Bella Beats Case Study

Ester Mkamba

2024-04-23

ASK

The business task at hand was to determine a strategy for business growth by analysing trends from other smart devices from a familiar company.

PREPARE

- The data sourced was shared with consent from the 30 participants who are the data owners. The privacy of the participants was preserved as anonymization was done by using only an assigned ID number instead of their names.
- Also, the source of data is open as it allows free access, usage and sharing of data.
- Speaking of the data itself, we have to be careful as it a secondary source of data however, it is still current since the source a Kaggle member, refreshes it. As of the current, it was updated a month ago and has been cited.
- The data is not comprehensive since only 8 of the 30 participants provided their weight information while only 24 provided sleep activity details out of 30 with some missing both sleep and weight information.
- The data might also not be reliable since some participants provided their weight information manually which is easy to get human error while the rest of the participants provided what we assume automatically recorded weight.
- As for the data being reliable and comprehensive we will judge with time.

PROCESS

I opted to use power pivot to merge data. I also had to recreate tables with appropriate data type using CAST and IF EXIST functions

- Because of the similarities in the contents, I opted to merge the hourly calories, hourly inyensitites and hourly steps using power pivot to get a new sheet and named it "Hourly activities".
- I imported the tables: dailyactivity, heart rate seconds, hourly activities, minuteMETs, minute sleep, sleep Day and weight log info.
- fter that I cleaned the dailyactivity as seen below by using the CAST function to give appropriate data type to the entries. The new table was named cleaned_dailyActivity using IF EXISTS function and its table used for analysis.
- I did the same for the weight log info and named the cleaned table cleanedWeightLog

ANALYSIS

After that I began my analysis using SQL

```
--1. Tracking their physical activity and averages
Select count (distinct Id) as user_trackingActivity,
AVG(TotalSteps) as average_steps,
AVG (TotalDistance) as average_distance,
AVG (Calories) as average_calories
From dbo.cleaned_dailyActivity

--2. Tracking the days each tracked their physical activities

Select distinct Id, count (ActivityDate) over (partition by Id) as days_recorded_activity
From cleaned_dailyActivity
Order by days_recorded_activity DESC

--3. Calculating the average minutes for each activity
Select
AVG (VeryActiveMinutes) as Average_veryactive_minutes,
AVG (FairlyActiveMinutes) as Average_fairlyactive_minutes,
AVG (LightlyActiveMinutes)/60 as Average_lightlyactive_hours,
AVG (SedentaryMinutes)/60 as Average_sedentary_hours
From cleaned_dailyActivity

Select *
From heartrate_seconds

--4. Tracking their heart rate
Select count (distinct Id) as user_heartrate,
AVG (Value) as average_heartrate,
MAX (Value) as maximum_heartrate,
MIN (Value) as minimum_heartrate
From dbo.heartrate_seconds

--5. Tracking sleep
select count (distinct Id) as number_users,
AVG (TotalMinutesAsleep) / 60.0 as Average_hours_asleep,
MIN (TotalMinutesAsleep) / 60.0 as minimum_hours_asleep,
MAX (TotalMinutesAsleep) / 60.0 as maximum_hours_asleep,
AVG (TotalTimeInBed)/60 as average_hours_bed
From sleepDay

select *
From cleaned_WeightLog

--6. Tracking their weight
Select count (distinct Id) as User_number,
AVG (WeightKg) as average_weight,
Max (WeightKg) as maximum_weight,
Min (WeightKg) as minimum_weight
From cleaned_WeightLog


--7.Determine time when users were mostly active
Select distinct ActivityHour as activity_time,
--calculating average intensity for every hour to determine at what time are most peope physically active
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 [Hourly_Activities ] JOIN minuteMETs ON [Hourly_Activities ].Id = minuteMETs.Id
AND
[Hourly_Activities ].ActivityHour = minuteMETs.ActivityMinute
Order by average_intensity DESC
```

SHARE

Chart showing the distribution of time by activeness

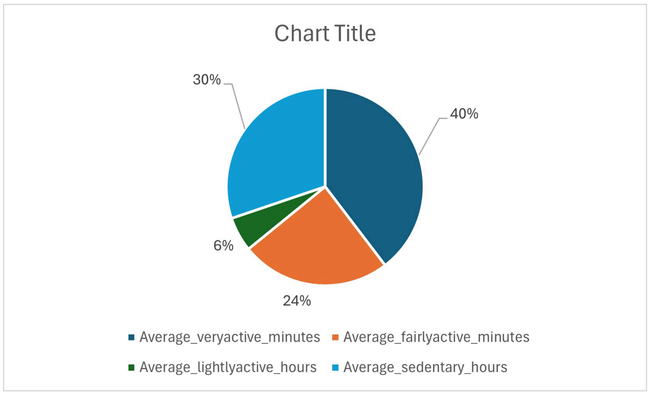


Chart showing steps versus calories

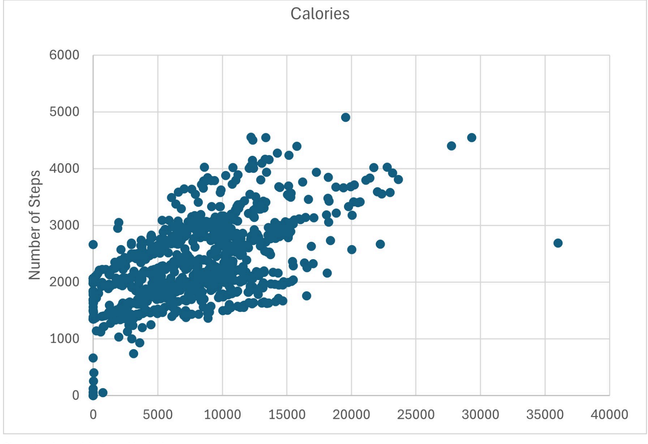
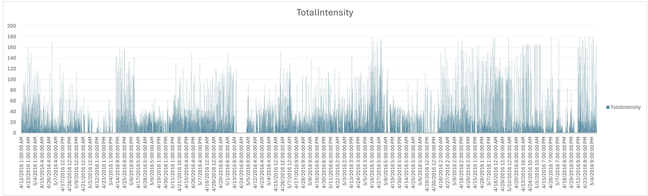


Chart showing activity intensities by hours



- Participants spent a majority of their time doing sedentary activities and lightly active activities and spending a very small amount of time (minutes)doing fairly active and very active activities
- Participants with the lowest calories had the highest number of steps
- Most intensity activites took place at around 9 PM and at 5 AM. Presuming that 5 in the morning is when majority of the participants get up to get ready for work and 9 PM as they are back from work and getting their workout schedule.

ACT

My recommendations are:

- To get more data that specifies their age and race and econoomic status so as to get accurate data for analysis
- For the Bella Beats device should have an alarm or alert to remind the participants to login their details as the data shows data entry decreases as the day goes by.
- Also, the company should emphasize on weight watching for its clients since only 8 out of the 33 participants logged in their weight details.