# 

# Bellabeat Fitness Data Analysis using only Excel

The case study below shows the six-step data analysis process per the Google Daa Analytics Course which include: Ask, Prepare, Process, Analyse, Share and Act.

Background: Bellabeat was founded in 2013, since then, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women. The company has 5 focus products: bellabeat app, bellabeat membership, time, spring and leaf. Bellabeat has the potential to become a larger player in the global smart device market. Our team have been asked to analyze smart device data to gain insight into how consumers are using their smart devices. These insights we will discover will help in guideing the perfect marketing strategy to be used for the company.

**ASK**

**Business task:** Analyze the data from the FitBit fitness tracker in order to gain insights into how consumers are using FitBit app and and also to discover trends in order to find the most suitable marketing strategy for Bellabeat.

**Primary stakeholders:** Sando Mur and Urška Sršen, Founders / Executive Team Members.

**Secondary stakeholders:** Bellabeat marketing team.

**PREPARE**

Information on Data Source:

1. Thirty (30) FitBit users consented to the submission of their personal tracking of data.

2. Data collected and anlysed included: physical activity recorded (minutes), heart rate, sleep monitoring, daily activity and steps.

3. [Data used for this analysis is publicly available on Kaggle](https://www.kaggle.com/datasets/arashnic/fitbit): FitBit Fitness Tracker Data is/ was stored in 18 csv files.

A good data source is Reliable, Original, Comprehensive, Current, and Cited.

* Reliability of this data is LOW and not reliable as it only has 30 respondents for analysis hence it does not represent the accurate sample size which will eventually lead to bias.
* Originality of the data is also LOW because it was sourced from a third party provider (Amazon Mechanical Turk)
* Comprehensitivity is MEDIUM because the data minute-level output for physical activity, heart rate, and sleep monitoring. Even though the data tracks many factors in the user activity and sleep, sample size is actually small as stated earlier and largeb part of the data was recorded during some select days of the week.
* Current status of this data is LOW because the data is 6 years old and may not be relevant
* Citation of this dat is LOW because the data was collected from third party, hence it is unkown or not fully known

Overall, the dataset is is of a low, poor or bad quality data and it is heavily not recommended to produce business recommendations based on the sourced data.

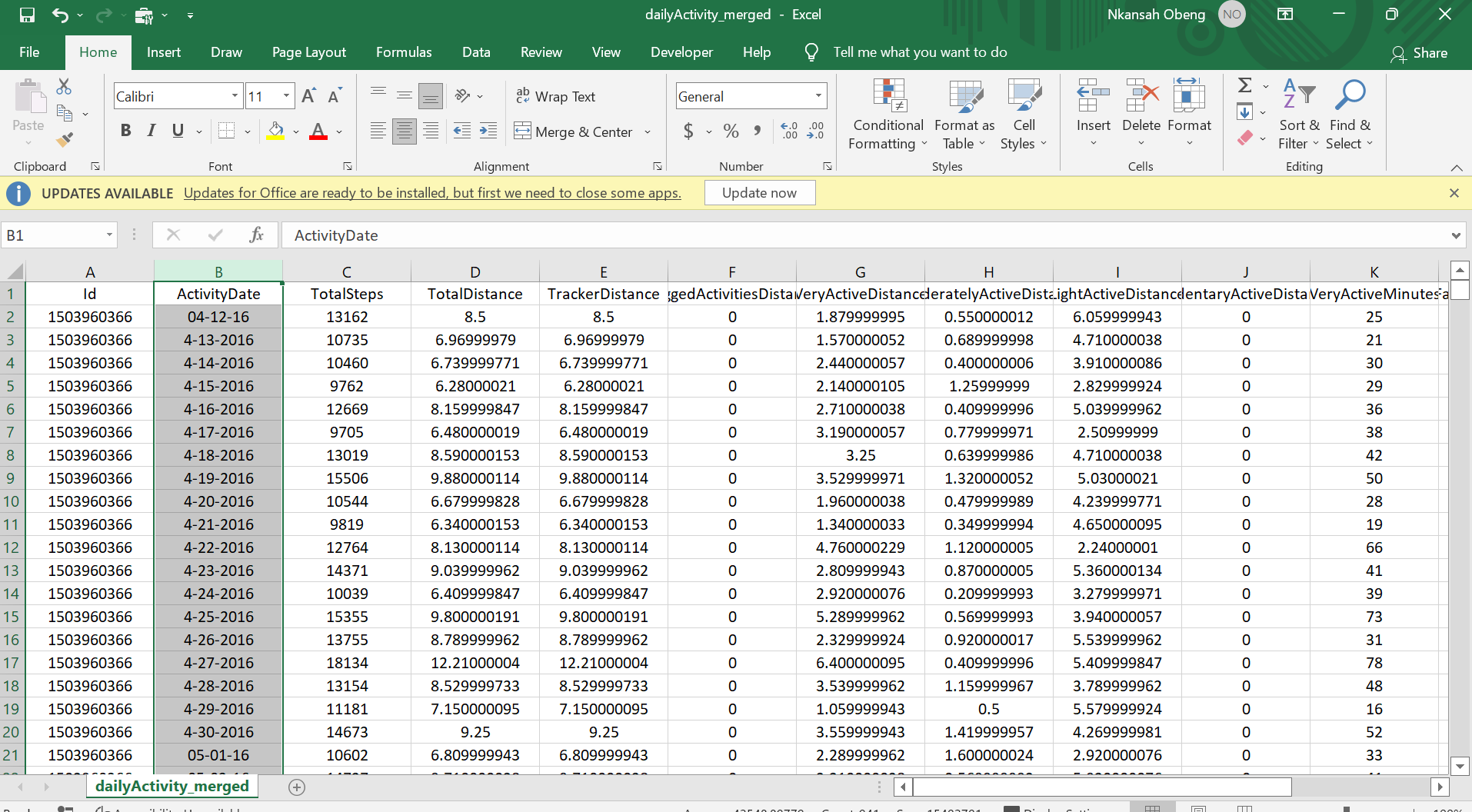
**Data selection**: “dailyActivity\_merged.csv” dataset was used for the data analysis.

**Tool:** Excel was used for data cleaning, transformation and visualisation.

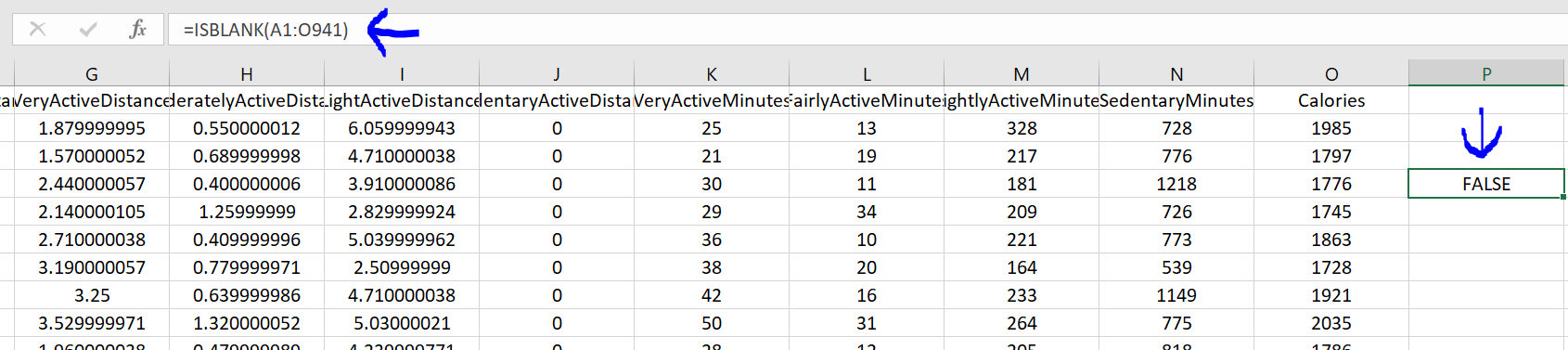
**PROCESS**

**Data cleaning and manipulation:**

1.Observing and familiarizing with the data

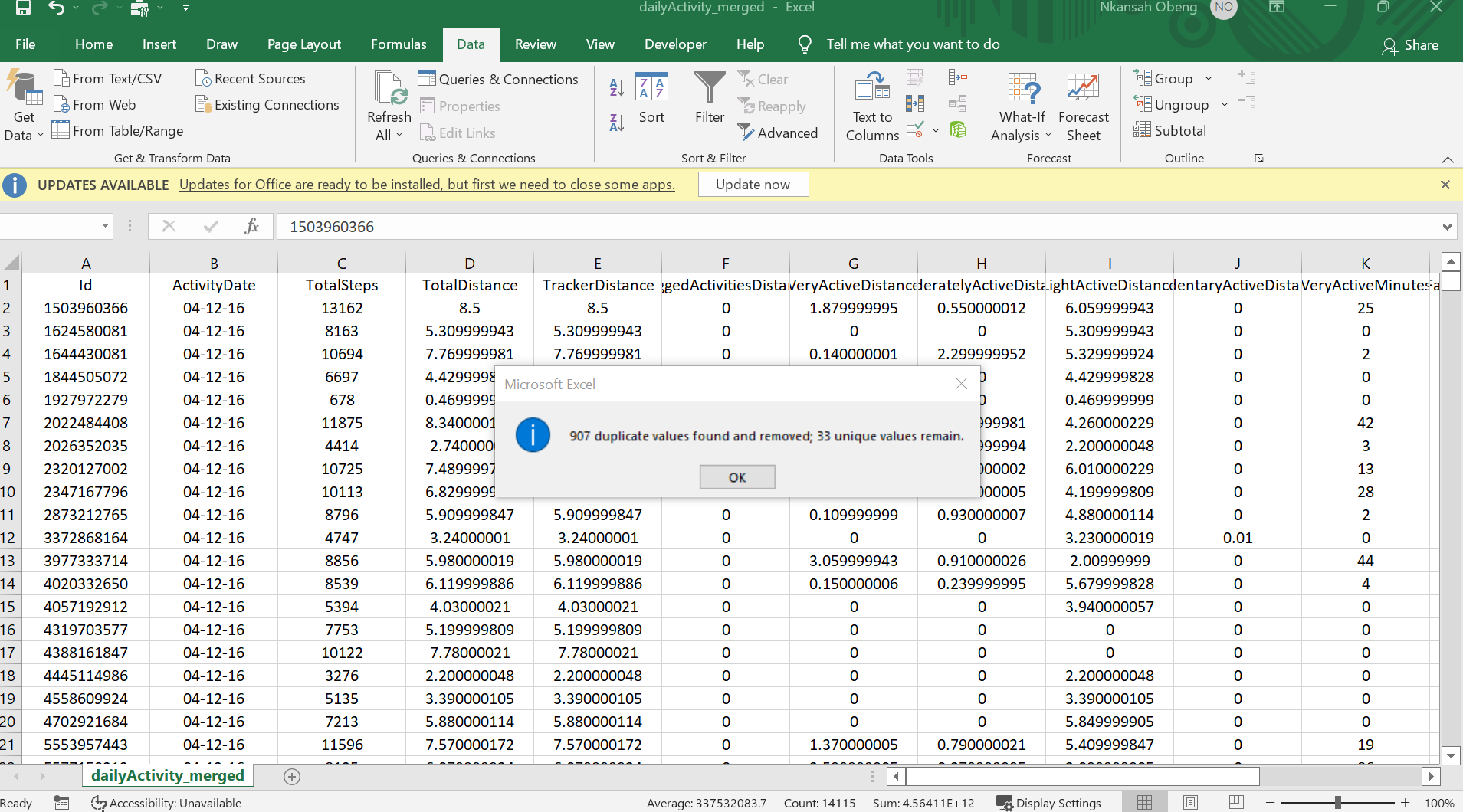
**Note: This is just a screenshot of part of the table. The full table is**[**here**](https://www.kaggle.com/datasets/arashnic/fitbit)**.**

2.Check for null or blank values. The whole data was sselected and ISBLANK function was used to deterine if there were any blank or missing values. As per the screenshot below. It returned a FALSE result which shows there were blank or missing value in the data to be analyzed. Formula is: =ISBLANK(Value)

**As you can see from the screenshot above, there is no NULL value.**

3.We are also going to count unique IDs to confirm whether data has 30 IDs as claimed by the survey using this method

* To filter for unique values, click Data > Sort & Filter > Advanced.
* To remove duplicate values, click Data > Data Tools > Remove Duplicates



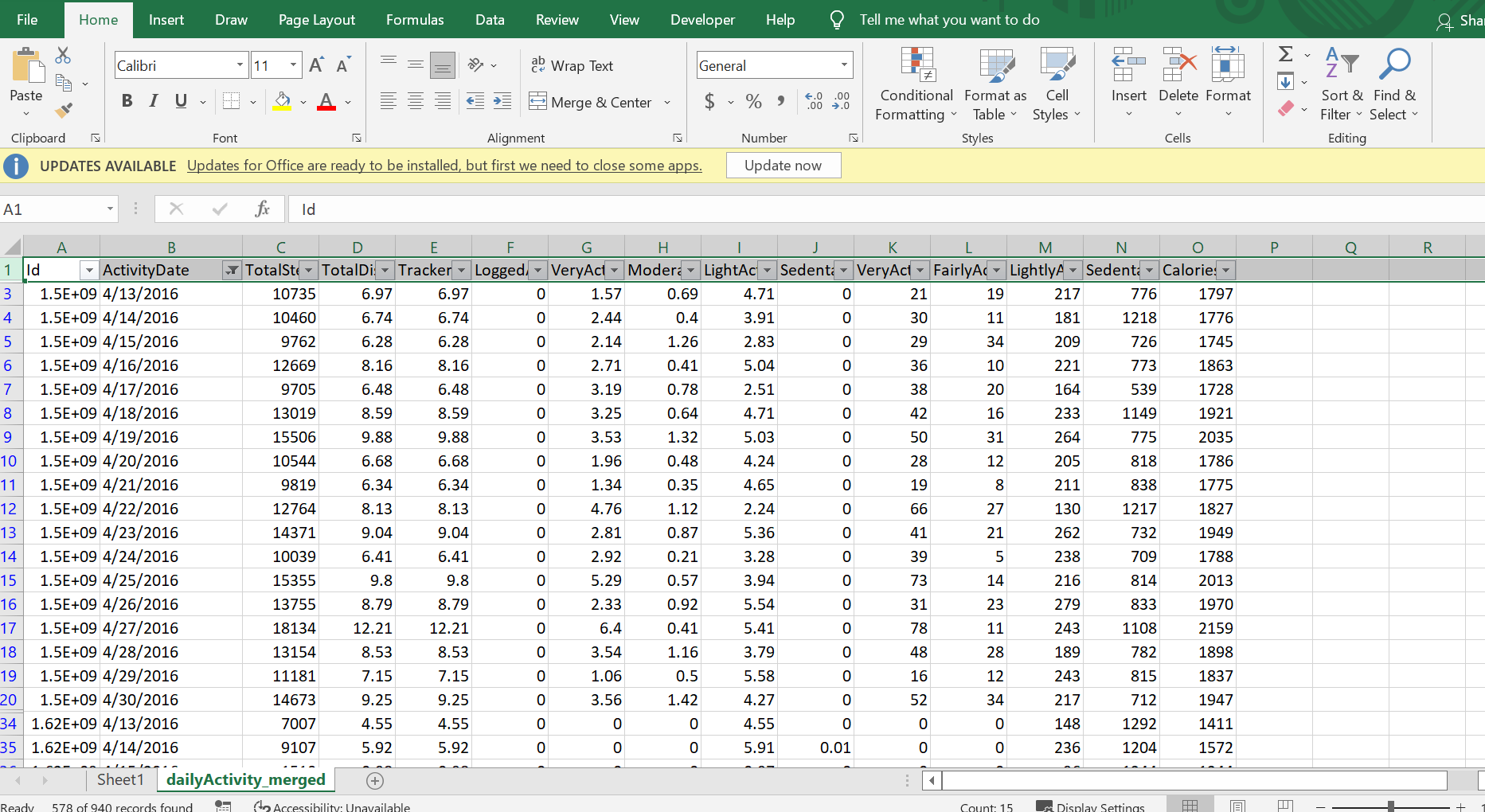
1From the above observations, we noted that:

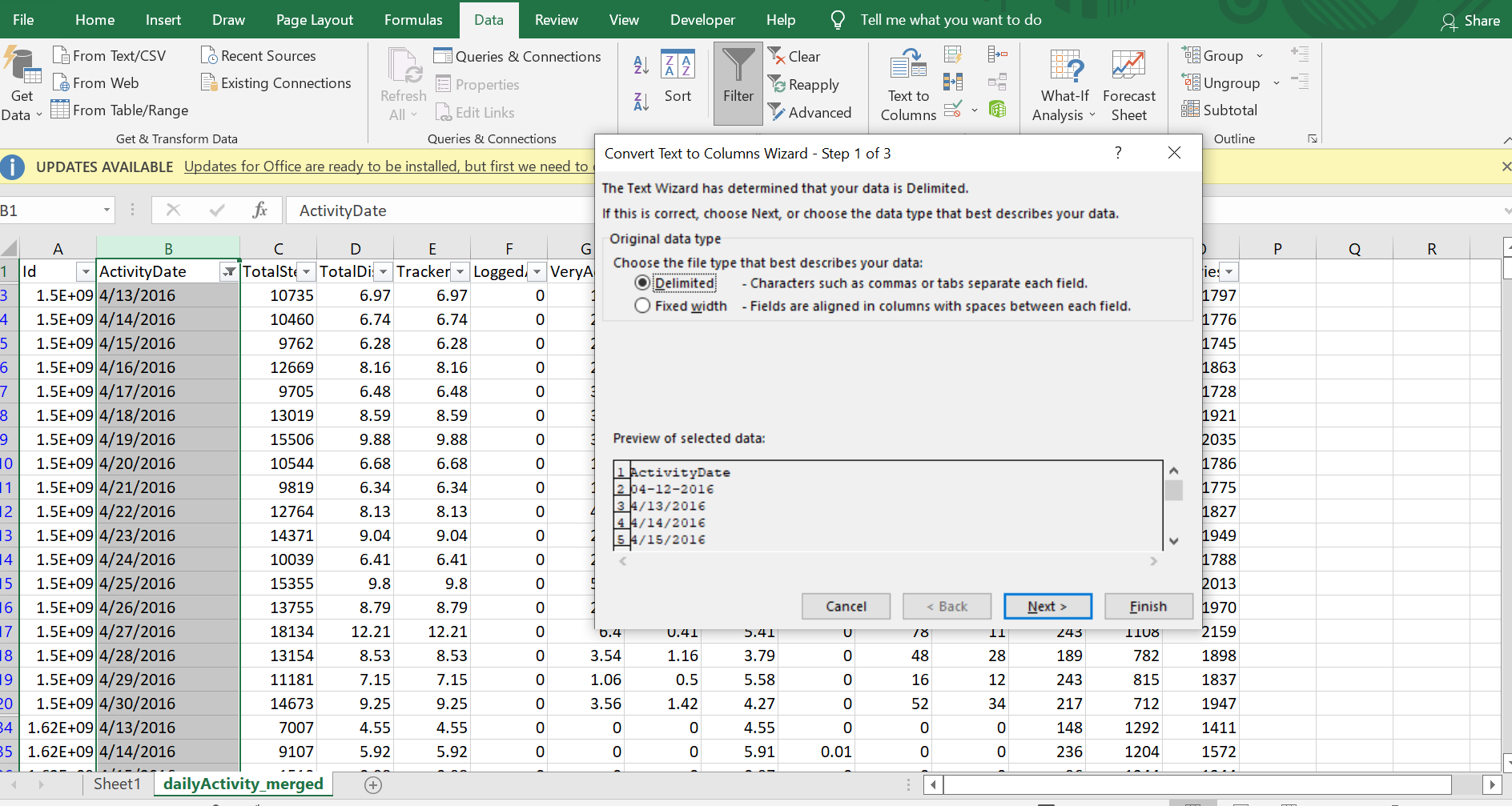
* There are Null or missing values as stated under the ‘Non-Null Count’.
* ActivityDate column is wrongly classified as general and was converted to be converted to short date type.
* There are 33 unique or different IDs, instead of 30 unique IDs as perceived. It may highly likely be because some users may have created other additional IDs during the data collection or survey period.

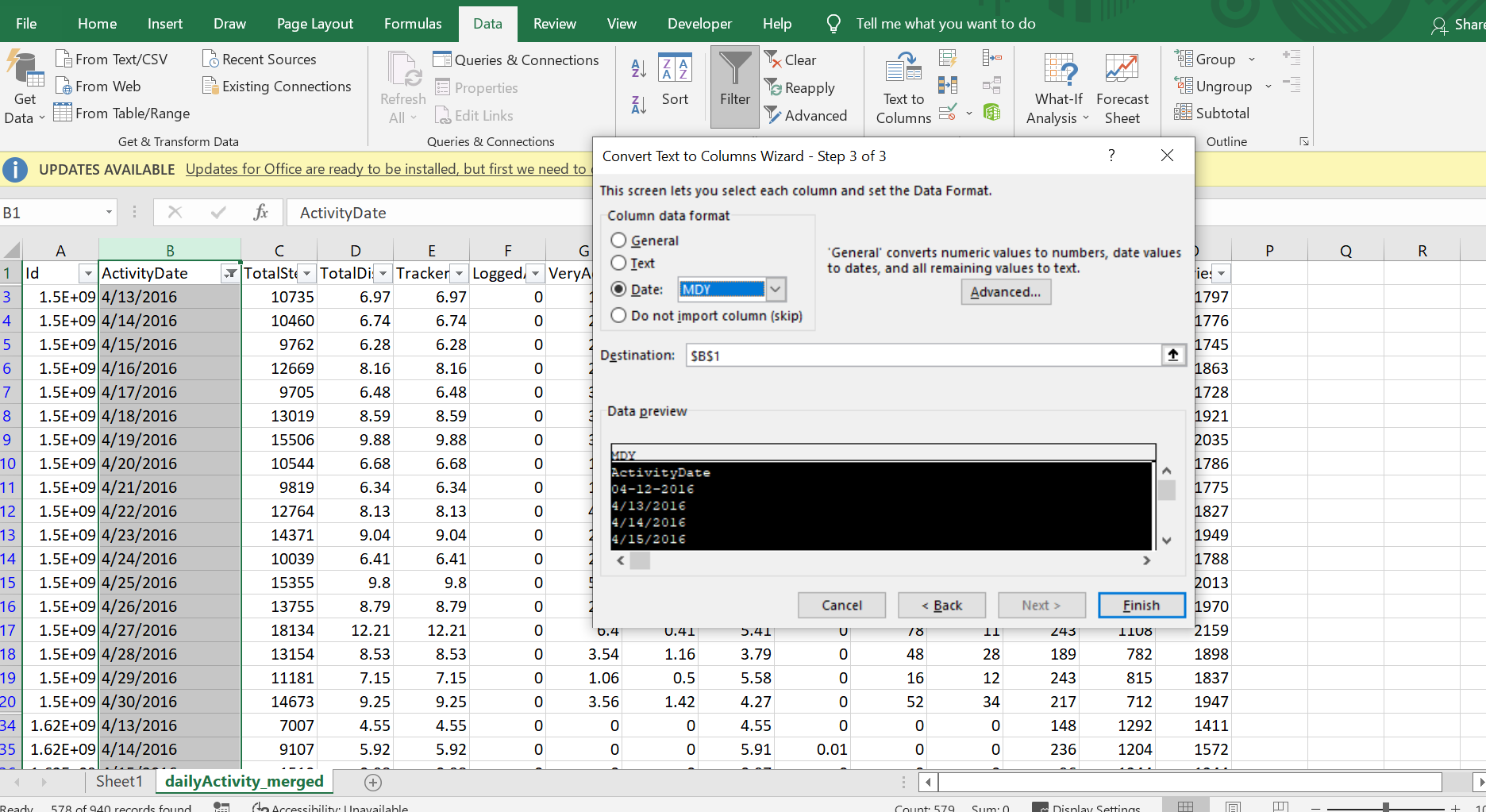
4.From the column ActivityDate, some of the dates are integers or numbers whilst others are in text or strings.

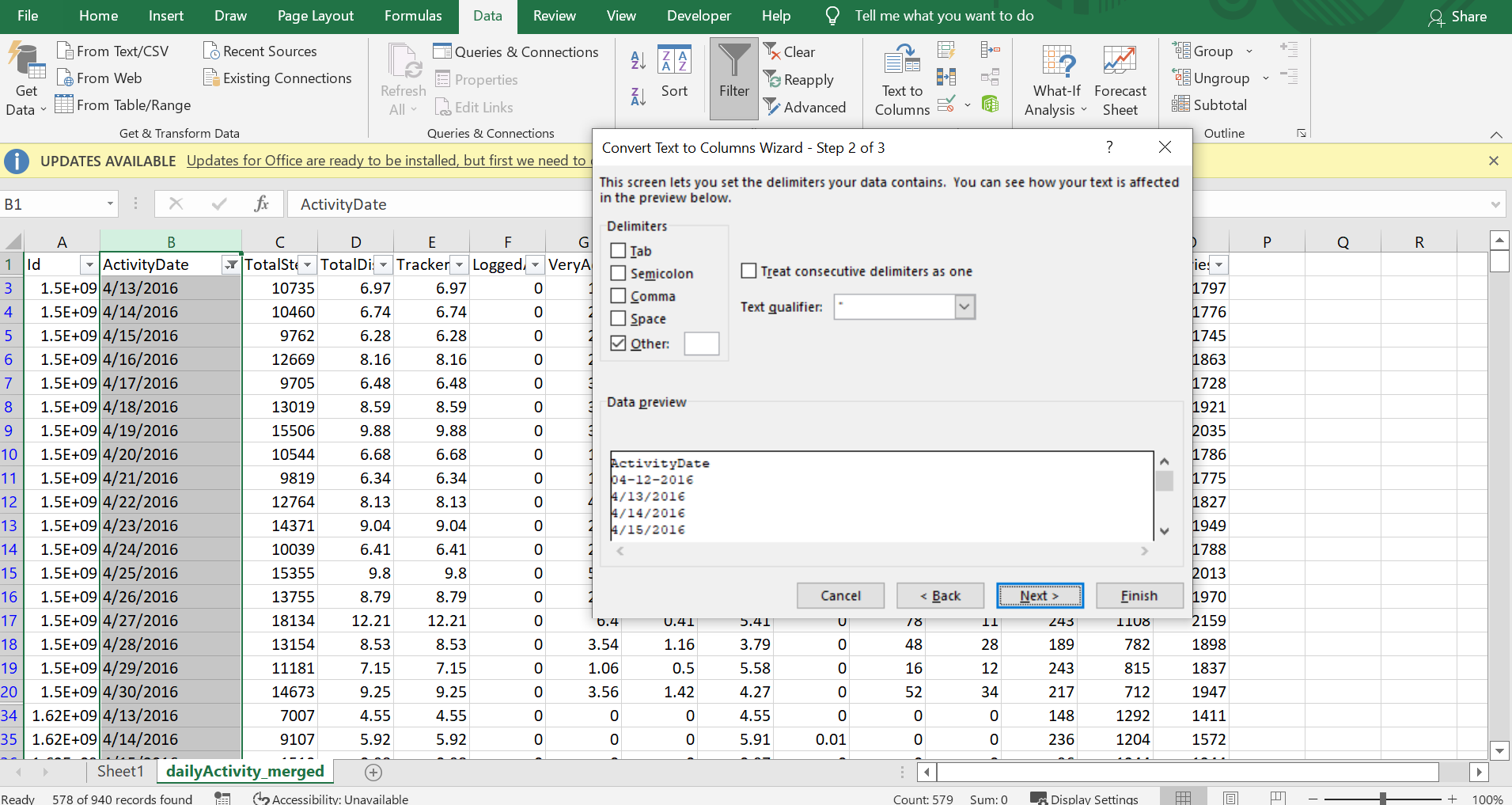
At this point we transform the raw dirty into a clean one through data manipulation.

1. Firstly date in ActivityDate column is changed from a text string to an integer.

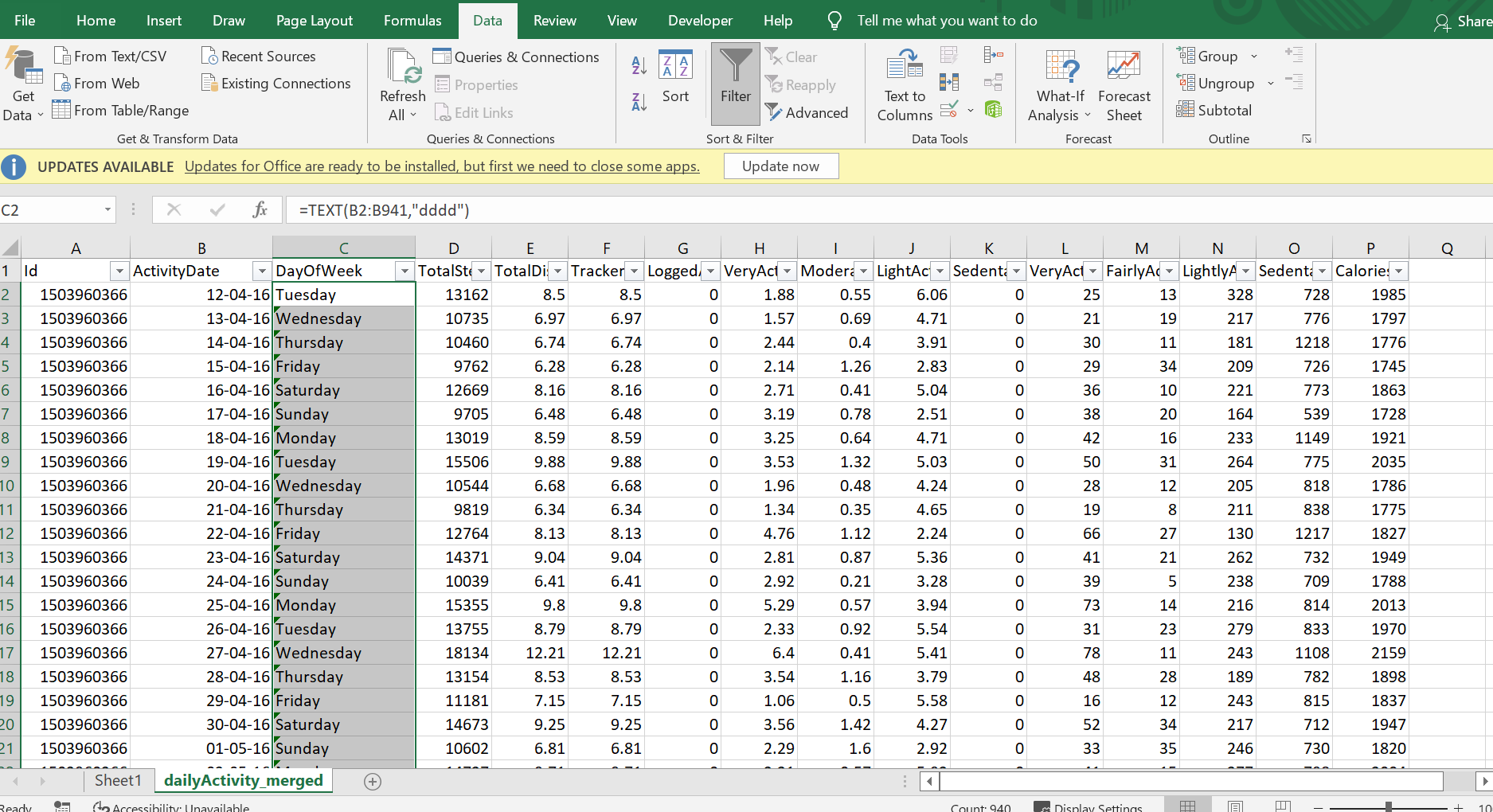




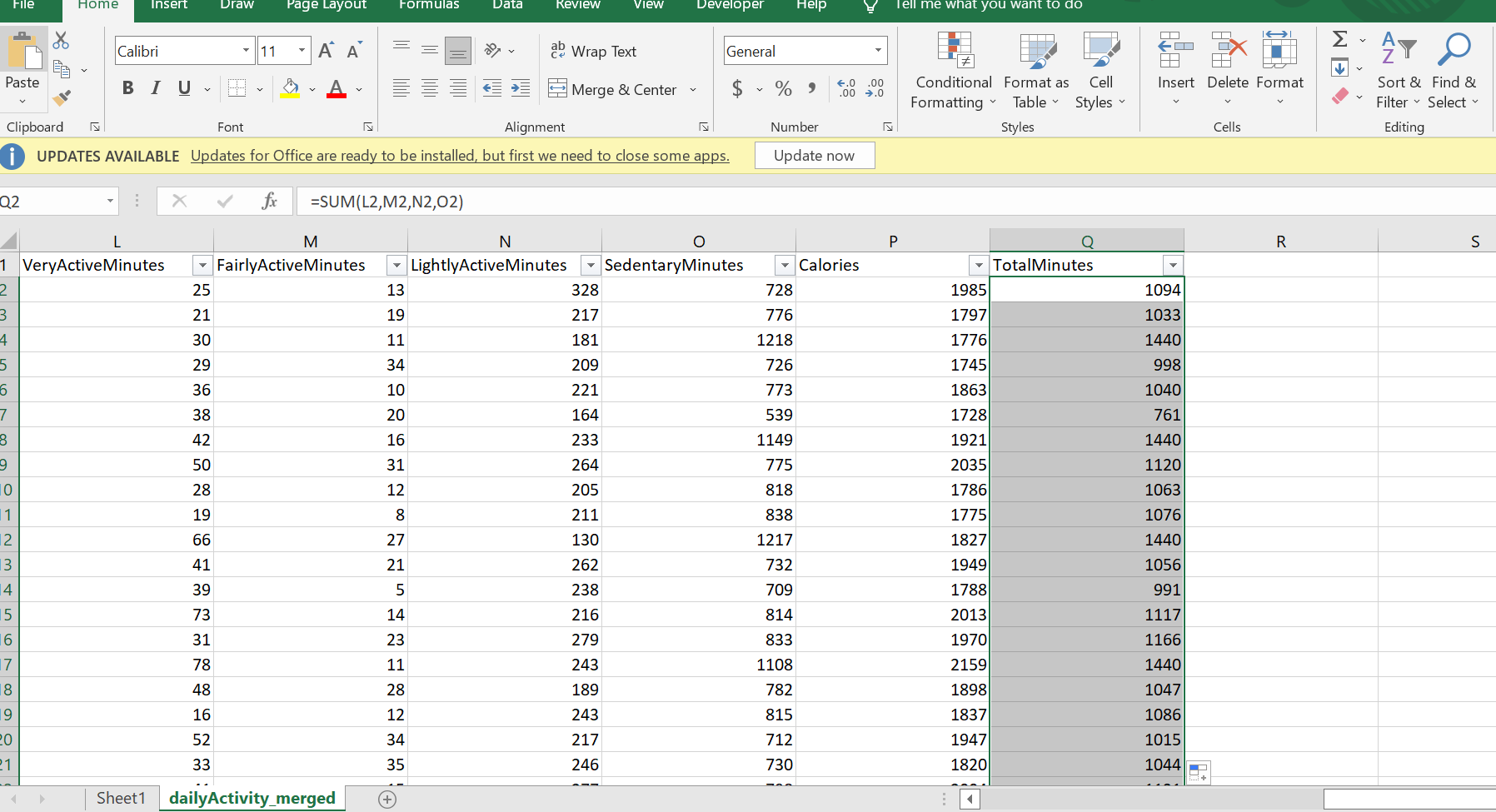




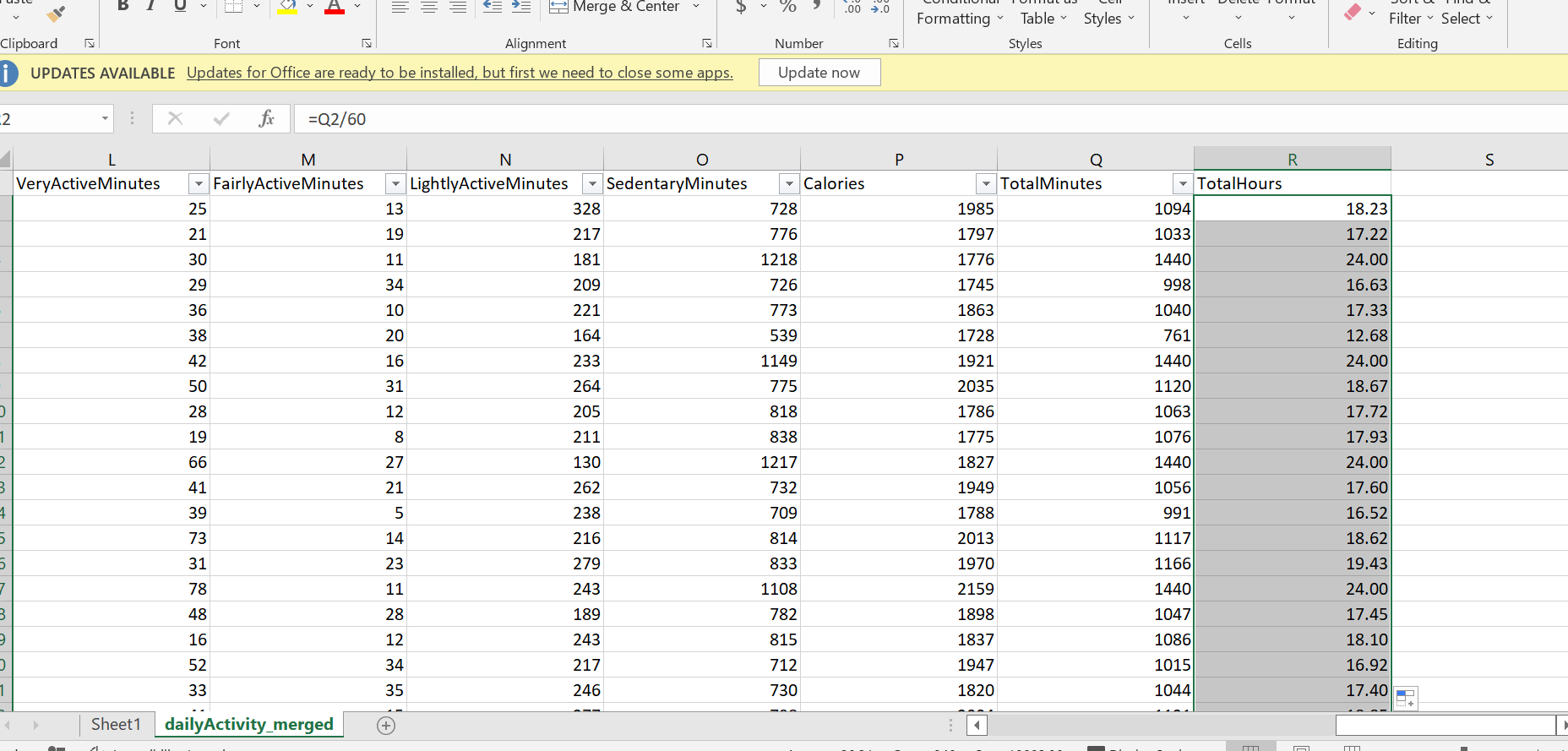
2.A new column labelled DayOfWeek will be created by transforming date in the form of day of the week for continuous analysis.



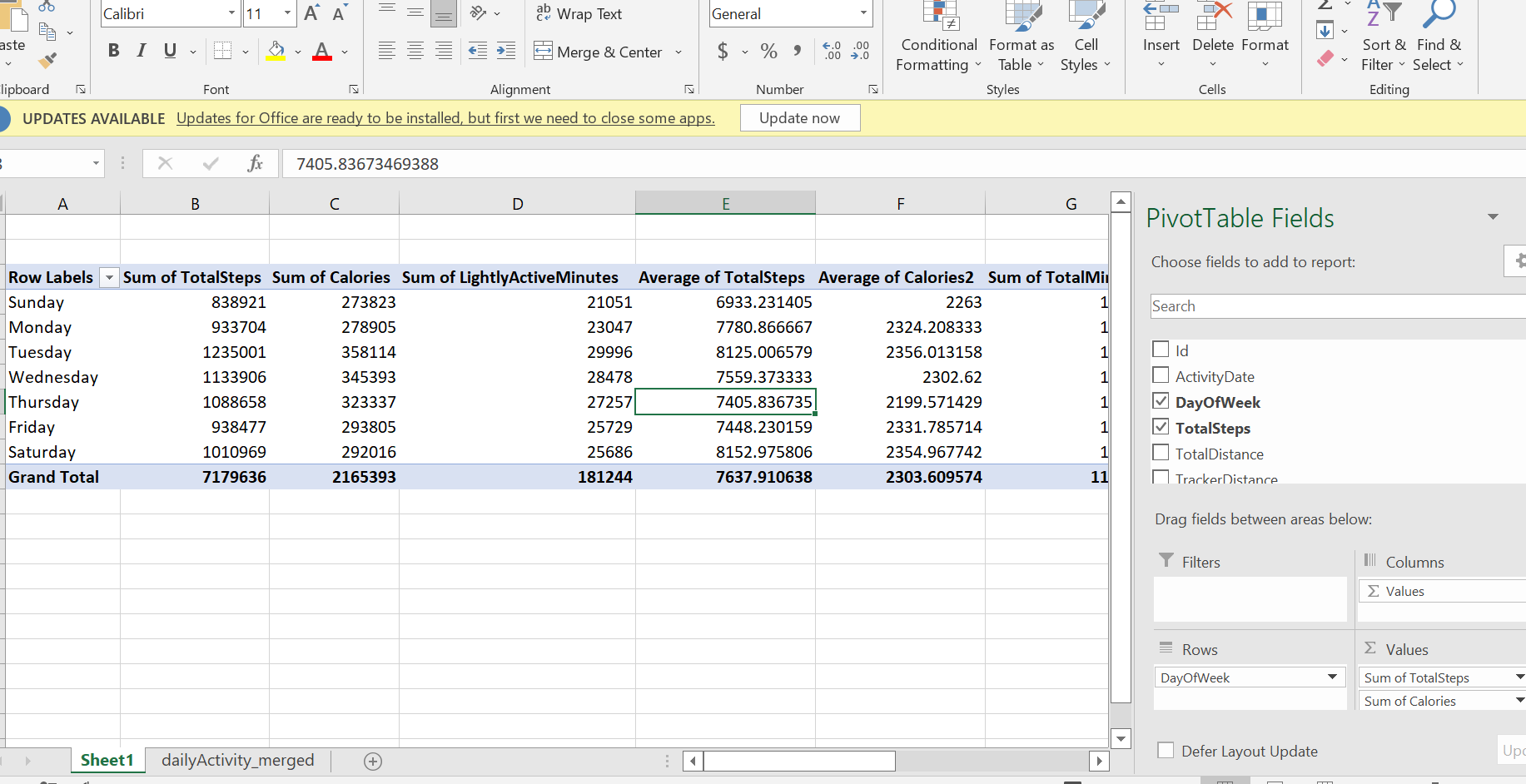
3.New column labelled TotalMins was created. This column contained the sum of the columns of VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes and SedentaryMinutes.



4.Another column is created and labelled TotalHours. This column contained integers derived from TotalMins column number of hours. This was done by dividing cells from TotalMins column by 60.



Data cleaning and manipulation is now completed. Hence, data is now readily available to be analyzed. A pivot table was created in another sheet to summarise what we want to analyse.



**ANALYZE**

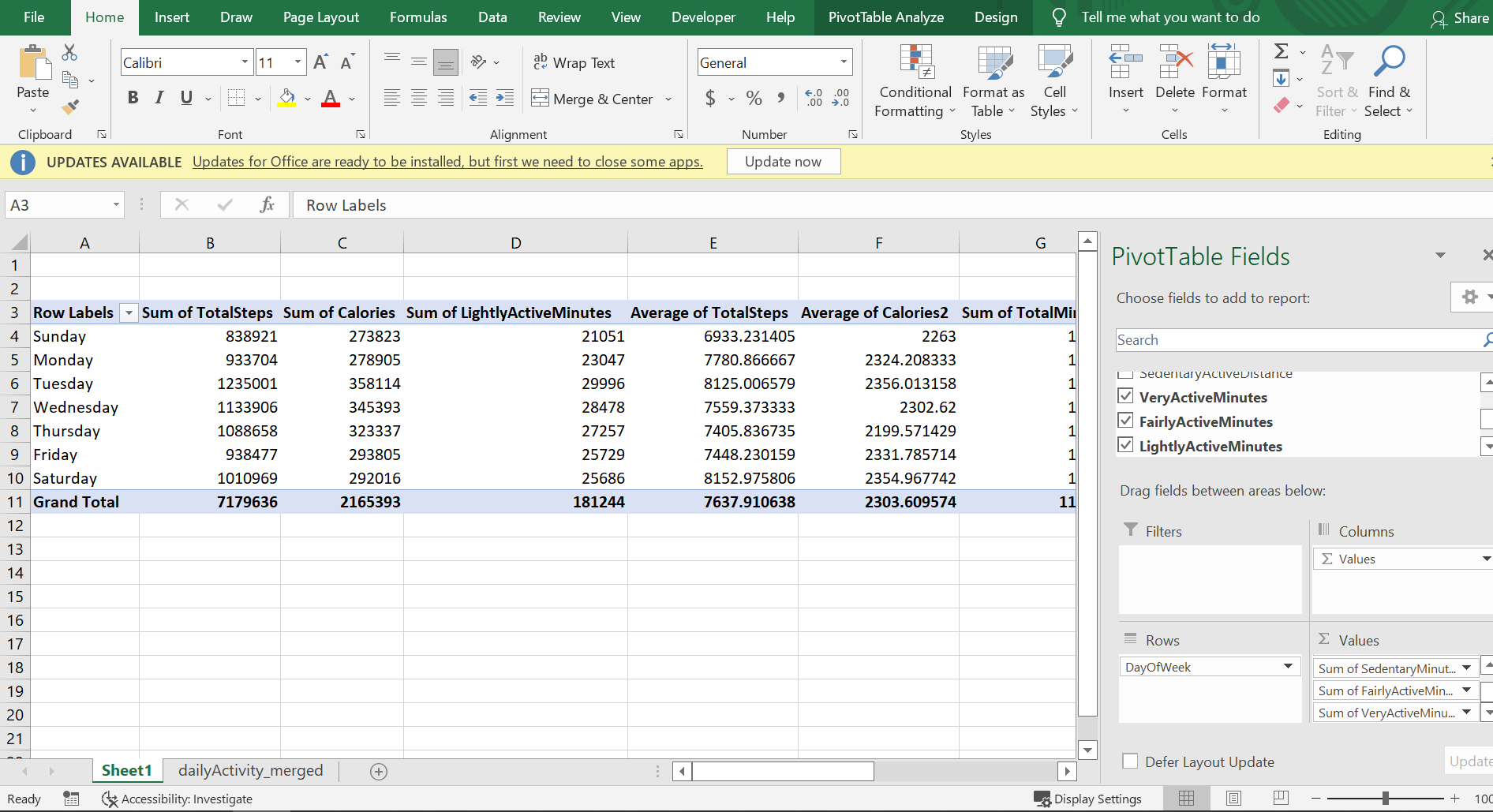
**Interpreting statistical findings**:

1. According to a health article MEDICAL NEWS TODAY at least ten thousand steps or 8 kilometers per day is needed to benefit from general health, fitness improvement and weight loss. According to our data gathered total sum of 7179636 Total steps were recorded by 33 users which averagly translates to 7637.9 steps or 5.4km only. This shows that the average steps taken is not adequate to have any health benefit from the indivuals involved in this tracking survey

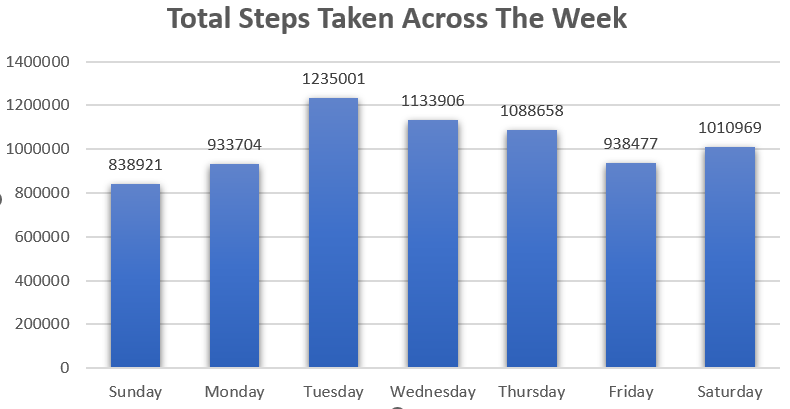
2.Sedentary track users were analyzed to be the majority group of individulas logging on with an average of 991 minutes which takes up a whooping 81% of total average minutes.

3. Several factors including as the weight, age, hormones, exercise, hormones and calorie intake affects how much calories will be burnt. These factors were not included in the data analysed hence a finite analysis cannot be deduced from th amount of calorie burned. Source: Health Line article

4.Sunday had the lowest steps and time and colouries burnt with Tuesdy being the highest.

****

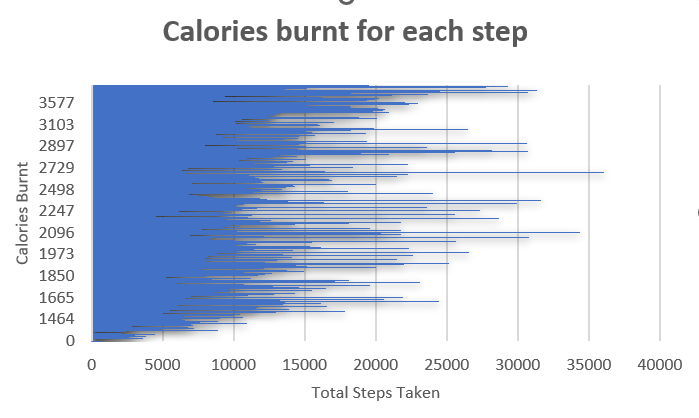
**SHARE**



Frequency of usage across the week:

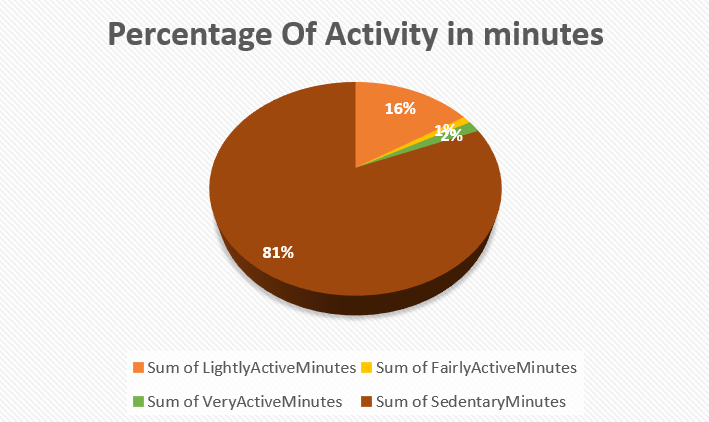
In the Bar graph above, we took a look at the total steps taken in each day across the week, this also represents the frequency of of use of the FitBit app in terms of days of the week.

1. Its evident that the Fitbit app was heavily used on Tuesdays, and its usage reduces and drops heavily on Fridays, picks back up slightly Saturday but then drops again on Sunday.
2. The weekends had an overall drop in use of the FiBit app.
3. The usage also reflects on the number of steps taken each day with Tuesday recording the highest number of steps taken.



From the clustered bar graph above, I analyzed that:

1. The intensity of calories burned is very high between the first step taken by users and the 15000th step. Burning of calories decreases after 15000th step taken.
2. It is a positive correlation.



Percentage of Activity in Minutes:

From the pie chart,

1. It indicates that users are use the FitBit app to log daily activities such as daily commute and not very active movements or running errands
2. Sedentary minutes takes the biggest part of the chart at 81%.
3. We can deduce that the FitBit App is rarely used to track fitness because of the very large amount of time sedentary minutes recorded compared to very active minutes of just 2%. This is very discouraging and disappointing because the FitBit app was creted and developed to encourage exercise and fitness.

**Act**

**Some of the trends identified from the data collected fom the FitBit Fitness App include:**

* Users rather tracked their activities during weekdays as prefferred to the weekends this is probably because the users spent more time outside their homes on weekdays for work and errands and stayed in their homes on weekends.
* Most of users representing 81% used the FitBit app to mostly track sedentary activities and not using it for tracking their excercises, fitness or health habits.

**How could these trends help influence Bellabeat marketing strategy?**

* Educational healthy style campaign can be created to pair with a point-reward incentive system. Users who are able to complete a whole week's exercise will bbe rewarded with points called Bellabeat points on products or memberships.
* Bellabeat can also use this Educational healthy style campaign to encourage users to have short active exercises during the week, longer sessions during the weekends especially on Sundays where the lowest steps were taken and most sedentary minutes were recorded.
* Bellabeat app can also create prompt notifications to encourage users to exercise and take their fitness serious on weekends.