title: "Bellabeat Case Study"

author: "Swati Garg"

date: "12/17/2021"

**About**

Bellabeat is a high-tech manufacturer of health-focused products for women. It 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.

This case study is about to find the market trends of health products and focus specifically on Bellabeat’s products and analyze smart device data to gain insight into how consumers are using their smart devices.

**ASK Phase**

**Business Task Summary**

I am analyzing one of the Bellabeat product known as Leaf , It is a 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. With this analysis I try to find the market trends and normal routine of customers with the leaf device and try to recommend some measures which will help Bellabeat in the planning of their market strategy.

**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.

**Prepare Phase**

**Data Sources**

This dataset generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Individual reports can be parsed by export session ID (column A) or timestamp (column B). Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences.

This dataset consists consist of 18 csvs , 3 are in wide form rest all are in narrow form.

I am using narrow form csv for my analysis.

**Limitations of the dataset**

1. This dataset contains the user id to identify users but it does not tell us the gender and age of the user which are the key factors of any fitness analysis.
2. Weight information varies from one day to another day.

**Data Consistency Parameters**

Besides the above mentioned limitations , the overall dataset looks good and concise :

1. Reliable: This dataset is generated by Amazon Mecahnical Turk which is quite reliable .
2. Original:Data is collected by fitness firm itself
3. Comprehensive : It consist of all information which is needed to answer the questions
4. Current: This dataset has taken between 03.12.2016-05.12.2016 which can be consider recent data source .
5. Cited: By pre checking data it confirmed it is cited and credible.

**Data Storage and analysis information**

I am using spreadsheet for storing and analyzing datasets and Tableau to create Viz for supporting my analysis .

Since the dataset is small hence I choose above tools for this case study.

**Download data**

I have downloaded dataset in zip format and extract it into a folder . It contains 18 csv files out of which I am using 7 files for analysis .

**Data Load**

1. dailyActivity\_merged.csv
2. dailyCalories\_merged.csv
3. dailyIntensities\_merged.csv
4. dailySteps\_merged.csv
5. hourlyCalories\_merged.csv
6. hourlyIntensities\_merged.csv
7. hourlySteps\_merged.csv

**Process Phase**

After loading data I will go ahead and check data organization and do cleaning process.

**Cleaning Data**

1. **Check for null values**

I applied filers to all categories and check whether there are any null values. There are multiples 0 (zero) values are present but this is correct data , I conclude it after checking related data from other files.

1. **Check for duplicate values**

I select columns and then go to Data tab and select remove duplicates .

There were no duplicate records present in any of the files.

1. **Check for column data type**

Date column has wrong data type so I change it to date data type through

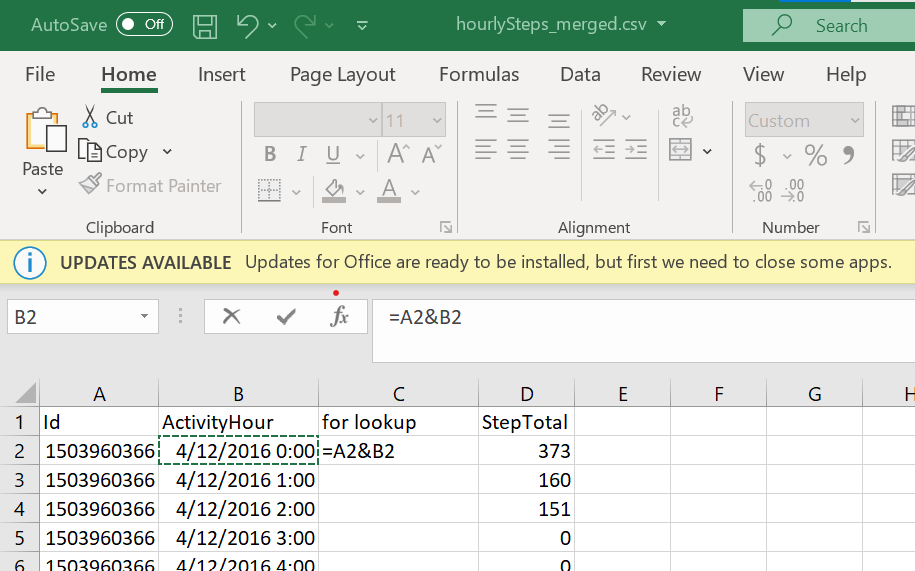
Select whole column>right click>select format cells>date

And then I checked all files and id the same for daily files but for hourly files there was time stamp as char instead of datetime so I changed the datatype.

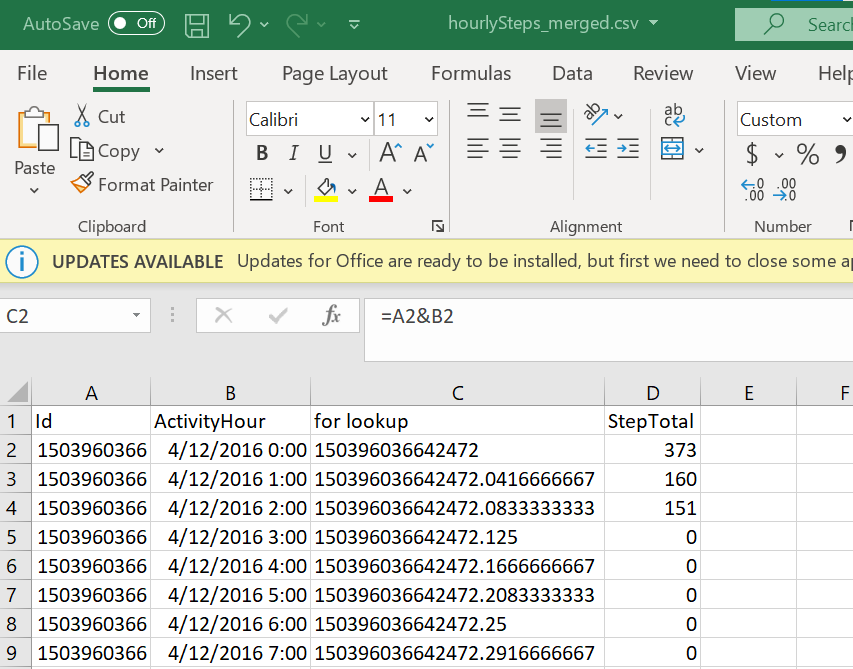
Select whole column>right click>select format cells>Custom>m/d/yyyy h:mm

**Reorganize data**

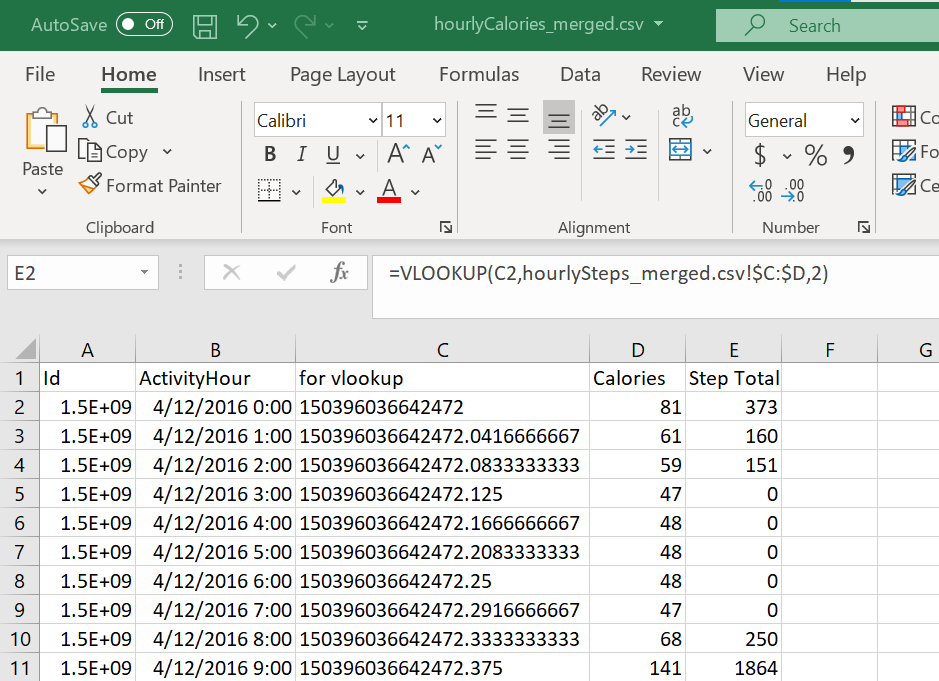
1. After doing above cleaning process I saved the daily data into a new file named **DailyActivity.csv**
2. I need to combine 3 files and save it to another file , for combining I will use vlookup on 2 categories Id and Activity Hour.
3. I created a column (for vlookup) which is the combination of two columns Id and Activity Hour in all 3 files.



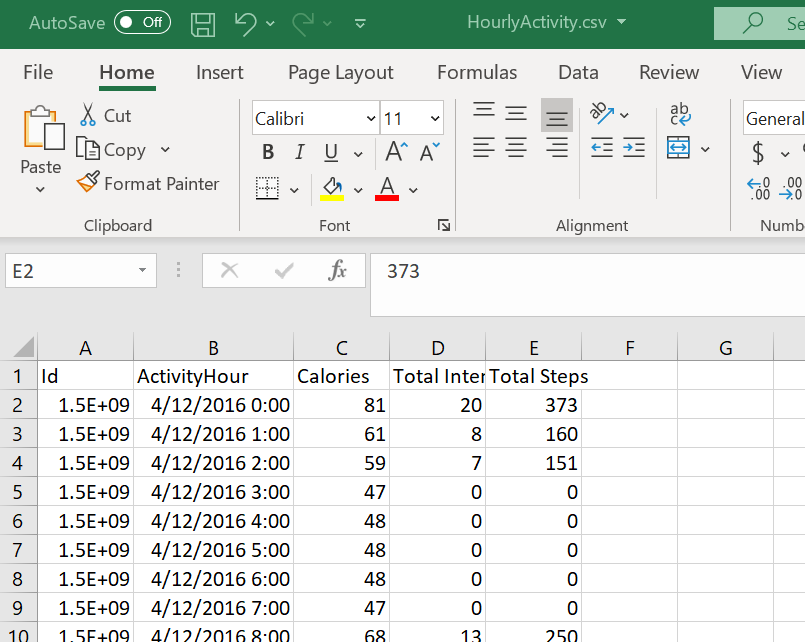
After applying formula below are the results



Now I will apply lookup and combine 3 files together



I saved final results in a file “HourlyActivity.csv”

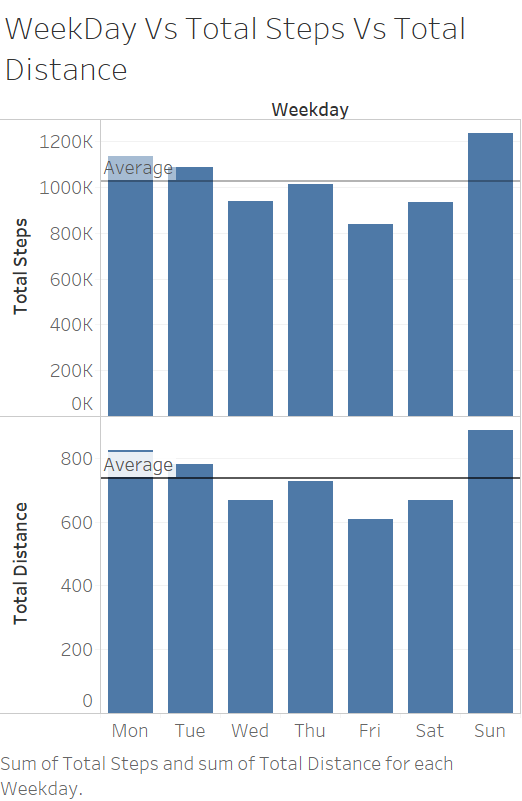


Now both files are ready for analysis .

I import both files in Tableau for viz

**Analysis Phase**

1. **Weekday Vs Total Steps Vs Total Distance**



As per the above graph people cover max steps and distance on Sunday . Since it is the only day when they are free and want to spend more time to their fitness.

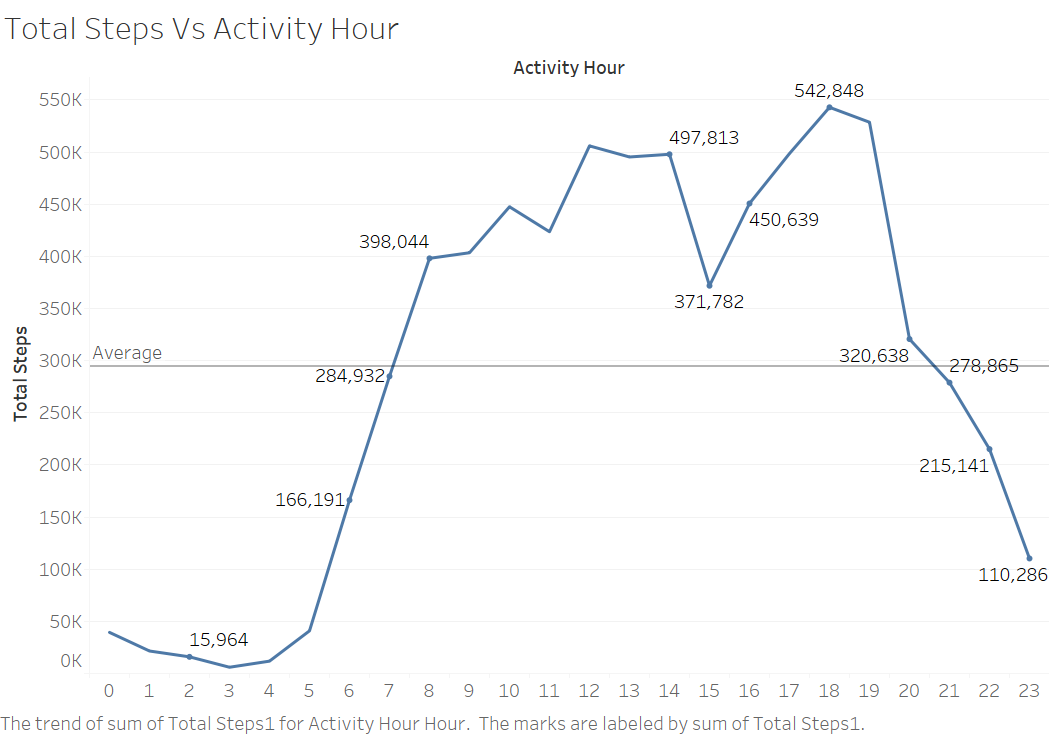
1. **Weekday Vs Total Steps Vs Calorie**



This graph clearly shows that the total steps and most calories burnt is on Sunday.

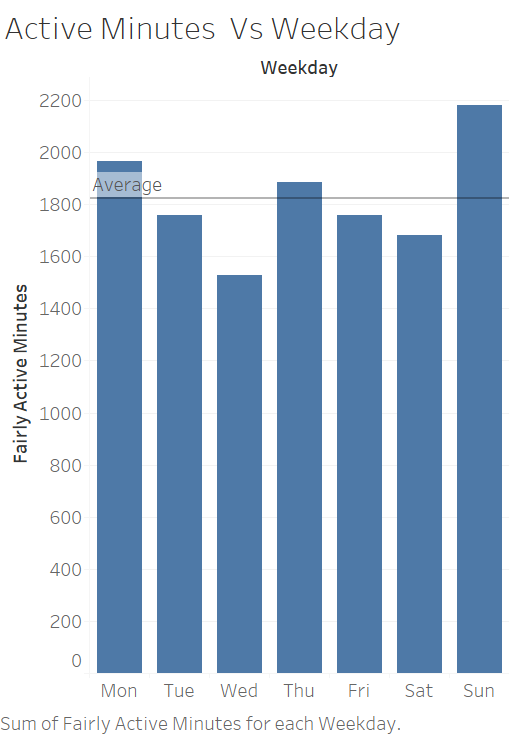
Besides calorie burnt depends on Active time as well the more the active time the more Calorie burnt.

1. **Total Steps Vs Activity Hour**



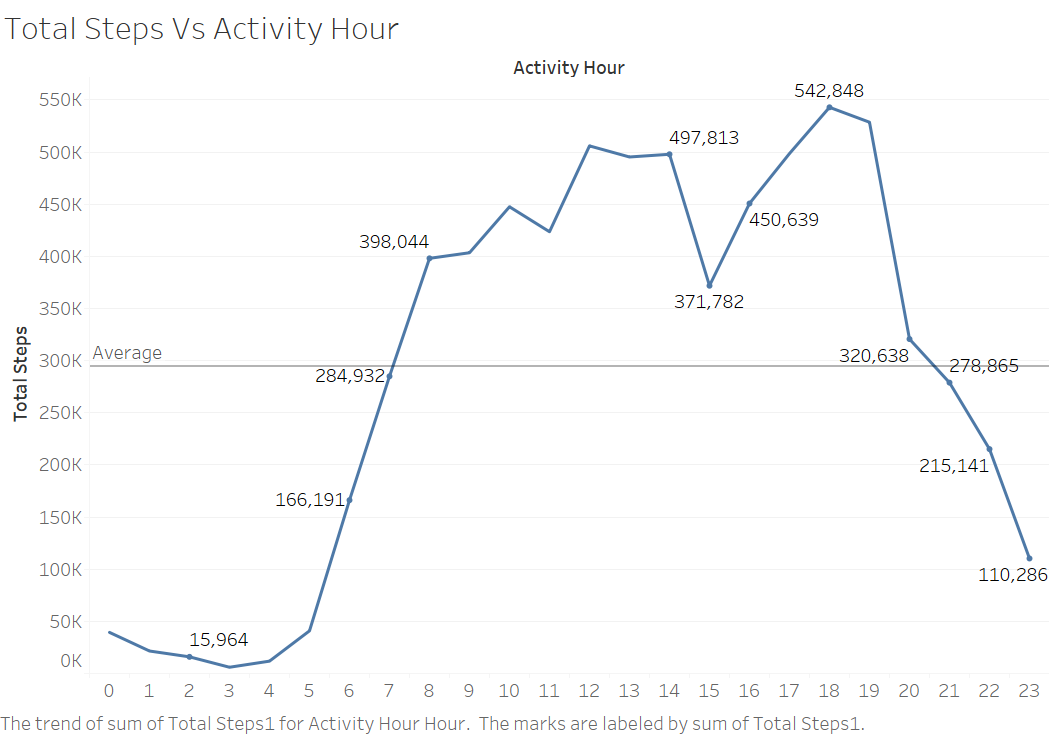
According to this viz , people likes evening walks as compared to morning walks .Evening at 6PM is the highest steps count hour.

1. **Active Minutes Vs Weekday**



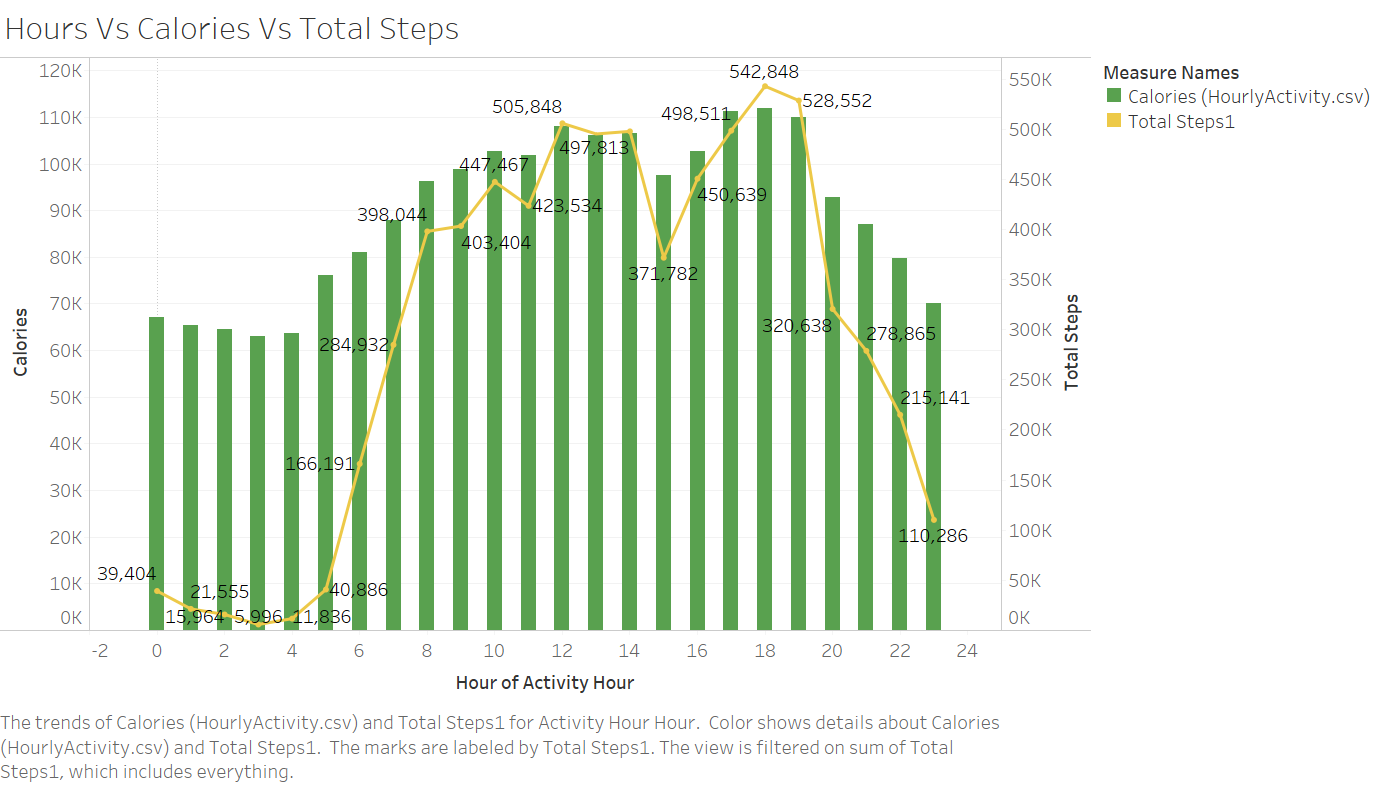
According to this graph People mostly active on Sunday but Monday is also above average due to first day of week at work.

1. **Total Steps Vs Activity Hour**



According to this graph most steps people covered are at 6PM , evening time of the day . It clearly suggest people often likes to take evening walks then morning walks .

1. **Hour of the day Vs Calorie Burnt Vs Total Steps**



This graph is showing the relationship between calorie burnt , total steps at particular hour of the day.

We can clearly see that Calorie burnt and total steps taken have no direct relationship . Calorie burnt is totally depend on the activity on that hour.

**Recommendations**

**Recommendations Based on Analysis**

1. **Evening time notifications**

We have seen through graphs that users are more active during evening hours so we can send notifications about their total steps taken and total target steps.

This makes them aware about their daily health routine.

1. **Integrate timing for other activities**

We have seen even that active hours and steps counts are not matching that means we need to upgrade our devices and add some more features including various activities like yoga, playing outdoor games , reading books .