# Data Cleaning Project

**About the company**  
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.

I began the data analysis process by identifying insights that could be derived from the data and the problems that those insights could solve. I downloaded and organized the data from Kaggle, a public database, and carefully examined it to determine which parts required cleaning. As the dataset was too large for Excel, I moved to MySQL for further analysis. I cleaned the data by searching for nulls and repeating data points, correctly formatting and creating new columns. After spending some time with the data, I organized it to make it easier to read and understand. I also joined and updated a few columns to reduce clutter and make the dataset ready for analysis.”

I discovered trends within the data. Although the dataset only had 33 people whose data I was cleaning, I did the best I could to find insights. I created multiple tables in Excel sheets for a summary of the data and moved the data to Tableau. In Tableau, I created visualizations to help explain my findings to the stakeholders. The next steps would be suggesting to the stakeholders to add additional datasets for users’ age, gender, and general location. This will help narrow down the demographics of who’s using their products and which locations they should target more with marketing and physical locations.

Code used in SQL Server to Clean and organize the data:

--Cleanig the Data

Select \*

From [GD Analytics].[dbo].[dailyActivity]

-- Finding Nulls and clearing or filling them in if information is available

select \*

From [GD Analytics].[dbo].[dailyActivity]

Where TotalDistance is null

--Creating a proper fromat for the Activity Date as "New Activity Date"

Select ActivityDate Convert(Date,ActivityDate)

From [GD Analytics].[dbo].[dailyActivity]

Update [GD Analytics].[dbo].[dailyActivity]

set ActivityDate = Convert(Date,ActivityDate)

Alter Table [GD Analytics].[dbo].[dailyActivity]

Add NewActivitydate = Convert(Date,ActivityDate)

--Combining very active minutes and Faily active minutes as Active minutes

Select Veryactiveminutes, FairlyActiveMinutes,

Veryactiveminutes + FairlyActiveMinutes As ActivityMins

From [GD Analytics].[dbo].[dailyActivity]

-- Creating a new colum of Combined activity minutes

Select \*

From [GD Analytics].[dbo].[dailyActivity]

--Adding a new Colum

Alter Table [GD Analytics].[dbo].[dailyActivity]

Add ActivityMins Varchar (50) Null

-- Updating new Colum

Update [GD Analytics].[dbo].[dailyActivity]

set ActivityMins = Veryactiveminutes + FairlyActiveMinutes

--Combinging VeryActive Distance and ModeratelyActive Distance

Alter Table [GD Analytics].[dbo].[dailyActivity]

Add Active\_Distance Varchar (50) Null

Update [GD Analytics].[dbo].[dailyActivity]

set Active\_Distance = VeryactiveDistance + ModeratelyActiveDistance

-- Combining all the minutes to find how many minutes in the day customers use the tracker

--Adding a column

Alter Table [GD Analytics].[dbo].[dailyActivity]

Add TotalUsageMins Varchar (50) Null

--Updating the column

Update [GD Analytics].[dbo].[dailyActivity]

set TotalUsageMins = VeryactiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes + SedentaryMinutes

Select Veryactiveminutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes,

Veryactiveminutes + FairlyActiveMinutes + LightlyActiveMinutes + SedentaryMinutes As TotalUsageMins

From [GD Analytics].[dbo].[dailyActivity]

------------------------------------------------------

Select \*

From [GD Analytics].[dbo].[dailyActivity]

Select Distinct ID,NewActivitydate,Totalsteps,Active\_Distance,Activitymins

From [GD Analytics].[dbo].[dailyActivity]

Order By Totalsteps

------------------------------------------------------

--creating general pivot tables

Select \*

From [GD Analytics].[dbo].[dailyActivity]

Select Distinct ID,NewActivitydate,Totalsteps,Active\_Distance,Activitymins

From [GD Analytics].[dbo].[dailyActivity]

Where ID =1503960366

Select AVG(TotalSteps) As AvgSteps

From [GD Analytics].[dbo].[dailyActivity]

Where ID = 1503960366

--ID,NewActivitydate,Totalsteps,Active\_Distance,Activitymins

Select AVG(TotalSteps) As AvgSteps

From [GD Analytics].[dbo].[dailyActivity]

Group by ID

Select AVG(TotalSteps) As AvgSteps

From [GD Analytics].[dbo].[dailyActivity]

Group by NewActivityDate

Select AVG(TotalDistance) As AvgDistance

From [GD Analytics].[dbo].[dailyActivity]

Group by ID

Select Distinct Id

From [GD Analytics].[dbo].[dailyActivity]

Group by Id

Select AVG(TotalSteps) As AvgSteps

From [GD Analytics].[dbo].[dailyActivity]

Where ID = 7007744171

Select Distinct Id,

From [GD Analytics].[dbo].[dailyActivity]