Proyecto análisis de datos Bellabeat

Bellabeat is a manufacturer of high technology devices focused on women's health. Despite being a small company, Bellabeat has the potential to become a bigger participant of the global smart device market.

Ask

Determine usage trends to create a marketing strategy focusing on one product otr function in order to establish Bellabeat as an important actor in the global market.

Prepare

The data used for this study was provided by Möbius, generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. In total, thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences. Data was stored in Google Sheets and Bigquery in a long format, it was also checked to see if any confidential information was stored in the dataset.

It was determined that the data concerning daily measurements would be used for analysis, in order to get a broader understanding of the participants behavior. Three datasets were used: Daily Activities, Sleep Day y WeightLog, in order to determine the usage of a daily activity tracker, sleep hours tracker and IBM/weight tracker.

Procesar

Data processing was performed simultaneously on Google Sheets and SQL. Despite having a small dataset, SQL was chosen due to combine the best parts of each data processing tools.
Data cleaning was performed on all data. The first steps were realized using Google Sheets, this included a general revision of the full dataset in order to get familiarized with the data before any operation was performed, eliminating duplicate data in every table, splitting the date-hour columns into separate columns and eliminating the hour column (due to it's nature and the qualities of the data, it was irrelevant to the proposed analysis) and checking the format and type of each data. Furthermore, some columns were dropped from each table due to the lack of further information or the redundancy of their presence. Once this first stage of data cleaning was finished, the results were three tables: **dailyActivity, weightLog, sleepDay**.

The second stage of the data cleaning process was performed using SQL. This mainly included double checking the format and type of each data, and the combination of the dailyActivity and sleepDay data into one table named **clean_data**.

```sql
CREATE TABLE Data.clean_Data as
SELECT
 act.Id,
 act.ActivityDate as Day,
 act.TotalSteps as Steps,
 act.Calories,
 act.LightlyActiveMinutes as LightActiveMin,
 act.FairlyActiveMinutes as FairActiveMin,
 act.VeryActiveMinutes as VeryActiveMin,
 act.SedentaryMinutes as SedentaryMin,
 sleep.TotalMinutesAsleep AS MinutesAsleep,
 sleep.TotalTimeInBed AS MinutesInBed
FROM `finalproject-358200.Data.dailyActivity` AS act
JOIN `finalproject-358200.Data.sleepDay`AS sleep
ON (act.Id=sleep.Id
AND act.ActivityDate=sleep.SleepDay)
ORDER BY Id
```

In order to perform analysis based on the averages of each day of the week, the Activity table was created. This new table would allow to analyze the behavior of all participants along the week in order to determine the most intensive days in terms of calories burned and steps taken).

```sql
CREATE TABLE Data.Activity AS
SELECT
Day,
ROUND(AVG(Calories), 1) AS avgCalories,
ROUND(AVG(Steps), 1) AS avgSteps
 FROM `finalproject-358200.Data.clean_Data`
 GROUP BY Day
 ORDER BY avgSteps DESC
```

Final data processing was performed by creating the Averages table, which would allow analysis on activity intensity and correlations between activity, calories, steps taken and the total time asleep for the participants.

```sql
CREATE TABLE Data.Averages AS
SELECT
Id,
ROUND(AVG(LightActiveMin),2) AS avgLightMin,
ROUND(AVG(FairActiveMin),2) AS avgFairMin,
ROUND(AVG(VeryActiveMin),2) AS avgVeryMin,
ROUND(AVG(SedentaryMin),2) AS avgSedentaryMin,
ROUND(AVG(Steps),2) AS avgSteps,
ROUND(AVG(Calories),2) AS avgCalories,
```

```
ROUND(AVG(MinutesAsleep),2) AS avgSleepMin
FROM
Data.clean_Data
GROUP BY
Id
```

Analyze & Share

The first step in our analysis consisted in finding the number of users that utilized each of the three functions.

| Function	Activity Tracker	Sleep Tracker	Weight Tracker
Total Users	33	24	8

With this information, we can assume that the less used function is the weight tracker, while the activity tracker and sleep tracker have a similar number of users. However, due to the limited amount of data present in our dataset, it is recommended to conduct a new study to gather new information. Due to these limitations, the information of the weight tracker is dropped from the data, to avoid misconceptions about the data, as it is considered that the sample was very small.

Average Calories & Steps by day of the week
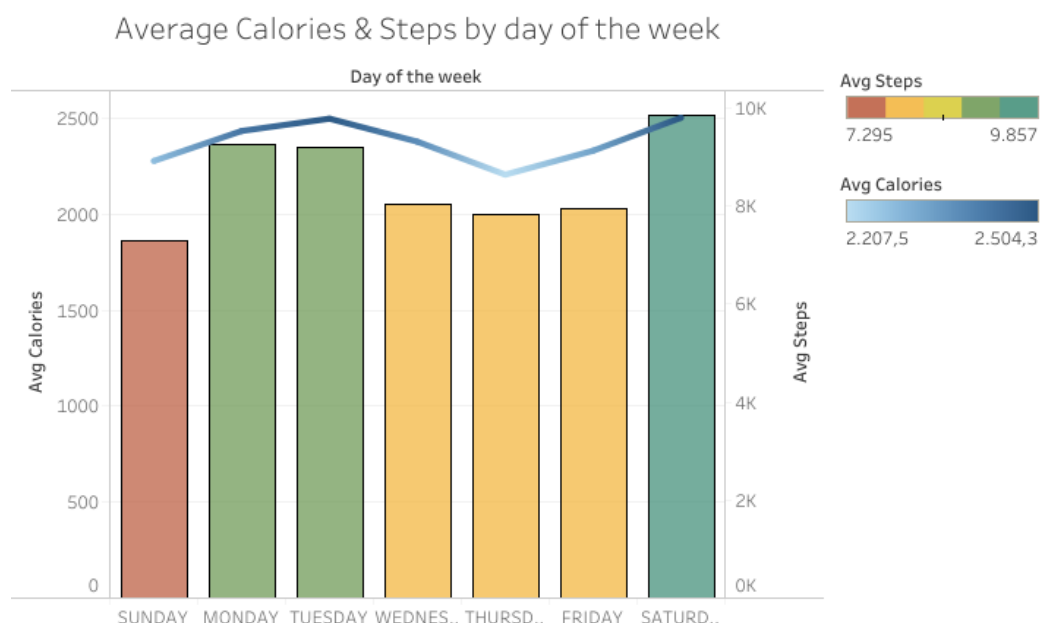
Based on the analysis of the calories and steps taken in average on each day of the week, we can identify a decreasing tendency of both calories and steps taken along the week, being the lowest point on sundays and the highest point on saturdays and mondays.

This tendency may be influenced by many different factors such as time availability and motivations throughout the week, as it´s common to start the week with a high motivation, only to find it decreased on consecutive days to do work stress or time consuming activities.

In terms of health, the World Health Organization recommends taking between 7,000 and 8,000 in order to stay healthy. While on average participants seem to have a good step count, Sundays and days in the middle of the week should be targeted in order to increase the usage of the smart devices.
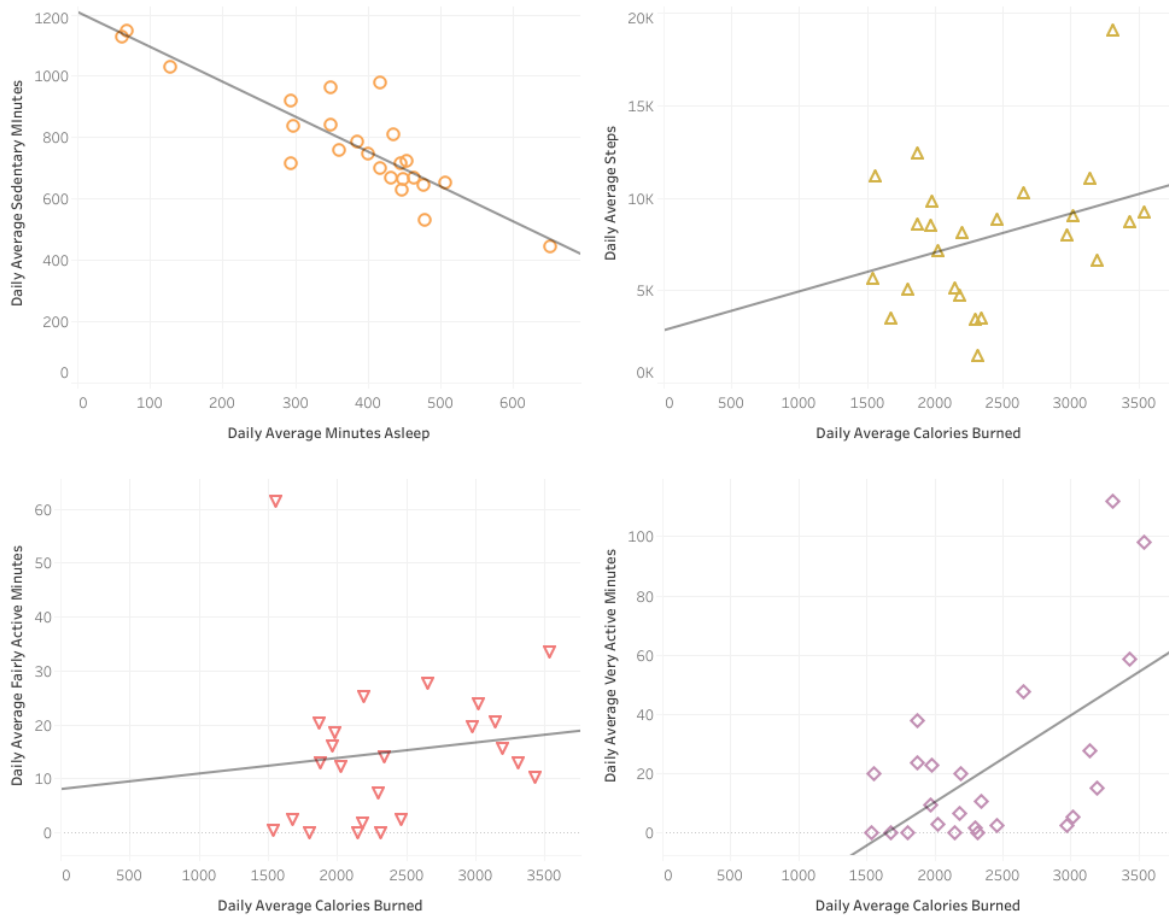
Analysis performed in the resulting dataset shows a clear relation between the daily average sedentary minutes and the average minutes asleep. The more active a person is (and in consequence, the less sedentary minutes per day), the sleep minutes increases, reaching between 6 and 8 hours of sleep, as recommended by the World Health Organization. This result should be taken as a marketing strategy, as being a more active person brings benefits to the daily rest.

Average calories burned and average steps taken show a little correlation between them. It's possible that the more steps taken, the more calories burned. Further analysis on this topic was performed via correlation of average fairly active minutes and very active minutes with the average calories burned. Out of the two, very active minutes show a better correlation with the calories burned, however it is recommended to make analysis with a bigger dataset in order to reassure this findings.

Act

The first recommendation to be made is based on the sample used on this analysis. It is highly recommended to reconduct the analysis when the dataset is increased. Based on the analysis made on the available data, we find that products that measure daily activity (such as the leaf and time products by Bellabeat) show a greater number of users. Due to this, in the first steps of the company should focus in establishing this two products in the global market, before expanding into more specialized products (such as the spring bottle).

Recommendations for the leaf and time marketing campaigns:
- Focus on improving sleep quality and establish the relation between decreasing sedentary time and the significant increase in the time asleep.
- Highlight the benefits of taking between 7,000 and 8,000 steps on average a day. Ads should have a higher programmation in the middle of the week and sundays, in order to improve quality of life.
- Highlight the relation between very active minutes a day and the total calories burned. Measurement of both categories by the Bellabeat products can help people with exercise schedules.