Big Data Visualization Assignment – 1

Participants –

Vijay Kumar (RA2111027010001)

Harshit Kumar (RA2111027010003)

Gaurang Ashava (RA2111027010007)

Section – AA1

Department - DSBS

Dataset – mobiledataset.csv

phone_na	brand	os	inches	resolution	battery	battery_ty	ram(GB)	announce	weight(g)	storage(GE	video_720	video_108	video_4K	video_8K	video_30f	video_60f	video_120	video_240	video_480	video_960	price(USD)
Y6II Comp	Huawei	Android 5.	5	720x1280	2200	Li-Po	2	########	140	16	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	120
K20 plus	LG	Android 7.	5.3	720x1280	2700	Li-Ion	2	########	140	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	100
P8 Lite (20	Huawei	Android 7.	5.2	1080x1920	3000	Li-Ion	4	########	147	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	420
Redmi Not	Xiaomi	Android 6.	5.5	1080x1920	4100	Li-Po	4	########	165	32	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	150
P10	Huawei	Android 7.	5.1	1080x1920	3200	Li-Ion	4	########	145	32	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	420
Xperia XA1	Sony	Android 7.	5	720x1280	2300	Li-Ion	3	########	143	32	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	140
P10 Lite	Huawei	Android 7.	5.2	1080x1920	3000	Li-Ion	4	########	146	32	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	420
P10 Plus	Huawei	Android 7.	5.5	1440x2560	3750	Li-Ion	6	########	165	64	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	170
Xperia XA1	Sony	Android 7.	6	1080x1920	2700	Li-Ion	4	########	188	32	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	250
X power2	LG	Android 7.	5.5	720x1280	4500	Li-Ion	2	########	164	16	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	170
Redmi Not	Xiaomi	Android 6.	5.5	1080x1920	4100	Li-Po	4	########	165	16	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	130
Xperia XZs	Sony	Android 7.	5.2	1080x1920	2900	Li-Ion	4	########	161	32	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	220
Xperia XZ F	Sony	Android 7.	5.46	3840x2160	3230	Li-Ion	4	########	195	64	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	320
Xperia L1	Sony	Android 7.	5.5	720x1280	2620	Li-Ion	2	########	180	16	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	110
A39	Oppo	Android 5.	5.2	720x1280	2900	Li-Ion	3	########	147	32	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	250
F3 Plus	Oppo	Android 6	6	1080x1920	4000	Li-Ion	6	########	185	64	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	250
Galaxy Xco	Samsung	Android 7.	5	720x1280	2800	Li-Ion	2	########	172	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	350
Galaxy J7 \	Samsung	Android 7.	5.5	720x1280	3300	Li-Ion	2	########	167	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	170
Galaxy S8	Samsung	Android 7.	6.2	1440x2960	3500	Li-Ion	6	########	173	64	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	350
K7 (2017)	LG	Android 6.	5	480x854	2500	Li-Ion	2	########	143	8	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	170
Y5 (2017)	Huawei	Android 6.	5	720x1280	3000	Li-Ion	2	########	150	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	280
Harmony	LG	Android 7.	5.3	720x1280	2800	Li-Ion	5	########	141	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	160
Mi 6	Xiaomi	Android 7.	5.15	1080x1920	3350	Li-Po	6	########	168	64	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	330
Y6 (2017)	Huawei	Android 6.	5	720x1280	3000	Li-Ion	2	########	150	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	240
Y7	Huawei	Android 7.	5.5	720x1280	4000	Li-Ion	2	########	165	16	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	90
nova 2 plu	Huawei	Android 7.	5.5	1080x1920	3340	Li-Po	4	########	169	64	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	160
nova 2	Huawei	Android 7.	5	1080x1920	2950	Li-Ion	4	########	143	64	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	140
Y3 (2017)	Huawei	Android 6.	5	480x854	2200	Li-Ion	1	########	175	8	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	230

Q) Define and Explain your Dataset fields.

phone_name: This field typically contains the name or model of the smartphone. It represents the specific identifier or name associated with each smartphone in the dataset.

brand: This field represents the brand or manufacturer of the smartphone. It identifies the company that produced the phone.

os: This field indicates the operating system that the smartphone uses, such as Android, iOS, or any other mobile operating system.

inches: This field represents the screen size of the smartphone, typically measured in inches diagonally. It specifies the physical size of the phone's display.

resolution: This field describes the screen resolution of the smartphone's display, typically given in pixels (width x height). For example, "1080x1920" represents a Full HD resolution of 1920 pixels in width and 1080 pixels in height.

battery: This field denotes the battery capacity of the smartphone, often measured in milliampere-hours (mAh). It indicates the amount of electric charge the battery can hold.

battery_type: This field specifies the type of battery used in the smartphone. Common battery types include Li-Po (Lithium Polymer) and Li-Ion (Lithium-Ion).

ram (GB): This field represents the amount of RAM (Random Access Memory) in the smartphone, typically measured in gigabytes (GB). RAM is used for temporary data storage and running applications.

announcement_date: This field indicates the date when the smartphone was officially announced or introduced to the market. The date is often in the format "dd-mm-yy" (day-month-year).

weight(g): This field specifies the weight of the smartphone in grams (g). It measures the mass of the device.

storage (GB): This field represents the internal storage capacity of the smartphone, often measured in gigabytes (GB). It indicates the amount of space available for storing apps, files, and data.

video_720p, video_1080p, video_4K, video_8K: These fields are typically Boolean values (TRUE/FALSE) and indicate whether the smartphone is capable of recording videos at various resolutions, including 720p, 1080p, 4K, and 8K.

video_30fps, video_60fps, video_120fps, video_240fps, video_480fps, video_960fps: Similar to the video resolution fields, these fields are typically Boolean values and indicate whether the smartphone supports video recording at different frame rates (frames per second).

price (USD): This field specifies the price of the smartphone in United States Dollars (USD) or another currency. It represents the cost of purchasing the smartphone.

Q 1) Which smartphone has the highest battery capacity in this dataset, and what is its capacity?

```
> smartphones <- read.csv("C:/Users/harsh/Documents/dataset.csv")
> max_battery <- max(smartphones$battery)
> smartphone_max_battery <- smartphones[smartphones$battery == max_battery, c("brand", "battery")]
> print(paste("The smartphone with the highest battery capacity is",
+             smartphone_max_battery$brand, "with a capacity of", smartphone_max_battery$battery))
[1] "The smartphone with the highest battery capacity is Honor with a capacity of 7250"
>
```

Q 2) What is the average screen size (in inches) of the smartphones in this dataset?

```
>
>
> avg_screen_size <- mean(smartphones$inches)
> print(paste("The average screen size of smartphones in this dataset is", round(avg_screen_size, 2), "inches"))
[1] "The average screen size of smartphones in this dataset is 6.42 inches"
>
```

Q 3) Among the Android 7.0 operating system smartphones, which one has the highest resolution?

```
> highest_resolution <- max(smartphones$resolution[smartphones$os == "Android 7.0"])
> smartphone_highest_resolution <- smartphones[smartphones$resolution == highest_resolution & smartphones$os == "Android 7.0", "brand"]
> print(paste("The smartphone with the highest resolution among Android 7.0 OS devices is", smartphone_highest_resolution))
[1] "The smartphone with the highest resolution among Android 7.0 OS devices is Lenovo"
> |
```

Q 4) What is the price range (in USD) for smartphones with 1080x1920 resolution and Android 7.0 operating system in this dataset?

```
 filtered_smartphones <- smartphones[smartphones$resolution == "1080x1920" & smartphones$os == "Android 7.0", "price(USD)"]
 price_range <- c(min(filtered_smartphones), max(filtered_smartphones))
arning messages:
: In min(filtered_smartphones) :
 no non-missing arguments to min; returning Inf
: In max(filtered_smartphones) :
 no non-missing arguments to max; returning -Inf
 print(paste("The price range for smartphones with 1080x1920 resolution and Android 7.0 OS is between",
             price_range[1], "USD and", price_range[2], "USD"))
1] "The price range for smartphones with 1080x1920 resolution and Android 7.0 OS is between Inf USD and -Inf USD"
 |
```

Q 5) What is the average weight (in grams) of smartphones announced in 2017?

```
 smartphones_announced_2017 <- smartphones[format(as.Date(smartphones$announcement_date, "%d-%m-%y"), "%Y") == "2017", ]
 average_weight_2017 <- mean(smartphones_announced_2017$weight)

 print(paste("The average weight of smartphones announced in 2017 is approximately", round(average_weight_2017, 2), "grams."))
1] "The average weight of smartphones announced in 2017 is approximately NA grams."
```

Q 6) What is the average battery capacity (in mAh) of smartphones with 1080p resolution and 4GB of RAM?

```
> filtered_smartphones <- smartphones[smartphones$resolution == "1080x1920" & smartphones$`ram(GB)` == 4, ]
>
> average_battery_capacity <- mean(filtered_smartphones$battery)
> print(paste("The average battery capacity of smartphones with 1080p resolution and 4GB of RAM is", round(average_battery_capacity, 2), "mA
h."))
[1] "The average battery capacity of smartphones with 1080p resolution and 4GB of RAM is NaN mAh."
>
```

Q 7) Among smartphones with Android 7.0, what is the median weight (in grams)?

```
android_7_smartphones <- smartphones[smartphones$os == "Android 7.0", ]


median_weight <- median(android_7_smartphones$`weight(g)`)
print(paste("Among smartphones with Android 7.0, the median weight is", median_weight, "grams."))
1] "Among smartphones with Android 7.0, the median weight is  grams."
```

Q 8) How many smartphones in the dataset support 8K video recording and have a battery capacity greater than 4000 mAh?

```
> # Filter for smartphones that support 8K video recording and have a battery capacity > 4000 mAh
> filtered_smartphones <- smartphones[smartphones$video_8K == TRUE & smartphones$battery > 4000, ]
>
> # Count the number of smartphones that meet both criteria
> count_8K_and_high_battery <- nrow(filtered_smartphones)
> print(paste("The number of smartphones in the dataset that support 8K video recording and have a battery capacity greater than 4000 mAh i
s", count_8K_and_high_battery))
[1] "The number of smartphones in the dataset that support 8K video recording and have a battery capacity greater than 4000 mAh is 0"
>
```

Q 9) What is the earliest announcement date in the dataset, and which smartphone was announced on that date?

```
# Find the earliest announcement date in the dataset
earliest_date <- min(smartphones$announcement_date)

# Find the smartphone(s) announced on the earliest date
earliest_announcements <- smartphones[smartphones$announcement_date == earliest_date, c("phone_name", "brand")]
print(paste("The earliest announcement date in the dataset is", earliest_date, "and the smartphone(s) announced on that date is/are:"))
1] "The earliest announcement date in the dataset is 2016-09-01 and the smartphone(s) announced on that date is/are:"
print(earliest_announcements)
  phone_name  brand
Y6II Compact Huawei
```

Q 10) Among smartphones with at least 32GB of storage, what is the average weight (in grams) for Sony brand smartphones?

```
# Find the earliest announcement date in the dataset
earliest_date <- min(smartphones$announcement_date)

# Find the smartphone(s) announced on the earliest date
earliest_announcements <- smartphones[smartphones$announcement_date == earliest_date, c("phone_name", "brand")]
print(paste("The earliest announcement date in the dataset is", earliest_date, "and the smartphone(s) announced on that date is/are:"))
1] "The earliest announcement date in the dataset is 2016-09-01 and the smartphone(s) announced on that date is/are:"
print(earliest_announcements)
  phone_name  brand
Y6II Compact Huawei
```

Statistical Analysis

Q) Mean

```
> mean = mean(smartphones$inches)
> print(mean)
[1] 6.42246
```

Q) Median

```
>
> median = median(smartphones$ram.GB.)
> print(median)
[1] 8
```

Graphs

average mobile phone price in US

Scatterplot of battery vs price

box plot of price

price in $

Boxplot of battery

battery