Research Question

2022-06-06

```r
options(knitr.duplicate.label = "allow")
```

Defining the question

Identifying which individuals are most likely to click on an online cryptography course

Defining the metric of success

The metric of success for this project will be identifying which individuals click on the course

Understanding the context

A Kenyan entrepreneur has created an online cryptography course and would want to advertise it on her blog. She currently targets audiences originating from various countries. She would now like to employ your services as a Data Science Consultant to help her identify which individuals are most likely to click on her ads.

Recordingg the experimental design

1. Reading and understanding the data
2. Clean the dataset
3. Perform univariate and bivariate analysis
4. Find insights and give a recommendation

Data relevance

The data provided is valid

```r
# Importing the dataset
df = read.csv("C:\\Users\\USER\\Documents\\Moringa School\\R Programming\\Fundermentals\\IP\\advertising
```

```r
# Preview the top of the dataset
head(df)
```

```
##   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
```

```r
# Preview the bottom of the dataset
tail(df)
```

```
##      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
```

```r
# Check the shape of the dataset
dim(df)
```

```
## [1] 1000   10
```

We have 1000 rows and 10 columns

```r
# Checking the datatype in the dataset
str(df)
```

```
## 'data.frame':    1000 obs. of  10 variables:
##  $ Daily.Time.Spent.on.Site: num  69 80.2 69.5 74.2 68.4 ...
##  $ Age                     : int  35 31 26 29 35 23 33 48 30 20 ...
##  $ Area.Income             : num  61834 68442 59786 54806 73890 ...
##  $ Daily.Internet.Usage    : num  256 194 236 246 226 ...
##  $ Ad.Topic.Line           : chr  "Cloned 5thgeneration orchestration" "Monitored national standardi
##  $ City                    : chr  "Wrightburgh" "West Jodi" "Davidton" "West Terrifurt" ...
##  $ Male                    : int  0 1 0 1 0 1 0 1 1 1 ...
##  $ Country                 : chr  "Tunisia" "Nauru" "San Marino" "Italy" ...
```

```
##  $ Timestamp               : chr  "2016-03-27 00:53:11" "2016-04-04 01:39:02" "2016-03-13 20:35:42"
##  $ Clicked.on.Ad           : int  0 0 0 0 0 0 0 1 0 0 ...
```

```r
# Checking for missing values
colSums(is.na(df))
```

```
## 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 null values in the dataset

```r
# Checking for duplicates in our dataset

duplicates_df = df[duplicated(df), ]
duplicates_df
```

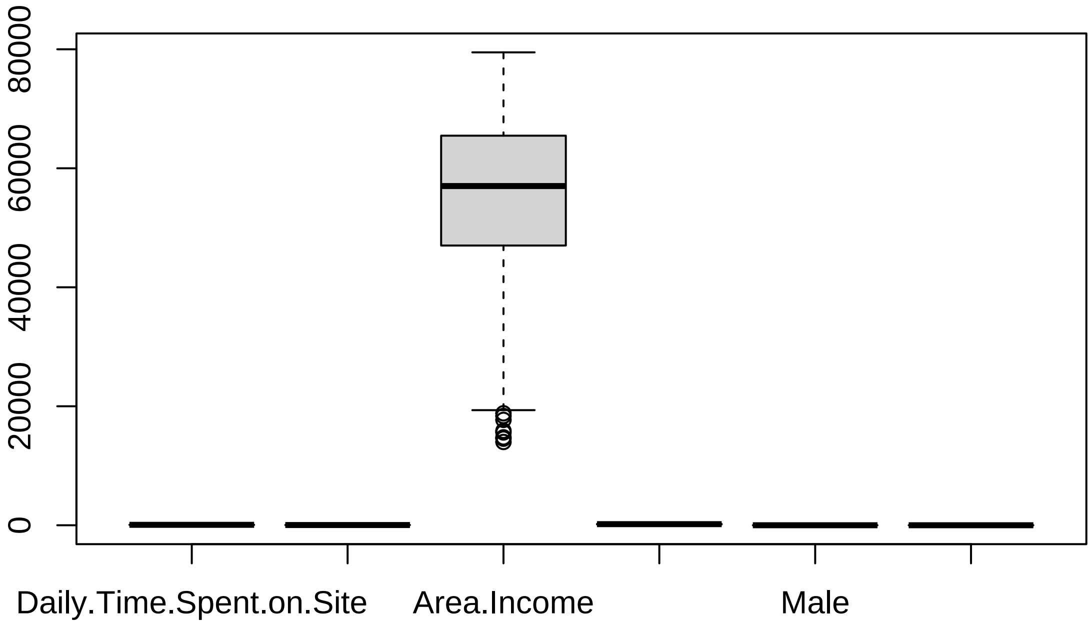
```
##  [1] Daily.Time.Spent.on.Site Age                      Area.Income
##  [4] Daily.Internet.Usage     Ad.Topic.Line            City
##  [7] Male                     Country                  Timestamp
## [10] Clicked.on.Ad
## <0 rows> (or 0-length row.names)
```

There are no duplicates in the data set

```r
# Checking for outliers using the boxplot
# Outline the numeric columns

numeric <- df[,unlist(lapply(df, is.numeric))]
head(numeric)
```

```
##   Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage Male
## 1                    68.95  35    61833.90               256.09    0
## 2                    80.23  31    68441.85               193.77    1
## 3                    69.47  26    59785.94               236.50    0
## 4                    74.15  29    54806.18               245.89    1
## 5                    68.37  35    73889.99               225.58    0
## 6                    59.99  23    59761.56               226.74    1
##   Clicked.on.Ad
## 1             0
## 2             0
## 3             0
## 4             0
## 5             0
## 6             0
```

```r
boxplot(numeric)
```




```r
## Univariate Analysis

# Mean
summary(numeric)
```

```
##  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
##       Male        Clicked.on.Ad
##  Min.   :0.000   Min.   :0.0
##  1st Qu.:0.000   1st Qu.:0.0
##  Median :0.000   Median :0.5
##  Mean   :0.481   Mean   :0.5
##  3rd Qu.:1.000   3rd Qu.:1.0
##  Max.   :1.000   Max.   :1.0
```

```r
# Checking the variance
```

```r
var(numeric)
```

```
##                          Daily.Time.Spent.on.Site           Age   Area.Income
## Daily.Time.Spent.on.Site              251.3370949 -4.617415e+01  6.613081e+04
## Age                                   -46.1741459  7.718611e+01 -2.152093e+04
## Area.Income                         66130.8109082 -2.152093e+04  1.799524e+08
```

```
## Daily.Internet.Usage                  360.9918827 -1.416348e+02  1.987625e+05
## Male                                   -0.1501864 -9.242142e-02  8.867509e+00
## Clicked.on.Ad                          -5.9331431  2.164665e+00 -3.195989e+03
##                          Daily.Internet.Usage        Male Clicked.on.Ad
## Daily.Time.Spent.on.Site         3.609919e+02 -0.15018639 -5.933143e+00
## Age                             -1.416348e+02 -0.09242142  2.164665e+00
## Area.Income                      1.987625e+05  8.86750903 -3.195989e+03
## Daily.Internet.Usage             1.927415e+03  0.61476667 -1.727409e+01
## Male                             6.147667e-01  0.24988889 -9.509510e-03
## Clicked.on.Ad                   -1.727409e+01 -0.00950951  2.502503e-01
```

```r
# Checking standard deviation
# For the age column
sd(df$Age)
```

```
## [1] 8.785562
```

```r
# For the area income column
sd(df$Area.Income)
```

```
## [1] 13414.63
```

```r
# For the internet usage column
sd(df$Daily.Internet.Usage)
```

```
## [1] 43.90234
```

```r
# Frequency table for age column
table(df$Age)
```

```
##
## 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44
##  6  6  6 13 19 21 27 37 33 48 48 39 60 38 43 39 39 50 36 37 30 36 32 26 23 21
## 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61
## 30 18 13 16 18 20 12 15 10  9  7  2  6  4  2  4  1
```

```r
# Frequency table for income column
table(df$Area.Income)
```

```
##
##  13996.5 14548.06  14775.5 15598.29  15879.1 17709.98 18368.57 18819.34
##        1        1        1        1        1        1        1        1
## 19345.36 19991.72 20592.99 20856.54 21644.91 21773.22 22205.74 22456.04
##        1        1        1        1        1        1        1        1
## 22473.08 23410.75 23821.72 23936.86 23942.61 23975.35 24030.06 24078.93
##        1        1        1        1        1        1        1        1
## 24316.61 24593.33  24852.9 25371.52 25408.21 25583.29 25598.75 25603.93
##        1        1        1        1        1        1        1        1
## 25682.65 25686.34 25739.09 25767.16 26023.99 26130.93 27073.27  27086.4
##        1        1        1        1        1        1        1        1
## 27241.11 27262.51 27508.41  27964.6 28019.09 28028.74 28186.65 28210.03
##        1        1        1        1        1        1        1        1
## 28265.81 28271.84 28275.48 28357.27 28387.42 28495.21 28679.93  29359.2
##        1        1        1        1        1        1        1        1
## 29398.61 29727.79  29875.8 30227.98 30487.48 30726.26    30976 31072.44
##        1        1        1        1        1        1        1        1
## 31087.54 31092.93 31215.88 31265.75 31281.01 31343.39 31523.09 31947.65
##        1        1        1        1        1        1        1        1
```

```
## 31998.72 32006.82 32252.38 32536.98 32549.95 32593.59  32635.7 32689.04
##        1        1        1        1        1        1        1        1
## 32708.94 32847.53 33147.19  33239.2 33258.09 33293.78 33502.57  33553.9
##        1        1        1        1        1        1        1        1
## 33601.84 33813.08 33951.63 33987.27 34127.21 34191.13 34191.23 34309.24
##        1        1        1        1        1        1        1        1
## 34418.09 34886.01 34903.67 34942.26 35253.98 35349.26 35350.55  35466.8
##        1        1        1        1        1        1        1        1
## 35521.88 35684.82 35764.49 36037.33 36424.94 36497.22 36752.24 36782.38
##        1        1        1        1        1        1        1        1
## 36834.04 36884.23 36913.51 37212.54 37334.78 37345.24 37345.34 37605.11
##        1        1        1        1        1        1        1        1
## 37713.23 37838.72 37908.29 38067.08 38260.89 38349.78 38427.66  38609.2
##        1        1        1        1        1        1        1        1
##  38641.2  38645.4 38745.29  38817.4 38987.42 39031.89 39131.53 39132.64
##        1        1        1        1        1        1        1        1
## 39193.45 39211.49 39552.49    39616 39699.13 39723.97 39799.73 39809.69
##        1        1        1        1        1        1        1        1
## 39840.55 39939.39 40135.06  40159.2 40182.84 40183.75 40243.82 40345.49
##        1        1        1        1        1        1        1        1
## 40468.53 40478.83 40763.13 40926.93 41059.64 41097.17 41229.16 41232.89
##        1        1        1        1        1        1        1        1
## 41335.84 41356.31 41417.27 41521.28 41547.62 41629.86 41768.13 41851.38
##        1        1        1        1        1        1        1        1
## 41866.55 41884.64 41920.79 42042.95 42078.89 42136.33  42162.9 42191.61
##        1        1        1        1        1        1        1        1
## 42251.59 42362.49 42415.72 42581.23 42650.32 42696.67 42760.22 42838.29
##        1        1        1        1        1        1        1        1
## 42861.42 42898.21 42907.89 42993.48  42995.8 43073.78 43111.41 43155.19
##        1        1        1        1        1        1        1        1
## 43241.19 43241.88 43299.63 43313.73 43386.07 43444.86 43450.11 43573.66
##        1        1        1        1        1        1        1        1
##  43662.1 43698.53 43708.88 43778.88 43870.51 43881.73 43974.49 44078.24
##        1        1        1        1        1        1        1        1
## 44174.25 44217.68 44248.52 44275.13 44304.13 44307.18 44490.09 44559.43
##        1        1        1        1        1        1        1        1
## 44893.71  45400.5 45465.25 45522.44 45580.92 45593.93 45632.51 45716.48
##        1        1        1        1        1        1        1        1
## 45800.48 45945.88 45959.86 46004.31 46024.29 46033.73 46132.18 46160.63
##        1        1        1        1        1        1        1        1
## 46179.97 46197.59 46239.14 46339.25 46403.18 46422.76 46473.14 46500.11
##        1        1        1        1        1        1        1        1
## 46557.92 46653.75 46693.76 46722.07 46737.34 46780.09 46868.53 46931.03
##        1        1        1        1        1        1        1        1
## 46964.11 46974.15 47051.02 47139.21 47160.53 47169.14 47258.59 47314.45
##        1        1        1        1        1        1        1        1
## 47338.94 47357.39 47391.95 47447.89 47510.42 47575.44  47638.3 47682.28
##        1        1        1        1        1        1        1        1
## 47708.42 47861.93 47929.83 47968.32 47997.75 48098.86 48206.04  48246.6
##        1        1        1        1        1        1        1        1
##  48335.2 48347.64 48376.14 48453.55 48467.68 48537.18 48554.45 48679.54
##        1        1        1        1        1        1        1        1
## 48758.92 48761.14 48826.14 48852.58 48867.36 48867.67 48913.07 48918.55
##        1        1        1        1        1        1        1        1
```

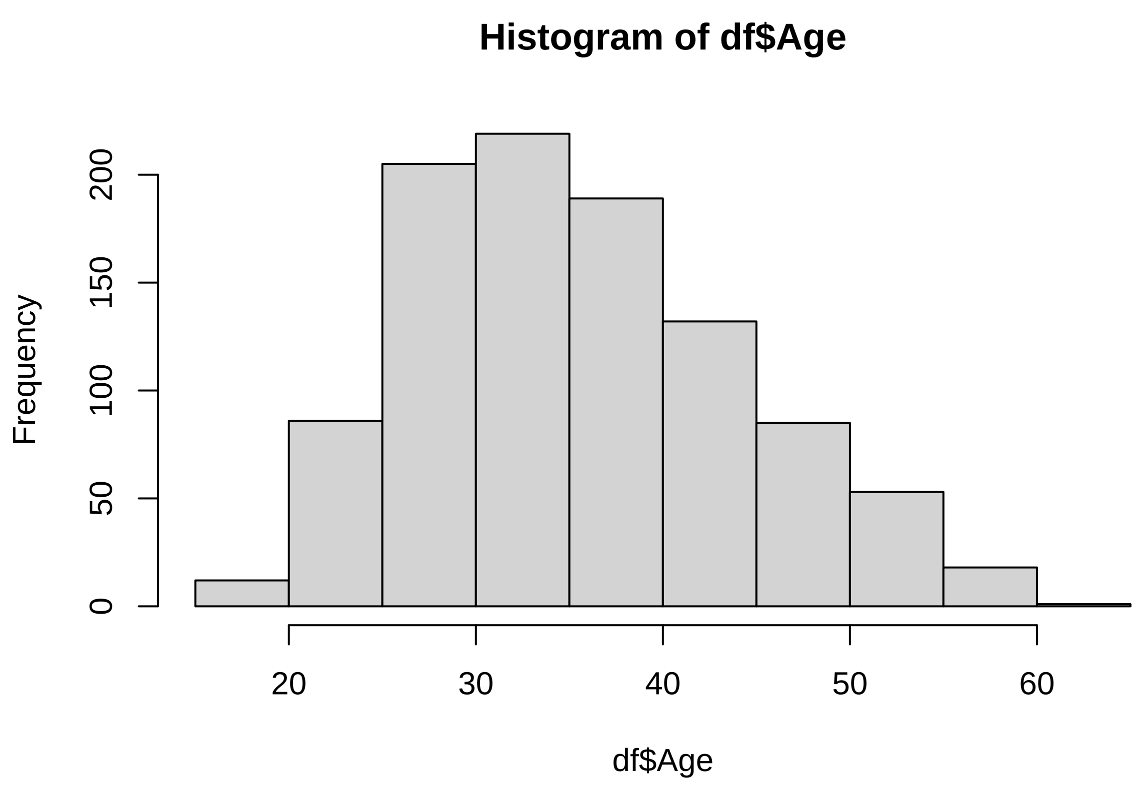
```
## 49030.03 49090.51 49101.67 49111.47  49158.5  49206.4 49269.98 49282.87 
##        1        1        1        1        1        1        1        1 
## 49309.14 49325.48 49457.48 49525.37 49544.41 49597.08 49742.83 49822.78 
##        1        1        1        1        1        1        1        1 
## 49850.52 49911.25 49942.66    49957 49995.63 50038.65 50055.33 50086.17 
##        1        1        1        1        1        1        1        1 
## 50147.72 50199.77 50216.01 50278.89 50333.72 50335.46 50337.93 50356.06 
##        1        1        1        1        1        1        1        1 
## 50439.49 50457.01 50468.36 50491.45 50506.44 50628.31  50666.5  50671.6 
##        1        1        1        1        1        1        1        1 
## 50711.68 50723.67 50760.23 50820.74 50950.24 50960.08 50971.73 50983.75 
##        1        1        1        1        1        1        1        1 
## 51013.37 51015.11 51049.47 51067.54 51119.93 51163.14 51171.23 51257.26 
##        1        1        1        1        1        1        1        1 
## 51315.38 51317.33 51363.16 51409.45 51463.17 51473.28 51501.38 51510.18 
##        1        1        1        1        1        1        1        1 
## 51512.66 51593.46 51600.47 51633.34 51636.12 51636.92 51662.24 51691.55 
##        1        1        1        1        1        1        1        1 
## 51739.63 51772.58 51812.71 51816.27 51824.01 51847.26 51864.77 51868.85 
##        1        1        1        1        1        1        1        1 
## 51869.87 51900.03 51920.49 51975.41    52011 52079.18 52097.32 52140.04 
##        1        1        1        1        1        1        1        1 
##  52177.4 52178.98 52182.23 52252.91 52261.73 52336.64  52340.1 52400.88 
##        1        1        1        1        1        1        1        1 
## 52416.18 52462.04 52520.75  52530.1 52563.22 52581.16 52656.13 52686.47 
##        1        1        1        1        1        1        1        1 
## 52691.79 52723.34 52736.33    52802 52802.58 52968.22 53012.94 53041.77 
##        1        1        1        1        1        1        1        1 
## 53042.51 53049.44 53058.91 53167.68 53185.34 53188.69 53223.58 53309.61 
##        1        1        1        1        1        1        1        1 
## 53336.76 53350.11 53412.32 53431.35 53441.69 53549.94 53575.48 53647.81 
##        1        1        1        1        1        1        1        1 
## 53673.08 53700.57 53767.12 53817.02 53852.85 53898.89 53922.43 54045.39 
##        1        1        1        1        1        1        1        1 
## 54106.21 54251.78  54286.1 54324.73 54429.17 54520.14 54541.56  54645.2 
##        1        1        1        1        1        1        1        1 
## 54725.87 54755.71 54773.99 54774.77 54787.37 54806.18 54875.95 54952.42 
##        1        1        1        1        1        1        1        1 
## 54989.93 55002.05 55015.08  55041.6 55121.65 55130.96 55187.85 55195.61 
##        1        1        1        1        1        1        1        1 
## 55316.97 55336.18 55353.41 55358.88 55368.67 55411.06 55424.24 55479.62 
##        1        1        1        1        1        1        1        1 
## 55499.69 55605.92 55642.32 55677.12 55764.43 55787.58 55901.12 55942.04 
##        1        1        1        1        1        1        1        1 
## 55984.89 55993.68 56067.38 56113.37 56129.89 56180.93 56194.56 56216.57 
##        1        1        1        1        1        1        1        1 
##  56242.7 56366.88 56369.74  56379.3 56394.82  56435.6 56457.01 56570.06 
##        1        1        1        1        1        1        1        1 
##  56593.8 56605.12 56637.59 56681.65 56683.32 56694.12 56725.47 56729.78 
##        1        1        1        1        1        1        1        1 
## 56735.14 56735.83 56759.48 56770.79 56782.18 56791.75 56884.74  56909.3 
##        1        1        1        1        1        1        1        1 
## 56974.51 56984.09 56986.73 57009.76 57014.84 57032.36 57179.91 57195.96 
##        1        1        1        1        1        1        1        1 
```

```
## 57260.41 57330.43 57425.87 57518.73 57519.64 57545.56    57587  57594.7
##        1        1        1        1        1        1        1        1
## 57667.99 57669.41 57691.95 57737.51 57739.03 57756.89 57777.11 57806.03
##        1        1        1        1        1        1        1        1
## 57844.96 57846.68 57868.44 57877.15 57887.64  57983.3 58019.64 58037.66
##        1        1        1        1        1        1        1        1
##  58114.3 58151.87 58183.04 58235.21 58287.86 58295.82 58337.18 58342.63
##        1        1        1        1        1        1        1        1
## 58348.41 58363.12 58443.99 58476.57 58526.04 58543.94 58576.12 58633.63
##        1        1        1        1        1        1        1        1
## 58638.75 58677.69 58776.67 58820.16 58847.07 58849.77 58909.36 58920.44
##        1        1        1        1        1        1        1        1
## 58953.01 58966.22 58996.12 58996.56 59047.91 59106.12 59144.02 59240.24
##        1        1        1        1        1        1        1        1
## 59243.46 59340.99 59397.89 59419.78 59422.47 59448.44 59457.52 59550.05
##        1        1        1        1        1        1        1        1
## 59593.56 59610.81 59621.02 59677.64 59683.16 59761.56 59784.18 59785.94
##        1        1        1        1        1        1        1        1
## 59797.64 59886.58 59967.19  59998.5 60015.57 60082.66 60151.77 60188.38
##        1        1        1        1        1        1        1        1
## 60192.72 60223.52 60248.97 60283.47 60283.98 60309.58 60315.19 60333.38
##        1        1        1        1        1        1        1        1
## 60372.64 60465.72 60514.05 60550.66 60575.99 60583.02 60637.62 60638.38
##        1        1        1        1        1        1        1        1
## 60641.09    60803 60803.37 60805.93 60812.77 60843.32 60845.55 60879.48
##        1        1        1        1        1        1        1        1
## 60938.73 60953.93 60968.62 60997.84 61004.51 61005.87  61009.1 61039.13
##        1        1        1        1        1        1        1        1
## 61067.58 61068.26  61117.5 61142.33 61161.29 61172.07 61227.59 61228.96
##        1        1        1        1        1        1        1        1
## 61230.03 61270.14 61275.18 61383.79  61389.5 61428.18 61467.33 61526.25
##        1        1        1        1        1        1        1        1
## 61601.05 61608.23 61610.05 61617.98 61625.87 61628.72 61652.53 61690.93
##        1        1        1        1        1        1        1        1
## 61747.98 61757.12 61770.34  61771.9 61806.31  61833.9 61840.26 61922.06
##        1        1        1        1        1        1        1        1
## 62053.37 62060.11  62109.8 62161.26 62204.93 62238.58 62312.23 62318.38
##        1        1        1        1        1        1        1        1
## 62330.75 62336.39 62378.05 62430.55  62463.7  62466.1 62475.99 62491.01
##        1        1        1        1        1        1        1        1
## 62572.88 62589.84 62657.53 62667.51 62669.59 62722.57  62729.4 62772.42
##        1        1        1        1        1        1        1        1
## 62784.85 62790.96 62792.43 62927.96  62939.5 63001.03 63006.14 63060.55
##        1        1        1        1        1        1        1        1
## 63071.34 63100.13 63102.19 63107.88 63109.74 63115.34 63126.96 63274.88
##        1        1        1        1        1        1        1        1
## 63296.87 63319.99 63336.85 63363.04  63373.7 63394.41 63429.18 63430.33
##        1        1        1        1        1        1        1        1
## 63450.96  63493.6 63497.62  63528.8 63551.67 63580.22 63649.04 63664.32
##        1        1        1        1        1        1        1        1
##  63727.5 63764.28 63879.72 63883.81 63891.29 63924.82  63936.5 63965.16
##        1        1        1        1        1        1        1        1
## 63966.72 63976.44 64008.55 64011.26 64021.55 64045.93 64122.36 64147.86
##        1        1        1        1        1        1        1        1
```

```
##  64188.5 64235.51 64238.71 64264.25 64267.88 64287.78 64395.85  64410.8
##        1        1        1        1        1        1        1        1
## 64433.99 64447.77 64564.07 64631.22 64654.66 64698.58  64775.1 64802.33
##        1        1        1        1        1        1        1        1
##    64828 64902.47 64927.19 64929.61 65044.59 65120.86 65172.22 65180.97
##        1        1        1        1        1        1        1        1
## 65186.58 65227.79 65229.13 65280.16 65421.39 65461.92 65496.78 65499.93
##        1        1        1        1        1        1        1        1
## 65576.05 65620.25 65653.47 65704.79 65756.36 65773.49 65791.17 65816.38
##        1        1        1        1        1        1        1        1
## 65826.53 65834.97 65856.74 65882.81 65883.39 65899.68 65953.76 65956.71
##        1        1        1        1        1        1        1        1
## 65963.37 66025.11 66027.31 66050.63 66107.84 66176.97 66187.58 66193.81
##        1        1        1        1        1        1        1        1
## 66198.66 66200.96 66217.31 66225.72 66262.59 66263.37 66265.34 66269.49
##        1        1        1        1        1        1        1        1
## 66281.46 66291.67  66345.1 66348.95 66359.32 66412.04 66429.84 66431.87
##        1        1        1        1        1        1        1        1
## 66504.16 66522.79  66524.8 66541.05 66572.39    66574 66618.21  66624.6
##        1        1        1        1        1        1        1        1
## 66629.61 66636.84 66691.23 66699.12 66744.65 66773.83 66784.81 66815.54
##        1        1        1        1        1        1        1        1
## 66861.67  66873.9 66929.03 66980.27 67033.34 67050.16 67058.72 67080.94
##        1        1        1        1        1        1        1        1
## 67113.46 67132.46 67186.54 67240.25 67279.06 67301.39 67307.43    67323
##        1        1        1        1        1        1        1        1
## 67384.31 67430.96 67432.49 67479.62 67511.86 67516.07 67526.92 67575.12
##        1        1        1        1        1        1        1        1
## 67633.44 67669.06 67682.32 67686.16 67714.82 67744.56 67781.31 67782.17
##        1        1        1        1        1        1        1        1
## 67866.95 67938.77 67990.84  68016.9 68030.18 68033.54 68094.85 68211.35
##        1        1        1        1        1        1        1        1
## 68305.91 68324.48 68333.01 68348.99 68357.96 68441.85 68447.17 68448.94
##        1        1        1        1        1        1        1        1
## 68519.96 68614.98  68713.7    68717 68737.75 68783.45 68787.09    68862
##        1        1        1        1        1        1        1        1
## 68863.95 68877.02 68962.32 69112.84 69285.69 69428.73 69438.04 69456.83
##        1        1        1        1        1        1        1        1
## 69476.42 69481.85 69562.46 69646.35 69710.51 69718.19 69758.31 69775.75
##        1        1        1        1        1        1        1        1
## 69784.85  69805.7 69868.48 69869.66 69874.18 70005.51 70012.83 70053.27
##        1        1        1        1        1        1        1        1
## 70179.11 70185.06 70203.74  70225.6 70232.95  70324.8 70377.23 70410.11
##        1        1        1        1        1        1        1        1
## 70449.04  70492.6 70495.64 70505.06 70510.59 70547.16  70575.6 70582.55
##        1        1        1        1        1        1        1        1
## 70592.81 70701.31 70783.94 70889.68 71055.22 71136.49 71157.05  71222.4
##        1        1        1        1        1        1        1        1
## 71228.44 71296.67 71384.57 71392.53 71455.62 71511.08 71718.51 71727.51
##        1        1        1        1        1        1        1        1
## 71881.84 72042.85 72154.68  72188.9 72196.29 72203.96 72209.99 72270.88
##        1        1        1        1        1        1        1        1
##  72272.9 72325.91 72330.57 72423.97 72524.86 72553.94 72683.35 72684.44
##        1        1        1        1        1        1        1        1
```

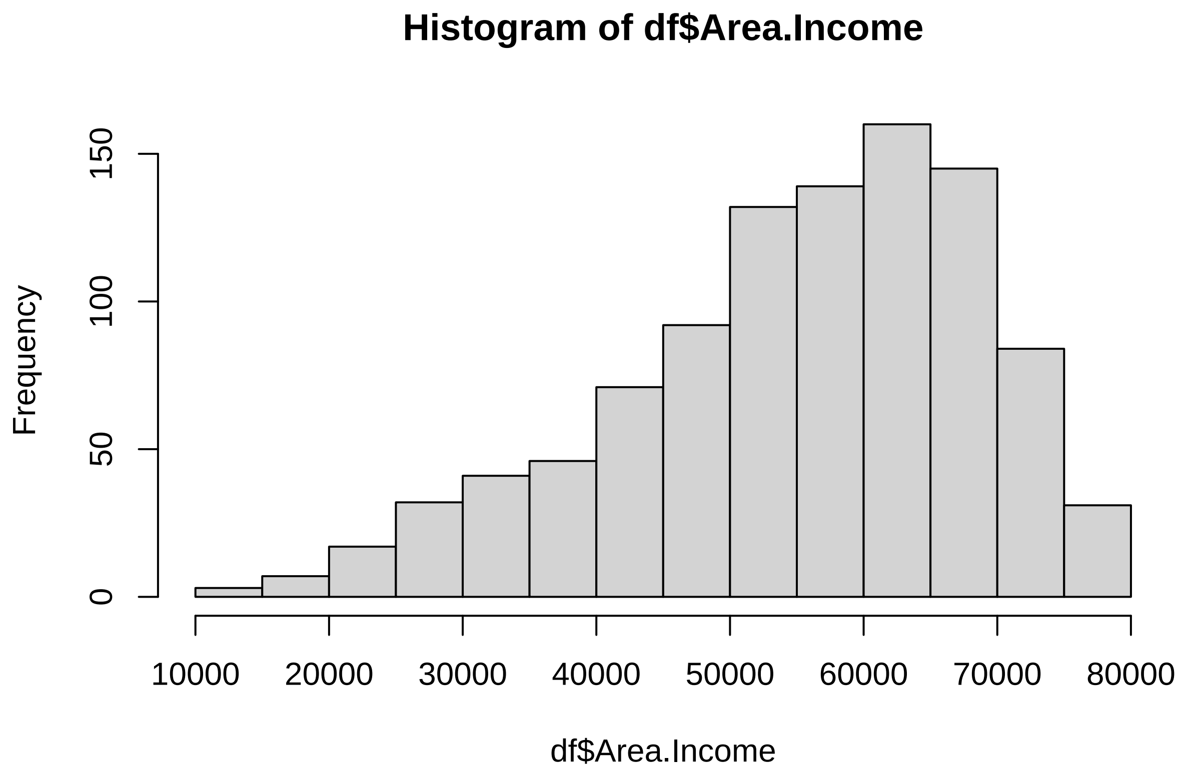
```
## 72707.87 72802.42 72948.76  73049.3 73104.47 73174.19 73207.15 73234.87
##        1        1        1        1        1        1        1        1
## 73347.67 73392.28 73413.87 73474.82 73538.09 73600.28 73608.99  73687.5
##        1        1        1        1        1        1        1        1
## 73863.25 73882.91 73884.48 73889.99  73910.9 73941.91 74024.61 74159.69
##        1        1        1        1        1        1        1        1
## 74166.24 74180.05 74430.08 74445.18 74535.94 74543.81 74623.27 74780.74
##        1        1        1        1        1        1        1        1
## 74903.41 75044.35  75180.2 75254.88 75265.96 75509.61 75524.78 75535.14
##        1        1        1        1        1        1        1        1
## 75560.65 75687.46 75769.82 75805.12 76003.47 76246.96 76368.31 76408.19
##        1        1        1        1        1        1        1        1
##  76435.3 76480.16 76560.59 76893.84 76984.21 77143.61 77220.42 77460.07
##        1        1        1        1        1        1        1        1
## 77567.85 77871.75 77988.71 78092.95  78119.5 78520.99 79332.33  79484.8
##        1        1        1        1        1        1        1        1
```

```
# Histogram showing the age
hist(df$Age)
```



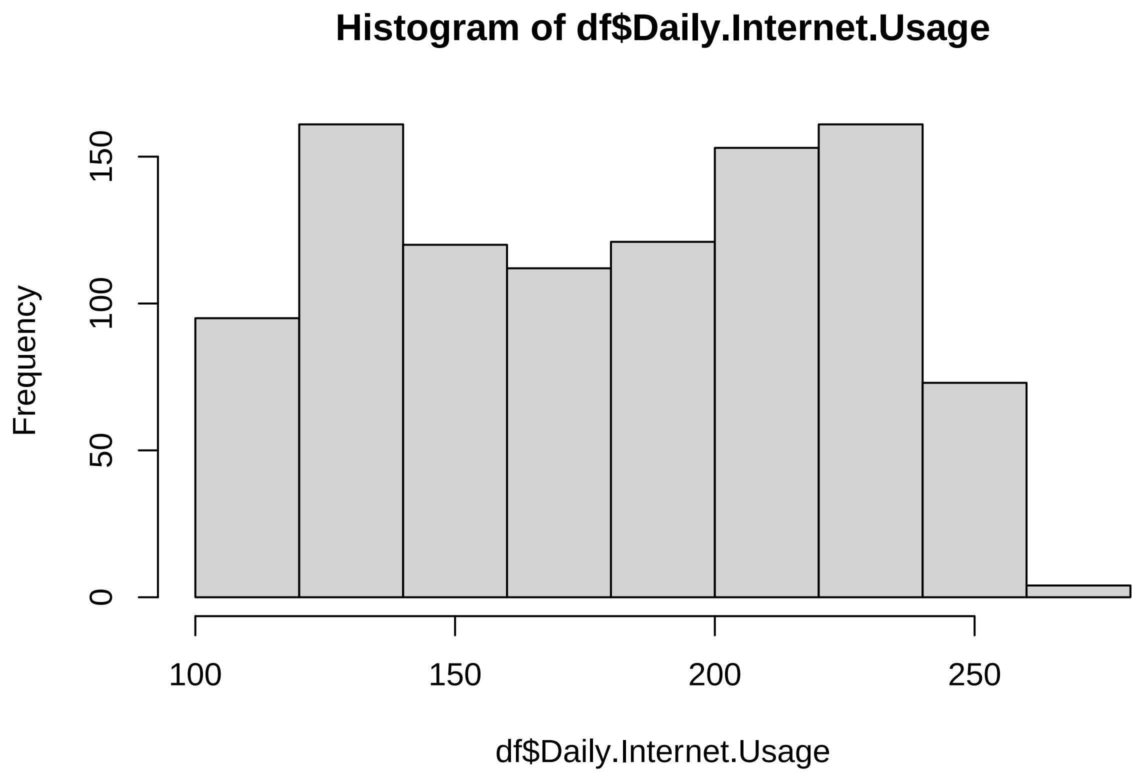
Most people range from the ages 25-40

```
# Histogram showing the income
hist(df$Area.Income)
```



Most people earn an income of 55000 to 70000

```
# Histogram showing the daily internet usage
hist(df$Daily.Internet.Usage)
```



The most frequency of internet usage is around 120

```r
# Frequency table for internet usage column
table(df$Daily.Internet.Usage)
```

```
##
## 104.78    105 105.04 105.15 105.22 105.63 105.69 105.71 105.86 105.94 106.04
##      1      1      1      1      1      1      1      1      1      1      1
## 106.86 106.96 107.19 107.56 107.92 108.03  108.1 108.15 108.16 108.17 108.18
##      1      1      1      1      1      1      1      1      1      1      1
## 108.25 108.27  108.7 108.85    109 109.04 109.07 109.22 109.29 109.34 109.77
##      1      1      1      1      1      1      1      1      1      1      1
## 109.98 110.25 110.57 110.66 110.68 110.84 110.93 111.02 111.59 111.63 111.71
##      1      1      1      1      1      1      1      1      1      1      1
##  111.8 111.94 112.19 112.52 112.72 113.12 113.53 113.69  113.7 113.75  113.8
##      1      1      1      1      1      1      2      1      1      1      1
## 114.53 114.69 114.85 115.26 115.35 115.37  115.6 115.79 115.91 116.07 116.19
##      1      1      1      1      1      1      1      1      2      1      1
## 116.27 116.38 116.53  117.3 117.33 117.35 117.66 117.75  118.1 118.16 118.27
##      1      1      1      2      1      1      1      1      1      1      1
## 118.39 118.45  118.6 118.69 119.03  119.2 119.27  119.3 119.32 119.47 119.65
##      1      1      1      1      1      1      1      2      1      1      1
## 119.84 119.86 119.93 120.06 120.12 120.25 120.37 120.46 120.49 120.63 120.75
##      1      1      1      2      1      1      1      1      1      1      1
## 120.85  120.9 120.95 121.05 121.07 121.24 121.28 121.57 121.81 122.02 122.04
##      1      1      1      1      1      1      1      1      1      1      1
## 122.31 122.45 122.59 123.08 123.13 123.22 123.24 123.25 123.28 123.51 123.62
```

```
##      1      1      1      1      1      1      1      1      1      1      1 
## 123.64 123.71 123.72  123.8 123.86 123.91 124.32 124.34 124.38 124.44 124.54 
##      1      1      1      1      1      1      1      1      1      1      1 
## 124.58 124.61 124.67 124.85 125.11 125.12  125.2 125.22 125.27 125.45 125.46 
##      1      1      1      1      1      1      1      1      1      2      1 
## 125.65 125.85 125.94 126.11 126.29 126.39 126.44 126.95 126.97 127.01 127.07 
##      1      1      1      1      1      1      1      1      1      1      1 
## 127.11  127.2 127.26 127.37 127.56 127.65 127.82 127.83    128 128.16 128.17 
##      1      1      1      1      1      1      1      1      1      1      1 
## 128.37 128.48 128.62 128.95 128.98 129.01 129.16 129.23 129.25 129.31 129.33 
##      1      1      1      1      1      1      1      1      1      1      1 
## 129.41  129.8 129.88  130.4 130.41 130.83 130.84 130.86 131.29 131.55 131.68 
##      1      1      1      1      1      1      1      1      1      1      1 
## 131.72 131.76 131.98 132.07 132.08 132.27 132.31 132.38 132.55 132.63 132.66 
##      1      1      1      1      1      1      1      2      1      1      1 
## 132.71 133.17 133.18  133.2 133.42 133.81  133.9 133.99 134.14 134.42 134.46 
##      1      1      1      1      1      1      1      1      1      1      1 
##  134.6 134.88 135.08 135.18 135.24 135.25 135.31 135.48 135.51 135.67 135.72 
##      1      1      1      1      2      1      1      1      1      1      1 
## 136.18 136.21  136.4 136.59 136.64 136.85 136.94 136.99  137.2 137.24 137.28 
##      2      1      1      1      1      1      1      1      1      1      1 
## 137.43 137.63 137.97 138.35 138.46 138.52 138.55 138.68 138.71 138.87 139.01 
##      1      1      1      2      1      1      1      1      1      1      1 
## 139.02 139.32 139.34 139.42 140.15 140.39 140.46 140.64 140.67 140.77 140.83 
##      1      1      1      1      1      1      1      1      1      1      1 
## 140.95 141.13 141.22 141.34 141.36 141.52 141.58 141.89 141.96 142.04 142.21 
##      1      1      1      1      1      1      1      1      1      1      1 
## 142.23 142.67 142.81 143.04 143.13 143.42 143.56 143.79 143.94 144.27 144.53 
##      1      1      1      1      1      1      1      1      1      1      1 
## 144.62 144.69 144.71 144.77 145.08 145.48 145.73 145.85 145.96 145.98 146.13 
##      1      1      1      1      1      1      1      1      1      1      1 
## 146.19 146.44  146.8 147.61 147.64 147.75 147.92 148.19 148.61 148.93  149.2 
##      1      1      1      1      1      1      1      1      1      1      1 
## 149.21 149.25 149.53 149.67 149.79  149.8 150.29 150.61 150.77 150.79  150.8 
##      1      1      1      1      1      1      1      1      1      1      1 
## 150.83 150.84 150.99 151.12 151.18 151.25 151.47 151.54 151.63 151.72 151.93 
##      1      1      1      1      1      1      1      1      1      1      1 
## 151.94 151.95 151.96 152.24 152.36 152.49 152.86 152.94 153.01 153.12 153.17 
##      1      1      1      1      1      1      1      1      1      1      1 
## 153.69 153.76 153.98    154 154.02 154.23 154.74 154.75 154.77 154.93 154.97 
##      1      1      1      1      1      1      1      1      1      1      1 
##  155.8 156.11  156.3 156.36 156.48 156.54 156.97 156.99 157.04 158.03 158.05 
##      1      1      1      1      1      1      1      1      1      1      1 
## 158.22 158.29 158.35 158.42 158.56  158.8 158.81 159.05 159.24 159.46  159.6 
##      2      1      1      1      1      1      1      1      1      1      1 
## 159.69 159.77 160.03 160.33 160.49 160.73 160.74 161.16 161.24 161.29 161.42 
##      1      1      1      1      1      1      1      2      1      1      1 
## 161.58 161.77 161.79 162.03 162.05 162.08 162.43 162.44 162.46 162.95    163 
##      1      1      1      1      1      1      1      2      1      1      1 
## 163.05 163.38 163.48 163.99 164.02 164.25 164.63 164.83 165.27 165.43 165.52 
##      1      1      1      1      1      2      1      1      1      1      1 
## 165.56 165.62 165.65 166.19 166.29 166.31 166.85 166.86 167.07 167.22 167.26 
##      1      1      1      1      1      1      1      1      1      2      1 
## 167.41 167.42 167.67 167.86 167.87    168 168.15 168.27 168.29 168.34 168.41 
```

```
##      1      1      1      1      1      1      1      1      1      1      1
## 168.92  169.1 169.18 169.23  169.4  169.5 169.88 170.04 170.13 170.49  170.9
##      1      1      1      1      2      1      1      1      1      1      1
## 171.07 171.23 171.24 171.31 171.54 171.62 171.72  171.9  172.1 172.57 172.58
##      1      1      1      1      1      1      1      1      1      1      1
## 172.81 172.83 173.01 173.05 173.43 173.49 173.75 174.55 174.88 175.14 175.17
##      1      1      1      1      1      1      1      1      1      1      1
## 175.37 175.43 176.28 176.52  176.7 176.73 176.78 176.98 177.46 177.55 177.78
##      1      1      1      1      1      1      1      1      1      1      1
## 178.35 178.51 178.69 178.75 178.85 178.92 179.04 179.58 179.82 180.42 180.47
##      1      1      1      2      1      1      1      1      1      1      1
## 180.77 180.88 181.02 181.11 181.25 182.11  182.2 182.65 182.84 183.42 183.43
##      1      1      1      1      1      1      1      2      1      1      1
## 183.48 183.82 183.85 184.03  184.1 184.23 184.88 184.94 184.98 185.45 185.46
##      1      1      1      1      1      1      1      1      1      1      1
## 185.47 185.85 186.37 186.48 186.98 187.03 187.09 187.36 187.53 187.64 187.76
##      1      1      1      1      1      1      1      1      1      1      1
## 187.95 188.27 188.32 188.56 189.91 190.05 190.08 190.12 190.17 190.25 190.41
##      1      1      1      1      1      1      1      1      1      1      1
## 190.71 190.75 190.84 190.95 191.14 191.17 191.26 191.78 191.82 192.27  192.5
##      1      1      1      2      1      1      1      1      1      1      1
## 192.57  192.6 192.81 192.85 192.93 193.15 193.29 193.58  193.6 193.63 193.77
##      1      1      1      1      1      1      1      1      1      1      1
##  193.8 193.97 194.23 194.37 194.44 194.56 194.62 194.83 194.95 195.07 195.31
##      1      1      2      1      1      1      1      1      1      1      1
## 195.36 195.54 195.56 195.68 195.69 195.89 195.91 195.93 196.17 196.23 196.61
##      1      1      1      1      1      1      1      1      1      1      1
## 196.71 196.76 196.77 196.83 197.15 197.33 197.66 197.93 198.11 198.13 198.24
##      1      1      1      1      1      1      1      1      1      1      1
##  198.3 198.32 198.45  198.5 198.56 198.72 198.79 198.86 199.08 199.25 199.29
##      1      1      1      1      1      1      1      1      1      1      1
## 199.39  199.4 199.43 199.62 199.76 199.79 200.22 200.23 200.28 200.55 200.58
##      1      1      1      1      1      1      1      1      1      1      1
## 200.59 200.71 201.04 201.15 201.24 201.26 201.29 201.54 201.58 202.12 202.16
##      1      1      1      2      1      1      1      1      1      1      1
## 202.18 202.25 202.34 202.61  202.7 202.77  202.9 203.23  203.3 203.44 203.74
##      1      1      1      1      1      1      1      1      1      1      1
## 203.84 203.87  203.9 204.02 204.22 204.27  204.4 204.47 204.52 204.56 204.79
##      1      1      1      1      1      1      1      1      1      1      1
## 204.82 204.86 205.38  205.5 205.64 205.71 205.84 206.79 206.98 207.17 207.18
##      1      1      1      1      1      1      1      1      1      1      1
## 207.27 207.44 207.48 207.53 207.87 207.96 208.01 208.02 208.05 208.21 208.23
##      1      1      1      1      1      1      1      1      1      1      1
## 208.24 208.36 208.58 208.76 209.25 209.64 209.72 209.82 209.91 209.93 210.23
##      1      1      1      1      1      1      1      1      1      1      1
## 210.26 210.27 210.39 210.46 210.53 210.54  210.6 210.87 211.12 211.17 211.38
##      1      1      1      1      1      1      1      1      1      1      1
## 211.39 211.56 211.64 211.65 211.83 211.87  212.3 212.38 212.56 212.58 212.59
##      1      1      1      1      1      2      1      1      1      1      1
## 212.67 212.79 212.87 212.88 212.92 213.36 213.38  213.7 213.75 213.96 214.06
##      1      1      1      1      1      1      1      1      1      1      1
## 214.08 214.23 214.33 214.38 214.42 214.49 214.53 214.74 215.04 215.18 215.25
##      1      1      1      1      2      1      1      1      1      2      1
## 215.29 215.44 215.93    216 216.01 216.03 216.24 216.49  216.5 216.57 216.87
```

```
##       1      1      1      1      1      1      1      1      1      1      1
##  217.1 217.37 217.66 217.68 217.79 217.85 218.17 218.22 218.49 218.61 218.79
##       1      1      1      1      1      1      1      1      1      1      1
##  218.8 218.97 219.49 219.55 219.69 219.72 219.75 219.79 219.91 219.94 219.98
##       1      1      1      1      1      2      1      1      1      1      1
## 220.05 220.08 220.48 220.92 221.18 221.21 221.51 221.53 221.59 221.79 221.94
##       1      1      1      1      1      1      1      1      1      1      1
## 221.98 222.08 222.11 222.25 222.26 222.35 222.63 222.72 222.75 222.77 222.87
##       1      1      2      1      1      1      1      1      1      1      1
## 222.91 223.03 223.09 223.16  223.2 223.28 223.93 224.01 224.07  224.2 224.23
##       1      1      1      2      1      1      1      1      1      1      1
## 224.44 224.58 224.82  224.9 224.92 224.98    225 225.02 225.05 225.23 225.24
##       1      1      1      1      1      1      1      1      1      1      1
## 225.29 225.34 225.47 225.58 225.76 225.87 225.97 225.99 226.11 226.45 226.49
##       1      1      1      1      1      1      1      1      1      1      1
## 226.54 226.64 226.69 226.74 226.79 227.37 227.53 227.56 227.63 227.72 227.73
##       1      1      1      1      1      1      1      1      1      1      1
## 228.03  228.7 228.76 228.78 228.81 228.94 229.12 229.19 229.22 229.88 229.99
##       1      1      1      1      2      1      1      1      1      1      1
## 230.14 230.18 230.36 230.52 230.77 230.78 230.87  230.9 230.91 230.93 230.95
##       1      1      2      1      1      1      1      1      1      1      1
## 231.07 231.21 231.28 231.37 231.38 231.42 231.48 231.49 231.54 231.59 231.85
##       1      1      1      1      1      1      1      1      1      1      1
## 231.87 231.91 231.94 231.95 232.21 232.54 232.68 232.78 233.04 233.36 233.56
##       1      1      1      1      1      1      1      1      1      1      1
##  233.6 233.61 233.65 233.85 233.93 234.23 234.26 234.64 234.72 234.75 234.81
##       1      1      1      1      1      1      1      1      1      2      1
## 235.01 235.28 235.29 235.35 235.56 235.78 235.94 235.97 235.99 236.08 236.15
##       1      2      1      1      1      1      1      1      1      1      1
## 236.19 236.29  236.5 236.64 236.72 236.75 236.87 236.96 237.34 237.39 238.06
##       1      1      1      1      1      1      1      2      1      1      1
##  238.1 238.43 238.45 238.58 238.63 238.99 239.22 239.32 239.52 239.76 239.94
##       1      1      1      1      1      1      1      1      1      1      1
## 240.09 240.63 240.64 240.95 241.03 241.36 241.38  241.5  241.8 242.37 242.59
##       1      1      1      1      1      1      1      1      1      1      1
## 243.37 243.61 244.23 244.34  244.4 244.55 244.87 244.91  245.5 245.76 245.78
##       1      1      1      1      1      1      1      1      1      1      1
## 245.89 246.06 246.29 246.44 246.72 247.01 247.05 247.31  247.9 248.12 248.16
##       1      1      1      1      1      1      2      1      1      1      1
## 248.19 248.23 248.51 249.45 249.54 249.81 249.99    250 250.03 250.11  250.2
##       1      1      1      1      1      1      1      1      1      1      1
## 250.32 250.35 250.36    251 251.08 252.07 252.36  252.6 252.77 253.17 253.48
##       1      1      1      1      1      1      1      1      1      1      1
## 254.05 254.34 254.57 254.59 254.65 254.94 255.07 255.57 255.61 256.09 256.39
##       1      1      1      1      1      1      1      1      1      1      1
##  256.4 258.06 258.26 258.62 259.76 261.02 261.52 267.01 269.96
##       2      1      1      1      1      1      1      1      1
```

```
## Bivariate analysis
# Getting the covarience

cov(numeric)
```

```
##                           Daily.Time.Spent.on.Site           Age   Area.Income
## Daily.Time.Spent.on.Site               251.3370949 -4.617415e+01  6.613081e+04
```

```
## Age                                   -46.1741459  7.718611e+01 -2.152093e+04
## Area.Income                         66130.8109082 -2.152093e+04  1.799524e+08
## Daily.Internet.Usage                  360.9918827 -1.416348e+02  1.987625e+05
## Male                                   -0.1501864 -9.242142e-02  8.867509e+00
## Clicked.on.Ad                          -5.9331431  2.164665e+00 -3.195989e+03
##                          Daily.Internet.Usage        Male Clicked.on.Ad
## Daily.Time.Spent.on.Site         3.609919e+02 -0.15018639 -5.933143e+00
## Age                             -1.416348e+02 -0.09242142  2.164665e+00
## Area.Income                      1.987625e+05  8.86750903 -3.195989e+03
## Daily.Internet.Usage             1.927415e+03  0.61476667 -1.727409e+01
## Male                             6.147667e-01  0.24988889 -9.509510e-03
## Clicked.on.Ad                   -1.727409e+01 -0.00950951  2.502503e-01
```

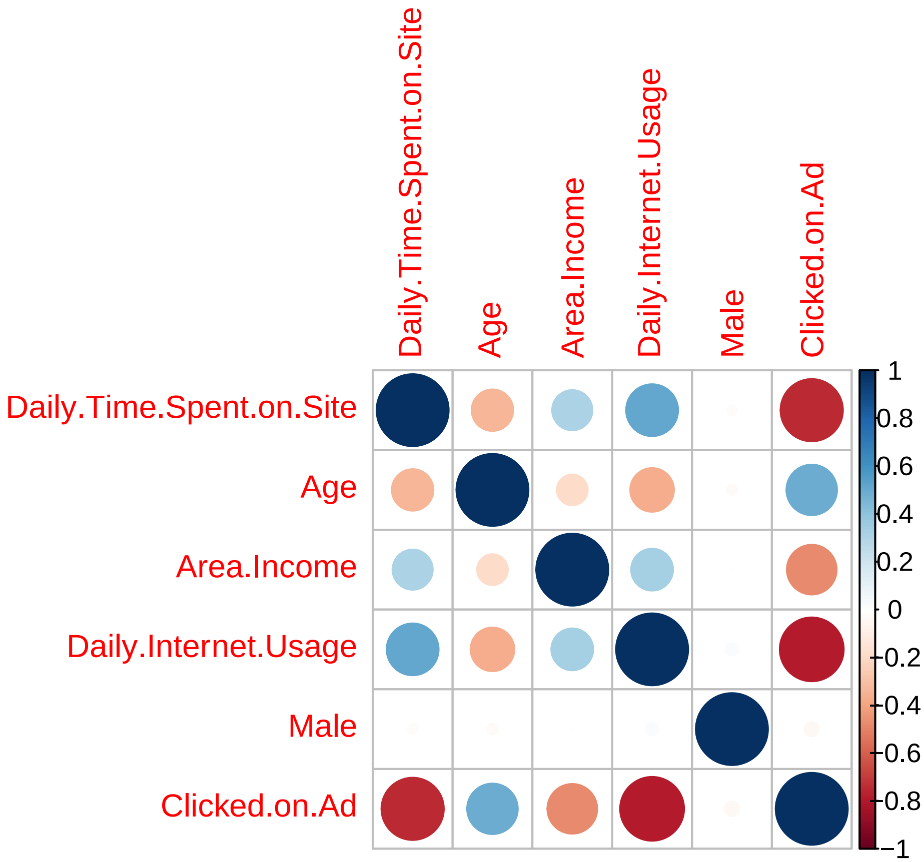
```
# Getting the correlation coefficient

matrix = cor(numeric)
matrix
```

```
##                          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
```

```
# getting the correlation matrix

library(corrplot)
```

```
## corrplot 0.92 loaded
```

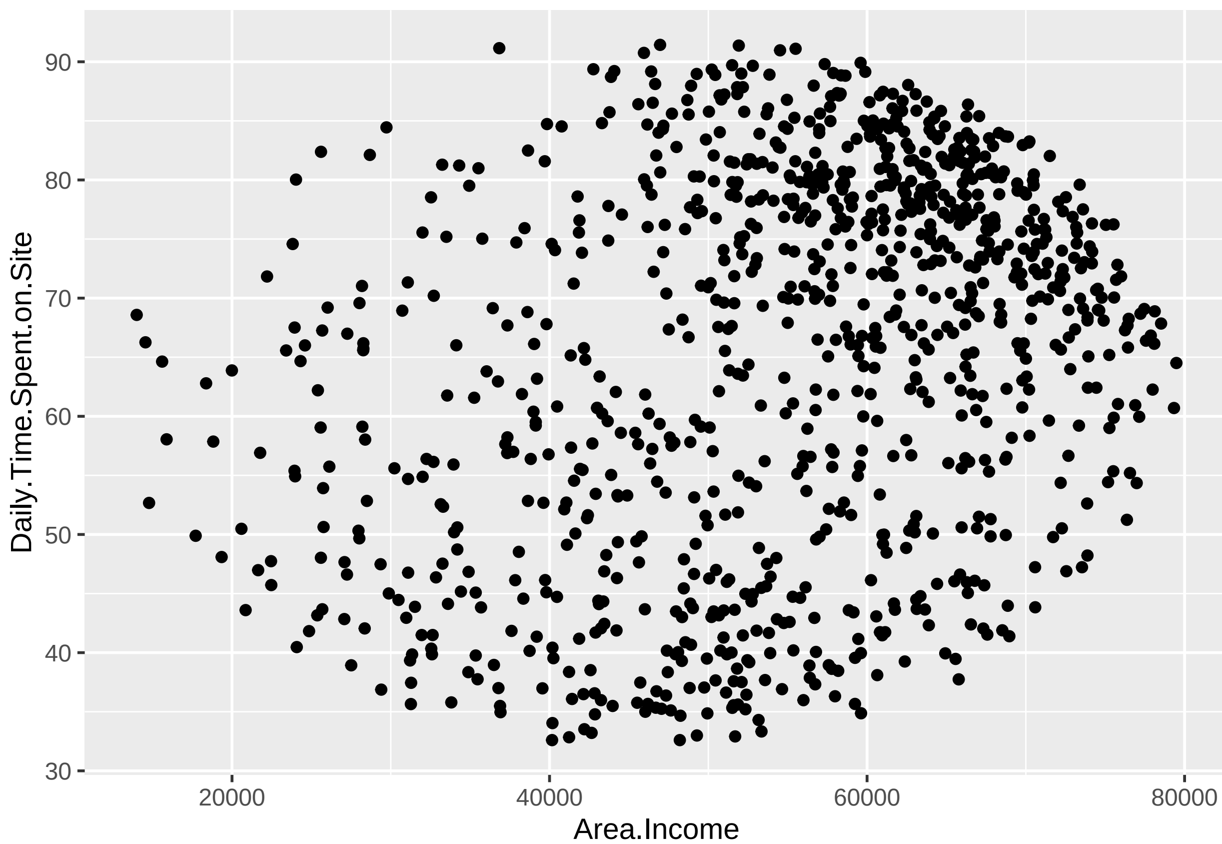
```
corrplot(matrix)
```




```r
# Change the data types of the categorical variables
df$'Clicked.on.Ad' <- as.factor(df$'Clicked.on.Ad')
df$Male <- as.factor(df$Male)
#Lets inspect the data type again
str(df)
```

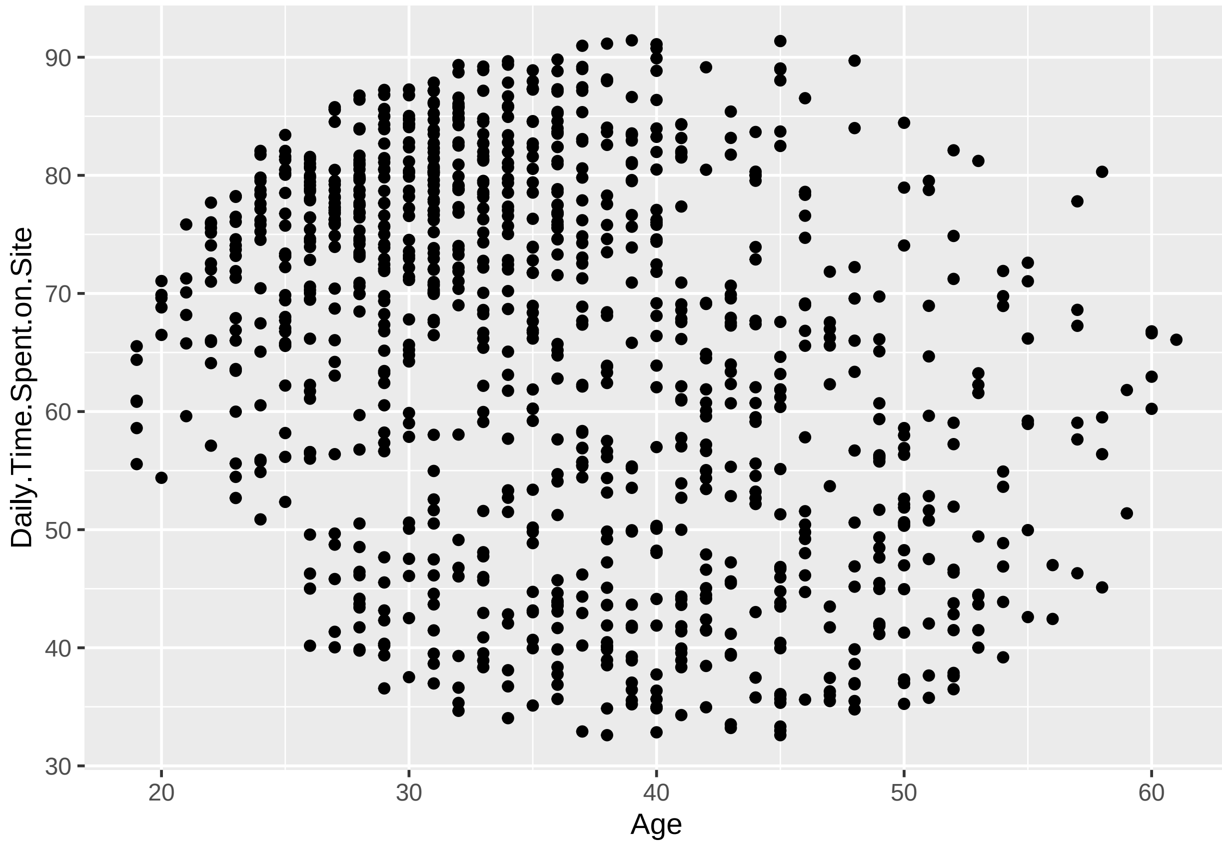
```
## 'data.frame':    1000 obs. of  10 variables:
##  $ Daily.Time.Spent.on.Site: num  69 80.2 69.5 74.2 68.4 ...
##  $ Age                     : int  35 31 26 29 35 23 33 48 30 20 ...
##  $ Area.Income             : num  61834 68442 59786 54806 73890 ...
##  $ Daily.Internet.Usage    : num  256 194 236 246 226 ...
##  $ Ad.Topic.Line           : chr  "Cloned 5thgeneration orchestration" "Monitored national standardi
##  $ City                    : chr  "Wrightburgh" "West Jodi" "Davidton" "West Terrifurt" ...
##  $ Male                    : Factor w/ 2 levels "0","1": 1 2 1 2 1 2 1 2 2 2 ...
##  $ Country                 : chr  "Tunisia" "Nauru" "San Marino" "Italy" ...
##  $ Timestamp               : chr  "2016-03-27 00:53:11" "2016-04-04 01:39:02" "2016-03-13 20:35:42"
##  $ Clicked.on.Ad           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...
```

```r
#scatter plot

library(ggplot2)


ggplot(df, aes(x=Area.Income, y=Daily.Time.Spent.on.Site)) +
  geom_point()
```



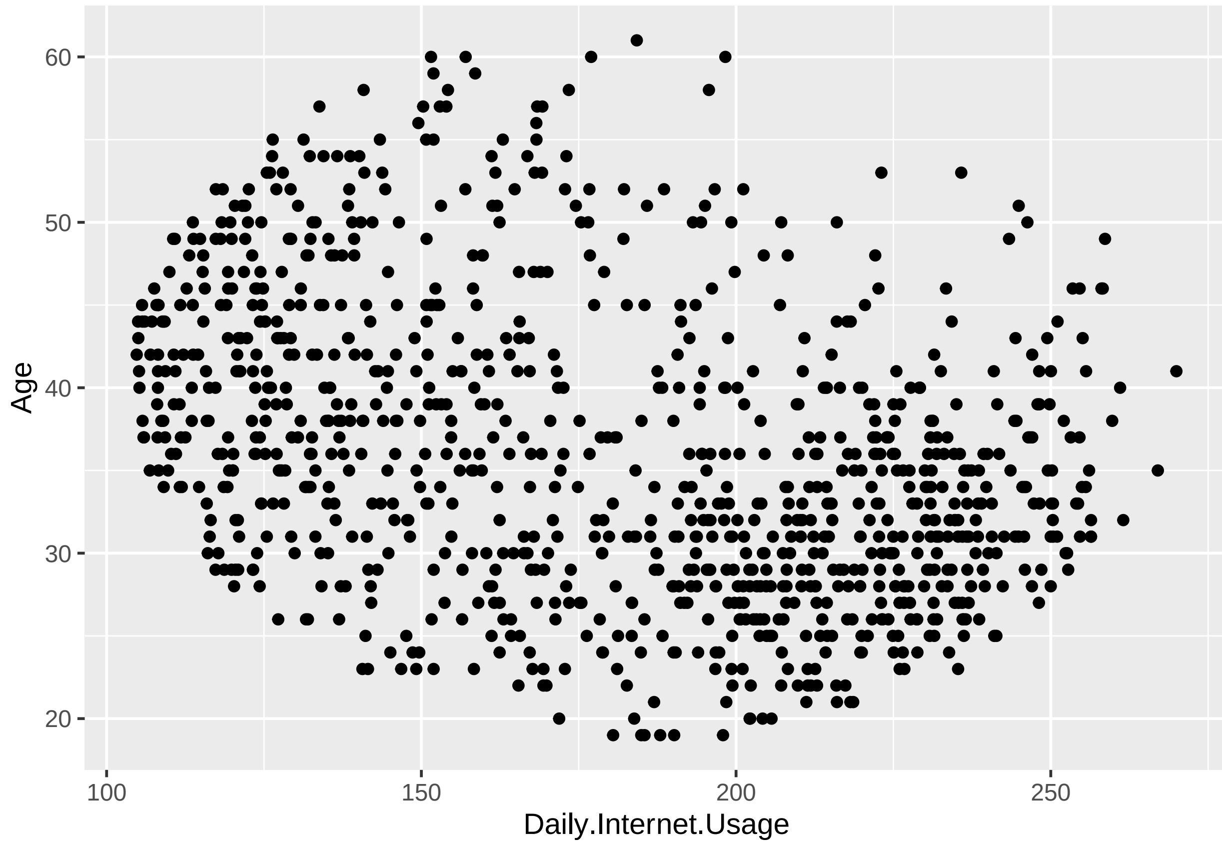
There is no correlation between daily time spent on site and the area income

```
ggplot(df, aes(x=Age, y=Daily.Time.Spent.on.Site)) +
  geom_point()
```



There is no correlation between daily time spent on site and the age

```
ggplot(df, aes(x=Daily.Internet.Usage
, y=Age
)) +
  geom_point()
```



There is no correlation between daily internet usage and the age

Modelling

```
head(df)
```

```
##   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
```

```r
# Dropping unecessary columns
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
df1 <- select(df, -c(Ad.Topic.Line
, Timestamp))
head(df1)
```

```
##   Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage           City
## 1                    68.95  35    61833.90               256.09    Wrightburgh
## 2                    80.23  31    68441.85               193.77      West Jodi
## 3                    69.47  26    59785.94               236.50       Davidton
## 4                    74.15  29    54806.18               245.89 West Terrifurt
## 5                    68.37  35    73889.99               225.58   South Manuel
## 6                    59.99  23    59761.56               226.74      Jamieberg
##   Male    Country Clicked.on.Ad
## 1    0    Tunisia             0
## 2    1      Nauru             0
## 3    0 San Marino             0
## 4    1      Italy             0
## 5    0    Iceland             0
## 6    1     Norway             0
```

```r
# Checking the structure if the data
str(df1)
```

```
## 'data.frame':    1000 obs. of  8 variables:
##  $ Daily.Time.Spent.on.Site: num  69 80.2 69.5 74.2 68.4 ...
##  $ Age                     : int  35 31 26 29 35 23 33 48 30 20 ...
##  $ Area.Income             : num  61834 68442 59786 54806 73890 ...
##  $ Daily.Internet.Usage    : num  256 194 236 246 226 ...
##  $ City                    : chr  "Wrightburgh" "West Jodi" "Davidton" "West Terrifurt" ...
##  $ Male                    : Factor w/ 2 levels "0","1": 1 2 1 2 1 2 1 2 2 2 ...
##  $ Country                 : chr  "Tunisia" "Nauru" "San Marino" "Italy" ...
##  $ Clicked.on.Ad           : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 2 1 1 ...
```

```r
# Label encode the categorical values in the dataframe
library(superml)
```

```
## Loading required package: R6
```

```r
label <- LabelEncoder$new()

df1$City <- label$fit_transform(df1$City)
df1$Country <- label$fit_transform(df1$Country)



head(df1)
```

```
##   Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage City Male
## 1                    68.95  35    61833.90               256.09    0    0
## 2                    80.23  31    68441.85               193.77    1    1
## 3                    69.47  26    59785.94               236.50    2    0
## 4                    74.15  29    54806.18               245.89    3    1
## 5                    68.37  35    73889.99               225.58    4    0
## 6                    59.99  23    59761.56               226.74    5    1
##   Country Clicked.on.Ad
## 1       0             0
## 2       1             0
## 3       2             0
## 4       3             0
## 5       4             0
## 6       5             0
```

We label encoded the categorical values so that they can be converted to a machine readable form

```
library(rpart)
library(mlbench)
library(caret)
```

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

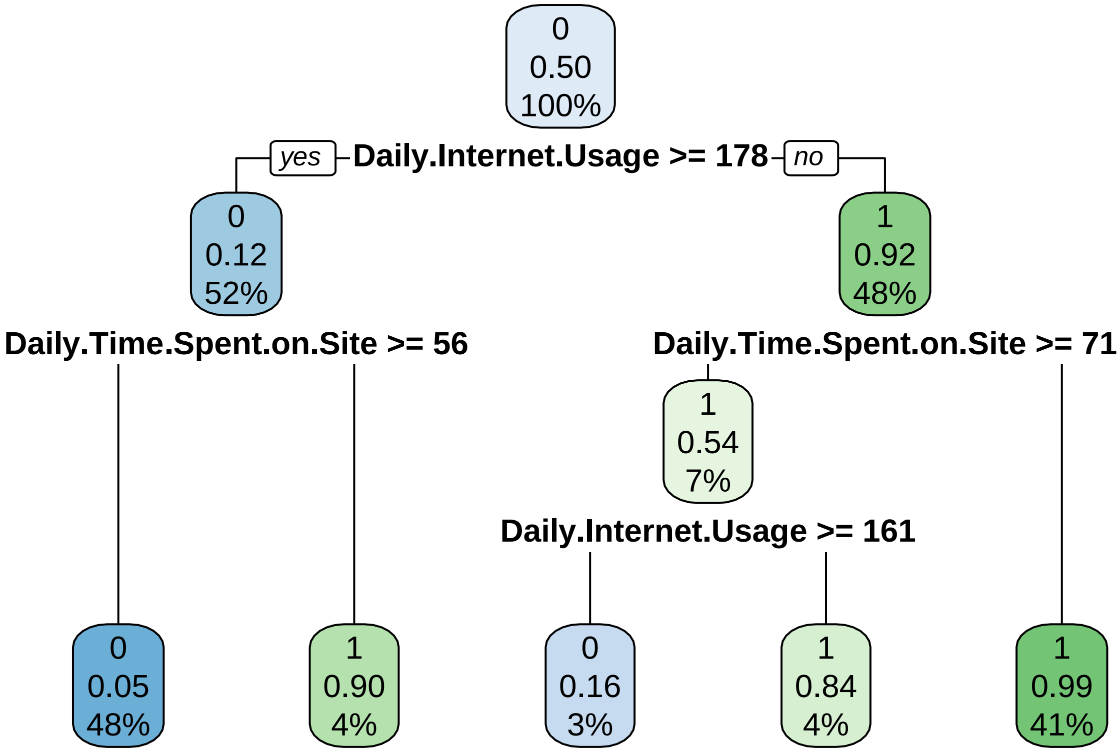
```
library(caretEnsemble)
```

```
##
## Attaching package: 'caretEnsemble'
```

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

```
# SVM
# Fitting the model
# Specifying the target and predictor variables
m <- rpart(Clicked.on.Ad ~ ., data = df1,
method = "class")
```

```
# Plotting the decision tree model
library(rpart.plot)
rpart.plot(m)
```




```
# Making predictions
# Printing the confusion matrix
p <- predict(m, df1, type ="class")
table(p, df1$'Clicked.on.Ad')
```

```
##
## p     0   1
##   0 485  28
##   1  15 472
```

The confusion matrix shows that we have 485 + 472 right predictions while 28 + 15 predictions are wrong

```
# Printing the Accuracy
mean(df1$'Clicked.on.Ad' == p)
```

```
## [1] 0.957
```

There is a 97% accuracy in the SVM model

```
# Splitting the data set into the Training set and Test set
library(caTools)
set.seed(123)
split = sample.split(df1$'Clicked.on.Ad', SplitRatio = 0.7)
training = subset(df1, split == TRUE)
test = subset(df1, split == FALSE)
```

```
library(e1071)
classifier = svm(formula = Clicked.on.Ad ~ .,
data = training,
```

```r
type = 'C-classification',
kernel = 'linear')
classifier
```

```
##
## Call:
## svm(formula = Clicked.on.Ad ~ ., data = training, type = "C-classification",
##     kernel = "linear")
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  1
##
## Number of Support Vectors:  64
```

```r
# Predicting the Test set results
y_pred = predict(classifier, newdata = test[-8])
```

```r
# Making the Confusion Matrix
library("RSNNS")
```

```
## Loading required package: Rcpp
```

```
##
## Attaching package: 'RSNNS'
```

```
## The following objects are masked from 'package:caret':
##
##     confusionMatrix, train
```

```r
confusionMatrix(y_pred,as.factor(test[,8]))
```

```
##        predictions
## targets   1   2
##       1 149   8
##       2   1 142
```

The confusion matrix shows us that we have 149 + 142 correct predictions while 1 + 8 predictions are wrong The model shows us that we have a 97% accuracy level

Conclusion:

From the analysis carried out above, we can see the factors that contribute to a user clicking an ad are: Gender, Area Income, Daily Time spent on the site Both the SVM and decision tree classifier have a 97% accuracy level so they can both be used for prediction

Recommendation

Both genders should be targeted More data should be provided to give more insight