Temple\_Analytics \_Challenge

Suraj Vidyadaran

Saturday, October 17, 2015

This Data analysis is done as a part of Temple University Analytics challenge to perform data analytics on big data. For more details about the challenge visit<http://ibit.temple.edu/analytics/>

The analysis is done in QVC data to answer the following questions:

#### 1)What are the products and product categories of QVC which sells best in the US market?

#### 2)How effective are QVC's campaigns in driving product sales?

#### 3)How effective are QVC's on-air broadcast and on-air personalities in driving product sales?

#### 4)What is the impact of QVC's social networking presence on driving product sales?

## Step 1- Read the data

**Read Customer Master Data**

customer\_master <- read.csv("F:/Analytics Challenge/full data/customer\_master.csv")

head(customer\_master)

## CUSTOMER\_NBR STATE ZIP\_CODE SHOPPER\_SEGMENT\_CODE EMAIL\_HOLD\_IND  
## 1 151386394 VA 23314 3 N  
## 2 151672483 NC 28717 3 N  
## 3 151548118 NY 10010 NULL N  
## 4 150138949 IN 46143 3 N  
## 5 152280775 CT 06708 4 N  
## 6 150365703 NE 68467 3 N  
## EMAIL\_JANUARAY\_ONLY\_HOLD\_IND NEW\_CUSTOMER\_IND  
## 1 N N  
## 2 N N  
## 3 N N  
## 4 N N  
## 5 N N  
## 6 N N

**Read the Email Campaign Data set**

email\_campaign <- read.csv("F:/Analytics Challenge/full data/email\_campaign.csv")

head(email\_campaign)

## CAMPAIGN\_DATE PRODUCT\_CATEGORY CAMPAIGN\_SPEND  
## 1 2015-01-01 All 7191.95  
## 2 2015-01-01 Accessories 1493.85  
## 3 2015-01-01 Jewelry 380.53  
## 4 2015-01-02 Apparel, Accessories & Jewelry 531.87  
## 5 2015-01-05 Apparel & Accessories 3788.44  
## 6 2015-01-08 All 8632.31  
## CAMPAIGN\_MONTH  
## 1 January  
## 2 January  
## 3 January  
## 4 January  
## 5 January  
## 6 January

**Read the Order Master Data set**

order\_master <- read.csv("F:/Analytics Challenge/full data/order\_master.csv")

**Read the Product airtime data**

product\_airtime <- read.csv("F:/Analytics Challenge/full data/product\_airtime.csv")

head(product\_airtime)

## PRODUCT\_NBR PRODUCT\_DESCRIPTION  
## 167 P150000009 Denim & Co. Original Waist Stretch Pants w/  
## 168 P150000009 Denim & Co. Original Waist Stretch Pants w/  
## 169 P150000009 Denim & Co. Original Waist Stretch Pants w/  
## 170 P150000009 Denim & Co. Original Waist Stretch Pants w/  
## 171 P150000010 Denim & Co. Stretch Leggings  
## 172 P150000010 Denim & Co. Stretch Leggings  
## PRODUCT\_CATEGORY BRAND\_NAME ONAIR\_DATE ONAIR\_START\_TMS  
## 167 Apparel Not Known 2015-01-01 2015-01-01 04:19:09  
## 168 Apparel Not Known 2015-01-01 2015-01-01 04:01:00  
## 169 Apparel Not Known 2015-01-01 2015-01-01 01:19:11  
## 170 Apparel Not Known 2015-01-01 2015-01-01 01:01:01  
## 171 Apparel Not Known 2015-01-01 2015-01-01 04:44:07  
## 172 Apparel Not Known 2015-01-01 2015-01-01 04:09:57  
## ONAIR\_END\_TMS ONAIR\_MINS HOST1 HOST2 ONAIR\_MONTH  
## 167 2015-01-01 04:28:18 9.17 January  
## 168 2015-01-01 04:01:36 0.62 January  
## 169 2015-01-01 01:28:18 9.13 Stauffer Stacey January  
## 170 2015-01-01 01:01:36 0.60 Stauffer Stacey January  
## 171 2015-01-01 04:49:29 5.38 January  
## 172 2015-01-01 04:10:47 0.85 January

**Read the Product Master Data**

product\_master <- read.csv("F:/Analytics Challenge/full data/product\_master.csv")

head(product\_master)

## PRODUCT\_NBR PRODUCT\_DESCRIPTION  
## 1 P150028635 Fitbit Flex Wireless Activity and Sleep Tracker  
## 2 P150025286 SquareTrade 3-Yr Service Contract/Accidental  
## 3 P150048858 Joan Rivers Crystal Pave' Orchid Brooch  
## 4 P150000250 Denim & Co. Original Waist Stretch Capri Pants  
## 5 P150018581 A-D philosophy no reason to hide duo  
## 6 P150039481 Stainless Steel Square Wheat Chain  
## PRODUCT\_CATEGORY BRAND\_NAME  
## 1 Fun & Leisure Fitbit  
## 2 Fun & Leisure SquareTrade  
## 3 Costume Jewelry Joan Rivers  
## 4 Apparel Denim & Co.  
## 5 Health/Beauty philosophy  
## 6 Jewelry Steel by Design

**Read the Social Network Data**

social <- read.csv("F:/Analytics Challenge/full data/social.csv")

head(social)

## ID SOURCE\_TYPE INTERACTION\_DATE  
## 1 1039 FORUMS 3/31/15 00:00:00 EST  
## 2 1040 FORUMS 3/31/15 00:00:00 EST  
## 3 1041 FORUMS 3/31/15 00:00:00 EST  
## 4 1042 FORUMS 3/31/15 00:00:00 EST  
## 5 1043 FORUMS 3/31/15 00:00:00 EST  
## 6 1044 FORUMS 3/31/15 00:00:00 EST  
## POST  
## 1 By that they showed that Auralux Skin Care is functioning in step with their created formula.  
## 2 Now, my grandson will be able to do his homeschool lessons without interruptions from pop-ups and being disconnected from the internet or not being able to connect at all.  
## 3 The company name is Bohus Biotech.  
## 4 so today I decided to google it to find out.  
## 5 Is there nowhere you could add a little room onto the back for your laundry room?  
## 6 Thank you for the clarity of thought and the Spirit-filled message.  
## SENTIMENT  
## 1 Neutral  
## 2 Positive  
## 3 Neutral  
## 4 Neutral  
## 5 Neutral  
## 6 Positive

## Step 2:Exploratory Data Analysis

**1)Customer Master Data**

#Generating the overall summary  
  
summary(customer\_master)

## CUSTOMER\_NBR STATE ZIP\_CODE   
## Min. :150000001 CA : 65556 08759 : 401   
## 1st Qu.:150983274 PA : 59583 11234 : 383   
## Median :151971891 NY : 57971 15601 : 375   
## Mean :151973373 TX : 45107 10314 : 371   
## 3rd Qu.:152961400 FL : 44904 11236 : 361   
## Max. :153949155 NJ : 32486 15642 : 357   
## (Other):444252 (Other):747611   
## SHOPPER\_SEGMENT\_CODE EMAIL\_HOLD\_IND EMAIL\_JANUARAY\_ONLY\_HOLD\_IND  
## 1 :108507 N:743539 N:734405   
## 2 : 38103 Y: 6320 Y: 15454   
## 3 :190711   
## 4 :127760   
## 5 : 39204   
## NULL:245574   
##   
## NEW\_CUSTOMER\_IND  
## N:656657   
## Y: 93202   
##   
##   
##   
##   
##

#Sales per US states  
  
Customer\_States<-as.data.frame(table(customer\_master$STATE))  
colnames(Customer\_States)<-c("State","Sales")  
Customer\_States<-Customer\_States[order(-Customer\_States$Sales),]  
Customer\_States

## State Sales  
## 11 CA 65556  
## 60 PA 59583  
## 55 NY 57971  
## 70 TX 45107  
## 18 FL 44904  
## 49 NJ 32486  
## 56 OH 32104  
## 38 MI 27753  
## 27 IL 26640  
## 45 NC 25141  
## 73 VA 23863  
## 20 GA 21057  
## 36 MD 20368  
## 28 IN 16137  
## 69 TN 15508  
## 34 MA 14510  
## 78 WA 13689  
## 32 LA 12532  
## 14 CT 12479  
## 40 MO 12387  
## 8 AZ 12305  
## 4 AL 11864  
## 65 SC 11620  
## 79 WI 11613  
## 31 KY 10991  
## 39 MN 8762  
## 13 CO 8410  
## 80 WV 7941  
## 57 OK 7733  
## 58 OR 7113  
## 54 NV 6953  
## 41 MS 6390  
## 30 KS 6105  
## 25 IA 5457  
## 6 AR 5013  
## 47 NE 4097  
## 51 NM 3905  
## 16 DE 3871  
## 48 NH 3416  
## 71 UT 3356  
## 37 ME 3121  
## 63 RI 2843  
## 26 ID 2584  
## 23 HI 2393  
## 42 MT 2295  
## 66 SD 1758  
## 46 ND 1590  
## 81 WY 1546  
## 15 DC 1500  
## 76 VT 1190  
## 3 AK 1075  
## 61 PR 670  
## 74 VI 421  
## 2 AE 58  
## 5 AP 28  
## 9 BC 27  
## 1 AB 14  
## 21 GU 7  
## 53 NULL 7  
## 62 QC 6  
## 35 MB 4  
## 43 N 4  
## 10 C 3  
## 7 AS 2  
## 17 F 2  
## 19 G 2  
## 33 M 2  
## 67 SK 2  
## 68 T 2  
## 77 W 2  
## 12 CN 1  
## 22 HA 1  
## 24 I 1  
## 29 KA 1  
## 44 NB 1  
## 50 NL 1  
## 52 NS 1  
## 59 P 1  
## 64 S 1  
## 72 V 1  
## 75 VM 1

#Calculate the percentage of sales per state  
Sales\_state\_percentage<-as.data.frame(prop.table(table(customer\_master$STATE))\*100)  
colnames(Sales\_state\_percentage)<-c("State","Sales\_percent")  
Sales\_state\_percentage<-Sales\_state\_percentage[order(-Sales\_state\_percentage$Sales\_percent),]  
Sales\_state\_percentage

## State Sales\_percent  
## 11 CA 8.7424435794  
## 60 PA 7.9458938280  
## 55 NY 7.7309200796  
## 70 TX 6.0153975614  
## 18 FL 5.9883258053  
## 49 NJ 4.3322811355  
## 56 OH 4.2813382249  
## 38 MI 3.7010958060  
## 27 IL 3.5526679016  
## 45 NC 3.3527636529  
## 73 VA 3.1823316117  
## 20 GA 2.8081279281  
## 36 MD 2.7162439872  
## 28 IN 2.1520045769  
## 69 TN 2.0681221403  
## 34 MA 1.9350304524  
## 78 WA 1.8255432021  
## 32 LA 1.6712475279  
## 14 CT 1.6641795324  
## 40 MO 1.6519105592  
## 8 AZ 1.6409751700  
## 4 AL 1.5821641135  
## 65 SC 1.5496246628  
## 79 WI 1.5486911539  
## 31 KY 1.4657422262  
## 39 MN 1.1684863421  
## 13 CO 1.1215441836  
## 80 WV 1.0589990918  
## 57 OK 1.0312605436  
## 58 OR 0.9485783327  
## 54 NV 0.9272409880  
## 41 MS 0.8521602061  
## 30 KS 0.8141530608  
## 25 IA 0.7277368145  
## 6 AR 0.6685256828  
## 47 NE 0.5463693841  
## 51 NM 0.5207645704  
## 16 DE 0.5162303846  
## 48 NH 0.4555523105  
## 71 UT 0.4475508062  
## 37 ME 0.4162115811  
## 63 RI 0.3791379446  
## 26 ID 0.3445981178  
## 23 HI 0.3191266625  
## 42 MT 0.3060575388  
## 66 SD 0.2344440755  
## 46 ND 0.2120398635  
## 81 WY 0.2061720937  
## 15 DC 0.2000376071  
## 76 VT 0.1586965016  
## 3 AK 0.1433602851  
## 61 PR 0.0893501312  
## 74 VI 0.0561438884  
## 2 AE 0.0077347875  
## 5 AP 0.0037340353  
## 9 BC 0.0036006769  
## 1 AB 0.0018670177  
## 21 GU 0.0009335088  
## 53 NULL 0.0009335088  
## 62 QC 0.0008001504  
## 35 MB 0.0005334336  
## 43 N 0.0005334336  
## 10 C 0.0004000752  
## 7 AS 0.0002667168  
## 17 F 0.0002667168  
## 19 G 0.0002667168  
## 33 M 0.0002667168  
## 67 SK 0.0002667168  
## 68 T 0.0002667168  
## 77 W 0.0002667168  
## 12 CN 0.0001333584  
## 22 HA 0.0001333584  
## 24 I 0.0001333584  
## 29 KA 0.0001333584  
## 44 NB 0.0001333584  
## 50 NL 0.0001333584  
## 52 NS 0.0001333584  
## 59 P 0.0001333584  
## 64 S 0.0001333584  
## 72 V 0.0001333584  
## 75 VM 0.0001333584

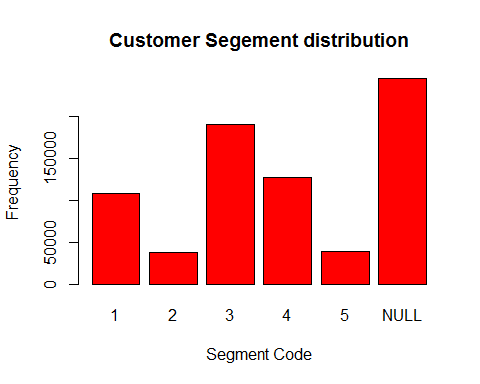
# Customer proportion in the code segment  
  
table(customer\_master$SHOPPER\_SEGMENT\_CODE)

##   
## 1 2 3 4 5 NULL   
## 108507 38103 190711 127760 39204 245574

segemnt\_code\_prop<-prop.table(table(customer\_master$SHOPPER\_SEGMENT\_CODE))  
segemnt\_code\_prop

##   
## 1 2 3 4 5 NULL   
## 0.14470320 0.05081355 0.25432915 0.17037870 0.05228183 0.32749357

plot(customer\_master$SHOPPER\_SEGMENT\_CODE,col="red",main="Customer Segement distribution",xlab="Segment Code",ylab="Frequency")



#Proportion of yes and no  
table(customer\_master$EMAIL\_HOLD\_IND)

##   
## N Y   
## 743539 6320

# No of new customer emails  
table(customer\_master$EMAIL\_JANUARAY\_ONLY\_HOLD\_IND)

##   
## N Y   
## 734405 15454

#Subsettiing customers with email sent  
customer\_email<-subset(customer\_master,customer\_master$EMAIL\_HOLD\_IND=="Y")  
head(customer\_email)

## CUSTOMER\_NBR STATE ZIP\_CODE SHOPPER\_SEGMENT\_CODE EMAIL\_HOLD\_IND  
## 87398 151042164 MD 21244 NULL Y  
## 155581 151208735 PA 19003 NULL Y  
## 199923 152286187 NY 13224 NULL Y  
## 200569 151682957 WA 98816 NULL Y  
## 206947 152302387 NY 12503 NULL Y  
## 232841 151848333 OH 44256 NULL Y  
## EMAIL\_JANUARAY\_ONLY\_HOLD\_IND NEW\_CUSTOMER\_IND  
## 87398 N Y  
## 155581 N Y  
## 199923 N Y  
## 200569 N Y  
## 206947 N N  
## 232841 N Y

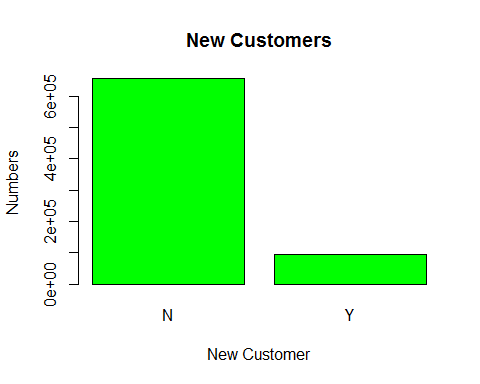
#If the emails are sent NULL shop segemnt code are more  
table(customer\_email$SHOPPER\_SEGMENT\_CODE)

##   
## 1 2 3 4 5 NULL   
## 274 77 5 2 29 5933

#Proportion of new customers  
table(customer\_master$NEW\_CUSTOMER\_IND)

##   
## N Y   
## 656657 93202

#Plot the new customers  
plot(customer\_master$NEW\_CUSTOMER\_IND,col="green",main="New Customers",xlab="New Customer",ylab="Numbers")



#Calculating the percentage of new customers  
percentage\_new\_customers<-(new\_customers/(new\_customers+old\_customers))\*100  
percentage\_new\_customers

## [1] 12.42927

**2)Email Campaign Data**

#Checking the basic summary and structure  
summary(email\_campaign)

## CAMPAIGN\_DATE PRODUCT\_CATEGORY CAMPAIGN\_SPEND   
## Min. :2015-01-01 All :13 Min. : 62.09   
## 1st Qu.:2015-01-24 Apparel : 9 1st Qu.: 891.89   
## Median :2015-02-12 Accessories : 7 Median : 1485.18   
## Mean :2015-02-13 Kitchen & Food : 7 Mean : 2647.54   
## 3rd Qu.:2015-03-05 Electronics : 5 3rd Qu.: 2499.10   
## Max. :2015-03-27 Apparel & Accessories: 4 Max. :19906.61   
## (Other) :16   
## CAMPAIGN\_MONTH   
## Length:61   
## Class :character   
## Mode :character   
##   
##   
##   
##

str(email\_campaign)

## 'data.frame': 61 obs. of 4 variables:  
## $ CAMPAIGN\_DATE : Date, format: "2015-01-01" "2015-01-01" ...  
## $ PRODUCT\_CATEGORY: Factor w/ 11 levels "Accessories",..: 2 1 10 5 4 2 2 8 11 3 ...  
## $ CAMPAIGN\_SPEND : num 7192 1494 381 532 3788 ...  
## $ CAMPAIGN\_MONTH : chr "January" "January" "January" "January" ...

#Changing the format of Date and generating a new attribute month  
email\_campaign$CAMPAIGN\_DATE<-as.Date(email\_campaign$CAMPAIGN\_DATE,"%m/%d/%Y")  
email\_campaign$CAMPAIGN\_MONTH<-months(email\_campaign$CAMPAIGN\_DATE)  
  
#Subsetting the data based on the months  
email\_campaign\_January<-subset(email\_campaign,email\_campaign$CAMPAIGN\_MONTH=="January")  
head(email\_campaign\_January)

## CAMPAIGN\_DATE PRODUCT\_CATEGORY CAMPAIGN\_SPEND  
## 1 2015-01-01 All 7191.95  
## 2 2015-01-01 Accessories 1493.85  
## 3 2015-01-01 Jewelry 380.53  
## 4 2015-01-02 Apparel, Accessories & Jewelry 531.87  
## 5 2015-01-05 Apparel & Accessories 3788.44  
## 6 2015-01-08 All 8632.31  
## CAMPAIGN\_MONTH  
## 1 January  
## 2 January  
## 3 January  
## 4 January  
## 5 January  
## 6 January

#Calculating the total money spent on email campaign in January  
total\_January<-sum(email\_campaign\_January$CAMPAIGN\_SPEND)  
total\_January

## [1] 53478.89

#Checking for which categories more money spent  
table(email\_campaign\_January$PRODUCT\_CATEGORY)

##   
## Accessories   
## 2   
## All   
## 5   
## Apparel   
## 3   
## Apparel & Accessories   
## 1   
## Apparel, Accessories & Jewelry   
## 1   
## Apparel, Accessories, Jewelry & Beauty   
## 0   
## Beauty   
## 0   
## Electronics   
## 2   
## Home & Garden   
## 1   
## Jewelry   
## 1   
## Kitchen & Food   
## 4

#Analysis of February  
email\_campaign\_February<-subset(email\_campaign,email\_campaign$CAMPAIGN\_MONTH=="February")  
head(email\_campaign\_February)

## CAMPAIGN\_DATE PRODUCT\_CATEGORY CAMPAIGN\_SPEND  
## 21 2015-02-01 Apparel & Accessories 2546.84  
## 22 2015-02-01 Electronics 2134.04  
## 23 2015-02-02 Apparel, Accessories & Jewelry 529.21  
## 24 2015-02-03 Apparel 617.92  
## 25 2015-02-04 Accessories 2309.40  
## 26 2015-02-04 Kitchen & Food 891.89  
## CAMPAIGN\_MONTH  
## 21 February  
## 22 February  
## 23 February  
## 24 February  
## 25 February  
## 26 February

total\_February<-sum(email\_campaign\_February$CAMPAIGN\_SPEND)  
total\_February

## [1] 46249.95

table(email\_campaign\_February$PRODUCT\_CATEGORY)

##   
## Accessories   
## 4   
## All   
## 4   
## Apparel   
## 5   
## Apparel & Accessories   
## 1   
## Apparel, Accessories & Jewelry   
## 1   
## Apparel, Accessories, Jewelry & Beauty   
## 0   
## Beauty   
## 2   
## Electronics   
## 1   
## Home & Garden   
## 1   
## Jewelry   
## 0   
## Kitchen & Food   
## 1

#Analysis of March  
email\_campaign\_March<-subset(email\_campaign,email\_campaign$CAMPAIGN\_MONTH=="March")  
total\_March<-sum(email\_campaign\_March$CAMPAIGN\_SPEND)  
total\_March

## [1] 61771.36

head(email\_campaign\_March)

## CAMPAIGN\_DATE PRODUCT\_CATEGORY CAMPAIGN\_SPEND  
## 41 2015-03-01 Beauty 1106.27  
## 42 2015-03-02 Apparel, Accessories, Jewelry & Beauty 1766.68  
## 43 2015-03-03 Apparel & Accessories 2181.10  
## 44 2015-03-03 Kitchen & Food 1236.26  
## 45 2015-03-05 All 8581.39  
## 46 2015-03-05 Home & Garden 1378.40  
## CAMPAIGN\_MONTH  
## 41 March  
## 42 March  
## 43 March  
## 44 March  
## 45 March  
## 46 March

table(email\_campaign\_March$PRODUCT\_CATEGORY)

##   
## Accessories   
## 1   
## All   
## 4   
## Apparel   
## 1   
## Apparel & Accessories   
## 2   
## Apparel, Accessories & Jewelry   
## 1   
## Apparel, Accessories, Jewelry & Beauty   
## 2   
## Beauty   
## 2   
## Electronics   
## 2   
## Home & Garden   
## 2   
## Jewelry   
## 2   
## Kitchen & Food   
## 2

**3)Product Master Data**

#Exploring the structure of product master  
str(product\_master)

## 'data.frame': 48780 obs. of 4 variables:  
## $ PRODUCT\_NBR : Factor w/ 48780 levels "P150000001","P150000002",..: 22948 20789 36470 196 16089 29627 31535 48329 48503 47464 ...  
## $ PRODUCT\_DESCRIPTION: Factor w/ 45420 levels "'Southern Charm' Bracelet by John Wind",..: 17400 37518 22310 13701 5338 38204 38170 18691 11655 12667 ...  
## $ PRODUCT\_CATEGORY : Factor w/ 19 levels "Accessories",..: 8 8 5 3 11 15 15 13 13 12 ...  
## $ BRAND\_NAME : Factor w/ 1885 levels "21 Day Fix","35 Degrees Below",..: 556 1565 829 392 1272 1585 1585 659 293 1177 ...

**4) Order Master Data**

#Explore the summary and structure  
summary(order\_master)  
str(order\_master)  
  
#Changing the order date format and creating a new attribute month  
order\_master$ORDER\_DATE<-as.Date(order\_master$ORDER\_DATE,"%Y-%m-%d")  
str(order\_master$ORDER\_DATE)  
order\_master$ORDER\_MONTH<-months(order\_master$ORDER\_DATE)

#Split and subset order master based on Months  
order\_January<-subset(order\_master,order\_master$ORDER\_MONTH=="January")  
order\_February<-subset(order\_master,order\_master$ORDER\_MONTH=="February")  
order\_March<-subset(order\_master,order\_master$ORDER\_MONTH=="March")  
  
rm(order\_master)

**5)Product Airtime Data**

#Analysing the structure and summary  
summary(product\_airtime)

## PRODUCT\_NBR   
## P150011229: 363   
## P150011250: 236   
## P150013562: 216   
## P150063673: 215   
## P150016256: 208   
## P150016809: 206   
## (Other) :62024   
## PRODUCT\_DESCRIPTION  
## IT Cosmetics Super-Size Celebration Foundation: 363   
## Cottage Farms Bud 'N Flower Booster Fertilizer: 251   
## Calista Tools Perfecter Pro Grip Heated Round : 236   
## IT Cosmetics CC Veil SPF 50 Foundation Cushion: 216   
## Laura Geller Baked Liquid Radiance Foundation : 208   
## IT Cosmetics CC Your Way to Radiant Skin! 5-pc: 206   
## (Other) :61988   
## PRODUCT\_CATEGORY BRAND\_NAME ONAIR\_DATE   
## Jewelry :12623 Not Known :11319 Min. :2015-01-01   
## Home Decor :12129 IT Cosmetics : 2206 1st Qu.:2015-01-26   
## Apparel :10847 Susan Graver : 2201 Median :2015-02-16   
## Health/Beauty: 9723 N/A : 2160 Mean :2015-02-15   
## Accessories : 9137 Cottage Farms: 1881 3rd Qu.:2015-03-10   
## Housewares : 2825 Denim & Co. : 1838 Max. :2015-03-31   
## (Other) : 6184 (Other) :41863   
## ONAIR\_START\_TMS ONAIR\_END\_TMS ONAIR\_MINS   
## 2015-01-01 00:19:21: 2 2015-01-01 00:19:20: 2 Min. : 0.020   
## 2015-01-01 06:41:32: 2 2015-01-01 06:41:31: 2 1st Qu.: 0.480   
## 2015-01-01 09:08:50: 2 2015-01-01 09:08:49: 2 Median : 1.100   
## 2015-01-01 13:00:00: 2 2015-01-01 12:59:59: 2 Mean : 3.794   
## 2015-01-01 14:49:12: 2 2015-01-01 13:25:44: 2 3rd Qu.: 6.220   
## 2015-01-01 16:00:00: 2 2015-01-01 13:58:20: 2 Max. :59.500   
## (Other) :63456 (Other) :63456   
## HOST1 HOST2 ONAIR\_MONTH   
## :30847 :60043 Length:63468   
## Nester Antonella: 1856 Faetsch Sharon : 524 Class :character   
## Stauffer Stacey : 1772 James-Dementri Pat: 488 Mode :character   
## Coffey Jennifer : 1736 Boesing Rachel : 476   
## Popaj Alberti : 1629 Hornback Nancy : 302   
## Stran Amy : 1627 Lindquist Kerstin : 275   
## (Other) :24001 (Other) : 1360

str(product\_airtime)

## 'data.frame': 63468 obs. of 11 variables:  
## $ PRODUCT\_NBR : Factor w/ 48780 levels "P150000001","P150000002",..: 8 8 8 8 9 9 9 9 701 701 ...  
## $ PRODUCT\_DESCRIPTION: Factor w/ 45420 levels "'Southern Charm' Bracelet by John Wind",..: 13703 13703 13703 13703 13999 13999 13999 13999 40537 40537 ...  
## $ PRODUCT\_CATEGORY : Factor w/ 19 levels "Accessories",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ BRAND\_NAME : Factor w/ 1885 levels "21 Day Fix","35 Degrees Below",..: 1177 1177 1177 1177 1177 1177 1177 1177 1623 1623 ...  
## $ ONAIR\_DATE : Date, format: "2015-01-01" "2015-01-01" ...  
## $ ONAIR\_START\_TMS : Factor w/ 62843 levels "2015-01-01 00:01:09",..: 122 117 44 32 130 120 55 38 343 341 ...  
## $ ONAIR\_END\_TMS : Factor w/ 62848 levels "2015-01-01 00:02:01",..: 124 116 45 31 129 119 58 37 342 341 ...  
## $ ONAIR\_MINS : num 9.17 0.62 9.13 0.6 5.38 0.85 5.38 0.83 0.05 6.12 ...  
## $ HOST1 : Factor w/ 27 levels "","Abiera Gabrielle",..: 1 1 23 23 1 1 23 23 1 1 ...  
## $ HOST2 : Factor w/ 20 levels "","Abiera Gabrielle",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ ONAIR\_MONTH : chr "January" "January" "January" "January" ...

#Merge product airtime with product master using product number  
product\_airtime<-merge(product\_master,product\_airtime,by="PRODUCT\_NBR")  
head(product\_airtime)

## PRODUCT\_NBR PRODUCT\_DESCRIPTION.x  
## 1 P150000002 Breezies Set of 6 Cotton Briefs with UltimAir  
## 2 P150000002 Breezies Set of 6 Cotton Briefs with UltimAir  
## 3 P150000002 Breezies Set of 6 Cotton Briefs with UltimAir  
## 4 P150000002 Breezies Set of 6 Cotton Briefs with UltimAir  
## 5 P150000002 Breezies Set of 6 Cotton Briefs with UltimAir  
## 6 P150000002 Breezies Set of 6 Cotton Briefs with UltimAir  
## PRODUCT\_CATEGORY.x BRAND\_NAME.x  
## 1 Accessories Not Known  
## 2 Accessories Not Known  
## 3 Accessories Not Known  
## 4 Accessories Not Known  
## 5 Accessories Not Known  
## 6 Accessories Not Known  
## PRODUCT\_DESCRIPTION.y PRODUCT\_CATEGORY.y  
## 1 Breezies Set of 6 Cotton Briefs with UltimAir Accessories  
## 2 Breezies Set of 6 Cotton Briefs with UltimAir Accessories  
## 3 Breezies Set of 6 Cotton Briefs with UltimAir Accessories  
## 4 Breezies Set of 6 Cotton Briefs with UltimAir Accessories  
## 5 Breezies Set of 6 Cotton Briefs with UltimAir Accessories  
## 6 Breezies Set of 6 Cotton Briefs with UltimAir Accessories  
## BRAND\_NAME.y ONAIR\_DATE ONAIR\_START\_TMS ONAIR\_END\_TMS  
## 1 Not Known 2015-03-28 2015-03-28 04:16:38 2015-03-28 04:17:06  
## 2 Not Known 2015-03-28 2015-03-28 04:01:12 2015-03-28 04:01:44  
## 3 Not Known 2015-03-28 2015-03-28 04:46:52 2015-03-28 04:47:18  
## 4 Not Known 2015-01-05 2015-01-05 12:01:38 2015-01-05 12:02:04  
## 5 Not Known 2015-01-16 2015-01-16 04:26:42 2015-01-16 04:34:13  
## 6 Not Known 2015-03-28 2015-03-28 07:01:12 2015-03-28 07:01:44  
## ONAIR\_MINS HOST1 HOST2 ONAIR\_MONTH  
## 1 0.48 Boesing Rachel March  
## 2 0.55 Boesing Rachel March  
## 3 0.45 Boesing Rachel March  
## 4 0.45 January  
## 5 7.53 Stauffer Stacey January  
## 6 0.55 March

#Changing the format of the date,sort it and create a new attribute month  
str(product\_airtime$ONAIR\_DATE)

## Date[1:63468], format: "2015-03-28" "2015-03-28" "2015-03-28" "2015-01-05" ...

product\_airtime$ONAIR\_DATE<-as.Date(product\_airtime$ONAIR\_DATE,format = "%Y-%m-%d")  
str(product\_airtime$ONAIR\_DATE)

## Date[1:63468], format: "2015-03-28" "2015-03-28" "2015-03-28" "2015-01-05" ...

product\_airtime<-product\_airtime[order(product\_airtime$ONAIR\_DATE),]  
product\_airtime$ONAIR\_MONTH<-months(product\_airtime$ONAIR\_DATE)  
  
# Subset product airtime into three groups based on months  
product\_airtime\_Jan<-subset(product\_airtime,product\_airtime$ONAIR\_MONTH=="January")  
product\_airtime\_Feb<-subset(product\_airtime,product\_airtime$ONAIR\_MONTH=="February")  
product\_airtime\_March<-subset(product\_airtime,product\_airtime$ONAIR\_MONTH=="March")