**How Can A Wellness Technology Play It Smart ?**

Case Study

Aarushi Srivastava

12.05.2022

**Introduction & Acknowledgement**

My name is Aarushi Srivastava and I am a graduate in Information Technology. I completed my Engineering degree back in 2020 and immediately after that I started working for an IT firm as a full stack developer where my roles and responsibilities were around development and cloud infrastructure delivery.

I was enjoying my first job but I was also missing the feeling of exploring my creative dimension .My interest were mostly inclined towards solving business problems and working on tangible real-world issues and that’s when I decided to learn about Data Analytics through a Certification program introduced by Google on Coursera platform. It was not the first time that I was dealing with data and numbers ; I always had a keen interest in data management and statistics and I got a distinction in Mathematics during my highschool. Furthermore,while I was doing my internship in 2019 ,my role was that of an administrative intern and there I collated spreadsheets of data by carrying out a first-hand survey among engineering students. There I learnt about data collection, data cleaning and organization for the first time.

I registered for the Google Data Analytics course in December 2021 and since then my learning journey has been extremely fulfilling and motivating . Currently I am left with the final milestone of the program which is the Capstone project. I chose the second Case Study which is a descriptive case for providing recommendations on marketing strategy for a Wellness Technology Company.

I would like to extend my token of thanks to all the discussion forums and various data analytics communities present on the internet for their support and spirit. I would also like to thank Google and the team for architecting the most lucrative course content and the excellent deliverables. Last but not the least, I would like to thank God and my parents for their relentless support and guidance which lead me to successfully complete the project.

**Business Problem and Company Background**

The company we are looking into is a high-tech company that manufactures health-focused smart products .They collect data on activity,sleep ,stress and reproductive health that helps in empowering women with knowledge about their own health and habits. The company was founded in 2013 by Sando Mur and Urška Sršen .

By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products are available through a growing number of online retailers in addition to their own e-commerce channel on their website. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates.

The analysis of Bellabeat’s available consumer data could reveal more opportunities for business growth.Hence, I am focussing on one Bellabeat product and analyzing the smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, high-level recommendations can be derived for how these trends can inform Bellabeat marketing strategy.

**Ask Phase**

In order to understand the business scenario in a more detailed way, I asked the following questions which lead to better understanding of the gaps in Bellabeat’s marketing strategy.

* Carefully monitor the trends in the user’s activities tracked by the Fitbit device and identify areas that are not included in the marketing scripts.
* Identified the activities that were of most interest for the customers and carefully studied their activity patterns.
* Understanding the demands and expected outputs for the stakeholders.
* Identifying the stakeholders .Urška Sršen, Bellabeat’s co-founder and Chief Creative Officer; Sando Mur, Mathematician and Bellabeat’s cofounder are the key stakeholders .

**Prepare Phase**

In this phase I will download and import the dataset. Organizing ,sorting and filtering of data will also be performed in this step. Sršen encourages to use public data that explores smart device users’ daily habits.

Downloading and Storing Data

**FitBit Fitness Tracker Data** (CC0: Public Domain, dataset made available through Mobius) - This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users’ habits.

Link for the dataset : [FitBit Fitness Tracker Data | Kaggle](https://www.kaggle.com/datasets/arashnic/fitbit)

The dataset consist of 18 CSV filed wherein data is stored in both long and wide format.

Importing the data

I will be using Kaggle Notebook for importing the data.First of all, I will be loading some R packages. Following are the packages imported.

library(tidyverse)

library(lubridate)

library(dplyr)

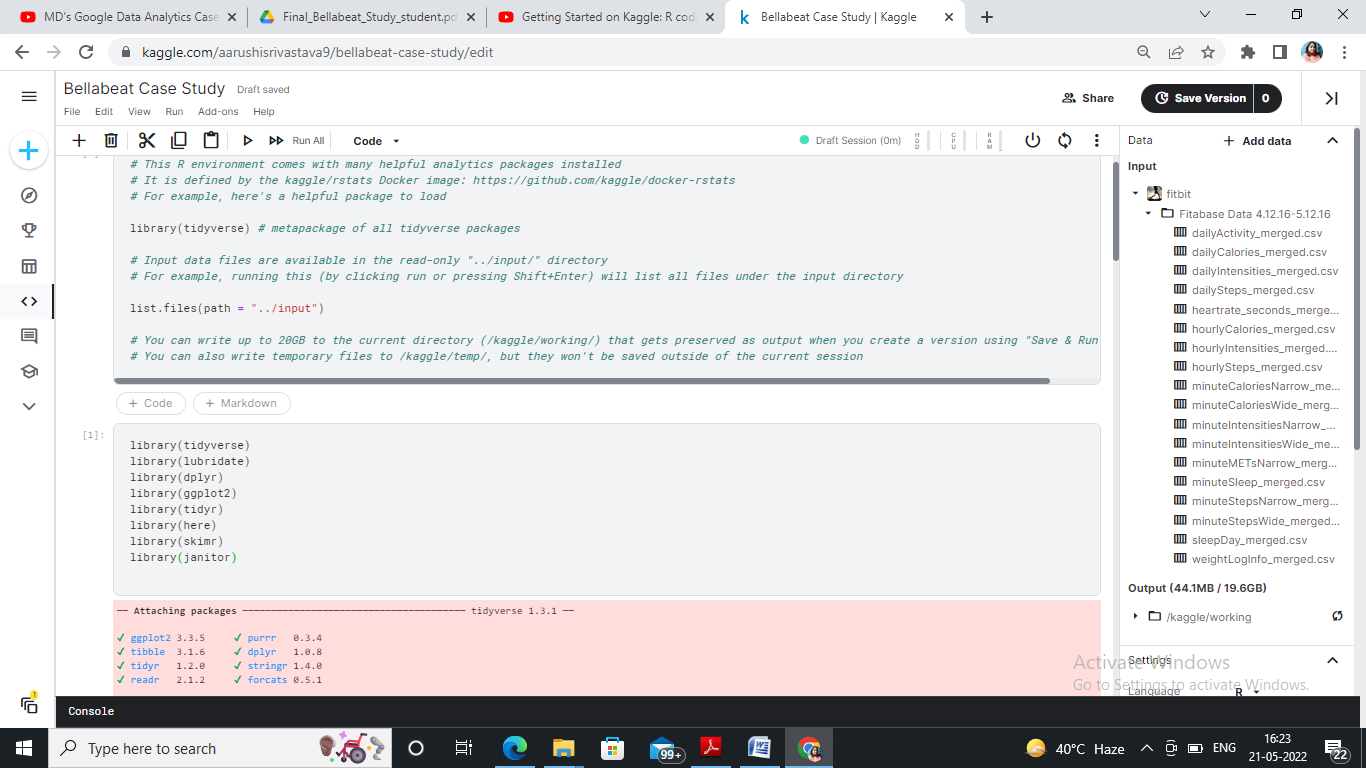
library(ggplot2)

library(tidyr)

library(here)

library(skimr)

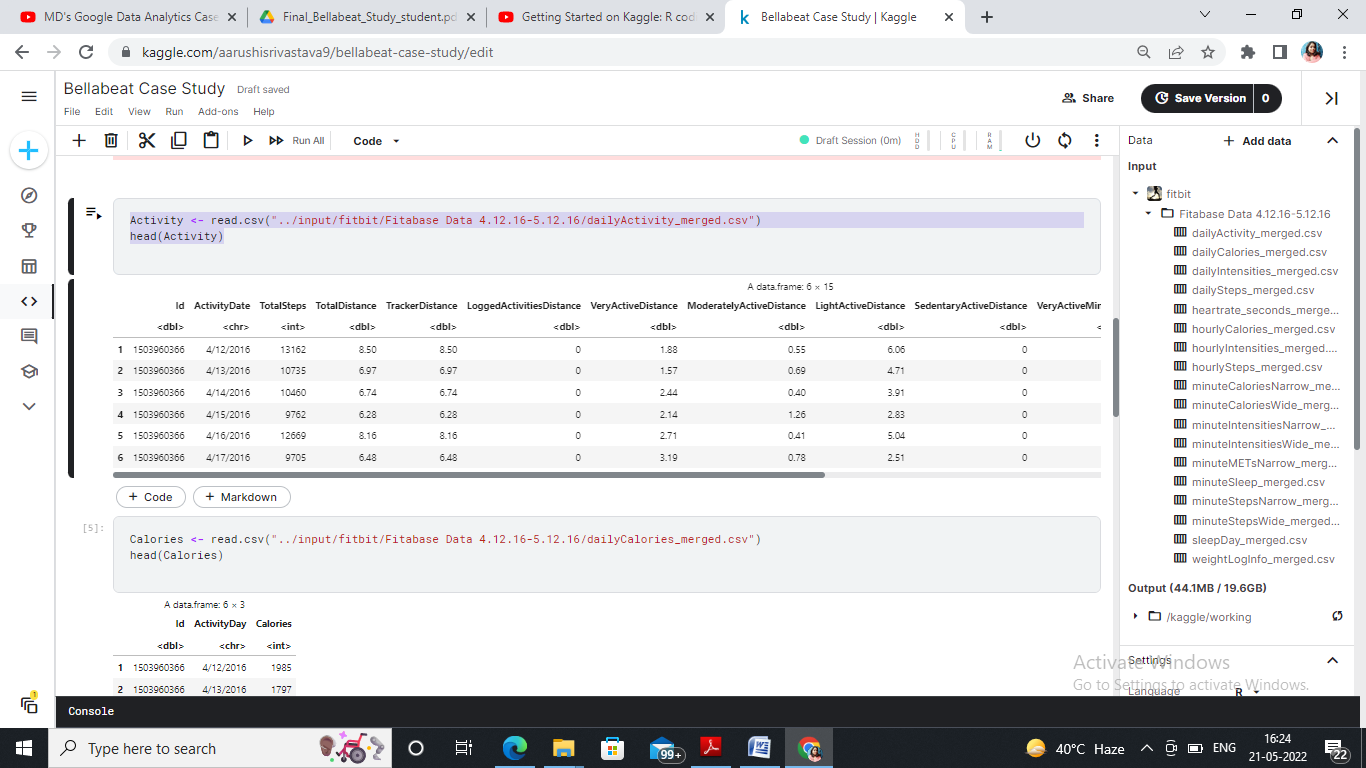
library(janitor)



* **dailyActivity\_merged.csv**

Activity <- read.csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyActivity\_merged.csv")

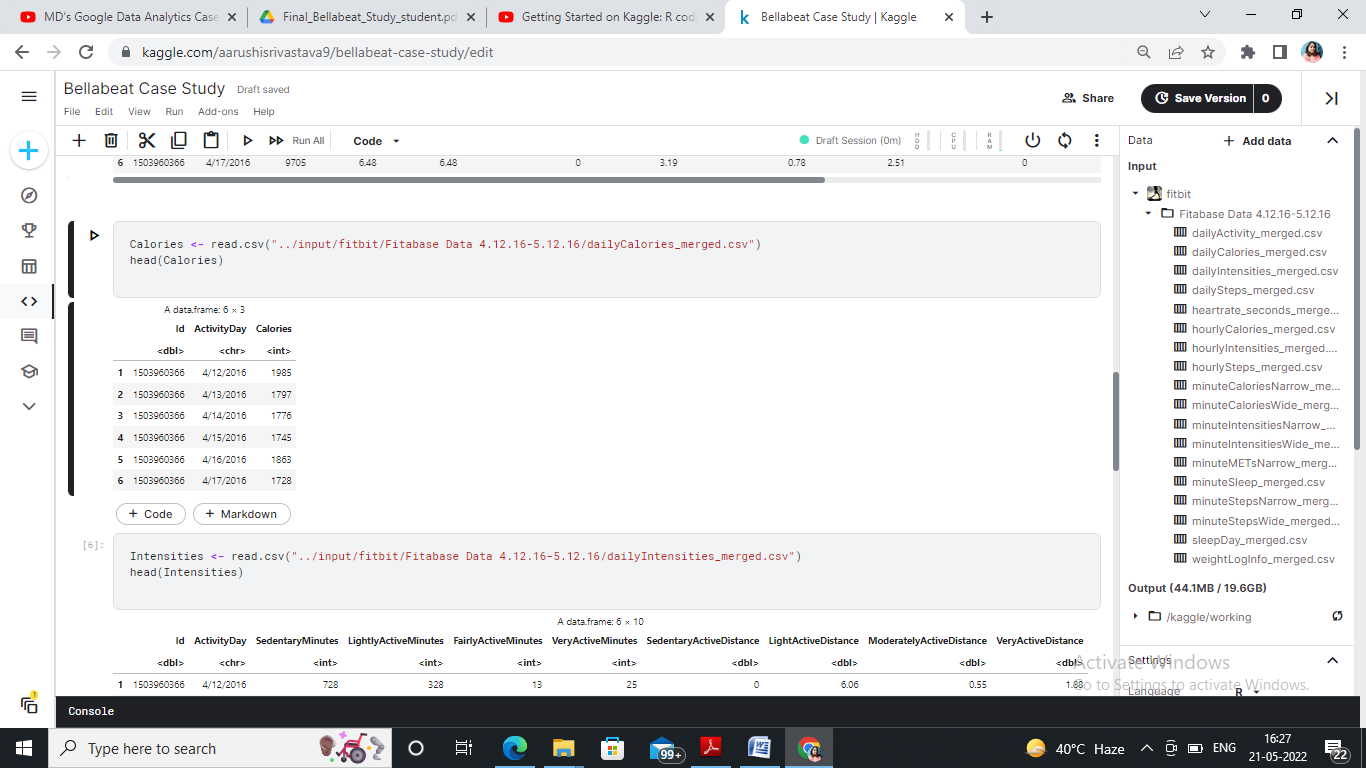
head(Activity)



* **dailyCalories\_merged.csv**

Calories <- read.csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyCalories\_merged.csv")

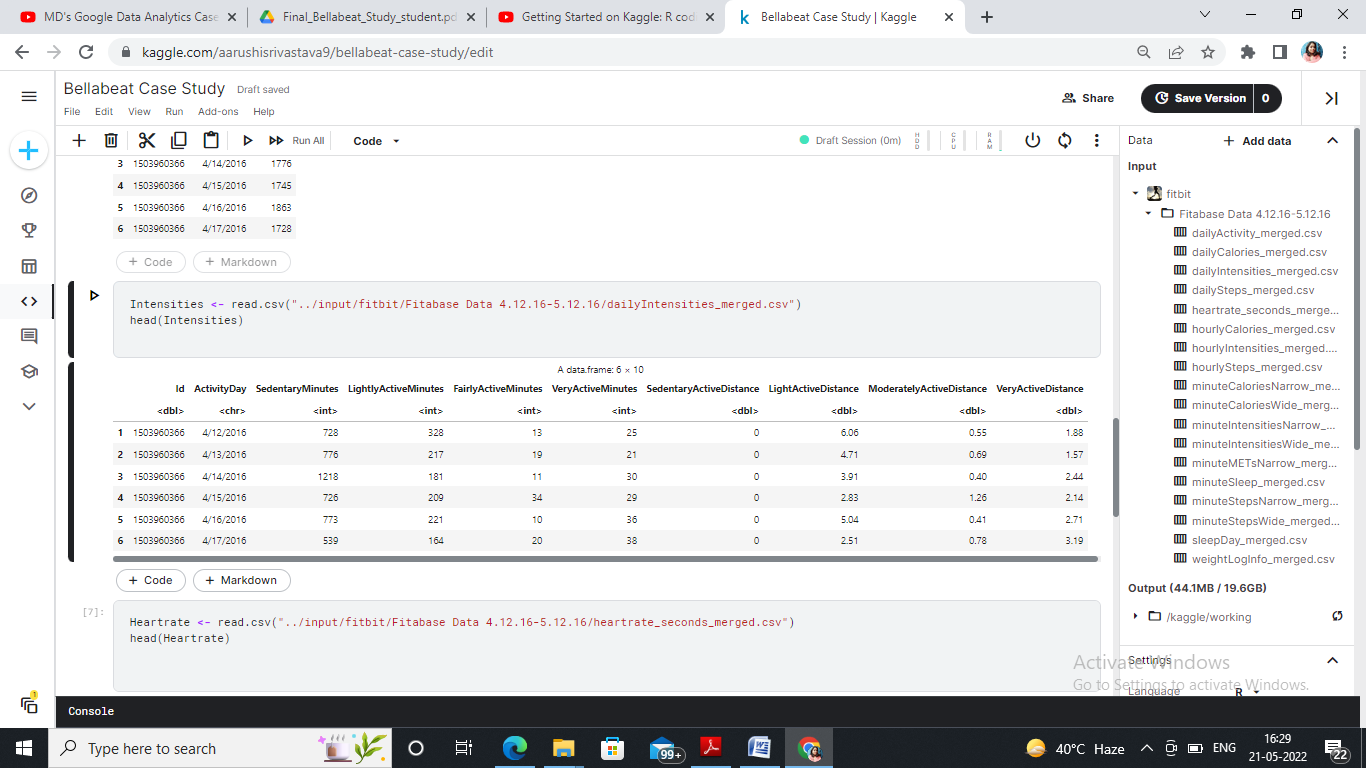
head(Calories)



* **dailyIntensities\_merged.csv**

Intensities <- read.csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyIntensities\_merged.csv")

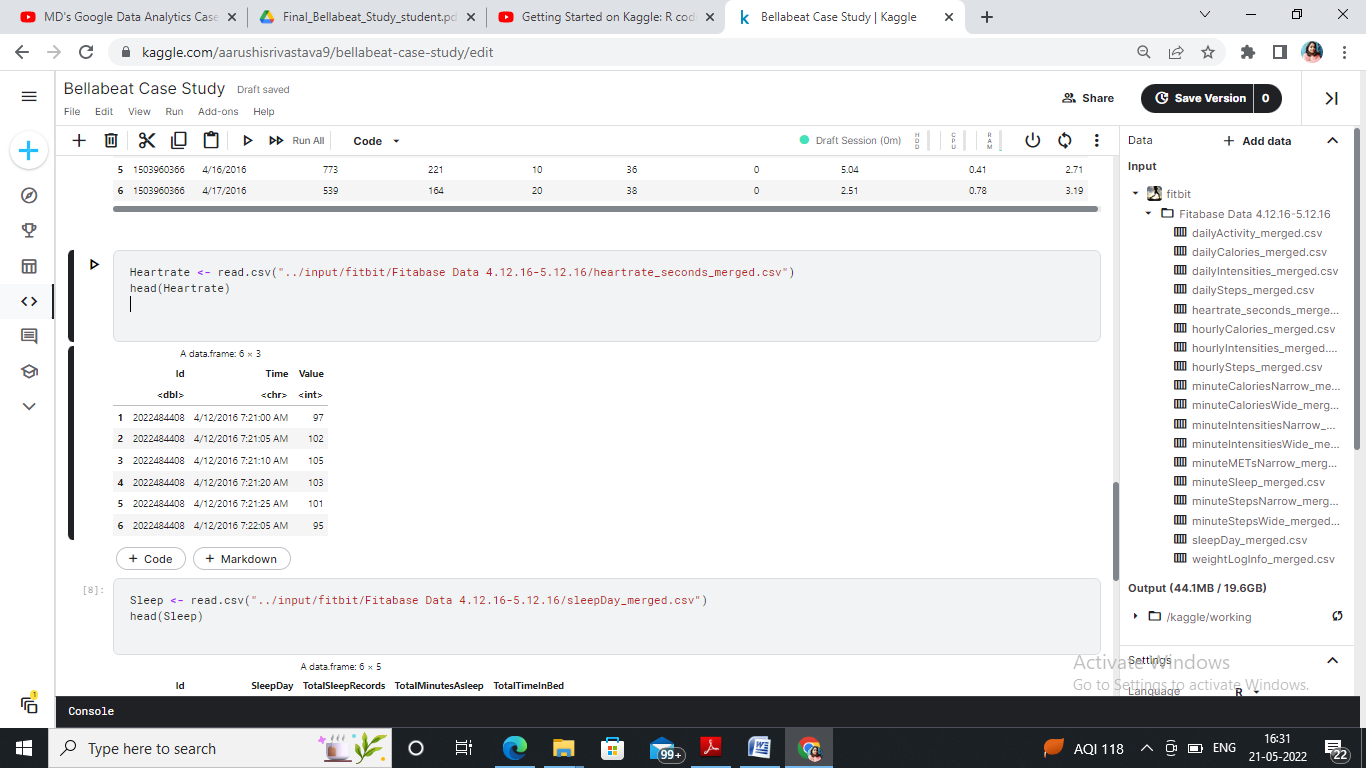
head(Intensities)



* **heartrate\_seconds\_merged.csv**

Heartrate <- read.csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/heartrate\_seconds\_merged.csv")

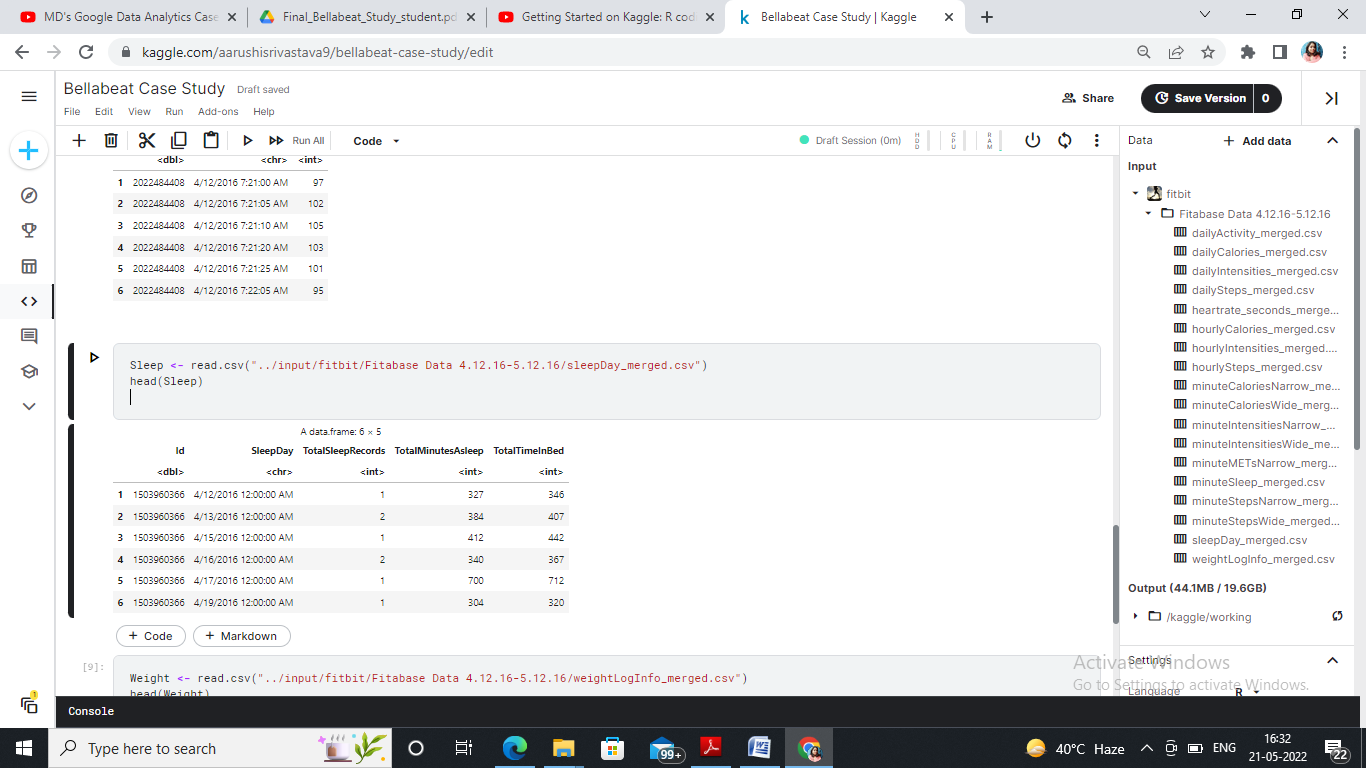
head(Heartrate)



* **sleepDay\_merged.csv**

Sleep <- read.csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/sleepDay\_merged.csv")

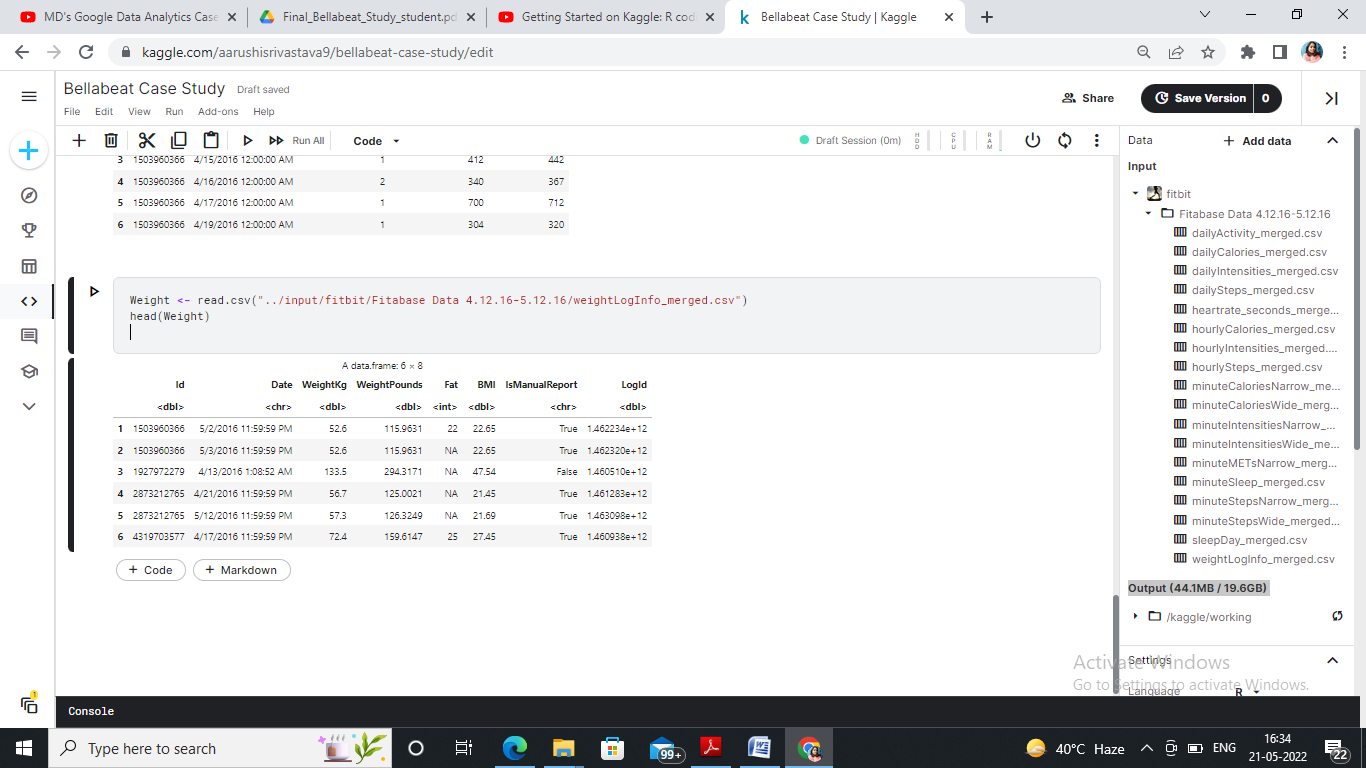
head(Sleep)



* **weightLogInfo\_merged.csv**

Weight <- read.csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/weightLogInfo\_merged.csv")

head(Weight)



So now, we can see that everything were imported correctly.

**Process Phase**

In this phase,I will be Processing, Cleaning and Organizing the dataset for analysis. I used functions like glimpse(),skim\_without\_charts to quickly review the data. To clean the names of the data,I have used clean\_names().

Data Cleaning steps involved with the respective datasets :

1. Daily Activity Dataset : Changed the Activity Date column to Short Date format . Changed the TotalDistance, TrackerDistance, LoggedActivitiesDistance, VeryActiveDistance, ModeratelyActiveDistance, LightActiveDistance and SedentaryActiveDistance columns to Number format with 2 decimal points for ensuring data consistency.
2. Daily Calories Dataset : Changed the Activity Date column to short date datatype.
3. Daily Intensies Dataset : Changed the Activity Date to short date format and SedentaryActiveDistance, LightActiveDistance, ModeratelyActiveDistance and VeryActiveDistance columns to number datatype.
4. Daily Steps Dataset : Modified the Activity Date column to date format.
5. Heartrate Dataset : Modified the Time column to time datatype.
6. Sleep Day Dataset : Changed the SleepDay column to short date format.
7. Weight Log Info Dataset : Changed the Date column to short date format and fixed the decimal value consistency for other columns.

**Analyse Phase**

At this stage, all the data is imported, stored and processed,thus, now I will begin the analysis of the data.

Finding the total number of distinct participants in each dataframe :

n\_distinct(daily\_activity$Id)

## [1] 33

n\_distinct(sleep\_day$Id)

## [1] 24

How many observations are there in each dataframe?

nrow(daily\_activity)

## [1] 940

nrow(sleep\_day)

## [1] 413

Here are some quick summary statistics we might want to know about each data frame.

For the daily activity dataframe:

daily\_activity %>%

select(TotalSteps,

TotalDistance,

SedentaryMinutes) %>%

summary()

## TotalSteps TotalDistance SedentaryMinutes

## Min. : 0 Min. : 0.000 Min. : 0.0

## 1st Qu.: 3790 1st Qu.: 2.620 1st Qu.: 729.8

## Median : 7406 Median : 5.245 Median :1057.5

## Mean : 7638 Mean : 5.490 Mean : 991.2

## 3rd Qu.:10727 3rd Qu.: 7.713 3rd Qu.:1229.5

## Max. :36019 Max. :28.030 Max. :1440.0

For the sleep dataframe:

sleep\_day %>%

select(TotalSleepRecords,

TotalMinutesAsleep,

TotalTimeInBed) %>%

summary()

## TotalSleepRecords TotalMinutesAsleep TotalTimeInBed

## Min. :1.000 Min. : 58.0 Min. : 61.0

## 1st Qu.:1.000 1st Qu.:361.0 1st Qu.:403.0

## Median :1.000 Median :433.0 Median :463.0

## Mean :1.119 Mean :419.5 Mean :458.6

## 3rd Qu.:1.000 3rd Qu.:490.0 3rd Qu.:526.0

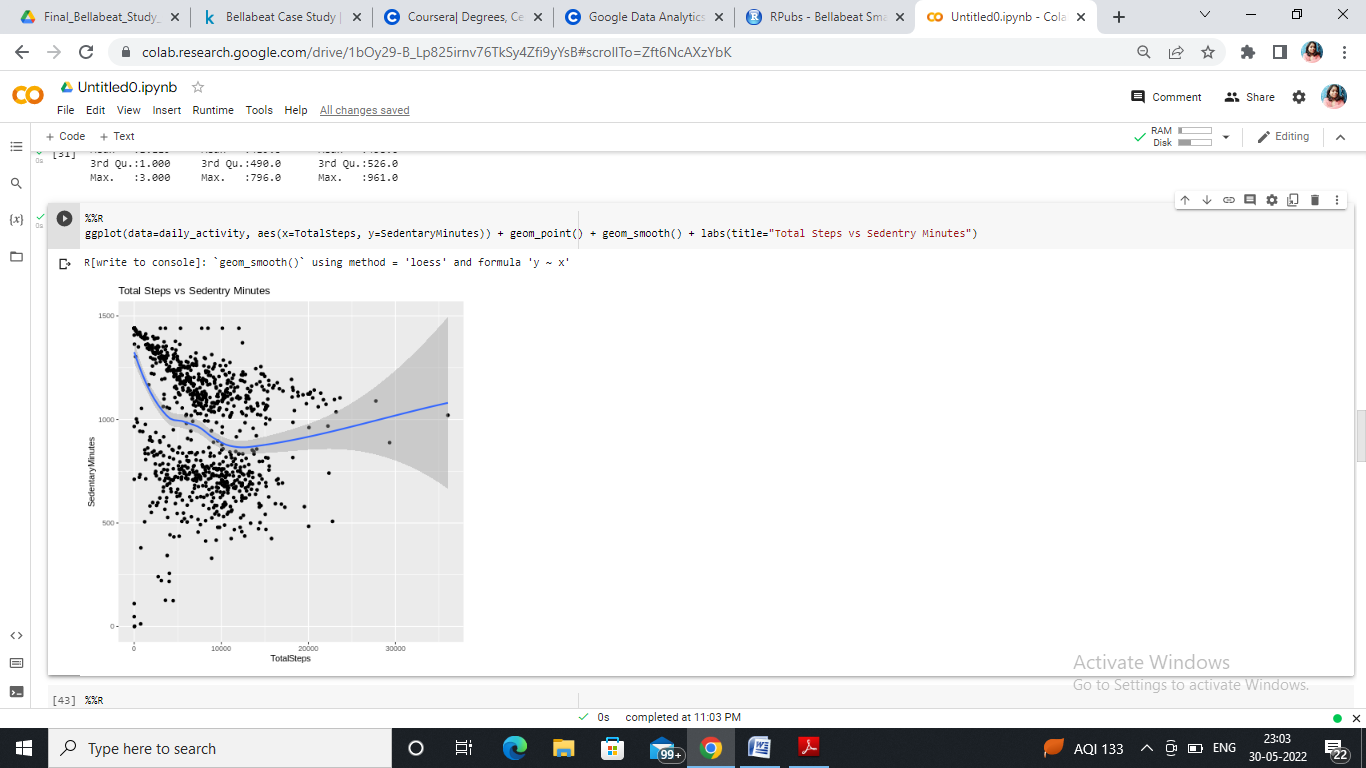
## Max. :3.000 Max. :796.0 Max. :961.0

**Key Findings**

* There are 33 distinct participants in Activity dataframe and only 24 participants in Sleep dataframe.
* The average stepcount is 7638 while the average sedentary mins are 991. Assuming the rate of stroll to be 50 steps per min for any age group, the active minutes will be 152.76 which is very low in comparison to the sedentary minutes.
* The average minutes asleep are 419.5 while average time in bed is 458.6 ,which means on an average people from all age groups spend 39.1 mins in bed trying to sleep which may go as high as 165 mins of inactivity.
* We realize from the demographics that a majority of the participants are lightly active with high sedentary time.Also, we figure out that the same can be a probable reason for participants to struggle falling asleep. Hence we can recommend marketing strategies to bridge the gap.

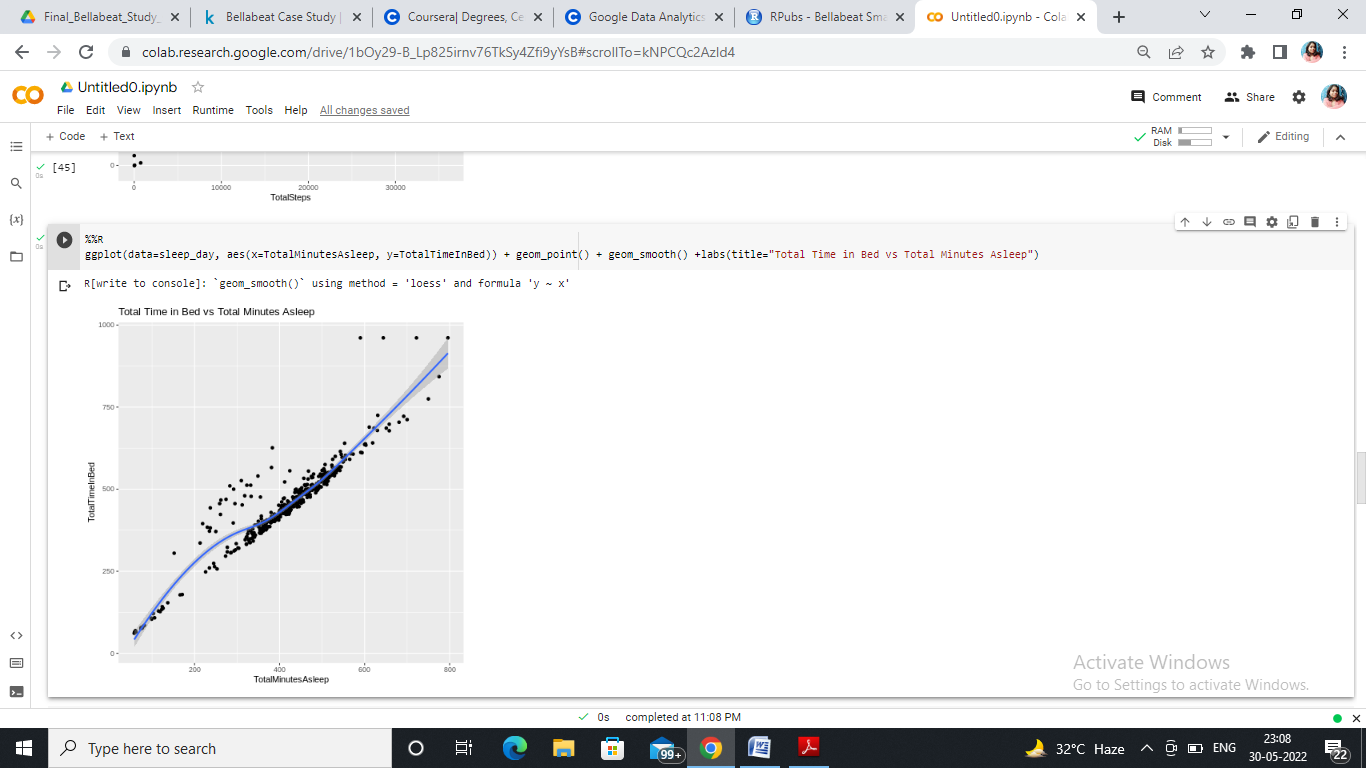
**Share and Act Phase**

**Relationship between steps and sedentary minutes.**



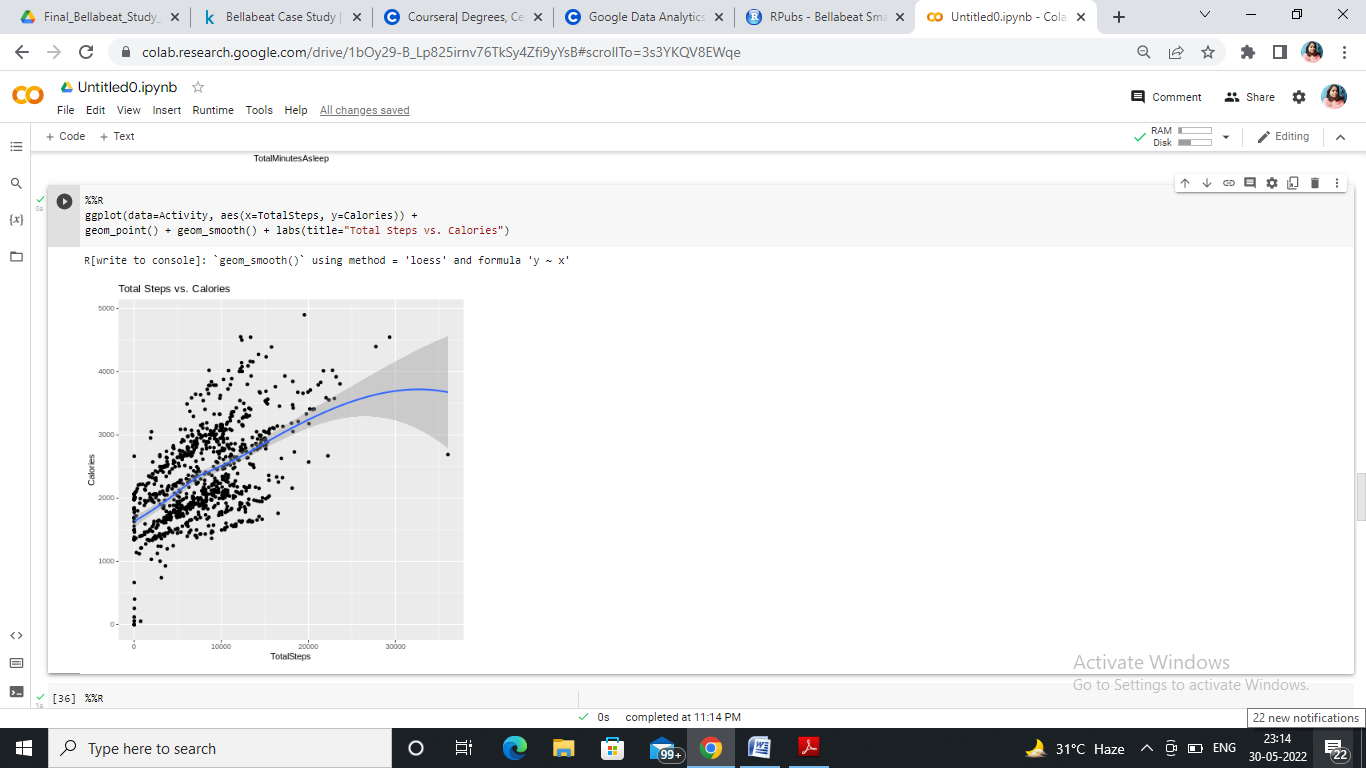
There is a negative correlation between steps and sedentary time. The lesser the steps ,the more the sedentary time .The insight drawn from here suggests the company to market steps recording devices to customers showing higher sedentary minutes.

**Relationship between Minutes Asleep and Time in Bed**



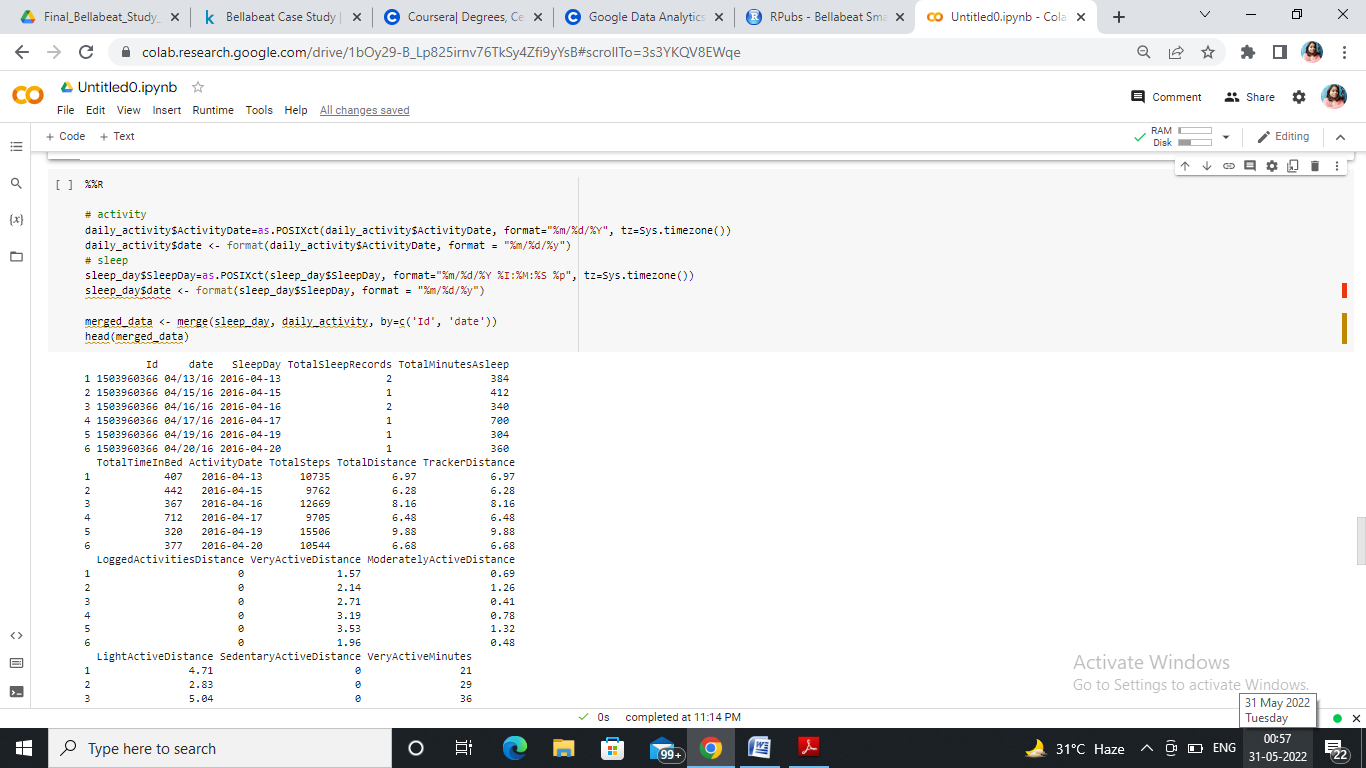
There is almost a linear relationship between Time in Bed and Minutes Asleep which indicates that the more people struggle falling asleep the more they sleep which is both bad for health. The company might want to suggest those customers to take up more strenuous activity so as to reduce time in bed and have a healthy sleep.

**Relationship between Total Steps vs Calories Burned**



From this visualization, we can make out that there is a positive correlation between No of Steps and Calories burned which suggests more activity on a daily basis to remain fit and active.

The really interesting pattern here is that there is a strong correlation between Total time in bed and total minutes asleep, but the higher than expected number of points that have a larger than expected amount of time in bed. Let’s examine this more closely, but first we need to merge these two files together. Because the date fields are not in the same format we have to clean them so that we can combine the two tables by their ID AND by date.



Id date SleepDay TotalSleepRecords TotalMinutesAsleep

1 1503960366 04/13/16 2016-04-13 2 384

2 1503960366 04/15/16 2016-04-15 1 412

3 1503960366 04/16/16 2016-04-16 2 340

4 1503960366 04/17/16 2016-04-17 1 700

5 1503960366 04/19/16 2016-04-19 1 304

6 1503960366 04/20/16 2016-04-20 1 360

TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance

1 407 2016-04-13 10735 6.97 6.97

2 442 2016-04-15 9762 6.28 6.28

3 367 2016-04-16 12669 8.16 8.16

4 712 2016-04-17 9705 6.48 6.48

5 320 2016-04-19 15506 9.88 9.88

6 377 2016-04-20 10544 6.68 6.68

LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance

1 0 1.57 0.69

2 0 2.14 1.26

3 0 2.71 0.41

4 0 3.19 0.78

5 0 3.53 1.32

6 0 1.96 0.48

LightActiveDistance SedentaryActiveDistance VeryActiveMinutes

1 4.71 0 21

2 2.83 0 29

3 5.04 0 36

4 2.51 0 38

5 5.03 0 50

6 4.24 0 28

FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories

1 19 217 776 1797

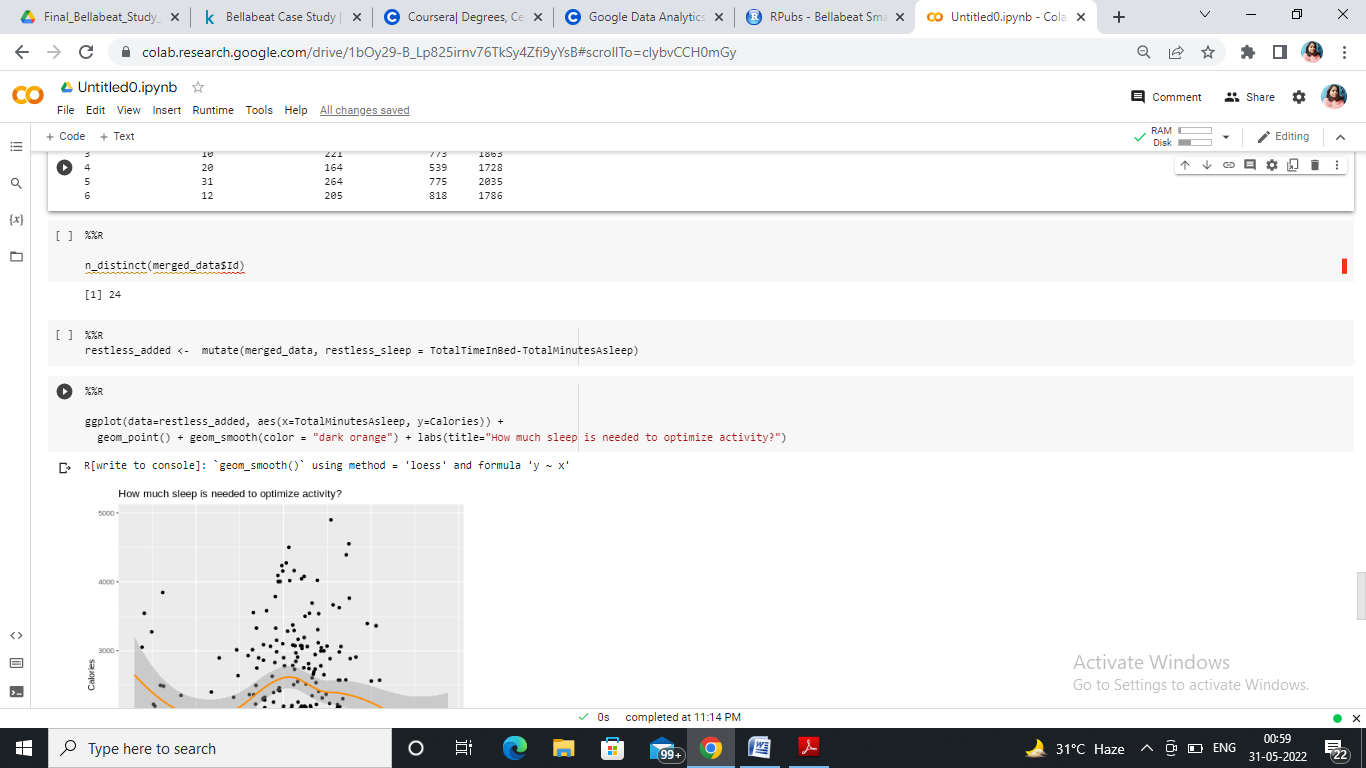
2 34 209 726 1745

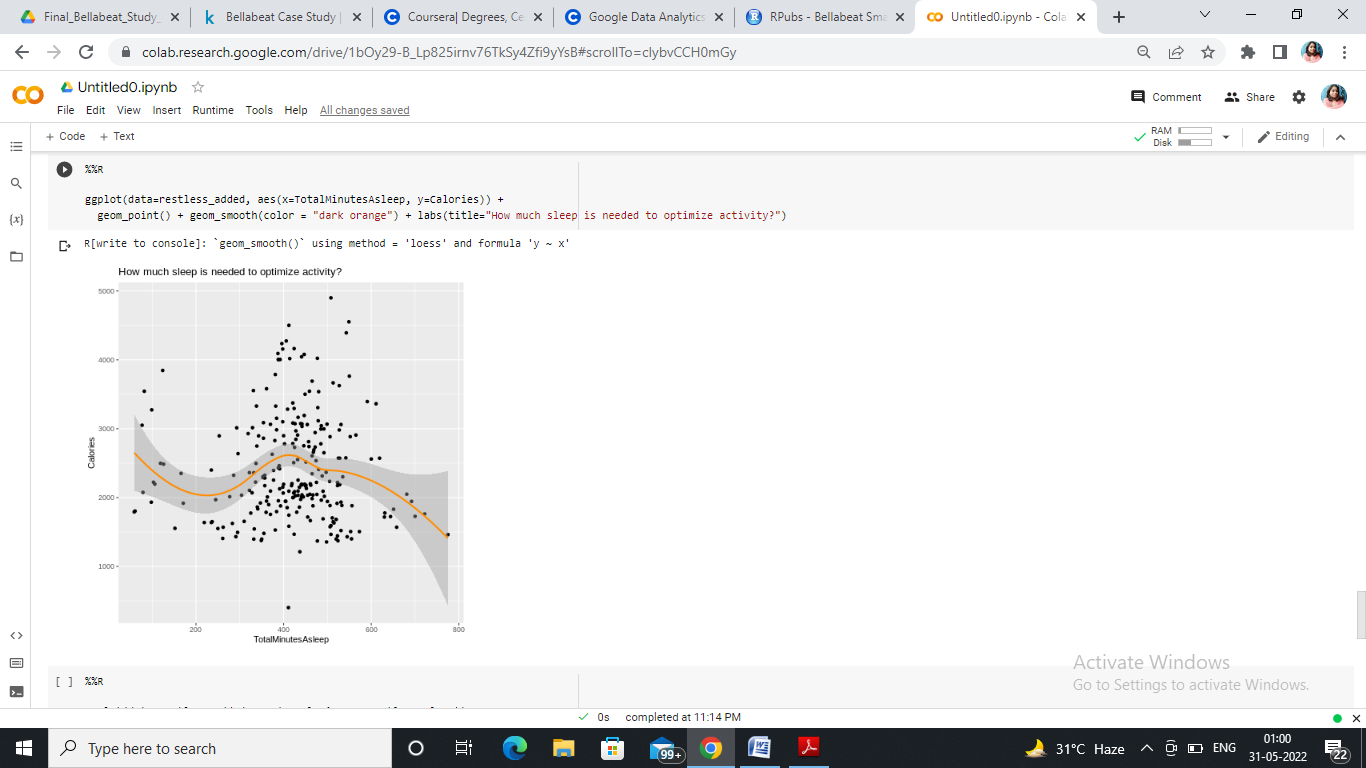
3 10 221 773 1863

4 20 164 539 1728

5 31 264 775 2035

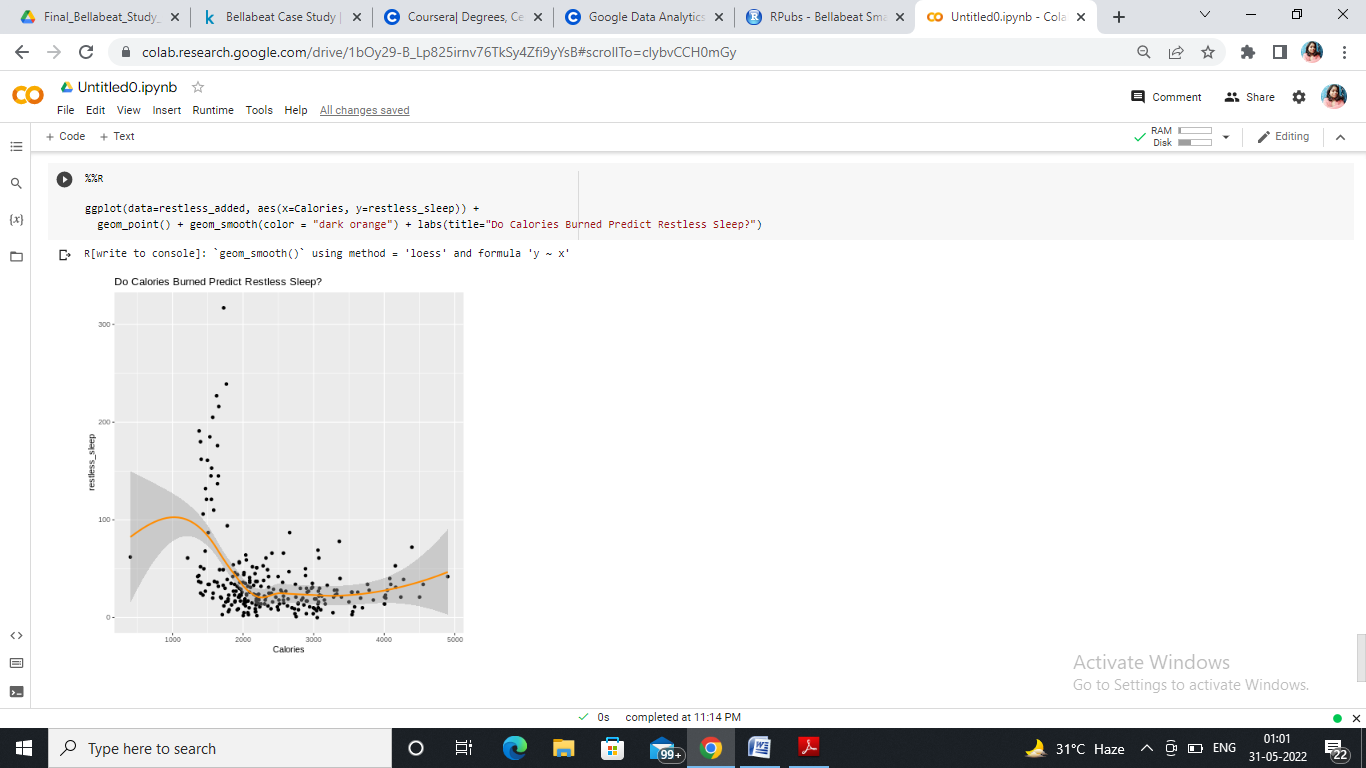
6 12 205 818 1786





It looks like just under 400 minutes (slightly less than 7 hours is where a peak of activity happens). Perhaps we could offer some functionality that begins by automatically setting an alarm for 6 hours and 40 minutes that also learns from individual variation and accommodates for that user based on empirical data?

In addition, we can look at the opposite. How much activity in terms of calories leads to the most efficient sleep, minimzing the amount of time spent in bed while not asleep?



This looks like a really interesting pattern that seems to decrease restless sleep up to around 2000 calories and then levels off. We could certainly push the user to try to achieve that threshold of activity throughout the day to improve their sleep.

**Conclusions & Recommandations for the Business**

So, collecting data on activity, sleep, stress, etc. will allow the company Bellabeat to empower the customers with knowledge about their own health and daily habits. The company Bellabeat is growing rapidly and quickly positioned itself as a tech-driven wellness company for their customers.

By analyzing the FitBit Fitness Tracker Data set, I found some insights that would help influence Bellabeat marketing strategy.

**Target Audience:**

People working full-time jobs and spending a lot of time at the computer and in the office and need fitness and daily activities to be in shape.The users are doing some light activity to stay healthy (according to the activity type analysis). And they need to improve their everyday activity to have more health benefits. And they might need some knowledge about developing healthy habits and motivation to keep them going.

**Message to the Company**

The Bellabeat app need to be a unique fitness activity app. By becoming a companion guide (like a friend) to its users and customers and help them balance their personal and professional life with healthy habits.

**Primary recommendations**

1. Users should try to get 6 hours and 40 minutes of sleep to optimize the amoutn of activity they can get in during a day. We suggest that Bellebeat use prompts to help a user achieve this amoutn of sleep.
2. To avoid restless sleep, a user should try to burn at least 2000 calories in a day. We suggest that the device offer prompts to encourage this level of activity throughout the day.
3. In concert, these two prompts have the potential to lead to greater customer loyalty and ultimately market share gains through “word of mouth” and specific campaigns championing this new functionality.