**Google Data Analytics Capstone Project**

The case study follows the six step of Data Analysis process:

INTRODUCTION:

Bellabeat is a high-tech company that manufactures health-focused smart products especially for women. The company has 5 focus products: bellabeat app, leaf, time, spring and bellabeat membership. 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. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

My Role: I am a junior data analyst working on the marketing analyst team at Bellabeat who have been asked to analyze smart device usage data to gain insight into how consumers use non-Bellabeat smart devices and provide recommendations.

ASK

* Business Goal: Analyze Fitbit data to identify trends on a single smart device usage and recommend marketing strategies for Bellabeat to become a global player.
* Stakeholders:
* Urška Sršen - Bellabeat cofounder and Chief Creative Officer
* Sando Mur - Bellabeat cofounder and key member of Bellabeat executive team
* Bellabeat Marketing Analytics team

PREPARE

The data source used for analysis in this case study is FitBit Fitness Tracker Data.  This dataset is stored in Kaggle and was made available for use by **Möbius**. Dataset in available on Kaggle as a public. The owner has dedicated the work to the public domain by waiving all his or her rights to the work worldwide under copyright law.

As per ROCCC approach:

* **Reliability and Original**: The dataset contains 18 CSV files. The data is from thirty Fitbit users who consented to the submission of personal tracker data, generated from a distributed survey via Amazon Mechanical Turk.
* **Comprehensive**: The data includes minute-level output for physical activity, heart rate, and sleep monitoring. While the data tracks many factors in the user activity and sleep, but the sample size is small, and most data is recorded during certain days of the week.
* **Current**: Data is from April 2016 to May 2016 which is not current and hence may not reflect the current user behaviour trends.
* **Cited**: Unknown

**Limitations**

* Data provides information of only 30 users and without any information about the demographics we cannot be sure if the data sample is representative of a whole population.
* Upon further investigation it was found that the set has 33 user data from daily activity, 24 from sleep and only 8 from weight. Some users did not record their data for tracking daily activity and sleep.
* The data does not have gender descriptions. As the Bellabeat focuses on women specific products, dataset representing women population would have been more relevant.
* Dataset is outed and records survey data only for 3 months. Most of the datapoints are incomplete and hence may not be comprehensive enough to form an accurate analysis.

PROCESS

Below I will describe step-by-step the process I used to for this project.

1. Examined the data for duplicates and error. Merged and all hourly activity in a single spreadsheet using Power Query
2. Uploaded data on MySQL and corrected datatypes
3. Created dashboard to share my review

Note: Due to small sample size, we will not be using weight and heart rate for analysis

Overview: I will first clean the data in Excel and then use MYSQL to process and transform the data. Finally, I will create a dashboard in Tableau to share my review and feedback with stakeholders.

Initial analysis shows that out of 33 candidates who used Fitbit to track activities of which only 8 used it to track weight and 14 to track heart rate respectively. Hence, we will be skipping them from further analysis.

All respondents used there tracking devices to record

* Total Steps
* Total calories burnt
* Distance and time covered while doing
  + Very active activities
  + Moderate active activities
  + Lightly active activities
  + Sedentary activities
* 24 (73%) respondents used devices to track sleeping patterns for an average of 17 days.
* 14 (42%) respondents recorded their heart rate while only 8 (24%) recorded their weight
* Checking activities and there average

--Calculating

--Number of users tracking physical activities and daily averages

SELECT

COUNT(DISTINCT Id) AS users\_tracking\_activity,

AVG(CAST(TotalSteps AS decimal)) AS average\_steps,

AVG(CAST(TotalDistance AS decimal)) AS average\_distance,

AVG(CAST(Calories AS decimal)) AS average\_calories

FROM

bellabeat.dbo.daily\_activity

--Number of users tracking heart rate

SELECT

COUNT(DISTINCT Id) AS users\_tracking\_heartrate,

AVG(Value) AS average\_heartrate,

MIN(Value) AS minimum\_heartrate,

MAX(Value) AS maximum\_heartrate

FROM

bellabeat.dbo.heartrate\_seconds

--Number of users tracking weight

SELECT

COUNT(DISTINCT Id) AS users\_tracking\_weight,

AVG (CAST(WeightKg AS decimal)) AS average\_weight,

MIN(CAST(WeightKg AS decimal)) AS minimum\_weight,

MAX(CAST(WeightKg AS decimal)) AS maximum\_weight

FROM

bellabeat.dbo.weightLogInfo

--Number of users tracking sleep

SELECT

COUNT(DISTINCT Id) AS users\_tracking\_sleep,

AVG(TotalMinutesAsleep)/60.0 AS average\_hours\_sleep,

MIN(TotalMinutesAsleep)/60.0 AS minimum\_hours\_sleep,

MAX(TotalMinutesAsleep)/60.0 AS maximum\_hours\_sleep,

AVG(TotalTimeInBed)/60 AS average\_hours\_inbed

FROM

bellabeat.dbo.sleep\_day

|  |  |  |  |
| --- | --- | --- | --- |
| User tracking activity | Average Steps | Average Distance | Average Calories |
| 33 | 7638 | 5.48 | 2303 |

|  |  |  |  |
| --- | --- | --- | --- |
| User tracking heartrate | Average heartrate | Minimum heartrate | Maximum Heartrate |
| 14 | 77 | 36 | 203 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| User tracking sleep | Average hours of sleep | Minimum hours of sleep | Maximum hours of sleep | Average hours in Bed |
| 24 | 6.9 | 0.9 | 13.2 | 7 |

* Calculating to check if there is any correlation between the activities and days

SELECT

ActivityDate,

AVG (Bellabeat.dbo.daily\_activity.TotalSteps) AS Steps\_Average,

AVG (Bellabeat.dbo.daily\_activity.SedentaryMinutes) AS Sedentary\_Average,

AVG(Bellabeat.dbo.daily\_activity.LightlyActiveMinutes) AS Lightly\_Average,

AVG (Bellabeat.dbo.daily\_activity.FairlyActiveMinutes) AS Fairly\_Average,

AVG(Bellabeat.dbo.daily\_activity.VeryActiveMinutes) AS VeryActive\_Average

FROM

Bellabeat.dbo.daily\_activity

WHERE

Bellabeat.dbo.daily\_activity.TotalSteps <> 0

GROUP BY ActivityDate

ORDER BY Steps\_Average;

There is no visible correlation between the number of steps, activities, and weekends/weekdays.

Cleaning daily activity and calculating average

Drop table

IF EXISTS(SELECT \*

FROM bellabeat.dbo.daily\_activity\_cleaned)

DROP TABLE bellabeat.dbo.daily\_activity\_cleaned

--Create table to convert datatypes

CREATE TABLE bellabeat.dbo.daily\_activity\_cleaned

(Id FLOAT, ActivityDate DATETIME2(7), TotalSteps INT, TotalDistance FLOAT, LoggedActiveDistance FLOAT,VeryActiveDistance FLOAT, ModeratelyActiveDistance FLOAT,

LightlyActiveDistance FLOAT, SedentaryActiveDistance FLOAT, VeryActiveMinutes INT, FairlyActiveMinutes INT, LightlyActiveMinutes INT, SedentaryActiveMinutes INT,

Calories FLOAT)

INSERT INTO bellabeat.dbo.daily\_activity\_cleaned

(Id, ActivityDate, TotalSteps, TotalDistance, LoggedActiveDistance,VeryActiveDistance, ModeratelyActiveDistance,

LightlyActiveDistance, SedentaryActiveDistance, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryActiveMinutes,

Calories)

SELECT

Id,

ActivityDate,

CAST(TotalSteps AS FLOAT) AS TotalSteps,

CAST(TotalDistance AS FLOAT) AS TotalDistance,

CAST(LoggedActivitiesDistance AS FLOAT) AS LoggedActiveDistance,

CAST(VeryActiveDistance AS FLOAT) AS VeryActiveDistance,

CAST(ModeratelyActiveDistance AS FLOAT) AS ModeratelyActiveDistance,

CAST(LightActiveDistance AS FLOAT) AS LightActiveDistance,

CAST(SedentaryActiveDistance AS FLOAT) AS SedentaryActiveDistance,

CAST(VeryActiveMinutes AS FLOAT) AS VeryActiveMinutes,

CAST(FairlyActiveMinutes AS FLOAT) AS FairlyActiveMinutes,

CAST(LightlyActiveMinutes AS FLOAT) AS LightlyActiveMinutes,

CAST(SedentaryMinutes AS FLOAT) AS SedentaryActiveMinutes,

CAST(Calories AS FLOAT) AS Calories

FROM bellabeat.dbo.daily\_activity

31 days of activities recorded

|  |  |  |  |
| --- | --- | --- | --- |
| Avg  Very Active Minutes | Avg  Fairly Active Minutes | Avg  Lightly Active Minutes | Avg  Sedentary Active Minutes |
| 21 | 13 | 192 | 991 |

* Determine when users were mostly active

--Determine when users were mostly active

SELECT\*

FROM bellabeat.dbo.hourly\_activity

SELECT

DISTINCT (CAST(ActivityHour AS TIME)) AS activity\_time,

AVG (TotalIntensity) OVER (Partition By DATEPART (HOUR,ActivityHour)) AS average\_intensity,

AVG(METs/10.0) OVER (Partition By DATEPART (HOUR, ActivityHour)) AS average\_METs

FROM

bellabeat.dbo.hourly\_activity AS hourly\_activity

JOIN bellabeat.dbo.minuteMETsNarrow AS METs

ON

hourly\_activity.Id = METs.Id AND

hourly\_activity.ActivityHour = Mets.ActivityMinute

ORDER BY

average\_intensity DESC



It can be seen from the table that the users were most active in evening between 5 to 7 PM.

ANALYSE

Key findings

* Most respondents used Fitbit to track steps and calories
* Not every user is using heartrate tracker, sleep tracker and weight log
* Some users didn’t track their steps every day
* Participants had 9 hours of sedentary time daily
* Most of the participants prefer light activity
* Participants were most active in the evenings, around 17:00 to 19:00 and fairly active in the afternoons
* No visible correlation between weekends/weekdays and number of steps

Recommendations

* A survey covering a larger population especially women respondents will help to get a better idea about the popularity of Fitbit
* Questionnaire like what is the best feature of Fitbit? What features needs improvement? How would the rate easy of using Fitbit?
* Above survey results will help to understand why the participants didn’t track their steps, heartrate, sleep, and weight every day.
* Bellabeat can raise campaigns to spotlight how increase in sedentary time can pose a risk to their health
* They can highlight the benefits of 8k+ steps per day and how they can be fit by increasing non-exercise activity thermogenesis **(NEAT)**