Case Study : How Can a Wellness Technology Company Play It Smart?

Tina Huang 2021-06-14

Scenario

Bellabeat, a high-tech company that manufactures health-focused smart products. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits.

The cofounder asks you to analyze smart device usage data in to gain insight into how consumers use non-Bellabeat smart devices. She then wants you to select one Bellabeat product to apply these insights and make recommendations for how these trends can inform Bellabeat marketing strategy.

Table of Content

Business Task

1. How could these trends help influence Bellabeat marketing strategy?
2. How could these trends apply to Bellabeat customers?
3. How do users track wellness data in smart device?

Analysis Summary

Use Microsoft SQL Server to clean and analyze the data and find the trends.

1. Calculate average steps, average of distance, average time of each activity type, and average calories from all users:

Average_Steps	Average_Distance	Average_VeryActiveMinutes	Average_FairlyActiveMinutes	Average_LightlyActiveMinutes	Average_SedentaryMinutes	Average_Calories
7637	5.49	21	13	192	991	2303

- 15 out of 33 users who have average steps exceeding the average steps of all users.
- 15 out of 33 users who have average distance exceeding the average distance of all users.
- 11 out of 33 users who have average VeryActiveMinutes exceeding the average VeryActiveMinutes of all users.
- 19 out of 33 users who have average SedentaryMinutes exceeding the average SedentaryMinutes of all users.

2. Find the users who have steps over 12,000 on 4/12/2016 and 5/11/2016:

Id	Date	TotalSteps
8877689391	2016/4/12	23186
8053475328	2016/4/12	18060
7007744171	2016/4/12	14172
1503960366	2016/4/12	13162

Id	Date	TotalSteps
8877689391	2016/5/11	21420
2022484408	2016/5/11	13272
1503960366	2016/5/11	12770
8053475328	2016/5/11	12209

- There are 3 users who have steps over 12,000 both on 4/12/2016 and 5/11/2016. The 3 users have slightly fewer steps after a month.

3. Calculate the percentage of each activity for user id = 2320127002 and 8877689391:

Id	Avg_Calories	Pct_VeryActive	Pct_ModeratelyActive	Pct_LightActive	Pct_SedentaryActive
2320127002	1724	0.0335	0.0307	0.9349	0
8877689391	3420	0.5023	0.0256	0.4684	0.0004

- User (Id = 8877689391) consumes more calories than User (Id = 2320127002) since the former has the most proportion in the type of VeryActive.

4. Calculate numbers of users for each active type:

About 50 percent of people have low activity.

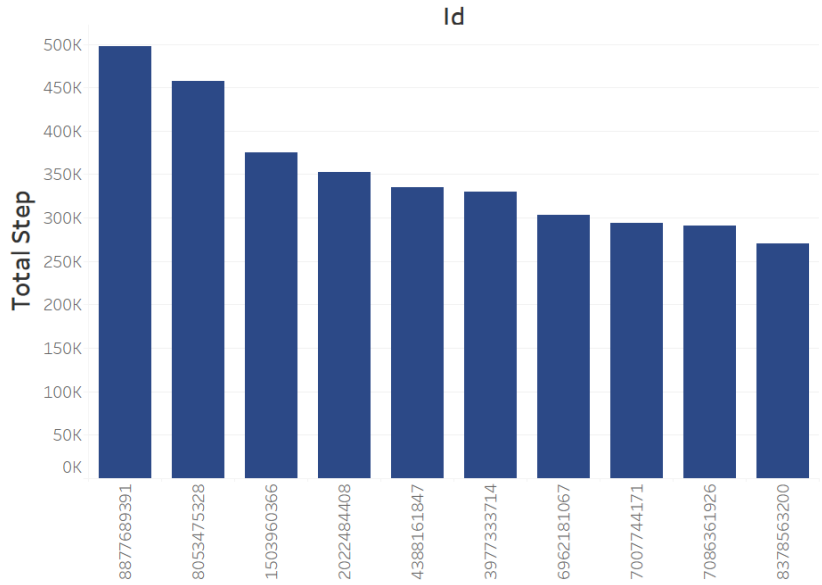
High Activity	Median Activity	Low activity
6	11	16

Definition:

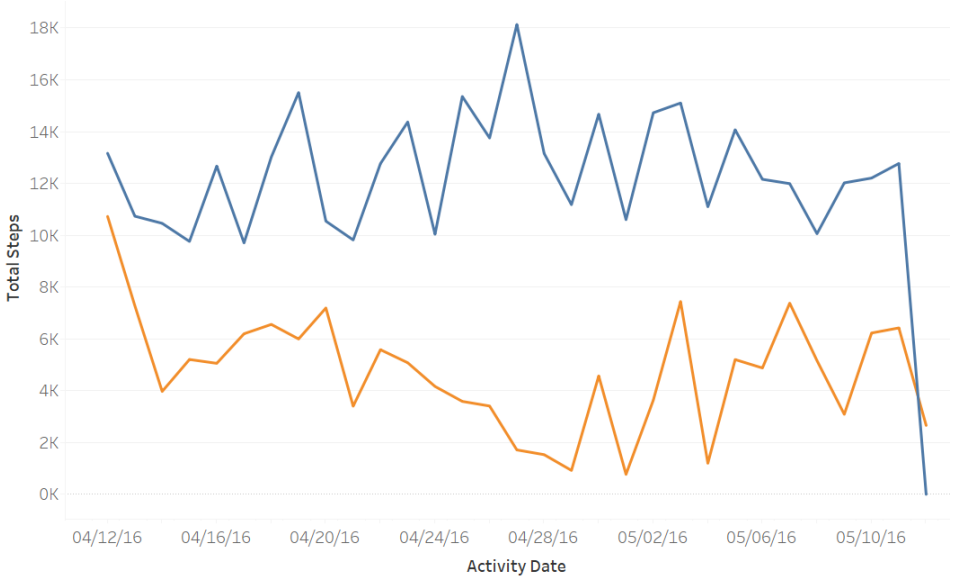
- High Activity: the average time of combination of VeryActiveMinutes and FairlyActiveMinutes more than or equal to 60 minutes.
- Median Activity: the average time of combination of VeryActiveMinutes and FairlyActiveMinutes more than or equal to 30 minutes but less than 60 minutes.
- Low Activity: the average time of combination of VeryActiveMinutes and FairlyActiveMinutes less than 30 minutes.

Visualizations and Insights

1. The maximum of total steps among all users is about 500,000. The smart device can list top 10 users with any metric for users to run a contest.



2. The two users have huge difference in total steps every day. Users can compare the total steps with friends every day to get motivation.

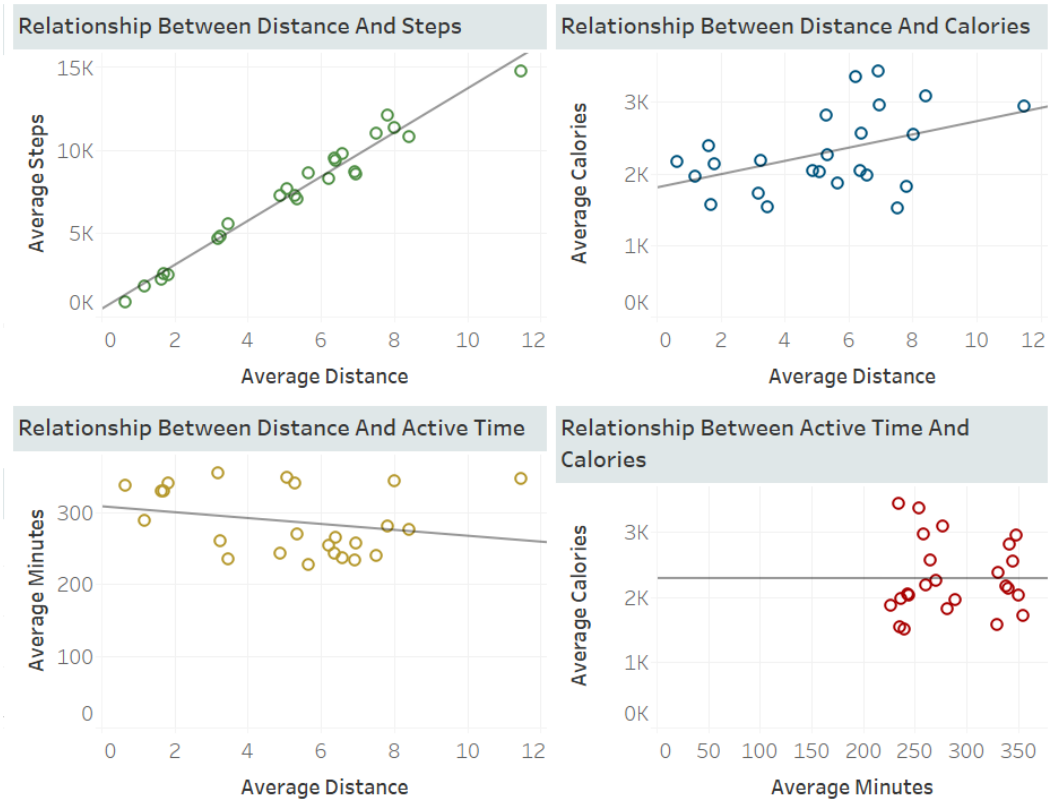


3. Relationships between two variables among the users:
   - There is a strong positive relationship between total distance and steps. (r = 0.9958)
   - There is a weak positive relationship between total distance and calories. (r = 0.4434)
   - There is a weak negative relationship between total distance and active time. (r = -0.2392, Active time = VeryActiveMinutes + FairlyActiveMinutes +

LightlyActiveMinutes + SedentaryMinutes)

If a user has more LightlyActiveMinutes and SedentaryMinutes for active time, then it is related to lower active distance. From the graph below, we know that users tend to have more LightlyActiveMinutes and SedentaryMinutes for active time.

- There is no relationship between active time and calories.



Recommendations

1. Create an alert function that gives alert to the users for sitting too long using the data of sedentary minutes. Most office workers are sedentary and confined to their workspace.
2. Track the calories consumed for every activity. The foundation will make users to know more about how the activity burns fats.
3. Provide the information of progress in steps and active time every week. Tracking the progress will make users realize how they progress every week and prompt users exercise more when the data shows low activity.
4. List top 10 users with any metric for users to realize who have better performance when there is a competition.
5. Create a graph to show daily activities to others. Share the activity with friends can help users inspire others. It’ s always fun to do exercise with friends.

6. Add a column to calculate the time difference between TotalMinutesAsleep and TotalTimeInBed. Users may want to know how long does it take to fall asleep.

Description of Data

1. Data source and security
   FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from thirty fitbit users from 4/12/2016 to 5/12/2016. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits.

Documentation of Cleaning or Manipulation of Data

1. Validate the data in column Id.
2. Identify duplicate rows and delete 3 duplicate records in table SleepDay.
3. Filter data by condition (is empty) to find null data.
4. Convert data types.