Bellabeat

Victoria

2024-11-21

BellaBeat Case Study

Loading Libraries

```
install.packages("reshape2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
library(ggplot2)
library(dplyr)
```

Loading the dataset

```
daily_activity <- read_csv("dailyActivity_merged.csv")
```

```
## Rows: 940 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
```

```
##
## 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 message.
```

Viewing First few rows

```
head(daily_activity)
```

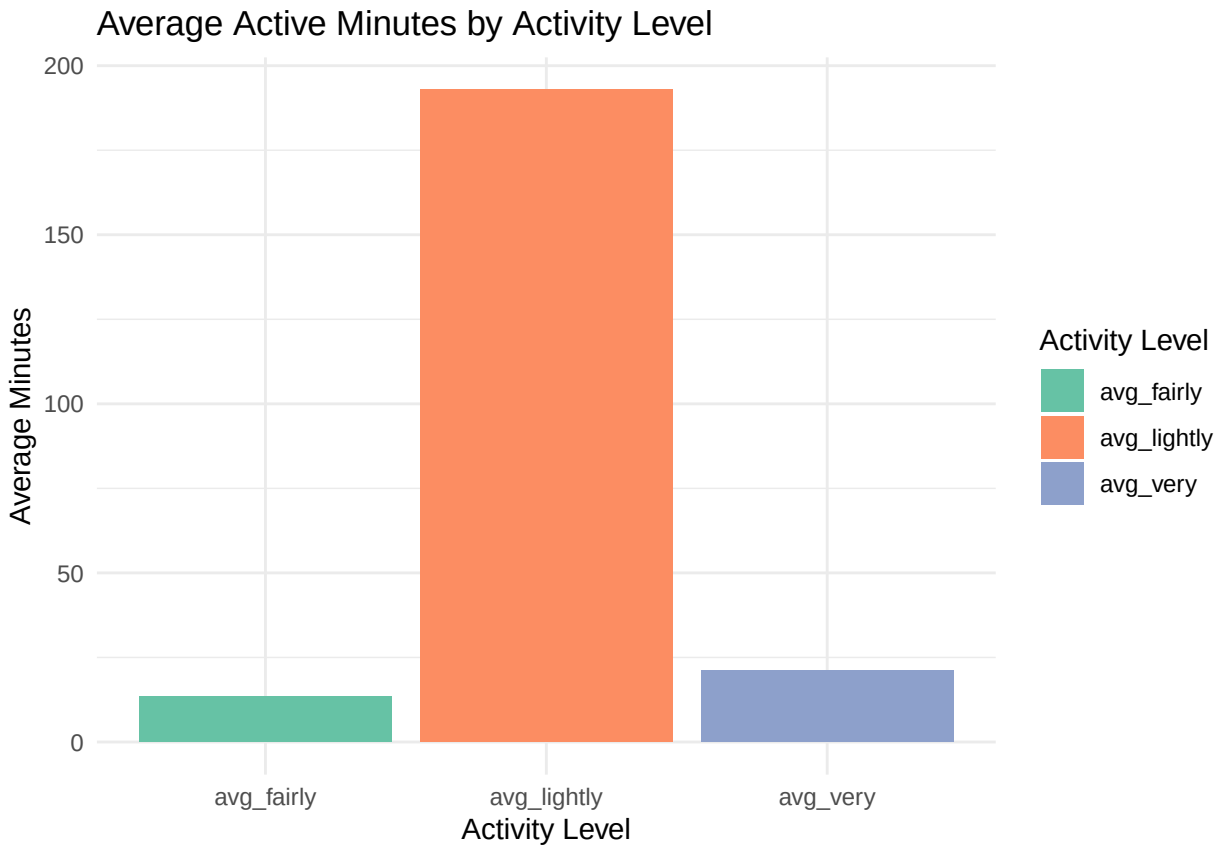
```
## # A tibble: 6 x 15
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
##        <dbl> <chr>              <dbl>         <dbl>           <dbl>
## 1 1503960366 4/12/2016         13162          8.5             8.5
## 2 1503960366 4/13/2016         10735          6.97            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
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

Plotting Daily Trends

```
daily_activity %>%
ggplot(aes(x = ActivityDate)) +
  geom_line(aes(y = TotalSteps, color = "Total Steps")) +
  geom_line(aes(y = Calories, color = "Calories")) +
  geom_line(aes(y = TotalDistance, color = "Total Distance")) +
  labs(title = "Daily Trends: Steps, Calories, and Distance",
       x = "Date", y = "Values",
       color = "Metric") +
  theme_minimal() +
  scale_color_manual(values = c("blue", "red", "green"))
```

Bar Plot

```
daily_activity %>%
  summarise(
    avg_sedentary = mean(SedentaryMinutes, na.rm = TRUE),
    avg_lightly = mean(LightlyActiveMinutes, na.rm = TRUE),
    avg_fairly = mean(FairlyActiveMinutes, na.rm = TRUE),
    avg_very = mean(VeryActiveMinutes, na.rm = TRUE)
  ) %>%
  gather(key = "Activity Level", value = "Average Minutes", -1) %>%
  ggplot(aes(x = `Activity Level`, y = `Average Minutes`, fill = `Activity Level`)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Active Minutes by Activity Level",
       x = "Activity Level", y = "Average Minutes") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set2")
```


Average Active Minutes by Activity Level
200
150
100
50
0
Average Minutes
avg_fairly
avg_lightly
avg_very
Activity Level
Activity Level
avg_fairly
avg_lightly
avg_very