

BellaBeat

Introduction

Bellabeat is a high-tech manufacturer of health-focused products for women. It is a successful small company in the global smart device market. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits.

Business Task

- Gain insight into smart device fitness data.
- Identify trends and potential opportunities.
- provide high-level recommendations on marketing strategy for the company.

Stakeholders

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

Prepare phase

The dataset used for this analysis is: [FitBit Fitness Tracker Data](#). This is public data that explores smart device users' daily habits. It contains personal fitness tracker from thirty fitbit users for a 1 month period. Data is stored in csv files sorted and arranged as below

- Hour and minute-level output for physical activity, heart rate and sleep monitoring.
- It includes information about daily activity, steps, and heart rate.
- Weight log information.
- Heart Rate record upto second-level.

Data aligns with ROCCC quality. Thirty eligible Fitbit users consented to use the data to explore users' habits. Hence the company has taken care of privacy, security, accessibility of data and data anonymization.

Process phase

I have used a combination of spreadsheet and SQL in data processing. Importing data is simple in a spreadsheet, download the files from Kaggle and store them in a local system. Arrange the files as per our convenience and open them in sheets to view and process. For SQL, create a project and add a dataset under it. Now create tables for each csv file and name them appropriately.

Inspecting and Pre-cleaning Activities

1. Check for blank cells using conditional formatting.
2. Remove duplicate entries.
3. Check the date format. Date format updated from AM/PM to UTC format before uploading in SQL.
4. Check if any column needs to be split or merged.
5. DailyCalories, DailyIntensity and DailyStep are included in DailyActivity, thus they will not be used in further analysis.
6. Columns "TotalDistance" and "TrackerDistance" have the exact same details so these two columns are duplicate of each other and we can consider one column for analysis.
7. SedentaryActiveDistance, LightActiveDistance, ModeratelyActiveDistance, VeryActiveDistance are added up to TotalDistance columns. Hence they can be ignored.
- 8.
9. Most people use their smart device to monitor their activity level, fewer people use it to monitor sleep and even fewer people log their weight using the device.
10. Below tables the transformed version of each other, hence we consider narrow data:
 - minuteIntensitiesNarrow_merged and minuteIntensitiesWide_merged
 - minuteStepsNarrow_merged and minuteStepsWide_merged
 - minuteCaloriesNarrow_merged and minuteCaloriesWide_merged
11. In SQL, preview data and column names.
12. Checked schema of all tables to check datatypes of each column is valid.

Analyse phase

1. Find the common column name in all tables using the below query.

```
SELECT column_name, COUNT(table_name) as Table_Count
FROM `capstone-bellabeat-334213.bellabeat.INFORMATION_SCHEMA.COLUMNS` GROUP BY
column_name LIMIT 1000;
```

2. There are 33 distinct users

```
select count(DISTINCT(Id)) from
`capstone-bellabeat-334213.bellabeat.dailyActivity_merged`;
```

3. Create a table to find the correlation between BMI,Body type, Steps walked,Active and sedentary minutes spent by smart device users. Body_type column is created based on BMI value

```
select distinct(a.Id),
ROUND(avg(TotalSteps),2) AS avg_steps,
ROUND(avg(VeryActiveMinutes+FairlyActiveMinutes+LightlyActiveMinutes),2) AS
Avg_ActiveMinutes,
ROUND(avg(SedentaryMinutes),2) AS SedentaryMinutes,
ROUND(avg(Calories),2) AS avg_Calories,
ROUND(avg(BMI),2) AS avg_BMI,
CASE
    WHEN ROUND(avg(BMI),2) < 18.5 THEN 'Under Weight'
    WHEN ROUND(avg(BMI),2) >= 18.5 and ROUND(avg(BMI),2) < 25 THEN 'Normal Weight'
    WHEN ROUND(avg(BMI),2) >= 25 and ROUND(avg(BMI),2) < 30 THEN 'Over weight'
    WHEN ROUND(avg(BMI),2) >= 30 THEN 'Obese '
END AS Body_type
from
`capstone-bellabeat-334213.bellabeat.weightLogInfo_merged` as a inner join
`capstone-bellabeat-334213.bellabeat.dailyActivity_merged` as b
on a.Id=b.Id
group by Id
order by avg_BMI;
```

Output of the query is BMI_Analysis table as below:

A	B	C	D	E	F	G
Id	avg_steps	Avg_ActiveMinutes	SedentaryMinutes	avg_Calories	avg_BMI	Body_type
2873212765	7555.77	328.23	1097.19	1916.97	21.57	Normal Weight
1503960366	12116.74	277.81	848.16	1816.42	22.65	Normal Weight
6962181067	9794.81	287.13	662.32	1982.03	24.03	Normal Weight
8877689391	16040.03	310.71	1112.87	3420.26	25.49	Over weight
4558609924	7685.13	309.06	1093.61	2033.26	27.21	Over weight
4319703577	7268.84	244.68	735.81	2037.68	27.41	Over weight
5577150313	8304.43	265.1	754.43	3359.63	28	Over weight
1927972279	916.13	40.68	1317.42	2172.81	47.54	Obese

There are only 8 users who have logged their BMI and weight details, based on the table normal weight and overweight users are more than obese.

4. Create a table merge daily activity and sleep day table to get insight of how users walk and sleep throughout the week.

```

SELECT
    FORMAT_DATE("%A",ActivityDate) as Date_format,
    CAST(AVG(TotalSteps) AS INT64 ) as Avg_steps,
    FLOOR(AVG(TotalMinutesAsleep)) as Avg_sleep_mins,
    FLOOR(AVG(TotalTimeInBed-TotalMinutesAsleep)) as Avg_awake_mins,
FROM
    `capstone-bellabeat-334213.bellabeat.dailyActivity_merged` as DA inner join
    `capstone-bellabeat-334213.bellabeat.sleepDay_merged` as SD
    on DA.Id=SD.Id and FORMAT_DATE("%A",ActivityDate)=FORMAT_DATE("%A",SleepDay)
Group by Date_format
ORDER BY Date_format
LIMIT 1000;

```

Output of the query is Weekday_steps_sleep_awake table as below:

Date_format	Avg_steps	Avg_sleep_mins	Avg_awake_mins
Friday	7790	405	39
Monday	9122	417	37
Saturday	9067	421	40
Sunday	6795	453	51
Thursday	7621	403	32
Tuesday	9048	403	38
Wednesday	7752	433	35

As per above data, users are less active on Sunday and tend to sleep more.

5. Create a merged table to get details of user activities with sleep details, create a custom column to categories type of sleep.

SELECT

```

ActivityDate,
FORMAT_DATE("%A",ActivityDate) as Weekday,
CAST(AVG(TotalSteps) AS INT64 ) as Avg_steps,
CAST(AVG(TotalDistance) AS INT64 ) as Avg_Distance,
CAST(AVG(VeryActiveMinutes) AS INT64 ) as Very_Active_mins,
CAST(AVG(FairlyActiveMinutes) AS INT64 ) as Fairly_Active_mins,
CAST(AVG(LightlyActiveMinutes) AS INT64 ) as Light_Active_mins,
CAST(AVG(SedentaryMinutes) AS INT64 ) as Sedentary_Active_mins,
CAST(AVG(Calories) AS INT64 ) as Calories,
CAST(AVG(VeryActiveMinutes)+AVG(FairlyActiveMinutes)+AVG(LightlyActiveMinutes) as
INT64) as Total_active_minutes,
FLOOR(AVG(TotalMinutesAsleep)) as Avg_sleep_mins,
FLOOR(AVG(TotalTimeInBed-TotalMinutesAsleep)) as Avg_awake_mins,
CASE
when AVG(TotalMinutesAsleep)/60 <7 then "Underslept"
when AVG(TotalMinutesAsleep)/60 >=7 and AVG(TotalMinutesAsleep)/60 <8 then
"Normal"
when AVG(TotalMinutesAsleep)/60 >=8 then "Overslept"
END as sleep_type
FROM
`capstone-bellabeat-334213.bellabeat.dailyActivity_merged` as DA inner join
`capstone-bellabeat-334213.bellabeat.sleepDay_merged` as SD
on DA.Id=SD.Id and FORMAT_DATE("%F",ActivityDate)=FORMAT_DATE("%F",SleepDay)

```

```

Group by ActivityDate
ORDER BY ActivityDate
LIMIT 1000;

```

Output of the query is Date_vs_others table as below:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ActivityDate	Weekday	Avg_steps	Avg_Distance	Very_Active_mins	Fairly_Active_mins	Light_Active_mins	Sedentary_Active_mins	Calories	Total_active_minutes	Avg_sleep_mins	Avg_awake_mins	sleep_type
2	2016-04-12	Tuesday	7506	5	27	9	194	750	2439	230	441	37	Normal
3	2016-04-13	Wednesday	6103	4	18	14	164	766	2208	196	430	41	Normal
4	2016-04-14	Thursday	7626	5	28	18	194	743	2426	240	445	34	Normal
5	2016-04-15	Friday	7472	5	18	14	200	777	2317	232	427	48	Normal
6	2016-04-16	Saturday	8615	6	32	17	227	710	2494	276	391	41	Underslept
7	2016-04-17	Sunday	6530	5	22	17	183	707	2252	222	464	45	Normal
8	2016-04-18	Monday	8611	6	31	24	190	743	2497	245	419	36	Underslept
9	2016-04-19	Tuesday	9902	7	36	22	227	706	2488	285	409	42	Underslept
10	2016-04-20	Wednesday	9148	7	31	18	210	720	2455	259	446	30	Normal
11	2016-04-21	Thursday	9698	7	29	19	208	791	2579	257	376	33	Underslept
12	2016-04-22	Friday	8377	6	31	19	222	739	2456	273	392	34	Underslept
13	2016-04-23	Saturday	11377	8	25	24	290	646	2549	338	430	33	Normal
14	2016-04-24	Sunday	7995	6	24	21	205	632	2310	250	487	48	Overslept
15	2016-04-25	Monday	9948	7	46	16	218	728	2543	280	395	36	Underslept
16	2016-04-26	Tuesday	9290	6	27	23	216	780	2444	266	369	31	Underslept
17	2016-04-27	Wednesday	8447	6	16	15	237	707	2391	267	427	43	Normal

This table shows that most of the users either undersleep or sleep normal hours.

- New table is created with the inner join of hourly steps and hourly intensity tables, also creating a new column to determine time of the day for activities or walking.

```

SELECT
    S.Id,
    CASE
        WHEN EXTRACT(HOUR from S.ActivityHour) >=6 AND EXTRACT(HOUR from S.ActivityHour)
<=12 THEN "Morning"
        WHEN EXTRACT(HOUR from S.ActivityHour) >12 AND EXTRACT(HOUR from S.ActivityHour)
<=4 THEN "Afternoon"
        WHEN EXTRACT(HOUR from S.ActivityHour) >4 AND EXTRACT(HOUR from S.ActivityHour)
<=8 THEN "Evening"
        ELSE "Night"
    END as Hours,
    FLOOR(Avg(StepTotal)) as TotalSteps,
    FLOOR(Avg(TotalIntensity)) as TotalIntensity

```

```

FROM
    `capstone-bellabeat-334213.bellabeat.hourlySteps_merged` as S inner join
    `capstone-bellabeat-334213.bellabeat.hourlyIntensities_merged` as I
    on S.Id=I.Id and S.ActivityHour=I.ActivityHour
group by
    Id,Hours
LIMIT 1000;

```

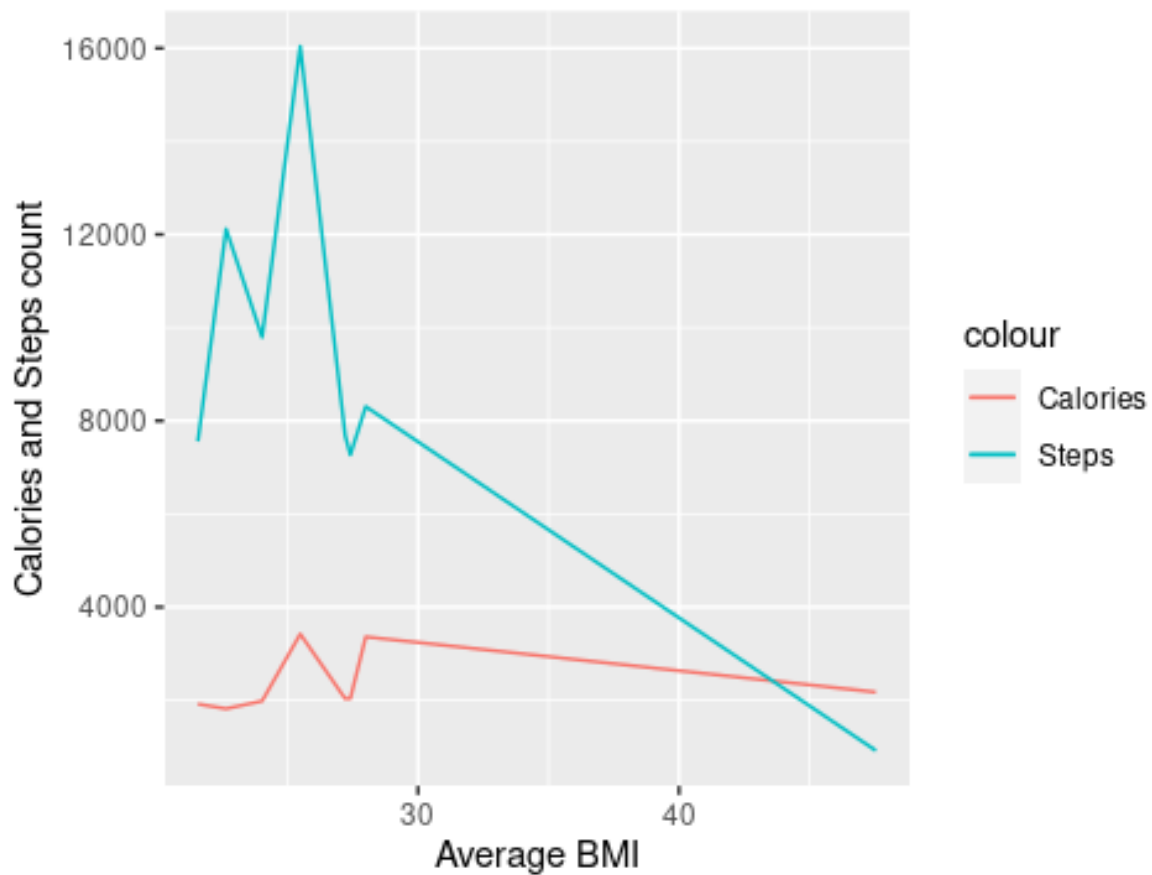
Output of the query is Steps_Intensity_hourly table as below

Id	Hours	TotalSteps	TotalIntensity
8053475328	Night	903	25
8053475328	Evening	0	0
8053475328	Morning	81	3
1644430081	Night	379	12
1644430081	Evening	0	0
1644430081	Morning	190	6
4558609924	Night	352	15
4558609924	Evening	0	0
4558609924	Morning	303	14
4319703577	Night	248	10
4319703577	Evening	68	2
4319703577	Morning	415	14
2320127002	Night	200	8

Share Phase

BMI vs Steps and Calories

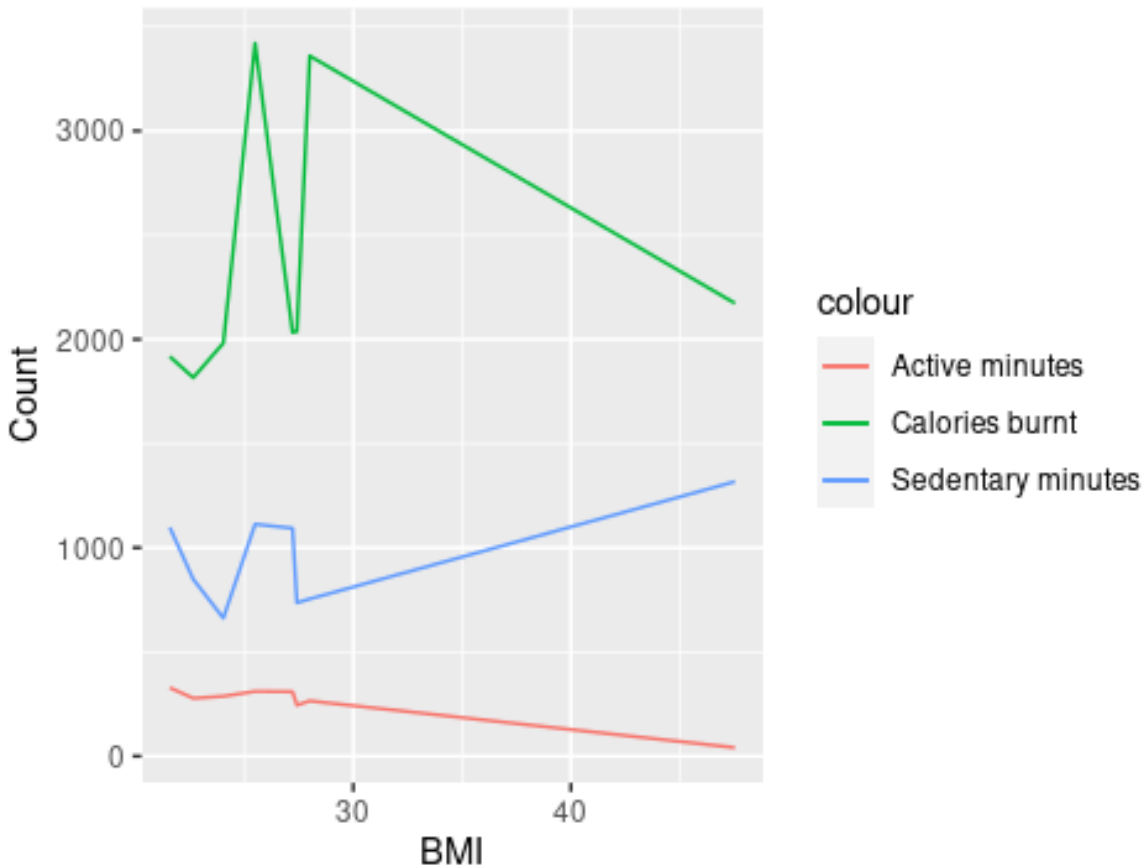
Using the BMI_Analysis table, when we plot BMI values against calories burnt, average steps walked by users. We notice calories burnt have negative correlation with BMI whereas sedentary minutes have positive correlation.



**Plot has Average values of all users

BMI vs Calories burnt, active and sedentary minutes

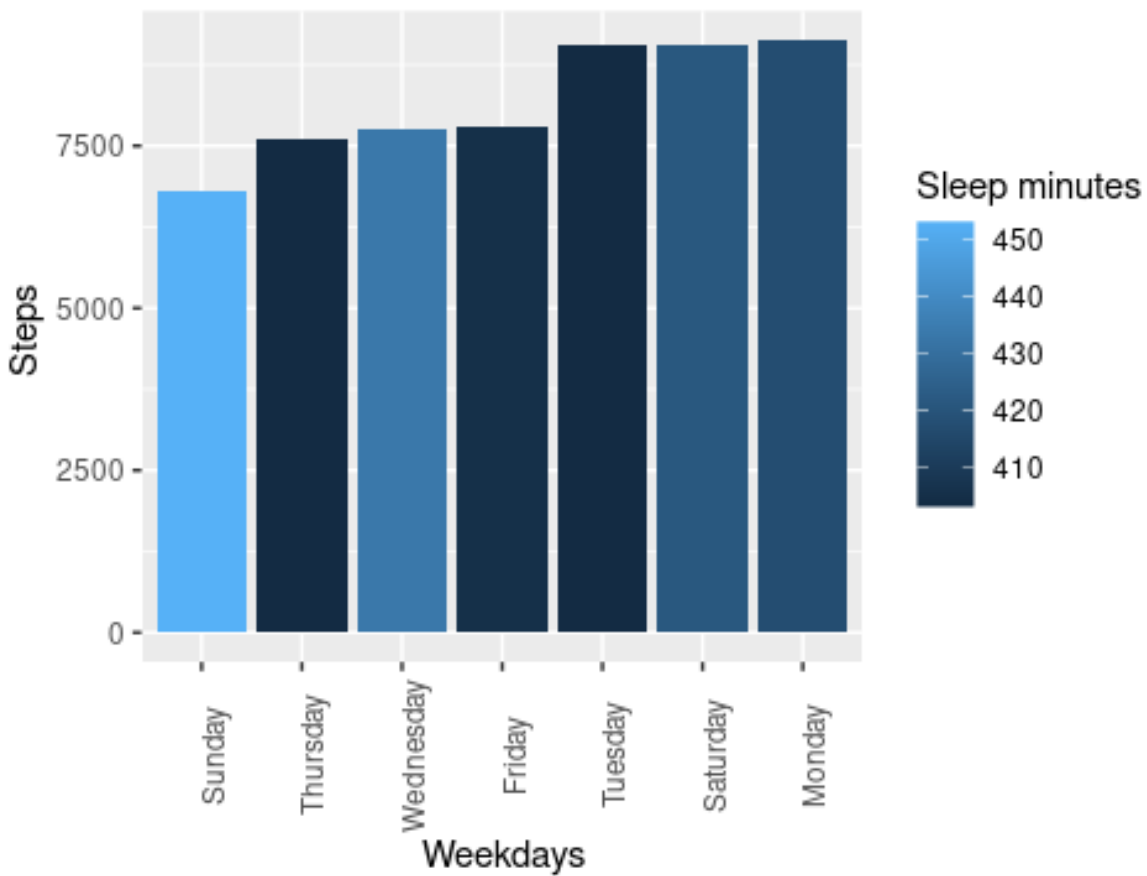
Using BMI_Analysis table, when we plot BMI values against Calories burnt, active minutes and sedentary minutes. We notice calories burnt and active minutes have negative correlation with BMI whereas sedentary minutes have positive correlation.



**Plot has Average values of all users

Average steps, sleep time vs days of week

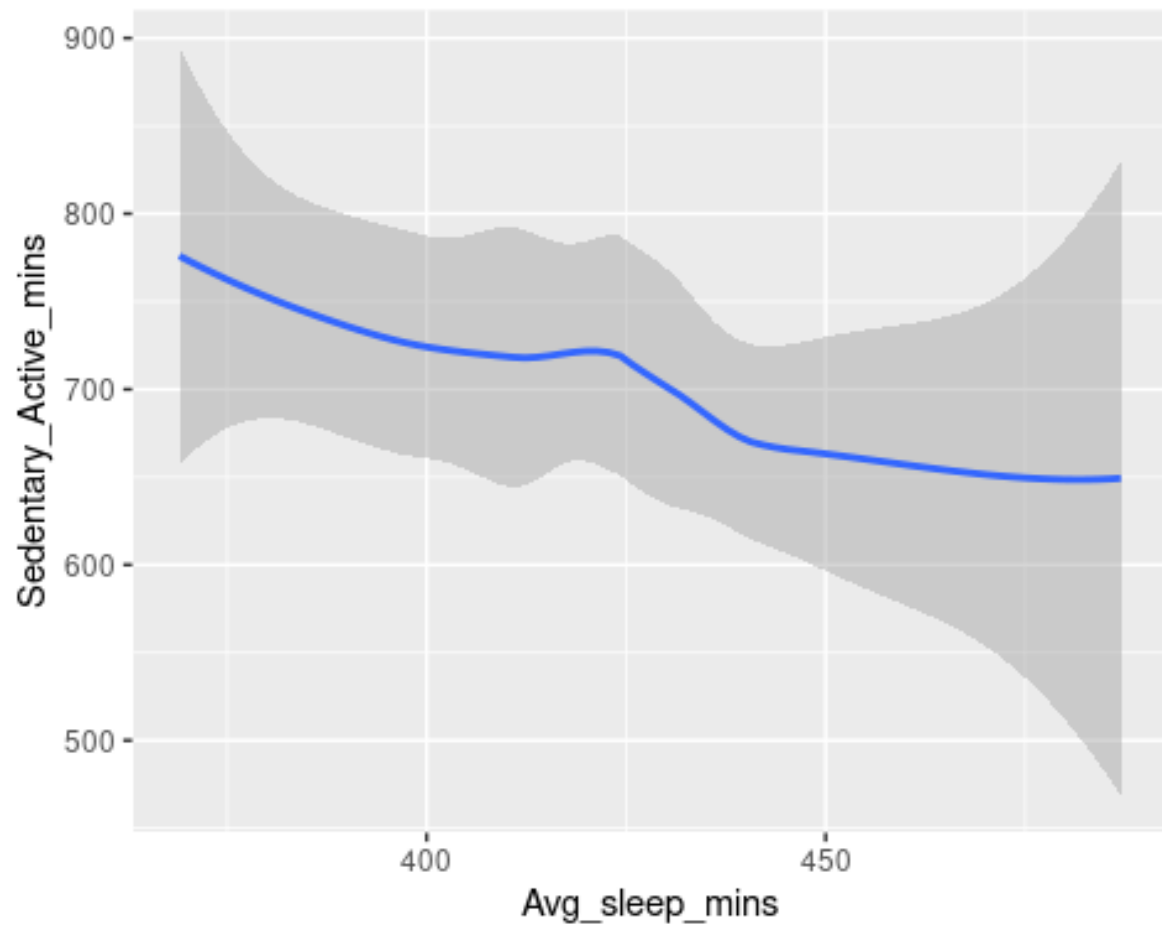
Weekday_steps_sleep_awake table data is plotted in the below graph. This shows how users like to spend their day of the week in terms of walk or sleep. We can notice most users are lazy on Sunday with more sleep time and less number of steps walked.



**Plot has Average values of all users

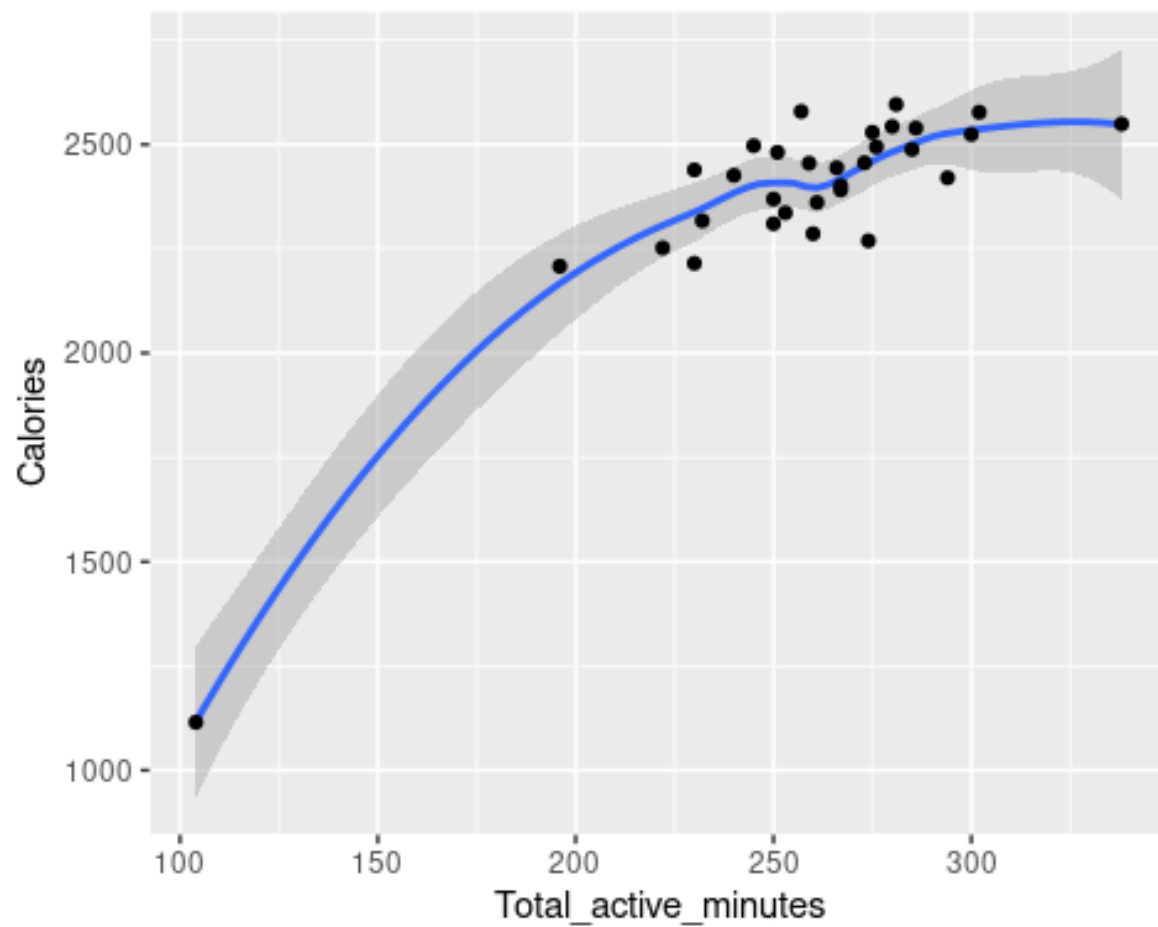
Average sleep time vs Sedentary minutes

Date_vs_others gives a clear picture on the impact of sedentary minutes on sleep time. We find that these two are negatively correlated. As we tend to spend more time on sitting, inactive or without exercise we won't have good sleep.



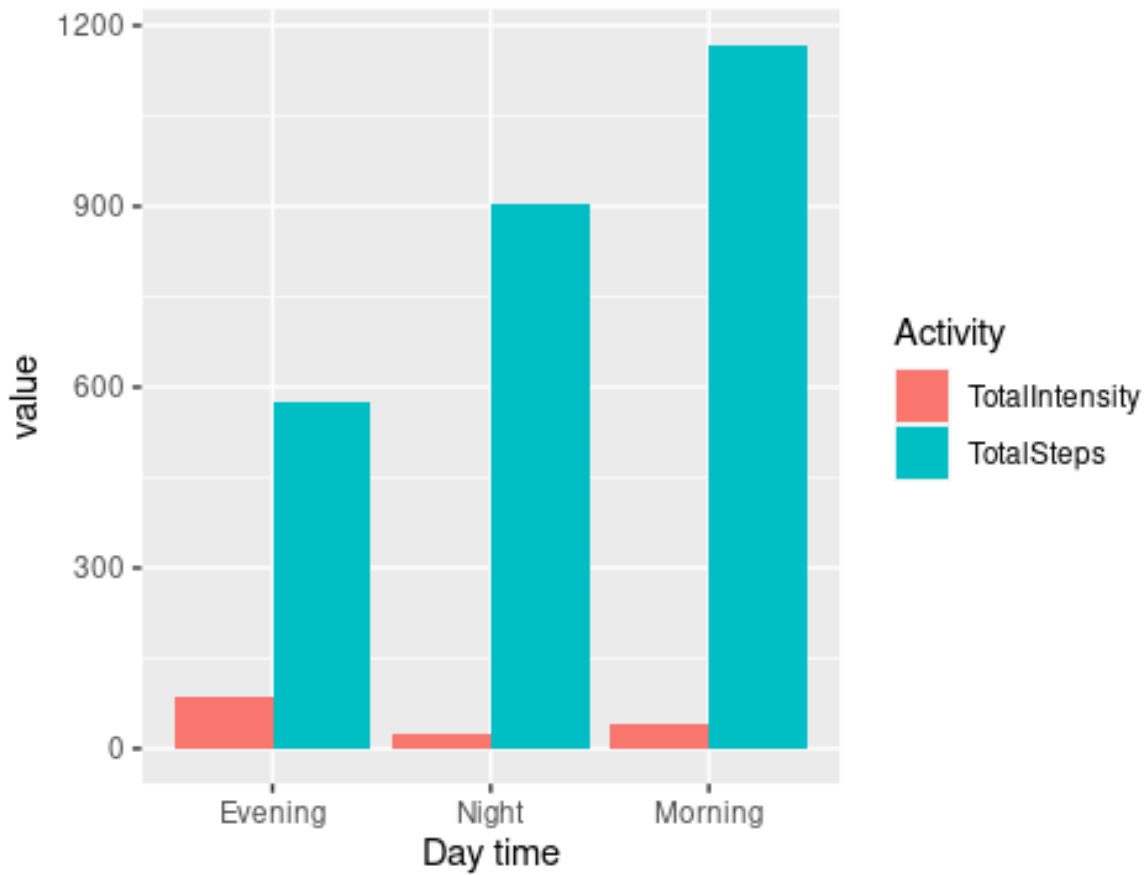
Calories burnt vs Active minutes

Using Date_vs_others we can determine there is a strong correlation between total active minutes and calories burnt by users.



Activity time in a day

Custom create table Steps_Intensity_hourly is used to get plotting of AVerage steps walked and average intensity activities performed by fitness device users during different times of the day.



**Average values taken for intensity and steps

Limitation

Since data has a small sample set of 33 distinct users, it gives us scope to further investigate why participants did not record daily sleep and weight log as compared to daily activity.

This dataset consists of records noted for only one month. For a more accurate analysis, we must have a bigger dataset.

Conclusion and Recommendation

- Based on our analysis, we notice only 24 % of users log their weight and BMI details. We must encourage more users to enter the same regularly.
- Only 72% of people tracked their sleep on a daily basis. Marketing team should make the customers aware of the benefits of tracking their sleep.
- Most users tend to get lazy on Sunday so Bellabeat can organize sunday-funday walking or running competitions among the users.
- People looking to burn more calories should spend more time being active rather than sedentary. This will lead to burning more calories and getting good sleep.
- Many users like to walk in the morning and do high intensity activities in the evening or night. Keep a weekly target of walk and high intensity activities, upon reaching refer users for free trial of bellabeat premium membership.
- Collaborate with local gyms for trial sessions for regular premium members. This can be used in advertisements.