FITBIT FITNESS TRACKER DATA

Prepared by Gizelle Felix 12.02.24-12.11.2024

BUSINESS OBJECTIVE: To focus on one of Bellabeat's products and analyze smart devices data to gain insight into how consumers are using their smart devices.

ASK

Pre-Questionnaire:

- Why are customers using the devices?

- How can these trends apply to Bellabeat consumers?
- What are the trends in smart device usage?
- How could these trends help influence Bellabeat Marketing strategy?

PREPARE

Source: https://www.kaggle.com/datasets/arashnic/fitbit

"This dataset was generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences"

To prepare for the analysis, I exported the Bellabeat dataset from Kaggle to a spreadsheet to determine what kind of information I will be working with. I previewed each dataset to see what kind of data I am working with and what kind of tools I will be using. Tools used for this case study was Kaggle, Google Sheets, Google BigQuery, and Tableau.

PROCESS

Following datasets were exported into a spreadsheet and BigQuery:
dailyActivity_merged.csv
weightLogInfo_merged.csv
sleepDay_merged.csv

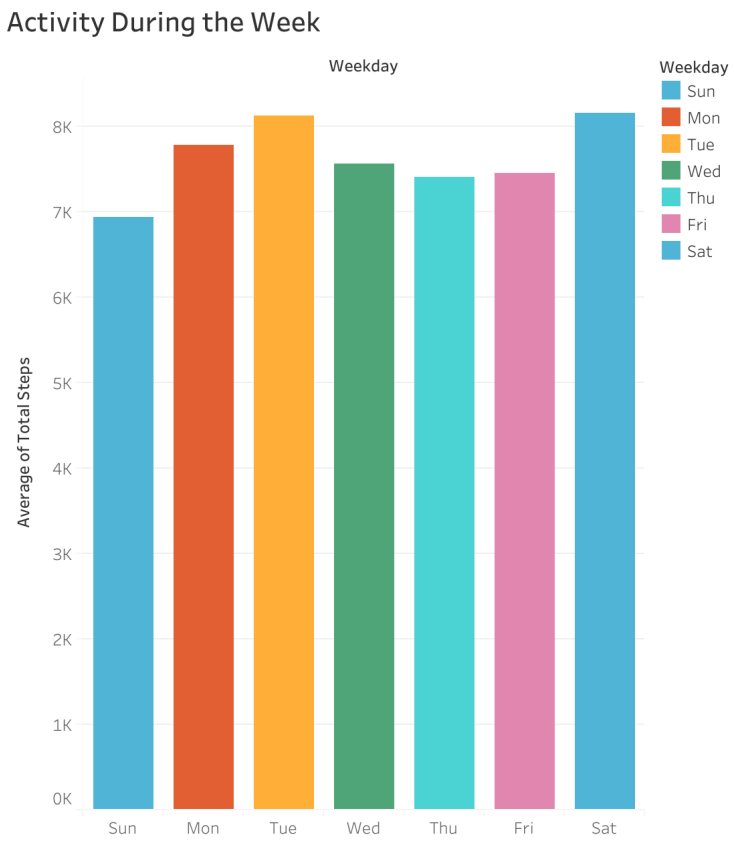
I examined by sorting and filtering for a better understanding of the datasets above. Other preparations included changing formats of the data, looking for inconsistencies, and incomplete data. Then switched from spreadsheet to SQL (BigQuery) after further examinations of the dataset from the spreadsheet.

ANALYZE

Activity during the weekdays

Importing the dailyActivity_merge.csv file on a spreadsheet. I used the function "(=TEXT(date),"ddd")" to figure out the days that the dates fall on. I created a pivot table averaging the Total Steps of all data to see the amount of activity used on a daily for the month 04.12-05.12.2016.

Weekday	AVERAGE of TotalSteps
Fri	7448.230159
Mon	7780.866667
Sat	8152.975806
Sun	6933.231405
Thu	7405.836735
Tue	8125.006579
Wed	7559.373333



(Image 1)

Narrowing my sample size

I decided to focus my analysis on users that have a log of >30 from the dailyActivity because I didn't want to have missing/null values

Determine the amount of usage or logs per ID

```
SELECT COUNT(Id) AS AmtofLogs, Id
FROM `projects-443602.bellabeat_data.dailyActivity`
GROUP BY Id
```

AmtofLogs	Id
31	1624580081
30	1644430081
31	2022484408
18	2347167796
30	3977333714
31	4319703577
31	4388161847
31	4702921684
30	5577150313
26	6775888955
31	6962181067
26	7007744171
31	7086361926
19	8253242879
31	8583815059
29	8792009665
31	1844505072
31	1927972279
31	2026352035
31	2320127002
31	2873212765
20	3372868164
31	4020332650
4	4057192912
31	4445114986
31	4558609924
31	5553957443
28	6117666160
29	6290855005
31	8053475328
31	8378563200
31	8877689391
31	1503960366

Based on the table above, there are users that don't use the devices daily for the time range of 4.12.16-5.12.16. I narrowed it down but also included values that I need to start my analysis, query as follows:

```
SELECT Id,
 AVG(TotalSteps) AS AvgSteps,
 AVG(TotalDistance) As AvgDistance,
 COUNT(Id) AS LogCount, AVG(Calories) AS AvgCalories
FROM `projects-443602.bellabeat_data.dailyActivity`
GROUP BY Id
HAVING
COUNT(Id) > 30
ORDER BY AVG(TotalSteps)
```

Id	AvgSteps	AvgDistance	LogCount	AvgCalories
1927972279	916.1290322580644	0.63451612308140759	31	2172.8064516129034
4020332650	2267.2258064516122	1.6261290389323431	31	2385.8064516129034
1844505072	2580.0645161290327	1.7061290368437776	31	1573.4838709677417
2320127002	4716.8709677419338	3.1877419044894566	31	1724.1612903225805
4445114986	4796.5483870967746	3.2458064402303384	31	2186.1935483870966
2026352035	5566.8709677419347	3.454838715253338	31	1540.6451612903224
1624580081	5743.9032258064517	3.9148387293661782	31	1483.3548387096776
8583815059	7198.5161290322576	5.6154838223611181	31	2732.0322580645166
4319703577	7268.8387096774195	4.8922580470361039	31	2037.6774193548383
2873212765	7555.7741935483873	5.1016128601566448	31	1916.9677419354837
4558609924	7685.1290322580644	5.0806451766721672	31	2033.258064516129
4702921684	8572.0645161290322	6.9551612830931147	31	2965.5483870967741
5553957443	8612.58064516129	5.6396774495801605	31	1875.6774193548385
8378563200	8717.709677419356	6.9135484618525309	31	3436.5806451612907
7086361926	9371.7741935483828	6.3880645078156268	31	2566.354838709678
6962181067	9794.8064516129034	6.585806477454403	31	1982.0322580645163
4388161847	10813.935483870966	8.39322589289757	31	3093.8709677419356
2022484408	11370.645161290324	8.0841934911666353	31	2509.9677419354839
1503960366	12116.741935483873	7.8096773855147843	31	1816.4193548387095
8053475328	14763.290322580646	11.475161198646791	31	2945.8064516129039

8877689391 16040.032258064515 13.212903138129944 31 3420.2580645161288

Total Users: 21

I exported this table into a new sheet in the spreadsheet.

Within the 21 users each were categorized based on their daily activities as follows:

INACTIVE <5,000 steps daily

LIGHT 5,000-7,455 steps daily
MODERATE 7,500-9,999 steps daily
ACTIVE >10,000 steps daily

According to Pallavi Suyog Uttekar,MD
<https://www.medicinenet.com/how_many_steps_a_day_is_considered_active/article.htm>

'Inactive Active Users'

```
SELECT Id, AvgSteps, AvgCalories
FROM `projects-443602.bellabeat_data.AverageActivity`
WHERE
AvgSteps < 5000
```

Id	AvgSteps	AvgCalories
1927972279	916.1290322580644	2172.8064516129034
4020332650	2267.2258064516122	2385.8064516129034
1844505072	2580.0645161290327	1573.4838709677417
2320127002	4716.8709677419338	1724.1612903225805
4445114986	4796.5483870967746	2186.1935483870966

'Lightly Active Users'

```
SELECT Id, AvgSteps, AvgCalories
FROM `projects-443602.bellabeat_data.AverageActivity`
WHERE
 AvgSteps BETWEEN 5000 AND 7455
```

Id	AvgSteps	AvgCalories
2026352035	5566.8709677419347	1540.6451612903224

1624580081	5743.9032258064517	1483.3548387096776
8583815059	7198.5161290322576	2732.0322580645166
4319703577	7268.8387096774195	2037.6774193548383

'Moderately Active Users'

```
SELECT Id, AvgSteps, AvgCalories
FROM `projects-443602.bellabeat_data.AverageActivity`
WHERE
 AvgSteps BETWEEN 7500 AND 9999
```

Id	AvgSteps	AvgCalories
2873212765	7555.7741935483873	1916.9677419354837
4558609924	7685.1290322580644	2033.258064516129
4702921684	8572.0645161290322	2965.5483870967741
5553957443	8612.58064516129	1875.6774193548385
8378563200	8717.709677419356	3436.5806451612907
7086361926	9371.7741935483828	2566.354838709678
6962181067	9794.8064516129034	1982.0322580645163

'Highly Active Users'

```
SELECT Id, AvgSteps, AvgCalories
FROM `projects-443602.bellabeat_data.AverageActivity`
WHERE
 AvgSteps > 10000
```

Id	AvgSteps	AvgCalories
4388161847	10813.935483870966	3093.8709677419356
2022484408	11370.645161290324	2509.9677419354839
1503960366	12116.741935483873	1816.4193548387095
8053475328	14763.290322580646	2945.8064516129039
8877689391	16040.032258064515	3420.2580645161288

Used these 21 users info to develop an analysis.

Activity and Calories

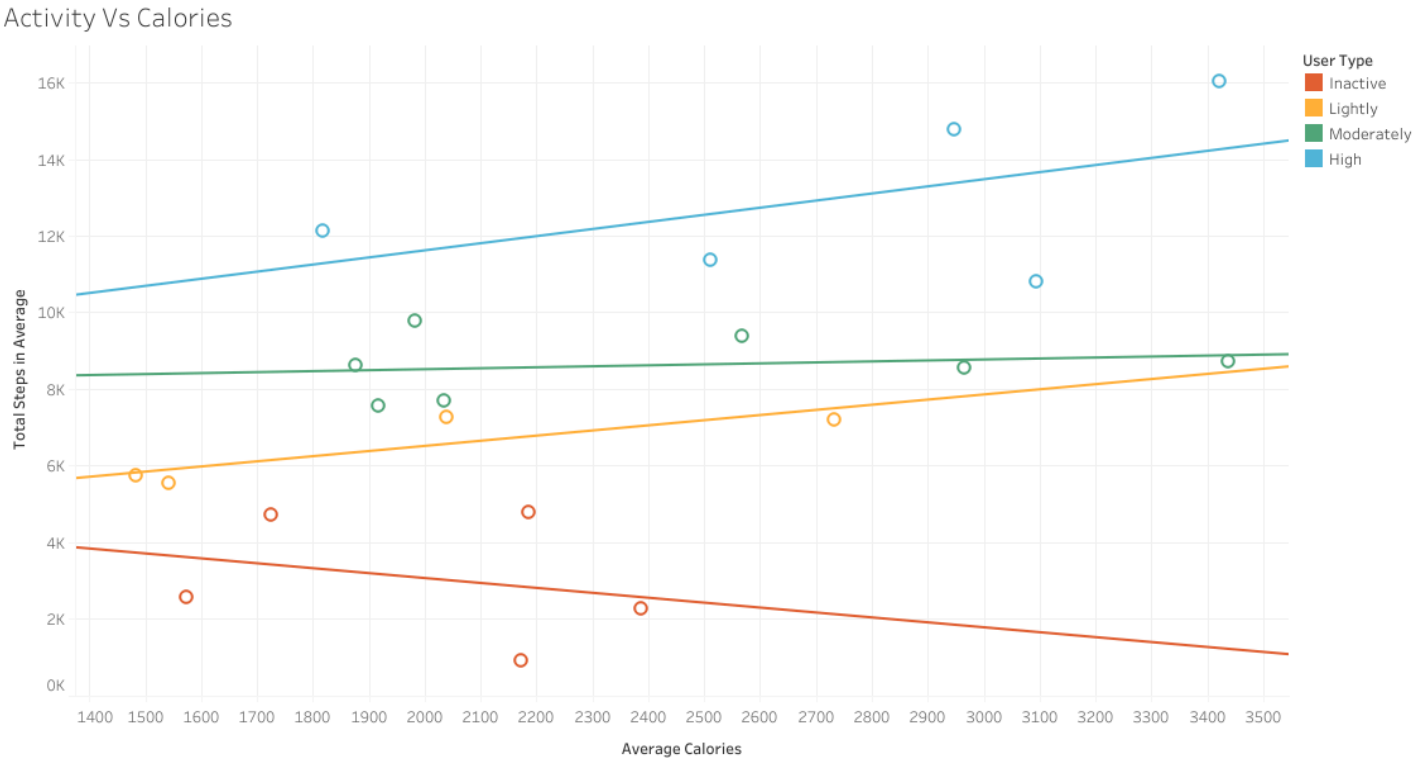
Find out if there's a correlation between Calories and Total Steps

Started with the query

```
SELECT
 Id,
 COUNT(Id) AS AmtofLogs,
 AVG(TotalSteps) AS AvgTotalSteps,
 AVG(TotalDistance) AS AvgTotalDistance,
 AVG(Calories) AS AvgCalories
FROM `projects-443602.bellabeat_data.dailyActivity`
GROUP BY Id
HAVING COUNT(Id) > 30;
```

Id	AmtofLogs	AvgTotalSteps	AvgTotalDistance	AvgCalories
1624580081	31	5743.9032258064517	3.9148387293661782	1483.3548387096776
2022484408	31	11370.645161290324	8.0841934911666353	2509.9677419354839
4319703577	31	7268.8387096774195	4.8922580470361039	2037.6774193548383
4388161847	31	10813.935483870966	8.39322589289757	3093.8709677419356
4702921684	31	8572.0645161290322	6.9551612830931147	2965.5483870967741
6962181067	31	9794.8064516129034	6.585806477454403	1982.0322580645163
7086361926	31	9371.7741935483828	6.3880645078156268	2566.354838709678
8583815059	31	7198.5161290322576	5.6154838223611181	2732.0322580645166
1844505072	31	2580.0645161290327	1.7061290368437776	1573.4838709677417
1927972279	31	916.1290322580644	0.63451612308140759	2172.8064516129034
2026352035	31	5566.8709677419347	3.454838715253338	1540.6451612903224
2320127002	31	4716.8709677419338	3.1877419044894566	1724.1612903225805
2873212765	31	7555.7741935483873	5.1016128601566448	1916.9677419354837
4020332650	31	2267.2258064516122	1.6261290389323431	2385.8064516129034
4445114986	31	4796.5483870967746	3.2458064402303384	2186.1935483870966
4558609924	31	7685.1290322580644	5.0806451766721672	2033.258064516129
5553957443	31	8612.58064516129	5.6396774495801605	1875.6774193548385
8053475328	31	14763.290322580646	11.475161198646791	2945.8064516129039

8378563200	31	8717.709677419356	6.9135484618525309	3436.5806451612907
8877689391	31	16040.032258064515	13.212903138129944	3420.2580645161288
1503960366	31	12116.741935483873	7.8096773855147843	1816.4193548387095



(Image 2)

It can be seen that there's an effect on steps on calories. Based on the average of the calories of each 21 users, the more steps the users input, the higher the calories they burn.

Sleep and Activity

Find the relationship between the amount of time it takes to sleep with users that are inactive, light, moderate, and highly active.

'Inactive Active Users'

```
SELECT
 AverageActivity.Id,
 AverageActivity.AvgSteps,
 sleepDay.TotalHoursAsleep,
```

```
 sleepDay.TotalHoursinBed,
 sleepDay.SleepDay AS Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
INNER JOIN
`projects-443602.bellabeat_data.sleepDay` AS sleepDay
ON
 AverageActivity.Id = sleepDay.Id
WHERE
 AverageActivity.AvgSteps < 5000
ORDER BY sleepDay.SleepDay
```

'Lightly Active Users'

```
SELECT
 AverageActivity.Id,
 AverageActivity.AvgSteps,
 sleepDay.TotalHoursAsleep,
 sleepDay.TotalHoursinBed,
 sleepDay.SleepDay AS Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
INNER JOIN
`projects-443602.bellabeat_data.sleepDay` AS sleepDay
ON
 AverageActivity.Id = sleepDay.Id
WHERE
 AverageActivity.AvgSteps BETWEEN 5000 AND 7455
ORDER BY sleepDay.SleepDay
```

'Moderately Active Users'

```
SELECT
 AverageActivity.Id,
 AverageActivity.AvgSteps,
 sleepDay.TotalHoursAsleep,
 sleepDay.TotalHoursinBed,
```

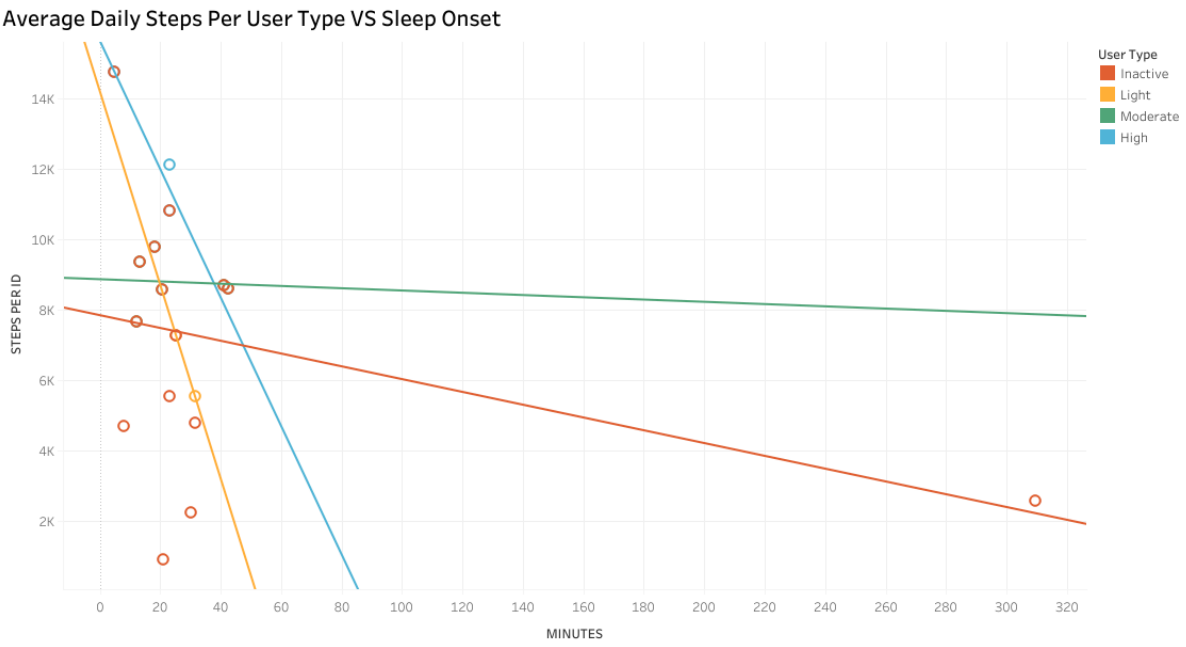
```
 sleepDay.SleepDay AS Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
INNER JOIN
`projects-443602.bellabeat_data.sleepDay` AS sleepDay
ON
 AverageActivity.Id = sleepDay.Id
WHERE
 AverageActivity.AvgSteps BETWEEN 7500 AND 9999
ORDER BY sleepDay.SleepDay
```

‘Highly Active Users’

```
SELECT
 AverageActivity.Id,
 AverageActivity.AvgSteps,
 sleepDay.TotalHoursAsleep,
 sleepDay.TotalHoursinBed,
 sleepDay.SleepDay AS Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
INNER JOIN
`projects-443602.bellabeat_data.sleepDay` AS sleepDay
ON
 AverageActivity.Id = sleepDay.Id
WHERE
 AverageActivity.AvgSteps > 10000
ORDER BY sleepDay.SleepDay
```

After gathering each table, I exported the tables into a spreadsheet. From the SleepDay data. I converted the TotalMinutesAsleep and TotalMinutesinBed into hours for easier understanding. Then labeled it as TotalHoursAsleep and TotalHoursin Bed. Proceeded with the function TotalHoursinBed - TotalHoursAsleep = MinutesSleepOnset (or ToSleepinMinutes in the table below)

Id	AVE of ToSleepinMinutes	AVE of TotalSteps	User Type
1503960366	22.872	12116.74194	High
4388161847	22.9826087	10813.93548	High
8053475328	4.8	14763.29032	High
1844505072	309.4	2580.064516	Inactive
1927972279	20.88	916.1290323	Inactive
2026352035	23.025	5566.870968	Inactive
2320127002	7.8	4716.870968	Inactive
4020332650	30.15	2267.225806	Inactive
4319703577	25.24615385	7268.83871	Inactive
4388161847	22.9826087	10813.93548	Inactive
4445114986	31.60714286	4796.548387	Inactive
4558609924	12.24	7685.129032	Inactive
4702921684	20.68888889	8572.064516	Inactive
5553957443	42.27096774	8612.580645	Inactive
6962181067	18.09677419	9794.806452	Inactive
7086361926	13.25	9371.774194	Inactive
8053475328	4.8	14763.29032	Inactive
8378563200	40.85806452	8717.709677	Inactive
2026352035	31.43571429	5566.870968	Light
4319703577	25.24615385	7268.83871	Light
4558609924	12.24	7685.129032	Moderate
4702921684	20.68888889	8572.064516	Moderate
5553957443	42.27096774	8612.580645	Moderate
6962181067	18.09677419	9794.806452	Moderate
7086361926	13.25	9371.774194	Moderate
8378563200	40.85806452	8717.709677	Moderate



(Image 3)

Above image shows each user type and the amount of time it takes for the users to fall asleep. Image shows that the highest activity users are able to fall asleep faster than the inactive users. It can be seen that a user from the inactive can take up to 300-325 minutes on average to fall asleep.

Activity and Weight(Kg)

Find out if the amount of activity per user has an effect with weight (Kg).

Following queries were inputted into BigQuery:

'Inactive Active Users'

```
SELECT
AverageActivity.Id,
AverageActivity.AvgSteps,
weightLogInfo.WeightKg,
weightLoginfo.Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
FULL JOIN
`projects-443602.bellabeat_data.weightLogInfo` AS weightLoginfo
ON
```

```
 AverageActivity.Id = weightloginfo.Id
WHERE
AverageActivity.AvgSteps < 5000
ORDER BY weightLoginfo.Date, weightLoginfo.Date
```

‘Lightly Active Users’

```
SELECT
AverageActivity.Id,
AverageActivity.AvgSteps,
weightLogInfo.WeightKg,
weightLoginfo.Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
FULL JOIN
`projects-443602.bellabeat_data.weightLogInfo` AS weightLoginfo
ON
 AverageActivity.Id = weightloginfo.Id
WHERE
AverageActivity.AvgSteps BETWEEN 5000 AND 7455
ORDER BY weightLoginfo.Id, weightLoginfo.Date
```

‘Moderately Active Users’

```
SELECT
AverageActivity.Id,
AverageActivity.AvgSteps,
weightLogInfo.WeightKg,
weightLoginfo.Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
FULL JOIN
`projects-443602.bellabeat_data.weightLogInfo` AS weightLoginfo
ON
 AverageActivity.Id = weightloginfo.Id
WHERE
AverageActivity.AvgSteps BETWEEN 7500 AND 9999
```

```
ORDER BY weightLoginfo.Date, weightLoginfo.Date
```

'Highly Active Users'

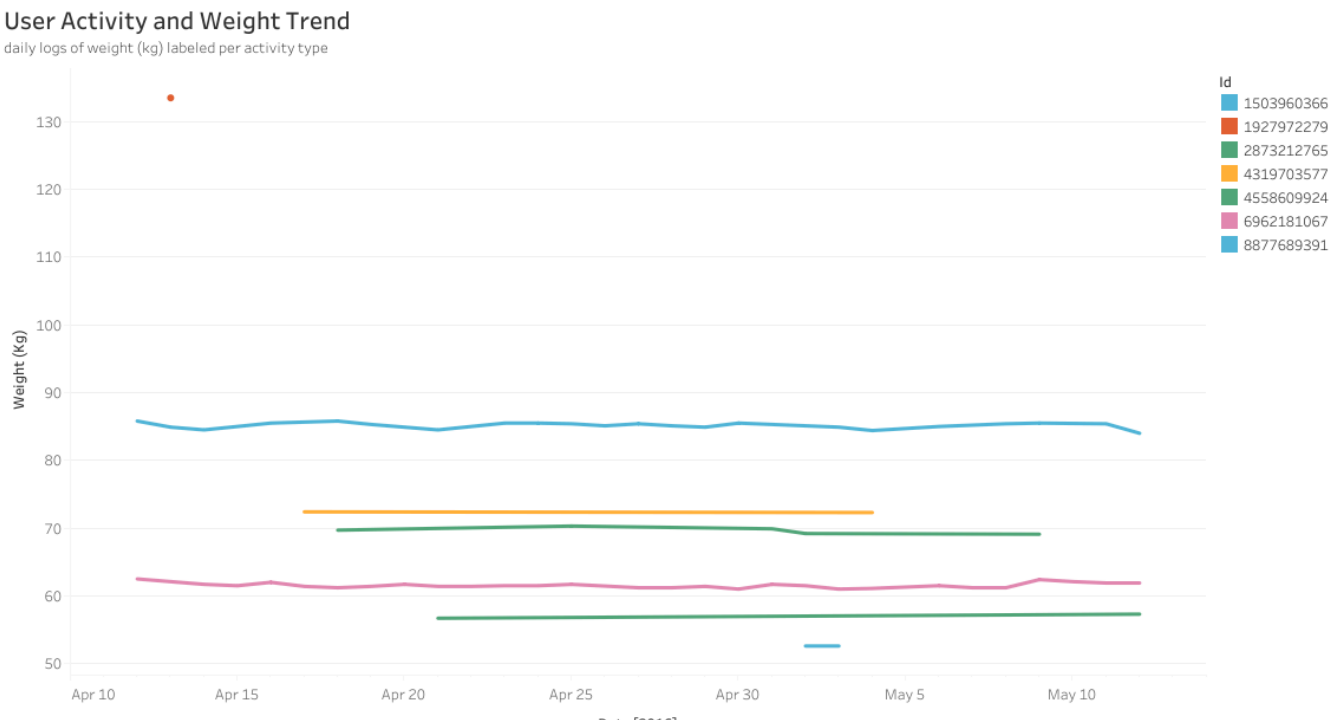
```
SELECT
AverageActivity.Id,
AverageActivity.AvgSteps,
weightLogInfo.WeightKg,
weightLoginfo.Date
FROM
`projects-443602.bellabeat_data.AverageActivity` AS AverageActivity
FULL JOIN
`projects-443602.bellabeat_data.weightLogInfo` AS weightLoginfo
ON
 AverageActivity.Id = weightloginfo.Id
WHERE
AverageActivity.AvgSteps > 10000
ORDER BY weightLoginfo.Id, weightLoginfo.Date
```

I exported the tables from these queries to a spreadsheet and created a pivot table.

Id	Date	Ave. of AvgSteps	Ave. of WeightKg
1503960366	2016-05-02	12116.74194	52.6
1503960366	2016-05-03	12116.74194	52.6
1927972279	2016-04-13	916.1290323	133.5
2873212765	2016-04-21	7555.774194	56.7
2873212765	2016-05-12	7555.774194	57.3
4319703577	2016-04-17	7268.83871	72.4
4319703577	2016-05-04	7268.83871	72.3
4558609924	2016-04-18	7685.129032	69.7
4558609924	2016-04-25	7685.129032	70.3
4558609924	2016-05-01	7685.129032	69.9
4558609924	2016-05-02	7685.129032	69.2
4558609924	2016-05-09	7685.129032	69.1
6962181067	2016-04-12	9794.806452	62.5
6962181067	2016-04-13	9794.806452	62.1

6962181067	2016-04-14	9794.806452	61.7
6962181067	2016-04-15	9794.806452	61.5
6962181067	2016-04-16	9794.806452	62
6962181067	2016-04-17	9794.806452	61.4
6962181067	2016-04-18	9794.806452	61.2
6962181067	2016-04-19	9794.806452	61.4
6962181067	2016-04-20	9794.806452	61.7
6962181067	2016-04-21	9794.806452	61.4
6962181067	2016-04-22	9794.806452	61.4
6962181067	2016-04-23	9794.806452	61.5
6962181067	2016-04-24	9794.806452	61.5
6962181067	2016-04-25	9794.806452	61.7
6962181067	2016-04-27	9794.806452	61.2
6962181067	2016-04-28	9794.806452	61.2
6962181067	2016-04-29	9794.806452	61.4
6962181067	2016-04-30	9794.806452	61
6962181067	2016-05-01	9794.806452	61.7
6962181067	2016-05-02	9794.806452	61.5
6962181067	2016-05-03	9794.806452	61
6962181067	2016-05-04	9794.806452	61.1
6962181067	2016-05-05	9794.806452	61.3
6962181067	2016-05-06	9794.806452	61.5
6962181067	2016-05-07	9794.806452	61.2
6962181067	2016-05-08	9794.806452	61.2
6962181067	2016-05-09	9794.806452	62.4
6962181067	2016-05-10	9794.806452	62.1
6962181067	2016-05-11	9794.806452	61.9
6962181067	2016-05-12	9794.806452	61.9
8877689391	2016-04-12	16040.03226	85.8
8877689391	2016-04-13	16040.03226	84.9
8877689391	2016-04-14	16040.03226	84.5
8877689391	2016-04-16	16040.03226	85.5
8877689391	2016-04-18	16040.03226	85.8

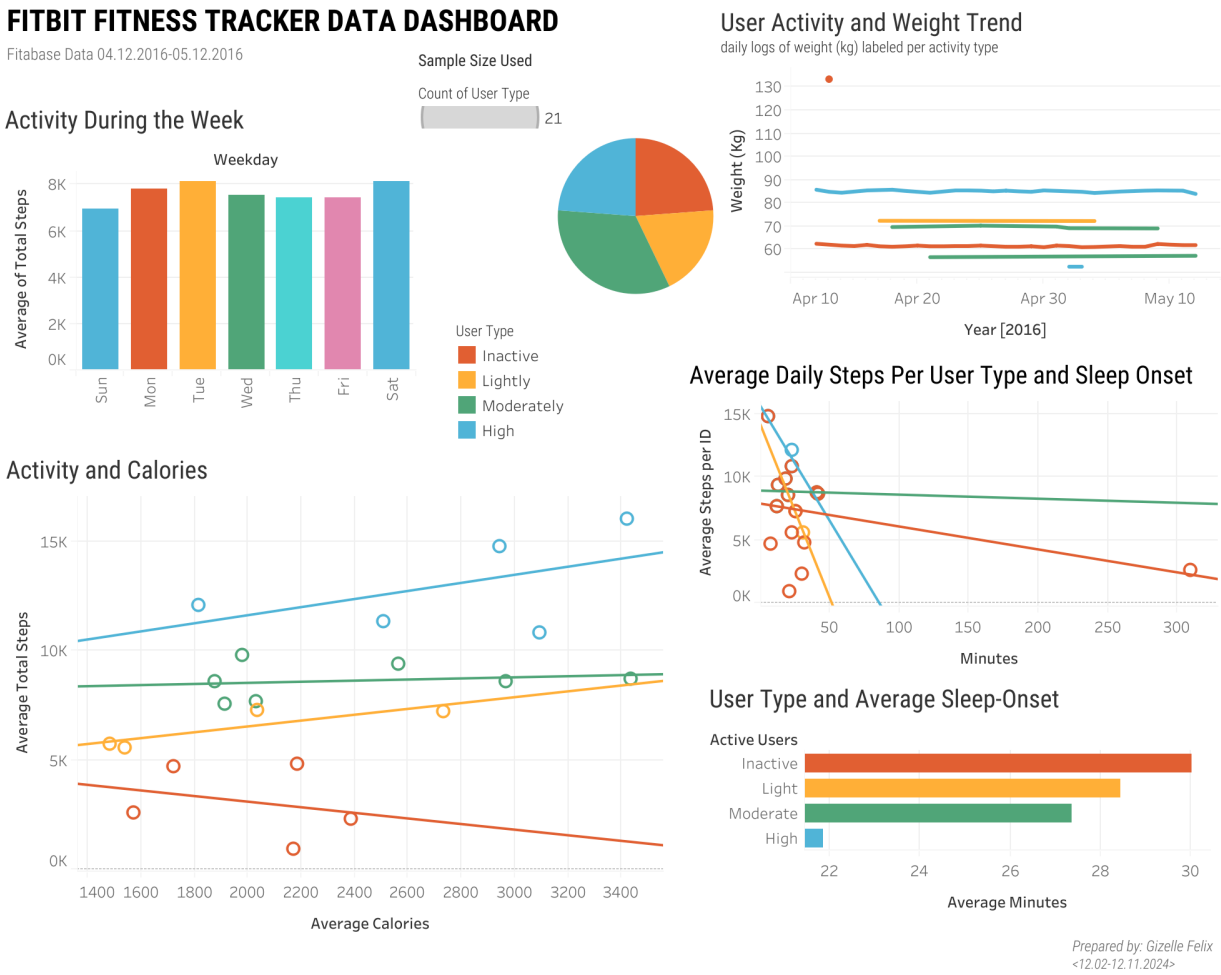
8877689391	2016-04-19	16040.03226	85.3
8877689391	2016-04-20	16040.03226	84.9
8877689391	2016-04-21	16040.03226	84.5
8877689391	2016-04-23	16040.03226	85.5
8877689391	2016-04-24	16040.03226	85.5
8877689391	2016-04-25	16040.03226	85.4
8877689391	2016-04-26	16040.03226	85.1
8877689391	2016-04-27	16040.03226	85.4
8877689391	2016-04-28	16040.03226	85.1
8877689391	2016-04-29	16040.03226	84.9
8877689391	2016-04-30	16040.03226	85.5
8877689391	2016-05-01	16040.03226	85.3
8877689391	2016-05-03	16040.03226	84.9
8877689391	2016-05-04	16040.03226	84.4
8877689391	2016-05-06	16040.03226	85
8877689391	2016-05-08	16040.03226	85.4
8877689391	2016-05-09	16040.03226	85.5
8877689391	2016-05-11	16040.03226	85.4
8877689391	2016-05-12	16040.03226	84



(Image 4)

Not all users had their weight logged but it can be seen that for the span of one month, there are changes to user's weights when users are consistent with their steps.

SHARE



ACT

Questions:

- Why are customers using the devices?
- How can these trends apply to Bellabeat consumers?
- What are the trends in smart device usage?
- How could these trends help influence Bellabeat Marketing strategy?

Insights and Recommendations:

Based on the trends from the data, it is seen that the users that use Bellabeat products as a guide and tool to fulfill their lifestyle goals. The insights gathered from the trends can be used as a source to implement marketing strategies to invite more users based on a user's needs.

Insights:

- More activity, the more calories
- Most active on Tuesdays and Saturdays
- Activity has an effect on user's sleep
- Consistent activity can have an effect on user weight

The more active a user is, the more calories they burn. My recommendation based on this trend is to help users fulfill their overall lifestyle goal. Use the user's activity trend to forecast what their data would look like according to the user's own lifestyle goals. This encourages the users to maintain their activity and usage of the products.

It can be seen that users are most active during Tuesdays and Saturdays. My recommendation for Bellabeat is to reward users for being active on those days. Which keeps users to stay active on those days. For the less active days, encourage users to fulfill their daily goals by reminding them to complete the user's goal for the day. Bellabeat can also implement a reward system to encourage more activity.

According to the data, the less activity a user has the more time it takes for a user to fall asleep. Poor sleep onset can have an overall effect on a person. From a website article published by Dr. Liji Thomas, MD, long-term sleep-onset insomnia can result in significant symptoms of severe sleep deprivation. These could include metabolic problems as well as psychological or psychiatric symptoms that could significantly affect one's quality of life.
(https://www.news-medical.net/health/Causes-of-Sleep-Onset-Insomnia.aspx)
My recommendation based on this trend is to challenge users to meet or pass their expected goal based on the user's needs. To do this, Bellabeat can cater to users that have problems with insomnia, or sleep onset and want to overall improve their sleep. By using user's data, Bellabeat can better track their activity and visually show the amount of activity a user has in relation to their sleep.

Use the reviews from the users on how to improve the devices to better engage with the users that utilize the products on the daily.

CHALLENGES

Challenges from this study were importing data into BigQuery and limited knowledge on utilizing tools.

LIMITATIONS

Limitations include short term data limitation, small sample size used to analyze data trends, and missing/null values in the datasets.