Lab-06

Nilanjana Dey

2024-01-03

# INTRODUCTION :

The following dataset comprises the various smartphones trending in the market and their respective features. The dataset has been taken from Kaggle.

The following columns are as follows:

brand\_name model  
price  
rating  
has\_5g  
has\_nfc  
has\_ir\_blaster  
processor\_brand num\_cores  
processor\_speed battery\_capacity fast\_charging\_available fast\_charging  
ram\_capacity  
internal\_memory  
screen\_size refresh\_rate  
num\_rear\_cameras num\_front\_cameras  
os : operating system primary\_camera\_rear primary\_camera\_front extended\_memory\_available extended\_upto  
resolution\_width resolution\_height

# The dataset is imported.  
  
library(readxl)

## Warning: package 'readxl' was built under R version 4.2.3

smartphones <- read\_excel("C:/Users/NILANJANA/Downloads/smartphones\_cleaned\_v6.xlsx")  
View(smartphones)

# There are 980 observations   
dim(smartphones)

## [1] 980 26

# the first 6 observations are displayed  
head(smartphones)

## # A tibble: 6 × 26  
## brand\_name model price rating has\_5g has\_nfc has\_ir\_blaster processor\_brand  
## <chr> <chr> <dbl> <dbl> <lgl> <lgl> <lgl> <chr>   
## 1 oneplus OnePlus… 54999 89 TRUE TRUE FALSE snapdragon   
## 2 oneplus OnePlus… 19989 81 TRUE FALSE FALSE snapdragon   
## 3 samsung Samsung… 16499 75 TRUE FALSE FALSE exynos   
## 4 motorola Motorol… 14999 81 TRUE FALSE FALSE snapdragon   
## 5 realme Realme … 24999 82 TRUE FALSE FALSE dimensity   
## 6 samsung Samsung… 16999 80 TRUE TRUE FALSE snapdragon   
## # ℹ 18 more variables: num\_cores <dbl>, processor\_speed <dbl>,  
## # battery\_capacity <dbl>, fast\_charging\_available <dbl>, fast\_charging <dbl>,  
## # ram\_capacity <dbl>, internal\_memory <dbl>, screen\_size <dbl>,  
## # refresh\_rate <dbl>, num\_rear\_cameras <dbl>, num\_front\_cameras <dbl>,  
## # os <chr>, primary\_camera\_rear <dbl>, primary\_camera\_front <dbl>,  
## # extended\_memory\_available <dbl>, extended\_upto <dbl>,  
## # resolution\_width <dbl>, resolution\_height <dbl>

# the last 6 observations are printed.  
tail(smartphones)

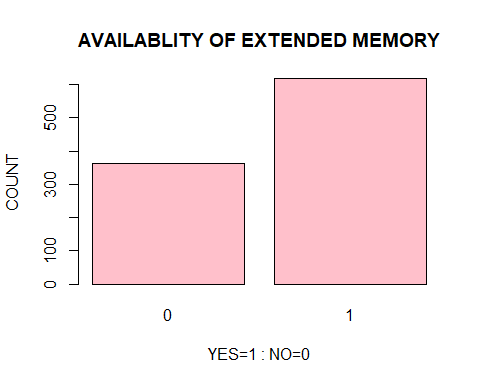
## # A tibble: 6 × 26  
## brand\_name model price rating has\_5g has\_nfc has\_ir\_blaster processor\_brand  
## <chr> <chr> <dbl> <dbl> <lgl> <lgl> <lgl> <chr>   
## 1 vivo Vivo X… 119990 NA TRUE TRUE FALSE snapdragon   
## 2 motorola Motoro… 34990 83 TRUE FALSE FALSE snapdragon   
## 3 honor Honor … 14990 75 TRUE FALSE FALSE snapdragon   
## 4 poco POCO X… 28990 85 TRUE TRUE TRUE dimensity   
## 5 motorola Motoro… 19990 80 TRUE TRUE FALSE snapdragon   
## 6 samsung Samsun… 24990 74 TRUE FALSE FALSE <NA>   
## # ℹ 18 more variables: num\_cores <dbl>, processor\_speed <dbl>,  
## # battery\_capacity <dbl>, fast\_charging\_available <dbl>, fast\_charging <dbl>,  
## # ram\_capacity <dbl>, internal\_memory <dbl>, screen\_size <dbl>,  
## # refresh\_rate <dbl>, num\_rear\_cameras <dbl>, num\_front\_cameras <dbl>,  
## # os <chr>, primary\_camera\_rear <dbl>, primary\_camera\_front <dbl>,  
## # extended\_memory\_available <dbl>, extended\_upto <dbl>,  
## # resolution\_width <dbl>, resolution\_height <dbl>

# HERE ARE DEALING WITH THE DESCRIPTIVE ANALYSIS OF THE ENTIRE DATASET.  
  
summary(smartphones)

## brand\_name model price rating   
## Length:980 Length:980 Min. : 3499 Min. :60.00   
## Class :character Class :character 1st Qu.: 12999 1st Qu.:74.00   
## Mode :character Mode :character Median : 19995 Median :80.00   
## Mean : 32521 Mean :78.26   
## 3rd Qu.: 35492 3rd Qu.:84.00   
## Max. :650000 Max. :89.00   
## NA's :101   
## has\_5g has\_nfc has\_ir\_blaster processor\_brand   
## Mode :logical Mode :logical Mode :logical Length:980   
## FALSE:431 FALSE:587 FALSE:821 Class :character   
## TRUE :549 TRUE :393 TRUE :159 Mode :character   
##   
##   
##   
##   
## num\_cores processor\_speed battery\_capacity fast\_charging\_available  
## Min. :4.000 Min. :1.200 Min. : 1821 Min. :0.0000   
## 1st Qu.:8.000 1st Qu.:2.050 1st Qu.: 4500 1st Qu.:1.0000   
## Median :8.000 Median :2.300 Median : 5000 Median :1.0000   
## Mean :7.772 Mean :2.427 Mean : 4818 Mean :0.8541   
## 3rd Qu.:8.000 3rd Qu.:2.840 3rd Qu.: 5000 3rd Qu.:1.0000   
## Max. :8.000 Max. :3.220 Max. :22000 Max. :1.0000   
## NA's :6 NA's :42 NA's :11   
## fast\_charging ram\_capacity internal\_memory screen\_size   
## Min. : 10.00 Min. : 1.00 Min. : 8 Min. :3.540   
## 1st Qu.: 18.00 1st Qu.: 4.00 1st Qu.: 64 1st Qu.:6.500   
## Median : 33.00 Median : 6.00 Median : 128 Median :6.580   
## Mean : 46.13 Mean : 6.56 Mean : 141 Mean :6.537   
## 3rd Qu.: 66.00 3rd Qu.: 8.00 3rd Qu.: 128 3rd Qu.:6.670   
## Max. :240.00 Max. :18.00 Max. :1024 Max. :8.030   
## NA's :211   
## refresh\_rate num\_rear\_cameras num\_front\_cameras os   
## Min. : 60.00 Min. :1.000 Min. :1.00 Length:980   
## 1st Qu.: 60.00 1st Qu.:2.000 1st Qu.:1.00 Class :character   
## Median : 90.00 Median :3.000 Median :1.00 Mode :character   
## Mean : 92.26 Mean :2.814 Mean :1.03   
## 3rd Qu.:120.00 3rd Qu.:3.000 3rd Qu.:1.00   
## Max. :240.00 Max. :4.000 Max. :2.00   
## NA's :4   
## primary\_camera\_rear primary\_camera\_front extended\_memory\_available  
## Min. : 2.00 Min. : 0.30 Min. :0.0000   
## 1st Qu.: 24.00 1st Qu.: 8.00 1st Qu.:0.0000   
## Median : 50.00 Median :16.00 Median :1.0000   
## Mean : 50.32 Mean :16.59 Mean :0.6306   
## 3rd Qu.: 64.00 3rd Qu.:16.00 3rd Qu.:1.0000   
## Max. :200.00 Max. :60.00 Max. :1.0000   
## NA's :5   
## extended\_upto resolution\_width resolution\_height  
## Min. : 32.0 Min. : 480 Min. : 480   
## 1st Qu.: 512.0 1st Qu.:1080 1st Qu.:1612   
## Median :1024.0 Median :1080 Median :2400   
## Mean : 736.1 Mean :1076 Mean :2215   
## 3rd Qu.:1024.0 3rd Qu.:1080 3rd Qu.:2408   
## Max. :2048.0 Max. :2460 Max. :3840   
## NA's :480

## VISUALIZATION OF SOME OF THE CATEGORICAL COLUMNS OF THE DATASET

# HERE WE ARE REPRESENTING "extended\_memory\_available" USING FREQUENCY PLOT  
  
x = table(smartphones$extended\_memory\_available)  
barplot(x,xlab = "YES=1 : NO=0",ylab="COUNT",main="AVAILABLITY OF EXTENDED MEMORY",col = "pink")



# THE PLOT BELOW SHOWS THAT THE NO. OF SMARTPHONES MADE HAVE THE FEATURE OF "EXTENDED MEMORY"

# Here we shall visualize the smartphones having different ratings using pie chart  
  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.3

library(dplyr)

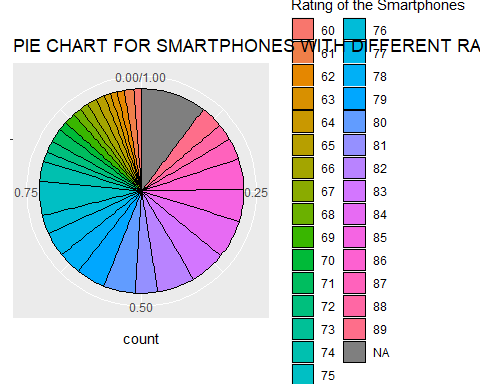
## Warning: package 'dplyr' was built under R version 4.2.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

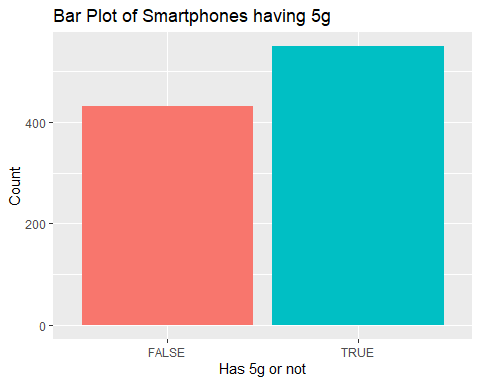
k=as.character(smartphones$rating)  
  
ggplot(smartphones, aes(x = "", fill = k)) +  
 geom\_bar(position = "fill", width = 1,color="black") +coord\_polar(theta = "y")+guides(fill = guide\_legend(title = "Rating of the Smartphones"))+theme(axis.title.y = element\_blank())+labs(title="PIE CHART FOR SMARTPHONES WITH DIFFERENT RATINGS")



# From the plot given below we can conclude that the count of different ratings is more or less same

# Here we are plotting the no. of smartphones having 5g and not having 5g.  
  
library(ggplot2)  
  
  
ggplot(smartphones, aes(x = smartphones$has\_5g, fill = smartphones$has\_5g)) +  
 geom\_bar(show.legend=FALSE) +  
 labs(title = "Bar Plot of Smartphones having 5g",x="Has 5g or not", y = "Count")

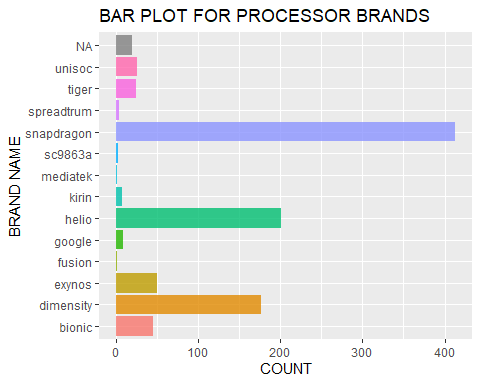
## Warning: Use of `smartphones$has\_5g` is discouraged.  
## ℹ Use `has\_5g` instead.  
## Use of `smartphones$has\_5g` is discouraged.  
## ℹ Use `has\_5g` instead.



# From the plot given below the no. of smartphones having 5g is more

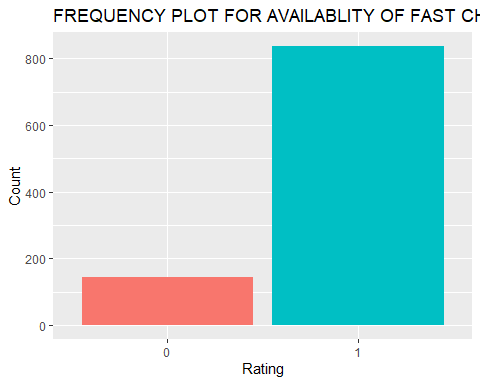
# Here we are plotting the frequency plot of different processor brands  
  
library(ggplot2)  
  
ggplot(smartphones, aes(x = smartphones$processor\_brand, fill = smartphones$processor\_brand)) +  
 geom\_bar(alpha=0.8,show.legend =FALSE ) + coord\_flip()+  
 labs(title = "BAR PLOT FOR PROCESSOR BRANDS", y = "COUNT", x = "BRAND NAME")

## Warning: Use of `smartphones$processor\_brand` is discouraged.  
## ℹ Use `processor\_brand` instead.  
## Use of `smartphones$processor\_brand` is discouraged.  
## ℹ Use `processor\_brand` instead.



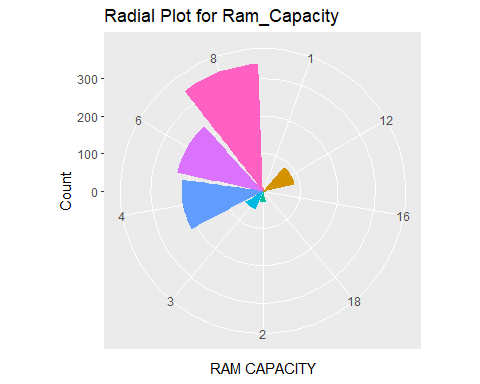
# The plot given below shows that 'snapdragon' is mostly used followed by 'helio' and 'dimesity'

# Here we are plotting the no. of smartphones having the fast charging facility or not  
library(ggplot2)  
  
m=as.character(smartphones$fast\_charging\_available)  
ggplot(smartphones, aes(x =m , fill = m)) +  
 geom\_bar(show.legend=FALSE) +  
 labs(title = "FREQUENCY PLOT FOR AVAILABLITY OF FAST CHARGING FACILITY", x = "Rating", y = "Count")



# The following plotted graph shows that fast charging facility is available in most of the mobiles

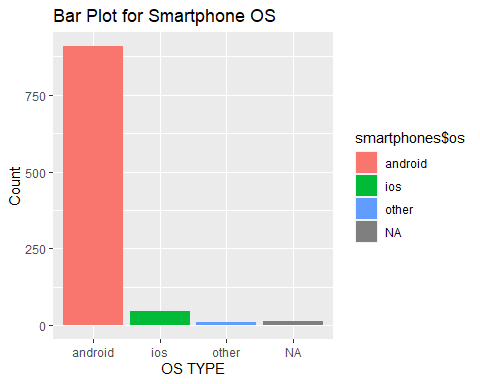
# Here we are representing the different kinds of ram\_capacity available in the smartphones using radial plot  
  
library(ggplot2)  
  
d=as.character(smartphones$ram\_capacity)  
  
ggplot(smartphones, aes(x = d, fill = d)) +  
 geom\_bar(show.legend=FALSE) + coord\_polar()+  
 labs(title = "Radial Plot for Ram\_Capacity", x = "RAM CAPACITY", y = "Count",color="RAM CAPACITY")



# The following graph indicates that the most of the smartphones have ram\_capacity = '8' followed by '6' and '4'

# Here we are plotting the frequency plot for different types of os used in smartphones  
  
library(ggplot2)  
  
  
ggplot(smartphones, aes(x = smartphones$os, fill = smartphones$os)) +  
 geom\_bar() +  
 labs(title = "Bar Plot for Smartphone OS ", x = "OS TYPE", y = "Count")

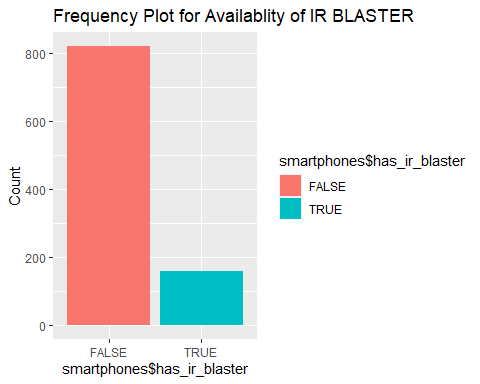
## Warning: Use of `smartphones$os` is discouraged.  
## ℹ Use `os` instead.  
## Use of `smartphones$os` is discouraged.  
## ℹ Use `os` instead.



# From the plot given below we conclude that 'android' is the most widely used os.

# Frequency plot for Availability of IR blaster  
library(ggplot2)  
  
  
ggplot(smartphones, aes(x = smartphones$has\_ir\_blaster, fill = smartphones$has\_ir\_blaster)) +  
 geom\_bar() +  
 labs(title = "Frequency Plot for Availablity of IR BLASTER",y = "Count")

## Warning: Use of `smartphones$has\_ir\_blaster` is discouraged.  
## ℹ Use `has\_ir\_blaster` instead.  
## Use of `smartphones$has\_ir\_blaster` is discouraged.  
## ℹ Use `has\_ir\_blaster` instead.



# The following plot indicates that the most of the smartphones have ir-blaster

# TEST FOR SINGLE POPULATION PROPORTION

# AIM : We shall consider one categorical column from the dataset and perform the single proportion test.

target variable = ‘has\_5g’

Objective : We shall test whether the proportion of smartphones having 5g facility is equal to the assumed value of 0.56.

target\_var=smartphones$has\_5g

# a sample of 300 observations is taken with replacement  
sample1=sample(target\_var,300,replace=TRUE)

# The sample taken , is converted into a dataframe  
s=as.data.frame(sample1)

# Now we get the contingency table showing the total no. of 'TRUE' and 'FALSE' values.  
library(dplyr)  
count(s,sample1)

## sample1 n  
## 1 FALSE 124  
## 2 TRUE 176

##HYPOTHESIS TESTING

BOTH-TAILED TEST

H0: P0=O.56 VS H1: P1!=0.56

n=300  
x=176  
prop.test(x, n, p = 0.56, alternative = "two.sided",  
 correct = TRUE)

##   
## 1-sample proportions test with continuity correction  
##   
## data: x out of n, null probability 0.56  
## X-squared = 0.76096, df = 1, p-value = 0.383  
## alternative hypothesis: true p is not equal to 0.56  
## 95 percent confidence interval:  
## 0.5285128 0.6425640  
## sample estimates:  
## p   
## 0. 5866667

# Interpretation : The p-value is more than 0.05 which indicates the acceptance of H0, i.e, our assumption is correct that the true proportion of smartphones having 5g is equal to the assumed value of 0.56 at 5% level of significance.

QUESTION 8:

Given : Sample size=600 , no. of students favoring online teaching=180

AIM : We want to test whether the proportion of students favoring online education is equal to the assumed value of ‘30% students favoring online education’

##HYPOTHESIS TESTING

BOTH-TAILED TEST

H0: P0=O.3 VS H1: P1!=0.3

n=600  
x=180  
prop.test(x, n, p = 0.3, alternative = "two.sided",  
 correct = TRUE)

##   
## 1-sample proportions test without continuity correction  
##   
## data: x out of n, null probability 0.3  
## X-squared = 0, df = 1, p-value = 1  
## alternative hypothesis: true p is not equal to 0.3  
## 95 percent confidence interval:  
## 0.2646995 0.3378452  
## sample estimates:  
## p   
## 0.3

# Interpretation : The p-value is 1 which is more than 0.05 which indicates the acceptance of H0, i.e, our assumption is correct.