Code

Mobile Device Usage & Behaviour Analysis

Abigail Asamoah

12-October-2024

Import the required libraries below

Hide

```
require(dplyr)
library(tidyr)
library(ggplot2)
library(readxl)
```

Read file and show the column names and data types, and number of observations

Hide

```
mob_data <- read.csv("../datasets/user_behavior_dataset.csv")

print("No of observations:")
```

```
[1] "No of observations:"
```

Hide

```
str(mob_data)
```

```
'data.frame':   700 obs. of  11 variables:
 $ User.ID                   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Device.Model              : chr  "Google Pixel 5" "OnePlus 9" "Xiaomi Mi 11" "Google
Pixel 5" ...
 $ Operating.System          : chr  "Android" "Android" "Android" "Android" ...
 $ App.Usage.Time..min.day.  : int  393 268 154 239 187 99 350 543 340 424 ...
 $ Screen.On.Time..hours.day.: num  6.4 4.7 4 4.8 4.3 2 7.3 11.4 7.7 6.6 ...
 $ Battery.Drain..mAh.day.   : int  1872 1331 761 1676 1367 940 1802 2956 2138 1957 ...
 $ Number.of.Apps.Installed  : int  67 42 32 56 58 35 66 82 75 75 ...
 $ Data.Usage..MB.day.       : int  1122 944 322 871 988 564 1054 1702 1053 1301 ...
 $ Age                       : int  40 47 42 20 31 31 21 31 42 42 ...
 $ Gender                    : chr  "Male" "Female" "Male" "Male" ...
 $ User.Behavior.Class       : int  4 3 2 3 3 2 4 5 4 4 ...
```

Hide

```
print(summary(mob_data))
```

```
    User.ID      Device.Model       Operating.System   App.Usage.Time..min.day. Screen.O
n.Time..hours.day. Battery.Drain..mAh.day.
 Min.   :  1.0   Length:700         Length:700         Min.   : 30.0            Min.   :
1.000             Min.   : 302.0
 1st Qu.:175.8   Class :character   Class :character   1st Qu.:113.2            1st Qu.:
2.500             1st Qu.: 722.2
 Median :350.5   Mode  :character   Mode  :character   Median :227.5            Median :
4.900             Median :1502.5
 Mean   :350.5                                         Mean   :271.1            Mean   :
5.273             Mean   :1525.2
 3rd Qu.:525.2                                         3rd Qu.:434.2            3rd Qu.:
7.400             3rd Qu.:2229.5
 Max.   :700.0                                         Max.   :598.0            Max.   :
12.000             Max.   :2993.0
 Number.of.Apps.Installed Data.Usage..MB.day.      Age           Gender          User.Be
havior.Class
 Min.   :10.00            Min.   : 102.0      Min.   :18.00   Length:700         Min.
:1.00
 1st Qu.:26.00            1st Qu.: 373.0      1st Qu.:28.00   Class :character   1st Q
u.:2.00
 Median :49.00            Median : 823.5      Median :38.00   Mode  :character   Median
:3.00
 Mean   :50.68            Mean   : 929.7      Mean   :38.48                      Mean
:2.99
 3rd Qu.:74.00            3rd Qu.:1341.0      3rd Qu.:49.00                      3rd Q
u.:4.00
 Max.   :99.00            Max.   :2497.0      Max.   :59.00                      Max.
:5.00
```

Show the number of rows and columns of data in the data set

Hide

```
no_rows <- as.character(nrow(mob_data))
no_columns <- as.character(ncol(mob_data))

paste("There are", no_rows, " rows and", no_columns,"columns in the dataset.")
```

```
[1] "There are 700  rows and 11 columns in the dataset."
```

Show the first 10 rows of the file for a quick review

Hide

```
mob_data_rows <- head(mob_data, 10)
print(mob_data_rows)
```

	User.ID <int>	Device.Model <chr>	Operating.System <chr>	App.Usage.Time..min.day. <int>
1	1	Google Pixel 5	Android	393
2	2	OnePlus 9	Android	268
3	3	Xiaomi Mi 11	Android	154
4	4	Google Pixel 5	Android	239
5	5	iPhone 12	iOS	187
6	6	Google Pixel 5	Android	99
7	7	Samsung Galaxy S21	Android	350
8	8	OnePlus 9	Android	543
9	9	Samsung Galaxy S21	Android	340
10	10	iPhone 12	iOS	424

1-10 of 10 rows | 1-5 of 11 columns

Review and find any missing values in the columns in the dataset

Hide

```
colSums(is.na(mob_data))
```

```
                  User.ID              Device.Model           Operating.System   App.U
sage.Time..min.day. Screen.On.Time..hours.day.
                       0                         0                          0
0                      0
   Battery.Drain..mAh.day.   Number.of.Apps.Installed       Data.Usage..MB.day.
Age                    Gender
                       0                         0                          0
0                      0
       User.Behavior.Class
                       0
```

Review and find any missing values in the columns in the dataset

Hide

```
duplicates <- mob_data[duplicated(mob_data),]
duplicates
```

0 rows | 1-5 of 11 columns

Que.1 Which operating System has the largest users?

Hide

```
#Grouping users by operating systems used
mob_data %>%
  group_by(Operating.System) %>%
  tally()
```

Operating.System <chr>	n <int>
Android	554
iOS	146

2 rows

Hide

NA

Que.2 What are the types of devices we have across the two operating systems?

Hide

```
library(dplyr)
mob_data %>%
  group_by(Device.Model) %>%
  tally()
```

Device.Model <chr>	n <int>
Google Pixel 5	142
OnePlus 9	133
Samsung Galaxy S21	133
Xiaomi Mi 11	146
iPhone 12	146

5 rows

Que.3 What is the average screen time hours per day?

Hide

```
print("The average screen hours in any given day:")
mean(mob_data$App.Usage.Time..min.day)
```

Que.4 Calculate the proportion of iOS users and identify number of apps by device model

Hide

```
mob_data %>%
  filter(Operating.System == "iOS") %>%
  group_by(Device.Model) %>%
  summarise(no_apps = (round(mean(Number.of.Apps.Installed), digits=0))) %>%
  arrange(desc(no_apps))
```

Que.5 Calculate the proportion of Android users and identify number of apps by device model

Hide

```
mob_data %>%
  filter(Operating.System == "Android") %>%
  group_by(Device.Model) %>%
  summarise(no_apps = (round(mean(Number.of.Apps.Installed), digits=0))) %>%
  arrange(desc(no_apps))
```

Que. 6 What is the total number of users by age group?

Hide

```
mob_data["age_group"] = cut(mob_data$Age, c(0, 16, 35, 48, 64, Inf), c("0-16", "16-35",
"35-48", "48-64", ">64"), include.lowest=TRUE)
mob_data %>%
  group_by(age_group) %>%
  tally()
```

Que. 7 What is the total of hours spent on screen by age group?

Hide

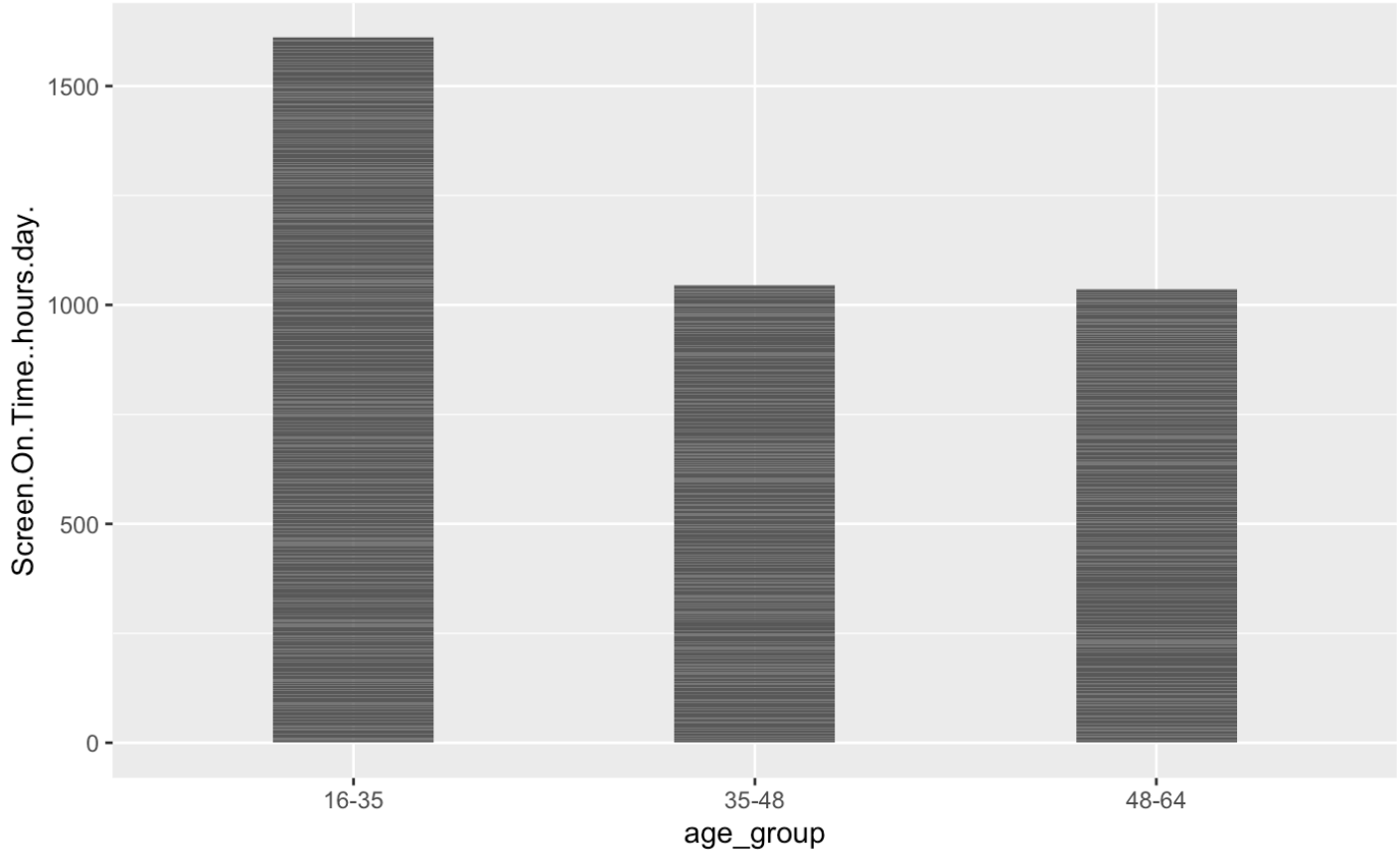
```
mob_data %>%
  group_by(age_group) %>%
  summarise(total_screen_time = (round(sum(Screen.On.Time..hours.day.), digits=0))) %>%
  arrange(desc(total_screen_time))
```

age_group <fctr>	total_screen_time <dbl>
16-35	1610
35-48	1044
48-64	1036

3 rows

Hide

```
ggplot(mob_data, aes(x=age_group, y=Screen.On.Time..hours.day.)) +
 geom_bar(stat = "identity", width=0.4)
```



Que. 8 What is the average battery drain time in a day for all device models?

Hide

```
mob_data %>%
  group_by(Device.Model) %>%
  summarise(battery_drain_time = (round(mean(Battery.Drain..mAh.day.), digits=0))) %>%
  arrange(desc(battery_drain_time))
```

Device.Model <chr>	battery_drain_time <dbl>
iPhone 12	1590
Xiaomi Mi 11	1529
OnePlus 9	1524
Samsung Galaxy S21	1505
Google Pixel 5	1476

5 rows

Hide

```
ggplot(mob_data, aes(x=Device.Model, y=Battery.Drain..mAh.day.)) +
 geom_bar(stat = "identity", width=0.4)
```