Supermart Sales Analysis

Chloe Shin

2025-09-13

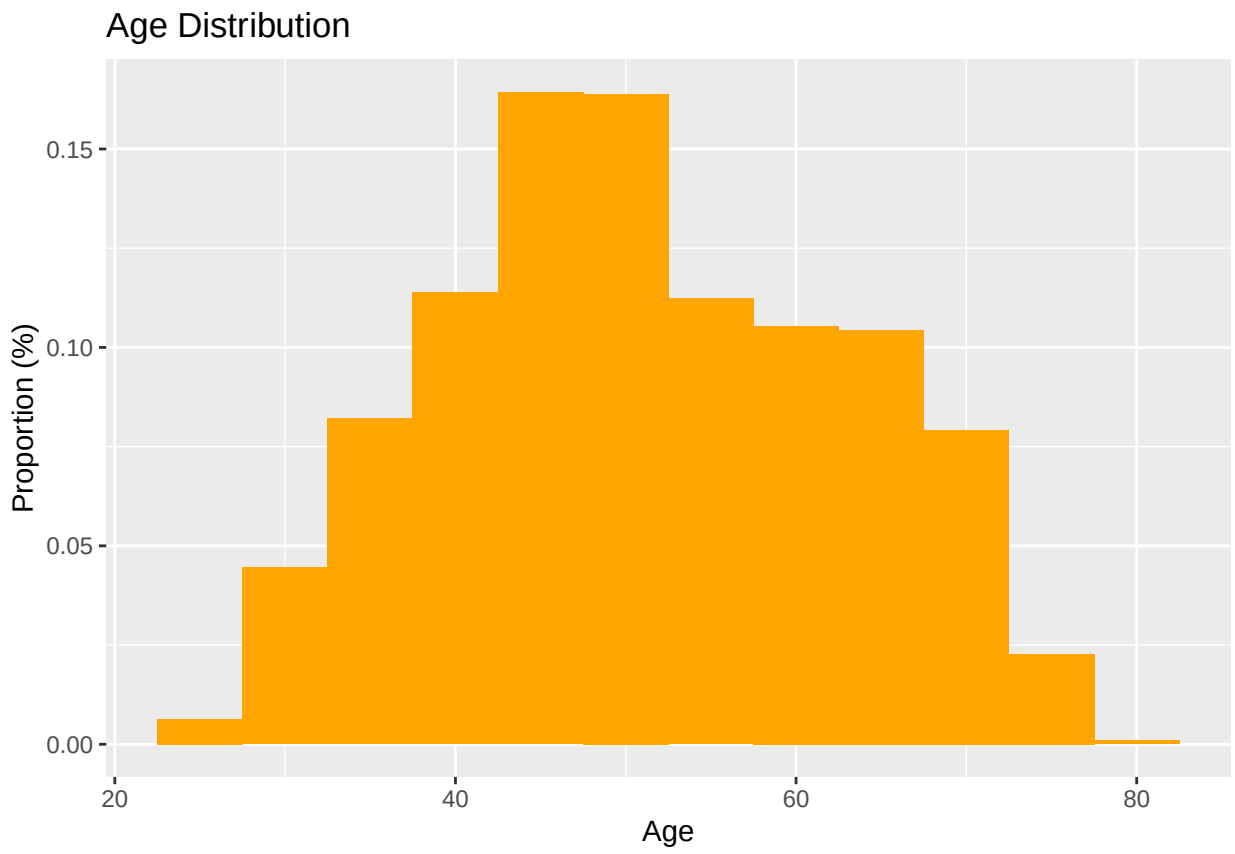
The dataset includes customer information of a supermarket that aims to predict the success of future marketing campaigns.

```
library(tidyverse)
library(tidymodels)
ifood <- read_csv("https://raw.githubusercontent.com/nailson/ifood-data-business-analyst-test/refs/heads
```

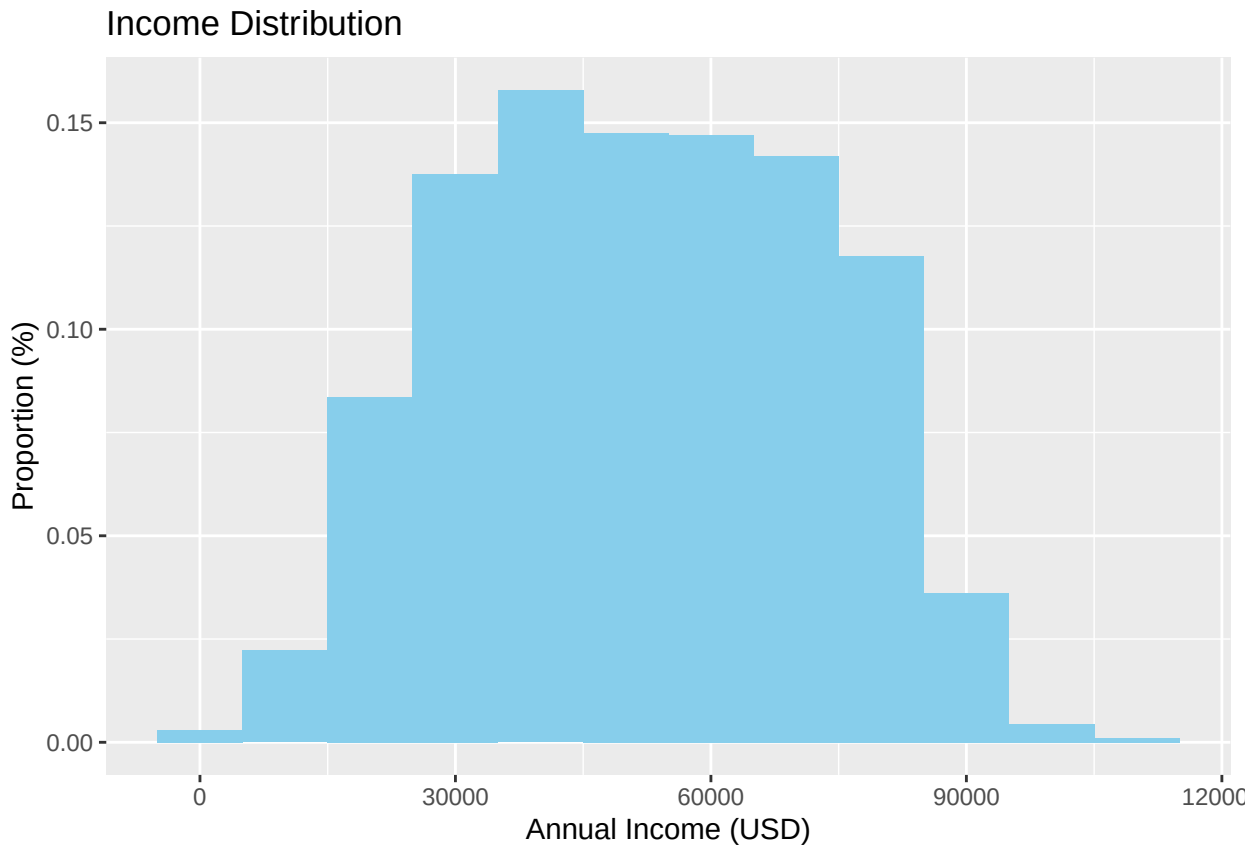
```
ifood <- ifood %>%
  distinct()
```

1. Customer profile

```
ifood %>%
  ggplot(aes(x = Age)) +
  geom_histogram(aes(y = after_stat(count) / sum(after_stat(count))),
                 binwidth = 5,
                 fill = 'orange') +
  labs(x = "Age",
       y = "Proportion (%)",
       title = "Age Distribution")
```




```
ifood %>%
  ggplot(aes(x = Income)) +
  geom_histogram(aes(y = after_stat(count) / sum(after_stat(count))),
                 binwidth = 10000,
                 fill = 'skyblue') +
  labs(x = "Annual Income (USD)",
       y = "Proportion (%)",
       title = "Income Distribution")
```

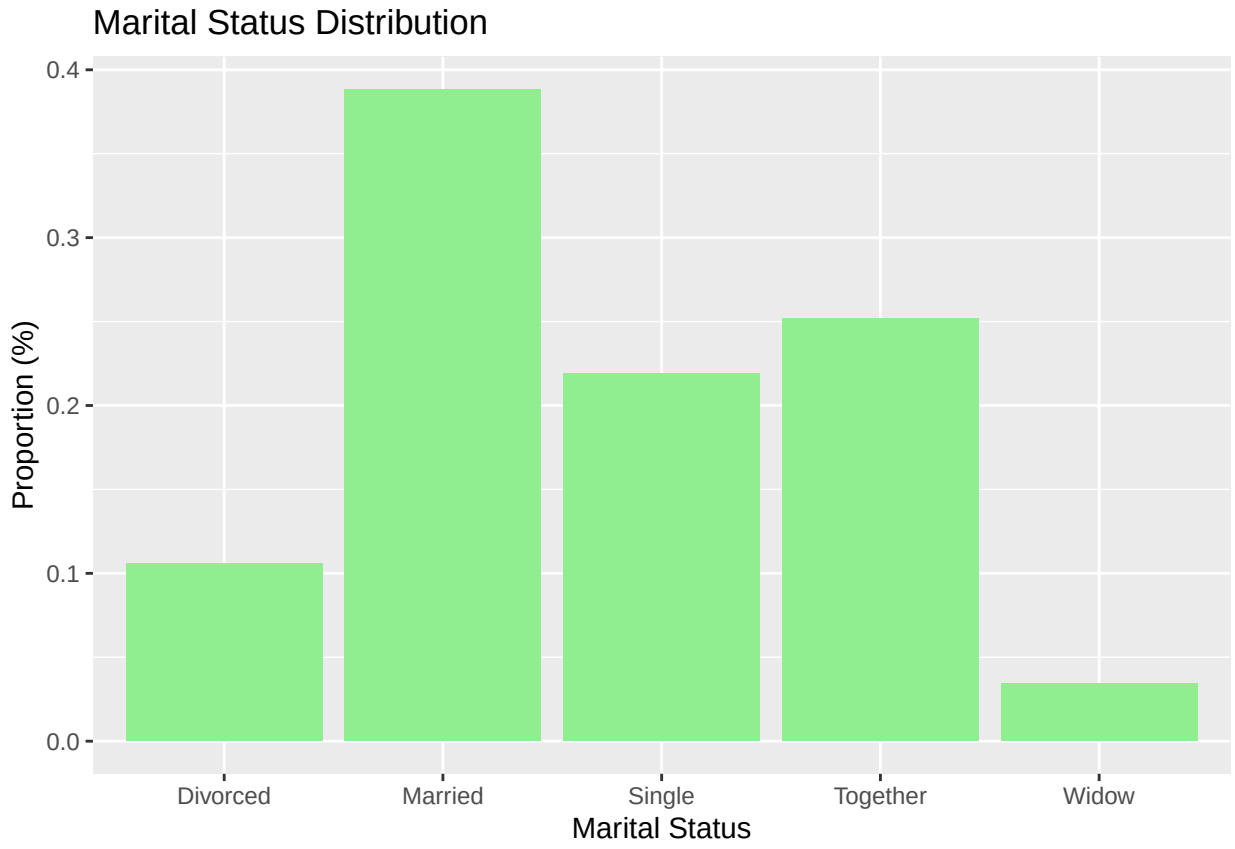



```
ifood_marriage <- ifood %>%
  pivot_longer(cols = c(marital_Widow,
                        marital_Divorced,
                        marital_Together,
                        marital_Single,
                        marital_Married),
               names_to = 'marital',
               values_to = 'value') %>%
  filter(value == 1)

ifood_marriage <- ifood_marriage %>%
  mutate (marital = recode(marital,
      marital_Divorced = "Divorced",
      marital_Married  = "Married",
      marital_Single   = "Single",
      marital_Together = "Together",
      marital_Widow    = "Widow")) %>%
  mutate(proportion = value / sum(value))

ifood_marriage %>%
  ggplot(aes(x = marital,
             y = proportion,
             fill = marital)) +
  geom_col(fill = 'lightgreen') +
  labs (x = 'Marital Status',
        y = 'Proportion (%)',
```

```
        title = 'Marital Status Distribution')
```

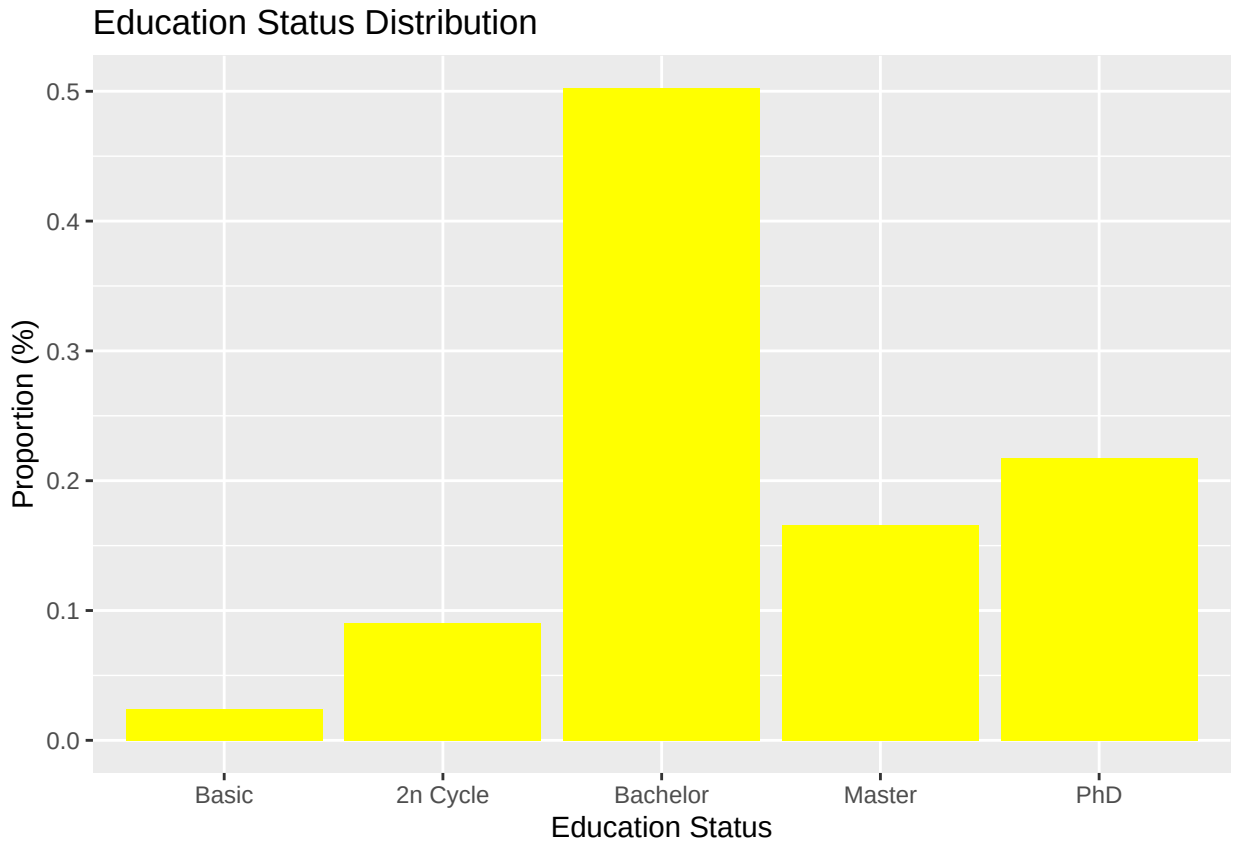



```
ifood_education <- ifood %>%
   pivot_longer(cols = c(`education_2n Cycle`,
                         education_Basic,
                         education_Graduation,
                         education_Master,
                         education_PhD ),
                names_to = 'education',
                values_to = 'value') %>%
  filter(value == 1)

ifood_education <- ifood_education %>%
  mutate (education = recode(education,
      `education_2n Cycle` = "2n Cycle",
      education_Basic  = "Basic",
      education_Graduation   = "Bachelor",
      education_Master = "Master",
      education_PhD    = "PhD")) %>%
      mutate(proportion = value / sum(value))

ifood_education %>%
  mutate(education = factor(education,
                            levels = c("Basic", "2n Cycle","Bachelor", "Master", "PhD"))) %>%
  ggplot(aes(x = education,
```

```
            y = proportion,
            fill = education)) +
  geom_col(fill = 'yellow') +
  labs (x = 'Education Status',
        y = 'Proportion (%)',
        title = 'Education Status Distribution')
```



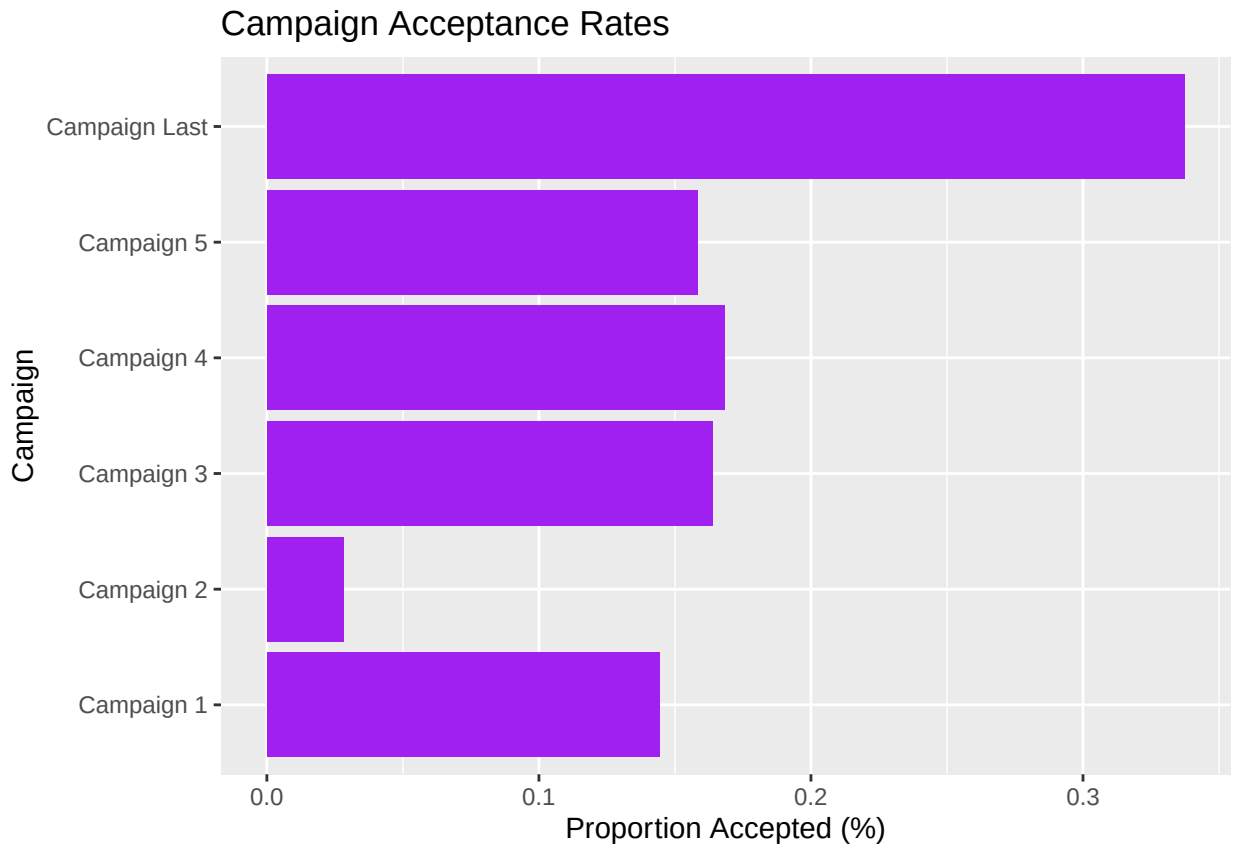
2. Assessing the success of previous marketing campaigns

```
ifood_campaign <- ifood[, c("AcceptedCmp1",
                            "AcceptedCmp2",
                            "AcceptedCmp3",
                            "AcceptedCmp4",
                            "AcceptedCmp5",
                            "Response")]

ifood_campaign <- ifood %>%
  summarise("Campaign 1" = sum(AcceptedCmp1),
            "Campaign 2" = sum(AcceptedCmp2),
            "Campaign 3" = sum(AcceptedCmp3),
            "Campaign 4" = sum(AcceptedCmp4),
            "Campaign 5" = sum(AcceptedCmp5),
            "Campaign Last" = sum(Response)) %>%
```

```
    pivot_longer(cols = everything(),
                 names_to = "campaign",
                 values_to = "value") %>%
    mutate(proportion = value / sum(value))

ifood_campaign %>%
  ggplot(aes(x = campaign,
             y = proportion)) +
  geom_col(fill = "purple") +
  coord_flip() +
  labs(
    x = "Campaign",
    y = "Proportion Accepted (%)",
    title = "Campaign Acceptance Rates"
  )
```



3. Campaign engagements

```
ifood_acceptcmp <- ifood %>%
  select(AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5, Response) %>%
  mutate(numaccept = rowSums(select(., AcceptedCmp1:AcceptedCmp5)),
         prevaccept = numaccept > 0)
```

```r
prop.table(table(ifood_acceptcmp$prevaccept, ifood_acceptcmp$Response), 1)
```

```
##
##                   0          1
##   FALSE 0.91410658 0.08589342
##   TRUE  0.59154930 0.40845070
```

```r
chisq.test(table(ifood_acceptcmp$prevaccept, ifood_acceptcmp$Response))
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(ifood_acceptcmp$prevaccept, ifood_acceptcmp$Response)
## X-squared = 266.18, df = 1, p-value < 2.2e-16
```

Among the customers who did not accept any of the previous campaigns, only 8.35% accepted the most recent campagin. On the other hand, among the customers who accepted at least one of the previous campaigns, 40.8% accepted the most recent campaign.
Previous engagement with campaigns significantly increases the likelihoood of responding to the most recent campaign.

4. Target demographic for the next campaign

```r
marital_response <- ifood_marriage %>%
  group_by(marital) %>%
  summarise(
    customers = n(),
    acceptance = mean(Response))

print(marital_response)
```

```
## # A tibble: 5 x 3
##   marital  customers acceptance
##   <chr>        <int>      <dbl>
## 1 Divorced       214      0.201
## 2 Married        785      0.117
## 3 Single         443      0.233
## 4 Together       509      0.108
## 5 Widow           70      0.257
```

```r
chisq.test(table(ifood_marriage$marital, ifood_marriage$Response))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(ifood_marriage$marital, ifood_marriage$Response)
## X-squared = 46.727, df = 4, p-value = 1.738e-09
```

```r
education_response <- ifood_education %>%
  group_by(education) %>%
  summarise(
    customers = n(),
    acceptance = mean(Response))

print(education_response)
```

```
## # A tibble: 5 x 3
##   education customers acceptance
##   <chr>         <int>      <dbl>
## 1 2n Cycle        183     0.120
## 2 Bachelor       1015     0.140
## 3 Basic            49     0.0408
## 4 Master          335     0.158
## 5 PhD             439     0.210
```

```r
chisq.test(table(ifood_education$education, ifood_education$Response))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(ifood_education$education, ifood_education$Response)
## X-squared = 18.43, df = 4, p-value = 0.001017
```

```r
ifood_predictors <- ifood %>%
  mutate(childnum = Kidhome + Teenhome)

model <- lm(Response ~ Income + Age + MntTotal + childnum + AcceptedCmpOverall, ifood_predictors)
summary(model)
```

```
##
## Call:
## lm(formula = Response ~ Income + Age + MntTotal + childnum +
##     AcceptedCmpOverall, data = ifood_predictors)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95613 -0.11669 -0.08023 -0.04483  0.97791
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.805e-01  3.710e-02   4.864 1.24e-06 ***
## Income             -2.110e-06  6.237e-07  -3.383  0.00073 ***
## Age                -4.582e-04  6.472e-04  -0.708  0.47910
## MntTotal            1.065e-04  2.449e-05   4.349 1.43e-05 ***
## childnum           -1.632e-02  1.144e-02  -1.427  0.15373
## AcceptedCmpOverall  2.032e-01  1.199e-02  16.951  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.3249 on 2015 degrees of freedom
## Multiple R-squared:  0.1917, Adjusted R-squared:  0.1897
## F-statistic: 95.58 on 5 and 2015 DF,  p-value: < 2.2e-16
```

Statistically significant predictors

i. Marital status: Customers living alone (i.e., divorced, single, widowed)

ii. Educational status: Customers with higher levels of education (i.e., PhD)

iii. Income: Customers with lower income

iv. Amount of purchase: Customers who spent greater amount of money on the supermarket's products

v. Number of accepted campaigns: Customers who accepted greater number of past campaigns

… are more likely to accept the most recent campaign