Research Goal:

To analysis phone Sales data set

Data Retrieval:

library(dplyr)

library(ggplot2)

phsale<-read.csv(file.choose(),header=T)

View(phsale)

Output:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Brand | Model | Color | Memory | Storage | Rating | Selling Price | Original Price |
| OPPO | A53 | Moonlight Black | 4 GB | 64 GB | 4.5 | 11990 | 15990 |
| OPPO | A53 | Mint Cream | 4 GB | 64 GB | 4.5 | 11990 | 15990 |
| OPPO | A53 | Moonlight Black | 6 GB | 128 GB | 4.3 | 13990 | 17990 |
| OPPO | A53 | Mint Cream | 6 GB | 128 GB | 4.3 | 13990 | 17990 |
| OPPO | A53 | Electric Black | 4 GB | 64 GB | 4.5 | 11990 | 15990 |
| OPPO | A53 | Electric Black | 6 GB | 128 GB | 4.3 | 13990 | 17990 |
| OPPO | A12 | Deep Blue | 4 GB | 64 GB | 4.4 | 10490 | 11990 |
| OPPO | A12 | Black | 3 GB | 32 GB | 4.4 | 9490 | 10990 |
| OPPO | A12 | Blue | 3 GB | 32 GB | 4.4 | 9490 | 10990 |
| OPPO | A12 | Flowing Silver | 3 GB | 32 GB | 4.4 | 9490 | 10990 |
| OPPO | A12 | Deep Blue | 3 GB | 32 GB | 4.4 | 9490 | 10990 |
| OPPO | A12 | Flowing Silver | 4 GB | 64 GB | 4.4 | 10490 | 11990 |
| OPPO | A53s 5G | Crystal Blue | 6 GB | 128 GB | 4.3 | 15990 | 16990 |
| OPPO | A53s 5G | Ink Black | 6 GB | 128 GB | 4.3 | 15990 | 16990 |
| OPPO | A12 | Blue | 4 GB | 64 GB | 4.4 | 10490 | 11990 |
| OPPO | A53s 5G | Crystal Blue | 8 GB | 128 GB | 4.3 | 17990 | 18990 |
| OPPO | A53s 5G | Ink Black | 8 GB | 128 GB | 4.3 | 17990 | 18990 |
| OPPO | A33 | Moonlight Black | 3 GB | 32 GB | 4.3 | 10490 | 12990 |
| OPPO | A31 | Lake Green | 4 GB | 64 GB | 4.3 | 11960 | 12990 |
| OPPO | A31 | Mystery Black | 4 GB | 64 GB | 4.3 | 11779 | 11919 |
| OPPO | A74 5G | Fantastic Purple | 6 GB | 128 GB | 4.3 | 18979 |  |
| OPPO | A11K | Flowing Silver | 2 GB | 32 GB | 4.2 | 8990 | 10990 |
| OPPO | F17 Pro | Matte Black | 8 GB | 128 GB | 4.3 | 19990 | 25990 |
| OPPO | A54 | Moonlight Gold | 4 GB | 128 GB | 4.3 | 15990 |  |
| OPPO | A54 | Moonlight Gold | 4 GB | 64 GB | 4.3 | 14990 |  |
| OPPO | Reno6 5G | Aurora | 8 GB | 128 GB | 4.3 | 29990 | 35990 |
| OPPO | Reno6 5G | Stellar Black | 8 GB | 128 GB | 4.3 | 29990 | 35990 |
| OPPO | F17 Pro | Magic Blue | 8 GB | 128 GB | 4.3 | 19990 | 25990 |
| OPPO | A54 | Starry Blue | 4 GB | 64 GB | 4.3 | 14990 |  |
| OPPO | A54 | Starry Blue | 4 GB | 128 GB | 4.3 | 15990 |  |
| OPPO | A54 | Moonlight Gold | 6 GB | 128 GB | 4.3 | 16990 | 17990 |

Data Preparation:

##str

str(phsale)

Output:

data.frame': 2647 obs. of 8 variables:

$ Brand : chr "OPPO" "OPPO" ...

$ Model : chr "A53" "A53" ...

$ Color : chr "Moonlight Black”

$ Memory : chr "4 GB" "4 GB" ...

$ Storage : chr "64 GB" "64 GB"..

$ Rating : num 4.5 4.5 4.3 4.3..

$ Selling.Price : int 11990 11990 ...

$ Original.Price: int 15990 15990 ...

##columns

colnames(phsale)

Output:

[1] "Brand" "Model" "Color"

[4] "Memory" "Storage" "Rating"

[7] "Selling.Price" "Original.Price"

###let's look descriptive statistics of the data

summary(phsale)

Output:

Brand Model

Length:2647 Length:2647

Class :character Class :character

Mode :character Mode :character

Color Memory

Length:2647 Length:2647

Class :character Class :character

Mode :character Mode :character

Storage Rating

Length:2647 Min. :0.000

Class :character 1st Qu.:4.000

Mode :character Median :4.300

Mean :4.002

3rd Qu.:4.400

Max. :5.000

Selling.Price Original.Price

Min. : 1000 Min. : 1599

1st Qu.: 9490 1st Qu.: 12999

Median : 14999 Median : 18999

Mean : 26461 Mean : 29853

3rd Qu.: 29998 3rd Qu.: 34999

Max. :179900 Max. :189999

NA's :3 NA's :1678

####checking null values

colSums(is.na(phsale))

Output:

Brand Model

1. 0

Color Memory

0 0

Storage Rating

3 0

Selling.Price Original.Price

3 1678

Data cleaning:

###storage

class(phsale$Storage)

Output:

[1] "character"

sum(is.na(phsale$Storage))

Output:

[1] 3

###replacing NA values

phsale$Storage[phsale$Storage==""]<-"8 GB"

phsale$Storage[is.na(phsale$Storage)]<-"8 GB"

###checking

sum(is.na(phsale$Storage))

Output:

[1] 0

### selling.price

class(phsale$Selling.Price)

Output:

[1] "integer"

sum(is.na(phsale$Selling.Price))

Output:

[1] 3

##replacing NA values with mean

m<-round(mean(phsale$Selling.Price,na.rm = TRUE),0)

phsale$Selling.Price[is.na(phsale$Selling.Price)]<-m

##checking

sum(is.na(phsale$Selling.Price))

Output:

[1] 0

####Original price

class(phsale$Original.Price)

Output:

[1] "integer"

sum(is.na(phsale$Original.Price))

Output:

[1] 1678

phsale%>%select(Brand,Selling.Price,Original.Price)

Output:

Brand Selling.Price Original.Price OPPO 17990 18990 OPPO 10490 12990

OPPO 11960 12990 OPPO 11779 11919 OPPO 18979 NA OPPO 8990 1099 OPPO 19990 25990 OPPO 15990 NA OPPO 14990 NA

OPPO 29990 35990 OPPO 29990 35990 OPPO 19990 25990 OPPO 14990 NA OPPO 15990 NA OPPO 16990 17990 OPPO 16990 17990

##replacing NA values

phsale$Original.Price[is.na(phsale$Original.Price)]<-phsale$Selling.Price[is.na(phsale$Original.Price)]

Output:

Brand Selling.Price Original.Price OPPO 17990 18990 OPPO 10490 12990 OPPO 11960 12990 OPPO 11779 11919 OPPO 18979 18979 OPPO 8990 10990 OPPO 19990 25990 OPPO 15990 15990 OPPO 14990 14990 OPPO 29990 35990 OPPO 29990 35990 OPPO 19990 25990

##checking

sum(is.na(phsale$Original.Price))

Output:

[1] 0

Data Transformation:

###memory

class(phsale$Memory)

Output:

[1] "character"

table(phsale$Memory)

####changing into factor

phsale$Memory<-as.factor(phsale$Memory)

class(phsale$Memory)

Output:

[1] "factor"

##storage

class(phsale$Storage)

Output:

[1] "character"

table(phsale$Storage)

phsale$Storage<-as.factor(phsale$Storage)

class(phsale$Storage)

Output:

[1] "factor"

#####renaming column names

phsale<-rename(phsale,Selling\_Price=Selling.Price)

phsale<-rename(phsale,Original\_Price=Original.Price)

Output:

Rating Selling\_Price Original\_Price

4.5 11990 15990

4.5 11990 15990

Exploratory data analysis:

##oppo brand with 128 gb storage

oppo\_128<-phsale%>%

select(Brand,Storage,Selling\_Price)%>%

filter(Brand=="OPPO" & Storage=="128 GB")

View(oppo\_128)

Output:

Brand Storage Selling\_Price

1 OPPO 128 GB 13990

2 OPPO 128 GB 13990

3 OPPO 128 GB 13990

4 OPPO 128 GB 15990

5 OPPO 128 GB 15990

6 OPPO 128 GB 17990

7 OPPO 128 GB 17990

##apple phones under 30000

apple\_30k<-phsale%>%

select(Brand,Model,Selling\_Price)%>%

filter(Brand=="Apple" & Selling\_Price<=30000)

View(apple\_30k)

Output:

Brand Model Selling\_Price

1 Apple iPhone 6s 25299

2 Apple iPhone 7 24999

3 Apple iPhone 7 24999

4 Apple iPhone 6s 25299

5 Apple iPhone 7 24999

6 Apple iPhone 7 24999

7 Apple iPhone 6s 25299

8 Apple iPhone 7 24999

### adding new column

phsale<-phsale%>%

mutate(price\_catg=ifelse(Selling\_Price<=30000,"low cost",

ifelse(Selling\_Price>30000 & Selling\_Price<=70000,

"medium cost","high cost")))

View(phsale)

Output:

Brand Selling\_Price price\_catg OPPO 11990 low cost

OPPO 11990 low cost

OPPO 13990 low cost OPPO 13990 low cost

OPPO 11990 low cost

OPPO 13990 low cost

OPPO 10490 low cost

##lenovo phones with 4.0 above rating and

## 64 gb storage

len\_64gb<-phsale%>%

select(Brand,Rating,Storage,Selling\_Price)%>%

filter(Brand=="Lenovo" & Rating>4.0 & Storage=="64 GB")

View(len\_64gb)

Output:

Brand Rating Storage Selling\_Price

1 Lenovo 4.1 64 GB 8499

2 Lenovo 4.1 64 GB 12999

3 Lenovo 4.1 64 GB 13499

4 Lenovo 4.1 64 GB 16999

5 Lenovo 4.1 64 GB 13999

6 Lenovo 4.1 64 GB 11999

7 Lenovo 4.1 64 GB 13999

8 Lenovo 4.1 64 GB 11999

9 Lenovo 4.1 64 GB 11999

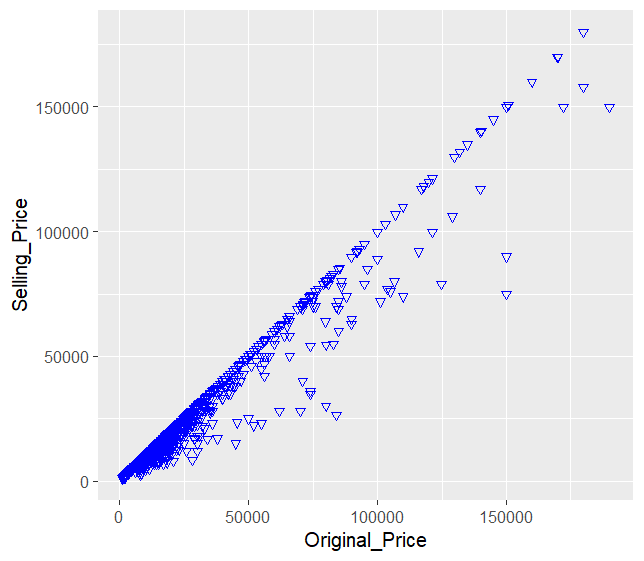
####relationship between selling price and

str(phsale)

ggplot(phsale,aes(x=Original\_Price,y=Selling\_Price))+

geom\_point(col="blue",pch=6)

Output:



It is positively correlated if original price

increases selling price also increases.

### most rated mobiles

tot\_r<-phsale%>%group\_by(Brand)%>%

summarise(tol\_ratings=sum(Rating>=4.0))

View(tot\_r)

Output:

Brand tol\_ratings

*<chr>* *<int>*

1 ASUS 78

2 Apple 311

3 GIONEE 53

4 Google Pixel 29

5 HTC 1

6 IQOO 5

7 LG 61

8 Lenovo 76

9 Motorola 87

10 Nokia 121

11 OPPO 229

12 POCO 20

13 SAMSUNG 611

14 Xiaomi 154

15 realme 253

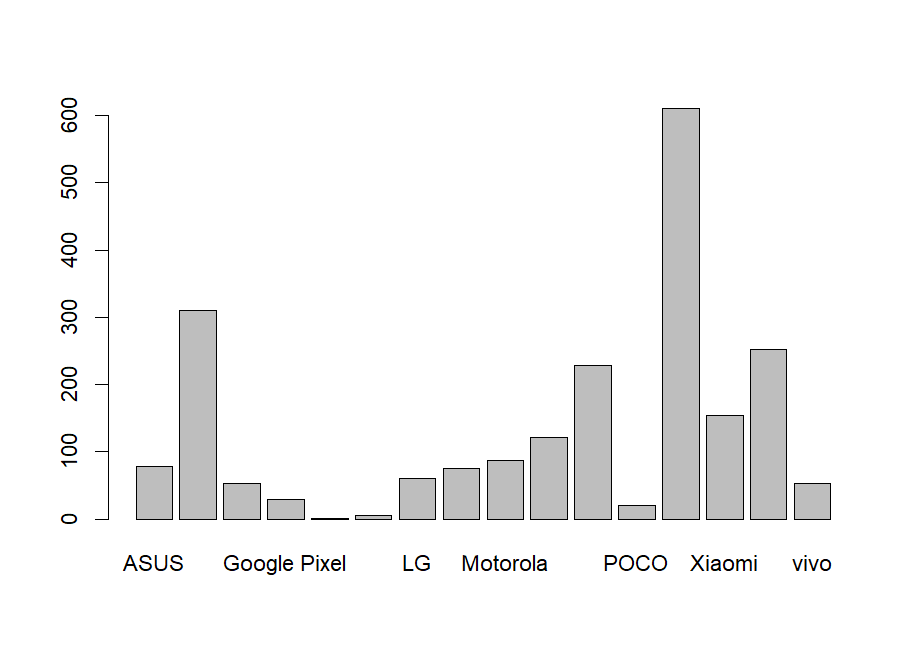
16 vivo 53

##barplot

barplot(tot\_r$tol\_ratings,

names.arg = tot\_r$Brand)

Output:



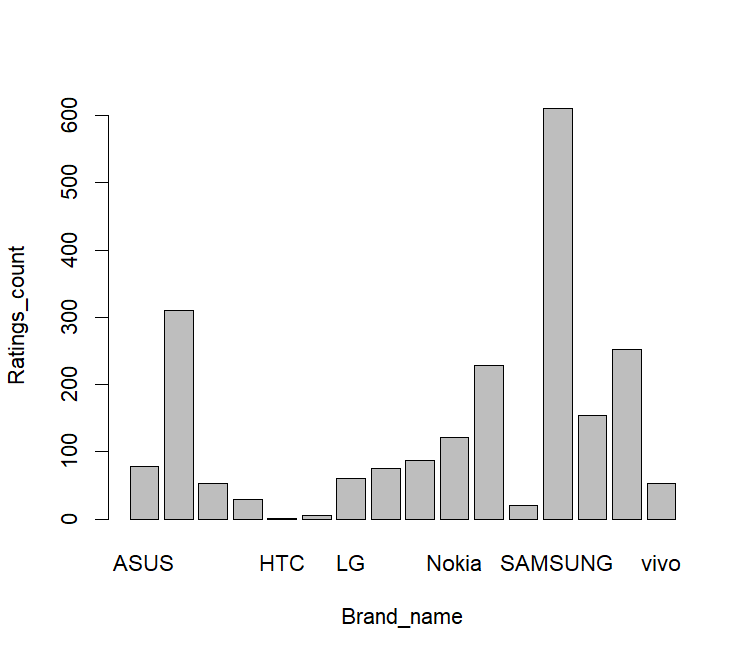
##adding x and y lab

barplot(tot\_r$tol\_ratings,

names.arg = tot\_r$Brand,

xlab="Brand\_name",ylab="Ratings\_count")

Output:



##adding title and color

barplot(tot\_r$tol\_ratings,

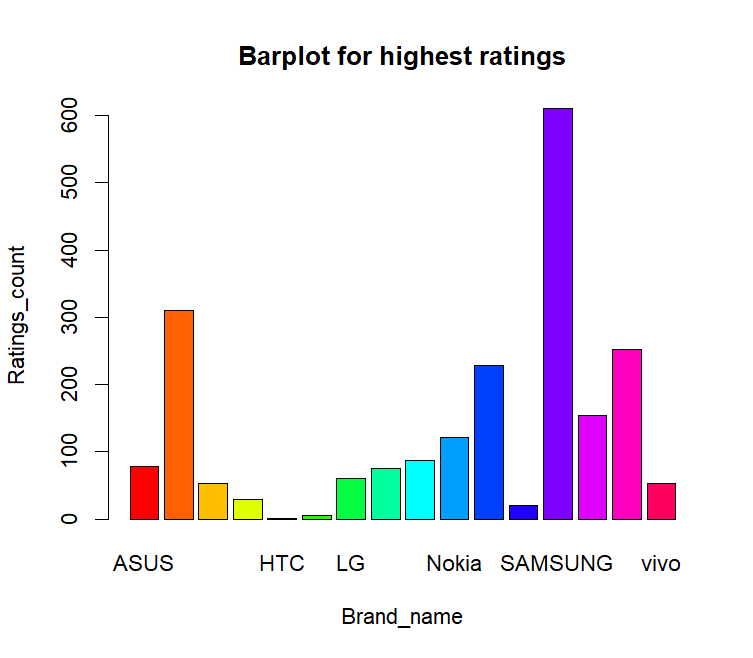
names.arg = tot\_r$Brand,

xlab="Brand\_name",ylab="Ratings\_count",

main="Barplot for highest ratings",

col=rainbow(length(tot\_r$Brand)))

Output:



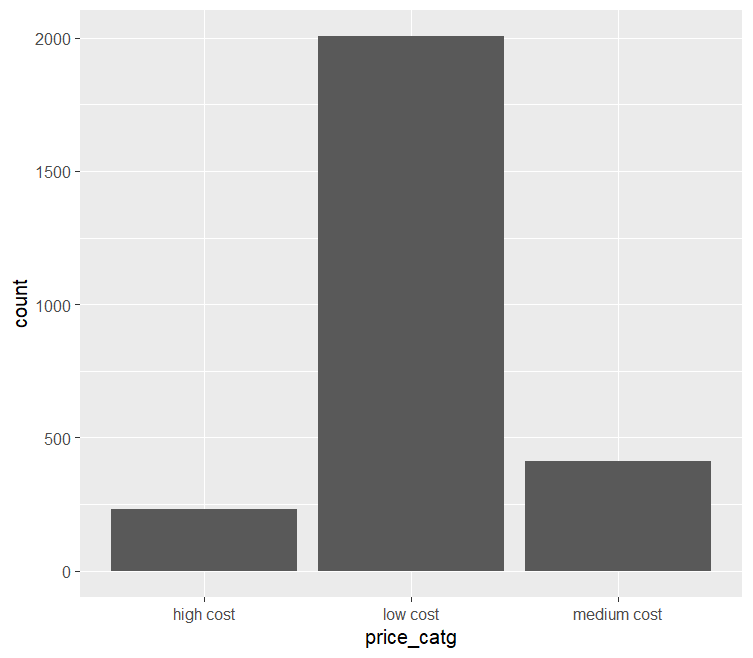
From the above graph most rated phones are

samsung and then apple

###barplot for rating division

ggplot(phsale,aes(x=price\_catg))+geom\_bar()

Output:

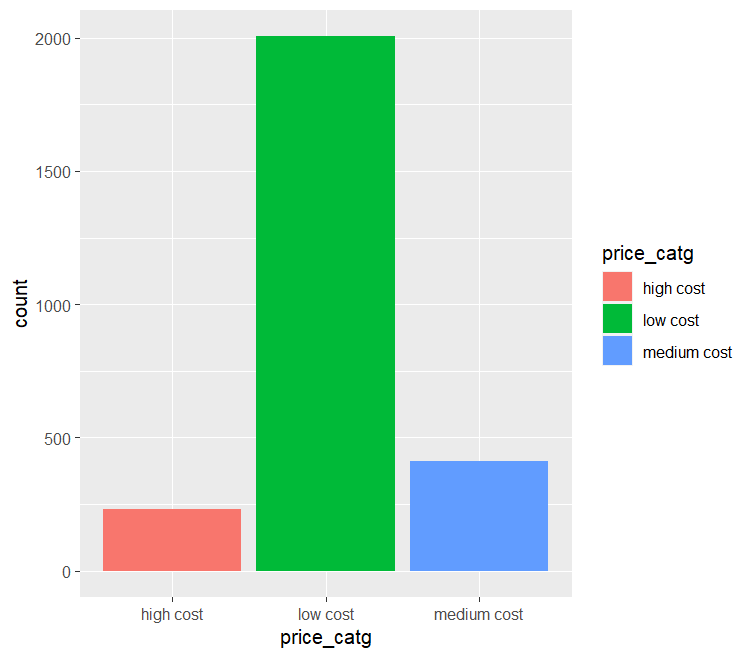


##adding color

ggplot(phsale,aes(x=price\_catg,fill=price\_catg))+

geom\_bar()

Output:



###visualizing statistical information of selling\_price column

ggplot(data=phsale,aes(x=Selling\_Price))+

geom\_boxplot(fill="palegreen")

output:



From the above graph the maximum rate of mobile phone is nearly

170000 and minimum rate is 1000 and the average rate of mobile

phones is in between 14000 to 15000.