BA_HW2

Problem 1:

```
url = "https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/JC-20
1709-citibike-tripdata.csv"
citibike = read.csv(url)
df = as.data.frame(citibike)
```

summary statistics for tripduration

```
summary(df$tripduration)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
##      61.0     238.0     355.0     756.9     610.0 2181628.0
```

```
var(df$tripduration)
```

```
## [1] 159480876
```

```
sd(df$tripduration)
```

```
## [1] 12628.57
```

```
range(df$tripduration)
```

```
## [1]      61 2181628
```

summary statistics for age

```
library(lubridate)
age <- function(dob, age.day = today(), units = "years", floor = TRUE) {
    calc.age = interval(dob, age.day) / duration(num = 1, units = units)
    if (floor) return(as.integer(floor(calc.age)))
    return(calc.age)
}
df$birth.year = as.Date(df$birth.year, "%Y")
df$age = age(df$birth.year)
summary(df$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   19.00   33.00   37.00   39.88   45.00  133.00    2384
```

```
var(df$age, na.rm=TRUE)
```

```
## [1] 100.9908
```

```
sd(df$age, na.rm=TRUE)
```

```
## [1] 10.04942
```

```
range(df$age, na.rm=TRUE)
```

```
## [1]  19 133
```

summary statistics for tripduration in minutes

```
df$tripduration_min = round((df$tripduration/60), digits = 2)
summary(df$tripduration_min)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
##     1.02     3.97     5.92    12.62    10.17 36360.47
```

```
var(df$tripduration_min)
```

```
## [1] 44300.25
```

```
sd(df$tripduration_min)
```

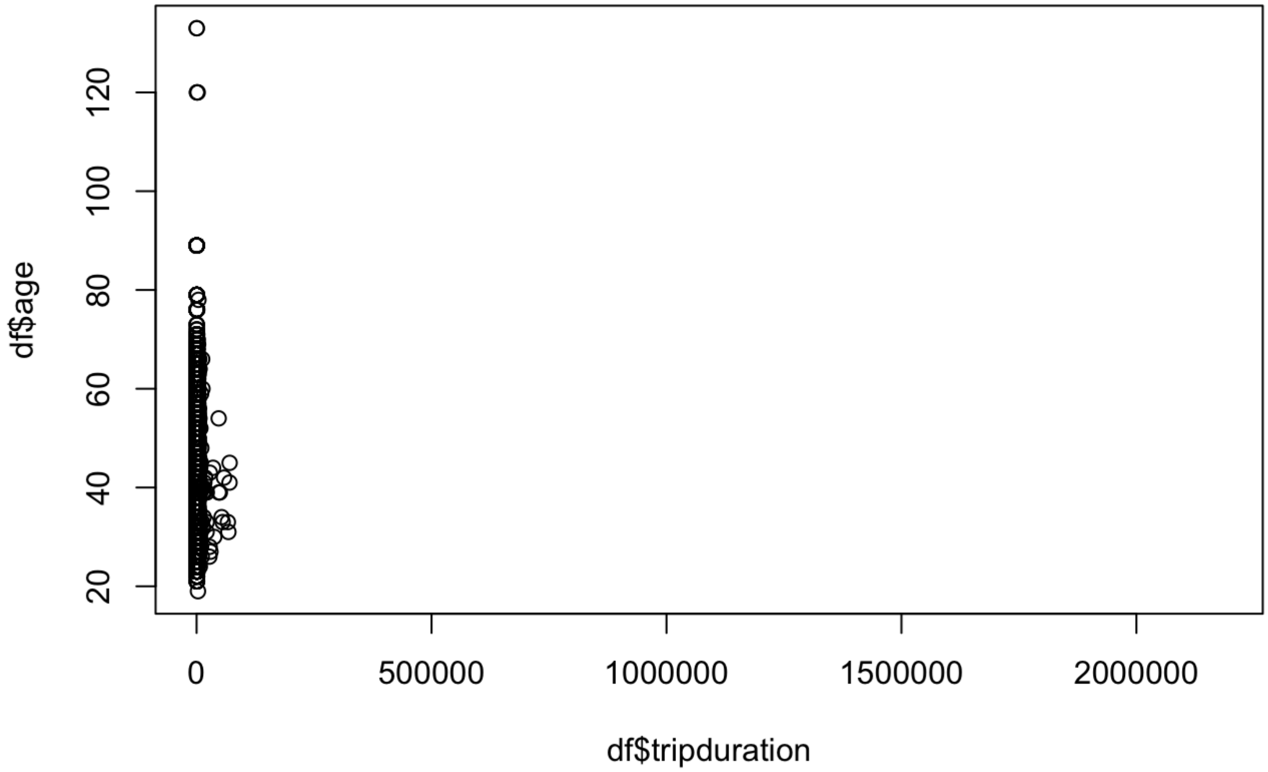
```
## [1] 210.4762
```

```
range(df$tripduration_min)
```

```
## [1]     1.02 36360.47
```

correlation between age and tripduration

```
plot(df$tripduration,df$age)
```



```
cor(df$tripduration, df$age, use = "complete.obs")
```

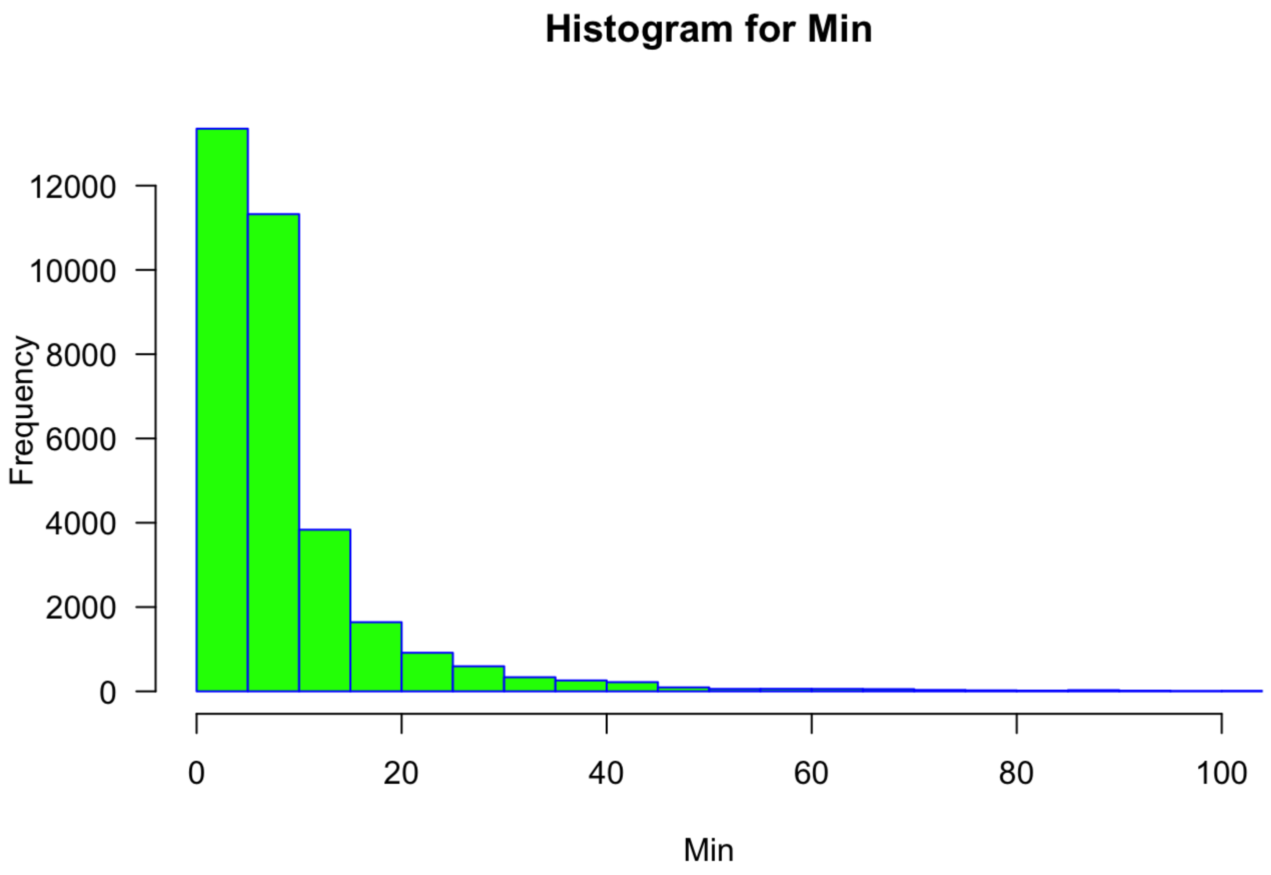
```
## [1] 0.007055148
```

What is the total revenue assuming all users riding bikes from 0 to 45 minutes pay 3 per ride and user exceeding 45 minutes pay an additional $2 per ride.

```
library(dplyr)
count_less45 = df %>% summarize(count = sum(df$tripduration_min<=45))
count_greater45 = df %>% summarize(count = sum(df$tripduration_min>45))
price_less45 = 3
price_greater45 = 5
total_revenue = (count_less45 * price_less45) + (count_greater45 * price_greater45)
total_revenue[1,1]
```

```
## [1] 100651
```

```
hist(df$tripduration_min,
     xlim=c(0,100), breaks=10000,
     xlab="Min",
     border="blue",
     col="green",
     las=1,
     main="Histogram for Min")
```



```
var(df$tripduration_min)
```

```
## [1] 44300.25
```

```
count_less45[1,1]/nrow(df)
```

```
## [1] 0.9804644
```

The variance of trip-duration is huge. The reason is that a small part of clients used citibikes for days that enlarged the range of data. However, more than 98% of clients' trip durations are less than 45 minutes. This means although the scope of trip-duration is broad, the significant citibike clients are those people who return the bikes within 45 min. In other words, the considerable clients of citibike are those people who pay $3 for their rides.

What does this mean for the pricing strategy?

As mentioned above, more than 98% of clients of citibike are those people who pay $3 for their rides. This means the dominating reason for charging an additional $2 for trips longer than 45min is not generating more revenues. Because the revenues generated from the additional fee is only a small portion of total revenue. This pricing stretegy is more likely to be a signal to the clients that “$3 per ride is a fair price” because most clients would return the bikes within 45 mins. For citibike, the price strategy of “$3 for all trips from 0 to 45 minutes” can help them to generate more revenues compared to “charge by every minute”. Under the current price strategy, the average revenue per minute is high at the start and gradually goes down. Similarly, the number of clients is also high at the start and gradually goes down. So, this pricing strategy maximized revenue because high average revenue corresponds to a large amount of clients.

What does this mean for inventory availability?

An additional 2 is charged to clients with a ride more than 45 mins indicates that citibike expected most clients return the bikes with 45 mins. This indicates the inventory of citybike is not large enough to support too many long trips. By charging an additional $2 for long trips, citibike forces some clients to return the bike instead of keep the bikes on their hands for a long period.

Prlblem 2:

```
url2 = "https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/zaga
t.CSV"
zagat = read.csv(url2)
df2 = as.data.frame(zagat)
```

Statistics of Central Tendency

```
sapply(df2[,2:5], mean)
```

```
##     Food    Decor  Service    Price
## 19.38667 15.72333 16.89667 36.55000
```

```
library(psych)
sapply(df2[,2:5], geometric.mean)
```

```
##     Food    Decor  Service    Price
## 19.01600 14.76890 16.52509 33.40387
```

```
sapply(df2[,2:5], harmonic.mean)
```

```
##     Food    Decor  Service    Price
## 18.62024 13.49590 16.15621 30.12219
```

```
sapply(df2[,2:5], median)
```

```
##    Food   Decor Service   Price
##      19      16      16      35
```

```
library(DescTools)
sapply(df2[,2:5], Mode)
```

```
##    Food   Decor Service   Price
##      21      13      16      28
```

Statistics of spread and dispersion of the ratings

```
sapply(df2[,2:5], range)
```

```
##      Food Decor Service Price
## [1,]    9     3       8     8
## [2,]   28    27      26    80
```

```
sapply(df2[,2:5], var)
```

```
##      Food     Decor   Service     Price
##  13.63594  24.50179  12.72173 221.31187
```

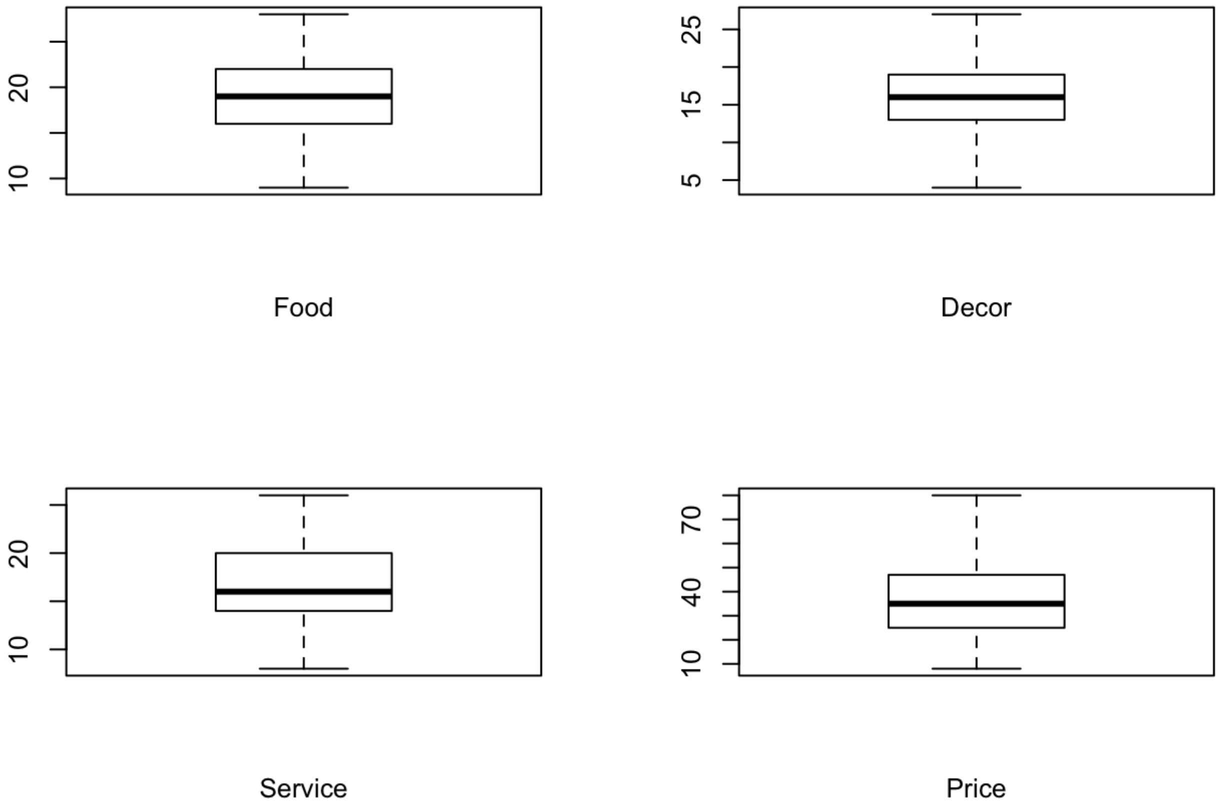
```
sapply(df2[,2:5], sd)
```

```
##      Food     Decor   Service     Price
##  3.692688  4.949929  3.566753 14.876554
```

```
sapply(df2[,2:5], IQR)
```

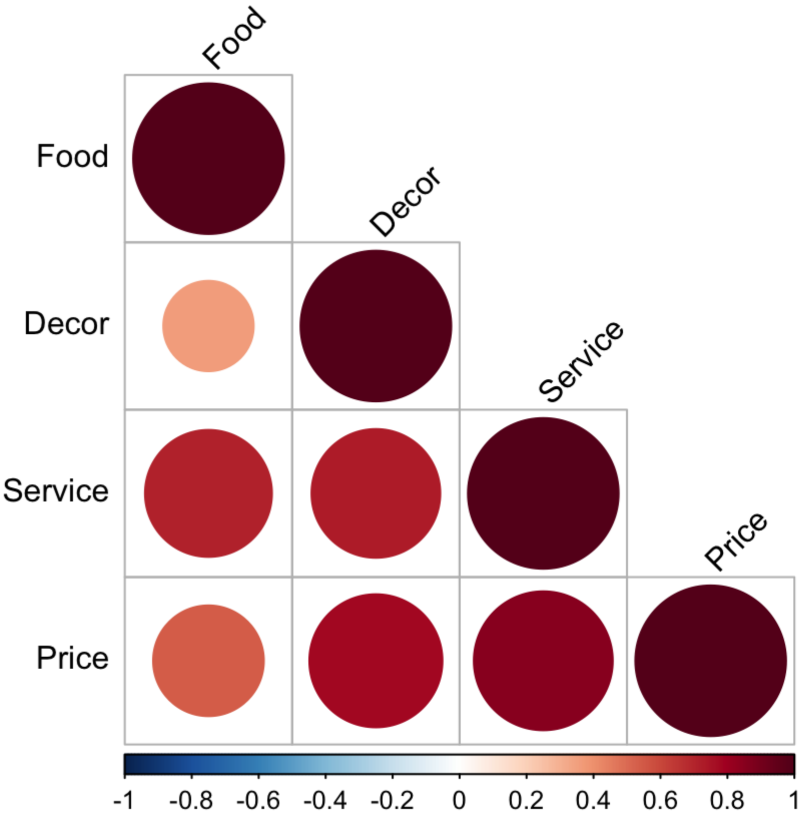
```
##    Food   Decor Service   Price
##       6       6       6      22
```

```
par(mfrow = c(2,2))
boxplot(df2$Food, xlab="Food", outline = F)
boxplot(df2$Decor, xlab="Decor", outline = F)
boxplot(df2$Service, xlab="Service", outline = F)
boxplot(df2$Price, xlab="Price", outline = F)
```



correlations between rating dimensions

```
source("http://www.sthda.com/upload/rquery_cormat.r")
rquery.cormat(df2[,2:5])
```



```
## $r
##         Food Decor Service Price
## Food       1
## Decor   0.36     1
## Service 0.71  0.73       1
## Price   0.54  0.78    0.85     1
##
## $p
##            Food   Decor Service Price
## Food          0
## Decor   9.3e-11       0
## Service   3e-47 1.9e-50       0
## Price   6.7e-24 2.4e-62 1.9e-84     0
##
## $sym
##         Food Decor Service Price
## Food    1
## Decor   .    1
## Service ,    ,     1
## Price   .    ,     +       1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1
```

weighted average (index) that computes scores for each restaurant. (weighted entropy method)

```
df2$Name <- NULL
min.max.norm <- function(x){
  (x-min(x))/(max(x)-min(x))
}

w1 = sapply(df2, min.max.norm)

first1 <- function(data)
{
  x <- c(data)
  for(i in 1:length(data))
    x[i] = data[i]/sum(data[])
  return(x)
}

df3 = apply(w1,2,first1)

first2 <- function(data)
{
  x <- c(data)
  for(i in 1:length(data)){
    if(data[i] == 0){
      x[i] = 0
    }else{
      x[i] = data[i] * log(data[i])
    }
  }
  return(x)
}

df4 = apply(df3, 2, first2)

k <- 1/log(length(df4[,1]))
d <- -k * colSums(df4)

d = 1-d

w = d/sum(d)
w
```

```
##      Food     Decor   Service     Price
## 0.1823203 0.2283491 0.2167842 0.3725464
```

What makes a business more profitable?

I believe the first thing to make a business more profitable is to find a balance between price and the number of customers each day. You do not want too many customers waiting outside, and you do not want too many empty tables in the restaurant. The second thing to make a business more profitable is to make the restaurant unique from the competitors.vMost people would like to try something they never met before. To distinguish yourself from others is also essential to attract new customers.

If you were hired to advise a new restaurant operator, what would you recommend in terms of the balance & trade-offs between food, decor, service, and price?

Based on my previous analysis, the weights of four criteria rank high to low as Price > Decor > Service > Food. So, the restaurants should put the most effort into managing their price strategy to attract more customers. Also, they need to put some efforts in service and decor because these two criteria have similar weights, which are around 20%.