How Can a Wellness Technology Company Play It Smart?

Hello there,
I have been learning Google data analytics for the past few months and with the capstone project I will be able to put my learning to use. Practice they'd say makes perfect.:)

Roadmap

In this data analytics capstone project, I'll be using Data Analysis Phases; Ask, Prepare, Process, Analyse, Share, and Act as guide in this journey.

Company Name: Belle beats
Vision: To become a larger player in the global smart device market.
Mission: To provide a wide variety of smart devices that help women track their habits and health.
Company Objectives
To help
Key Stakeholders: Urska Srsen, Sando Mur
Team: Bellabeat Marketing Analytics

Products

Bellabeat app: The Bellabeat app (Fitbit) provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions.

Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.

Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide insights into daily wellness.

Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that users are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track users' hydration levels.

I am a junior data analyst working as a marketing analyst at Beabeat a manufacturer of health fitness devices for women. Our products were designed to help users mainly women to improve their health, wellness and fitness hence, reducing health-related diseases.

One of our company's stakeholders is concerned about the growth of the business, he came up with the idea that analysing the company's smart device fitness data could help unlock new growth opportunities for the company.

Being part of the marketing analytics team, that collects, analyses, and reports data that helps guide Bellabeat's marketing strategy. I have been assigned to work on this idea and focus on one of Bellabeat's products, and then analyse the smart device data usage to gain insight into how consumers are already using their smart devices.

Follow me as I analyse and give data-driven recommendations to my Stakeholders

Roadmap 1: ASK

In the ask stage, I will be asking relevant questions to understand what I'll be doing, tackle the business problems and to understand what the stakeholders want to achieve with my findings.

What are my stakeholders saying about the problem?
They want me to get the data of the company's smart devices, and analyse their usability to gain insight into how consumers are already using the devices.

How can I solve this business problem?
The best product for solving the business problem is the Bellabeat app. The app is connected with the other three products (Leaf, Time and Spring). Hence, it's easier to get adequate data on the Bellabeat app.

Business Task
Analyse the data from the Bellabeat app to gain insight into how consumers are already using their smart devices. The insight from this analysis will be used to generate recommendations for how trends in the data analysis can help Belle-beats make informed marketing strategy decisions.

Deliverables

I will document a report that includes;
- Clear summary of the business task
- Description of all data sources used
- Documentation of any cleaning or manipulation of data
- A summary of my analysis
- Supporting visualizations and key findings
- High-level content recommendations based on my analysis

Roadmap 1: PREPARE

Considerations
- The data was collewcted from 30 fitbit users that consented to the submissi0on of the information.
- It's assumed that women only submitted the information because the fitbit app is for women. However, It's unclear the gender that submitted the data.
- The data is open, available and accessible to the public.
- The data does display metaphase so it's not difficult to tell the source of the data, how and when it was created and what's all about. It's updated 2 years ago and has a public domain license.
- The dataset contain information about daily activity, steps and heart rate that can be udes to explore users habits.

Problem

Data source is out-dated and may not be relevant in today's fitness because human behavouirs has changed since 2016. For instances, the COVID'19 pandemic coupled with working from home have changed human activities exponentially. Therefore, the data is not relevant to today's data.

Organization of data: It's well organised and it's easy to identify them in group.

Importing Data

I downloaded and Imported three(3) data sets.csv to R studio. The data imported are;
- Daily activity_merged (wide data of 940 entries)
- Dailysteps_merged (long data of 940 entries)
- Minutesleep_merged (long data of 188,521 entries)

All data sorted and filter on R studio

Roadmap 1: PROCESS

Process stage is where we get to clean the data, making sure it's free of errors.

I'll be using R programmining language to process my data because R is easy to manipulate, It helps with faster calculation and visualization. The R supports larger data sets, It's has easier project organization and better for accuracy, it create better graphics and easier to learn and use.

Intergrity
Having checked that the data is correct and complete, I am certain of the integrity of the data for my analysis.

What steps have you taken to ensure that your data is clean?
- I check for incorrect and repeated data entry and remove them since they are irrelevant to our analysis
- I checked for extra spaces
- I check for bias in the data, ensuring it doesn't affect our result

I have verify that the data is clean and ready for analysis
Now I install the following packages and load them
(Here, Skimr, Janitor, dplyr, Tidyverse, ggplot2)

"Here" packages - It makes referencing files easier
"Skimr" package - makes summarising data easy and allow you skim quickly
"Janitor" package - it functions is to clean data
"dplyr" package - It has read-csv that loads csv.files
"Ggplot2" package - For visualization

Have you documented your cleaning process so you can review and share those results?

```
> str(dailyActivity_merged)
Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame':         940 obs. of  15 variables:
 $ Id                      : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
 $ ActivityDate            : chr  "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
 $ TotalSteps              : num  13162 10735 10460 9762 12669 ...
 $ TotalDistance           : num  8.5 6.97 6.74 6.28 8.16 ...
 $ TrackerDistance         : num  8.5 6.97 6.74 6.28 8.16 ...
 $ LoggedActivitiesDistance: num  0 0 0 0 0 0 0 0 0 0 ...
 $ VeryActiveDistance      : num  1.88 1.57 2.44 2.14 2.71 ...
 $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
 $ LightActiveDistance     : num  6.06 4.71 3.91 2.83 5.04 ...
 $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 0 ...
 $ VeryActiveMinutes       : num  25 21 30 29 36 38 42 50 28 19 ...
 $ FairlyActiveMinutes     : num  13 19 11 34 10 20 16 31 12 8 ...
 $ LightlyActiveMinutes    : num  328 217 181 209 221 164 233 264 205 211 ...
 $ SedentaryMinutes        : num  728 776 1218 726 773 ...
 $ Calories                : num  1985 1797 1776 1745 1863 ...
 - attr(*, "spec")=List of 3
  ..$ cols   :List of 15
  .. ..$ Id                      : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ ActivityDate            : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
  .. ..$ TotalSteps              : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ TotalDistance           : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ TrackerDistance         : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ LoggedActivitiesDistance: list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ VeryActiveDistance      : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ ModeratelyActiveDistance: list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ LightActiveDistance     : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ SedentaryActiveDistance : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
```

```
> #str(dataframe_name)
> str(dailySteps_merged_1_1_)
Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame':       940 obs. of  3 variables:
 $ Id         : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
 $ ActivityDay: chr  "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
 $ StepTotal  : num  13162 10735 10460 9762 12669 ...
 - attr(*, "spec")=List of 3
  ..$ cols   :List of 3
  .. ..$ Id         : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ ActivityDay: list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
  .. ..$ StepTotal  : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  ..$ default: list()
  .. ..- attr(*, "class")= chr [1:2] "collector_guess" "collector"
  ..$ skip   : int 1
  ..- attr(*, "class")= chr "col_spec"
> str(minuteSleep_merged_1_)
Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame':       188521 obs. of  4 variables:
 $ Id   : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
 $ date : chr  "4/12/2016 2:47:30 AM" "4/12/2016 2:48:30 AM" "4/12/2016 2:49:30 AM" "4/12/2016 2:50:30 AM"
...
 $ value: num  3 2 1 1 1 1 1 2 2 2 ...
 $ logId: num  1.14e+10 1.14e+10 1.14e+10 1.14e+10 1.14e+10 ...
 - attr(*, "spec")=List of 3
  ..$ cols   :List of 4
  .. ..$ Id   : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ date : list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_character" "collector"
  .. ..$ value: list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  .. ..$ logId: list()
  .. .. ..- attr(*, "class")= chr [1:2] "collector_double" "collector"
  ..$ default: list()
  .. ..- attr(*, "class")= chr [1:2] "collector_guess" "collector"
  ..$ skip   : int 1
  ..- attr(*, "class")= chr "col_spec"
```

After a brief view of the output, there are a few issues that we need to address:

- The naming of the column names

- Daily_activity, ActivityDate and Time is formatted as CHR not as a date format

To change this format, we use the clean_name () function

```
> clean_names(dailyActivity_merged)
            id activity_date total_steps total_distance tracker_distance logged_activities_distance
1  1503960366    2020-04-12        13162           8.50             8.50                           0
2  1503960366    2020-04-13        10735           6.97             6.97                           0
3  1503960366    2020-04-14        10460           6.74             6.74                           0
4  1503960366    2020-04-15         9762           6.28             6.28                           0
5  1503960366    2020-04-16        12669           8.16             8.16                           0
6  1503960366    2020-04-17         9705           6.48             6.48                           0
7  1503960366    2020-04-18        13019           8.59             8.59                           0
8  1503960366    2020-04-19        15506           9.88             9.88                           0
9  1503960366    2020-04-20        10544           6.68             6.68                           0
10 1503960366    2020-04-21         9819           6.34             6.34                           0
11 1503960366    2020-04-22        12764           8.13             8.13                           0
12 1503960366    2020-04-23        14371           9.04             9.04                           0
13 1503960366    2020-04-24        10039           6.41             6.41                           0
14 1503960366    2020-04-25        15355           9.80             9.80                           0
15 1503960366    2020-04-26        13755           8.79             8.79                           0
16 1503960366    2020-04-27        18134          12.21            12.21                           0
17 1503960366    2020-04-28        13154           8.53             8.53                           0
18 1503960366    2020-04-29        11181           7.15             7.15                           0
19 1503960366    2020-04-30        14673           9.25             9.25                           0
20 1503960366    2020-05-01        10602           6.81             6.81                           0
21 1503960366    2020-05-02        14727           9.71             9.71                           0
22 1503960366    2020-05-03        15103           9.66             9.66                           0
23 1503960366    2020-05-04        11100           7.15             7.15                           0
24 1503960366    2020-05-05        14070           8.90             8.90                           0
25 1503960366    2020-05-06        12159           8.03             8.03                           0
26 1503960366    2020-05-07        11992           7.71             7.71                           0
27 1503960366    2020-05-08        10060           6.58             6.58                           0
28 1503960366    2020-05-09        12022           7.72             7.72                           0
29 1503960366    2020-05-10        12207           7.77             7.77                           0
30 1503960366    2020-05-11        12770           8.13             8.13                           0
```

```
> clean_names(dailySteps_merged_1_1_)
             id activity_day step_total
1    1503960366    4/12/2016      13162
2    1503960366    4/13/2016      10735
3    1503960366    4/14/2016      10460
4    1503960366    4/15/2016       9762
5    1503960366    4/16/2016      12669
6    1503960366    4/17/2016       9705
7    1503960366    4/18/2016      13019
8    1503960366    4/19/2016      15506
9    1503960366    4/20/2016      10544
10   1503960366    4/21/2016       9819
11   1503960366    4/22/2016      12764
12   1503960366    4/23/2016      14371
13   1503960366    4/24/2016      10039
14   1503960366    4/25/2016      15355
15   1503960366    4/26/2016      13755
16   1503960366    4/27/2016      18134
17   1503960366    4/28/2016      13154
18   1503960366    4/29/2016      11181
19   1503960366    4/30/2016      14673
20   1503960366     5/1/2016      10602
21   1503960366     5/2/2016      14727
22   1503960366     5/3/2016      15103
23   1503960366     5/4/2016      11100
24   1503960366     5/5/2016      14070
25   1503960366     5/6/2016      12159
26   1503960366     5/7/2016      11992
27   1503960366     5/8/2016      10060
28   1503960366     5/9/2016      12022
29   1503960366    5/10/2016      12207
30   1503960366    5/11/2016      12770
```

```
> clean_names(minuteSleep_merged_1_)
             id                    date value       log_id
1    1503960366 4/12/2016 2:47:30 AM      3 11380564589
2    1503960366 4/12/2016 2:48:30 AM      2 11380564589
3    1503960366 4/12/2016 2:49:30 AM      1 11380564589
4    1503960366 4/12/2016 2:50:30 AM      1 11380564589
5    1503960366 4/12/2016 2:51:30 AM      1 11380564589
6    1503960366 4/12/2016 2:52:30 AM      1 11380564589
7    1503960366 4/12/2016 2:53:30 AM      1 11380564589
8    1503960366 4/12/2016 2:54:30 AM      2 11380564589
9    1503960366 4/12/2016 2:55:30 AM      2 11380564589
10   1503960366 4/12/2016 2:56:30 AM      2 11380564589
11   1503960366 4/12/2016 2:57:30 AM      3 11380564589
12   1503960366 4/12/2016 2:58:30 AM      3 11380564589
13   1503960366 4/12/2016 2:59:30 AM      3 11380564589
14   1503960366 4/12/2016 3:00:30 AM      3 11380564589
15   1503960366 4/12/2016 3:01:30 AM      3 11380564589
16   1503960366 4/12/2016 3:02:30 AM      2 11380564589
17   1503960366 4/12/2016 3:03:30 AM      1 11380564589
18   1503960366 4/12/2016 3:04:30 AM      1 11380564589
19   1503960366 4/12/2016 3:05:30 AM      1 11380564589
20   1503960366 4/12/2016 3:06:30 AM      1 11380564589
21   1503960366 4/12/2016 3:07:30 AM      1 11380564589
22   1503960366 4/12/2016 3:08:30 AM      1 11380564589
23   1503960366 4/12/2016 3:09:30 AM      1 11380564589
24   1503960366 4/12/2016 3:10:30 AM      1 11380564589
25   1503960366 4/12/2016 3:11:30 AM      1 11380564589
26   1503960366 4/12/2016 3:12:30 AM      1 11380564589
27   1503960366 4/12/2016 3:13:30 AM      1 11380564589
28   1503960366 4/12/2016 3:14:30 AM      1 11380564589
29   1503960366 4/12/2016 3:15:30 AM      1 11380564589
30   1503960366 4/12/2016 3:16:30 AM      1 11380564589
```

Summary

```
> summary(dailyActivity_merged$total_steps)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0    3790    7406    7638   10727   36019
> summary(dailyActivity_merged$very_active_minutes)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    4.00   21.16   32.00  210.00
> summary(dailyActivity_merged$fairly_active_minutes)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    6.00   13.56   19.00  143.00
> summary(dailyActivity_merged$lightly_active_minutes)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0   127.0   199.0   192.8   264.0   518.0
> summary(dailyActivity_merged$sedentary_minutes)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0.0   729.8  1057.5   991.2  1229.5  1440.0
> summary(dailyActivity_merged$calories)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0    1828    2134    2304    2793    4900
> summary(dailySteps_merged_1_1_$step_total)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0    3790    7406    7638   10727   36019
> summary(minuteSleep_merged_1_$value)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.000   1.000   1.096   1.000   3.000
> |
```

Now we move to visualization in R

Using ggplot2 to make high quality customizable plots of the data

Creating relationship using scatter plot

Next, let's examine the relationship between the following:

Relationship between sedentary minutes and total steps