INTRODUCTION

This case study is the Capstone Project of **Google Data Analytics Professional Certificate**. **The 6 steps of Data Analysis** are used to present this analysis.

Title: **Bellabeat Case Study**

Author: Monish

Date: 21 August 2024

STEP 1: ASK

1.0 Background

Bellabeat, founded in 2013, is a leading manufacturer of health-focused smart products designed specifically for women. The company's mission is to inspire and empower women by providing them with valuable insights into their health and daily habits.

Urška Sršen, the co-founder and Chief Creative Officer of Bellabeat, believes that analysing data from other health-tracking devices, such as Fitbit, could uncover valuable insights and opportunities for Bellabeat's growth.

1.2 Business Task

Analyse Fitbit Fitness Tracker data to understand consumer usage patterns. The goal is to uncover trends and insights that can inform Bellabeat's marketing strategy.

1.3 Business Objectives

• Identify Trends: What are the significant trends in how consumers use Fitbit?
• Apply to Bellabeat: How can these trends be relevant to Bellabeat's current and potential customers?
• Influence Strategy: How can these insights guide Bellabeat's marketing efforts?

1.4 Deliverables

• Summary of the Business Task: Provide a clear and concise overview of the task.
• Data Sources: Describe all data sources used in the analysis.
• Data Cleaning & Manipulation: Document any cleaning or data manipulation processes applied.
• Analysis Summary: Summarize the analysis process and findings.
• Visualizations & Key Findings: Include supporting visualizations and highlight key insights.
• Content Recommendations: Offer high-level recommendations for content strategy based on the analysis.

1.5 Key Stakeholders

• Urška Sršen: Co-founder and Chief Creative Officer of Bellabeat
• Sando Mur: Co-founder and key member of the Bellabeat executive team
• Bellabeat Marketing Analytics Team: A team of data analysts to explore the trends and guide the Bellabeat's marketing strategy.

STEP 2: PREPARE

2.1 Information on Data Source

The data for this analysis is sourced from Kaggle, specifically the "FitBit Fitness Tracker Data" dataset, which comprises 18 CSV files. This dataset was generated by respondents through a distributed survey via Amazon Mechanical Turk between March 12, 2016, and May 12, 2016. The dataset includes data from 30 Fitbit users who consented to share their personal tracker data. The collected data spans various metrics, including:

• Physical activity (measured in minutes)
• Heart rate
• Sleep monitoring
• Daily activity
• Steps taken

2.2 Limitations of the Data Set

While this dataset offers valuable insights, it does come with several limitations:

• Data Age: The data was collected in 2016, and users' daily activity, fitness, sleep habits, diet, and food consumption may have changed significantly since then. This makes the data potentially less relevant to current trends.
• Sample Size: The sample consists of only 30 female Fitbit users, which is not representative of the broader female population.
• Data Integrity: Since the data was collected via a survey, it's challenging to verify its accuracy or integrity.

2.3 Is the Data ROCCC?

To evaluate the quality of the dataset, I assessed it against the ROCCC criteria, which stands for

Reliable, Original, Comprehensive, Current, and Cited.

• Reliable: LOW - The dataset is not highly reliable due to its small sample size (only 30 respondents).
• Original: LOW - The data is not original; it was collected by a third-party provider (Amazon Mechanical Turk).
• Comprehensive: MEDIUM - The dataset covers many of the parameters that Bellabeat's products also track, but it may not be exhaustive.
• Current: LOW - The data is over five years old, making it less relevant for current analysis.
• Cited: LOW - The data comes from a third party, and its origin is not well-documented.

Overall Assessment: The dataset is of low quality, and its limitations suggest that it may not be ideal for generating reliable business recommendations.

2.4 Data Selection

For the analysis, I selected the dailyActivity_merged.csv file, as it provides a comprehensive overview of users' daily activities, which aligns well with the objectives of this analysis.

STEP 3: PROCESS

We are using PostgreSQL to prepare and process the data.

3.1 Preparing the Environment

• Install all the required files to run PostgreSQL

3.2 Importing data set

• Import all the CSV files into PostgreSQL and prepare the tables for subsequent analysis.

3.3 Data cleaning and manipulation

Steps for data cleaning:

1. Observe and familiarise with data

2. Check for null or missing values

3. Perform sanity check of data

Creating a duplicate table of dailyactivity as dailyactivity_staging to proceed with the next operations.

Previewing using SQL command to show the first 100 rows of dailyactivity_staging to familiarise with the data.

SELECT * FROM public.dailyactivity_staging LIMIT 100;

Analyse the data and check for null values

SELECT * FROM public.dailyactivity
WHERE id = null

This step is repeated to each and every columns.

Null values are deleted using SQL command:

DELETE FROM your_table_name
WHERE your_column_name IS NULL;

From the above observation, noted that:

1. There is no typo, Null or missing values.
2. Data frame has 457 rows and 15 columns.
3. There are 35 unique IDs, instead of 30 unique IDs as expected from 30 fitness tracker users.

The following data manipulation is performed:
1. Convert ActivityDate to date datatype.
2. Create new column DayOfTheWeek by separating the date into day of the week for further analysis.
3. Create new column TotalMins being the sum of VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes and SedentaryMinutes.
4. Create new column TotalHours by converting TotalMins to number of hours.

--Step 1: Convert ActivityDate to date datatype.

ALTER TABLE public.dailyactivity_staging
ALTER COLUMN ActivityDate TYPE DATE
USING CAST(ActivityDate AS DATE);

--Step 2: Create new column DayOfTheWeek by separating the date into day of the week for further analysis.

ALTER TABLE public.dailyactivity_staging
ADD COLUMN DayOfTheWeek VARCHAR;

--Step 3: Add values to the recently created column

UPDATE public.dailyactivity_staging
SET DayOfTheWeek = TO_CHAR(DayOfTheWeek, 'Day');

--Step 4: Create new column TotalMins by adding the VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes and SedentaryMinutes for further analysis.

```
ALTER TABLE public.dailyactivity_staging
ADD COLUMN TotalMins BIGINT;

UPDATE public.dailyactivity_staging
SET TotalMins = ROUND(VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes +
SedentaryMinutes);
```

--Step 5: Create new column TotalHours by converting TotalMins into Hours for further analysis.

```
ALTER TABLE public.dailyactivity_staging
ADD COLUMN TotalHours BIGINT;

UPDATE public.dailyactivity_staging
SET TotalHours = ROUND(TotalMins/60);
```

Data cleaning and manipulation is completed. Hence, data is now ready to be analysed.

STEP 4: ANALYZE

4.1 Perform calculations

Pulling the statistics of dailyactivity_staging for analysis by:
• Sum
• Average
• Count
• Max
• Min

Statistical findings show that:

1. Sedentary users are the majority logging on average 991 minutes or 20 hours making up 81% of total average minutes.
2. Average calories burned is 2,189 calories equivalent to 0.6 pound. Could not interpret into detail as calories burned depend on several factors such as the age, weight, daily tasks, exercise, hormones and daily calorie intake
3. The calories burned and total steps walked on Tuesday, Wednesday and Thursday are decreased and the maximum calories burned and total steps walked is on Saturday.
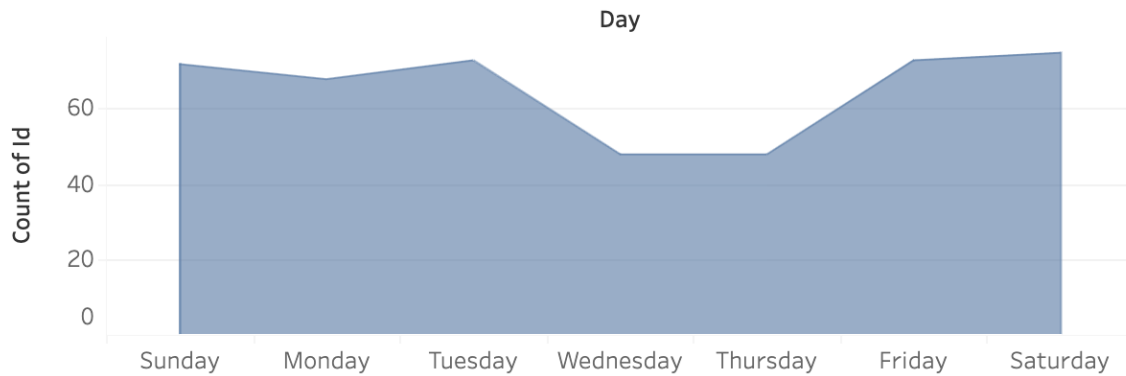

STEP 5: SHARE
In this step, we are creating visualizations and communicating our findings based on our analysis using Tableau.

5.1 Data Visualisation and Findings

**Frequency of usage across the week**

User Login Frequency

In this area chart, we are looking at the frequency of FitBit app usage in terms of days of the week.

1. We discovered that users prefer or remember (giving them the doubt of benefit that they forgotten) to track their activity on the app during midweek from Tuesday to Friday.

2. Noting that the frequency dropped on Friday and continue on weekends and Monday.

Percentage of Activity in Minutes

Activity Minutes Distribution by Intensity
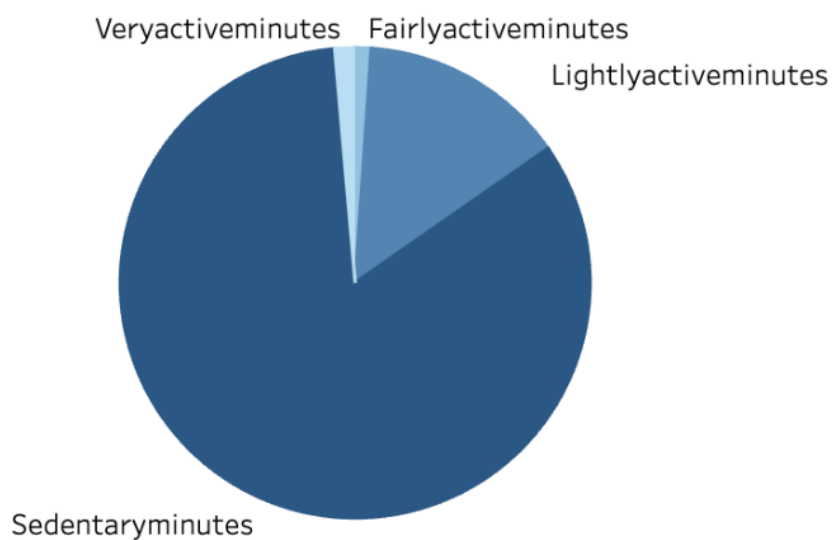
Daily Activity Intensity Levels
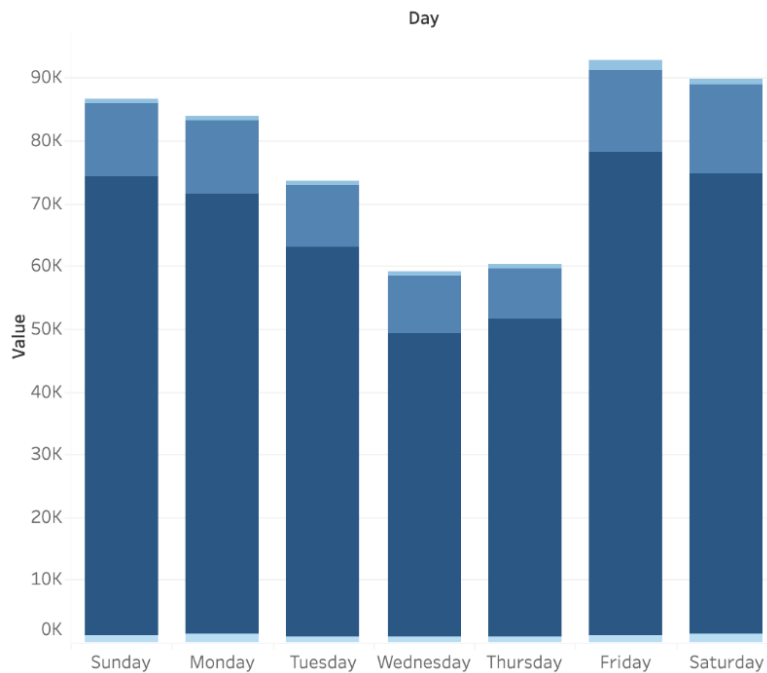

As seen from the pie chart,

Sedentary minutes takes the biggest slice.

This indicates that users are using the FitBit app to log daily activities such as daily commute, inactive movements (moving from one spot to another) or running errands.

App is rarely being used to track fitness (i.e. running) as per the minor percentage of fairly active activity (1.1%) and very active activity (1.7%). This is highly discouraging as FitBit app was developed to encourage fitness.

Calories burned for every step taken

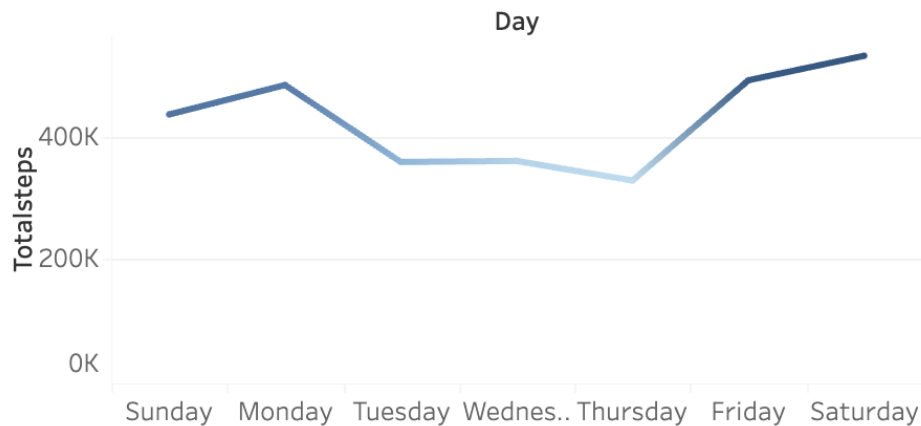## Correlation Between Steps and Calorie Burn
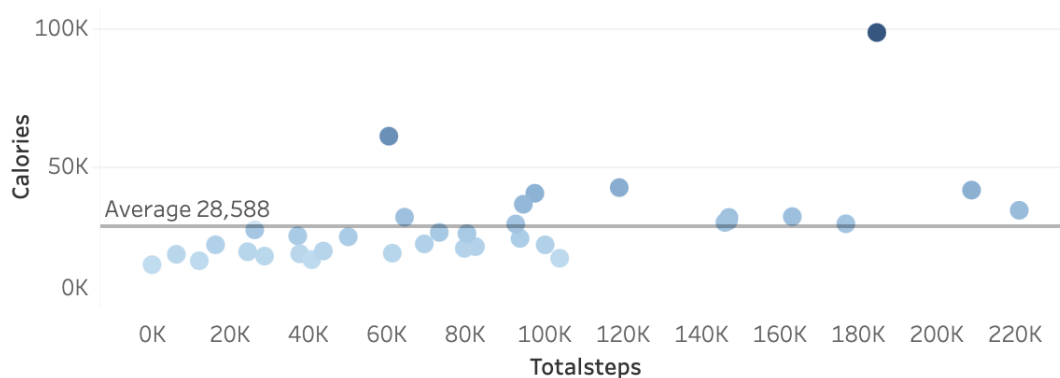
From the chart, we discovered that:

It is a positive correlation.

As the total steps increases the calories also increases.

Daily Trends in Calorie Burn

The line chart indicates the total steps walked and calories burned on weekends are higher than the weak days.

Tuesday, Wednesday and Thursday has the lowest step count and calories burnt, this represents the decrease in motivation for the users in every mid-section of the week.

STEP 6: ACT

In this final step, I will deliver insights and provide recommendations based on the analysis conducted.

Revisit Business Questions and High-Level Recommendations

1. What are the trends identified?

• A significant majority (81.3%) of users are primarily using the Fitbit app to track sedentary activities rather than their overall health habits.

• Users tend to track their activities more during weekdays compared to weekends, possibly because they are more active and spend more time outside on weekdays, whereas weekends might be more relaxed with indoor activities.

2. How could these trends apply to Bellabeat customers?

• Bellabeat and Fitbit both develop products focused on providing women with detailed insights into their health, habits, and fitness data. These identified trends in health and fitness tracking can be directly applied to Bellabeat customers, encouraging them to adopt healthier habits.

3. How could these trends help influence Bellabeat's marketing strategy?

• The Bellabeat marketing team could leverage these insights by educating users about the benefits of fitness and providing suggestions for various types of exercises. For example, they could promote simple 10-minute exercises for weekdays and suggest more intense workouts for weekends, along with information on calorie intake and burn rates.

• Additionally, the Bellabeat app could introduce notifications during weekends to encourage users to stay active, countering the tendency to be less active on these days.

Tableau Dashboard Link:

https://public.tableau.com/views/BellabeatGoogleCapstoneProject/INTENSITY?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link