BELLABEAT

Summary:

Bellabeat is a high-tech company that manufactures health-focused smart products for women. By collecting comprehensive data on activity, sleep, stress, and reproductive health, Bellabeat empowers women with valuable insights and knowledge about their personal health and lifestyle habits. This data-driven approach enables women to make informed decisions and take proactive steps towards improving their well-being and helping in product development.

Business Summary:

The business task at hand is to analyze the FitBit Data provided by Mobius, which consists of personal fitness tracking data of 30 registered users and gain insight into how people are using their smart devices. By exploring these trends and patterns, the goal is to inform recommendations to BellaBeat, which will help them shape marking strategies.

Ask Phase:

1. *Stakeholders*
   - **Urskan Srsen:** Bellabeat's co-founder and Chief Creative Officer
   - **Sando Mur:** Mathematician and Bellabeat's cofounder; a key member of the Bellabeat executive team
   - **Bellbeat marketing analytics team**
2. *Products*
   - Bellabeat app: Provides health data to users
   - Leaf: An accessory that tracks activity, sleep, and stress
   - Time: A wellness watch that tracks activity, sleep, and stress
   - Spring: A smart water bottle that ensures proper hydration
3. *Questions*
   - What are the most common activities tracked by users on their smart devices?
   - What are some trends in smart device usage?
   - Are there correlations between physical activity levels and other metrics like calories burned?
   - Are there any trends or patterns in users' sleep patterns or sleep quality?
4. *Limitations*
   - The data represents only Two months of data, making it potentially less comprehensive for capturing long-term trends.
   - Furthermore, the data was collected in 2016, hence not reflecting to current business conditions.
   - Also, the data does have missing values, affecting the accuracy of the analysis.

Prepare Phase:

The Fitbit Fitness Tracking Data, provided by Mobius and accessible through Kaggle (link: https://www.kaggle.com/datasets/arashnic/fitbit), forms the foundation of this project. This comprehensive dataset was collected through a distributed survey conducted via Amazon Mechanical Turk from 03.12.2016 to 05.12.2016. It comprises personal fitness tracker data from thirty Fitbit users, offering granular minute-level insights into physical activity, heart rate, and sleep patterns.

- All the data is stored in CSV FORMAT.
- Despite not being current the data is reliable, original and comprehensive.
1. *Selected Datasets*
   - DailyActivity_merged
   - SleepDay_merged
   - MinuteCaloriesNarrow_merged
   - Heartrate_seconds_merged

Process Phase:

There are multiple tools which I used to process the data:

1. Excel
2. SQL Server 2014
3. Power BI

NOTE:

- As my subscription to BigQuery has expired, I used SQL Server 2014 and Excel to analyze data and finally made the visual presentations with the help of Power BI.

To ensure the cleanliness of the data, a systematic approach was followed to check the column type as needed. The *CAST* function was employed to convert various columns into their appropriate formats, such as converting Date to DATE/Time and ID from VARCHAR to INT. Furthermore, I removed Duplicate values using *DISTINCT/GROUP BY* statements and *Remove Duplicate* options in Excel. These steps were crucial to ensure consistency. Additionally, I sorted data values using *ORDER BY* and *Sort* Tool in Excel. By using *MIN, MAX, AVG* and *STDEV* functions to check if there is any outlier data which has to be removed. Since some of the files were huge to open in Excel, I opened all documents in SQL SERVER 2014.

Details of Cleaning Process:

- DailyCalories_merged: Converted ActivityDate into Date format and converted Id from VarChar to INT using the *CAST* Function. Checked for duplicate data and deleted them.
- SleepDay_merged: Converted Sleepday VARCHAR to DATE using the *CAST* Function. Checked for duplicate data and deleted them.
- MinuteCaloriesNarrow_merged: Converted ActivityMinute VARCHAR to DATETIME.
- Heartrate_seconds_merged: Converted Id from VarChar to INT, Converted Time from VARCHAR to DATETIME and removed some inconsistency present in the data (where

one unique id at a specific time had 2 different values). Checked for duplicate data and deleted them.

1) Importing Datasets

To import the datasets, we use the SQL Server Import and Export Wizard.
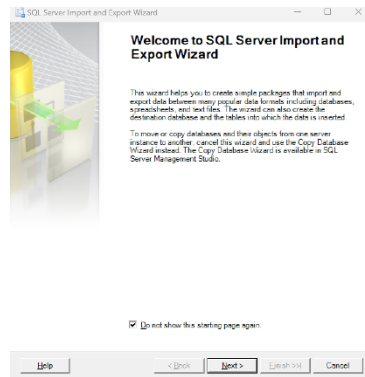
2) Preview Dataset

We use the *SELECT* and *TOP* statements for this task. An example query would be as follows:

```
SELECT TOP 50 [Id]
,[ActivityDate]
,[TotalSteps]
,[TotalDistance]
,[TrackerDistance]
,[LoggedActivitiesDistance]
,[VeryActiveDistance]
,[ModeratelyActiveDistance]
,[LightActiveDistance]
,[SedentaryActiveDistance]
,[VeryActiveMinutes]
,[FairlyActiveMinutes]
,[LightlyActiveMinutes]
,[SedentaryMinutes]
,[Calories]
FROM [Project].[dbo].[dailyActivity_merged]
```

To check the name of columns and the data type we use the *SELECT, Where* and *information_schema.columns*. An Example query would be as follows:

```
SELECT column_name as 'Column Name', data_type as 'Data Type',
        character_maximum_length as 'Max Length'
FROM Project.information_schema.columns
WHERE table_name = 'minuteIntensitiesNarrow_merged';
```

Where project is the database name and minuteIntensitiesNarrow_merged is our table name.

We can check the total number of rows present in a table by clicking on properties (of the table) -> storage where the information is provided or we could count each row with *COUNT(*)*.

In SQL Server the easiest way to remove duplicate data is to use the *DISTINCT* keyword or else we could remove duplicate data using *ROW_NUMBER()* and *Common Table Expressions (CTE).*
An example query would be as follows:
WITH cte *AS* (*SELECT* aid,time1,calories, *ROW_NUMBER*() *OVER* (*PARTITION BY* aid,time1,calories *ORDER BY* aid,time1,calories) rownum
FROM Project.dbo.Duplicate12)
DELETE FROM cte *WHERE* rownum>1

Where duplicate12 was a table made with duplicate values to try it out.

Analyze Phase:

Before we start analysing the data, let's aggregate the data so that it can be useful and accessible. To aggregate the data we could try answering the following questions hence making it easier to analyze:

- Are there correlations between physical activity levels and other metrics like sleep quality or calories burned?
- Are there any trends or patterns in users' sleep patterns or sleep quality?
- Does the day of the week have an influence on the total calories burned?
- How does the heart rate fluctuate with respect to time in a day?
- How does sleep vary with respect to the days?
- How does the sleep duration vary across the week?

By analyzing these questions, we can merge data from the data set to visualize and analyze the data better.

NOTE:
We will answer all of those questions, along with a few questions in the upcoming section.

After analyzing the data and seeing the question I have decided to merge the following tables to acquire the following tables:

1. Relationship_table= The table is made for the purpose of enabling slicer/filter features in Power BI. However, it is important to note that not all IDs are present in all tables. As a result, when applying filters, some values in the graph may appear empty for certain IDs. This table was created using the *DISTINCT* keyword in order to get all IDs.

```sql
SELECT DISTINCT Id
FROM Project.dbo.dailyActivity_merged
```

2. Sleep_pattern= This table provides insight into the sleep quality and activity levels of each individual on different days. It aims to gather and organize data about daily activity and sleep patterns.

```
SELECT DISTINCT DA.Id,CAST(ActivityDate AS DATE) AS ActivityDate,TotalSteps,TotalDistance,
SedentaryMinutes,TotalMinutesAsleep,TotalTimeInBed,TotalSleepRecords,

     CASE
        WHEN (CAST(([TotalMinutesAsleep]) AS INT))<=420 THEN 'Poor Sleep'
        WHEN (CAST(([TotalMinutesAsleep]) AS INT))>420 AND (CAST(([TotalMinutesAsleep]) AS INT))<=540 THEN 'Good Sleep'
        ELSE 'Excess Sleep'
     END  AS Sleep_Pattern,
(CAST(TotalTimeInBed AS INT)-CAST(TotalMinutesAsleep AS INT)) AS Idle_minutes_on_bed
FROM Project.dbo.dailyActivity_merged AS DA
INNER JOIN Project.dbo.sleepDay_merged AS SD ON DA.Id=SD.Id AND ActivityDate=SleepDay
```

3. HeartRate= The resulting table provides insights into heart rate measurements and total calorie expenditure for specific time intervals. This query aims to retrieve heart rate and calorie data, calculate the total calories for each combination of Id, Time, and Value, and order the results by Time and HR.Id.

```
SELECT HR.Id,CAST(Time AS DATETIME) AS Date_Time,Value,
SUM(CAST(Calories AS FLOAT)) AS Total_Calories
FROM Project.dbo.heartrate_seconds_merged AS HR
INNER JOIN Project.dbo.minuteCaloriesNarrow_merged AS MC ON MC.Id=HR.Id AND
CAST(MC.ActivityMinute AS DATETIME)=HR.Time
GROUP BY HR.id,Time,Value
ORDER BY Time,HR.Id
```

The aggregated data is utilized to gain insights by creating visualizations in Power BI. The following steps were taken to enhance the data in Power BI:

- Transforming the Data: In the transform window of Power BI, three additional columns were added to the sleep_pattern table. The first column represents the name of the day, the second column represents the number of the day and the third column contained if the day was a weekday or weekend.
- Differentiating Weekdays and Weekends: To identify whether a specific day falls on a weekend or weekday, an IF statement was implemented.
- Calculating Total Active Minutes: The total active minutes were computed by combining the VeryActiveMinutes and SedentaryMinutes with lightly active minutes. This calculation provides an overall measure of the time spent actively engaging in physical activities.

Now let's answer some of those questions.

1) *What are the most common activities tracked by users on their smart devices?*
   i) The most common activities tracked by users are Total Steps, Total Distance Covered, Calories Burned, Total Sedentary Minutes, Total Minutes Asleep, Weight, Heart Rate and total time in bed. This analysis could provide valuable insights into what new could be incorporated into our devices.

2) *What is the percentage of time that users spend engaged in activities other than sleeping compared to the time spent asleep while in bed?*
   i) This analysis has the potential to offer valuable insights into users who may be experiencing insomnia. There are 2 columns "TotalTimeInBed" and TotalMinutesAsleep" which will help us answer this question.

Sum of Total_Min_Asleep and Sum of Idle_minutes on the bed
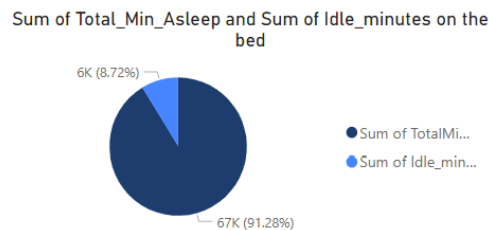
ii)

iii) Based on the analysis, it appears that, on average, the majority of users tend to be asleep while in bed. The data suggests that a significant portion of the total time spent in bed is dedicated to sleep rather than engaging in other activities.

3. *How does the sleep pattern vary with respect to Weekdays and Weekends?*
   i)   It is recommended that adults should aim for at least 7 hours of sleep per night. We use the CASE expression in SQL SERVER to get data about users who have excess, good and poor sleep.
   ii) We proceed by classifying users into these 3 categories:
        (1) Less than 7 hours is Inadequate Sleep
        (2) More than 7 hours but less than 9 hours is Good Sleep
        (3) Greater than 9 hours is Excess Sleep.
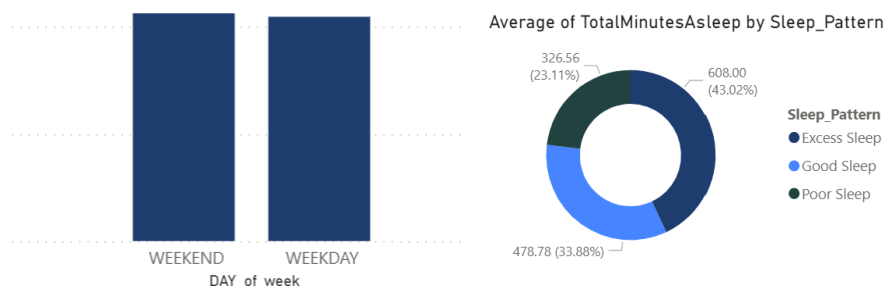
iii)

   iv) Based on the analysis of the bar chart, it is evident that there is a marginal variation in sleep duration between weekdays and weekends. This finding shows that users prefer to rest more on weekends.
   v)   The Donute chart shows a significant portion of users in the dataset are not achieving an adequate amount of sleep.

4. *How does the heart rate fluctuate with respect to time in a day?*
   i)   The normal heart rate of a human is around 60-100 beats per minute.
   ii) By doing this analysis we can get important information about the cardiovascular health of an individual and notify individuals if the heartbeats are irregular.
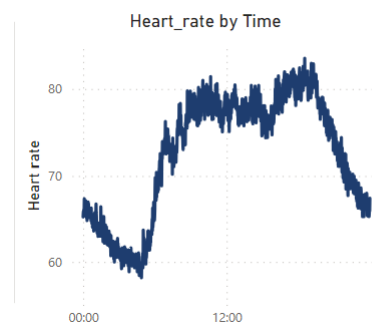

iii)

iv) The heart rate of all individuals does fall under this category. (Can further be examined in the DASHBOARD)

5. *Are there any correlations between physical activity levels and other metrics like sleep quality or calories burned?*

   i) Analyzing correlations between physical activity levels and other metrics such as sleep quality or calories burned can help us to inform health and fitness planning, performance improvement strategies, and give out sleep quality assessments.

   ii)

   TotalSteps by Calories

   Sum of TotalDistance by TotalMinutesAsleep
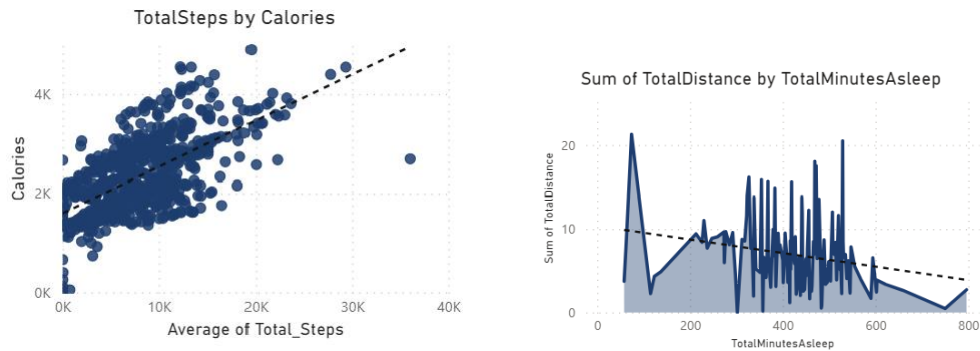
   iii) The scatter plot demonstrates a positive correlation between the number of steps taken and the calories burned. As the step count increases, there is a clear trend of higher calories burned. However, there is a negative correlation between the total distance travelled and the calories burned.

6. *How does the day of the week influence the total calories burned?*

   i) This analysis provides insights into the impact of daily routines making It valuable to individuals looking to optimize their calorie burn through lifestyle adjustments. Furthermore, it can be used to promote more physical activity on specific days.

   ii)

   Calories by Day Name
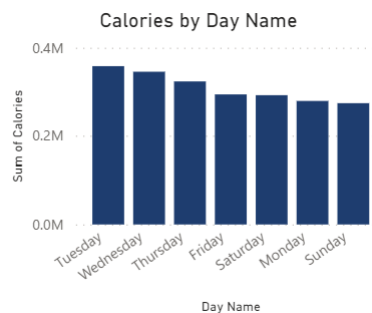
   iii) After analyzing the data, it was found that the maximum amount of calories was burned on Tuesday, followed by Wednesday and Thursday. These findings align with the expectation that weekdays, particularly mid-week days, tend to have a higher calorie burn due to regular work and activity routines. Conversely, Sunday exhibited the least amount of burned calories, which could be attributed to a more relaxed or less active day for many individuals.

   iv) Interestingly, Monday was observed to have the second-to-last position in terms of burnt calories. This finding may indicate a potential dip in energy expenditure at the start of the week.

7. *To what extent are users engaged in physical activity throughout the given time frame?*

   i) Analyzing the level of user activity throughout the given time frame can provide valuable insights into the physical fitness and lifestyle of the individuals.

Average of Total_Active_Minutes by Month and Day
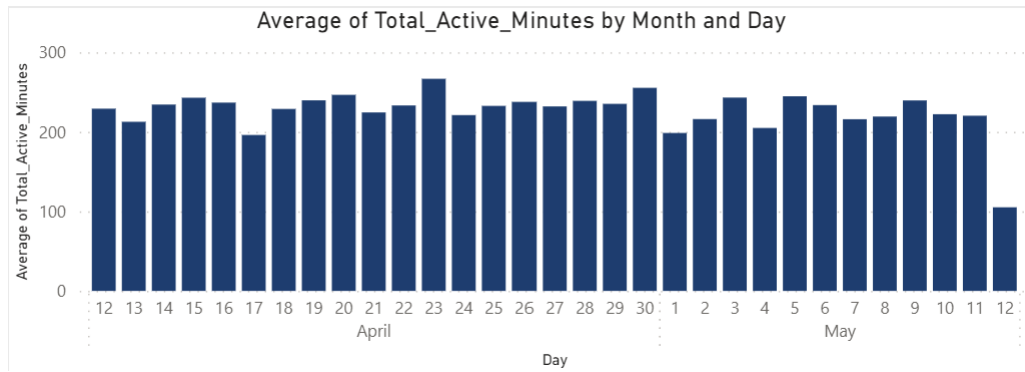
ii)

iii) By examining the activity levels throughout the dataset, it is evident that users generally maintain consistent levels of activity. However, an exception can be observed on May 12th, where there is a notable deviation from the overall pattern.

Act Phase:

Based on the important findings from the analysis of the fitness company's data and focusing on the Leaf product, we can make the following suggestions to the stakeholders:

1) Sleep and Sleep Quality
   - Enhance the sleep-tracking capabilities of the Leaf and provide users with comprehensive insights into their sleep patterns, including sleep stages, efficiency metrics, and personalized sleep improvement recommendations.
   - Educate users about the importance of sufficient and quality sleep, offering resources, tips, and personalized guidance within the app to help users establish healthy sleep routines and overcome common sleep issues.
2) Cardiovascular Health Monitoring
   - Implement features that analyze heart rate fluctuations throughout the day to help identify any irregularities and provide important information about users' heart health.
   - Implement SOS features into the Leaf.
3) Correlation between steps taken and calories burned
   - Use the positive correlation between steps taken and calories burned to motivate users to increase their physical activity levels. Offer rewards, challenges, and personalized goals to encourage users to reach and exceed their daily step targets.
4) Health and Fitness Planning
   - Promote sleep education, Educate users about the importance of sufficient and quality sleep.
   - Address sedentary behaviour and emphasize the importance of reducing sedentary behaviour during waking hours

Dashboard Layout:

The attached pictures showcase an interactive dashboard layout designed to provide users with a visually engaging and user-friendly interface for data exploration and analysis. The dashboard offers various interactive elements and visualizations to present information in an intuitive and actionable manner.

Page 1

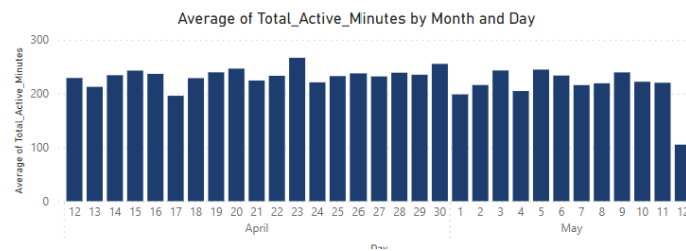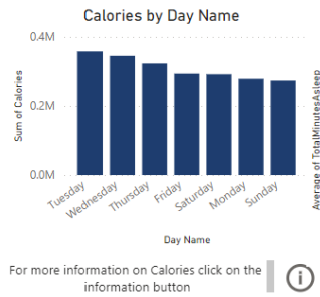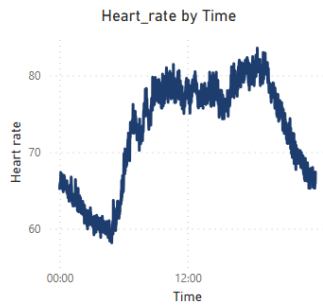Page 2

Filter

Id

- [ ] 1503960366
- [ ] 1624580081
- [ ] 1644430081
- [ ] 1844505072
- [ ] 1927972279
- [ ] 2022484408
- [ ] 2026352035
- [ ] 2320127002
- [ ] 2347167796
- [ ] 2873212765
- [ ] 3372868164
- [ ] 3977333714
- [ ] 4020332650
- [ ] 4057192912
- [ ] 4319703577
- [ ] 4388161847
- [ ] 4445114986

TotalMinutesAsleep by SleepDay

Sum of TotalDistance by TotalMinutesAsleep
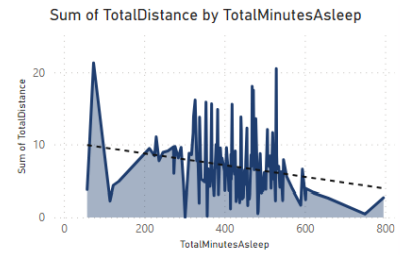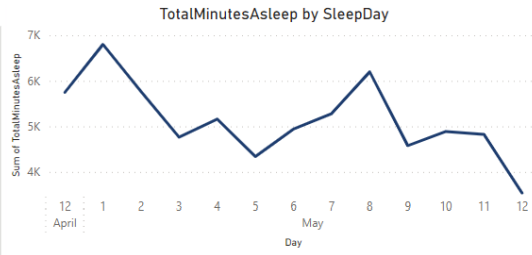
Average of TotalMinutesAsleep by Sleep_Pattern

326.56 (23.11%)
608.00 (43.02%)
478.78 (33.88%)

Sleep_Pattern
- ● Excess Sleep
- ● Good Sleep
- ● Poor Sleep

Sum of Total_Min_Asleep and Sum of Idle_minutes on the bed

6K (8.72%)
67K (91.28%)

- ● Sum of TotalMi...
- ● Sum of Idle_min...

Page 3

Done By Siddhanth Subbaiah