

R codes for the project

```
getwd()
setwd("C:/Users/navee/Documents")

#1. dataInput
ecomm.df <- read.csv("ecommerce-data_A4.csv")
str(ecomm.df)
View(ecomm.df)
head(ecomm.df)

#2. Prop table for profile and gender
with(ecomm.df, prop.table(table(profile ,gender)))
with(ecomm.df[ecomm.df$gender=="Male" | ecomm.df$gender=="Female",],
      prop.table(table(profile , gender),
                    margin = 1))

#3. Purchased recently
table(ecomm.df$profile, ecomm.df$purchasedWhen)

#Visualization
ecomm.df$profile <- factor(ecomm.df$profile)
library(lattice)
histogram(~profile | purchasedWhen, data = ecomm.df,
          scales=c(list(x=list(rot=30)), main=list(label="Purchased According to the profile", cex=1.5)),
          xlab=list(label="profile", cex=1),
          ylab=list(label="Purchases", cex=1.3),
          par.settings = list(axis.line = list(col = 0)))

ecomm.df$profile <- factor(ecomm.df$profile)
ecomm.df$purchasedWhen <- factor(ecomm.df$purchasedWhen)
```

#4 Create a contingency table of purchases by profile

```
purchase_counts <- table(ecomm.df$profile, ecomm.df$purchasedWhen)
```

Display the counts

```
print(purchase_counts)
```

Visualization using base R graphics

```
barplot(purchase_counts, beside = TRUE, legend.text = TRUE,  
        xlab = "Purchased When", ylab = "Count",  
        main = "Purchases According to Profile")
```

#5 Split the data into groups based on profile and gender

```
purchase_counts <- table(ecomm.df$profile, ecomm.df$gender, ecomm.df$behavAnySale)
```

Find the combination of profile and gender with the highest number of purchases

```
max_purchase_combination <- which.max(purchase_counts)
```

Calculate the total number of observations

```
total_observations <- nrow(ecomm.df)
```

Calculate the purchase rate for each combination of profile and gender

```
purchase_rates <- purchase_counts / total_observations
```

Find the combination of profile and gender with the highest purchase rate

```
max_purchase_rate_combination <- which.max(purchase_rates)
```

Display the results

```
print("Combination with the highest number of purchases:")
```

```
print(dimnames(purchase_counts)[[1]][max_purchase_combination[1]])
```

```
print(dimnames(purchase_counts)[[2]][max_purchase_combination[2]])
```

```
print("Combination with the highest purchase rate:")  
print(dimnames(purchase_rates)[[1]][max_purchase_rate_combination[1]])  
print(dimnames(purchase_rates)[[2]][max_purchase_rate_combination[2]])
```