1p1_MarketingDataPreparation.R

atchirc

Mon Apr 10 00:39:40 2017

```
# ****************************************************************************
#             MARKET MIX  MODELLING
#
#       PGDDA ( IIIT Bangalore )
#       April 2017
#       AtchiReddy (atchireddi@gmail.com)
#
#       Marketing DATA PREPARATION
#
# ****************************************************************************
```

```
# ****************************************************************************
#                   LOAD LIRRARY ----
# ****************************************************************************
library(ggplot2)


# ****************************************************************************
#                   LOAD DATA ----
# ****************************************************************************
# Can't use xlsx/xlconnect package which needs 64bit java version

# . . . .   ProductList ----
productList_data      <-
        read.csv("../input/ProductList.csv", stringsAsFactors = FALSE,
                 na.strings=c('\\N'))

# . . . .   Media Investment ----
mediaInvestment_data  <-
        read.csv("../input/MediaInvestment.csv", stringsAsFactors = FALSE)

# . . . .  Special Sale Event ----

specialSale_data      <-
        read.csv("../input/SpecialSale.csv", stringsAsFactors = FALSE)

# . . . .   Monthly NPS ----
monthlyNPS_data       <-
        read.csv("../input/MonthlyNPSscore.csv", stringsAsFactors = FALSE )


# ****************************************************************************
#                   DATA PREPARATION ----
# ****************************************************************************

# . . . .   ProductList ----
str(productList_data)
```

```
## 'data.frame':    74 obs. of  3 variables:
##  $ Product  : chr  NA "AmplifierReceiver" "AudioMP3Player" "Binoculars" ...
##  $ Frequency: int  5828 4056 112892 14599 2879 987 2269 17523 41307 15660 ...
##  $ Percent  : num  0.4 0.2 6.8 0.9 0.2 0.1 0.1 1.1 2.5 0.9 ...
```

```
atchircUtils::naSummary(productList_data)
```

```
##        Vars NAS     class   perNAS
## 2 Frequency   0   integer 0.000000
## 3   Percent   0   numeric 0.000000
## 1   Product   1 character 1.351351
```

```
productList_data <- na.omit(productList_data)

# . . . . . .  Correct Data types ----
productList_data$Frequency <- as.integer(productList_data$Frequency)

# . . . .   Media Investment ----
str(mediaInvestment_data)
```

```
## 'data.frame':    12 obs. of  12 variables:
##  $ Year             : int  2015 2015 2015 2015 2015 2015 2016 2016 2016 2016 ...
##  $ Month            : int  7 8 9 10 11 12 1 2 3 4 ...
##  $ Total.Investment : num  17.1 5.1 96.3 170.2 51.2 ...
##  $ TV               : num  0.2 0 3.9 6.1 4.2 5.4 4.4 2.6 9.3 5.2 ...
##  $ Digital          : num  2.5 1.3 1.4 12.6 1.3 3.1 0.5 1.9 2.1 0.9 ...
##  $ Sponsorship      : num  7.4 1.1 62.8 84.7 14.2 56.7 4.2 11.7 41.6 24.3 ...
##  $ Content.Marketing: num  0 0 0.6 3.4 0.2 1.1 0.9 0.6 0.4 0 ...
##  $ Online.marketing : num  1.3 0.1 16.4 24.4 19.6 22.5 22.9 19.9 18.4 16.5 ...
##  $ Affiliates       : num  0.5 0.1 5 7 6.6 6.8 7.4 6.5 6.2 5.7 ...
##  $ SEM              : num  5 2.5 6.2 31.9 5.2 11.2 4.2 4.9 5.2 4.2 ...
##  $ Radio            : num  NA NA NA NA NA NA 2.7 NA 0.9 NA ...
##  $ Other            : num  NA NA NA NA NA NA 27.1 NA 15.9 NA ...
```

```
# . . . . . .  Missing Values ----
mediaInvestment_data[is.na(mediaInvestment_data)] <- 0   # zero investment

# . . . .   SPecialSale ----
str(specialSale_data)
```

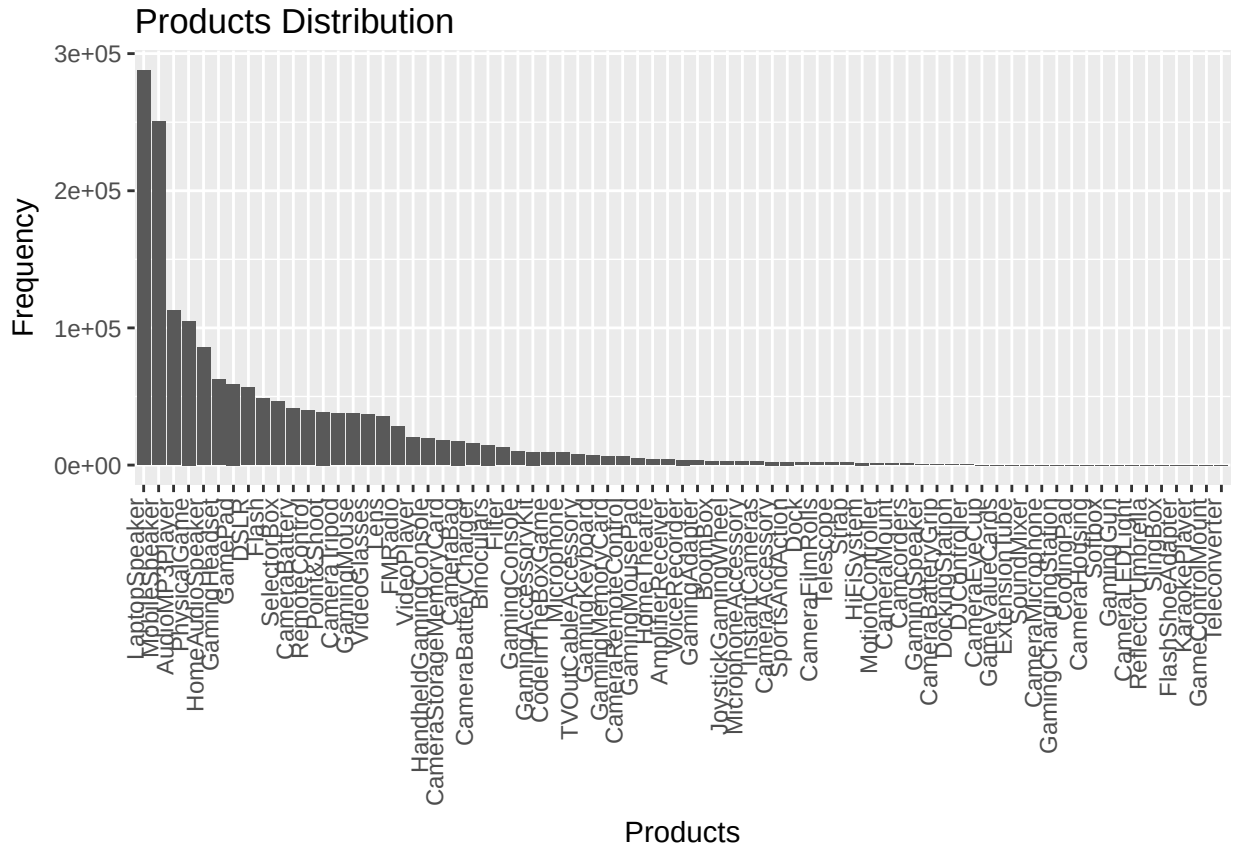
```
## 'data.frame':    44 obs. of  2 variables:
##  $ Day         : chr  "7/18/2015" "7/19/2015" "8/15/2015" "8/16/2015" ...
##  $ SaleOccasion: chr  "Eid_RathaYatraSale" "Eid_RathaYatraSale" "IndependenceSale" "IndependenceSale
```

```
specialSale_data$Day          <- as.Date(specialSale_data$Day, format = "%m/%d/%Y")
specialSale_data$week <- atchircUtils::nweek(specialSale_data$Day,origin = as.Date("2015-07-01"))


# . . . .   Monthly NPS ----
str(monthlyNPS_data)
```

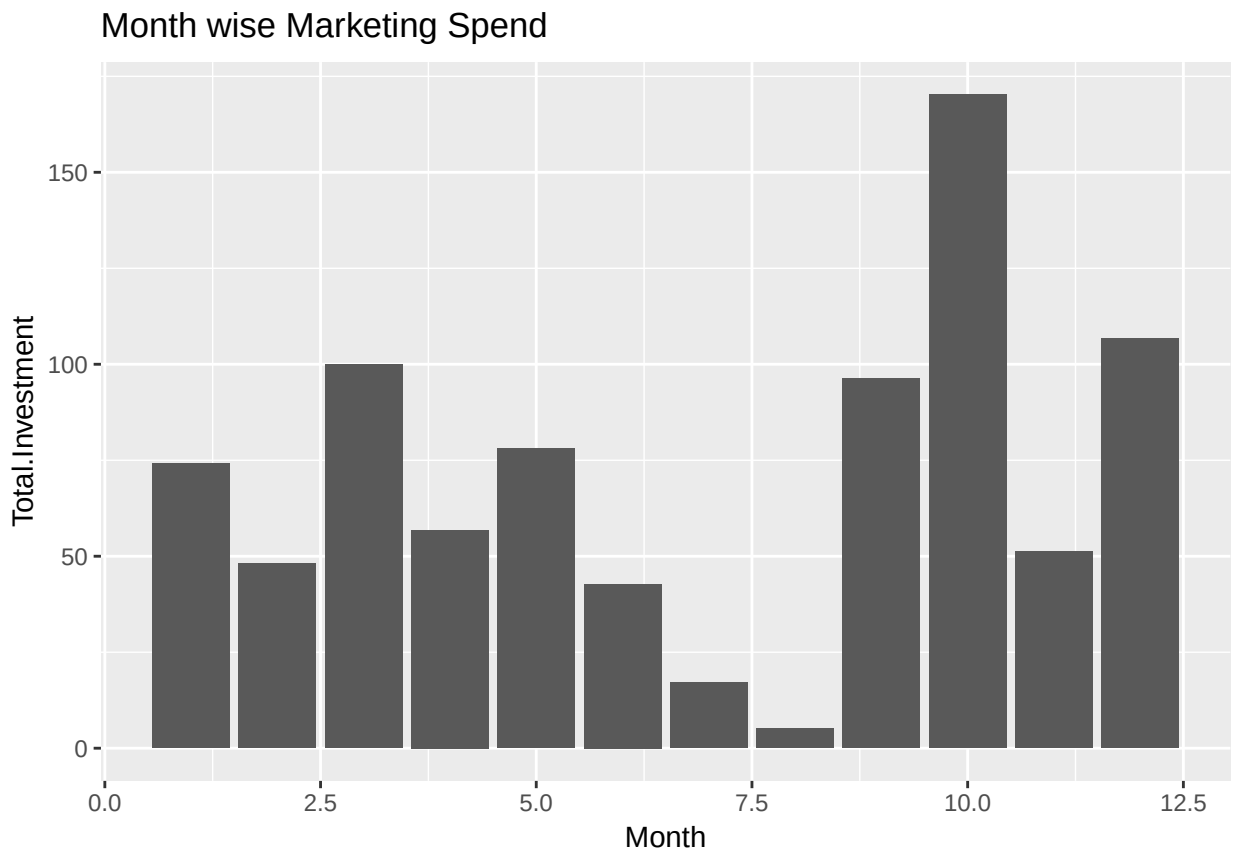
```
## 'data.frame':    12 obs. of  2 variables:
##  $ Month: int  7 8 9 10 11 12 1 2 3 4 ...
##  $ NPS  : num  54.6 60 46.9 44.4 47 45.8 47.1 50.3 49 51.8 ...
```

```
# ***************************************************************************
#                   EDA ----
```

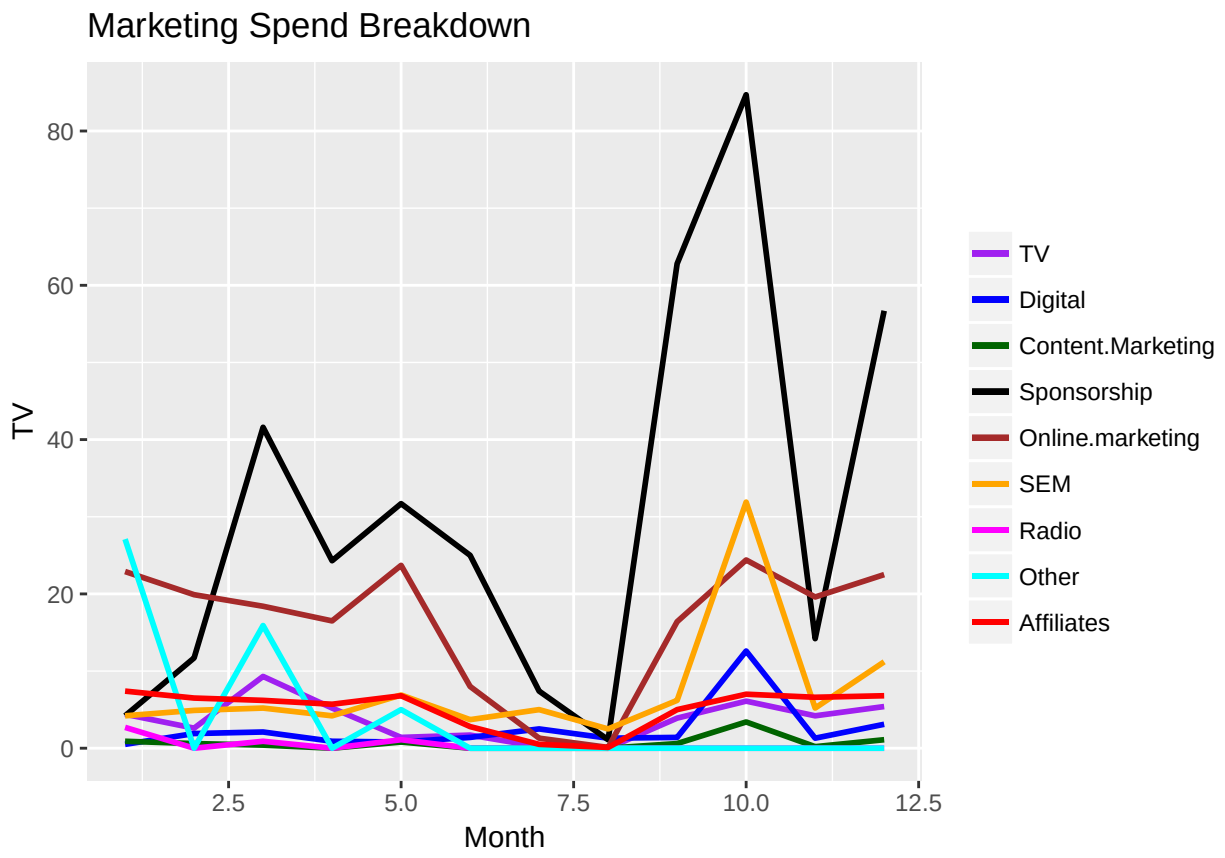
```
# ***************************************************************************

# . . . .   Product List ----
plt <- ggplot(productList_data, aes(x=reorder(Product,-Frequency),Frequency)) +
        geom_bar(stat = 'identity') +
        theme(axis.text.x = element_text(angle = 90, hjust = 1,vjust = 0)) +
        labs(x='Products',title='Products Distribution')

plt
```



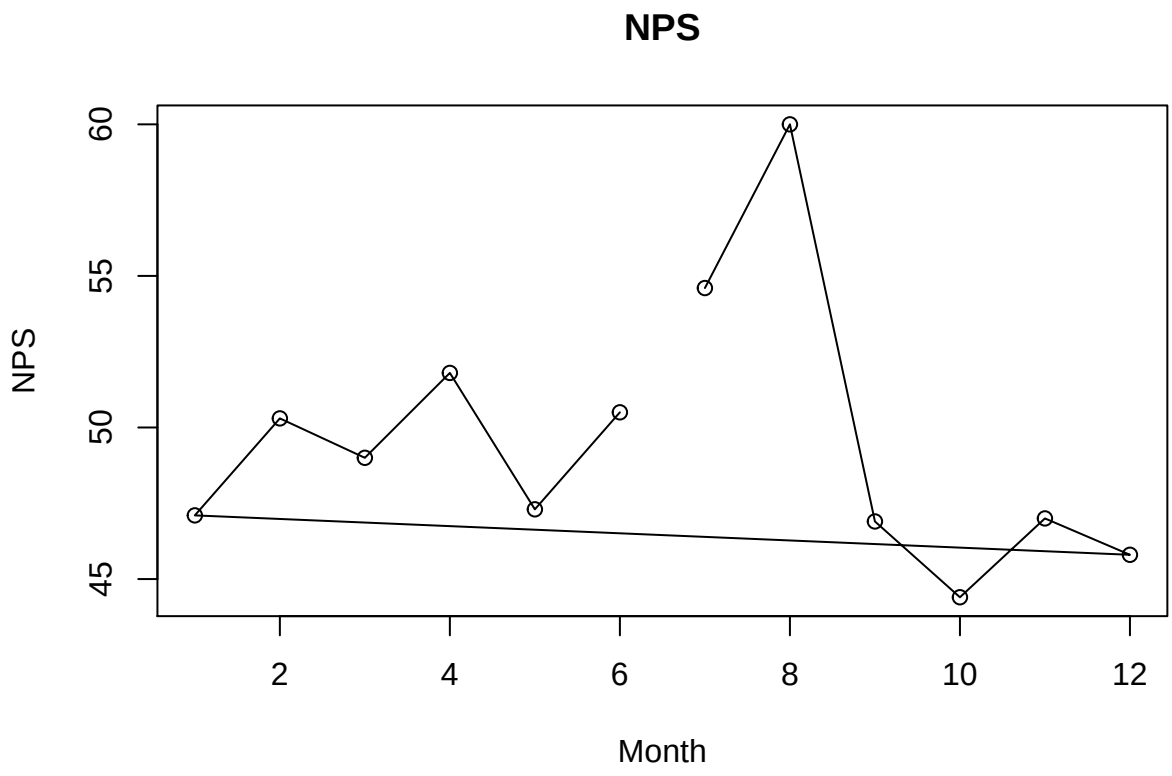

```
# . . . .   Media Investment ----

# Month wise Media Investment
plt <- ggplot(mediaInvestment_data, aes(Month))
plt <- plt + geom_bar(aes(y=Total.Investment),stat = "identity")
plt <- plt + labs(title="Month wise Marketing Spend")
plt
```




```
# Channel wise Media Investment Breakdown
plt <- ggplot(mediaInvestment_data, aes(Month))
plt <- plt + geom_line(aes(y=TV, colour= "TV"), size=1)
plt <- plt + geom_line(aes(y=Digital, colour="Digital"), size=1)
plt <- plt + geom_line(aes(y=Content.Marketing, colour="Content.Marketing"),size=1)
plt <- plt + geom_line(aes(y=Sponsorship, colour="Sponsorship"),size=1)
plt <- plt + geom_line(aes(y=Online.marketing, colour="Online.marketing"),size=1)
plt <- plt + geom_line(aes(y=SEM, colour="SEM"),size=1)
plt <- plt + geom_line(aes(y=Radio, colour="Radio"),size=1)
plt <- plt + geom_line(aes(y=Other, colour="Other"),size=1)
plt <- plt + geom_line(aes(y=Affiliates, colour="Affiliates"),size=1)
plt <- plt + scale_colour_manual("",
                breaks = c("TV", "Digital","Content.Marketing","Sponsorship",
                           "Online.marketing","SEM","Radio","Other","Affiliates"),
                values = c("red", "dark green","blue","brown","cyan","magenta",
                           "orange","black","purple"))
plt + labs(title="Marketing Spend Breakdown")
```




```
# . . . .   NPS ----
plot(monthlyNPS_data,main="NPS")
lines(monthlyNPS_data,main="NPS")
```




```
# ********************************************************************************
#                     FEATURE ENGINEERING ----
# ********************************************************************************

# . . . .   Product List ----

# . . . . . . . .  KPI : Type of product ----
# Based on # of items sold, categorize items either Fast moving, rare moving

productList_data$productSales <- "rareMoving"
productList_data[productList_data$Percent>2,'productSales'] <- "MediumMoving"
productList_data[productList_data$Percent>5,'productSales'] <- "FastMoving"


str(productList_data)
```

```
## 'data.frame':    73 obs. of  4 variables:
##  $ Product     : chr  "AmplifierReceiver" "AudioMP3Player" "Binoculars" "BoomBox" ...
##  $ Frequency   : int  4056 112892 14599 2879 987 2269 17523 41307 15660 401 ...
##  $ Percent     : num  0.2 6.8 0.9 0.2 0.1 0.1 1.1 2.5 0.9 0 ...
##  $ productSales: chr  "rareMoving" "FastMoving" "rareMoving" "rareMoving" ...
##  - attr(*, "na.action")=Class 'omit'  Named int 1
##   .. ..- attr(*, "names")= chr "1"
```

```
str(mediaInvestment_data)
```

```
## 'data.frame':    12 obs. of  12 variables:
```

```
##  $ Year             : int  2015 2015 2015 2015 2015 2015 2016 2016 2016 2016 ...
##  $ Month            : int  7 8 9 10 11 12 1 2 3 4 ...
##  $ Total.Investment : num  17.1 5.1 96.3 170.2 51.2 ...
##  $ TV               : num  0.2 0 3.9 6.1 4.2 5.4 4.4 2.6 9.3 5.2 ...
##  $ Digital          : num  2.5 1.3 1.4 12.6 1.3 3.1 0.5 1.9 2.1 0.9 ...
##  $ Sponsorship      : num  7.4 1.1 62.8 84.7 14.2 56.7 4.2 11.7 41.6 24.3 ...
##  $ Content.Marketing: num  0 0 0.6 3.4 0.2 1.1 0.9 0.6 0.4 0 ...
##  $ Online.marketing : num  1.3 0.1 16.4 24.4 19.6 22.5 22.9 19.9 18.4 16.5 ...
##  $ Affiliates       : num  0.5 0.1 5 7 6.6 6.8 7.4 6.5 6.2 5.7 ...
##  $ SEM              : num  5 2.5 6.2 31.9 5.2 11.2 4.2 4.9 5.2 4.2 ...
##  $ Radio            : num  0 0 0 0 0 0 2.7 0 0.9 0 ...
##  $ Other            : num  0 0 0 0 0 0 27.1 0 15.9 0 ...
```

```
str(specialSale_data)
```

```
## 'data.frame':    44 obs. of  3 variables:
##  $ Day         : Date, format: "2015-07-18" "2015-07-19" ...
##  $ SaleOccasion: chr  "Eid_RathaYatraSale" "Eid_RathaYatraSale" "IndependenceSale" "IndependenceSale
##  $ week        : num  3 4 7 8 8 9 9 10 16 16 ...
```

```
str(monthlyNPS_data)
```

```
## 'data.frame':    12 obs. of  2 variables:
##  $ Month: int  7 8 9 10 11 12 1 2 3 4 ...
##  $ NPS  : num  54.6 60 46.9 44.4 47 45.8 47.1 50.3 49 51.8 ...
```

```
# ****************************************************************************
#                    Save Data ----
# ****************************************************************************

write.csv(productList_data,'../intrim/productList.csv')
write.csv(mediaInvestment_data,'../intrim/mediaInvestment.csv')
write.csv(specialSale_data,'../intrim/specialSale.csv')
write.csv(monthlyNPS_data,'../intrim/monthlyNPS.csv')
```

‘‘‘