BellaBeat Smart Device Data Analysis (Google Capstone Project)

Dillon O'Rourke

2022-03-12

Disclaimer

This case study is a project for my online portfolio as part of the Google Data Analytics professional certificate and is entirely fictional.

Introduction

I am a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company.

The purpose of this project;

Is to use data to gain insights into customer usage of the non-company products and identify new growth opportunities for the company.
An analysis of smart device usage could identify trends which could be investigated further to explore if they can be applied to Bellabeat customers. Applying these trends and behaviors to the company's marketing strategy could improve future sales revenue.

The key business tasks;

- Analyse smart device data and gain insight into how consumers are using their devices.
  - What are some trends in smart device usage?
  - How could these trends apply to Bellabeat customers?
  - How could these trends help influence Bellabeat marketing strategy?
- Create a presentation of my analysis and high-level recommendations for how these trends can inform Bellabeat marketing strategy to the executive team.

Assumptions or theories

Prodcuts

- **Bellabeat app:** The Bellabeat app 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. The Bellabeat app connects to their line of smart wellness products.

- **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 you with insights into your daily wellness.
- **Spring:** This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.
- **Bellabeat membership:** Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

The Data

1. FitBit Fitness Tracker Data.

Personal fitness tracker data collected from 30 consenting fitbit users in 2016. It includes minute level output for exercise, heart rate and daily activity.

- **Source:** External 3rd party data.
- **Type:** Structured continuous & discrete mostly nominal quantitative data.
- **Structure:** Time-series data for each individual user in several tables. Data ranges from daily to minute timestamps. Heartrate, calories, steps, sleep and weight information are provided by the data.

The *dailyActivity* table will be analysed first as it will provide the most high level veiw look for trends in daily data and then hourly or minute data can be delved into to provide a higher resolution picture into a trend.

- The data runs from the 12th of March to the 12th of April 2016. Therefore there should be 32 days of data to work with.

In [21]:
```
import pandas as pd
import os

#Defining the data folder and creating a list of files in it using data_files
data_dir = r'C:\Users\Dillon\Documents\Coding Learning\Google Data Analytics
data_files = os.listdir(data_dir)

#creating a dataframe from each file in the os dir
dailyActivity = pd.read_csv(data_dir + '\dailyActivity_merged.csv')

print(len(pd.unique(dailyActivity['ActivityDate'])), 'days')
```

```
32 days
```

Problems

- The Fitbit data used was collected from only 30 users. This is a very small sample size when considering that there are around 30 million users. Our sample size therefore represents roughly 1 millionth or 0.0001% of the population.
- In addition to the above, the data was collected in 2016 so it is outdated. Were this study carried out a year ago during a pandemic and lockdowns, the data might even be misleading as people's exercise and eating habits are likely not the same during a pandemic.

- During the exploration of the daliyactivity table, 35 unique IDs were found. This is an issue because it was stated that Fitbit data was collected from 30 users and yet there are 35 unique user IDs in the first table.

You need to identify what the major problems are, explain how you have analyzed the problem, and present any facts you are using to support your findings.

In [8]:
```
print(len(pd.unique(dailyActivity['Id'])), ' Unique IDs in dailyActivity tabl
```

```
35  Unique IDs in dailyActivity table
```

- During the exploration of the daliyactivity table, it was found that data was not logged for each of the 32 days for everyone.

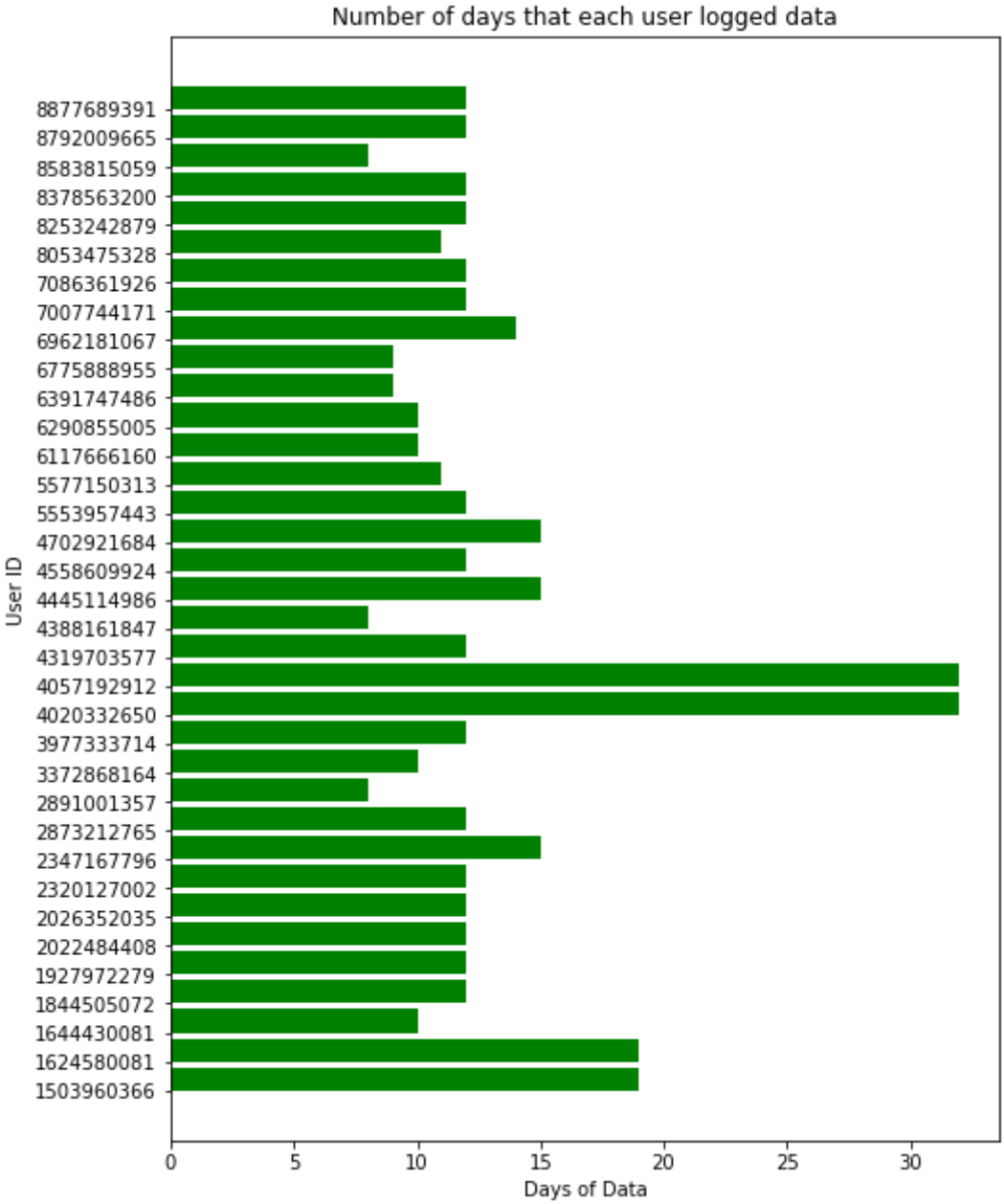
In [20]:
```
import matplotlib.pyplot as plt
import numpy as np

daily_actchk = dailyActivity.groupby(['Id']).agg(['nunique'])

dates = list(daily_actchk.iloc[:, 0])
IDs = list(pd.unique(dailyActivity['Id']))

IDss = list()
for i in IDs:
    IDss.append(str(i))
fig = plt.figure()
ax = fig.add_axes([0,0,1,2])
x = IDss
y = dates
plt.barh(IDss, y, align='edge', color = 'green')
plt.xlabel("Days of Data")
plt.ylabel("User ID")
plt.title("Number of days that each user logged data")
plt.show()

print('Average number of days of data: ', int(np.around(np.average(dates), 0)
```




```
Average number of days of data:  13
```

Solutions

Outline a solution that would alleviate the problem and have a few alternatives in mind to show that you have given the case study considerable thought. Don’t forget to include pros and cons for each solution.

Conclusion

End your presentation by summarizing key takeaways of all of the problem-solving you conducted, highlighting what you have learned from this.

Next steps

Choose the best solution and propose recommendations for the client or business to take. Explain why you made your choice and how this will a ect the scenario in a positive way. Be specific and include what needs to be done, who should enforce it, and when.

References

[1] FitBit Fitness Tracker Data, Furberg, Robert; Brinton, Julia; Keating, Michael ; Ortiz, Alexa https://zenodo.org/record/53894#.YMoUpnVKiP9

In [ ]: