BELLABEAT GOOGLE DATA ANALYTICS CAPSTONE PROJECT USING MYSQL

Executive Summary

This executive summary highlights key findings from the analysis of Fitbit user data and provides actionable recommendations for enhancing user engagement and improving the Fitbit app features. The data analysis process followed the phases of Ask, Prepare, Process, Analyze, and Act, utilizing SQL for data cleaning and analysis.

The analysis revealed that a subset of 21 individuals used the Fitbit app daily, indicating potential for increasing overall user engagement. To encourage regular app usage, a discount program targeting infrequent users is recommended.

The weight tracking feature exhibited low utilization, with only 8 users making use of it, and only 2 users specifically tracking their fitness using the "Fat" metric. Two strategies are proposed to address this: improving the correlation between weight tracking and other features through enhanced visualizations or removing the feature entirely to prioritize more popular app functionalities.

An anomaly in step tracking data was identified, where only half of the users' step data was recorded in the app. To address this, implementing a notification or alarm feature within the app is suggested to remind users to carry their devices and ensure accurate step tracking.

Throughout the analysis, SQL was employed for data cleaning and analysis, following the Ask, Prepare, Process, Analyze, and Act framework.

Implementing these recommendations will help drive user engagement, improve feature utilization, and enhance the Fitbit app experience.

The revised executive summary provides a more concise overview of the analysis and recommendations, ensuring a more suitable length for an executive summary.

Introduction:

Bellabeat case study is a Capstone Project of Google Data Analytics Certificate on Coursera.
Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.
By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on **their website**.

Stakeholders:

> **Urška Sršen:** Bellabeat's cofounder and Chief Creative Officer
> **Sando Mur:** Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
> **Bellabeat marketing analytics team:** A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy.

Scenario

> I am junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global **smart device** market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. I have been asked to focus on one of Bellabeat's products and analyze smart device data to gain insight into how consumers are using their smart devices. The insights will then help guide marketing strategy for the company. You will present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat's marketing strategy.

Goal:

> **Find new growth opportunities for the company.**

Business Task:

- Analyze smart device usage data to gain insight into how people are already using their smart devices.
- present your analysis to the Bellabeat executive team along with your high-level recommendations for Bellabeat's marketing strategy.

ASK

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

PREPARE

Sršen has provided below public data to use that explores smart device users' daily habits.

● FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): 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.

Our data is stored in multiple csv file in long format. I merged some of the files with same primary key for example "hourly_calories_intensity_step" csv using 3 different csv.

Each file has different information such as heartrate per second, calories per second, intensity per second, steps per second, sleep hour and weight.

This dataset is made available under CC0 may be protected by copyright and related or neighboring rights ("Copyright and Related Rights"). License

Limitation:
the datasets have inputs of only 33 unique users. This tells us that the data is not comprehensive. Of the 33 users only 8 entered weight, 12 heart rate and only 24 users for sleep entries.
Data collected from year 2016. Users' daily activity, fitness and sleeping habits, diet and food consumption may have changed since then, hence data may not be timely or relevant.
As data is collected in a survey, hence unable to ascertain the integrity or accuracy of data.

Is Data ROCCC?

A good data source is ROCCC which stands for **R**eliable, **O**riginal, **C**omprehensive, **C**urrent, and **C**ited.

1. Reliable - LOW - Not reliable as it only has 30 respondents.
2. Original - LOW - Third party provider (Amazon Mechanical Turk)
3. Comprehensive - MED - Parameters match most of Bellabeat's products' parameters
4. Current - LOW - Data is 5 years old and is not relevant.
5. Cited - LOW - Data collected from third party, hence unknown.

.

PROCESS

Load dataset in MYSQL.

Datasets Selected

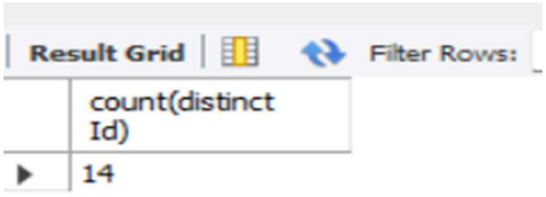
For the case study analysis, the following datasets were chosen:

- dailyactivity_merged
- heartrate_sec
- hourly_calories_intensity_steps
- met_min
- sleepday_merged
- weight_merged

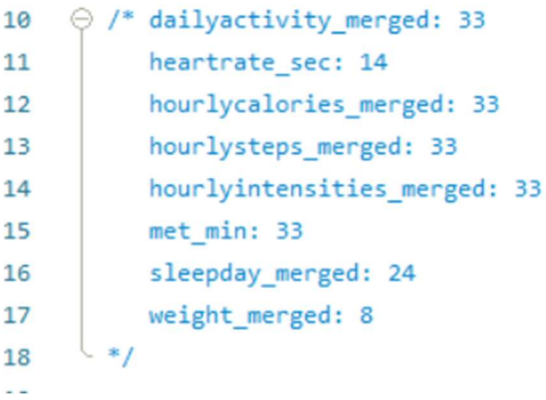
Data Exploration:

1. **Checking number of unique users in each table.**

```
select count(distinct Id) from heartrate_sec;
```



Similarly using same queries for all the table, we got the following results:



2. **Checking whether we have same users on not in different tables.**

```
SELECT COUNT(DISTINCT dailyactivity_merged.id)
FROM dailyactivity_merged
INNER JOIN weight_merged
ON dailyactivity_merged.id = weight_merged.id;
```

Result Grid | Filter Rows:

COUNT(DISTINCT dailyactivity_merged.id)
8

```
/*  heartrate_sec has same 14 users which are in dailyactivity)merged table.
    24 same for sleepday_merged
    8 same for weight_merged
*/
```

3. **Checking Duplicates**

```
SELECT ID, count(*)
FROM dailyactivity_merged
GROUP BY ID, ActivityDate
HAVING COUNT(*) > 1;
```

Result Grid | Filte

ID	count(*)

```
/* no duplicates for this table, but we have found some
   users don't use smart device for all 30 days.
   So now check how many distict days categories we have.
*/
```

```
SELECT ID, COUNT(*)
FROM dailyactivity_merged
GROUP BY ID
HAVING COUNT(*) > 1;
```

ID	COUNT(*)
2320127002	31
2347167796	18
2873212765	31
3372868164	20
3977333714	30
4020332650	31
4057192912	4

```sql
-- checking duplicates for weitht_merged table
SELECT ID, COUNT(LogId)
FROM weight_merged
GROUP BY ID
HAVING COUNT(*) > 1;

/* This is suggesting that only two users have
   significant entry in weight table which is 24 and 30 times
   a month.
 */
```

Result Grid | Filter Rows:

ID	COUNT(LogId)
2873212765	2
4319703577	2
4558609924	5
6962181067	30
8877689391	24

```sql
SELECT id, count(*)
FROM weight_merged
GROUP BY ID, LogId
HAVING COUNT(*) > 1;

/* no items in the result. Hence, we can say that there
   are no duplicate rows in weight_merged table.
*/
```

```sql
-- met_min
select ID, count(*)
FROM met_min
GROUP BY ID, ActivityMinute
HAVING COUNT(*) > 1;
/* no duplicates */
```

Result Grid

ID	count(*)

```sql
select *
from sleepday_merged
group by SleepDay,id, TotalSleepRecords, TotalMinutesAsleep, TotalT
having count(*)>1;

/* there are 3 duplicate rows in above data, so now we need to
   remove them
*/
```

4. **Joining all the tables with hourly data for better analysis.**

```sql
select hcm.id, hcm.activityhour, hcm.Calories, him.Totalintensity,
       him.averageIntensity, hsm.steptotal
from hourlycalories_merged as hcm
inner join hourlyintensities_merged as him
on hcm.id = him.id and hcm.activityhour = him.activityhour
inner join hourlysteps_merged as hsm
on him.id = hsm.id and him.activityhour = hsm.activityhour;

/* now the total number of rows should be same as individual tables
then  only we can say there were no missing values in all three
tables and all 3 tables has same primary composite key which is
combination of id and activityhour.
So, let's check total rows for this new table.
*/

select count(*)
from hourlycalories_merged as hcm
inner join hourlyintensities_merged as him
on hcm.id = him.id and hcm.activityhour = him.activityhour
inner join hourlysteps_merged as hsm
on him.id = hsm.id and him.activityhour = hsm.activityhour;

-- now it is 22099 which is same as individual tables.
/* Hence, we can say that our merge is successful. Thus, we can
   create a new table now and then delete old hourly tables.
*/


drop table hourlycalories_merged;
drop table hourlyintensities_merged;
drop table hourlysteps_merged;
```

5. **Checking for null values in dailyactivity_merged table**

```
select * from dailyactivity_merged
where id is null;
```

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

Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDis

ANALYZE

1. **How frequently people are using their devices.**

```
SELECT distinct(count(*)) As Logged_days, id
FROM dailyactivity_merged
GROUP BY ID;
```

Result Grid | Filter Rows:

Logged_days	id
31	1503960366
31	1624580081
30	1644430081
31	1844505072
31	1927972279

```
create temporary table mytemp As (
SELECT distinct(count(*)) As Logged_days, id
FROM dailyactivity_merged
GROUP BY ID
);

select Logged_days, count(id) As people_count
from mytemp
Group by Logged_days;
```

Result Grid | Filter Rows:

Logged_days	people_count
31	21
30	3
18	1
20	1
4	1
28	1
29	2
26	2
19	1

2. **Weight table has only 8 entries, which means less people prefer to use their devices to log weight related information.**

```
SELECT COUNT(DISTINCT dailyactivity_merged.id)
FROM dailyactivity_merged
INNER JOIN weight_merged
ON dailyactivity_merged.id = weight_merged.id;
```

Result Grid | Filter Rows:

COUNT(DISTINCT dailyactivity_merged.id)
8

3. **Analyze Anomaly in Weight_merged Table**

```
select * from weight_merged;
```

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

	WeightKg	WeightPounds	Fat	BMI	IsManualRe
9:59 PM	52.59999847	115.9631465	22	22.64999962	TRUE
9:59 PM	52.59999847	115.9631465		22.64999962	TRUE
8:52 AM	133.5	294.31712		47.54000092	FALSE
59:59 PM	56.70000076	125.0021043		21.45000076	TRUE
59:59 PM	57.29999924	126.3248746		21.69000053	TRUE
59:59 PM	72.40000153	159.6146812	25	27.45000076	TRUE
59:59 PM	72.30000305	159.3942223		27.37999916	TRUE

```
/*now there is a fat column in weight table which is almost
   empty, so we can conclude that fat feature is not being
   used in the fitbit. only 2 out of 67 user use this featur
*/
```

```
select count(*) from weight_merged where fat!="";
```

Result Grid

count(*)
2

4. **Checking Anomaly in Total Steps**

```
-- check totalSteps column

select count(id), totalSteps
from dailyactivity_merged
where totalSteps=0;

select count(distinct id)
from dailyactivity_merged
where totalSteps=0;
```

Result Grid | Filter Row

	count(id)	totalSteps
▶	77	0

Result Grid

	count(distinct id)
▶	15

```
/* there are total 15 distict users (almost half of the users)
   who do not use their device
   consistenlty for counting steps. because there are total
   77 time when the Total steps taken by the user is 0.
   Resons might be, their is not enough battery backup of device.
   or long charging time or maybe some usere might not find this
   accurate.
   */
```

Act

Smart device usage data insights

Observations:

- Notably, not all individuals utilize the device daily. Specifically, out of the total user base, only 21 individuals consistently use the device every single day of the month.
- The weight tracking feature is not widely utilized by users, with only 8 individuals making use of this feature. Among these 8 users, only 2 individuals employ the "Fat" metric to track their fitness.
- Although the Step feature appears to be crucial for fitness tracking, an anomaly has been detected in the Fitbit data. It has been observed that only half of the users' data is currently recorded in the app. Interestingly, there have been 77 instances where the app has recorded 0 steps, involving a total of 15 distinct users out of the 30 users in total.

Strategies to Enhance User Engagement and Improve Fitbit App Features:

- **"** To encourage regular usage of the Fitbit app among infrequent users, we can implement a discount program tailored specifically for those individuals. This initiative aims to cultivate a habit of using the app consistently. By providing

incentives and rewards, we can motivate users to engage with the Fitbit app on a regular basis.

- Alternatively, we can enhance the correlation between the weight tracking feature and other aspects of the app, presenting the information in a visually appealing and user-friendly manner. This improved visualization will help users appreciate the value and benefits of this feature, potentially increasing its popularity.

- On the other hand, if the weight tracking feature does not resonate with the majority of users, we can consider removing it from the app altogether. By reallocating our focus and resources to other beloved features, we can further enhance the user experience and cater to the preferences and needs of our user base.

- Regarding the anomaly in step recording, there could be various reasons why users were unable to track their steps accurately. For instance, their device may have been uncharged or had limited battery life, or they may have forgotten to carry the device while walking. To address this, we can incorporate a notification or alarm feature within the Fitbit app. This customizable feature would allow users to set reminders according to their preferences, ensuring they remember to carry their device with them for accurate step tracking.