Thinx.R

jyothi

Sat May 20 07:22:00 2017

```
 orders_export_1 <- read.csv('/Users/jyothi/Desktop/thinx/orders_export_1.csv', comment.char="~")
 orders_export_2 <- read.csv('/Users/jyothi/Desktop/thinx/orders_export_2.csv', comment.char="~")
 orders_export <- read.csv('/Users/jyothi/Desktop/thinx/orders_export.csv', comment.char="~")
#View(orders_export_2)
# Merging three datasets
mergedf <- rbind( orders_export, orders_export_1,orders_export_2 )
#View(mergedf)
#  Remove # sign before Name field

mergedf$Name <- substring(mergedf$Name, 2)
mergedf$Billing.Zip <- substring(mergedf$Billing.Zip, 2)
mergedf$Shipping.Zip <- substring(mergedf$Shipping.Zip, 2)

#mergedf[["Subtotal"]][is.na(mergedf[["Subtotal"]])] <- 0
subDf <- subset(mergedf, select=c("Name", "Created.at","Lineitem.name","Lineitem.price","Lineitem.quant
#View(subDf)
#hiphugger<- subset(subDf ,subDf$Lineitem.name = 'Hiphugger ')
#subDf[Lineitem.name %like% "Hiphugger"]
#df1 <- read.csv('/Users/jyothi/Desktop/thinx.csv')
p1 <- 'Hiphugger'
df1 <- subset(subDf, grepl(p1,Lineitem.name ) )
#View(df1)

#hiphuggerprice <- subset(df1 ,df1$Lineitem.price != 34)
#View(hiphuggerprice)
summary(df1)
```

```
##      Name                              Created.at
##  Length:57081       2015-12-14 11:15:27 -0500:   14
##  Class :character   2016-03-04 15:44:05 -0500:   13
##  Mode  :character   2016-01-12 10:22:11 -0500:   10
##                     2016-02-08 20:14:40 -0500:    8
##                     2016-02-25 17:06:11 -0500:    7
##                     2015-12-30 11:50:05 -0500:    7
##                     (Other)                 :57022
##                Lineitem.name   Lineitem.price Lineitem.quantity
##  Hiphugger - M / Black :15130   Min.   :34     Min.   : 1.000
##  Hiphugger - S / Black :11311   1st Qu.:34     1st Qu.: 1.000
##  Hiphugger - L / Black : 9781   Median :34     Median : 1.000
##  Hiphugger - XL / Black: 4593   Mean   :34     Mean   : 1.448
##  Hiphugger - XS / Black: 3228   3rd Qu.:34     3rd Qu.: 2.000
##  Hiphugger - M / Beige : 3029   Max.   :34     Max.   :41.000
##  (Other)              :10009
##  Lineitem.discount Lineitem.fulfillment.status   Lineitem.sku
##  Min.   :  0.000   fulfilled:56942             TXHH0103:15130
##  1st Qu.:  0.000   pending  :  139             TXHH0102:11311
##  Median :  0.000                               TXHH0104: 9781
```

```
##  Mean   :  2.978                                  TXHH0105: 4593
##  3rd Qu.:  3.400                                  TXHH0101: 3228
##  Max.   :160.590                                  TXHH0203: 3029
##  NA's   :1                                        (Other) :10009
```

```
#write.csv(df1, file = "/Users/jyothi/Desktop/thinx.csv")
df1$Created.date <- as.Date(df1$Created.at ,format= "%Y-%m-%d %H:%M:%S")
df1 <- na.omit(df1)
#View(df1)
attach(df1)
df1$PAD <- with(df1, (Lineitem.price  -(Lineitem.discount/Lineitem.quantity)))
df1$Order.price <- with(df1, (Lineitem.price*Lineitem.quantity)-Lineitem.discount)
library(dplyr)
```

```
##
## 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
```

```
library(lubridate)
PR<- df1 %>% group_by(Created.at.month=floor_date(Created.date, "month")) %>%
  summarize(totalsales=sum(Order.price) )
QR<- df1 %>% group_by(Created.at.month=floor_date(Created.date, "month")) %>%
  summarize(noofitems=sum(Lineitem.quantity) )
DR<- df1 %>% group_by(Created.at.month=floor_date(Created.date, "month")) %>%
  summarize(totaldiscount=sum(Lineitem.discount) )
#View(GR)
QR
```

```
## Source: local data frame [4 x 2]
##
##   Created.at.month noofitems
##              (date)     (int)
## 1       2015-12-01     15782
## 2       2016-01-01     17359
## 3       2016-02-01     18559
## 4       2016-03-01     30979
```

```
DR
```

```
## Source: local data frame [4 x 2]
##
##   Created.at.month totaldiscount
##              (date)         (dbl)
## 1       2015-12-01         35.70
## 2       2016-01-01      22702.08
## 3       2016-02-01      53921.02
## 4       2016-03-01      93336.80
```

```r
#df2 <- subset(df1, format.Date(Created.at, "%d")==31)
#View(df2)
summary(df1)
```

```
##      Name                              Created.at
##  Length:57080       2015-12-14 11:15:27 -0500:   14
##  Class :character   2016-03-04 15:44:05 -0500:   13
##  Mode  :character   2016-01-12 10:22:11 -0500:   10
##                     2016-02-08 20:14:40 -0500:    8
##                     2016-02-25 17:06:11 -0500:    7
##                     2015-12-30 11:50:05 -0500:    7
##                     (Other)                  :57021
##                 Lineitem.name   Lineitem.price Lineitem.quantity
##  Hiphugger - M / Black :15129   Min.   :34     Min.   : 1.000
##  Hiphugger - S / Black :11311   1st Qu.:34     1st Qu.: 1.000
##  Hiphugger - L / Black : 9781   Median :34     Median : 1.000
##  Hiphugger - XL / Black: 4593   Mean   :34     Mean   : 1.448
##  Hiphugger - XS / Black: 3228   3rd Qu.:34     3rd Qu.: 2.000
##  Hiphugger - M / Beige : 3029   Max.   :34     Max.   :41.000
##  (Other)               :10009
##  Lineitem.discount Lineitem.fulfillment.status   Lineitem.sku
##  Min.   :  0.000   fulfilled:56941              TXHH0103:15129
##  1st Qu.:  0.000   pending  :  139              TXHH0102:11311
##  Median :  0.000                                TXHH0104: 9781
##  Mean   :  2.978                                TXHH0105: 4593
##  3rd Qu.:  3.400                                TXHH0101: 3228
##  Max.   :160.590                                TXHH0203: 3029
##                                                 (Other) :10009
##   Created.date             PAD          Order.price
##  Min.   :2015-12-01   Min.   : 5.704   Min.   :  24.79
##  1st Qu.:2016-01-10   1st Qu.:30.600   1st Qu.:  34.00
##  Median :2016-02-13   Median :34.000   Median :  34.00
##  Mean   :2016-02-08   Mean   :32.381   Mean   :  46.27
##  3rd Qu.:2016-03-11   3rd Qu.:34.000   3rd Qu.:  57.80
##  Max.   :2016-03-31   Max.   :34.000   Max.   :1394.00
##
```

```r
sum_df <- cbind(PR,QR,DR)
sum_df
```

```
##   Created.at.month totalsales Created.at.month noofitems Created.at.month
## 1       2015-12-01   536552.3       2015-12-01     15782       2015-12-01
## 2       2016-01-01   567503.9       2016-01-01     17359       2016-01-01
## 3       2016-02-01   577085.0       2016-02-01     18559       2016-02-01
## 4       2016-03-01   959949.2       2016-03-01     30979       2016-03-01
##   totaldiscount
## 1         35.70
## 2      22702.08
## 3      53921.02
## 4      93336.80
```

```r
#Finding monthly
sum_df<- sum_df[ -c(3 , 5) ]
sum_df
```

```
##   Created.at.month totalsales noofitems totaldiscount
## 1       2015-12-01   536552.3     15782         35.70
## 2       2016-01-01   567503.9     17359      22702.08
## 3       2016-02-01   577085.0     18559      53921.02
## 4       2016-03-01   959949.2     30979      93336.80
```

```r
sum_df["totalrevenue"]  <- sum_df$totalsales - sum_df$totaldiscount
sum_df
```

```
##   Created.at.month totalsales noofitems totaldiscount totalrevenue
## 1       2015-12-01   536552.3     15782         35.70     536516.6
## 2       2016-01-01   567503.9     17359      22702.08     544801.8
## 3       2016-02-01   577085.0     18559      53921.02     523164.0
## 4       2016-03-01   959949.2     30979      93336.80     866612.4
```

```r
price.byitem <- list()
 for(i in 0:length(sum_df)-1){
   price.byitem[i] <- sum_df$totalrevenue[i]/sum_df$noofitems[i]
   print(price.byitem[i])
 }
```

```
## list()
## list()
## [[1]]
## [1] 33.99548
##
## [[1]]
## [1] 31.3844
##
## [[1]]
## [1] 28.18923
##
## [[1]]
## [1] 27.97419
```

```r
sum_df["discout.unit"] <- sum_df$totaldiscount /sum_df$noofitems
sum_df
```

```
##   Created.at.month totalsales noofitems totaldiscount totalrevenue
## 1       2015-12-01   536552.3     15782         35.70     536516.6
## 2       2016-01-01   567503.9     17359      22702.08     544801.8
## 3       2016-02-01   577085.0     18559      53921.02     523164.0
## 4       2016-03-01   959949.2     30979      93336.80     866612.4
##   discout.unit
## 1  0.002262071
## 2  1.307798836
## 3  2.905383911
## 4  3.012905517
```

```r
price.byitem <- data.frame(matrix(unlist(price.byitem), nrow=4, byrow=T))
#View(price.byitem )
sum_df = cbind(sum_df, price.byitem)
colnames(sum_df)[7] <- "price.byitem"
sum_df
```

```
##   Created.at.month totalsales noofitems totaldiscount totalrevenue
## 1       2015-12-01   536552.3     15782         35.70     536516.6
## 2       2016-01-01   567503.9     17359      22702.08     544801.8
## 3       2016-02-01   577085.0     18559      53921.02     523164.0
## 4       2016-03-01   959949.2     30979      93336.80     866612.4
##   discout.unit price.byitem
## 1  0.002262071     33.99548
## 2  1.307798836     31.38440
## 3  2.905383911     28.18923
## 4  3.012905517     27.97419
```

```r
per.price <- list()
per.quantity <- list()
sum_df
```

```
##   Created.at.month totalsales noofitems totaldiscount totalrevenue
## 1       2015-12-01   536552.3     15782         35.70     536516.6
## 2       2016-01-01   567503.9     17359      22702.08     544801.8
## 3       2016-02-01   577085.0     18559      53921.02     523164.0
## 4       2016-03-01   959949.2     30979      93336.80     866612.4
##   discout.unit price.byitem
## 1  0.002262071     33.99548
## 2  1.307798836     31.38440
## 3  2.905383911     28.18923
## 4  3.012905517     27.97419
```

```r
### Price Elasticity for 4 months.
for(i in 2:4){
  per.price <- ((sum_df$price.byitem[i]-  sum_df$price.byitem[i-1] )/ sum_df$price.byitem[i])*100
  per.quantity <- ((sum_df$noofitems[i]-  sum_df$noofitems[i-1] )/ sum_df$noofitems[i])*100
  print( per.quantity / per.price)
}
```

```
## [1] -1.091947
## [1] -0.5704478
## [1] -52.1538
```

```r
# There is increase in demand for the month '2016-03-01'.  May be this is due to  either price point ma
# or may be some other factors like advertisement or some other factors.
```