

CAPSTONE PROJECT-FITBIT FITNESS DATASET ANALYSIS WITH SQL & POWER BI



Introduction

Bellabeat is 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. This project focuses on one of the Bellabeat's product.

It analyzes the smart device data

1. To gain insight into how consumers are using their smart device the insights discovered will then help guide marketing strategy for the company, (which is the **BUSINESS TASK** in this project).

About the company

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. 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. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates. Sršen knows that an analysis of Bellabeat's available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are

already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

```
--PREPARE PHASE--
--1.I will be using a dataset provided by Mobius from Kaggle for the
analysis of the usage of fitbit device tracker.
--2.I will be checking the data for consistency, duplicates and Null
values etc.
--3.I will use SQL JOINS to join the tables that have relating data
--4.I will use SQL AGGREGATE FUNCTIONS to count, sum or average columns
from the tables
--5.I will analyze the data in SQL
--6. I will perform data visualization on POWER BI. This is to leverage
the integration of SQL with other application (Power BI)
--7. I will create a database called Bellabeat in MS SQL SERVER and
import all the tables.
```

```
CREATE DATABASE Bellabeat
USE Bellabeat
```

```
-- checking the columns of each table–Familiarizing with datasets--
```

```
SELECT *
FROM dailyActivity_merged
SELECT *
FROM dailyCalories_merged
SELECT *
FROM dailyIntensities_merged
SELECT *
FROM sleepDay_merged
SELECT *
FROM weightLogInfo_merged
```

```
--ASK PHASE--
```

```
--How many people were selected to take part in the data collection
process?--
```

```
SELECT COUNT(DISTINCT(ID)) AS POPULATION
FROM [dbo].[dailyActivity_merged]
```

The screenshot shows a SQL query window with the following text:

```
SELECT COUNT(DISTINCT(ID)) AS POPULATION
FROM [dbo].[dailyActivity_merged]
```

Below the query window, the 'Results' tab is active, displaying a single row of data:

	POPULATION
1	33

--How many days was the activity carried out?--

```
SELECT COUNT(DISTINCT(ActivityDate)) AS NumberofDays
FROM [dbo].[dailyActivity_merged]
```

The screenshot shows a SQL query window with the following text:

```
SELECT COUNT(DISTINCT(ActivityDate)) AS NumberofDays
FROM [dbo].[dailyActivity_merged]
```

Below the query window, the 'Results' tab is active, displaying a single row of data:

	NumberofDays
1	31

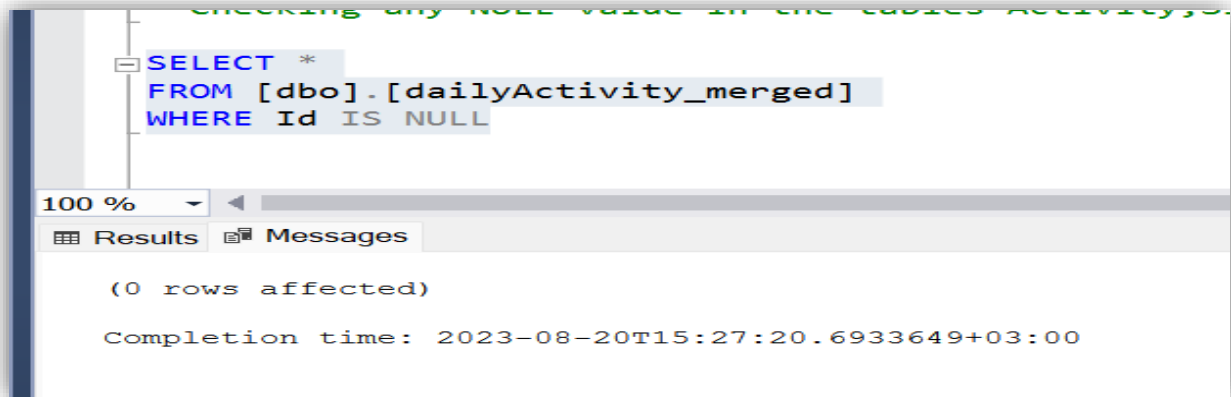
--What distance was covered by each ID?--

```
SELECT Id,Sum(TotalDistance) AS Distance
FROM [dbo].[dailyActivity_merged]
GROUP BY Id
ORDER BY Distance DESC
```

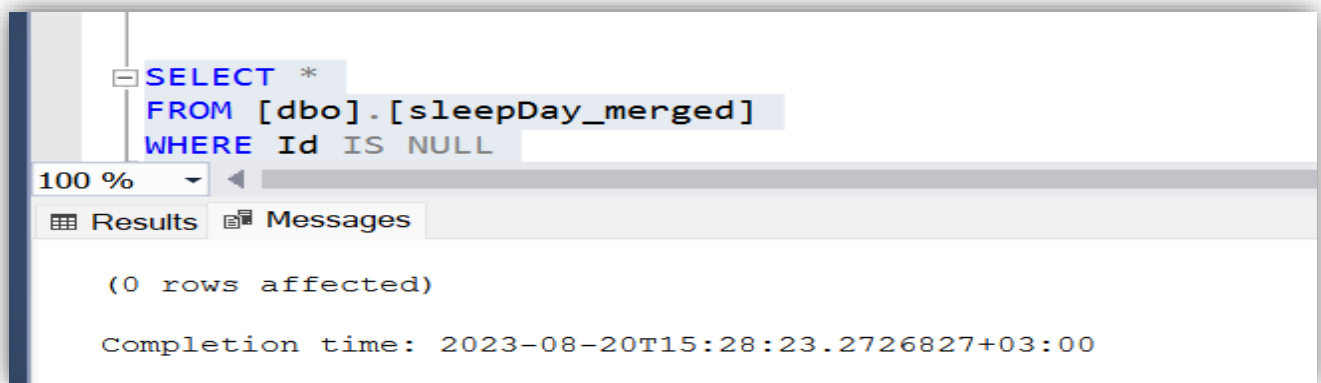
--PROCESS PHASE--

--Checking any NULL value in the tables Activity,Sleep & Weight info Logged?--

```
SELECT *  
FROM [dbo].[dailyActivity_merged]  
WHERE Id IS NULL
```



```
SELECT *  
FROM [dbo].[sleepDay_merged]  
WHERE Id IS NULL
```



```
SELECT *  
FROM [dbo].[weightLogInfo_merged]  
WHERE Id IS NULL
```

```
SELECT *
FROM [dbo].[weightLogInfo_merged]
WHERE Id IS NULL

--there is no NULL values in the tables--
```

100 %

Results Messages

(0 rows affected)

Completion time: 2023-08-20T15:29:44.8869140+03:00

--there is no NULL values in the tables--

--checking for Null IDs in the datasets--

```
SELECT COUNT(DISTINCT (Id))
FROM [dbo].[dailyActivity_merged]
```

```
SELECT COUNT(DISTINCT (Id))
FROM [dbo].[dailyCalories_merged]
```

```
SELECT COUNT(DISTINCT (Id))
FROM [dbo].[dailyIntensities_merged]
```

--check activity with calories

```
SELECT COUNT(dailyActivity_merged.Id)
```

```
FROM dailyActivity_merged
```

```
INNER JOIN dailyCalories_merged ON dailyActivity_merged.Id =
dailyCalories_merged.Id
```

```
AND dailyActivity_merged.ActivityDate =
dailyCalories_merged.ActivityDay
```

```
AND dailyActivity_merged.Calories = dailyCalories_merged.Calories
```

--check activity with steps

```
SELECT COUNT(dailyActivity_merged.Id)
```

```
FROM dailyActivity_merged
```

```
INNER JOIN minuteStepsNarrow_merged ON dailyActivity_merged.Id =  
minuteStepsNarrow_merged.Id
```

```
AND dailyActivity_merged.ActivityDate =  
minuteStepsNarrow_merged.ActivityMinute
```

```
AND dailyActivity_merged.TotalSteps = minuteStepsNarrow_merged.Steps
```

```
--check activity with intensities
```

```
SELECT COUNT(dailyActivity_merged.Id)
```

```
FROM dailyActivity_merged
```

```
INNER JOIN dailyIntensities_merged ON dailyActivity_merged.Id =  
dailyIntensities_merged.Id
```

```
AND dailyActivity_merged.ActivityDate =  
dailyIntensities_merged.ActivityDay
```

```
AND dailyActivity_merged.SedentaryMinutes =  
dailyIntensities_merged.SedentaryMinutes
```

```
AND dailyActivity_merged.LightlyActiveMinutes =  
dailyIntensities_merged.LightlyActiveMinutes
```

```
AND dailyActivity_merged.FairlyActiveMinutes =  
dailyIntensities_merged.FairlyActiveMinutes
```

```
AND dailyActivity_merged.VeryActiveMinutes =  
dailyIntensities_merged.VeryActiveMinutes
```

```
AND dailyActivity_merged.SedentaryActiveDistance =  
dailyIntensities_merged.SedentaryActiveDistance
```

```
AND dailyActivity_merged.LightActiveDistance =  
dailyIntensities_merged.LightActiveDistance
```

```
AND dailyActivity_merged.ModeratelyActiveDistance =  
dailyIntensities_merged.ModeratelyActiveDistance
```

```
AND dailyActivity_merged.VeryActiveDistance =  
dailyIntensities_merged.VeryActiveDistance
```

--After inspection, We find out that Calories, intensities and steps data's column all match with activity data.

--In this case I will use activity data to continue the process

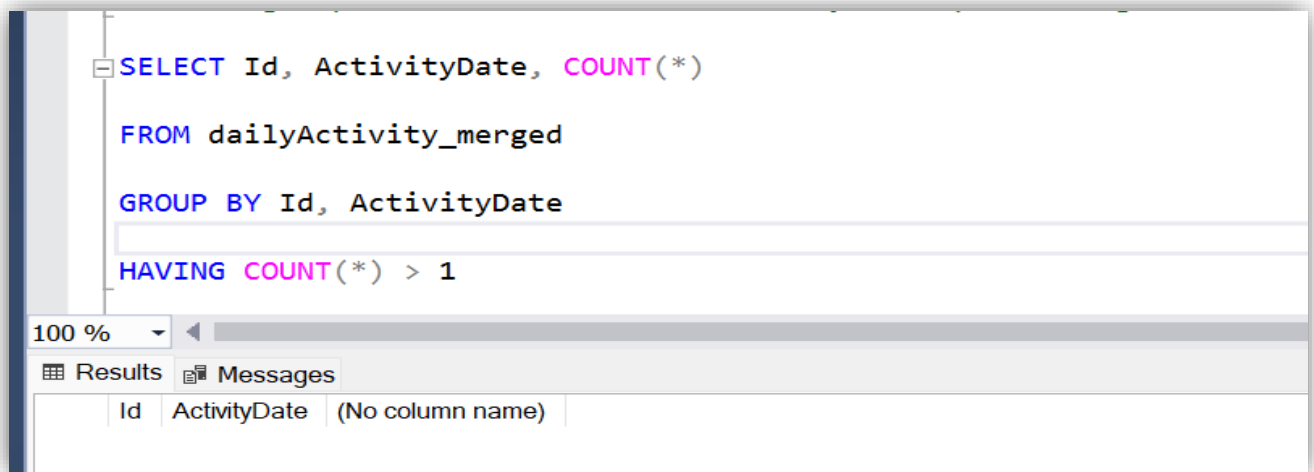
--Finding duplicates of row from activity, sleep and weight

```
SELECT Id, ActivityDate, COUNT(*)
```

```
FROM dailyActivity_merged
```

```
GROUP BY Id, ActivityDate
```

```
HAVING COUNT(*) > 1
```



--no duplicates

```
SELECT Id, Date, COUNT(*)
```

```
FROM weightLogInfo_merged
```

```
GROUP BY Id, Date
```

```
HAVING COUNT(*) > 1
```



```

SELECT Id, Date, COUNT(*)
FROM weightLogInfo_merged
GROUP BY Id, Date

```

100 %

Results Messages

	Id	Date	(No column name)
--	----	------	------------------

--No duplicates

```

SELECT Id, SleepDay, COUNT(*)
FROM sleepDay_merged
GROUP BY Id, SleepDay
HAVING COUNT(*) > 1

```

```

SELECT Id, SleepDay, COUNT(*)
FROM sleepDay_merged
GROUP BY Id, SleepDay
HAVING COUNT(*) > 1

```

100 %

Results Messages

	Id	SleepDay	(No column name)
1	8378563200	2016-04-25 00:00:00.0000000	2
2	4388161847	2016-05-05 00:00:00.0000000	2
3	4702921684	2016-05-07 00:00:00.0000000	2

```

--Having duplicates
--Removing duplicates

--Remove duplicate sleep data and

CREATE TABLE Bellabeat.SleepDay_mergedNew

AS

SELECT DISTINCT *

FROM Bellabeat.SleepDay_merged

--Checking for NULL values in Activity data,Sleep and Weight

SELECT *
FROM dailyActivity_merged
WHERE Id IS NULL
--there is no NULL in Activity table

SELECT *
FROM weightLogInfo_merged
WHERE Id IS NULL
--there is no NULL in Weight table

SELECT *
FROM sleepDay_merged
WHERE Id IS NULL
--there is no NULL in Sleep table

--ANALYZE AND SHARE PHASE--

--Find how active ppl use device

SELECT Id, Count(Id) as total_activity,

CASE

WHEN COUNT(Id) BETWEEN 21 and 31 THEN 'Active User'

WHEN COUNT(Id) BETWEEN 11 and 20 THEN 'Moderate User'

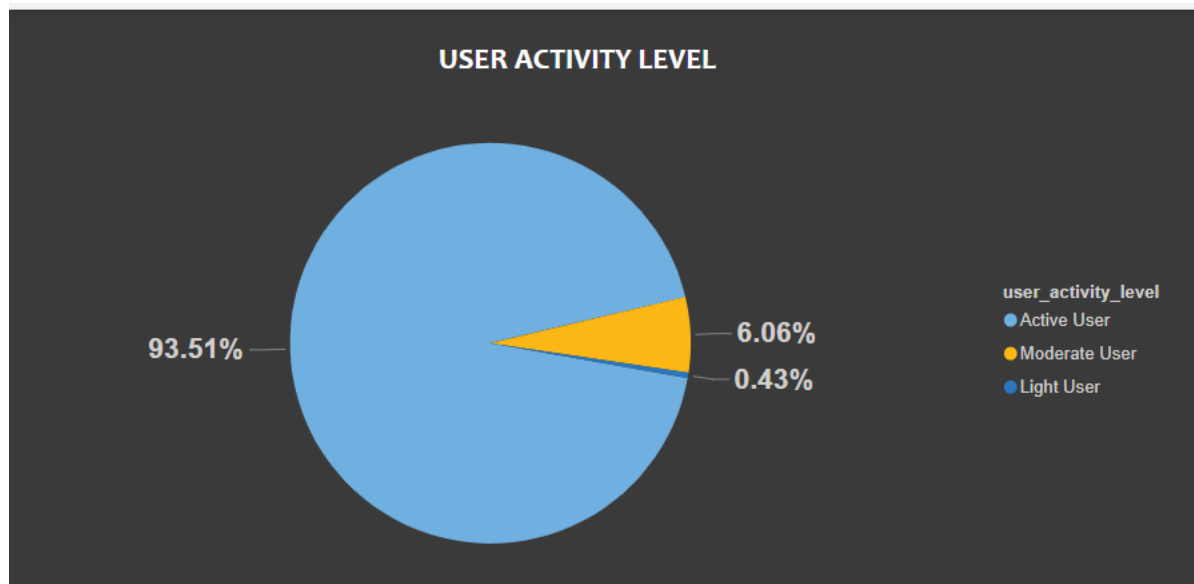
WHEN COUNT(Id) BETWEEN 0 and 10 THEN 'Light User'

```

```

END AS user_activity_level
FROM dailyActivity_merged
GROUP BY Id
ORDER BY total_activity DESC

```



--we find out 93% of fitbit users are active user who use the tracker 21-31 days a month.

--6% of fitbit users are moderate user who use the tracker 11-20 days a month.

--0.43% of fitbit users are light user who use the tracker 0-10 days a month.

--Find the average of activity minutes per week

--Find the average of activity minutes per week

```

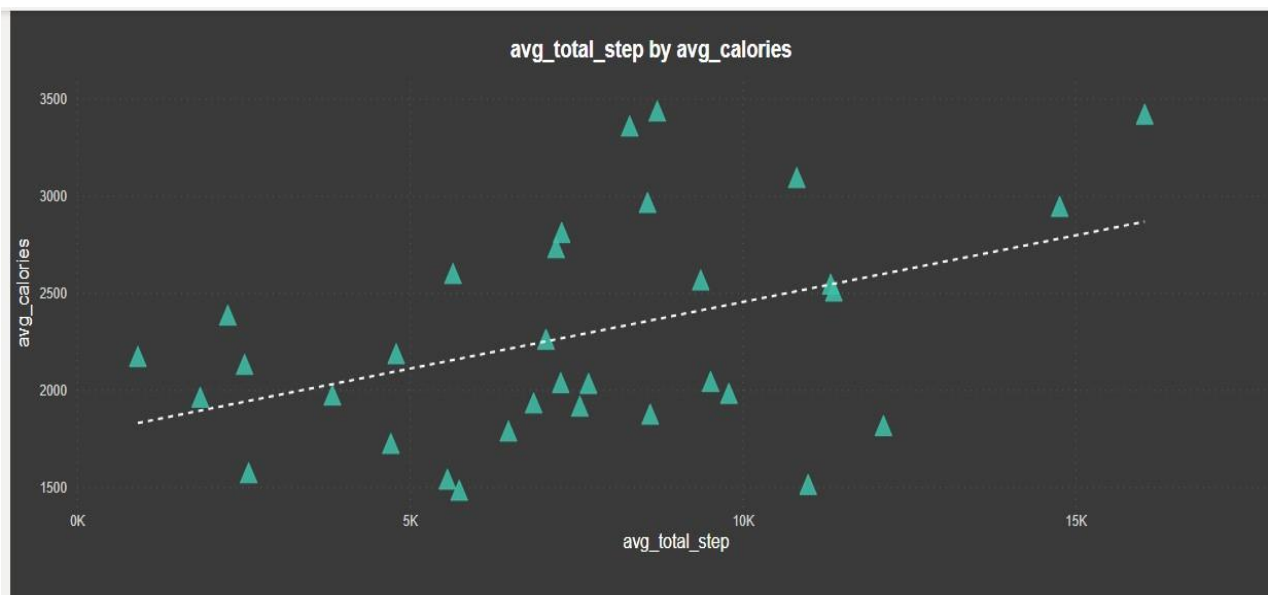
SELECT AVG(VeryActiveMinutes) AS VeryActive,
       AVG(FairlyActiveMinutes) AS FairlyActive,

```

```
    AVG(LightlyActiveMinutes) AS LightlyActive,  
    AVG(SedentaryMinutes) AS SedentaryMinutes  
FROM dailyActivity_merged
```

--Finding average total step vs calories

```
SELECT Id,  
  
    AVG(Calories) as avg_calories,  
  
    AVG(TotalSteps) as avg_total_step  
FROM dailyActivity_merged  
  
GROUP BY Id  
  
ORDER BY avg_calories DESC
```



--There exist a positive relationship between steps and amount of calories burnt.

--Meaning the more the steps a user made the more the amount of calories they burnt.

```
--finding average weight with respect to sleep minutes
SELECT sleepDay_merged.Id,

AVG(sleepDay_merged.TotalMinutesAsleep) as avg_sleepingtime,

AVG(weightLogInfo_merged.BMI) as avg_BMI

FROM sleepDay_merged
FULL JOIN weightLogInfo_merged
ON sleepDay_merged.Id=weightLogInfo_merged.Id

GROUP BY sleepDay_merged.Id

ORDER BY avg_BMI DESC
```



```
--High BMI is visible in users with averagly more sleeping time (high
body fatness)
--But majority of users did not give enough data on weight and BMI, only
6 users is not enough to make a decision
--Going forward i cannot use this data to make any decision
```

```
--Another correlation test
--finding average sleep time and amount of calories burnt
```

```
SELECT sleepDay_merged.Id,

AVG(sleepDay_merged.TotalMinutesAsleep) as avg_sleepingtime,
```

```

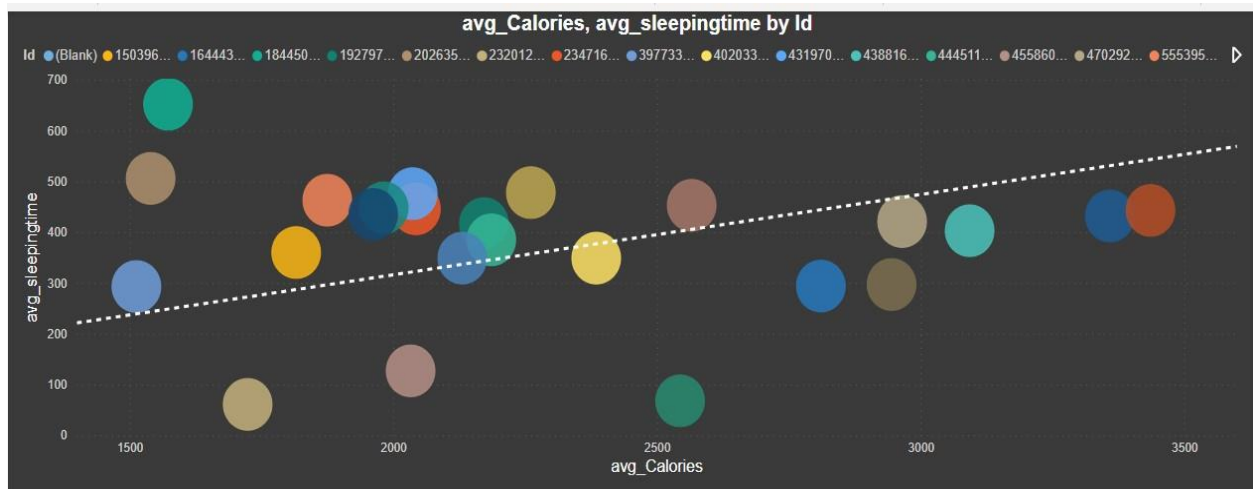
AVG(dailyCalories_merged.Calories) as avg_Calories

FROM sleepDay_merged
FULL JOIN dailyCalories_merged
ON sleepDay_merged.Id=dailyCalories_merged.Id

GROUP BY sleepDay_merged.Id

ORDER BY avg_Calories DESC

```



--There is major relationship between the number of minutes a user slept and the number of calories they burnt
 --meaning longer sleeping users burnt the least calories and vice versa

--ACT PHASE--

- The company can consider collecting data from many users or come up with an APP linked to their internal data source and to every device to help in collection of sufficient data.
- A population of only 33 users is quite small and this narrows the confidence of making broader decisions

- I can recommend that the most active users be rewarded as most loyal customers for their loyalty and continued use of the devices
- The company can as well consider tracking the performance data for other products i.e. spring bottle, Time, Bellabeat APP etc.
- The company can consider using internal data and not external i.e. from Mobius
- 5.The company can also analyze the sales generated by the Fitbit and ascertain whether they are making profits or to determine the lucrativeness of its production.
- 6.I would finally recommend the dataset to have gender specified so that the variability in the outcome can be grouped by gender

By Cynthia Dalmas-Data analyst
END...