Sandya

Case study 2\_Excel and SQL

**Ask Phase:**

**About Company**

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. 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, launched multiple products, and also invested year-round in online platforms.

**Key Stakeholder**

* 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

**Products**

* Bellabeat app: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.
* Leaf: Bellabeat’s classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
* Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat application to provide you with insights into your daily wellness.
* Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.
* Bellabeat membership: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health, beauty, and mindfulness based on their lifestyle and goals.

**Data used:**

* Dataset : [Fitbit Fitness Tracker Data](https://www.kaggle.com/datasets/arashnic/fitbit)
* Date description : [Fitabase data dictionary](https://www.fitabase.com/media/1930/fitabasedatadictionary102320.pdf)

**Import CSV files into SQL**

* I used my SQL work bench.

The dataset has 18 CSV files. The data also follow a ROCCC approach:

Reliability: The data is from 30 FitBit users who consented to the submission of personal tracker data and generated by from a distributed survey via Amazon Mechanical Turk.

Original: The data is from 30 FitBit users who consented to the submission of personal tracker data via Amazon Mechanical Turk.

Comprehensive: Most data is recorded during certain days of the week.

Current: Data is from March 2016 to May 2016. Data is not current.

Cited: The information was appropriately cited in Kaggle with the License-CC0 1.0 Universal (CC0 1.0) Public Domain Dedication

**Cleaning:**

**All tables:**

* Check for extra spaces.
* Check length of id columns
* Check for inconsistencies across tables.

**DailyActivity\_merged:**

* Deleted all the distance columns from dailyActivity\_merged table.
* Added a new column named totalactive\_min by adding three active minute columns.
* Added a day column

**sleepDay\_merged:**

* Added column day using formula =text(B2,”dddd”)
* Added column time\_bed by substracting TotalMinutesAsleep from TotalTimeInBed

**weightLogInfo\_merged:**

* Divide Date\_time into date and time using excel.
* Delete fat, logId,WeightKg columns
* Rounded the WeightPounds to 3 decimal places and named weight.
* Added a day column
* Change isManualReport to manual\_report and replaced true with manual and false with automatic.

**Analysis:**

From SQL:

* There are 6 common rows in all the three tables.
* Sleep has all the rows as in activity, activity has all the rows as weight but sleep and weight only has 6 rows in common.
* Created a new table containing avg activities for the columns in activity table, avg\_sleep for columns in sleep table and avg\_weight for weight table.
* Imported all the newly created tables to **excel** for further analysis.

From Excel:

* From avg\_activity, I found that all the people except 9 recorded their activity for 31, 31 days.
* Lowest total\_avg\_dist was 0.63 km and that was recorded by person with 31 days record and his avg active min is 40.6 min.
* The highest total\_avg\_dist was 13.21 kms and total active time is 310.71
* From sleep\_activity table, only 7 people recorded their night activity for 28,31,32 days.
* The highest avg sleep was 10.86 hours and that was recorded by person with 3 days record.

From SQL:

* I combined avg\_activty and avg\_sleep to know relation bw those two tables. I used “join”. It retrieved the columns but join was not working in sql workbench so I saved the output as csv file, opened it in excel and deleted all the sllep\_id with null values and finally deleted sleep\_id column.
* From weight table I creates avg\_weight table.
* Max avg weight was 294.317 lbs and min was 115.963
* Only 8 users logged their weight for total of 31 days and in those 8 people only 1,2 or 3 people recorded their weights. From this we can assume that people are only checking their weight once in while.
* I added hr\_asleep to sleep table to convert minute sleep time to hours.
* From sleep table, people are sleeping most on Sunday, Wednesday, Saturday.

From Excel:

* From activity table, people gradually stopped after 26 days. We can see decrease to 24 people after may 10.
* From sleep table I created a pivot table containing dates and avg hr sleep. People didn’t record their sleep consistently.
* But from the data recorded we can say that people nearly slept for 7 hours everyday.
* From activity table, I created a pivot chart and found that people are walking most on Saturday, tuesday, Monday and active most on Saturday, Friday, Tuesday
* People are walking less on Sunday.
* Created a weight diff table to calculate diff bw weight on start\_date and end\_date.
* There are 8 people. People who recorded 1 or 2 days showed 0 % weight % change. People who recorded for 30 days showed the most decrease in weight. Only one person who recorded for 22 days showed increase in weight.

From SQL:

* Created a table common\_ids to save all the ids common between sleep and weight in turn in activity table.
* Combined avg\_activity, avg\_sleep, avg\_weight, weight\_diff basedon ids in common\_ids and saved in a table.
* From the table, I found that all the ids that are common in all the three tables(activity, sleeo, weight) -> recorded their activity for 30 or 31 days,

->2 people recorded their sleep for 5 days and others more than 25 days.

-> only one person recorded their weight for 30 days, all others for 1,2,5 days.

From this we can say that, people liked checking their weight for only few times in a month.

Summary:

* People are burning more calories the more they are active or the more they are walking.
* There is a positive correlation between activity and calories burned.
* There is also positive correlation between avg steps taken and avg calories burned.
* People started not to log their activity after some days.
* According to the logs recorded, max avg weight is recorded on Wednesday, Sunday, Saturday.
* Number of people recording weight are very low. People are only checking their weight once in a while.
* Only 2 people recorded their weight for 24, 30 days in the 8 people that recorded weight.
* There is a positive correlation between very\_active\_min activity and total min asleep. That means people are going to sleep faster when there is more activity.
* There is a negative correlation between calories burned and vg total min not asleep. That means people are sleeping fast and sound when they burned more calories.
* There are ups and downs in sleep log activity. That means people are checking their sleep activity once in few days not everyday.

**Recommendations:**

* Encourage people wear fit bit everyday by giving them rewards and points
* Give people a task everyday.
* Add features that can give a work out plan or diet plan based on the activity.
* Add memberships for exercise coaching.
* Add features that can record blood pressure, heart beat, etc.,
* People are showing good results when they are exercising actively, give them plan to be more active more time in a day.
* Send notification to tale break or drink water.
* Make the design of the fit bit as a sport watch and also as normal day to day watch so that it can go with any kind of clothing style.
* Send notification to remind them to sleep.
* Let them add their initial weight and their goal weight. Give them a plan and instructions to reach that goal weight. Praise when they reached their goal weight.
* Let the watches sync with others to that friends can workout together.
* Promote in the direction of health and well being rather than solely focusing on weight loss.

**Visualizations:**

[**power bi\_Sandya**](https://app.powerbi.com/groups/me/reports/b5257b3f-00ce-41da-87a0-e480883b906c/ReportSection?experience=power-bi)