BELLABEAT CAPSTONE PROJECT

Background

Bellabeat is a high-tech company that manufactures health-focused smart products and is founded by Urska Srsen (Chief Creative Officer) and Sando Mur (Mathematician and key member of Bellabeat executive team). For this mock project, I am working as a Junior Data Analyst in Bellabeat's marketing analyst team and have been given the task to analyse smart device usage data to unlock new growth opportunities for the company. The results from the analysis will be presented to the company's executive team along with high-level recommendations for Bellabeat's marketing strategy.

PHASE 1: ASK

In this phase, a clear business task statement needs to be determined to align to and guide the entire project. I have been given a task by Srsen to look into non-Bellabeat smart devices user data and see what insights can be discovered and applied on one of Bellabeat products. With this in mind, I have came up with the following business task statement:

To determine which products within the Bellabeat ecosystem to market to its existing customers by analyzing trends from publicly available data of non-Bellabeat smart devices and gaining insights from the data to drive Bellabeat's marketing strategy for that Bellabeat product.

PHASE 2: PREPARE

For this analysis, I used a publically available dataset on Fitbit Fitness tracker data, provided by Mobius on Kaggle. The dataset consists of thirty eligible Fitbit users who consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It also includes information about daily activity, steps, and hear rate that can be used to explore users' habits.

Out of the 18 datasets provided, I chose 6 datasets for my analysis, which are the daily data (i.e. daily activity, intensity, sleep, steps, calories) and user weight log.

In [1]:
```
library(tidyverse)
daily_activity <- read_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")
daily_sleep <- read_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
user_weight_log <- read_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/weightLogInfo_merged.csv")
daily_steps <- read_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailySteps_merged.csv")
daily_intensities <- read_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyIntensities_merged.csv")
daily_calories <- read_csv("../input/fitbit/Fitabase Data 4.12.16-5.12.16/dailyCalories_merged.csv")
```

```
── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1
──

✓ ggplot2 3.3.5     ✓ purrr   0.3.4
✓ tibble  3.1.5     ✓ dplyr   1.0.7
✓ tidyr   1.1.4     ✓ stringr 1.4.0
✓ readr   2.0.2     ✓ forcats 0.5.1

── Conflicts ────────────────────────────────────────── tidyverse_conflicts()
──
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()

Rows: 940 Columns: 15

── Column specification ────────────────────────────────────────────────────
──
Delimiter: ","
chr  (1): ActivityDate
dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesD
i...


ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this messa
ge.

Rows: 413 Columns: 5

── Column specification ────────────────────────────────────────────────────
──
Delimiter: ","
chr (1): SleepDay
dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed


ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this messa
ge.

Rows: 67 Columns: 8

── Column specification ────────────────────────────────────────────────────
──
Delimiter: ","
chr (1): Date
dbl (6): Id, WeightKg, WeightPounds, Fat, BMI, LogId
lgl (1): IsManualReport


ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this messa
ge.

Rows: 940 Columns: 3

── Column specification ────────────────────────────────────────────────────
```

```
──
Delimiter: ","
chr (1): ActivityDay
dbl (2): Id, StepTotal


ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this messa
ge.

Rows: 940 Columns: 10

── Column specification ────────────────────────────────────────────────────────
──
Delimiter: ","
chr (1): ActivityDay
dbl (9): Id, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMinutes, V
e...


ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this messa
ge.

Rows: 940 Columns: 3

── Column specification ────────────────────────────────────────────────────────
──
Delimiter: ","
chr (1): ActivityDay
dbl (2): Id, Calories


ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this messa
ge.
```

Before proceeding with the analysis, there are a few limitations to the data that should be highlighted:

1. SAMPLE SIZE: The sample size from this dataset involves only 30 individuals, which is just enough to be statiscally significant but would be insufficient to drive a key business marketing decision. The insights uncovered from this dataset should be explored further with a much larger dataset, preferably from Bellabeat's own user data.
2. TIMEFRAME: The datasets only covers a period from 12 March to 12 May 2016, which is just 62 days. Longer periods of at least 6 months data are needed as they are useful in establishing habits and behaviour in individuals.
3. MISSING METRICS: Some key user stats such as gender, age, height, weight and ethnicity, which are useful to the overall analysis, are not available.
4. BIAS: The nature of how this datset is obtained may create a bias towards more active and participative users as these users volunteer their data to be tracked and published in a public database.
5. FITBIT AND BELLABEAT: This dataset comes from Fitbit devices which may consists of users from all genders (due to the absence of gender data, the exact gender composition cannot be determined) while Bellabeat devices are marketed mostly towards female.

PHASE 3: PROCESS

The first cleaning process I did was to convert the dates within the dataset to datetime format and also standardised the column name of the dates to 'Date'. With the dates and column name now standardised across the datasets, I can merge all the daily datasets into a single dataset later on for easier analysis.

In [2]:
```
daily_activity$ActivityDate <- as.Date(daily_activity$ActivityDate, format = "%m/%d/%Y")
daily_sleep <- daily_sleep %>% mutate(SleepDay = as.Date(SleepDay,format="%m/%d/%Y"))
daily_activity <- rename(daily_activity, Date = ActivityDate)
daily_sleep <- rename(daily_sleep, Date = SleepDay)
```

PHASE 4 & 5: ANALYZE AND SHARE

The first thing I wanted to do after cleaning my data was to understand the users who provided the data. To do this, I group the data within each daily dataset by user's unique ID, average the metrics (i.e. steps, intensities, calories, sleep and BMI) across the dates and merge these data together into a single dataframe.

In [3]:
```
user_avg_BMI <- user_weight_log %>% group_by(Id) %>% summarise(avgBMI = mean(BMI))
user_avg_activeness <- daily_activity %>% group_by(Id) %>% summarise(avgSteps = mean(TotalSteps),avgVA = mean(VeryActiveMinutes),avgFA = mean(FairlyActiveMinutes),avgLA = mean(LightlyActiveMinutes),avgSed=mean(SedentaryMinutes),avgCalories = mean(Calories))
user_avg_sleep <- daily_sleep %>% group_by(Id) %>% filter(n() > 3) %>% summarise(avgSleep = mean(TotalMinutesAsleep))
user_summary <- merge(merge(user_avg_activeness,user_avg_sleep,by="Id",all=TRUE),user_avg_BMI,by="Id",all=TRUE)
user_analysis <-user_summary[,c("Id","avgSteps","avgVA","avgFA","avgLA","avgSed","avgCalories","avgSleep","avgBMI")]
#options(repr.plot.width = 20, repr.plot.height = 20)
#ggpairs(user_analysis)
#avgPctActiveMins = mean((VeryActiveMinutes + FairlyActiveMinutes+LightlyActiveMinutes)/(VeryActiveMinutes + FairlyActiveMinutes+LightlyActiveMinutes+SedentaryMinutes)), avgPctHighActive = mean((VeryActiveMinutes + FairlyActiveMinutes)/(VeryActiveMinutes + FairlyActiveMinutes+LightlyActiveMinutes))
print(user_analysis)
summary(user_analysis)
```

```
           Id  avgSteps       avgVA      avgFA     avgLA    avgSed avgCalorie
s
1  1503960366 12116.742 38.70967742 19.1612903 219.93548  848.1613    1816.41
9
2  1624580081  5743.903  8.67741935  5.8064516 153.48387 1257.7419    1483.35
5
3  1644430081  7282.967  9.56666667 21.3666667 178.46667 1161.8667    2811.30
0
4  1844505072  2580.065  0.12903226  1.2903226 115.45161 1206.6129    1573.48
4
5  1927972279   916.129  1.32258065  0.7741935  38.58065 1317.4194    2172.80
6
6  2022484408 11370.645 36.29032258 19.3548387 257.45161 1112.5806    2509.96
8
7  2026352035  5566.871  0.09677419  0.2580645 256.64516  689.4194    1540.64
5
8  2320127002  4716.871  1.35483871  2.5806452 198.19355 1220.0968    1724.16
1
9  2347167796  9519.667 13.50000000 20.5555556 252.50000  687.1667    2043.44
4
10 2873212765  7555.774 14.09677419  6.1290323 308.00000 1097.1935    1916.96
8
11 3372868164  6861.650  9.15000000  4.1000000 327.90000 1077.5500    1933.10
0
12 3977333714 10984.567 18.90000000 61.2666667 174.76667  707.5333    1513.66
7
13 4020332650  2267.226  5.19354839  5.3548387  76.93548 1237.2581    2385.80
6
14 4057192912  3838.000  0.75000000  1.5000000 103.00000 1217.2500    1973.75
0
15 4319703577  7268.839  3.58064516 12.3225806 228.77419  735.8065    2037.67
7
16 4388161847 10813.935 23.16129032 20.3548387 229.35484  836.6774    3093.87
1
17 4445114986  4796.548  6.61290323  1.7419355 209.09677  829.9032    2186.19
4
18 4558609924  7685.129 10.38709677 13.7096774 284.96774 1093.6129    2033.25
8
19 4702921684  8572.065  5.12903226 26.0322581 237.48387  766.4194    2965.54
8
20 5553957443  8612.581 23.41935484 13.0000000 206.19355  668.3548    1875.67
7
21 5577150313  8304.433 87.33333333 29.8333333 147.93333  754.4333    3359.63
3
22 6117666160  7046.714  1.57142857  2.0357143 288.35714  796.2857    2261.14
3
23 6290855005  5649.552  2.75862069  3.7931034 227.44828 1193.0345    2599.62
1
24 6775888955  2519.692 11.00000000 14.8076923  40.15385 1299.4231    2131.76
9
25 6962181067  9794.806 22.80645161 18.5161290 245.80645  662.3226    1982.03
2
26 7007744171 11323.423 31.03846154 16.2692308 280.73077 1055.3462    2544.00
0
27 7086361926  9371.774 42.58064516 25.3548387 143.83871  850.4516    2566.35
5
28 8053475328 14763.290 85.16129032  9.5806452 150.96774 1148.0000    2945.80
```

```
6
29 8253242879  6482.158 20.52631579 14.3157895 116.89474 1287.3684    1788.00
0
30 8378563200  8717.710 58.67741935 10.2580645 156.09677  716.1290    3436.58
1
31 8583815059  7198.516  9.67741935 22.1935484 138.29032 1267.2258    2732.03
2
32 8792009665  1853.724  0.96551724  4.0344828  91.79310 1060.4828    1962.31
0
33 8877689391 16040.032 66.06451613  9.9354839 234.70968 1112.8710    3420.25
8
   avgSleep   avgBMI
1  360.2800 22.65000
2        NA       NA
3  294.0000       NA
4        NA       NA
5  417.0000 47.54000
6        NA       NA
7  506.1786       NA
8        NA       NA
9  446.8000       NA
10       NA 21.57000
11       NA       NA
12 293.6429       NA
13 349.3750       NA
14       NA       NA
15 476.6538 27.41500
16 403.1250       NA
17 385.1786       NA
18 127.6000 27.21400
19 421.1429       NA
20 463.4839       NA
21 432.0000 28.00000
22 478.7778       NA
23       NA       NA
24       NA       NA
25 448.0000 24.02800
26       NA       NA
27 453.1250       NA
28       NA       NA
29       NA       NA
30 443.3438       NA
31       NA       NA
32 435.6667       NA
33       NA 25.48708
```

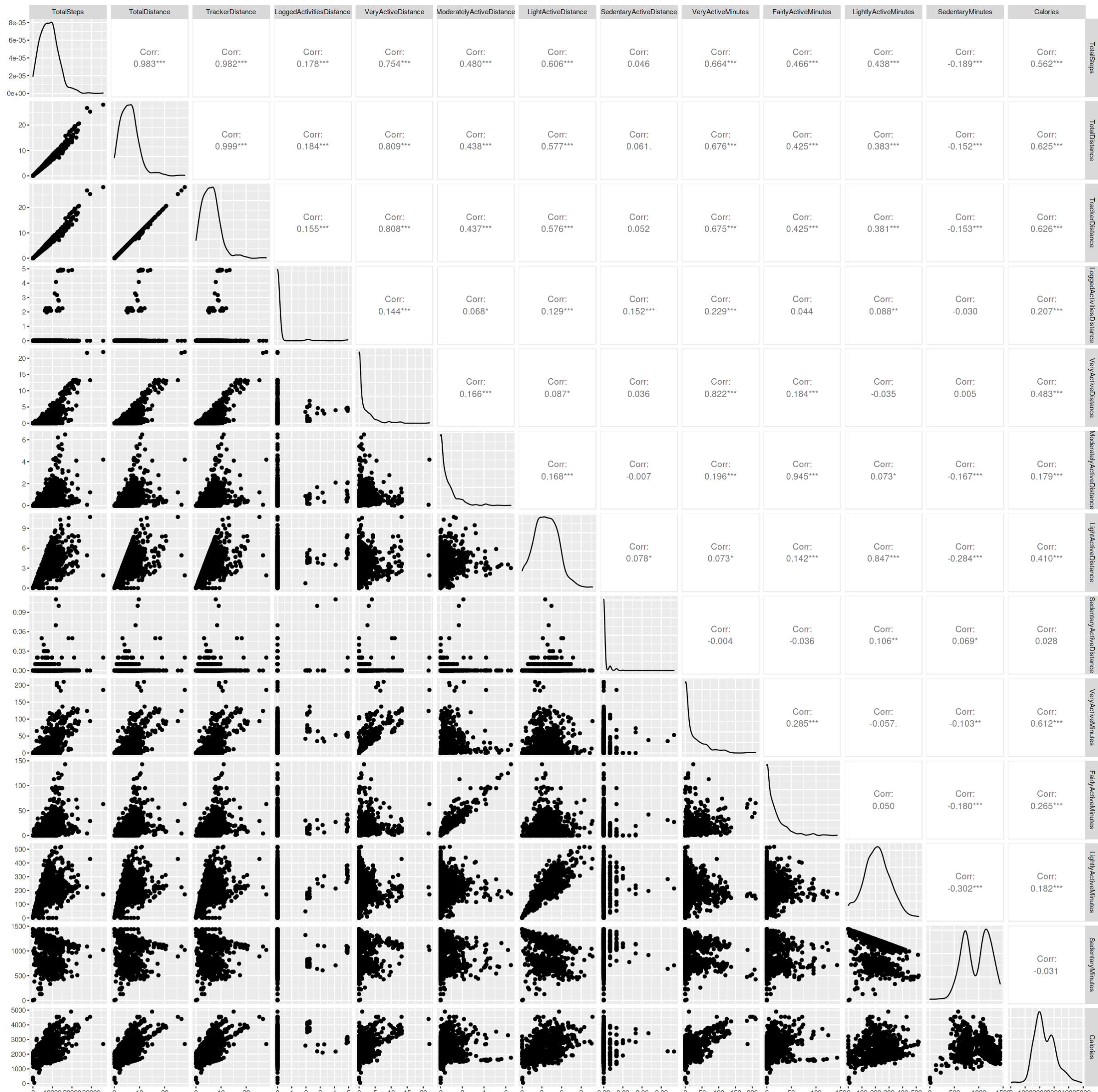
```
       Id               avgSteps           avgVA              avgFA        
 Min.   :1.504e+09   Min.   :  916.1   Min.   : 0.09677   Min.   : 0.2581  
 1st Qu.:2.347e+09   1st Qu.: 5566.9   1st Qu.: 3.58065   1st Qu.: 4.0345  
 Median :4.445e+09   Median : 7283.0   Median :10.38710   Median :12.3226  
 Mean   :4.857e+09   Mean   : 7519.3   Mean   :20.30877   Mean   :13.2602  
 3rd Qu.:6.962e+09   3rd Qu.: 9519.7   3rd Qu.:23.41935   3rd Qu.:19.3548  
 Max.   :8.878e+09   Max.   :16040.0   Max.   :87.33333   Max.   :61.2667  
                                                                           
     avgLA            avgSed         avgCalories      avgSleep    
 Min.   : 38.58   Min.   : 662.3   Min.   :1483   Min.   :127.6  
 1st Qu.:143.84   1st Qu.: 766.4   1st Qu.:1917   1st Qu.:372.7  
 Median :206.19   Median :1077.5   Median :2132   Median :432.0  
 Mean   :191.52   Mean   : 999.2   Mean   :2282   Mean   :401.9  
 3rd Qu.:245.81   3rd Qu.:1206.6   3rd Qu.:2600   3rd Qu.:450.6  
 Max.   :327.90   Max.   :1317.4   Max.   :3437   Max.   :506.2  
                                                  NA's   :14     
     avgBMI     
 Min.   :21.57  
 1st Qu.:23.68  
 Median :26.35  
 Mean   :27.99  
 3rd Qu.:27.56  
 Max.   :47.54  
 NA's   :25     
```

Who are our users?

1. ID: There are 33 unique ID for the datasets. However, this differs from the stated sample size in the dataset description of 30 individuals. There is a chance that two or more different fitness devices could come from the same individual. For the purpose of my analysis, I will treat each unique ID as a separate individual.
2. STEPS: The 33 users in the sample size are lightly active individuals with an average of 7,519 steps per day. However, this falls below the recommended 10,000 steps per day recommendation by most health professionals and organisation.
3. ACTIVITY INTENSITY (in Minutes): Among the three activity intensities, the users spend a big proportion of active time on Light Activity at an average of 191.52 minutes or about 3 hours 12 minutes per day.
4. CALORIES: The users burned an average of 2,282 calories per day. A conclusion on the health benefits cannot be made with just the calories number by itself without considering the gender, age, height, weight, and dietary calorie intake of the individual.
5. SLEEP: Among the available user sleep data available and also after excluding those that only key in less than 3 days worth of data (22 users), the users have an average of about 406 minutes of sleep, or about 6 hours 45 minutes. This is lower than the recommended sleep duration of at least 7 hours per night (https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html (https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html)).
6. BMI: Out of the sample size of 33 users, only 8 users actually keyed in their BMI data. From this small sample, the average BMI was found to be 27.99, which falls under the 'Overweight' category.

In [4]:
```
library("GGally")
activity_pair_analysis <- daily_activity[,c(3:15)]
activity_pair_analysis <- subset(activity_pair_analysis,TotalSteps != 0)
options(repr.plot.width = 20, repr.plot.height = 20)
ggpairs(activity_pair_analysis)
```

```
Registered S3 method overwritten by 'GGally':
  method from
  +.gg   ggplot2
```



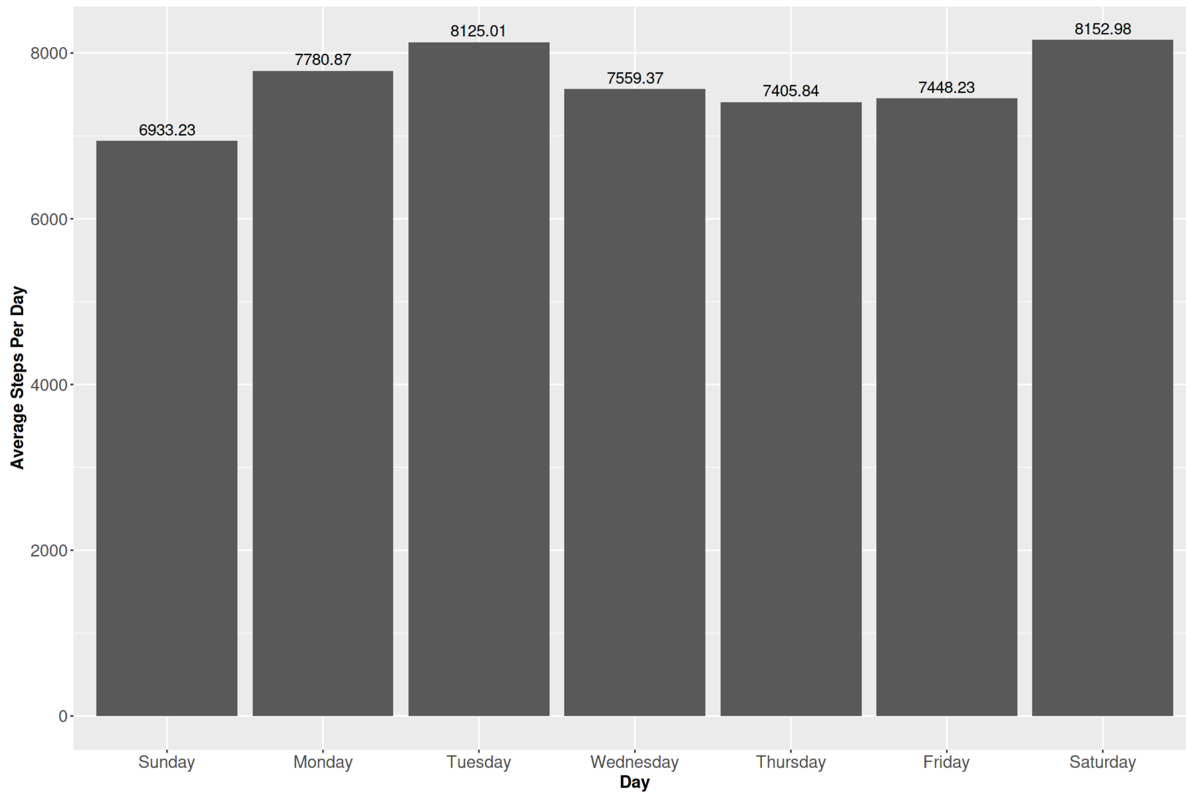
Pairwise analysis of the Daily Activity dataset

1. Activity Minutes (i.e. Very Active, Fairly Active and Lightly Active Minutes) and its correspondonding Activity Distance (i.e. Very Active, Moderately Active and Lightly Active Distance) have a high correlation (above 0.8). Due to the high correlation and Activity Minutes by intensity providing a clearer and direct link to health benefits than Activity Distance by intensity (30 minutes of medium intensity exercise vs 5km of medium intensity exercise), I will be focusing on the former rather than the latter for my analysis.
2. The data for TotalDistance and TrackerDistance have very high correlation, showing that they are nearly identical (close to 1) and could essentially be the same metrics with insignificant differences. Based on the metadata available from Fitabase (https://www.fitabase.com/media/1930/fitabasedatadictionary102320.pdf (https://www.fitabase.com/media/1930/fitabasedatadictionary102320.pdf)), TrackerDistance is distance tracked only by the Fitbit device while TotalDistance would also include user input distance.
3. Total calories burned has the highest correlation with Total Distance and Very Active Minutes, and lowest correlation with Sedentary Minutes. This is in line with the notion that longer and more intense exercise burns the most calories.

Analysing the Activity by Days of the Week

Next, I look at the Daily Activity data grouped by the days of the week. I assigned the days of the week to the dates within the dataframe and group the data by it. Then, I get the average of the key metrics (i.e. Steps, Distance, Activity Intensity Minutes) by day.

In [5]:
```
daily_activity$Day <- weekdays(daily_activity$Date)
daysOrder <- c("Sunday","Monday","Tuesday","Wednesday","Thursday","Friday","Saturday")
step_intensity_analysis <- daily_activity %>% group_by(Day) %>% summarise(avgSteps=mean(TotalSteps),avgDistance=mean(TotalDistance),avgVAMins = mean(VeryActiveMinutes), avgFAMins = mean(FairlyActiveMinutes), avgLAMins = mean(LightlyActiveMinutes), avgSedMins = mean(SedentaryMinutes), avgCal = mean(Calories))
step_intensity_analysis$Day <- factor(step_intensity_analysis$Day, levels = daysOrder)
step_intensity_analysis <- step_intensity_analysis[order(step_intensity_analysis$Day),]
options(repr.plot.width = 15, repr.plot.height = 10)
ggplot(step_intensity_analysis,aes(x=Day,y=avgSteps))+geom_col()+geom_text(aes(label=sprintf("%0.2f",round(avgSteps,digits=2))),size=5,vjust = -0.5)+theme(axis.text=element_text(size=15),
        axis.title=element_text(size=15,face="bold"))+labs(y="Average Steps Per Day")
```



First thing I did was to look at the steps data by day:

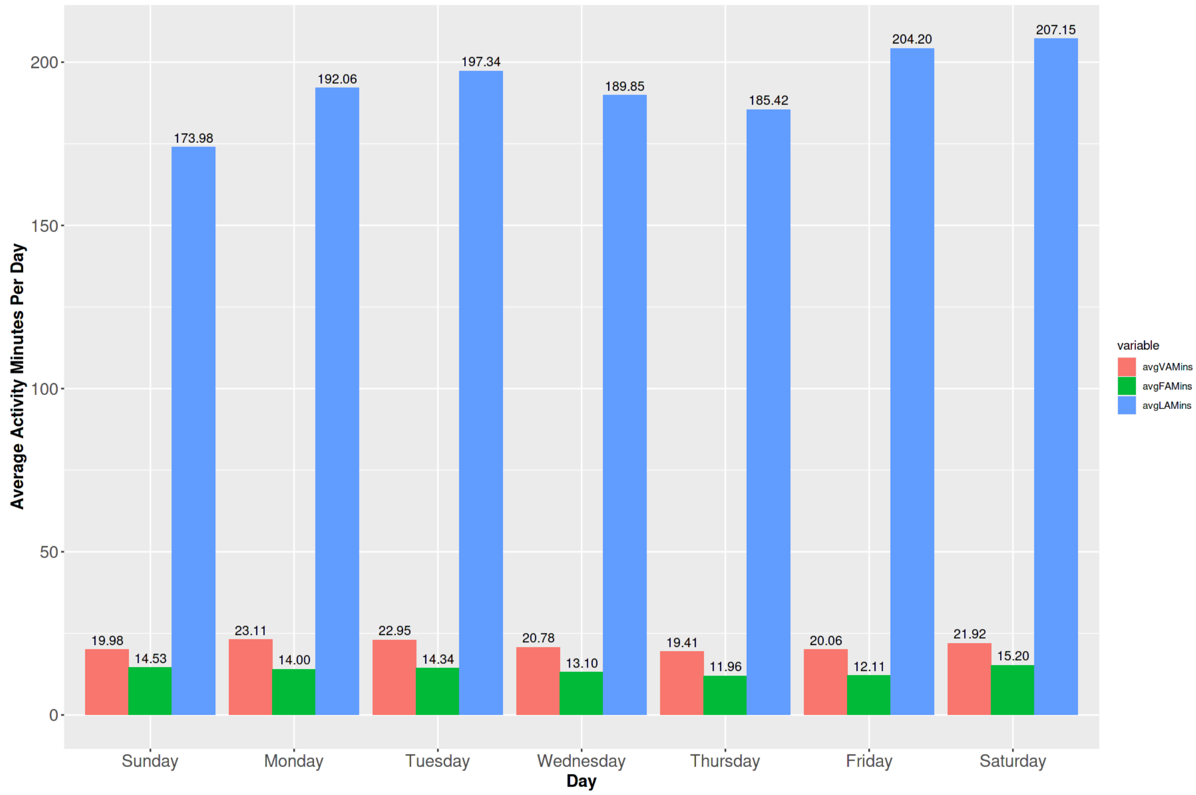
1. From the steps data, it can be seen that Saturday clocks in the highest number of steps while Sunday is the lowest. This could be attributed to Saturdays being a non-working day and individuals are free to move about either for leisure and/or to dedicate more time into physical activities. Many regarded Sundays as rest days, thus the low average steps.
2. There is also a significant rise in the average steps after Sunday, peaking at about 8,125 steps on Tuesday before reducing to about 7,500 steps for the rest of the weekdays. This could be due to the fact that Monday can be seen to best day to start their physical activity routine after resting on Sunday, with momentum helping to push the average steps up on Tuesday.
3. The reduction that happens after Tuesday could be due to individuals 'normalising' their physical activity after Tuesday spike.

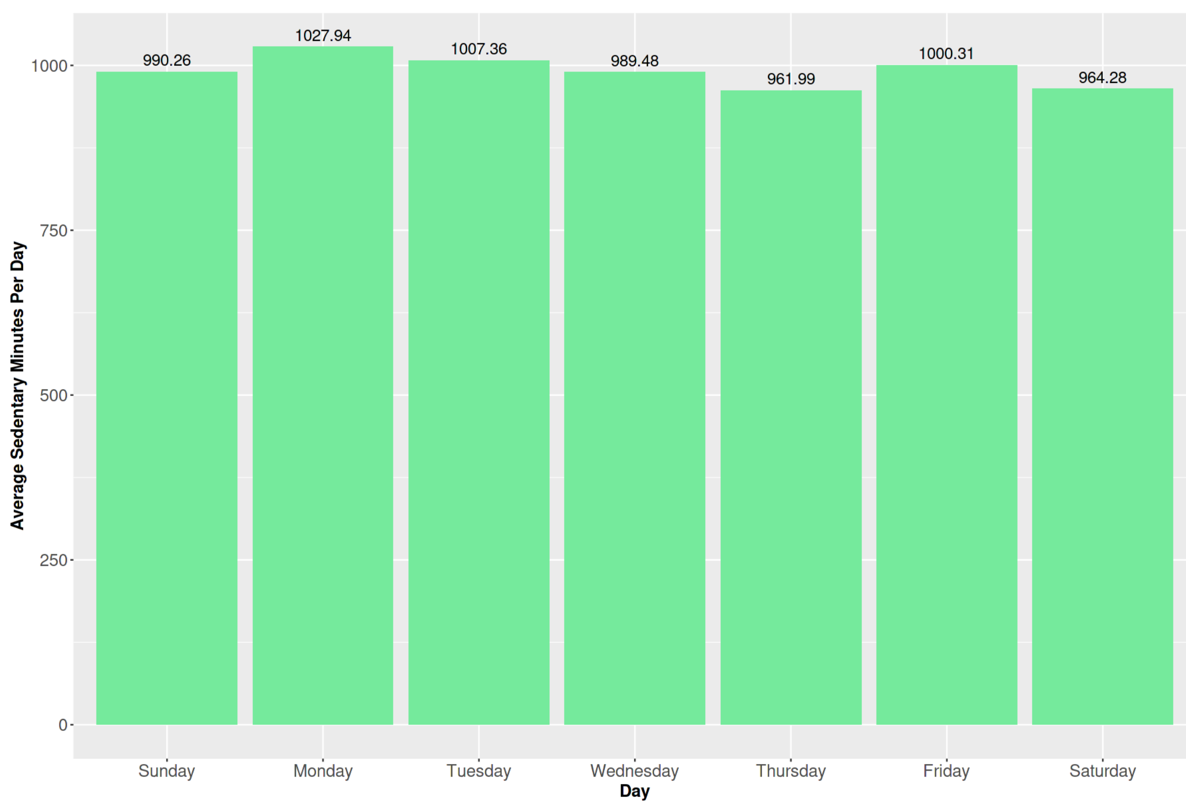
In [6]:
```
library(reshape2)
intensity_analysis <- step_intensity_analysis[,!(names(step_intensity_analysis
) %in% c("avgSteps","avgDistance","avgSedMins","avgCal"))]
intensity_visual_analysis <- melt(intensity_analysis,id.vars = 'Day')
options(repr.plot.width = 15, repr.plot.height = 10)
ggplot(intensity_visual_analysis,aes(x=Day,y=value,fill=variable))+geom_bar(st
at='identity',position='dodge')+geom_text(aes(label=sprintf("%0.2f",round(valu
e,digits=2))),size=4,vjust = -0.5,position=position_dodge(.9))+theme(axis.text
=element_text(size=15),
        axis.title=element_text(size=15,face="bold"))+labs(y="Average Activity
Minutes Per Day")
ggplot(step_intensity_analysis,aes(x=Day,y=avgSedMins))+geom_col(fill = "#75ea
9c")+geom_text(aes(label=sprintf("%0.2f",round(avgSedMins,digits=2))),size=5,v
just = -0.5)+theme(axis.text=element_text(size=15),
        axis.title=element_text(size=15,face="bold"))+labs(y="Average Sedentar
y Minutes Per Day")
ggplot(step_intensity_analysis,aes(x=Day,y=avgCal))+geom_col(fill="#f44336")+g
eom_text(aes(label=sprintf("%0.2f",round(avgCal,digits=2))),size=5,vjust = -0.
5)+theme(axis.text=element_text(size=15),
        axis.title=element_text(size=15,face="bold"))+labs(y="Average Calories
Burned Per Day")
```

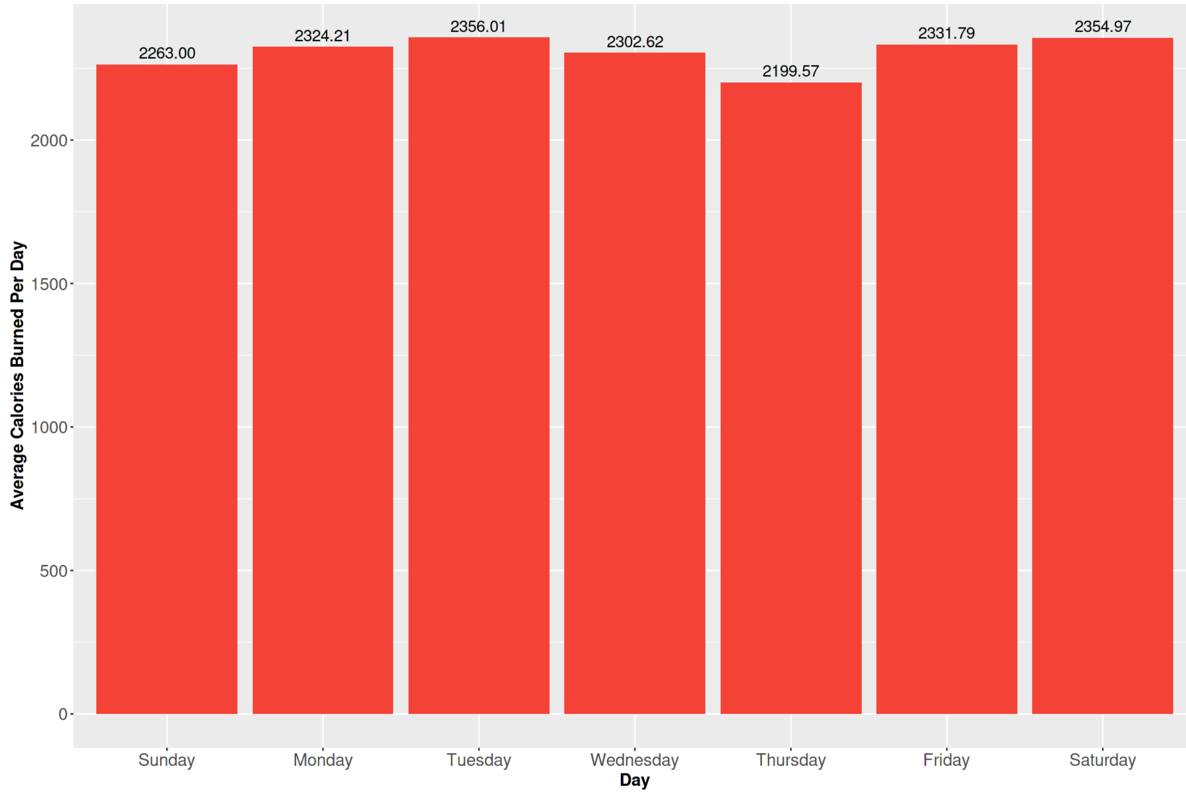
```
Attaching package: ‘reshape2’


The following object is masked from ‘package:tidyr’:

    smiths
```







I then look at the Average Activity Intensity Minutes, Average Sendentary Minutes and Average Calories Burned by Days of the Week:

1. The differences in time spent on Light Intensity activity over Fairly Active and Very Active Intensity activities are more prominent from the chart above.
2. The trend for the Light Intensity activity minutes per day and Average Calories Burned per day roughly mimics the trend in the average steps per day.
3. As for sedentary activity minutes, there is not much difference that can be observed between the days of the week.

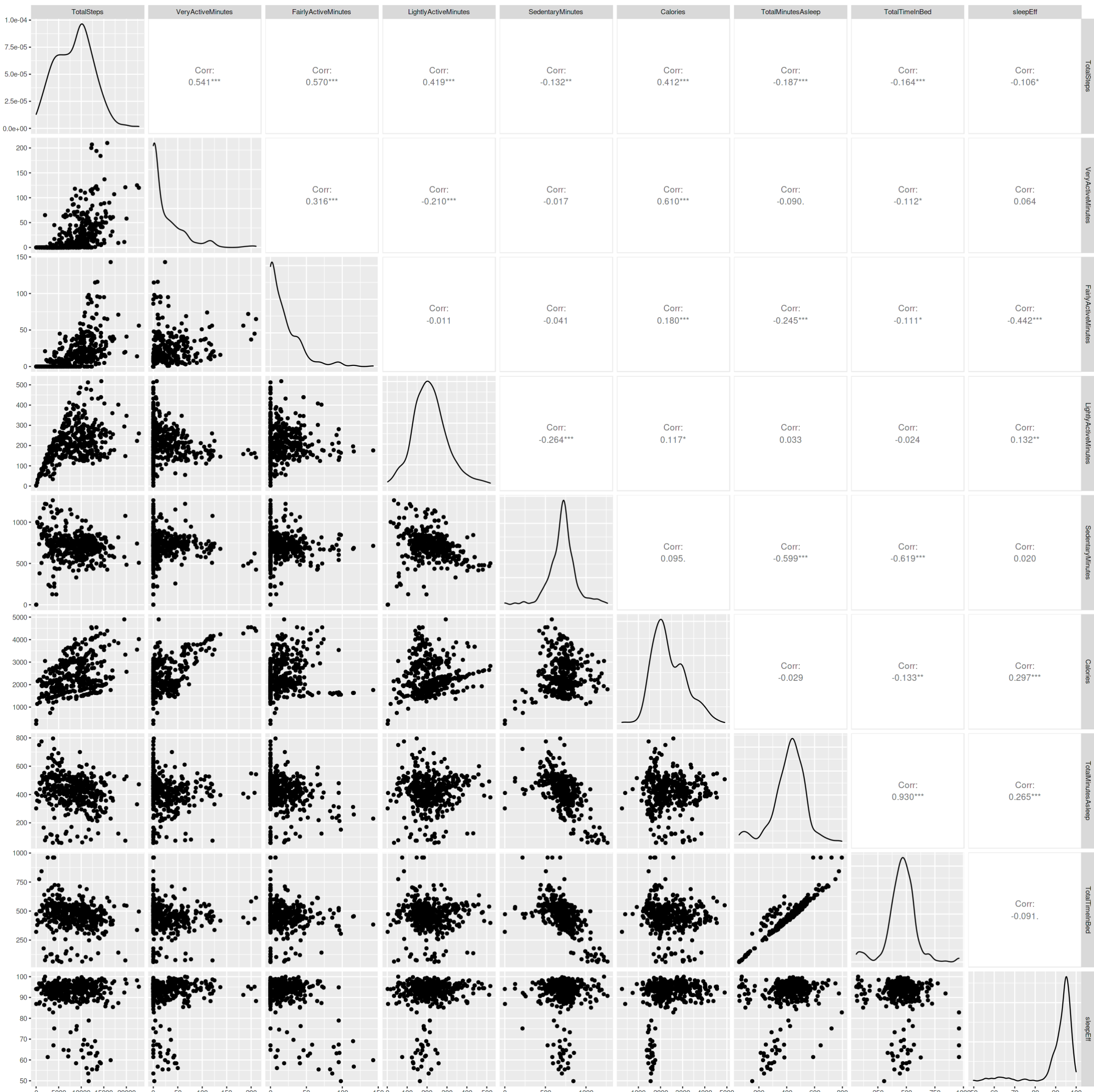
Merging Daily Activity with Daily Sleep

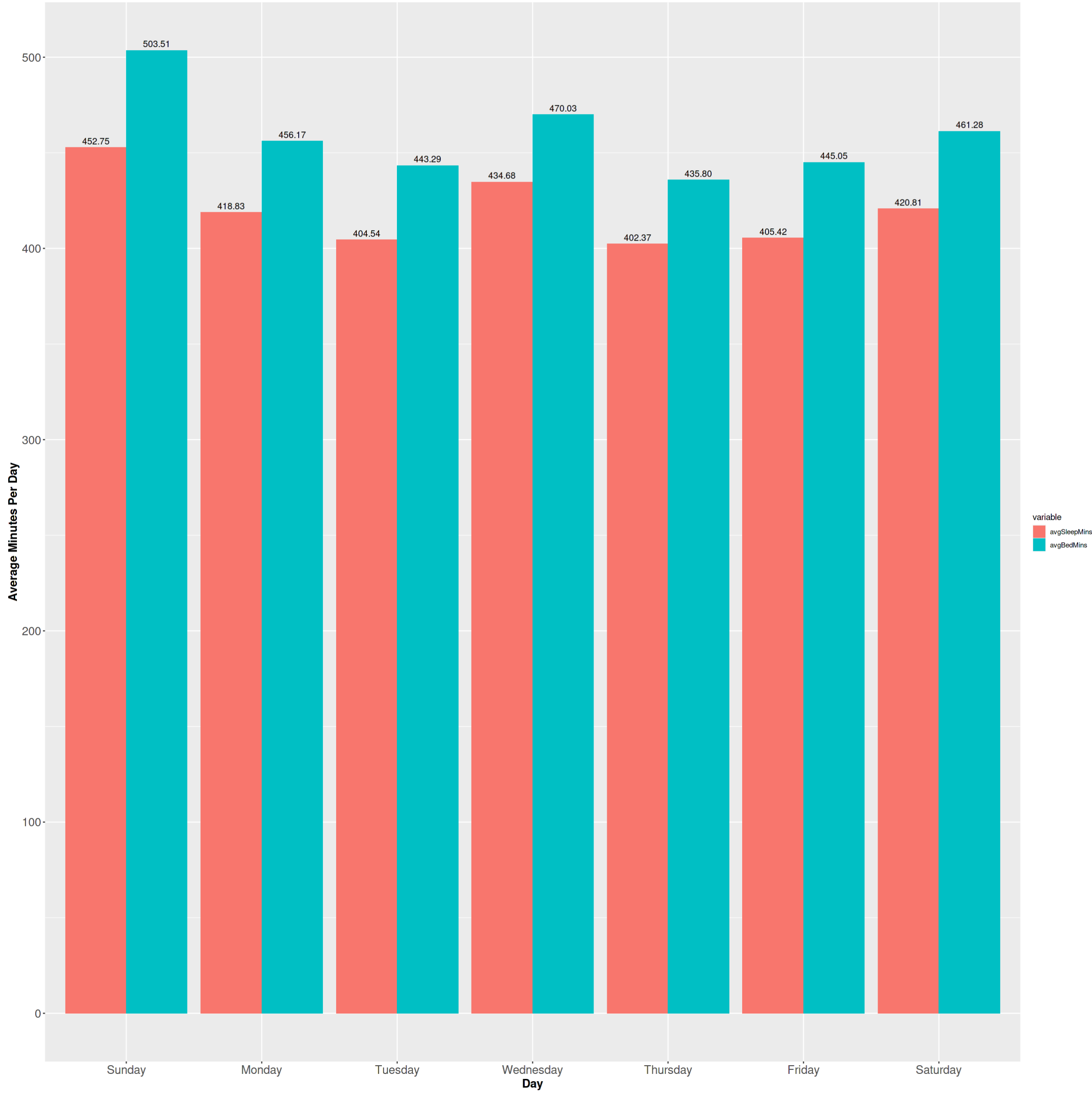
Next, I decided to see the relationship between daily activity dataset and the daily sleep dataset. I merged both datasets together and excluding days/users where there are no corresponding sleep data. Then I added a new column called 'sleepEff' (i.e. Sleep Efficiency) which demonstrate how much time is spent asleep over time spent in bed in percentage form.

In [7]:
```
merge_data <- merge(daily_activity,daily_sleep, by = c("Id","Date"), all.y = TRUE)
merge_data <- merge_data %>% add_column(sleepEff=(merge_data$TotalMinutesAsleep/merge_data$TotalTimeInBed)*100)
sleep_pair_analysis <- merge_data[,c("TotalSteps","VeryActiveMinutes","FairlyActiveMinutes","LightlyActiveMinutes","SedentaryMinutes","Calories","TotalMinutesAsleep","TotalTimeInBed","sleepEff")]
options(repr.plot.width = 20, repr.plot.height = 20)
summary(merge_data)
ggpairs(sleep_pair_analysis)

daily_sleep$Day <- weekdays(daily_sleep$Date)
sleep_analysis <- daily_sleep %>% group_by(Day) %>% summarise(avgSleepMins=mean(TotalMinutesAsleep),avgBedMins=mean(TotalTimeInBed))
sleep_analysis$Day <- factor(sleep_analysis$Day, levels = daysOrder)
sleep_analysis <- sleep_analysis[order(sleep_analysis$Day),]
sleep_visual_analysis <- melt(sleep_analysis,id.vars = 'Day')
ggplot(sleep_visual_analysis,aes(x=Day,y=value,fill=variable))+geom_bar(stat='identity',position='dodge')+geom_text(aes(label=sprintf("%0.2f",round(value,digits=2))),size=4,vjust = -0.5,position=position_dodge(.9))+theme(axis.text=element_text(size=15),
        axis.title=element_text(size=15,face="bold"))+labs(y="Average Minutes Per Day")
```

```
       Id                  Date              TotalSteps    TotalDistance
 Min.   :1.504e+09   Min.   :2016-04-12   Min.   :   17   Min.   : 0.010
 1st Qu.:3.977e+09   1st Qu.:2016-04-19   1st Qu.: 5206   1st Qu.: 3.600
 Median :4.703e+09   Median :2016-04-27   Median : 8925   Median : 6.290
 Mean   :5.001e+09   Mean   :2016-04-26   Mean   : 8541   Mean   : 6.039
 3rd Qu.:6.962e+09   3rd Qu.:2016-05-04   3rd Qu.:11393   3rd Qu.: 8.030
 Max.   :8.792e+09   Max.   :2016-05-12   Max.   :22770   Max.   :17.540
 TrackerDistance  LoggedActivitiesDistance VeryActiveDistance
 Min.   : 0.010   Min.   :0.0000           Min.   : 0.00
 1st Qu.: 3.600   1st Qu.:0.0000           1st Qu.: 0.00
 Median : 6.290   Median :0.0000           Median : 0.57
 Mean   : 6.034   Mean   :0.1131           Mean   : 1.45
 3rd Qu.: 8.020   3rd Qu.:0.0000           3rd Qu.: 2.37
 Max.   :17.540   Max.   :4.0817           Max.   :12.54
 ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
 Min.   :0.0000           Min.   :0.010       Min.   :0.0000000
 1st Qu.:0.0000           1st Qu.:2.540       1st Qu.:0.0000000
 Median :0.4200           Median :3.680       Median :0.0000000
 Mean   :0.7502           Mean   :3.807       Mean   :0.0009201
 3rd Qu.:1.0400           3rd Qu.:4.930       3rd Qu.:0.0000000
 Max.   :6.4800           Max.   :9.480       Max.   :0.1100000
 VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
 Min.   :  0.00    Min.   :  0.00      Min.   :  2.0        Min.   :   0.0
 1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:158.0        1st Qu.: 631.0
 Median :  9.00    Median : 11.00      Median :208.0        Median : 717.0
 Mean   : 25.19    Mean   : 18.04      Mean   :216.9        Mean   : 712.2
 3rd Qu.: 38.00    3rd Qu.: 27.00      3rd Qu.:263.0        3rd Qu.: 783.0
 Max.   :210.00    Max.   :143.00      Max.   :518.0        Max.   :1265.0
    Calories        Day            TotalSleepRecords TotalMinutesAsleep
 Min.   : 257   Length:413         Min.   :1.000     Min.   : 58.0
 1st Qu.:1850   Class :character   1st Qu.:1.000     1st Qu.:361.0
 Median :2220   Mode  :character   Median :1.000     Median :433.0
 Mean   :2398                      Mean   :1.119     Mean   :419.5
 3rd Qu.:2926                      3rd Qu.:1.000     3rd Qu.:490.0
 Max.   :4900                      Max.   :3.000     Max.   :796.0
 TotalTimeInBed     sleepEff
 Min.   : 61.0   Min.   : 49.84
 1st Qu.:403.0   1st Qu.: 91.22
 Median :463.0   Median : 94.31
 Mean   :458.6   Mean   : 91.68
 3rd Qu.:526.0   3rd Qu.: 96.07
 Max.   :961.0   Max.   :100.00
```





A pairwise analysis was then done again but this time with the sleep data included. Based on the above, it can be observed that:

1. TotalMinutesAsleep on average is about 419 minutes or just about 7 hours of sleep. However when looking at the day to day breakdown, I found that the average time asleep for 4 out of the 5 weekdays are less than 420 minutes or 7 hours.
2. The sleep efficiency among the users clocked in an average of 91% efficiency, which is considered very good ([https://jcsm.aasm.org/doi/10.5664/jcsm.5498 (https://jcsm.aasm.org/doi/10.5664/jcsm.5498)](https://jcsm.aasm.org/doi/10.5664/jcsm.5498)).
3. There is a positive correlation between Calories burned against Sleep Efficiency. This suggests that the more calories they burned, the better their sleep quality.
4. There is also a significant negative correlation between sedentary minutes and total minutes asleep, showing that people with higher sedentary minutes tend to get less sleep.

PHASE 6: ACT

Key Insights From This Study

Being physically active is important to the overall health of the individuals and has been shown to reduce risk of cardiovascular disease, type-2 diabetes, improve our sleep, etc. The WHO recommends healthy adults from age 18 to 64 to do 150 to 300 minutes per week of moderate intensity aerobic physical activity or at least 75 to 150 minutes per week of vigorous intensity exercise or a combination of both. However the data from this project indicates that most users on average fall below the recommended 10,000 steps per day and tend to do gravitate towards lighter activities. While this is not necessary bad as the users are not sedentary individuals, with an average of 7,500 steps per day and 190 minutes per day of light activities, more improvements could be made.

Sleep data also show that the users tend to get lesser sleep during weekday nights, lesser than the recommended 7 hours sleep for healthy adults, which could be attributed to a busier lifestyle during working weekday nights. On the other hand, the sleep quality on average is excellent and also positively correlated with calories burned, indicating that users who burned more calories for the day may fall asleep faster and have deeper sleeps. The lack of BMI data is also evident, indicating that users tend to not manually self-log their own data.

Source: https://www.who.int/news-room/fact-sheets/detail/physical-activity#:~:text=Regular%20physical%20activity%20is%20proven,of%20life%20and%20well%2Dbeing (https://www.who.int/news-room/fact-sheets/detail/physical-activity#:~:text=Regular%20physical%20activity%20is%20proven,of%20life%20and%20well%2Dbeing)

Recommendations

Companies that produce activity trackers that tracks steps, activity and sleep are in abundance but I believe what makes a wellness technology company like Bellabeat stands out is the ability to motivate and encourage its users to go the extra mile and improve their health voluntary. This is the reason why I am recommending that the Bellabeat membership program or Coach (https://bellabeat.com/coach/ (https://bellabeat.com/coach/)) should be marketed aggressively to its users. Being like a personalised coach, the program takes a holistic approach to the overall health of its users by recommending tailor-made activities and customized meal plans based on their profile and their set goals, helping users with their sleep and beauty, etc. This not only helps the users to achieve a decent amount of physical activity per week and achieve higher quality sleep everyday but it also encourages self logging of data such as weight, height and also nutritional information. Additionally, the popularity and success of Coach may draw more people into the Bellabeat ecosystem, directly increasing the sales of other Bellabeat products.