sleepday,totalsleeprecords,totalminutesasleep,totaltimeinbed) as row_num
114    from sleepday_merged;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed	row_num
1503960366	4/12/2016 12:00:00 AM	1	327	346	1
1503960366	4/13/2016 12:00:00 AM	2	384	407	1
1503960366	4/15/2016 12:00:00 AM	1	412	442	1
1503960366	4/16/2016 12:00:00 AM	2	340	367	1
1503960366	4/17/2016 12:00:00 AM	1	700	712	1
1503960366	4/19/2016 12:00:00 AM	1	304	320	1
1503960366	4/20/2016 12:00:00 AM	1	360	377	1
1503960366	4/21/2016 12:00:00 AM	1	325	364	1
1503960366	4/23/2016 12:00:00 AM	1	361	384	1
1503960366	4/24/2016 12:00:00 AM	1	430	449	1
1503960366	4/25/2016 12:00:00 AM	1	277	323	1
1503960366	4/26/2016 12:00:00 AM	1	245	274	1

Result 465 ×

```
116 •  with cte as (
117    select *,
118    row_number() over (partition by id,sleepday,totalsleeprecords,totalminutesasleep,totaltimeinbed) as row_num
119    from sleepday_merged)
120    select*
121    from cte
122    where row_num > 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed	row_num
4388161847	5/5/2016 12:00:00 AM	1	471	495	2
4702921684	5/7/2016 12:00:00 AM	1	520	543	2
8378563200	4/25/2016 12:00:00 AM	1	388	402	2

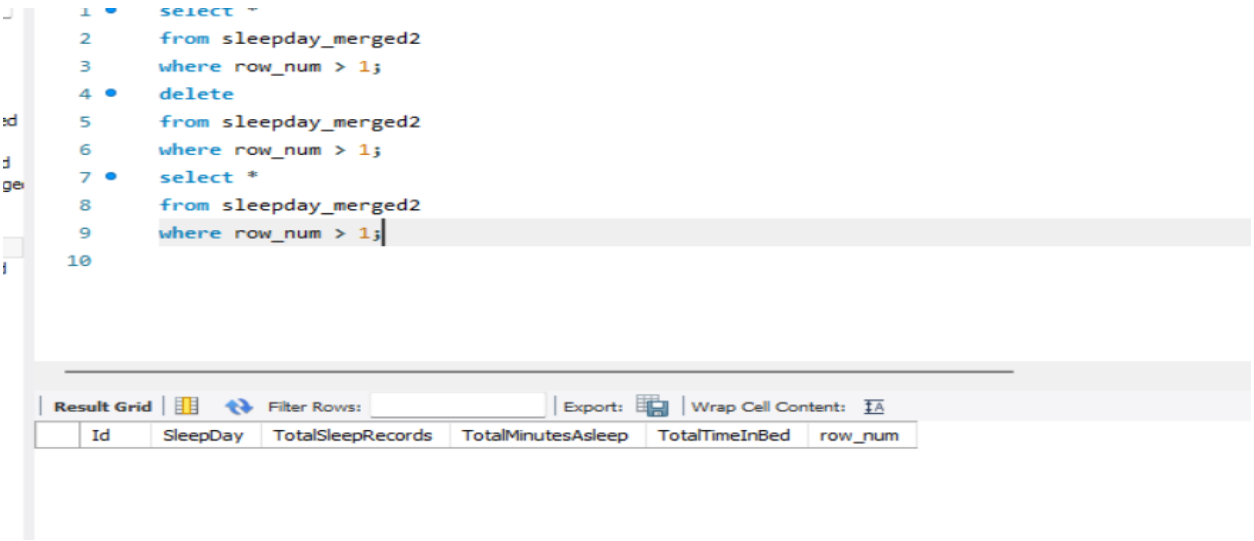
```
123 •  select * from sleepday_merged
124    where id = 4388161847 and TotalMinutesAsleep = 471;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed
4388161847	5/5/2016 12:00:00 AM	1	471	495
4388161847	5/5/2016 12:00:00 AM	1	471	495

Removing Duplicates:

To remove duplicates I did not make any changes in the raw table. I created a copy of the table and then removed the duplicates.



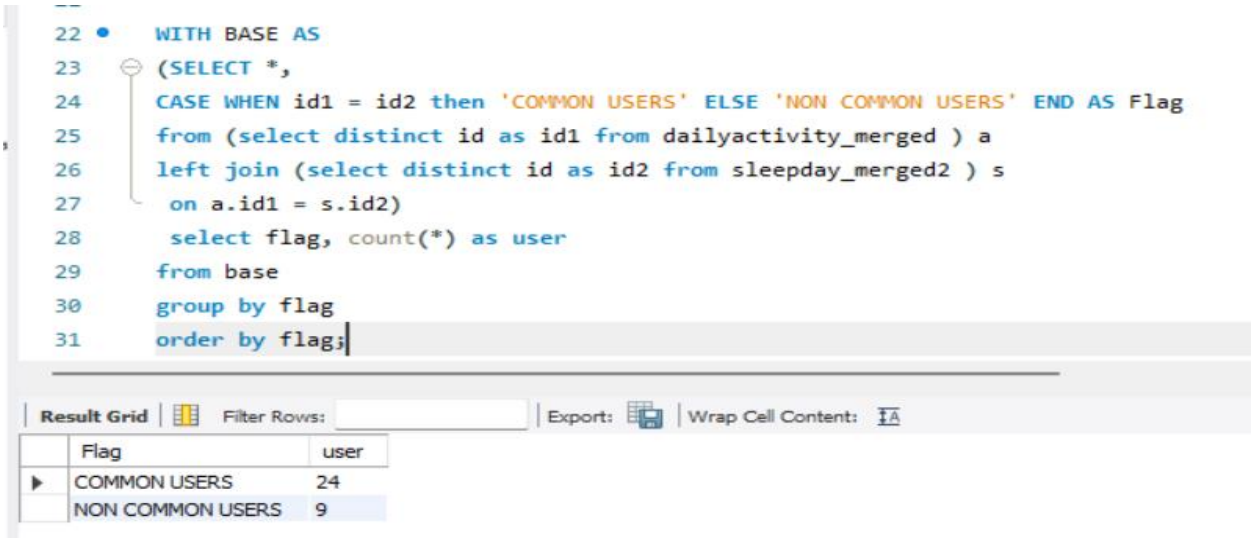
```
select *
from sleepday_merged2
where row_num > 1;
delete
from sleepday_merged2
where row_num > 1;
select *
from sleepday_merged2
where row_num > 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed	row_num

 # To check if same users exist in all tables.

The tables were checked for common users. All the tables had common users having count of 33 except sleepday_merged2 with 9 non-common users.



```
WITH BASE AS
(SELECT *,
CASE WHEN id1 = id2 then 'COMMON USERS' ELSE 'NON COMMON USERS' END AS Flag
from (select distinct id as id1 from dailyactivity_merged ) a
left join (select distinct id as id2 from sleepday_merged2 ) s
 on a.id1 = s.id2)
 select flag, count(*) as user
from base
group by flag
order by flag;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Flag	user
COMMON USERS	24
NON COMMON USERS	9

The dataset called *dailyActivity_merged* contains daily calories, intensities, and steps, which made the datasets dedicated specifically to those data redundant for this analysis.

So, I used only three tables for my analysis

Dailyactivity_merged, Sleepday_merged, weightlogininfo_merged.

Data Transformation and Integration

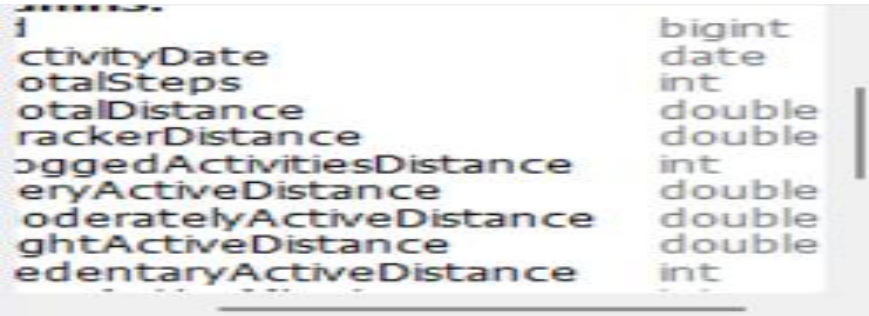
The date columns in all the tables is in Text format. I changed the data type from string to data by following steps..

```
75 •  select`ActivityDate`,
76     str_to_date(`ActivityDate`,'%m/%d/%Y') as Activitydate2
77     from dailyactivity_merged2;
78
79 •  update dailyactivity_merged2
80     set `ActivityDate` = str_to_date(`ActivityDate`,'%m/%d/%Y');
81
82 •  select *
83     from dailyactivity_merged2;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance
1503960366	2016-04-12	13162	8.5	8.5	0
1503960366	2016-04-13	10735	6.96999979	6.96999979	0
1503960366	2016-04-14	10460	6.739999771	6.739999771	0
1503960366	2016-04-15	9762	6.28000021	6.28000021	0

```
•  ALTER TABLE dailyactivity_merged2
   MODIFY COLUMN `ActivityDate` DATE;
```



Creating Combined Activity and Sleep Table:

```
CREATE TABLE DailyActivity_Sleep as
with dailyactivity_cte as( select id, activitydate, activityday2 as day_week,totalsteps,totaldistance,veryactivedistance,moderatelyactivedistar
veryactiveminutes,fairlyactiveminutes,lightlyactiveminutes,sedentaryminutes,calories,
case
    when totalsteps < 5000 then' SEDENTARY'
    when totalsteps between 5000 and 9999 then 'LIGHTLY ACTIVE'
    when totalsteps between 10000 and 12500 then 'MODERATELY ACTIVE'
    ELSE 'VERY ACTIVE'
    END AS ACTIVITYLEVEL
    from dailyactivity_merged2),
    sleepday_cte as (select id, sleepday as sleepdate,totalsleeprecords,totalminutesasleep,totaltimeinbed
    from sleepday_merged2)
    select
    A.*,B.TOTALSLEEPRECORDS,B.TOTALMINUTESASLEEP,B.TOTALTIMEINBED
    FROM Dailyactivity_cte A
    left join sleepday_cte B
    ON A.id=B.id and A.activitydate =B.Sleepdate;
```

Creating Combined Activity_ Sleep_weightTable:

```
CREATE TABLE dailyactivity_sleep_weight as
 select A.*, B.WEIGHTKG,B.BMI
 FROM dailyactivity_sleep A
LEFT JOIN weightloginfo_merged2 B ON A.id = B.id and A.activitydate = B.date;

select* from dailyactivity_sleep_weight;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

id	activitydate	day_week	totalsteps	totaldistance	veryactivedistance	moderatelyactivedistance	lightactivedistance	veryactiveminutes	fairlyactiveminutes	lightlya
1503960366	2016-04-12	TUESDAY	13162	8.5	1.879999995	0.550000012	6.059999943	25	13	328
1503960366	2016-04-13	WEDNESDAY	10735	6.96999979	1.570000052	0.689999998	4.710000038	21	19	217
1503960366	2016-04-14	THURSDAY	10460	6.739999771	2.440000057	0.400000006	3.910000086	30	11	181
1503960366	2016-04-15	FRIDAY	9762	6.28000021	2.140000105	1.25999999	2.829999924	29	34	209
1503960366	2016-04-16	SATURDAY	12669	8.159999847	2.710000038	0.409999996	5.039999962	36	10	221
1503960366	2016-04-17	SUNDAY	9705	6.480000019	3.190000057	0.779999971	2.50999999	38	20	164
1503960366	2016-04-18	MONDAY	13019	8.590000153	3.25	0.639999986	4.710000038	42	16	233
1503960366	2016-04-19	TUESDAY	15506	9.880000114	3.529999971	1.320000052	5.03000021	50	31	264
1503960366	2016-04-20	WEDNESDAY	10544	6.679999828	1.960000038	0.479999989	4.239999771	28	12	205

dailyactivity_sleep_weight 35 ×

IMPUTING NULL VALUES IN THE FINAL TABLE

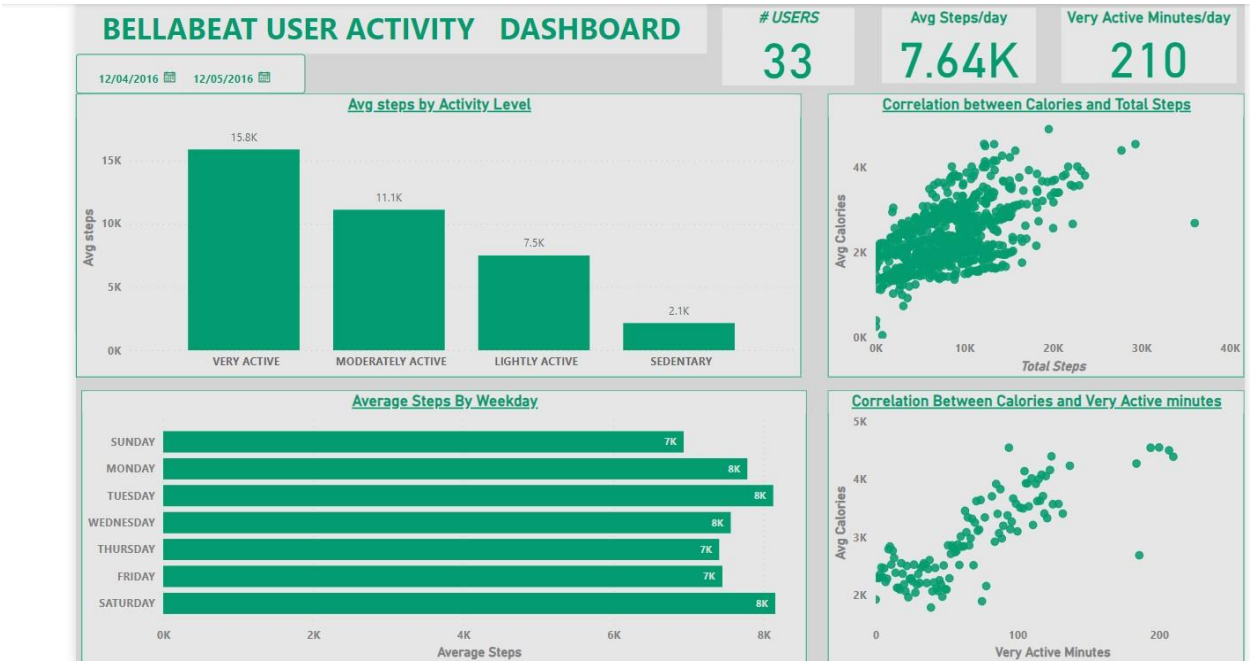
```
251 • UPDATE dailyactivity_sleep_weight
252   set totalminutesasleep = coalesce(totalminutesasleep,0);
253 • UPDATE dailyactivity_sleep_weight
254   set totaltimeinbed = coalesce(totaltimeinbed,0);
255 • UPDATE dailyactivity_sleep_weight
256   set weightkg = coalesce(weightkg,0);
257 • UPDATE dailyactivity_sleep_weight
258   set BMI = coalesce(BMI,0);
259 • UPDATE dailyactivity_sleep_weight
260   set calories = coalesce(calories,0);
261
```

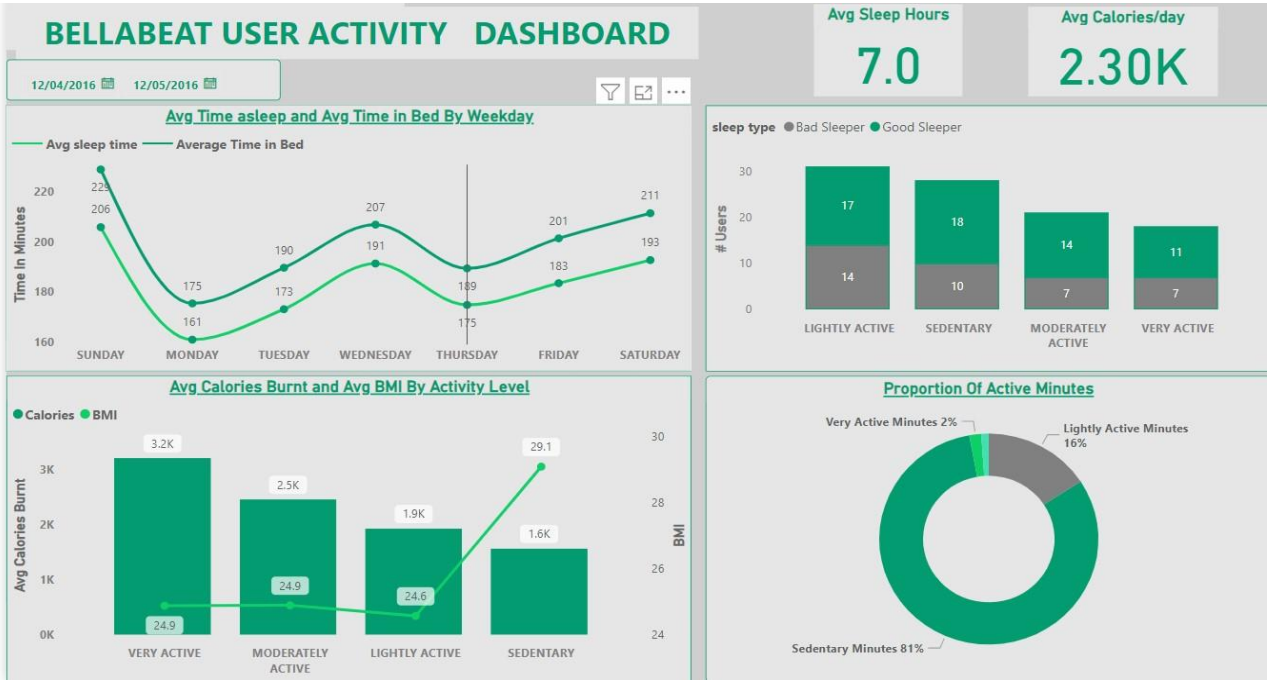
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

totalsteps	totaldistance	veryactivedistance	moderatelyactivedistance	lightactivedistance	veryactiveminutes	fairlyactiveminutes	lightlyactiveminutes	sedentaryminutes	calories
13162	8.5	1.879999995	0.550000012	6.059999943	25	13	328	728	1985
10735	6.96999979	1.570000052	0.689999998	4.710000038	21	19	217	776	1797
10460	6.739999771	2.440000057	0.400000006	3.910000086	30	11	181	1218	1776
9762	6.28000021	2.140000105	1.25999999	2.829999924	29	34	209	726	1745
12669	8.159999847	2.710000038	0.409999996	5.039999962	36	10	221	773	1863
9705	6.480000019	3.190000057	0.779999971	2.50999999	38	20	164	539	1728
13019	8.590000153	3.25	0.639999986	4.710000038	42	16	233	1149	1921
15506	9.880000114	3.529999971	1.320000052	5.03000021	50	31	264	775	2035

The final table dailyactivity_sleep_weight was exported from MYSQL as csv file and dashboard was created using PowerBI.

Insights and Analysis





CONCLUSION

After performing the collection, transformation, cleaning, organisation and analysis of the given datasets, I can infer

- Sedentary minutes took up most participants' days and were consistent throughout the week.
- Participants, on average, slept the most and took the fewest steps on Sundays.
- Participants took the most steps on Tuesdays and Saturdays.
- On average, participants slept about 7 hours per night.
- Users who were engaged in "very active minutes" burnt more calories.
- There is a direct correlation between the number of steps taken and the calories burnt.
- The healthy BMI range is between 18.5 and 24.9, overweight is 25-29.9, obesity is 30-39.9, and severe obesity is over 40. Very Active users have a BMI of 24.9 while sedentary users have BMI of 29.1.

Recommendations

Based on the analysis of the Bellabeat data, the following recommendations are made to improve user health and engagement:

1. **Encourage Reduced Sedentary Behavior:** Since sedentary minutes occupy most participants' days and are consistent throughout the week, introduce features or reminders to encourage regular movement and reduce prolonged inactivity. Implementing short activity prompts or stretching exercises could be beneficial
2. **Promote Physical Activity on Sundays:** Participants tend to sleep the most and take the fewest steps on Sundays. Encourage light physical activities or wellness challenges specifically targeted for Sundays to balance rest and activity.
3. **Leverage High Activity Days:** Participants take the most steps on Tuesdays and Saturdays. Use these insights to plan and promote community challenges, events, or new features that leverage these high activity days to boost engagement and motivation.
4. **Encourage Consistent Sleep Patterns:** With an average sleep duration of about 7 hours per night, ensure users are aware of the importance of consistent and adequate sleep. Consider incorporating sleep health tips and tracking features that promote good sleep hygiene.
5. **Highlight the Benefits of Active Minutes:** Users engaged in "very active minutes" burn more calories and have a healthier BMI. Promote the benefits of very active minutes through success stories, personalized fitness goals, and rewards for reaching specific activity milestones.
6. **Emphasize the Importance of Step Count:** There is a direct correlation between the number of steps taken and the calories burnt. Develop initiatives that focus on increasing

daily step counts, such as step challenges, virtual walk events, or integrating step tracking with other wellness apps.

7. **Tailor Health Interventions Based on BMI Insights:** The data shows a healthy BMI range between 18.5 and 24.9, with very active users having a BMI of 24.9 and sedentary users having a BMI of 29.1. Tailor health interventions and personalized coaching based on users' BMI, encouraging those in the higher BMI ranges to engage in more physical activities and adopt healthier lifestyles.

By implementing these recommendations, Bellabeat can enhance user engagement, promote healthier lifestyles, and ultimately contribute to the overall well-being of its users.