fitbit project

john myers

2/21/2022

The purpose of this project is to see if there are associations with using a fitbit device. We will look at sleep, calories, number of steps and distance just to name a few. We hope to see that the fitbit device is working correctly so we check number of steps vs calories to make sure we see an increase in calories with increase in steps.

Load libraries

**library**(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4

## v tibble 3.1.6 v dplyr 1.0.8

## v tidyr 1.2.0 v stringr 1.4.0

## v readr 2.1.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --

## x dplyr::filter() masks stats::filter()

## x dplyr::lag() masks stats::lag()

**library**(lubridate)

##

## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':

##

## date, intersect, setdiff, union

**library**(ggcorrplot)

Import the datasets

data\_day <- read.csv("Fitabase Data 4.12.16-5.12.16/dailyActivity\_merged.csv")

head(data\_day)

## Id ActivityDate TotalSteps TotalDistance TrackerDistance

## 1 1503960366 4/12/2016 13162 8.50 8.50

## 2 1503960366 4/13/2016 10735 6.97 6.97

## 3 1503960366 4/14/2016 10460 6.74 6.74

## 4 1503960366 4/15/2016 9762 6.28 6.28

## 5 1503960366 4/16/2016 12669 8.16 8.16

## 6 1503960366 4/17/2016 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

data\_sleep <- read.csv("Fitabase Data 4.12.16-5.12.16/sleepDay\_merged.csv")

head(data\_sleep)

## Id SleepDay TotalSleepRecords TotalMinutesAsleep

## 1 1503960366 4/12/2016 12:00:00 AM 1 327

## 2 1503960366 4/13/2016 12:00:00 AM 2 384

## 3 1503960366 4/15/2016 12:00:00 AM 1 412

## 4 1503960366 4/16/2016 12:00:00 AM 2 340

## 5 1503960366 4/17/2016 12:00:00 AM 1 700

## 6 1503960366 4/19/2016 12:00:00 AM 1 304

## TotalTimeInBed

## 1 346

## 2 407

## 3 442

## 4 367

## 5 712

## 6 320

data\_c <- read.csv("Fitabase Data 4.12.16-5.12.16/dailyCalories\_merged.csv")

head(data\_c)

## Id ActivityDay Calories

## 1 1503960366 4/12/2016 1985

## 2 1503960366 4/13/2016 1797

## 3 1503960366 4/14/2016 1776

## 4 1503960366 4/15/2016 1745

## 5 1503960366 4/16/2016 1863

## 6 1503960366 4/17/2016 1728

Check missing values

data\_day %>% is.na() %>% colMeans()

## Id ActivityDate TotalSteps

## 0 0 0

## TotalDistance TrackerDistance LoggedActivitiesDistance

## 0 0 0

## VeryActiveDistance ModeratelyActiveDistance LightActiveDistance

## 0 0 0

## SedentaryActiveDistance VeryActiveMinutes FairlyActiveMinutes

## 0 0 0

## LightlyActiveMinutes SedentaryMinutes Calories

## 0 0 0

data\_sleep %>% is.na() %>% colMeans()

## Id SleepDay TotalSleepRecords TotalMinutesAsleep

## 0 0 0 0

## TotalTimeInBed

## 0

data\_c %>% is.na() %>% colMeans()

## Id ActivityDay Calories

## 0 0 0

Convert data column to datetime

data\_day <- data\_day %>%

mutate(ActivityDate = as.Date(ActivityDate, format='%m/%d/%Y'))

data\_sleep <- data\_sleep %>%

mutate(SleepDay = as.Date(str\_remove(SleepDay,' 12:00:00 AM' ), format='%m/%d/%Y' ))

data\_c <- data\_c %>%

mutate(ActivityDay = as.Date(ActivityDay, format='%m/%d/%Y'))

This graph shows the individual TotalSteps out of the 33 subjects.Notice the differences bewtween the participants.

data\_day %>%

ggplot((aes(x=ActivityDate , y=TotalSteps ))) + geom\_line() + facet\_wrap(~Id, scales = 'free\_y') +

labs(title = "33 participants number of steps each")

Table

Description automatically generated

You can see that the further the step distance, the greater amount of calories are used.This shows the fitbit device is working properly in recording data.

data\_day %>%

ggplot(aes(x=TotalDistance , y=Calories)) + geom\_point() + labs(title = "steps vs calories" )

Chart, scatter chart

Description automatically generated

This graph shows there is a weak correlation with TotalMinuetsAsleep and TotalSteps where the least amount of steps is associated with greater amount of sleep.This was not expected.

data\_day %>%

left\_join(data\_sleep, by = c('ActivityDate'= 'SleepDay', 'Id' = 'Id')) %>%

ggplot(aes(x=TotalMinutesAsleep, y=TotalSteps)) + geom\_point() + geom\_smooth(method='lm') +

labs(title = "minuets asleeps vs total step count")

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 530 rows containing non-finite values (stat\_smooth).

## Warning: Removed 530 rows containing missing values (geom\_point).

Chart, scatter chart

Description automatically generated

This correlation plot looks at the patterns of the different columns and shows the more active a subject is the greater amount of distance they go.This is expected.

data\_day %>%

left\_join(data\_sleep, by = c('ActivityDate'= 'SleepDay', 'Id' = 'Id')) %>%

keep(is.numeric) %>%

select(-Id) %>%

cor(use = 'pairwise.complete.obs') %>%

ggcorrplot(type = 'lower', lab=T, digits=1)

Chart

Description automatically generated with medium confidence

This graph shows no correlation between Activity Day and number of steps.

data\_day %>%

group\_by(ActivityDate) %>%

summarise(TotalSteps = mean(TotalSteps)) %>%

slice(-nrow(.)) %>%

ggplot((aes(x=ActivityDate , y=TotalSteps ))) + geom\_col()+

labs(title = "day of activity and number of steps")

Chart, bar chart

Description automatically generated

This graph shows that LightlyActiveMinuets was by far the most common Activity state versus Fairly Active and Very Active.

data\_day %>%

group\_by(ActivityDate = wday(ActivityDate, label=T)) %>%

summarise\_all('mean') %>%

select(ActivityDate, VeryActiveMinutes, LightlyActiveMinutes, FairlyActiveMinutes) %>%

gather(key, val, -ActivityDate) %>%

ggplot((aes(x=ActivityDate , y= val ))) + geom\_col(position = 'dodge') + facet\_wrap(~key) +

labs(title = "activity state in amount of time")

Chart, bar chart, histogram

Description automatically generated

After examining the results we find that the fitbit device was recording calories burned and number of steps correctly as well as distance vs number of steps. The calories increased with the number of steps. The longer sleep time associated with less steps was not expected and could be do to an error in the data collection or a true correlation within the data.

Reccomendations for a company interested in their own fitbit device:

We know that the cusumer wears the device during the day and night, so we could look for how many nights versus days each cusumer uses the device for further research.

No day of the time period showed a higher activity level when compared with all the participants. However Sunday was a less active day for most. This could be due to Saturday, which was one of the more active days, needing Sunday to be a rest day.