Advertisement Analysis — Week 13 IP

1. Define the question

1.1 Main question — which individuals are more likely to click on cryptpgraphy course advertisement?

1.1.1 Other Research Questions:

1.2 Metric for Success

1.3 Understanding the context

A Kenyan entrepreneur, who has previously ran a course advertisement on her blog would like to know: which kinds of individuals are more likely to click on her advertisements posted on the blog. From the initial advertisement, the entrepreneur was able to collect some data and would like a Data Science Consultant to analyze the given data and report recommendations. The findings from the analysis will be crucial in determining which kinds of individuals to target before running another online course ad on the blog.

1.4 Experimental Design

The approach for this project will include:

1. Importing the necessary libraries
2. Reading the data
3. Checking the data
4. Tidying up the data
5. Implementing the solution using Univariate and Bivariate
6. Conclusion and Recommendations

1.5 Data Relevance

I think that the data was relevant, however, I also think that we could have used more columns to add onto our analysis.

2. Importing libraries

```
library(tidyverse)
```

```
## -- Attaching packages ------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ---------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(ggcorrplot)
library(caret)
```

```
## Loading required package: lattice
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:purrr':
##
##     lift
```

```
#install.packages("kernlab")
library(kernlab)
```

```
##
## Attaching package: 'kernlab'
```

```
## The following object is masked from 'package:purrr':
##
##     cross
```

```
## The following object is masked from 'package:ggplot2':
##
##     alpha
```

```
#install.packages("e1071")
library(e1071)
# #install.packages("klaR")
# library(klaR)
```

3. Reading the data

```
# Imported data from the directory where it was saved
advertising <- read.csv("~/Moringa School/R Programming/R datasets/advertising.csv")
view(advertising)
```

4. Checking the data

```
# let's get glimpse of how the data looks like
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <int> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <int> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, ...
```

Our data has 1,000 records and 10 columns. Some columns have integer, float and character data types. clicked.on.ad is our target variable. The target variable and male columns are listed as integers, but they are both categorical. We will need to convert these columns data types to the appropriate data types during cleaning.

```
#checking the top of the data
head(advertising)
```

```
##   Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage
## 1                    68.95  35    61833.90               256.09
## 2                    80.23  31    68441.85               193.77
## 3                    69.47  26    59785.94               236.50
## 4                    74.15  29    54806.18               245.89
## 5                    68.37  35    73889.99               225.58
## 6                    59.99  23    59761.56               226.74
##                          Ad.Topic.Line           City Male    Country
## 1    Cloned 5thgeneration orchestration    Wrightburgh    0    Tunisia
## 2    Monitored national standardization      West Jodi    1      Nauru
## 3      Organic bottom-line service-desk       Davidton    0 San Marino
## 4 Triple-buffered reciprocal time-frame West Terrifurt    1      Italy
## 5         Robust logistical utilization   South Manuel    0    Iceland
## 6       Sharable client-driven software      Jamieberg    1     Norway
##             Timestamp Clicked.on.Ad
## 1 2016-03-27 00:53:11             0
## 2 2016-04-04 01:39:02             0
## 3 2016-03-13 20:35:42             0
## 4 2016-01-10 02:31:19             0
## 5 2016-06-03 03:36:18             0
## 6 2016-05-19 14:30:17             0
```

```
#checking the bottom of the data
tail(advertising)
```

```
##      Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage
## 995                     43.70  28    63126.96               173.01
## 996                     72.97  30    71384.57               208.58
## 997                     51.30  45    67782.17               134.42
## 998                     51.63  51    42415.72               120.37
## 999                     55.55  19    41920.79               187.95
## 1000                    45.01  26    29875.80               178.35
##                            Ad.Topic.Line          City Male
## 995         Front-line bifurcated ability  Nicholasland    0
## 996         Fundamental modular algorithm     Duffystad    1
## 997       Grass-roots cohesive monitoring   New Darlene    1
## 998          Expanded intangible solution South Jessica    1
## 999  Proactive bandwidth-monitored policy   West Steven    0
## 1000      Virtual 5thgeneration emulation   Ronniemouth    0
##                      Country           Timestamp Clicked.on.Ad
## 995                  Mayotte 2016-04-04 03:57:48             1
## 996                  Lebanon 2016-02-11 21:49:00             1
## 997  Bosnia and Herzegovina 2016-04-22 02:07:01             1
## 998                 Mongolia 2016-02-01 17:24:57             1
## 999                Guatemala 2016-03-24 02:35:54             0
## 1000                  Brazil 2016-06-03 21:43:21             1
```

After looking at the head and tail of the data, there are a few assumptions on the data we need to make

before we proceed. These assumptions include:

1. Daily time spend on site column is in minutes.
2. Income is in United States Dollars($).
3. Daily internet usage units is in megabytes. Also assuming that the megabytes refers to data used on the blog site.
4. For male column: if 0 that means not male, but if 1 then yes it's a male.
5. Clicked on ad column: if 1 that means yes somebody clicked on the ad, but if 0 then the add was not clicked.

```
# checking for unique elements in columns if interest
# starting with our target variable
unique(advertising$Clicked.on.Ad)
```

```
## [1] 0 1
```

There are two elements 0 for not clicked and 1 for clicked(based on the assumption we made above)

```
unique(advertising$Male)
```

```
## [1] 0 1
```

There is two elements in the male column. 0 for not a male and 1 for male

```
#unique(advertising$Country)
```

In total there is a total of 237 countries. From the countries list, I see funny country names e.g Nauru, Wallis & Futuna, Holy see. Where did these countries come from? Note: country output not added here because the list is too long

5. Tidying the data

5.1 Fixing data types

```
# converting male column to character
advertising$Male <- as.character(advertising$Male)
```

```
# converting clicked on ad column to character
advertising$Clicked.on.Ad <- as.character(advertising$Clicked.on.Ad)
```

```
# confirming that the data types have been converted
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <chr> "0", "1", "0", "1", "0", "1", "0", "1", "1...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

Male and Clicked on ad have been converted to their appropriate data types

5.2 Missing values

```
# checking for missing values in each column
colSums(is.na(advertising))
```

```
## Daily.Time.Spent.on.Site                      Age             Area.Income
##                        0                        0                        0
##     Daily.Internet.Usage            Ad.Topic.Line                     City
##                        0                        0                        0
##                     Male                  Country                Timestamp
##                        0                        0                        0
##            Clicked.on.Ad
##                        0
```

There are no missing values in our dataset

5.3 Duplicates

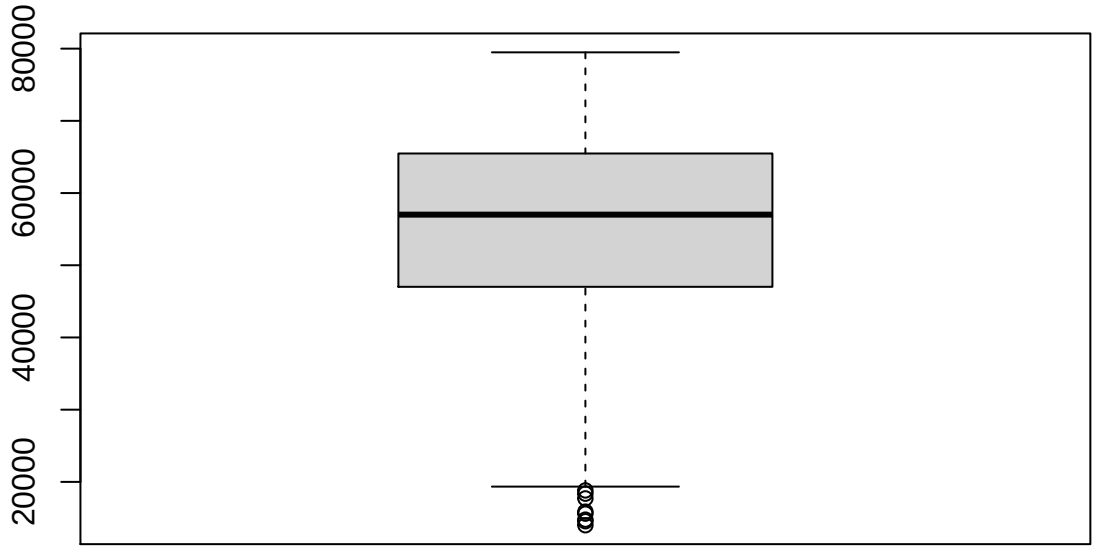
```
# checking for duplicates
advert_duplicates <- advertising[duplicated(advertising),]
advert_duplicates
```

```
##  [1] Daily.Time.Spent.on.Site Age                      Area.Income
##  [4] Daily.Internet.Usage     Ad.Topic.Line            City
##  [7] Male                     Country                  Timestamp
## [10] Clicked.on.Ad
## <0 rows> (or 0-length row.names)
```

There are no duplicates in our dataset

5.4 Outliers

```
# checking for outliers in numerical variables using boxplots
# we will separate area income from other numerical variables because of the difference in scale
boxplot(advertising$Area.Income)
```



We see some outliers from Income column on the lower side of the whisker. We won't be deleting these outliers because we establish that these are true observations. The low income could be as a result of many reasons. An explanation for these observations is that — in every country there are certain areas that have a lot of poor people, which leads to low area income.

```
advertising %>%
  class()
```

```
## [1] "data.frame"
```

```
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <chr> "0", "1", "0", "1", "0", "1", "0", "1", "1...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

```
num_cols <- select(advertising, Daily.Time.Spent.on.Site, Age, Daily.Internet.Usage)
num_cols
```

```
##      Daily.Time.Spent.on.Site Age Daily.Internet.Usage
```

```
## 1                68.95  35              256.09
## 2                80.23  31              193.77
## 3                69.47  26              236.50
## 4                74.15  29              245.89
## 5                68.37  35              225.58
## 6                59.99  23              226.74
## 7                88.91  33              208.36
## 8                66.00  48              131.76
## 9                74.53  30              221.51
## 10               69.88  20              183.82
## 11               47.64  49              122.02
## 12               83.07  37              230.87
## 13               69.57  48              113.12
## 14               79.52  24              214.23
## 15               42.95  33              143.56
## 16               63.45  23              140.64
## 17               55.39  37              129.41
## 18               82.03  41              187.53
## 19               54.70  36              118.39
## 20               74.58  40              135.51
## 21               77.22  30              224.44
## 22               84.59  35              226.54
## 23               41.49  52              164.83
## 24               87.29  36              209.93
## 25               41.39  41              167.22
## 26               78.74  28              204.79
## 27               48.53  28              134.14
## 28               51.95  52              129.23
## 29               70.20  34              119.20
## 30               76.02  22              209.82
## 31               67.64  35              267.01
## 32               86.41  28              207.48
## 33               59.05  57              169.23
## 34               55.60  23              212.58
## 35               57.64  57              133.81
## 36               84.37  30              201.58
## 37               62.26  53              125.45
## 38               65.82  39              221.94
## 39               50.43  46              119.32
## 40               38.93  39              162.08
## 41               84.98  29              202.61
## 42               64.24  30              252.36
## 43               82.52  32              198.11
## 44               81.38  31              212.30
## 45               80.47  25              204.86
## 46               37.68  52              172.83
## 47               69.62  20              202.25
## 48               85.40  43              198.72
## 49               44.33  37              123.72
## 50               48.01  46              119.93
## 51               73.18  23              196.71
## 52               79.94  28              225.29
## 53               33.33  45              193.58
## 54               50.33  50              133.20
```

```
## 55                62.31  47              119.30
## 56                80.60  31              177.55
## 57                65.19  36              150.61
## 58                44.98  49              129.31
## 59                77.63  29              239.22
## 60                41.82  41              156.36
## 61                85.61  27              183.43
## 62                85.84  34              192.93
## 63                72.08  29              169.50
## 64                86.06  32              178.92
## 65                45.96  45              141.22
## 66                62.42  29              198.50
## 67                63.89  40              105.22
## 68                35.33  32              200.22
## 69                75.74  25              215.25
## 70                78.53  34              131.72
## 71                46.13  31              139.01
## 72                69.01  46              222.63
## 73                55.35  39              153.17
## 74                33.21  43              167.07
## 75                38.46  42              145.98
## 76                64.10  22              215.93
## 77                49.81  35              120.06
## 78                82.73  33              238.99
## 79                56.14  38              113.53
## 80                55.13  45              111.71
## 81                78.11  27              209.25
## 82                73.46  28              222.75
## 83                56.64  38              115.91
## 84                68.94  54              138.71
## 85                70.79  31              184.10
## 86                57.76  41              105.15
## 87                77.51  36              200.55
## 88                52.70  34              118.60
## 89                57.70  34              109.07
## 90                56.89  37              109.29
## 91                69.90  43              138.35
## 92                55.79  24              149.67
## 93                70.03  26              227.72
## 94                50.08  40              125.85
## 95                43.67  31              166.29
## 96                72.84  26              238.63
## 97                45.72  36              154.02
## 98                39.94  41              156.30
## 99                35.61  46              158.22
## 100               79.71  34              211.65
## 101               41.49  53              169.18
## 102               63.60  23              235.28
## 103               89.91  40              194.23
## 104               68.18  21              218.17
## 105               66.49  20              202.16
## 106               80.49  40              229.12
## 107               72.23  25              241.03
## 108               42.39  42              150.99
```

```
## 109                47.53  30              135.18
## 110                74.02  32              210.54
## 111                66.63  60              176.98
## 112                63.24  53              235.78
## 113                71.00  22              211.87
## 114                46.13  46              123.64
## 115                69.00  32              221.21
## 116                76.99  31              244.34
## 117                72.60  55              162.95
## 118                61.88  42              112.19
## 119                84.45  50              207.18
## 120                88.97  45              152.49
## 121                86.19  31              210.26
## 122                49.58  26              231.94
## 123                77.65  27              212.79
## 124                37.75  36              225.24
## 125                62.33  43              127.11
## 126                79.57  31              230.93
## 127                80.31  44              127.07
## 128                89.05  45              206.98
## 129                70.41  27              223.03
## 130                67.36  37              233.56
## 131                46.98  50              175.37
## 132                41.67  36              132.55
## 133                51.24  36              176.73
## 134                75.70  29              215.44
## 135                43.49  47              127.83
## 136                49.89  39              160.03
## 137                38.37  36              140.46
## 138                38.52  38              137.28
## 139                71.89  23              172.81
## 140                75.80  38              146.19
## 141                83.86  31              190.25
## 142                37.51  30              163.00
## 143                55.60  44              124.38
## 144                83.67  44              234.26
## 145                69.08  41              210.60
## 146                37.47  44              141.89
## 147                56.04  49              128.95
## 148                70.92  41              108.16
## 149                49.78  46              152.24
## 150                68.61  57              150.29
## 151                58.18  25              176.28
## 152                78.54  35              172.10
## 153                37.00  48              158.22
## 154                65.40  33              247.31
## 155                79.52  27              183.48
## 156                87.98  38              222.11
## 157                44.64  36              127.01
## 158                41.73  28              202.18
## 159                80.46  27              207.96
## 160                75.55  36              159.24
## 161                76.32  35              195.31
## 162                82.68  33              222.77
```

```
## 163                72.01  31              251.00
## 164                75.83  24              162.44
## 165                41.28  50              140.39
## 166                34.66  32              194.83
## 167                66.18  55              143.42
## 168                86.06  31              219.72
## 169                59.59  42              104.78
## 170                86.69  34              198.56
## 171                43.77  52              138.55
## 172                71.84  47              199.79
## 173                80.23  31              196.23
## 174                74.41  26              163.05
## 175                63.36  48              137.43
## 176                71.74  35              227.56
## 177                60.72  44              105.69
## 178                72.04  22              199.43
## 179                44.57  31              133.17
## 180                85.86  34              208.23
## 181                39.85  38              145.96
## 182                84.53  27              168.34
## 183                62.95  60              157.04
## 184                67.58  41              255.61
## 185                85.56  29              210.46
## 186                46.88  54              136.64
## 187                46.31  57              153.98
## 188                77.95  31              233.65
## 189                84.73  30              153.76
## 190                39.86  36              145.85
## 191                50.08  30              123.91
## 192                60.23  35              106.86
## 193                60.70  49              110.57
## 194                43.67  53              143.79
## 195                77.20  33              254.05
## 196                71.86  32              116.53
## 197                44.78  45              137.24
## 198                78.57  36              239.32
## 199                73.41  31              201.26
## 200                77.05  27              191.14
## 201                66.40  40              214.42
## 202                69.35  29              252.77
## 203                35.65  40              172.58
## 204                70.04  31              183.85
## 205                69.78  29              218.79
## 206                58.22  29              120.90
## 207                76.90  28              212.67
## 208                84.08  30              187.36
## 209                59.51  58              140.83
## 210                40.15  38              134.88
## 211                76.81  28              217.85
## 212                41.89  38              163.38
## 213                76.87  27              235.35
## 214                67.28  43              155.80
## 215                81.98  40              229.22
## 216                66.01  23              151.95
```

```
## 217                61.57  53              125.94
## 218                53.30  34              111.94
## 219                34.87  40              200.23
## 220                43.60  38              170.49
## 221                77.88  37              254.57
## 222                75.83  27              200.59
## 223                49.95  39              136.59
## 224                60.94  41              154.97
## 225                89.15  42              171.07
## 226                78.70  30              133.99
## 227                57.35  29              119.84
## 228                34.86  38              154.75
## 229                70.68  31              199.08
## 230                76.06  23              201.04
## 231                66.67  33              228.03
## 232                46.77  32              136.40
## 233                62.42  38              143.94
## 234                78.32  28              239.52
## 235                37.32  50              199.25
## 236                40.42  45              133.90
## 237                76.77  36              123.51
## 238                65.65  30              158.05
## 239                74.32  33              128.17
## 240                73.27  32              234.75
## 241                80.03  44              150.84
## 242                53.68  47              115.26
## 243                85.84  32              192.85
## 244                85.03  30              204.52
## 245                70.44  24              178.75
## 246                81.22  53              223.09
## 247                39.96  45              146.13
## 248                57.05  41              269.96
## 249                42.44  56              168.27
## 250                62.20  25              161.16
## 251                76.70  36              222.25
## 252                61.22  45              119.03
## 253                84.54  33              204.02
## 254                46.08  30              164.63
## 255                56.70  48              123.13
## 256                81.03  28              201.15
## 257                80.91  32              231.42
## 258                40.06  38              138.68
## 259                83.47  39              226.11
## 260                73.84  31              121.05
## 261                74.65  28              212.56
## 262                60.25  35              109.77
## 263                59.21  35              144.62
## 264                43.02  44              125.22
## 265                84.04  38              244.55
## 266                70.66  43              120.95
## 267                70.58  26              136.94
## 268                72.44  34              230.14
## 269                40.17  26              171.31
## 270                79.15  26              203.23
```

```
## 271                44.49  53              168.00
## 272                73.04  37              221.79
## 273                76.28  33              254.34
## 274                68.88  37              179.58
## 275                73.10  28              242.37
## 276                47.66  29              156.54
## 277                87.30  35              216.87
## 278                89.34  32              177.78
## 279                81.37  26              156.48
## 280                81.67  28              196.76
## 281                46.37  52              144.27
## 282                54.88  24              148.61
## 283                40.67  35              133.18
## 284                71.76  35              237.39
## 285                47.51  51              130.41
## 286                75.15  22              212.87
## 287                56.01  26              127.26
## 288                82.87  37              213.36
## 289                45.05  42              141.36
## 290                60.53  24              167.22
## 291                50.52  31              171.62
## 292                84.71  32              210.23
## 293                55.20  39              159.46
## 294                81.61  33              228.76
## 295                71.55  36              163.99
## 296                82.40  36              218.97
## 297                73.95  35              238.58
## 298                72.07  31              226.45
## 299                80.39  31              214.74
## 300                65.80  25              231.49
## 301                69.97  28              250.00
## 302                52.62  50              176.52
## 303                39.25  39              152.36
## 304                77.56  38              130.83
## 305                33.52  43              165.56
## 306                79.81  24              178.85
## 307                84.79  33              214.53
## 308                82.70  35              231.07
## 309                84.88  32              186.48
## 310                54.92  54              161.16
## 311                76.56  34              221.53
## 312                69.74  49              243.37
## 313                75.55  22              169.40
## 314                72.19  33              250.35
## 315                84.29  41              232.54
## 316                73.89  39              110.68
## 317                75.84  21              186.98
## 318                73.38  25              236.19
## 319                80.72  31              186.37
## 320                62.06  44              105.00
## 321                51.50  34              135.31
## 322                90.97  37              180.77
## 323                86.78  30              170.13
## 324                66.18  35              243.61
```

```
## 325  84.33  41  240.95
## 326  36.87  36  195.91
## 327  34.78  48  208.21
## 328  76.84  32  231.59
## 329  67.05  25  220.92
## 330  41.47  31  219.79
## 331  80.71  26  200.58
## 332  80.09  31  214.08
## 333  56.30  49  135.24
## 334  79.36  34  245.78
## 335  86.38  40  188.27
## 336  38.94  41  142.67
## 337  87.26  35  184.03
## 338  75.32  28  233.60
## 339  74.38  40  220.05
## 340  65.90  22  211.39
## 341  36.31  47  168.92
## 342  72.23  48  115.35
## 343  88.12  38  230.91
## 344  83.97  28  205.50
## 345  61.09  26  131.68
## 346  65.77  21  218.61
## 347  81.58  25  199.39
## 348  37.87  52  188.56
## 349  76.20  37  178.51
## 350  60.91  19  184.94
## 351  74.49  28  237.34
## 352  73.71  23  211.38
## 353  78.19  30  228.81
## 354  79.54  44  217.68
## 355  74.87  52  126.97
## 356  87.09  36  221.98
## 357  37.45  47  167.86
## 358  49.84  39  111.59
## 359  51.38  59  158.56
## 360  83.40  34  207.87
## 361  38.91  33  150.80
## 362  62.14  41  110.93
## 363  79.72  28  193.80
## 364  73.30  36  135.72
## 365  69.11  42  231.48
## 366  71.90  54  140.15
## 367  72.45  29  195.36
## 368  77.07  40  261.02
## 369  74.62  36  217.79
## 370  82.07  25  205.38
## 371  58.60  50  113.70
## 372  36.08  45  151.47
## 373  79.44  26  206.79
## 374  41.73  47  144.71
## 375  73.19  25  203.74
## 376  77.60  24  197.33
## 377  89.00  37  222.26
## 378  69.20  42  123.80
```

```
## 379  67.56  31  125.45
## 380  81.11  39  248.19
## 381  80.22  30  224.58
## 382  43.63  41  123.25
## 383  77.66  29  168.15
## 384  74.63  26  235.99
## 385  49.67  27  153.69
## 386  80.59  37  224.23
## 387  83.49  33  190.75
## 388  44.46  42  132.66
## 389  68.10  40  227.73
## 390  63.88  38  136.85
## 391  78.83  36  234.64
## 392  79.97  44  216.00
## 393  80.51  28  200.28
## 394  62.26  26  202.77
## 395  66.99  47  124.44
## 396  71.05  20  204.22
## 397  42.05  51  174.55
## 398  50.52  28  219.69
## 399  76.24  40  198.32
## 400  77.29  27  201.24
## 401  35.98  47  165.52
## 402  84.95  34  230.36
## 403  39.34  43  148.93
## 404  87.23  29  202.12
## 405  57.24  52  117.35
## 406  81.58  41  248.16
## 407  56.34  50  139.02
## 408  48.73  27  142.04
## 409  51.68  49  258.62
## 410  35.34  45  152.86
## 411  48.09  33  180.42
## 412  78.68  29  208.05
## 413  68.82  20  205.64
## 414  56.99  40  108.15
## 415  86.63  39  209.64
## 416  41.18  43  129.25
## 417  71.03  32  120.85
## 418  72.92  29  217.10
## 419  77.14  24  184.88
## 420  60.70  43  192.60
## 421  34.30  41  160.74
## 422  83.71  45  220.48
## 423  53.38  35  120.06
## 424  58.03  31  129.33
## 425  43.59  36  132.31
## 426  60.07  42  120.75
## 427  54.43  37  154.74
## 428  81.99  33  230.90
## 429  60.53  29  123.28
## 430  84.69  31  231.85
## 431  88.72  32  211.87
## 432  88.89  35  218.80
```

```
## 433                69.58  43              255.07
## 434                85.23  36              212.92
## 435                83.55  39              221.18
## 436                56.66  42              139.42
## 437                56.39  27              248.12
## 438                76.24  27              214.42
## 439                57.64  36              110.25
## 440                78.18  23              167.67
## 441                46.04  32              147.92
## 442                79.40  35              236.87
## 443                36.44  39              147.64
## 444                53.14  38              109.00
## 445                32.84  40              171.72
## 446                73.72  32              256.40
## 447                38.10  34              214.38
## 448                73.93  44              218.22
## 449                51.87  50              119.65
## 450                77.69  22              169.88
## 451                43.41  28              160.73
## 452                55.92  24              145.08
## 453                80.67  34              239.76
## 454                83.42  25              183.42
## 455                82.12  52              201.15
## 456                66.17  33              238.45
## 457                43.01  35              127.37
## 458                80.05  25              219.94
## 459                64.88  42              129.80
## 460                79.82  26              223.28
## 461                48.03  40              134.60
## 462                32.99  45              177.46
## 463                74.88  27              175.17
## 464                36.49  52              196.61
## 465                88.04  45              191.17
## 466                45.70  33              151.12
## 467                82.38  35              159.60
## 468                52.68  23              149.20
## 469                65.59  47              121.81
## 470                65.65  25              224.92
## 471                43.84  36              167.42
## 472                67.69  37              216.57
## 473                78.37  24              207.27
## 474                81.46  29              231.54
## 475                47.48  31              141.34
## 476                75.15  33              219.49
## 477                78.76  24              219.98
## 478                44.96  50              132.71
## 479                39.56  41              143.13
## 480                39.76  28              196.83
## 481                57.11  22              207.17
## 482                83.26  40              187.76
## 483                69.42  25              213.38
## 484                50.60  30              129.88
## 485                46.20  37              119.30
## 486                66.88  35              119.47
```

```
## 487                83.97  40              158.42
## 488                76.56  30              213.75
## 489                35.49  48              159.77
## 490                80.29  31              244.87
## 491                50.19  40              117.30
## 492                59.12  33              124.54
## 493                59.88  30              193.63
## 494                59.70  28              120.25
## 495                67.80  30              117.75
## 496                81.59  35              223.16
## 497                81.10  29              216.49
## 498                41.70  39              126.95
## 499                73.94  27              173.49
## 500                58.35  37              132.63
## 501                51.56  46              124.85
## 502                79.81  37              253.17
## 503                66.17  26              228.70
## 504                58.21  37              105.94
## 505                66.12  49              113.80
## 506                80.47  42              215.18
## 507                77.05  31              236.64
## 508                49.99  41              121.07
## 509                80.30  58              173.43
## 510                79.36  33              234.72
## 511                57.86  30              166.86
## 512                70.29  26              231.37
## 513                84.53  33              215.18
## 514                59.13  44              106.04
## 515                81.51  41              250.03
## 516                42.94  37              130.40
## 517                84.81  32              233.93
## 518                82.79  34              132.08
## 519                59.22  55              126.39
## 520                35.00  40              151.25
## 521                46.61  42              136.18
## 522                63.26  29              120.46
## 523                79.16  32              202.90
## 524                67.94  43              128.16
## 525                79.91  32              230.18
## 526                66.14  41              165.27
## 527                43.65  39              138.87
## 528                59.61  21              198.45
## 529                46.61  52              156.99
## 530                89.37  34              162.03
## 531                65.10  49              118.10
## 532                53.44  42              108.17
## 533                79.53  51              244.91
## 534                91.43  39              209.91
## 535                73.57  30              212.38
## 536                78.76  32              208.02
## 537                76.49  23              181.11
## 538                61.72  26              218.49
## 539                84.53  35              236.29
## 540                72.03  34              230.95
```

```
## 541                77.47  36              222.91
## 542                75.65  39              247.90
## 543                78.15  33              194.37
## 544                63.80  38              108.70
## 545                76.59  29              211.64
## 546                42.60  55              168.29
## 547                78.77  28              211.83
## 548                83.40  39              235.01
## 549                79.53  33              236.72
## 550                73.89  35              229.99
## 551                75.80  36              224.90
## 552                81.95  31              208.76
## 553                56.39  58              154.23
## 554                44.73  35              127.56
## 555                38.35  33              145.48
## 556                72.53  37              223.93
## 557                56.20  49              114.85
## 558                79.67  28              226.79
## 559                75.42  26              164.25
## 560                78.64  31              235.28
## 561                67.69  44              109.22
## 562                38.35  41              144.69
## 563                59.52  44              251.08
## 564                62.26  37              166.19
## 565                64.75  36              117.66
## 566                79.97  26              185.45
## 567                47.90  42              114.53
## 568                80.38  30              238.06
## 569                64.51  42              190.71
## 570                71.28  37              246.72
## 571                50.32  40              125.65
## 572                72.76  33              240.63
## 573                72.80  35              249.54
## 574                74.59  23              158.35
## 575                46.66  45              118.16
## 576                48.86  54              134.46
## 577                37.05  39              142.81
## 578                81.21  36              233.04
## 579                66.89  23              208.24
## 580                68.11  38              231.21
## 581                69.15  46              112.72
## 582                65.72  36              120.12
## 583                40.04  27              161.58
## 584                68.60  33              135.08
## 585                56.16  25              164.25
## 586                78.60  46              254.59
## 587                78.29  38              252.07
## 588                43.83  45              129.01
## 589                77.31  32              238.10
## 590                39.86  28              161.24
## 591                66.77  25              141.13
## 592                57.20  42              110.66
## 593                73.15  25              211.12
## 594                82.07  24              193.97
```

```
## 595                49.84  38              135.24
## 596                43.97  36              156.97
## 597                77.25  27              231.38
## 598                74.84  37              246.44
## 599                83.53  36              204.56
## 600                38.63  48              222.11
## 601                84.00  48              136.21
## 602                52.13  50              118.27
## 603                71.83  40              135.48
## 604                78.36  24              196.77
## 605                50.18  35              127.82
## 606                64.67  51              138.35
## 607                69.50  26              203.84
## 608                65.22  30              240.09
## 609                62.06  40              116.27
## 610                84.29  30              160.33
## 611                32.91  37              181.02
## 612                39.50  31              148.19
## 613                75.19  31              245.76
## 614                76.21  31              228.94
## 615                67.76  31              242.59
## 616                40.01  53              161.77
## 617                52.70  41              109.34
## 618                68.41  38              259.76
## 619                35.55  39              151.18
## 620                74.54  24              219.75
## 621                81.75  24              190.08
## 622                87.85  31              210.27
## 623                60.23  60              151.54
## 624                87.97  35              149.25
## 625                78.17  27              192.27
## 626                67.91  23              146.80
## 627                85.77  27              191.78
## 628                41.16  49              150.83
## 629                53.54  39              108.03
## 630                73.94  26              236.15
## 631                63.43  29              236.75
## 632                84.59  36              241.80
## 633                70.13  31              224.98
## 634                40.19  37              136.99
## 635                58.95  55              131.29
## 636                35.76  51              195.07
## 637                59.36  49              110.84
## 638                91.10  40              198.13
## 639                61.04  41              149.21
## 640                74.06  23              225.99
## 641                64.63  45              158.80
## 642                81.29  28              219.72
## 643                76.07  36              235.56
## 644                75.92  22              182.65
## 645                78.35  46              253.48
## 646                46.14  28              137.97
## 647                44.33  41              120.63
## 648                46.43  28              137.20
```

```
## 649                66.04  27              199.76
## 650                84.31  29              225.87
## 651                83.66  38              175.14
## 652                81.25  33              222.35
## 653                85.26  32              224.07
## 654                86.53  46              233.36
## 655                76.44  26              224.20
## 656                52.84  43              122.31
## 657                85.24  31              182.84
## 658                74.71  46              258.06
## 659                82.95  39              201.29
## 660                76.42  26              223.16
## 661                42.04  49              182.11
## 662                46.28  26              228.78
## 663                48.26  50              122.45
## 664                71.03  55              150.77
## 665                81.37  33              215.04
## 666                58.05  32              195.54
## 667                75.00  29              230.36
## 668                79.61  31              235.97
## 669                52.56  31              250.36
## 670                62.18  33              126.44
## 671                77.89  26              201.54
## 672                66.08  61              184.23
## 673                89.21  33              210.53
## 674                49.96  55              151.94
## 675                77.44  28              210.39
## 676                82.58  38              225.23
## 677                39.36  29              161.79
## 678                47.23  38              149.80
## 679                87.85  34              153.01
## 680                65.57  46              130.86
## 681                78.01  26              200.71
## 682                44.15  28              141.96
## 683                43.57  36              125.20
## 684                76.83  28              192.81
## 685                42.06  34              131.55
## 686                76.27  27              226.69
## 687                74.27  37              247.05
## 688                73.27  28              216.24
## 689                74.58  36              230.52
## 690                77.50  28              225.34
## 691                87.16  33              197.15
## 692                87.16  37              231.95
## 693                66.26  47              179.04
## 694                65.15  29              117.30
## 695                68.25  33              198.86
## 696                73.49  38              244.23
## 697                39.19  54              173.05
## 698                80.15  25              214.49
## 699                86.76  28              189.91
## 700                73.88  29              233.61
## 701                58.60  19              197.93
## 702                69.77  54              132.27
```

```
## 703                87.27  30              204.27
## 704                77.65  28              208.01
## 705                76.02  40              219.55
## 706                78.84  26              217.66
## 707                71.33  23              169.40
## 708                81.90  41              225.47
## 709                46.89  48              176.78
## 710                77.80  57              152.94
## 711                45.44  43              119.27
## 712                69.96  31              214.06
## 713                87.35  35              158.29
## 714                49.42  53              128.00
## 715                71.27  21              216.03
## 716                49.19  38              123.08
## 717                39.96  35              138.52
## 718                85.01  29              192.50
## 719                68.95  51              185.85
## 720                67.59  45              113.69
## 721                75.71  34              246.06
## 722                43.07  36              137.63
## 723                39.47  43              163.48
## 724                48.22  40              214.33
## 725                76.76  25              230.77
## 726                78.74  27              234.75
## 727                67.47  24              225.05
## 728                81.17  30              231.91
## 729                89.66  34              171.23
## 730                79.60  28              227.37
## 731                65.53  19              190.17
## 732                61.87  35              250.20
## 733                83.16  41              194.95
## 734                44.11  41              121.24
## 735                56.57  26              131.98
## 736                83.91  29              222.87
## 737                79.80  28              229.88
## 738                71.23  52              122.59
## 739                47.23  43              210.87
## 740                82.37  30              207.44
## 741                43.63  38              135.25
## 742                70.90  28              190.95
## 743                71.90  29              193.29
## 744                62.12  37              105.86
## 745                67.35  29              118.69
## 746                57.99  50              124.58
## 747                66.80  29              248.51
## 748                49.13  32              120.49
## 749                45.11  58              195.69
## 750                54.35  42              164.02
## 751                61.82  59              151.93
## 752                77.75  31              240.64
## 753                70.61  28              190.12
## 754                82.72  31              179.82
## 755                76.87  36              212.59
## 756                65.07  34              227.53
```

```
## 757                56.93  37              111.80
## 758                48.86  35              128.37
## 759                36.56  29              195.89
## 760                85.73  32              147.75
## 761                75.81  40              229.19
## 762                72.94  31              190.84
## 763                53.63  54              126.29
## 764                52.35  25              147.61
## 765                52.84  51              121.57
## 766                51.58  33              115.91
## 767                42.32  29              187.09
## 768                55.04  42              106.96
## 769                68.58  41              171.54
## 770                85.54  27              175.43
## 771                71.14  30              224.82
## 772                64.38  19              180.47
## 773                88.85  40              213.96
## 774                66.79  60              198.30
## 775                32.60  45              185.47
## 776                43.88  54              166.85
## 777                56.46  26              151.63
## 778                72.18  30              225.02
## 779                52.67  44              191.26
## 780                80.55  35              219.91
## 781                67.85  41              202.70
## 782                75.55  36              123.71
## 783                80.46  29              230.78
## 784                82.69  29              167.41
## 785                35.21  39              154.00
## 786                36.37  40              144.53
## 787                74.07  22              165.43
## 788                59.96  33              197.66
## 789                85.62  29              195.68
## 790                40.88  33              136.18
## 791                36.98  31              167.87
## 792                35.49  47              170.04
## 793                56.56  26              204.47
## 794                36.62  32              162.44
## 795                49.35  49              119.86
## 796                75.64  29              204.82
## 797                79.22  27              198.79
## 798                77.05  34              236.08
## 799                66.83  46              196.17
## 800                76.20  24              228.81
## 801                56.64  29              123.24
## 802                53.33  34              111.63
## 803                50.63  50              142.23
## 804                41.84  49              139.32
## 805                53.92  41              125.46
## 806                83.89  28              180.88
## 807                55.32  43              127.65
## 808                53.22  44              108.85
## 809                43.16  35              156.11
## 810                67.51  43              127.20
```

```
## 811  43.16  29  143.04
## 812  79.89  30  241.38
## 813  84.25  32  170.90
## 814  74.18  28  203.87
## 815  85.78  34  232.78
## 816  80.96  39  225.00
## 817  36.91  48  159.69
## 818  54.47  23  141.52
## 819  81.98  34  212.88
## 820  79.60  39  194.23
## 821  57.51  38  105.71
## 822  82.30  31  232.21
## 823  73.21  30  252.60
## 824  79.09  32  209.72
## 825  68.47  28  226.64
## 826  83.69  36  192.57
## 827  83.48  31  222.72
## 828  43.49  45  124.67
## 829  66.69  35  108.27
## 830  48.46  49  132.38
## 831  42.51  30  144.77
## 832  42.83  34  132.38
## 833  41.46  42  128.98
## 834  45.99  33  124.61
## 835  68.72  27  225.97
## 836  63.11  34  254.94
## 837  49.21  46  115.60
## 838  55.77  49  117.33
## 839  44.13  40  128.48
## 840  57.82  46  107.56
## 841  72.46  40  113.53
## 842  61.88  45  108.18
## 843  78.24  23  199.29
## 844  74.61  38  231.28
## 845  89.18  37  224.01
## 846  44.16  42  133.42
## 847  55.74  37  124.34
## 848  88.82  36  169.10
## 849  70.39  32  261.52
## 850  59.05  52  118.45
## 851  78.58  33  250.11
## 852  35.11  35  158.03
## 853  60.39  45  108.25
## 854  81.56  26  213.70
## 855  75.03  34  255.57
## 856  50.87  24  190.41
## 857  82.80  30  223.20
## 858  78.51  25  205.71
## 859  37.65  51  161.29
## 860  83.17  43  244.40
## 861  91.37  45  182.65
## 862  68.25  29  220.08
## 863  81.32  25  165.65
## 864  76.64  39  241.50
```

```
## 865                74.06  50              246.29
## 866                39.53  33              142.21
## 867                86.58  32              195.93
## 868                90.75  40              216.50
## 869                67.71  25              225.76
## 870                82.41  36              222.08
## 871                45.82  27              171.24
## 872                76.79  27              235.94
## 873                70.05  33              203.44
## 874                72.19  32              250.32
## 875                77.35  34              167.26
## 876                40.34  29              173.75
## 877                67.39  44              107.19
## 878                68.68  34              187.03
## 879                81.75  43              249.45
## 880                66.03  22              217.37
## 881                47.74  33              154.93
## 882                79.18  31              236.96
## 883                86.81  29              199.62
## 884                41.53  42              158.81
## 885                70.92  39              249.81
## 886                46.84  45              123.22
## 887                44.40  53              140.95
## 888                52.17  44              115.37
## 889                81.45  31              205.84
## 890                54.08  36              111.02
## 891                76.65  31              238.43
## 892                54.39  20              171.90
## 893                37.74  40              190.95
## 894                69.86  25              241.36
## 895                85.37  36              194.56
## 896                80.99  26              207.53
## 897                78.84  32              235.29
## 898                77.36  41              115.79
## 899                55.46  37              108.10
## 900                35.66  45              151.72
## 901                50.78  51              122.04
## 902                40.47  38              203.90
## 903                45.62  43              121.28
## 904                84.76  30              178.69
## 905                80.64  26              221.59
## 906                75.94  27              236.96
## 907                37.01  50              216.01
## 908                87.18  31              193.60
## 909                56.91  50              146.44
## 910                75.24  24              226.49
## 911                42.84  52              182.20
## 912                67.56  47              109.98
## 913                34.96  42              160.49
## 914                87.46  37              211.56
## 915                41.86  39              128.62
## 916                34.04  34              174.88
## 917                54.96  42              113.75
## 918                87.14  31              199.40
```

```
## 919                78.79  32              215.29
## 920                65.56  25              181.25
## 921                81.05  34              245.50
## 922                55.71  37              112.52
## 923                45.48  49              129.16
## 924                47.00  56              149.53
## 925                59.64  51              153.12
## 926                35.98  45              150.79
## 927                72.55  22              202.34
## 928                91.15  38              184.98
## 929                80.53  29              187.64
## 930                82.49  45              130.84
## 931                80.94  36              239.94
## 932                61.76  34              114.69
## 933                63.30  38              116.19
## 934                36.73  34              149.79
## 935                78.41  33              248.23
## 936                83.98  36              194.62
## 937                63.18  45              107.92
## 938                50.60  48              135.67
## 939                32.60  38              190.05
## 940                60.83  19              185.46
## 941                44.72  46              123.86
## 942                78.76  51              162.05
## 943                79.51  39              125.11
## 944                39.30  32              145.73
## 945                64.79  30              116.07
## 946                89.80  36              198.24
## 947                72.82  34              191.82
## 948                38.65  31              154.77
## 949                59.01  30              178.75
## 950                78.96  50              193.15
## 951                63.99  43              138.46
## 952                41.35  27              162.46
## 953                62.79  36              231.87
## 954                45.53  29              141.58
## 955                51.65  31              249.99
## 956                54.55  44              109.04
## 957                35.66  36              172.57
## 958                69.95  28              247.01
## 959                79.83  29              234.23
## 960                85.35  37              161.42
## 961                56.78  28              124.32
## 962                78.67  26              195.56
## 963                70.09  21              211.17
## 964                60.75  42              247.05
## 965                65.07  24              233.85
## 966                35.25  50              194.44
## 967                37.58  52              176.70
## 968                68.01  25              188.32
## 969                45.08  38              125.27
## 970                63.04  27              159.05
## 971                40.18  29              151.96
## 972                45.17  48              132.07
```

```
## 973                    50.48  50              162.43
## 974                    80.87  28              203.30
## 975                    41.88  40              126.11
## 976                    39.87  48              139.34
## 977                    61.84  45              105.63
## 978                    54.97  31              116.38
## 979                    71.40  30              166.31
## 980                    70.29  31              254.65
## 981                    67.26  57              168.41
## 982                    76.58  46              258.26
## 983                    54.37  38              140.77
## 984                    82.79  32              234.81
## 985                    66.47  31              256.39
## 986                    72.88  44              125.12
## 987                    76.44  28              232.68
## 988                    63.37  43              105.04
## 989                    89.71  48              204.40
## 990                    70.96  31              256.40
## 991                    35.79  44              165.62
## 992                    38.96  38              140.67
## 993                    69.17  40              123.62
## 994                    64.20  27              227.63
## 995                    43.70  28              173.01
## 996                    72.97  30              208.58
## 997                    51.30  45              134.42
## 998                    51.63  51              120.37
## 999                    55.55  19              187.95
## 1000                   45.01  26              178.35
```

There are no observed outliers in these columns. Based on the analysis, so far, our data is pretty clean and we are ready to begin solution implementation using Univariate and Bivariate analysis.

6. Implementing the solution using Univariate and Bivariate

6.1 Univariate analysis

```
#checking for statistical summaries of numerical data. This will allow us to get a summary of: mean
#median, min, max, range, and quantiles
#first select numerical variables
num <- select(advertising, Daily.Time.Spent.on.Site, Age, Area.Income,Daily.Internet.Usage)
glimpse(num)
```

6.1.1 Measures of central tendency and measure of dispersion for numerical variables

```
## Rows: 1,000
## Columns: 4
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
```

```
summary(num)
```

```
##  Daily.Time.Spent.on.Site      Age         Area.Income    Daily.Internet.Usage
##  Min.   :32.60            Min.   :19.00   Min.   :13996   Min.   :104.8
##  1st Qu.:51.36            1st Qu.:29.00   1st Qu.:47032   1st Qu.:138.8
```

```
##  Median :68.22           Median :35.00   Median :57012   Median :183.1
##  Mean   :65.00           Mean   :36.01   Mean   :55000   Mean   :180.0
##  3rd Qu.:78.55           3rd Qu.:42.00   3rd Qu.:65471   3rd Qu.:218.8
##  Max.   :91.43           Max.   :61.00   Max.   :79485   Max.   :270.0
```

On average daily time spent on the blog is 65 minutes.

Average age that visit the site is about 36 years old, min age = 19 years and max 61 years.

Average area income is $55,000.

The average daily internet usage on the site is 180 megabytes.

From the statistical summary, we don't see anything strange with the numerical columns.

```
# Variance
# Daily.Time.Spent.on.Site column
# sapply(num, var)
```

```
# standard deviation
# sapply(num, sd)
```

From the standard deviations we see the spreads of the numerical data from their means.

```
# hist(num$Daily.Time.Spent.on.Site, xlab= "Daily time spent on site(min)", ylab="Frequency",
#      main="Histogram of daily time spent on site")
```

6.1.2 Histograms From the graph, we see that the most popular times spent on the site is between 65-85 minutes. The distribution looks like a multi-nomial distribution with left skewness.

```
# hist(num$Age, xlab= "Age distribution(years)", ylab="Frequency",
#      main="Histogram of age distribution")
```

Majority of the people visiting the entrepreneur's blog are between the age of 25-45 years old. We don't know how many of them clicked on the ad, but it's likely that since they are the majority that they will have more clicks. We will check on this during bivariate analysis. The data is almost a normal distribution with a bit of skewness to the right.

```
# hist(num$Area.Income, xlab= "Area Income($)", ylab="Frequency",
#      main="Histogram of Area Income distribution")
```

The plot shows an obvious left skewness. This is not unusual. Earlier from the boxplot, we saw outliers on the lower side of the whisker. From the plot people who visited the blog are from high income areas($50,000 to 75,000). This makes sense because with high income people are able to easily obtain power and internet needed to visit the blog. Unlike in the low income areas, people can not afford internet as they have to deal with more pressing needs.

```
# hist(num$Daily.Internet.Usage, xlab= "Daily.Internet.Usage(megabytes)", ylab="Frequency",
#      main="Histogram of daily internet usage")
```

We observe a bimodal distribution here. We have a group of people who majorly use between 125 to about 140 megabytes on the blog, while another group that majorly use between 200 to about 240 megabytes on the blog.

```
advertising %>%
    group_by(Clicked.on.Ad) %>%
    summarize(frequency = n())
```

6.1.3 Univariate analysis of categorical variables using frequency distribution table

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 2
##   Clicked.on.Ad frequency
##   <chr>             <int>
## 1 0                   500
## 2 1                   500
```

The distribution of the number of people who clicked on adds and the ones that did not click on ad is equal. This is great because it means that our data is balanced.

```
advertising %>%
    group_by(Male) %>%
    summarize(frequency = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 2
##   Male  frequency
##   <chr>     <int>
## 1 0           519
## 2 1           481
```

We have majority non males visiting the blog.

```
advertising %>%
    group_by(Country) %>%
    summarize(frequency = n()) %>%
    ungroup() %>%
    arrange(desc(frequency)) %>%
    head(20)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 20 x 2
##    Country                frequency
##    <chr>                      <int>
##  1 Czech Republic                 9
##  2 France                         9
##  3 Afghanistan                    8
##  4 Australia                      8
##  5 Cyprus                         8
##  6 Greece                         8
##  7 Liberia                        8
##  8 Micronesia                     8
##  9 Peru                           8
## 10 Senegal                        8
## 11 South Africa                   8
## 12 Turkey                         8
## 13 Albania                        7
## 14 Bahamas                        7
```

```
## 15 Bosnia and Herzegovina         7
## 16 Burundi                        7
## 17 Cambodia                       7
## 18 Eritrea                        7
## 19 Ethiopia                       7
## 20 Fiji                           7
```

Here we have top 20 most popular countries where the most individuals visit the blog.

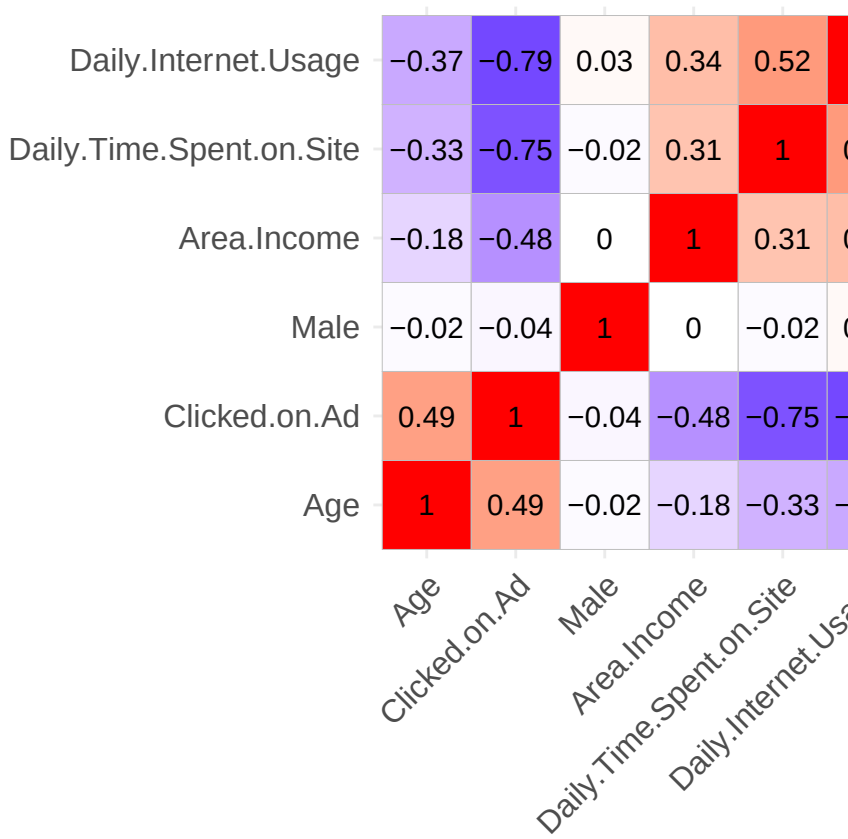
6.2 Bivariate Analysis

```
advertising$Male <- as.integer(advertising$Male)
advertising$Clicked.on.Ad <- as.integer(advertising$Clicked.on.Ad)
num_var <- select(advertising, Daily.Time.Spent.on.Site, Age, Area.Income, Daily.Internet.Usage, Male, C
#cor(advertising)
cor(num_var, method = "pearson")
```

6.2.1 Correlation calculation using pearson method

```
##                           Daily.Time.Spent.on.Site         Age  Area.Income
## Daily.Time.Spent.on.Site               1.00000000 -0.33151334  0.310954413
## Age                                   -0.33151334  1.00000000 -0.182604955
## Area.Income                            0.31095441 -0.18260496  1.000000000
## Daily.Internet.Usage                   0.51865848 -0.36720856  0.337495533
## Male                                  -0.01895085 -0.02104406  0.001322359
## Clicked.on.Ad                         -0.74811656  0.49253127 -0.476254628
##                           Daily.Internet.Usage         Male Clicked.on.Ad
## Daily.Time.Spent.on.Site           0.51865848 -0.018950855   -0.74811656
## Age                               -0.36720856 -0.021044064    0.49253127
## Area.Income                        0.33749553  0.001322359   -0.47625463
## Daily.Internet.Usage               1.00000000  0.028012326   -0.78653918
## Male                               0.02801233  1.000000000   -0.03802747
## Clicked.on.Ad                     -0.78653918 -0.038027466    1.00000000
```

```
advertising %>%
    select_if(is.numeric) %>%
    cor %>%
    ggcorrplot(lab = TRUE, hc.order = TRUE)
```



6.2.2 Correlation matrix visualization

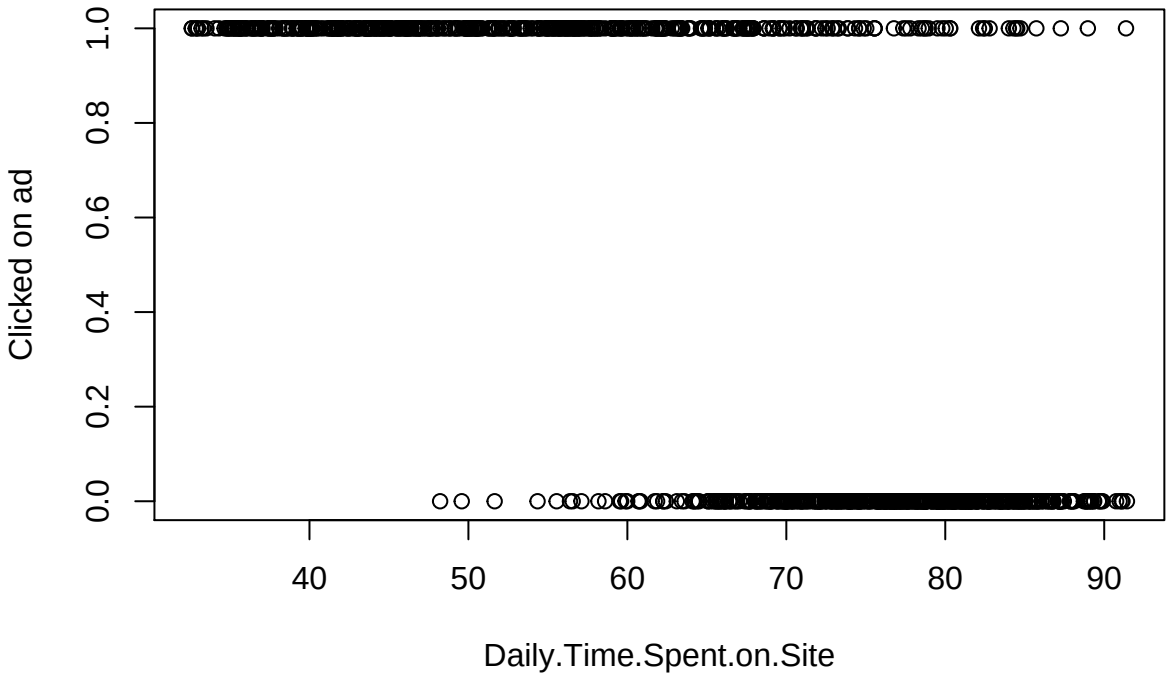
Here we get to compare correlations between the target variable (clicked on ad) with other variables. Observations:

1. The target variable has a positive moderate correlation to age (corr coefficient of 0.49).
2. The target variable is weakly correlated to male column. This means that there is a very weak relationship between a person being male or not and clicking the ad on entrepreneur's blog.
3. The target variable is moderately negatively correlated to Area Income. Meaning that the target variable are usually moving in opposite direction. Which can be interpreted as: the lower income areas have high click rate than low income areas.
4. The target variable is strongly negatively correlated to daily time spent on site. Here we see a strong relationship among the two variable where the two variable are moving in the opposite direction. This could also mean that individuals spending less time on the blog are more likely to click on the ad.
5. We see similar relationship with daily internet usage and target variable as we did with daily time spent on the site. This could also mean that individuals with less internet usage are likely to click on the ads.

6.2.3 Scatter plots

1. Clicked on ad with Daily.Time.Spent.on.Site scatter plot.

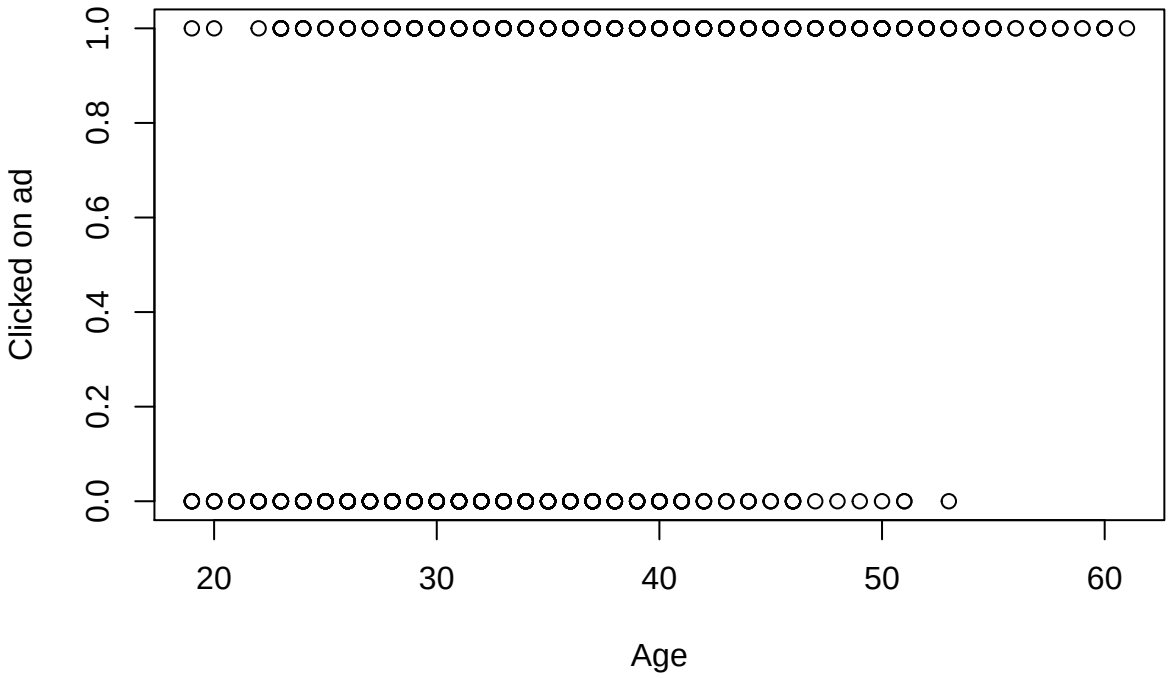
```
plot(advertising$Daily.Time.Spent.on.Site, advertising$Clicked.on.Ad, xlab="Daily.Time.Spent.on.Site", y
```



From the scatter plot, we don't see any relationship between the target variable and daily time spent on site column.

2. Clicked on ad with Age scatter plot

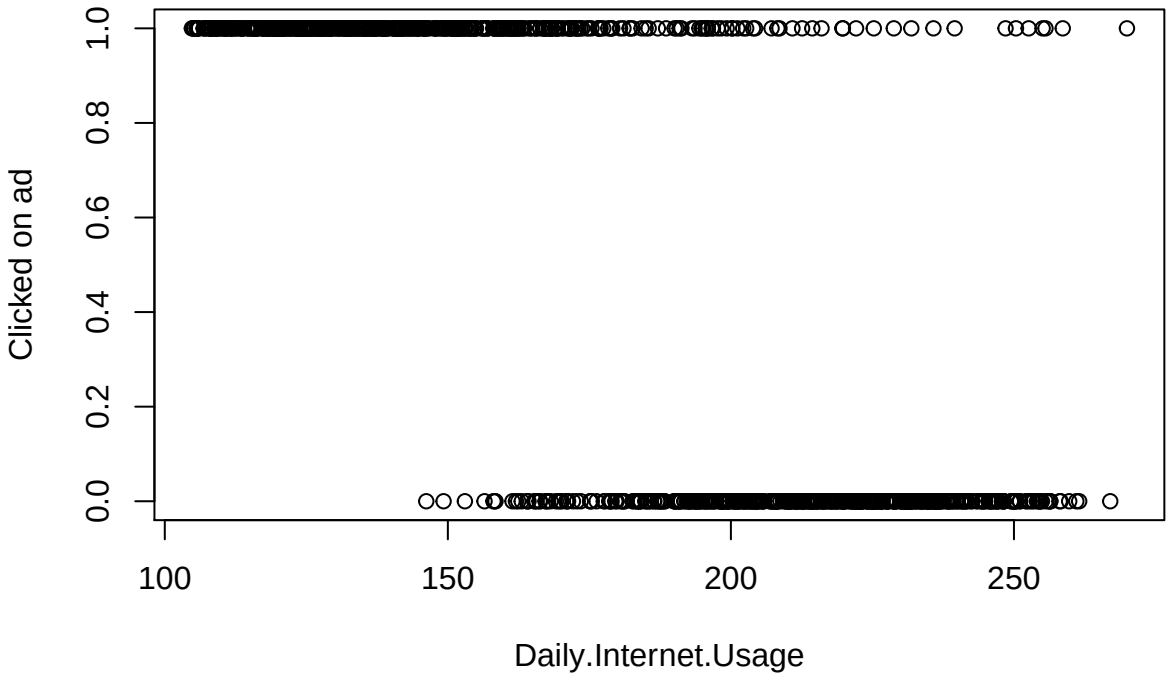
```
plot(advertising$Age, advertising$Clicked.on.Ad, xlab="Age", ylab="Clicked on ad")
```



From the scatter plot, we dont see any relationship between the target variable and age column

3. Clicked on ad with Daily.Internet.Usage scatter plot

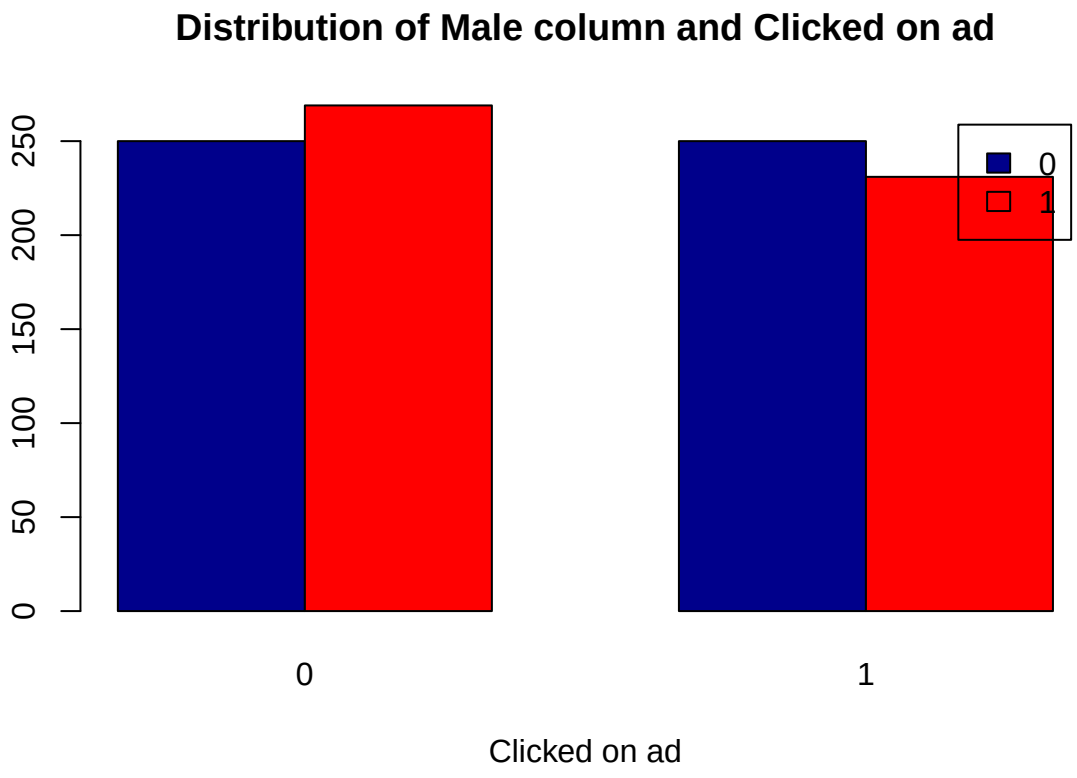
```
plot(advertising$Daily.Internet.Usage, advertising$Clicked.on.Ad, xlab="Daily.Internet.Usage", ylab="Cl
```



There are no insights from the scatterplots as shown above.

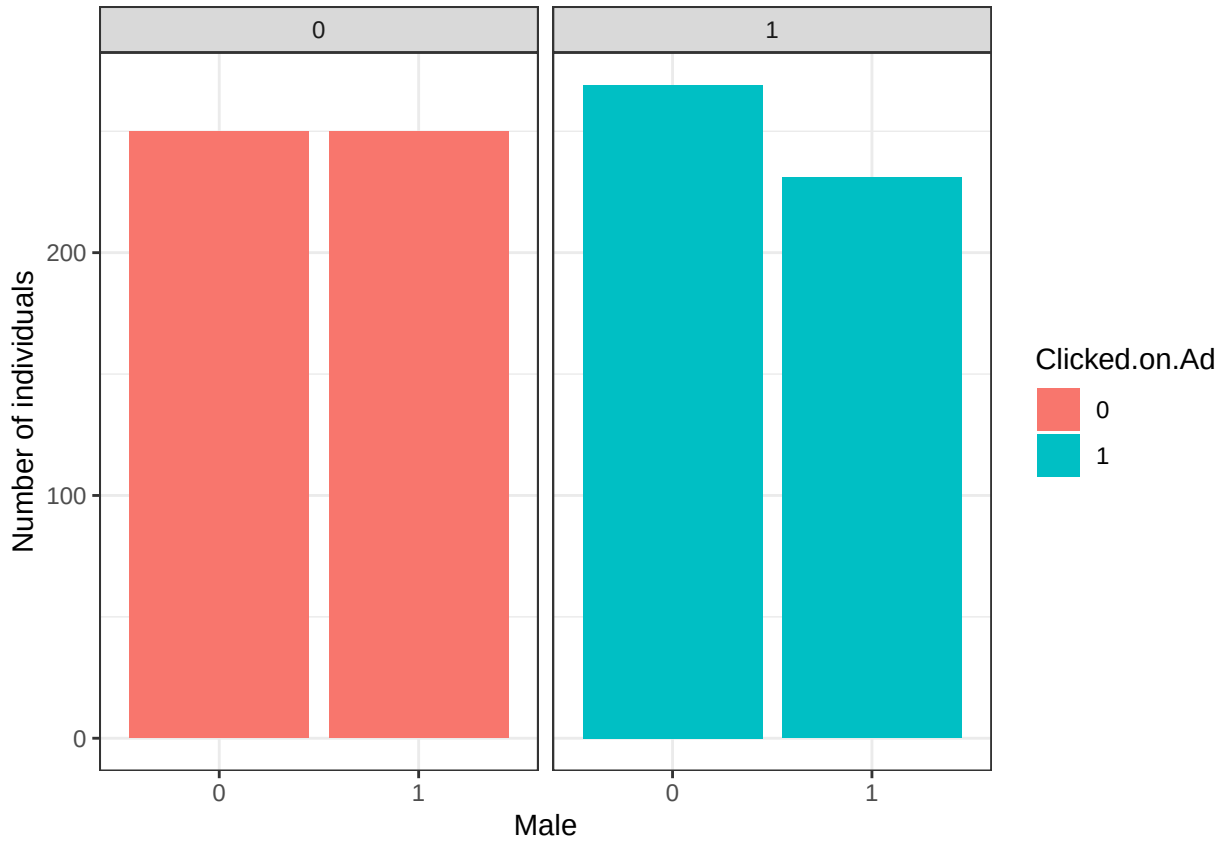
6.2.4 More bivariate analysis Visualization of Male column with target variable

```
counts <- table(advertising$Clicked.on.Ad, advertising$Male)
barplot(counts, main="Distribution of Male column and Clicked on ad",
  xlab="Clicked on ad",col=c("darkblue","red"),
  legend = rownames(counts), beside=TRUE)
```



There was a slightly higher number of non males who clicked on the ad as shown by the frequency table below.

```
advertising$Male <- as.factor(advertising$Male)
advertising$Clicked.on.Ad <- as.factor(advertising$Clicked.on.Ad)
ggplot(advertising, aes(x=Male, fill = Clicked.on.Ad))+
  theme_bw()+
  geom_bar()+
  facet_wrap(~Clicked.on.Ad)+
  labs(y="Number of individuals")
```




```
advertising %>%
    group_by(Clicked.on.Ad, Male) %>%
    summarize(frequency = n())
```

```
## `summarise()` regrouping output by 'Clicked.on.Ad' (override with `.groups` argument)
```

```
## # A tibble: 4 x 3
## # Groups:   Clicked.on.Ad [2]
##   Clicked.on.Ad Male  frequency
##   <fct>         <fct>     <int>
## 1 0             0           250
## 2 0             1           250
## 3 1             0           269
## 4 1             1           231
```

Comparing countries with click on ad.

```
advertising %>%
    group_by(Clicked.on.Ad, Country) %>%
    summarize(frequency = n())
```

```
## `summarise()` regrouping output by 'Clicked.on.Ad' (override with `.groups` argument)
```

```
## # A tibble: 430 x 3
## # Groups:   Clicked.on.Ad [2]
##    Clicked.on.Ad Country                                    frequency
##    <fct>         <chr>                                          <int>
##  1 0             Afghanistan                                        3
##  2 0             Albania                                            3
```

```
##  3 0             Algeria                                                3
##  4 0             American Samoa                                         2
##  5 0             Angola                                                 3
##  6 0             Anguilla                                               3
##  7 0             Antarctica (the territory South of 60 deg S)           1
##  8 0             Antigua and Barbuda                                    1
##  9 0             Argentina                                              1
## 10 0             Armenia                                                2
## # ... with 420 more rows
```

```
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <fct> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, ...
```

```
advertising$Clicked.on.Ad <- as.factor(advertising$Clicked.on.Ad)
```

```
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <fct> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, ...
```

```
advertising$Clicked.on.Ad <- as.character(advertising$Clicked.on.Ad)
```

```
advertising %>% mutate(Clicked.on.Ad=Clicked.on.Ad %>% as.character) %>% glimpse()
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
```

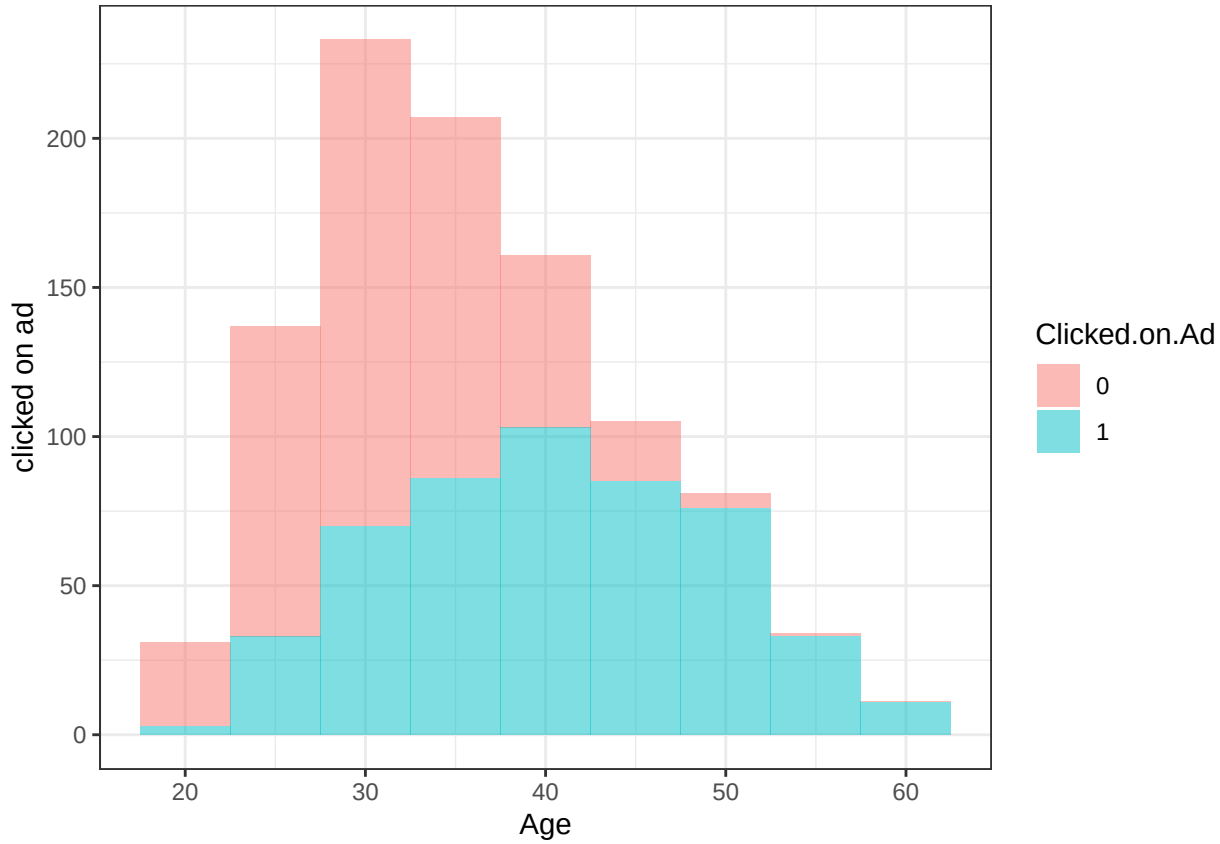
```
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

```
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

Histogram of Age vs clicked on Ad

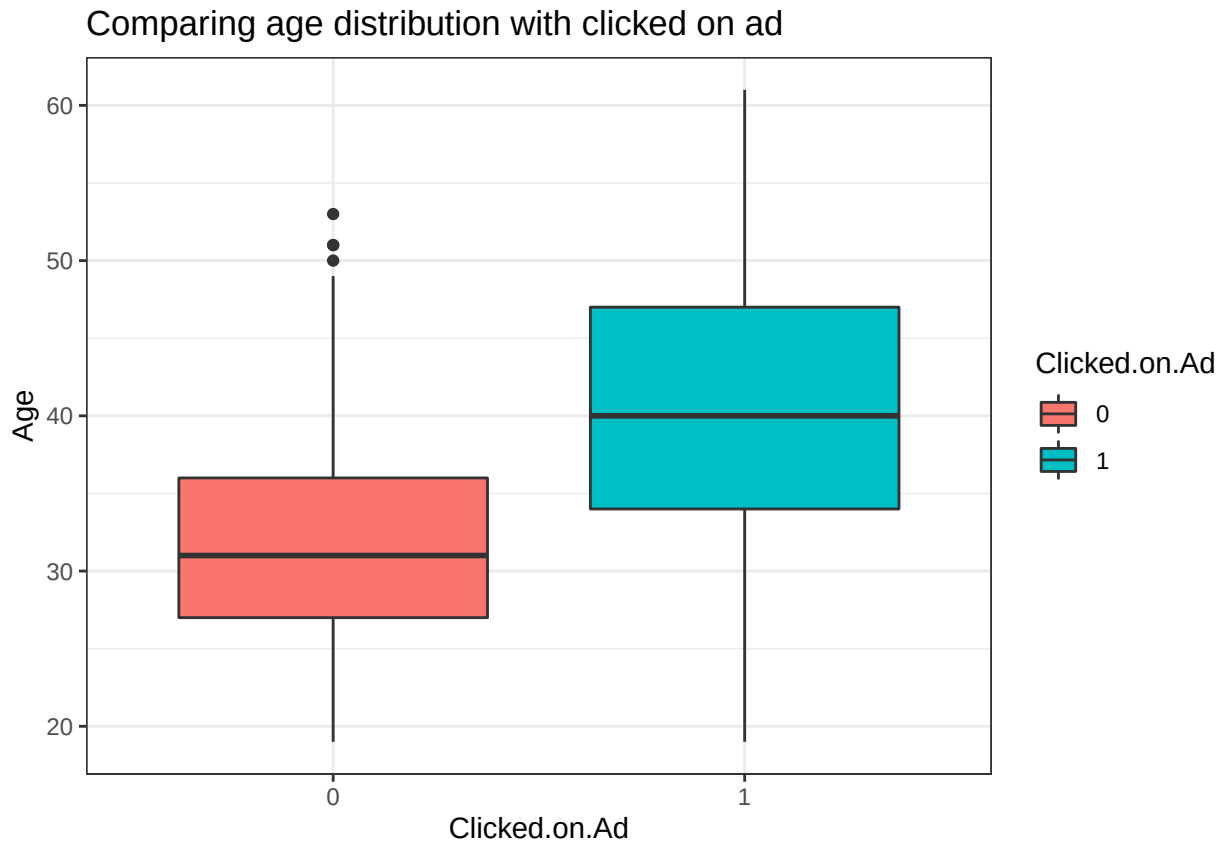
```
ggplot(advertising, aes(x=Age, fill=Clicked.on.Ad))+
  theme_bw()+
  geom_histogram(binwidth=5, alpha=0.5)+
  labs(y='clicked on ad')
```



Boxplot of Age Vs clicked on ad

```
ggplot(advertising, aes(y=Age, x=Clicked.on.Ad, fill=Clicked.on.Ad))+
  theme_bw()+
  geom_boxplot()+
```

```
  labs(y='Age', title='Comparing age distribution with clicked on ad')
```

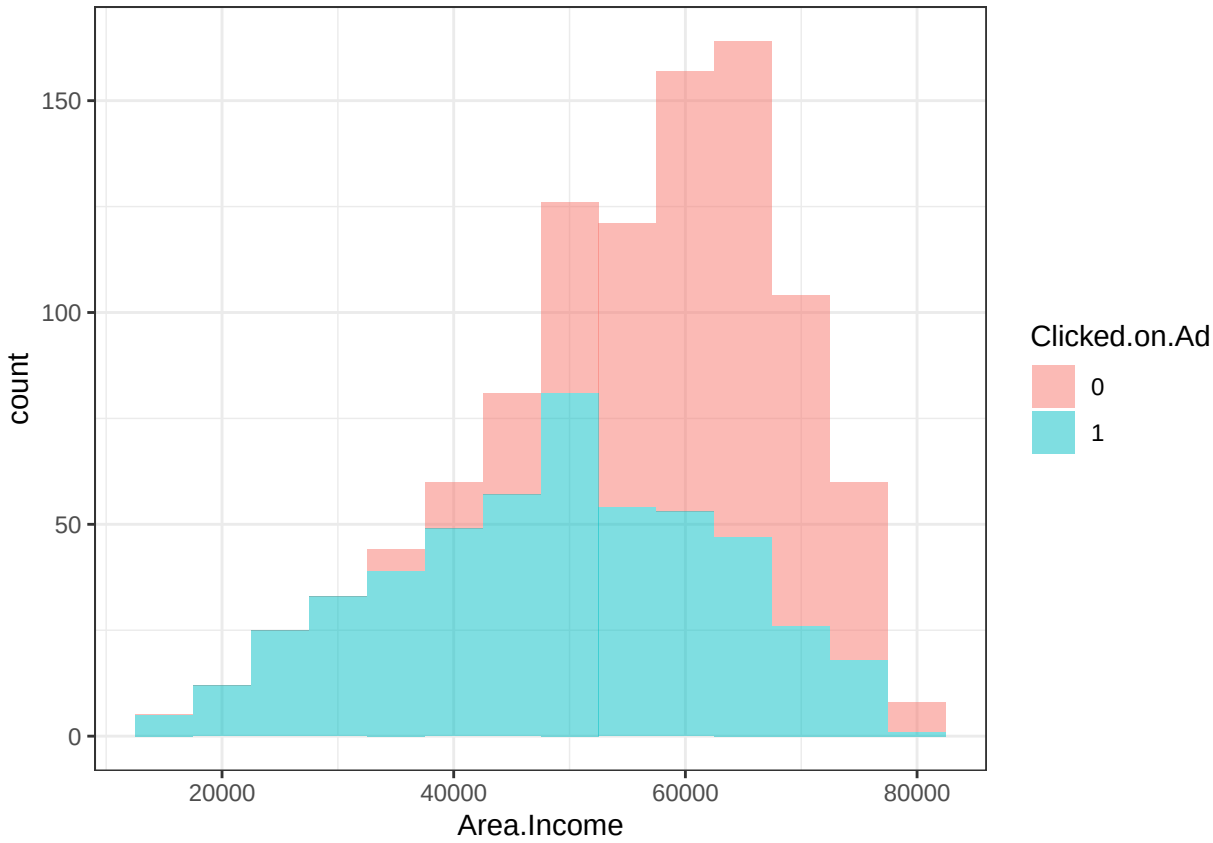



```
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

Histogram of Area Income Vs Clicked on Ad

```
ggplot(advertising, aes(x=Area.Income, fill=Clicked.on.Ad))+
  theme_bw()+
  geom_histogram(binwidth=5000, alpha=0.5)
```

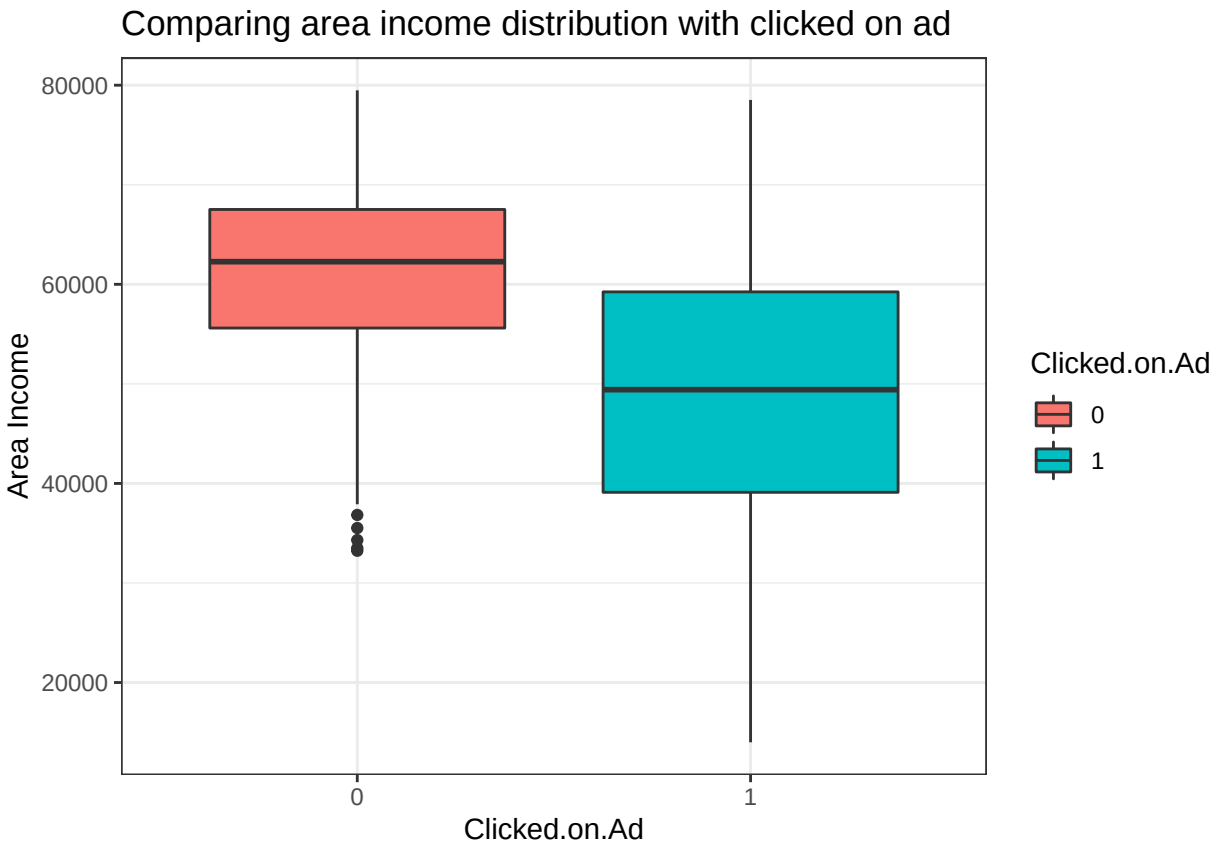



```
  labs(y='clicked on ad')
```

```
## $y
## [1] "clicked on ad"
##
## attr(,"class")
## [1] "labels"
```

Box plot of Area Income Vs clicked on Ad

```
ggplot(advertising, aes(y=Area.Income, x=Clicked.on.Ad, fill=Clicked.on.Ad))+
  theme_bw()+
  geom_boxplot()+
  labs(y='Area Income', title='Comparing area income distribution with clicked on ad')
```

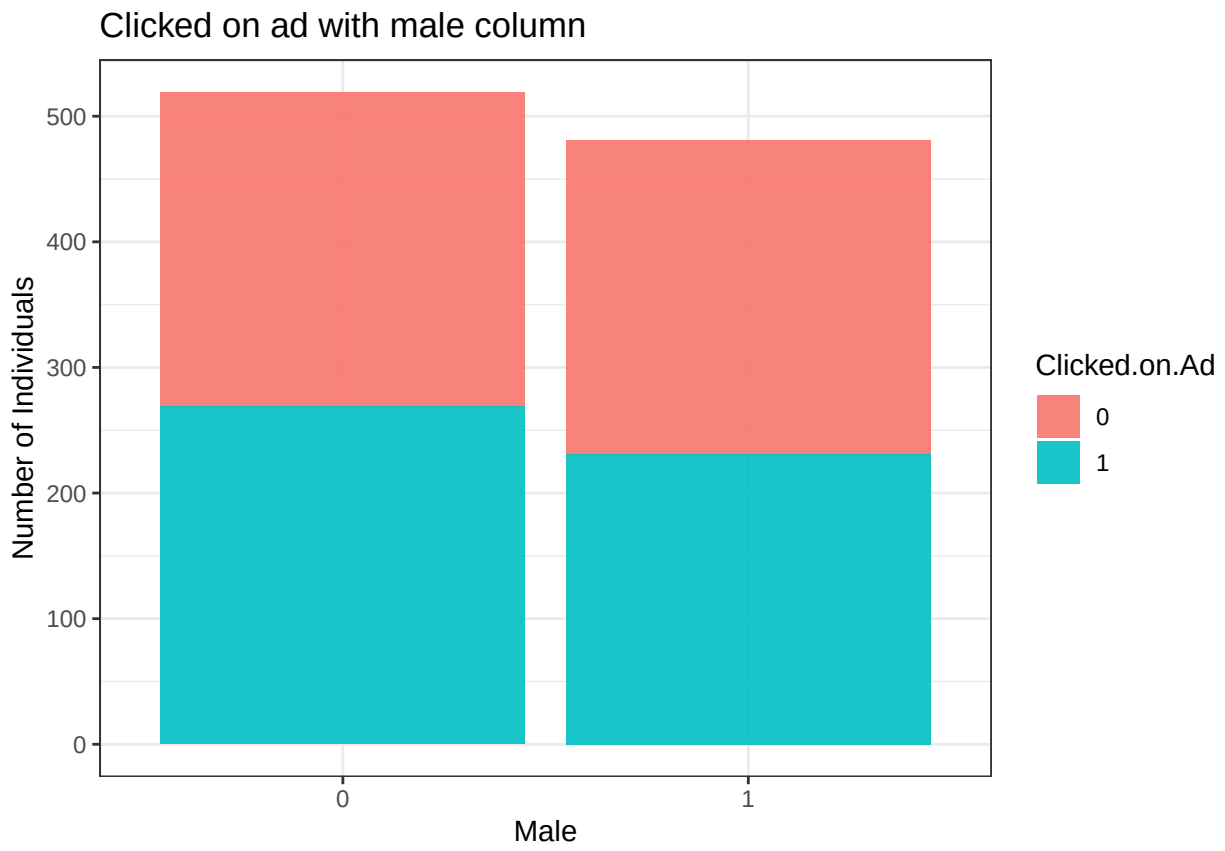



```
advertising$Male <- advertising$Male %>% as.character()
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <chr> "0", "1", "0", "1", "0", "1", "0", "1", "1...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

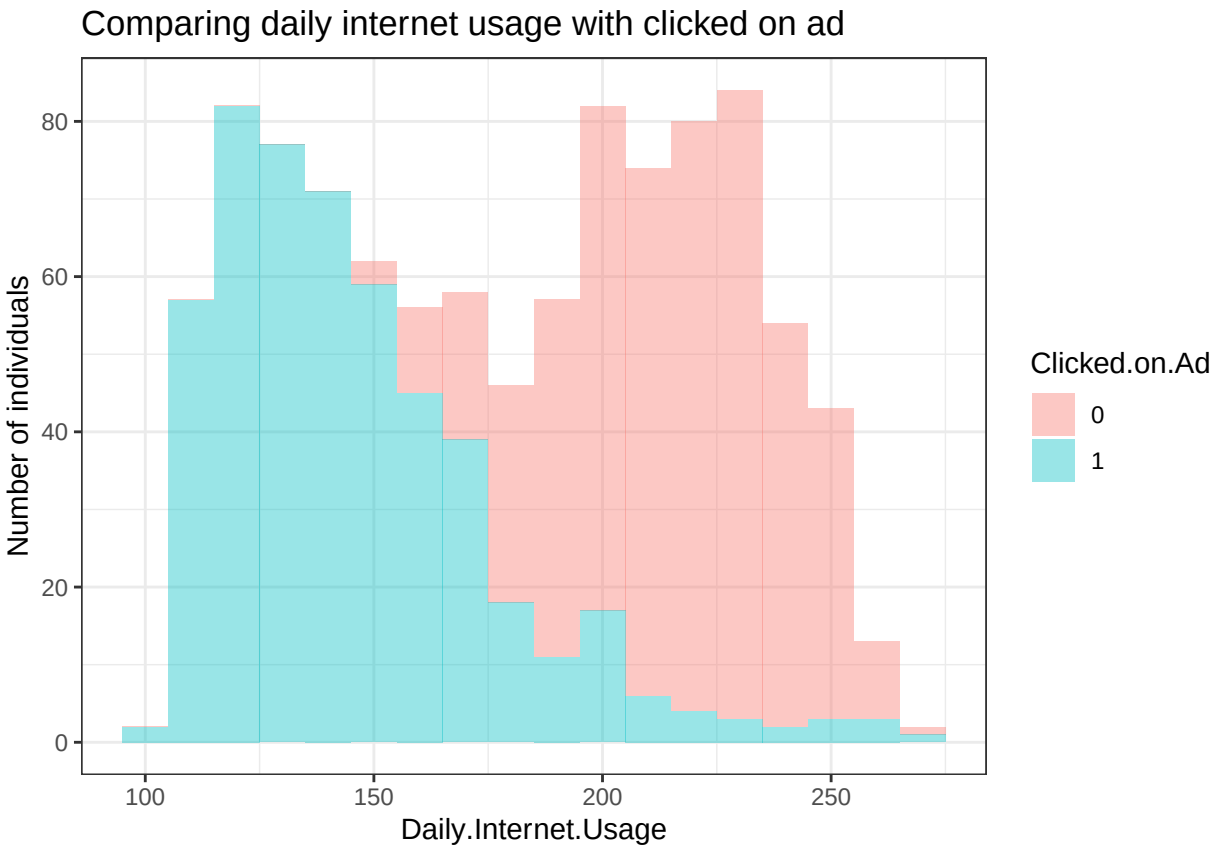
Boxplot of make vs Clicked on Ad

```
ggplot(advertising, aes(x=Male, fill=Clicked.on.Ad))+
  geom_bar(alpha=0.9)+
  theme_bw()+
  labs(title='Clicked on ad with male column', y= 'Number of Individuals')
```



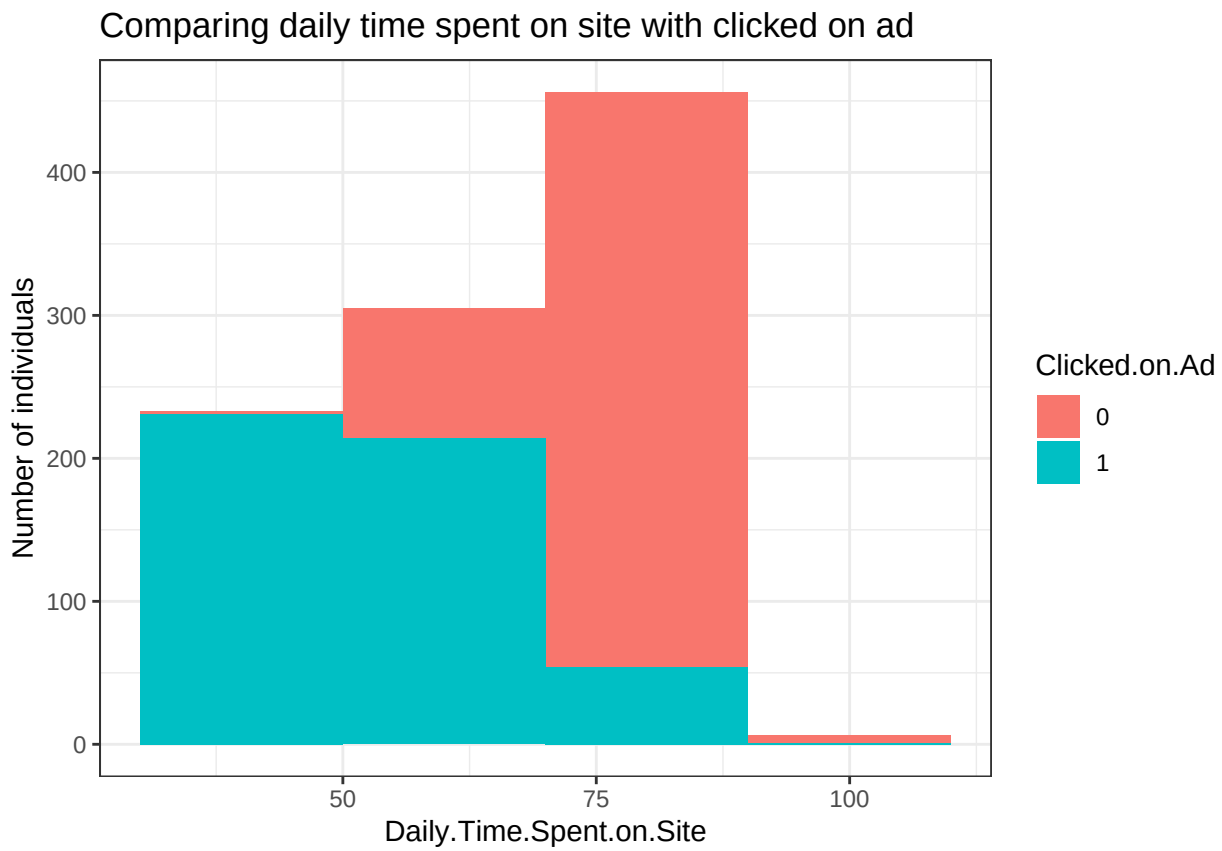
histogram of Daily internet usage vs Clicked on Ad

```
ggplot(advertising, aes(x=Daily.Internet.Usage, fill= Clicked.on.Ad))+
  theme_bw()+
  geom_histogram(binwidth=10, alpha=0.4)+
  labs(title='Comparing daily internet usage with clicked on ad', y='Number of individuals')
```



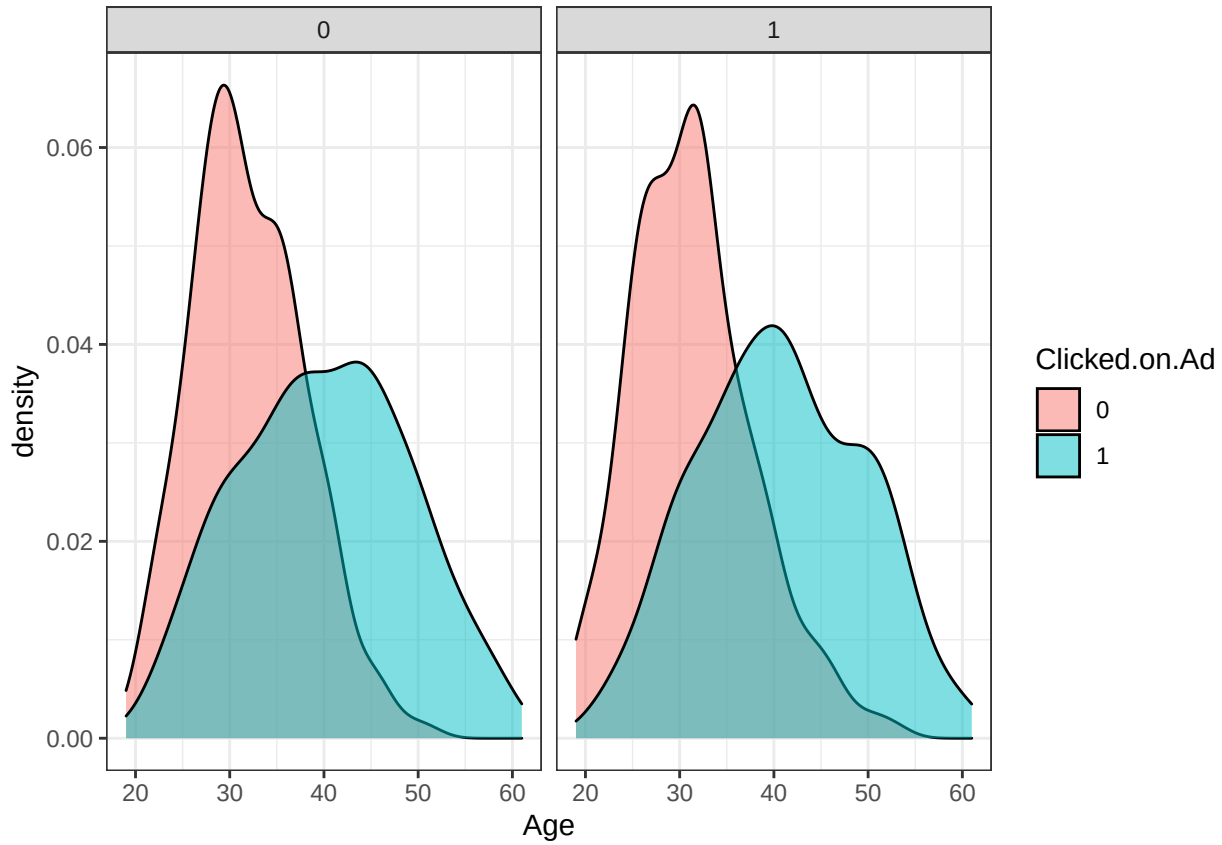
Histogram of daily time spent on site by clicked on ad

```
ggplot(advertising, aes(x=Daily.Time.Spent.on.Site, fill=Clicked.on.Ad))+
  theme_bw()+
  geom_histogram(binwidth = 20, alpha=10)+
  labs(y='Number of individuals', title='Comparing daily time spent on site with clicked on ad')
```



Distribution plots for Age vs clicked on Ad

```
ggplot(advertising, aes(x=Age, fill=Clicked.on.Ad))+
  theme_bw()+
  facet_wrap(~Male)+
  geom_density(alpha=0.5)
```



7. Supervised machine Learning ### 7.1 SVM #### 7.1.1 Preprocessing

```
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Ad.Topic.Line            <chr> "Cloned 5thgeneration orchestration", "Mon...
## $ City                     <chr> "Wrightburgh", "West Jodi", "Davidton", "W...
## $ Male                     <chr> "0", "1", "0", "1", "0", "1", "0", "1", "1...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Timestamp                <chr> "2016-03-27 00:53:11", "2016-04-04 01:39:0...
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

```
#dropping unnecessary columns
advertising$Ad.Topic.Line <- NULL
advertising$City <- NULL
advertising$Timestamp <- NULL
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 7
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
```

```
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Male                     <chr> "0", "1", "0", "1", "0", "1", "0", "1", "1...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Clicked.on.Ad            <chr> "0", "0", "0", "0", "0", "0", "0", "1", "0...
```

```
#converting categorical variables to factors
advertising[["Male"]] = factor(advertising[["Male"]])
advertising[["Clicked.on.Ad"]] = factor(advertising[["Clicked.on.Ad"]])
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 7
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <chr> "Tunisia", "Nauru", "San Marino", "Italy",...
## $ Clicked.on.Ad            <fct> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, ...
```

```
advertising[["Country"]] = as.integer(as.factor(advertising[["Country"]]))
```

```
glimpse(advertising)
```

```
## Rows: 1,000
## Columns: 7
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <int> 216, 148, 185, 104, 97, 159, 146, 13, 83, ...
## $ Clicked.on.Ad            <fct> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, ...
```

```
#splitting data into train and test. test size = 0.3
df.svm <- advertising #creating a copy of data frame to run SVM
intrain <- createDataPartition(y = advertising$Clicked.on.Ad, p= 0.7, list = FALSE)
training <- advertising[intrain,]
testing <- advertising[-intrain,]
```

```
#checking the dimensions of our training and test dataset
dim(training)
```

7.1.2 SVM application

```
## [1] 700   7
```

```
dim(testing)
```

```
## [1] 300   7
```

```
#controlling the computational overheads before modeling using train control() function
train.control <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
#fitting our model
svm_Linear <- train(Clicked.on.Ad ~., data = training, method = "svmLinear",
trControl=train.control,
```

```
preProcess = c("center", "scale"),
tuneLength = 10)
```

```
#making predictions
test_pred <- predict(svm_Linear, newdata = testing)
test_pred
```

```
##   [1] 0 0 0 0 1 0 0 1 0 1 1 0 0 0 1 0 1 1 1 1 0 1 0 0 1 0 0 1 0 1 1 0 1 1 1 0 1
##  [38] 1 1 0 1 0 1 1 1 1 0 1 1 0 0 1 1 0 1 1 1 0 1 0 1 0 0 1 1 1 0 0 1 0 1 1 1 1
##  [75] 0 1 1 1 1 0 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0 1 0
## [112] 0 1 1 1 1 0 0 1 0 1 1 0 1 0 1 1 0 0 0 0 0 0 1 0 1 1 0 0 1 1 1 0 0 1 0 1 0
## [149] 1 0 0 0 1 0 0 0 0 1 1 1 0 1 0 1 0 1 0 0 1 1 0 1 0 1 0 1 1 0 0 0 1 1 1 0 0
## [186] 1 1 1 1 0 0 0 1 1 0 0 1 1 1 0 0 1 1 0 1 1 0 0 1 0 1 1 0 0 0 1 1 0 0 0 1 0
## [223] 0 1 0 1 0 1 0 0 1 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1 0 0 0 0
## [260] 0 1 0 1 1 1 0 1 1 1 1 0 1 1 0 1 1 0 0 0 1 1 1 0 1 1 1 0 0 1 0 0 0 1 1 0 0
## [297] 0 0 0 1
## Levels: 0 1
```

```
#evaluating the model using confusion matrix
confusionMatrix(table(test_pred, testing$Clicked.on.Ad))
```

```
## Confusion Matrix and Statistics
##
##
## test_pred   0   1
##         0 147   4
##         1   3 146
##
##                Accuracy : 0.9767
##                  95% CI : (0.9525, 0.9906)
##     No Information Rate : 0.5
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9533
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9800
##             Specificity : 0.9733
##          Pos Pred Value : 0.9735
##          Neg Pred Value : 0.9799
##              Prevalence : 0.5000
##          Detection Rate : 0.4900
##    Detection Prevalence : 0.5033
##       Balanced Accuracy : 0.9767
##
##        'Positive' Class : 0
##
```

SVM model has performed really well with an accuracy score of 97%. From the confusion matrix 145 individuals were predicted to be individuals who did not click on the ad, where as, 146 were correctly predicted to have clicked on the ad.

8. Challenging the solution

We will challenge our solution using Naive Bayes algorithm

8.1 Naive Bayes

```
#creating a copy of data frame for Naive Bayes
df.nb <- advertising
glimpse(df.nb)
```

```
## Rows: 1,000
## Columns: 7
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
## $ Country                  <int> 216, 148, 185, 104, 97, 159, 146, 13, 83, ...
## $ Clicked.on.Ad            <fct> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, ...
```

```
#splitting data into training and test. we will use a test size of 0.3
indxTrain <- createDataPartition(y = df.nb$Clicked.on.Ad,p = 0.7,list = FALSE)
training <- df.nb[indxTrain,]
testing <- df.nb[-indxTrain,]
```

```
#Checking dimensions of the split

prop.table(table(df.nb$Clicked.on.Ad)) * 100
```

```
##
##  0  1
## 50 50
```

```
prop.table(table(training$Clicked.on.Ad)) * 100
```

```
##
##  0  1
## 50 50
```

```
prop.table(table(testing$Clicked.on.Ad)) * 100
```

```
##
##  0  1
## 50 50
```

```
# Creating variables that hold predictor and target variables
x = training[,-6]
y = training$Clicked.on.Ad
```

```
glimpse(df.nb)
```

```
## Rows: 1,000
## Columns: 7
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99, ...
## $ Age                      <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20, 49...
## $ Area.Income              <dbl> 61833.90, 68441.85, 59785.94, 54806.18, 73...
## $ Daily.Internet.Usage     <dbl> 256.09, 193.77, 236.50, 245.89, 225.58, 22...
## $ Male                     <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, ...
```

```
## $ Country                  <int> 216, 148, 185, 104, 97, 159, 146, 13, 83, ...
## $ Clicked.on.Ad            <fct> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, ...
```

```
# building naive bayes model
# nb.model <- train(x,y,'nb',trControl=trainControl(method='cv',number=10))
```

```
# # making predictions
# Predict <- predict(nb.model,newdata = testing )
#
# # evaluating the model
# > confusionMatrix(Predict, testing$Clicked.on.Ad )
```

9. Recommendations

1. From the correlation matrix we saw a strong negative correlation between target variable and area income. This means that individuals who live in areas of low income were likely to click on the add compared to individuals from high income areas. We recommend that the entrepreneur focuses on areas of low income because the individuals from this areas will likely click on the ad.
2. We also saw that the target variable was strongly negatively correlated to daily time spent on the blog. It is possible that individuals spending less time on the blog go into the blog to check on any new course ads. We recommend that the entrepreneur find ways of attracting the attention of the individuals who spend longer times in the blog so that they can also click on her ads.
3. Finally, from the correlation matrix we noticed that the target variable was very weakly correlated to the male column. This means that there is no strong relationship between the gender and the target variable so the entrepreneur can continue with the advertisement without worrying about gender bias.
4. The best performing model for this project with an accuracy of 97.0 % is SVM