Bellabeat Case Study

Amarnath

2024-02-05

Business Objective.

Analyze smart device usage data to gain insight into how consumers use non-Bellabeat smart devices. Select one Bellabeat product to apply these insights and provide recommendations to the marketing team.

Preparing the Data.

*FitBit Fitness Tracker Data: Minute-level details from 30 users on physical activity, heart rate, and sleep. Valuable for health and lifestyle insights. CC0: Public Domain, made available through Mobius.

*Lets use R for Data Analysis Process.

```r
install.packages("tidyverse")
```

Loading Necessary Packages.

```
##
## The downloaded binary packages are in
##  /var/folders/8y/qnvkl4p95515fyd6yh7z1q800000gn/T//RtmpjEFyTL/downloaded_packages
```

```r
install.packages("ggplot2")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8y/qnvkl4p95515fyd6yh7z1q800000gn/T//RtmpjEFyTL/downloaded_packages
```

```r
install.packages("dplyr")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8y/qnvkl4p95515fyd6yh7z1q800000gn/T//RtmpjEFyTL/downloaded_packages
```

```r
install.packages("ggpubr")
```

```
##
## The downloaded binary packages are in
##  /var/folders/8y/qnvkl4p95515fyd6yh7z1q800000gn/T//RtmpjEFyTL/downloaded_packages
```

```r
library(tidyverse)
library(ggplot2)
library(dplyr)
```

```r
DailyActivity <- read.csv("/Users/amarnathreddychinthapalli/Desktop/Bellabeat - casestudy/Fitabase Data

DailySleep <- read.csv("/Users/amarnathreddychinthapalli/Desktop/Bellabeat - casestudy/Fitabase Data 4.
```

Loading Dataset.

```r
str(DailyActivity)
```

Structure of Data Frames.

```
## '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              : int  13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
##  $ 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       : int  25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes     : int  13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes    : int  328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes        : int  728 776 1218 726 773 539 1149 775 818 838 ...
##  $ Calories                : int  1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

```r
str(DailySleep)
```

```
## 'data.frame':    413 obs. of  5 variables:
##  $ Id                : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr  "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00 AM"
##  $ TotalSleepRecords : int  1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: int  327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : int  346 407 442 367 712 320 377 364 384 449 ...
```

Data Cleaning.

*We have noticed that Dates in all data frames is in Character type, lets change it to Date type in the format of "YYYY-MM-DD".

```r
DailyActivity$ActivityDate <- as.Date(DailyActivity$ActivityDate, format = "%m/%d/%Y")
DailySleep$SleepDay <- as.Date(DailySleep$SleepDay, format = "%m/%d/%Y")
```

*Lets see the preiew of the data frames.

```r
head(DailyActivity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-14      10460          6.74            6.74
## 4 1503960366   2016-04-15       9762          6.28            6.28
## 5 1503960366   2016-04-16      12669          8.16            8.16
## 6 1503960366   2016-04-17       9705          6.48            6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.44                     0.40
## 4                        0               2.14                     1.26
## 5                        0               2.71                     0.41
## 6                        0               3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                3.91                       0                30
## 4                2.83                       0                29
## 5                5.04                       0                36
## 6                2.51                       0                38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728
```

```r
head(DailySleep)
```

```
##           Id   SleepDay TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 1503960366 2016-04-12                 1                327            346
## 2 1503960366 2016-04-13                 2                384            407
## 3 1503960366 2016-04-15                 1                412            442
## 4 1503960366 2016-04-16                 2                340            367
## 5 1503960366 2016-04-17                 1                700            712
## 6 1503960366 2016-04-19                 1                304            320
```

Data Manipulation.

*Lets create a dataframe with users activity who have burned more than 2000 calories.

```r
filtered_df = DailyActivity %>%
  filter(DailyActivity$Calories > 2000)
```

*Lets Merge DailyActivity and DailySleep Data frames to find out if there is any correlation between calories burnt and sleep time.

*First, lets change activitydate and sleepday column names to same name 'Date'.

```r
colnames(DailyActivity)[colnames(DailyActivity) == "ActivityDate"] <- "Date"
colnames(DailySleep)[colnames(DailySleep) == "SleepDay"] <- "Date"
```

*Lets merge only the required columns into a new dataframe.

```r
sub_DailyActivity <- select(DailyActivity, Id, Date, Calories, TotalSteps)
sub_DailySleep <- select(DailySleep, Id, Date, TotalMinutesAsleep)
Activity_Sleep <- merge(sub_DailyActivity, sub_DailySleep, by = c("Id","Date"))
```

*Lets create a dataframe to analyse the perfomance of user daily with respect to weekdays.

```r
WeekdaysPerformance <- Activity_Sleep %>%
  mutate(Weekday = factor(weekdays(Date), levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Fri
```

```r
str(Activity_Sleep)
```

```
## 'data.frame':    413 obs. of  5 variables:
##  $ Id                : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ Date              : Date, format: "2016-04-12" "2016-04-13" ...
##  $ Calories          : int  1985 1797 1745 1863 1728 2035 1786 1775 1949 1788 ...
##  $ TotalSteps        : int  13162 10735 9762 12669 9705 15506 10544 9819 14371 10039 ...
##  $ TotalMinutesAsleep: int  327 384 412 340 700 304 360 325 361 430 ...
```

```r
summary(Activity_Sleep[,c("TotalSteps", "TotalMinutesAsleep", "Calories")])
```

```
##    TotalSteps    TotalMinutesAsleep    Calories
##  Min.   :   17   Min.   : 58.0      Min.   : 257
##  1st Qu.: 5206   1st Qu.:361.0      1st Qu.:1850
##  Median : 8925   Median :433.0      Median :2220
##  Mean   : 8541   Mean   :419.5      Mean   :2398
##  3rd Qu.:11393   3rd Qu.:490.0      3rd Qu.:2926
##  Max.   :22770   Max.   :796.0      Max.   :4900
```

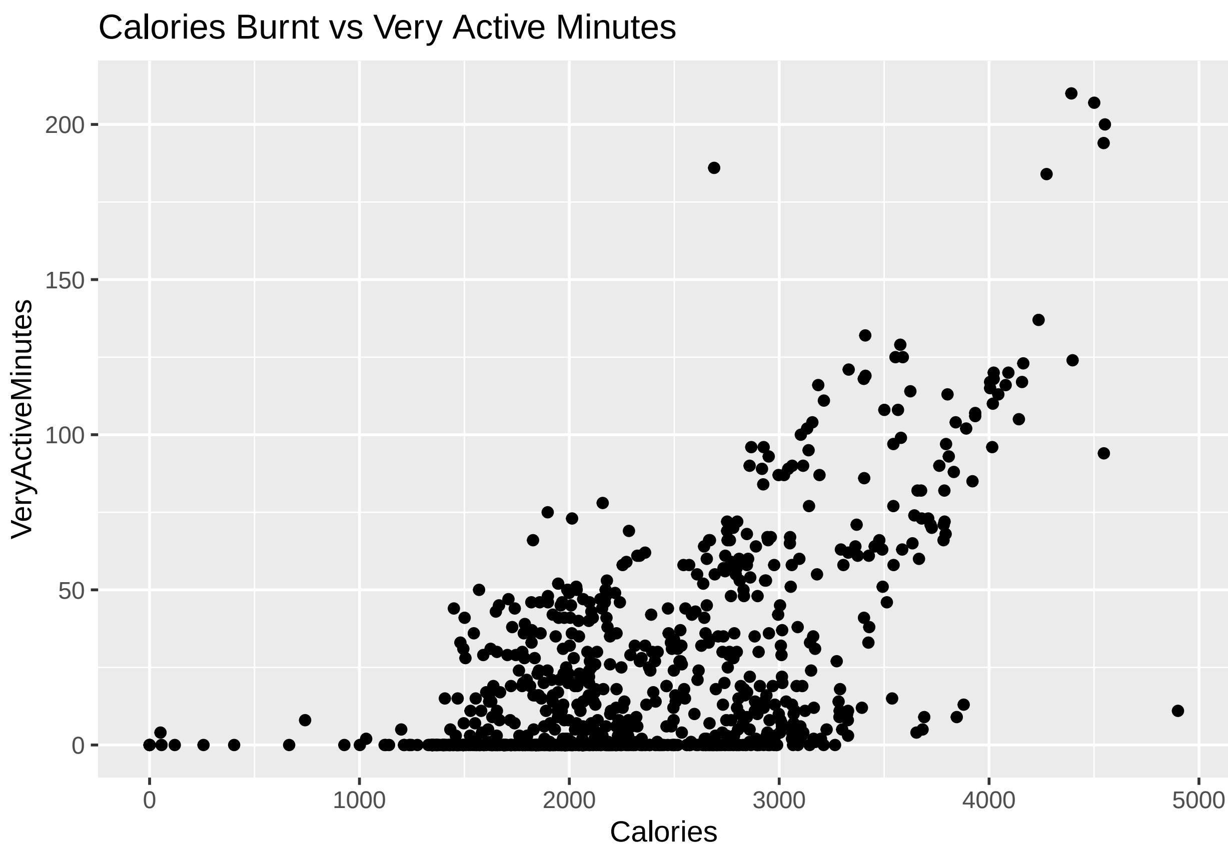
```r
str(WeekdaysPerformance)
```

```
## 'data.frame':    413 obs. of  6 variables:
##  $ Id                : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ Date              : Date, format: "2016-04-12" "2016-04-13" ...
##  $ Calories          : int  1985 1797 1745 1863 1728 2035 1786 1775 1949 1788 ...
##  $ TotalSteps        : int  13162 10735 9762 12669 9705 15506 10544 9819 14371 10039 ...
##  $ TotalMinutesAsleep: int  327 384 412 340 700 304 360 325 361 430 ...
##  $ Weekday           : Factor w/ 7 levels "Monday","Tuesday",..: 2 3 5 6 7 2 3 4 6 7 ...
```

Data Analysis and Visualization.

```
library(ggplot2)
ggplot(data = DailyActivity)+
geom_point(mapping = aes(x = Calories, y = VeryActiveMinutes))+
labs(title = "Calories Burnt vs Very Active Minutes")
```

Analysing impact of Active Minutes on Calories.

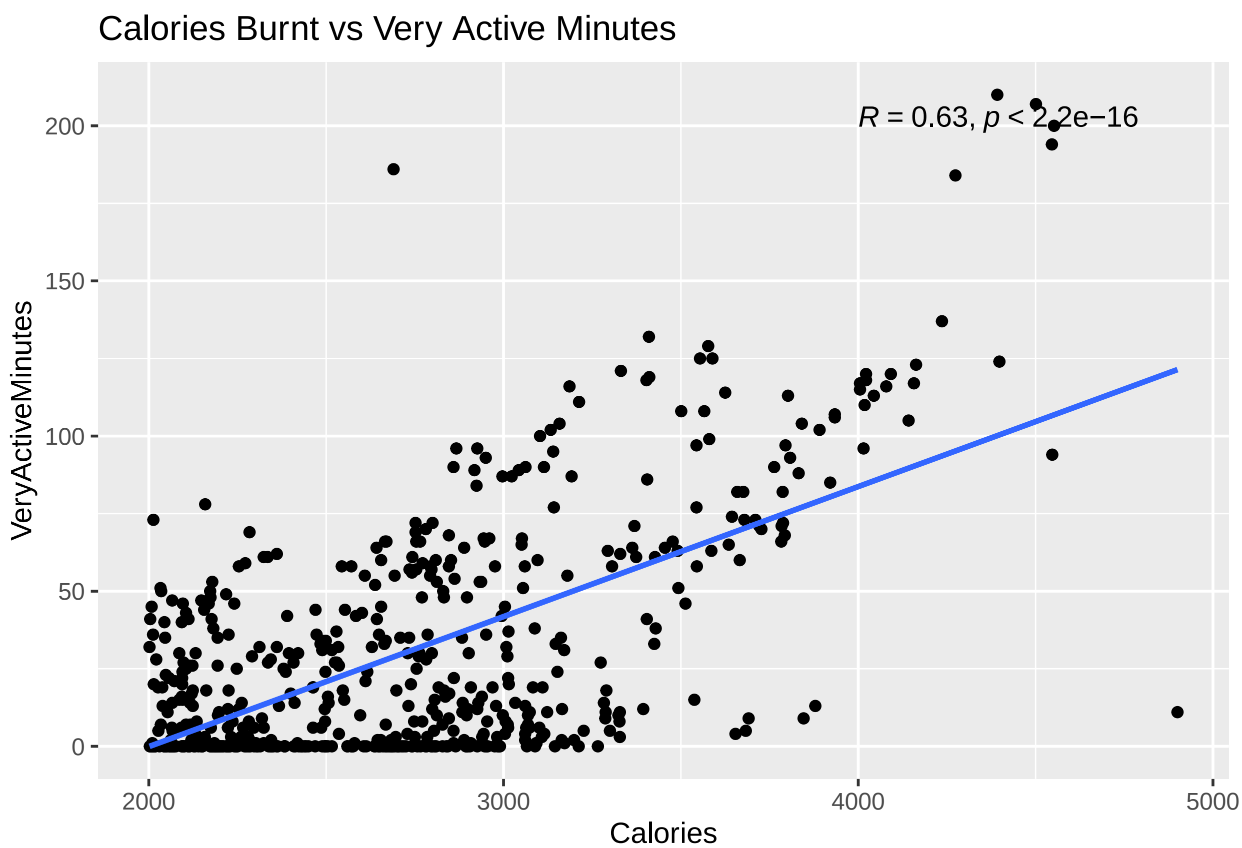


*Users who are burning more than 2000 calories have more very active minutes.

*Lets explore this further and see how Very Active Minutes have an impact on users who burn more than 3000 calories.

```
library(ggplot2)
library(ggpubr)
ggplot(data = filtered_df, aes(x = Calories, y = VeryActiveMinutes)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  stat_cor(label.x = 4000)+
  labs(title = "Calories Burnt vs Very Active Minutes")
```

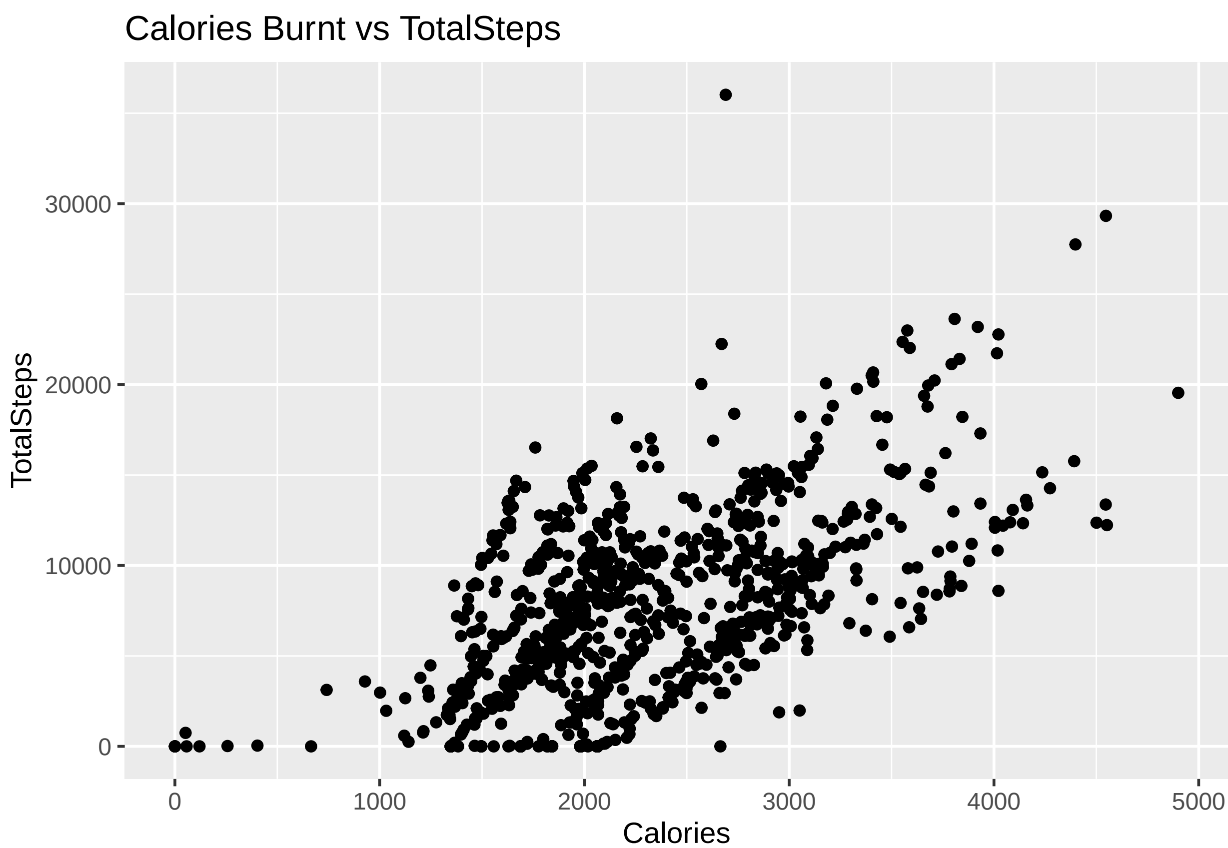
```
## `geom_smooth()` using formula = 'y ~ x'
```



*As we can see there is a strong correlation between VeryAciveMinutes and Calories(>2000).

```
ggplot(DailyActivity)+
  geom_point(mapping = aes(x=Calories, y=TotalSteps))+
  labs(title = "Calories Burnt vs TotalSteps")
```

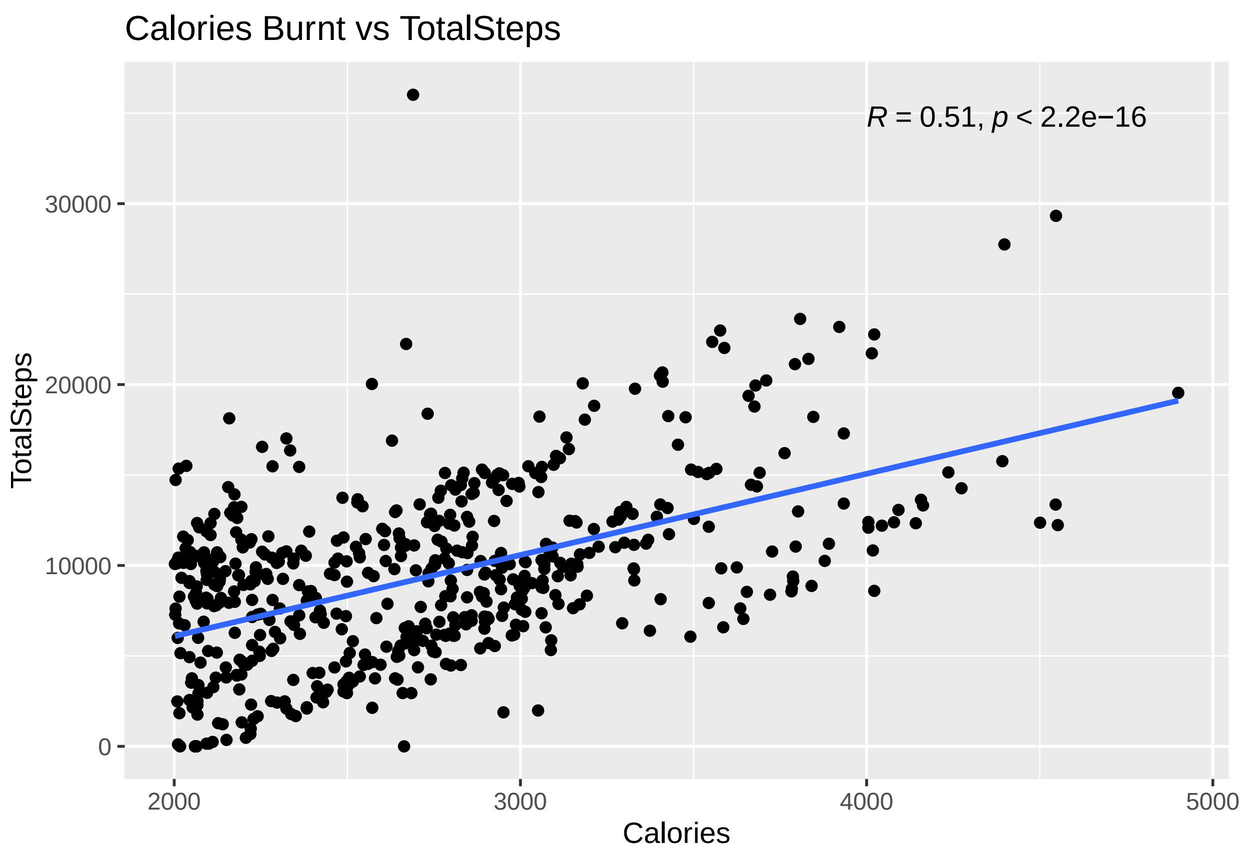
Analysing impact of steps taken in a day on Calories.



*Lets explore this further.

```
ggplot(filtered_df,  aes(x=Calories, y=TotalSteps))+
  geom_point()+
  geom_smooth(method ="lm", se=FALSE)+
  stat_cor(label.x = 4000)+
  labs(title = "Calories Burnt vs TotalSteps")
```

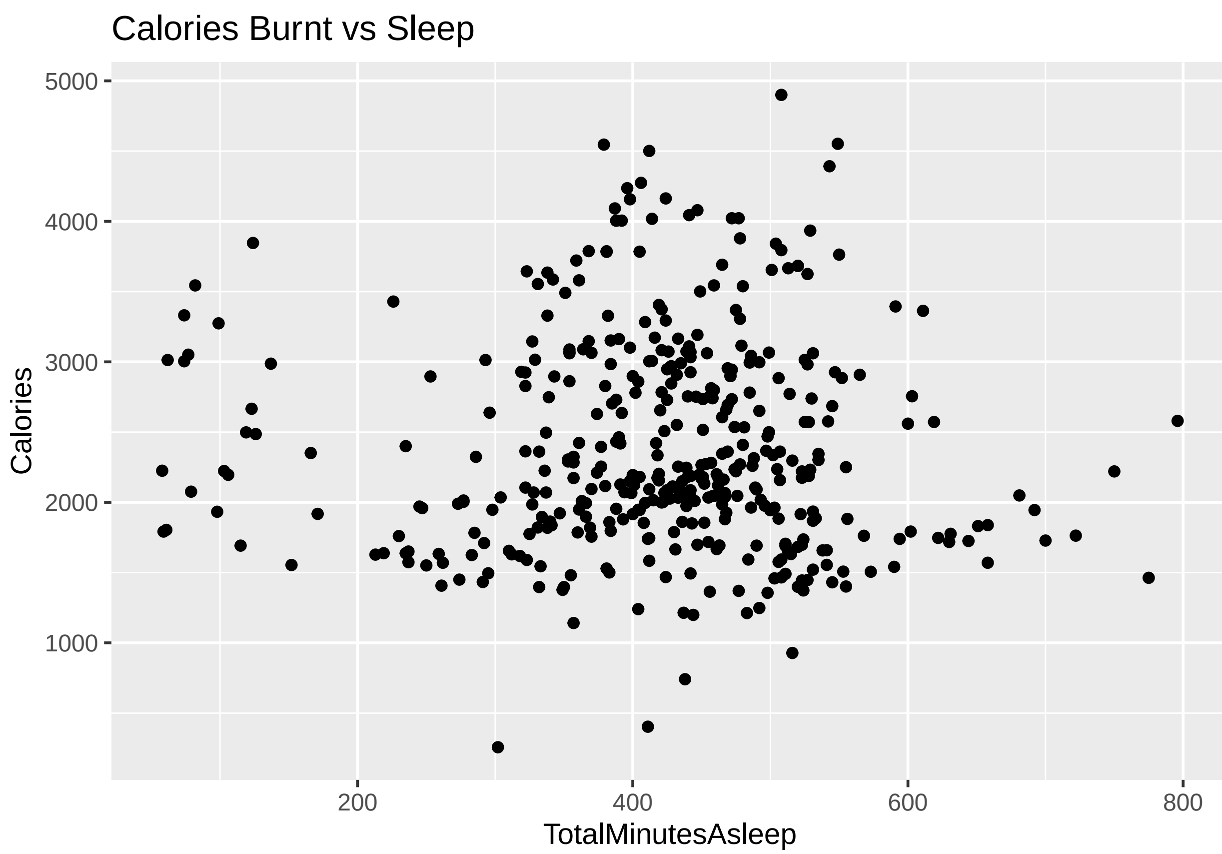
```
## `geom_smooth()` using formula = 'y ~ x'
```



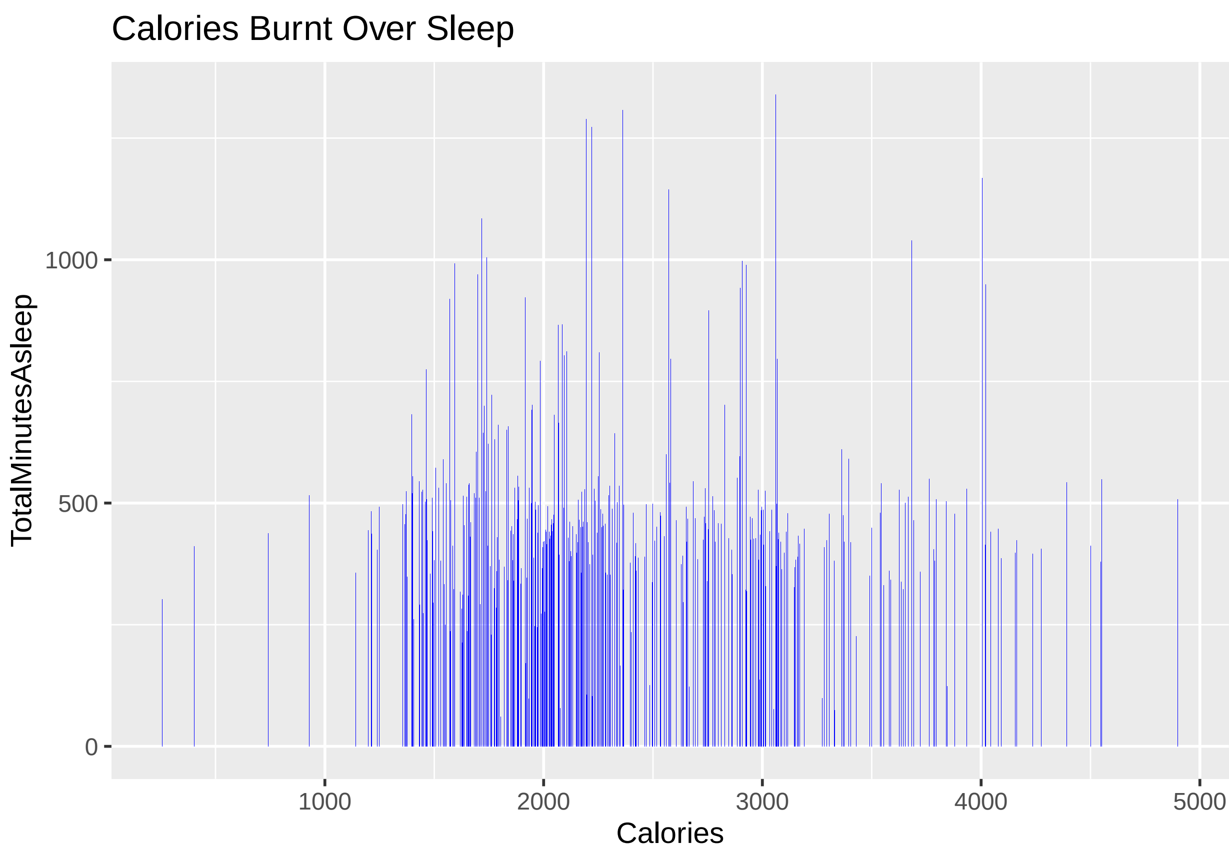
*As expected there is a moderate correlation between totalsteps taken and Calories(>2000)

```r
ggplot(Activity_Sleep, aes(x=TotalMinutesAsleep, y = Calories)) +
  geom_point()+
  labs(title = "Calories Burnt vs Sleep")
```

Analysing impact of sleep on Calories.



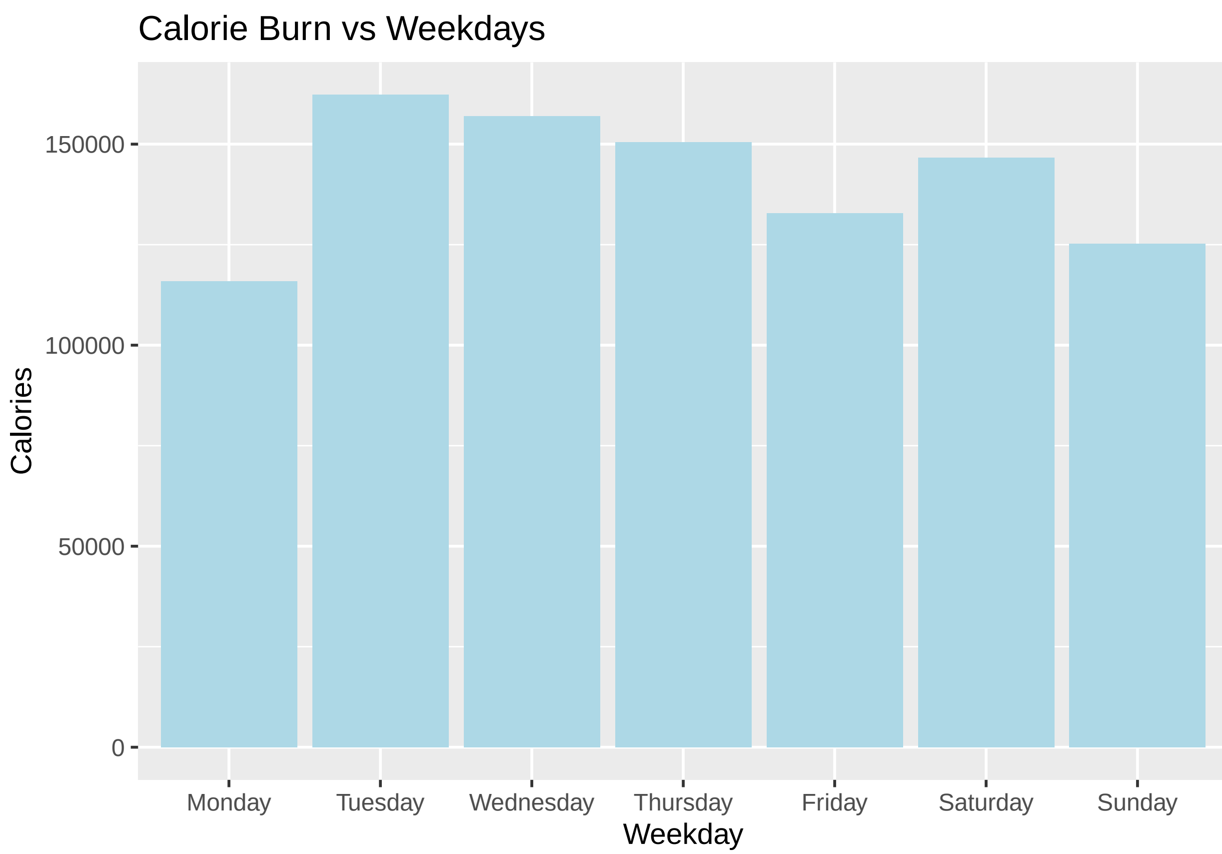

```r
ggplot(Activity_Sleep, aes(x = Calories, y = TotalMinutesAsleep)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Calories Burnt Over Sleep")
```



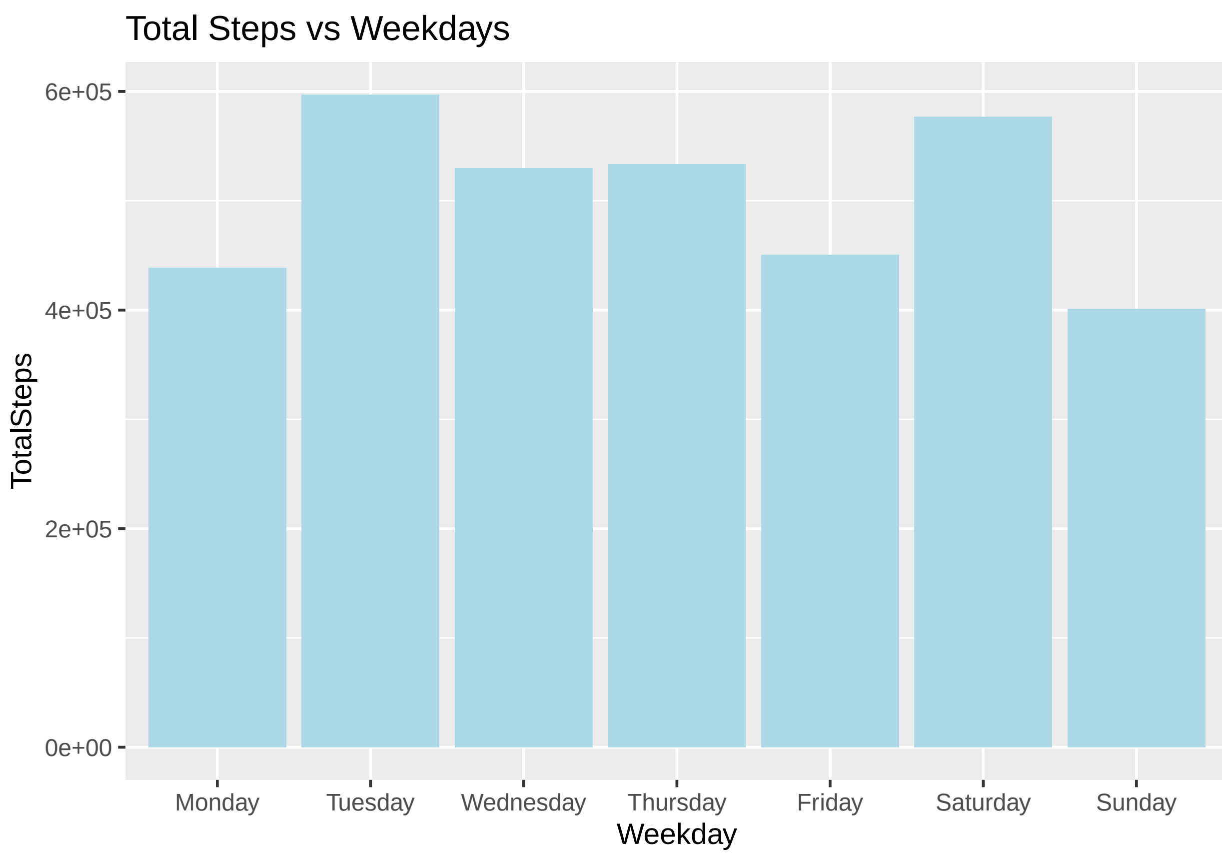
*Users typically burning more calories tend to sleep between 325 and 550 minutes.

```
ggplot(WeekdaysPerformance, aes(x=Weekday, y=Calories))+
  geom_bar(stat = "identity", fill="lightblue")+
  labs(title = "Calorie Burn vs Weekdays")
```

Analysing the user's performance on a weekday basis.




```r
ggplot(WeekdaysPerformance, aes(x=Weekday, y=TotalSteps))+
  geom_bar(stat = "identity", fill="lightblue")+
  labs(title = "Total Steps vs Weekdays")
```



*After examining both graphs, we observe that users tend to burn fewer calories and take fewer steps on Mondays, Fridays, and Sundays. This pattern could be attributed to the beginning and end of the week.

Recommendations to the Marketing Team.

*If a user aims to increase calorie burn, maintaining more active minutes exceeding 50 and total steps surpassing 8000 is highly beneficial. Therefore, we could advise the marketing team to set calorie burn goals and support users by sending targeted notifications.

*Individuals typically burning more calories tend to sleep between 325 and 550 minutes. Hence, we could suggest the marketing team send notifications regarding sleep duration to encourage calorie expenditure.

*Users exhibit lower calorie expenditure and step counts on Mondays, Fridays, and Sundays, potentially due to the start and end of the week. Consequently, we suggest the marketing team consider sending motivational messages and challenges to encourage activity during these days.