

Project_STA518

Cleyde Murillo, Jr.

10/13/2021

Loading in libraries

```
library(tidyverse)
library(data.table)
library(readr)
```

```
# Loading in Customer data
```

```
customers <- read_table2("marketing_campaign.csv")
```

```
## Warning: 'read_table2()' was deprecated in readr 2.0.0.
## Please use 'read_table()' instead.
```

```
## Warning: 221 parsing failures.
## row col expected actual file
## 11 -- 29 columns 28 columns 'marketing_campaign.csv'
## 20 -- 29 columns 30 columns 'marketing_campaign.csv'
## 28 -- 29 columns 28 columns 'marketing_campaign.csv'
## 38 -- 29 columns 30 columns 'marketing_campaign.csv'
## 44 -- 29 columns 28 columns 'marketing_campaign.csv'
## ... ..
## See problems(...) for more details.
```

```
summary(customers)
```

```
##           ID           Year_Birth      Education      Marital_Status
##  Min.      :    0      Min.      :1893      Length:2240      Length:2240
## 1st Qu.: 2828      1st Qu.:1959      Class :character      Class :character
## Median : 5458      Median :1970      Mode  :character      Mode  :character
## Mean      : 5592      Mean      :1969
## 3rd Qu.: 8428      3rd Qu.:1977
## Max.      :11191      Max.      :1996
##
##      Income      Kidhome      Teenhome      Dt_Customer
## Length:2240      Min.      :    0      Length:2240      Length:2240
## Class :character 1st Qu.:    0      Class :character      Class :character
## Mode  :character Median :    0      Mode  :character      Mode  :character
##                      Mean      : 4253
##                      3rd Qu.:    1
##                      Max.      :96547
```

```

##
##      Recency              MntWines              MntFruits              MntMeatProducts
## Length:2240             Min.   : 0.0             Min.   : 0.00             Min.   : 0.0
## Class :character        1st Qu.: 25.0             1st Qu.: 2.00            1st Qu.: 14.0
## Mode  :character        Median : 138.0            Median : 9.00             Median : 57.0
##                               Mean   : 288.5            Mean   : 42.39            Mean   : 156.8
##                               3rd Qu.: 476.5            3rd Qu.: 38.00           3rd Qu.: 211.5
##                               Max.    :1493.0           Max.    :1215.00          Max.    :1725.0
##
## MntFishProducts MntSweetProducts MntGoldProds NumDealsPurchases
## Min.   : 0.00     Min.   : 0.00     Min.   : 0.00     Min.   : 0.000
## 1st Qu.: 3.00     1st Qu.: 1.00     1st Qu.: 7.00     1st Qu.: 1.000
## Median : 13.00     Median : 9.00     Median : 22.00     Median : 2.000
## Mean   : 45.34     Mean   : 28.44     Mean   : 42.45     Mean   : 6.326
## 3rd Qu.: 55.00     3rd Qu.: 35.00     3rd Qu.: 54.00     3rd Qu.: 4.000
## Max.   :974.00     Max.   :362.00     Max.   :321.00     Max.   :246.000
##
## NumWebPurchases NumCatalogPurchases NumStorePurchases NumWebVisitsMonth
## Min.   : 0.000     Min.   : 0.000     Min.   : 0.000     Min.   : 0.000
## 1st Qu.: 2.000     1st Qu.: 1.000     1st Qu.: 3.000     1st Qu.: 3.000
## Median : 3.000     Median : 2.000     Median : 5.000     Median : 6.000
## Mean   : 3.929     Mean   : 2.817     Mean   : 5.503     Mean   : 5.278
## 3rd Qu.: 6.000     3rd Qu.: 4.000     3rd Qu.: 8.000     3rd Qu.: 7.000
## Max.   :27.000     Max.   :28.000     Max.   :13.000     Max.   :20.000
##
## AcceptedCmp3 AcceptedCmp4 AcceptedCmp5 AcceptedCmp1
## Min.   :0.00000     Min.   :0.00000     Min.   :0.00000     Min.   :0.00000
## 1st Qu.:0.00000     1st Qu.:0.00000     1st Qu.:0.00000     1st Qu.:0.00000
## Median :0.00000     Median :0.00000     Median :0.00000     Median :0.00000
## Mean   :0.5545     Mean   :0.07679     Mean   :0.07277     Mean   :0.06161
## 3rd Qu.:0.00000     3rd Qu.:0.00000     3rd Qu.:0.00000     3rd Qu.:0.00000
## Max.   :9.0000     Max.   :1.00000     Max.   :1.00000     Max.   :1.00000
##
## AcceptedCmp2 Complain Z_CostContact Z_Revenue
## Min.   :0.00000     Min.   :0.00000     Min.   : 0.000     Min.   : 0.00
## 1st Qu.:0.00000     1st Qu.:0.00000     1st Qu.: 3.000     1st Qu.:11.00
## Median :0.00000     Median :0.00000     Median : 3.000     Median :11.00
## Mean   :0.01875     Mean   :0.03661     Mean   : 2.809     Mean   :10.18
## 3rd Qu.:0.00000     3rd Qu.:0.00000     3rd Qu.: 3.000     3rd Qu.:11.00
## Max.   :1.00000     Max.   :3.00000     Max.   :11.000     Max.   :11.00
##
## Response
## Min.   : 0.000
## 1st Qu.: 0.000
## Median : 0.000
## Mean   : 1.132
## 3rd Qu.: 0.000
## Max.   :11.000
## NA's   :21

```

Variable Dictionary:

```
df2n <- customers %>%
  subset(!(Education %in% c('Basic', 'Graduation', 'Master', 'PhD')) | is.na(Response))

customers %>%
  filter()
```

```
## # A tibble: 2,240 x 29
##       ID Year_Birth Education Marital_Status Income Kidhome Teenhome Dt_Customer
##   <dbl>   <dbl> <chr>      <chr>      <chr>   <dbl> <chr>   <chr>
## 1  5524     1957 Graduat~ Single      58138     0 0     04-09-2012
## 2  2174     1954 Graduat~ Single      46344     1 1     08-03-2014
## 3  4141     1965 Graduat~ Together    71613     0 0     21-08-2013
## 4  6182     1984 Graduat~ Together    26646     1 0     10-02-2014
## 5  5324     1981 PhD      Married     58293     1 0     19-01-2014
## 6  7446     1967 Master   Together    62513     0 1     09-09-2013
## 7   965     1971 Graduat~ Divorced    55635     0 1     13-11-2012
## 8  6177     1985 PhD      Married     33454     1 0     08-05-2013
## 9  4855     1974 PhD      Together    30351     1 0     06-06-2013
## 10 5899     1950 PhD      Together     5648     1 1     13-03-2014
## # ... with 2,230 more rows, and 21 more variables: Recency <chr>,
## #   MntWines <dbl>, MntFruits <dbl>, MntMeatProducts <dbl>,
## #   MntFishProducts <dbl>, MntSweetProducts <dbl>, MntGoldProds <dbl>,
## #   NumDealsPurchases <dbl>, NumWebPurchases <dbl>, NumCatalogPurchases <dbl>,
## #   NumStorePurchases <dbl>, NumWebVisitsMonth <dbl>, AcceptedCmp3 <dbl>,
## #   AcceptedCmp4 <dbl>, AcceptedCmp5 <dbl>, AcceptedCmp1 <dbl>,
## #   AcceptedCmp2 <dbl>, Complain <dbl>, Z_CostContact <dbl>, Z_Revenue <dbl>,
## #   Response <dbl>
```