

Internet Marketing Analysis

Maureen Gatu

27/08/2021

Overview

Targeted advertising is a form of advertising, including online advertising, that is directed towards an audience with certain traits, based on the product or person the advertiser is promoting. These traits can either be demographic with a focus on race, economic status, sex, age, generation, level of education, income level, and employment, or there can be a psychographic focus which is based on the consumer values, personality, attitude, opinion, lifestyle and interest. This focus can also entail behavioral variables, such as browser history, purchase history, and other recent online activities. Targeted advertising is focused on certain traits and consumers who are likely to have a strong preference. These individuals will receive messages instead of those who have no interest and whose preferences do not match a particular product's attributes. This eliminates waste.

Defining the Question

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. In the past, she ran ads to advertise a related course on the same blog and collected data in the process. 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.

Metric of Success

Successfully identify the characteristics of individuals most likely to click on the ads.

Research Design

1. Defining the question
2. Loading and previewing the data
3. Data Cleaning
4. Data Analysis
5. Recommendations
6. Conclusion

The Data

Loading the Data

#setting up the enviroment

```
getwd()
```

```
## [1] "E:/Deasktop/Documents"
```

#Locating the dataset

```
setwd("A:\\PROGRAMMING WITH R\\advert")
```

#Loading the dataset

```
advt <- read.csv("advertising.csv", TRUE, ",")
```

Previewing the Data

```
#Preview the top of the dataset
```

```
head(advt)
```

##	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			

```
#Preview thr bottom of the dataset
```

```
tail(advt)
```

##	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
```

Getting information about the dataset

Size of the dataset

#Size of the dataset

```
dim(advt)
```

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

The dataset has 1000 rows and 10 columns

Viewing the column names

```
names(advt)
```

```
## [1] "Daily.Time.Spent.on.Site" "Age"
## [3] "Area.Income"             "Daily.Internet.Usage"
## [5] "Ad.Topic.Line"           "City"
## [7] "Male"                    "Country"
## [9] "Timestamp"               "Clicked.on.Ad"
```

Checking the variables data types

```
sapply(advt, class)
```

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

Checking the number of unique entries in each variable

```
print(advt %>% summarise_all(n_distinct))
```

```
##      Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage
## Ad.Topic.Line
## 1              900  43          1000          966
```

```
1000
##   City Male Country Timestamp Clicked.on.Ad
## 1  969    2      237      1000            2
```

Data Cleaning

Checking for duplicates

#Checking the duplicates using duplicated.data.frame() function

```
dim(advt)

## [1] 1000   10

table(duplicated.data.frame(advt))

##
## FALSE
## 1000
```

The dataset has no duplicated records

Check for missing values

```
null <- advt[!complete.cases(advt),]           #Give total number of rows with
missing values
dim(null)

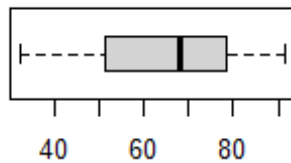
## [1]  0 10
```

The data has no incomplete rows i.e there are no missing values in the dataset. ####

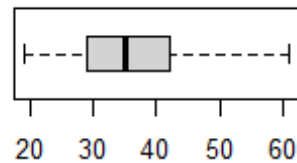
Outliers using the boxplot

```
par(mfrow = c(2,2))
for (i in 1:4){
  boxplot(advt[,i], main = names(advt)[i], horizontal = TRUE)
}
```

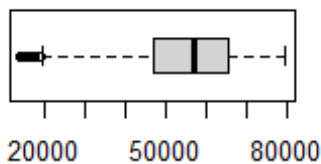
Daily.Time.Spent.on.Site



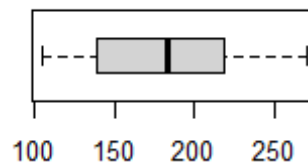
Age



Area.Income



Daily.Internet.Usage



There are a few outliers in the Area.income variable. We preview these outliers using the quantile method

```
lower_bound <- quantile(advt$Area.Income, 0.025)

# get the data index
lower_ind <- which(advt$Area.Income < lower_bound)

#Preview these data
lower <- advt[lower_ind, ]
lower                                     #Since the outlier is in regards to income, we
choose to retain it.

##      Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage
## 17                55.39    37    23936.86         129.41
## 20                74.58    40    23821.72         135.51
## 97                45.72    36    22473.08         154.02
## 131               46.98    50    21644.91         175.37
## 136               49.89    39    17709.98         160.03
## 220               43.60    38    20856.54         170.49
## 241               80.03    44    24030.06         150.84
## 310               54.92    54    23975.35         161.16
## 390               63.88    38    19991.72         136.85
## 411               48.09    33    19345.36         180.42
## 511               57.86    30    18819.34         166.86
## 603               71.83    40    22205.74         135.48
## 606               64.67    51    24316.61         138.35
```

## 641	64.63	45	15598.29	158.80
## 666	58.05	32	15879.10	195.54
## 680	65.57	46	23410.75	130.86
## 693	66.26	47	14548.06	179.04
## 769	68.58	41	13996.50	171.54
## 779	52.67	44	14775.50	191.26
## 810	67.51	43	23942.61	127.20
## 881	47.74	33	22456.04	154.93
## 902	40.47	38	24078.93	203.90
## 909	56.91	50	21773.22	146.44
## 953	62.79	36	18368.57	231.87
## 973	50.48	50	20592.99	162.43
##	Ad.Topic.Line			City Male
## 17	Customizable multi-tasking website			West Dylanberg 0
## 20	Advanced 24/7 productivity			Millertown 1
## 97	Versatile homogeneous capacity			Williammouth 1
## 131	Down-sized well-modulated archive			East Michelleberg 0
## 136	Enhanced system-worthy application			East Michele 1
## 220	Virtual bandwidth-monitored initiative			North Ricardotown 0
## 241	Automated static concept			Christinetown 0
## 310	Extended interactive model			Roberttown 0
## 390	Upgradable even-keeled hardware			Kristintown 0
## 411	Balanced motivating help-desk			West Travismouth 0
## 511	Horizontal modular success			Estesfurt 0
## 603	Diverse background ability			Costaburgh 1
## 606	Horizontal incremental website			Andersonfurt 1
## 641	Triple-buffered high-level Internet solution			Isaacborough 1
## 666	Total asynchronous architecture			Sanderstown 1
## 680	Implemented asynchronous application			Reginamouth 0
## 693	Optional full-range projection			Matthewtown 1
## 769	Exclusive discrete firmware			New Williamville 1
## 779	Persevering 5thgeneration knowledge user			New Hollyberg 0
## 810	Digitized homogeneous core			Lake Faith 0
## 881	Open-source 5thgeneration leverage			Henryland 1
## 902	Sharable 5thgeneration access			Fraziershire 0
## 909	Team-oriented executive core			West Randy 0
## 953	Total coherent archive			New James 1
## 973	Switchable real-time product			Dianaville 0
##	Country			Timestamp Clicked.on.Ad
## 17	Palestinian Territory			2016-01-30 19:20:41 1
## 20	Russian Federation			2016-02-27 04:43:07 1
## 97	Hong Kong			2016-04-19 15:14:58 1
## 131	Lithuania			2016-05-04 09:00:24 1
## 136	Belize			2016-04-16 12:09:25 1
## 220	Chile			2016-01-11 07:36:22 1
## 241	Afghanistan			2016-07-23 14:47:23 1
## 310	Saint Pierre and Miquelon			2016-06-13 13:59:51 1
## 390	Madagascar			2016-02-29 23:56:06 1
## 411	Heard Island and McDonald Islands			2016-05-28 12:38:37 1
## 511	Algeria			2016-07-08 17:14:01 1

```
## 603          Rwanda 2016-02-18 22:42:33      1
## 606          Togo  2016-02-14 16:33:29      1
## 641    Azerbaijan 2016-06-12 03:11:04      1
## 666    Tajikistan 2016-02-12 10:39:10      1
## 680      Belgium  2016-04-15 15:07:17      1
## 693      Lebanon  2016-04-25 19:31:39      1
## 769    El Salvador 2016-07-06 12:04:29      1
## 779      Jersey   2016-05-19 06:37:38      1
## 810    Western Sahara 2016-04-29 14:10:00      1
## 881      Saint Lucia 2016-05-14 14:49:05      1
## 902      Burundi  2016-07-22 07:44:43      1
## 909    Norfolk Island 2016-04-01 05:17:28      1
## 953      Luxembourg 2016-05-30 20:08:51      1
## 973      Malawi    2016-05-16 18:51:59      1
```

Range of the outlier income

```
range(lower$Area.Income)
```

```
## [1] 13996.50 24316.61
```

The outliers in the dataset are in the Area.Income variable which includes income between 13,996.50 upto 24,316.61. Since a person's income does affect their actions when interacting with website, the outliers are kept in orders to well see the influence of person's income on whether they clicked on an ad or not.

Exploratory Data Analysis

Univariate Analysis

Descriptive statistics of the dataset

```
summary(advt[c(1:10)])
```

```
##  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
##  Ad.Topic.Line      City      Male      Country
##  Length:1000      Length:1000      Min.   :0.000      Length:1000
##  Class :character      Class :character      1st Qu.:0.000      Class :character
##  Mode  :character      Mode  :character      Median :0.000      Mode  :character
##                               Mean   :0.481
##                               3rd Qu.:1.000
##                               Max.   :1.000
##  Timestamp      Clicked.on.Ad
```

```
## Length:1000      Min.   :0.0
## Class :character  1st Qu.:0.0
## Mode  :character  Median :0.5
##                      Mean  :0.5
##                      3rd Qu.:1.0
##                      Max.   :1.0
```

```
describe(advt)
```

```
## advt
##
## 10 Variables      1000 Observations
```

```
## -----
## Daily.Time.Spent.on.Site
##      n missing distinct    Info      Mean      Gmd      .05      .10
##    1000      0      900      1        65    18.11    37.58    41.34
##      .25      .50      .75      .90      .95
##    51.36    68.22    78.55    83.89    86.20
##
## lowest : 32.60 32.84 32.91 32.99 33.21, highest: 90.97 91.10 91.15 91.37
## 91.43
## -----
```

```
## -----
## Age
##      n missing distinct    Info      Mean      Gmd      .05      .10
##    1000      0      43    0.999    36.01    9.943    23.95    26.00
##      .25      .50      .75      .90      .95
##    29.00    35.00    42.00    49.00    52.00
##
## lowest : 19 20 21 22 23, highest: 57 58 59 60 61
## -----
```

```
## -----
## Area.Income
##      n missing distinct    Info      Mean      Gmd      .05      .10
##    1000      0    1000      1    55000    15037    28275    35223
##      .25      .50      .75      .90      .95
##   47032    57012    65471    70506    73601
##
## lowest : 13996.50 14548.06 14775.50 15598.29 15879.10
## highest: 78092.95 78119.50 78520.99 79332.33 79484.80
## -----
```

```
## -----
## Daily.Internet.Usage
##      n missing distinct    Info      Mean      Gmd      .05      .10
##    1000      0      966      1      180    50.63    113.5    120.5
##      .25      .50      .75      .90      .95
##    138.8    183.1    218.8    236.2    246.7
##
## lowest : 104.78 105.00 105.04 105.15 105.22, highest: 259.76 261.02 261.52
```


267.01 269.96

```
## -----
-----
## Ad.Topic.Line
##      n missing distinct
##    1000      0      1000
##
## lowest : Adaptive 24hour Graphic Interface      Adaptive asynchronous
attitude      Adaptive context-sensitive application Adaptive
contextually-based methodology Adaptive demand-driven knowledgebase
## highest: Visionary client-driven installation      Visionary maximized
process improvement Visionary mission-critical application Visionary multi-
tasking alliance      Visionary reciprocal circuit
## -----
-----
## City
##      n missing distinct
##    1000      0      969
##
## lowest : Adamsbury      Adamside      Adamsstad      Alanview
Alexanderfurt
## highest: Youngburgh      Youngfort      Yuton      Zacharystad
Zacharyton
## -----
-----
## Male
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000      0      2      0.749      481      0.481      0.4998
##
## -----
-----
## Country
##      n missing distinct
##    1000      0      237
##
## lowest : Afghanistan      Albania      Algeria      American
Samoa      Andorra
## highest: Wallis and Futuna Western Sahara      Yemen      Zambia
Zimbabwe
## -----
-----
## Timestamp
##      n missing distinct
##    1000      0      1000
##
## lowest : 2016-01-01 02:52:10 2016-01-01 03:35:35 2016-01-01 05:31:22 2016-
01-01 08:27:06 2016-01-01 15:14:24
## highest: 2016-07-23 05:21:39 2016-07-23 06:18:51 2016-07-23 11:46:28 2016-
07-23 14:47:23 2016-07-24 00:22:16
## -----
```

```

-----
## Clicked.on.Ad
##      n missing distinct      Info      Sum      Mean      Gmd
##    1000         0         2      0.75      500      0.5      0.5005
##
## -----
-----

#Change the data type of the categorical variables to factor for analysis
names <- c(5:10)
advt[,names] <- lapply(advt[,5:10], factor)
glimpse(advt)

## Rows: 1,000
## Columns: 10
## $ Daily.Time.Spent.on.Site <dbl> 68.95, 80.23, 69.47, 74.15, 68.37, 59.99,
## 88.~
## $ Age <int> 35, 31, 26, 29, 35, 23, 33, 48, 30, 20,
## 49, 3~
## $ Area.Income <dbl> 61833.90, 68441.85, 59785.94, 54806.18,
## 73889~
## $ Daily.Internet.Usage <dbl> 256.09, 193.77, 236.50, 245.89, 225.58,
## 226.7~
## $ Ad.Topic.Line <fct> Cloned 5thgeneration orchestration,
## Monitored~
## $ City <fct> Wrightburgh, West Jodi, Davidton, West
## Terrif~
## $ Male <fct> 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0,
## 0, ~
## $ Country <fct> Tunisia, Nauru, San Marino, Italy,
## Iceland, N~
## $ Timestamp <fct> 2016-03-27 00:53:11, 2016-04-04 01:39:02,
## 201~
## $ Clicked.on.Ad <fct> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0,
## 1, ~

summary(advt[,c(5:10)])

##           Ad.Topic.Line           City           Male
## Adaptive 24hour Graphic Interface : 1 Lisamouth : 3 0:519
## Adaptive asynchronous attitude : 1 Williamsport : 3 1:481
## Adaptive context-sensitive application : 1 Benjaminchester: 2
## Adaptive contextually-based methodology: 1 East John : 2
## Adaptive demand-driven knowledgebase : 1 East Timothy : 2
## Adaptive uniform capability : 1 Johnstad : 2
## (Other) :994 (Other) :986
##           Country           Timestamp           Clicked.on.Ad
## Czech Republic: 9 2016-01-01 02:52:10: 1 0:500
## France : 9 2016-01-01 03:35:35: 1 1:500
## Afghanistan : 8 2016-01-01 05:31:22: 1
## Australia : 8 2016-01-01 08:27:06: 1

```

```
## Cyprus      : 8   2016-01-01 15:14:24: 1
## Greece      : 8   2016-01-01 20:17:49: 1
## (Other)     :950 (Other)              :994
```

City: Lisamuth and Williamsport were the top 2 cities in the data set with both appearing 3 times in the data frame.

Country: Czech Republic and France were the two most popular countries with both appearing 9 times in the data.

Gender: 481 were male and 519 we female.

Clicked.on.Ad: half of the add titles were clicked on.

Mean

```
sapply(advt[,c(1:4)], mean)
```

```
## Daily.Time.Spent.on.Site      Age      Area.Income
##           65.0002           36.0090      55000.0001
##   Daily.Internet.Usage
##           180.0001
```

Standard deviation

```
sapply(advt[,c(1:4)], sd)
```

```
## Daily.Time.Spent.on.Site      Age      Area.Income
##           15.853615           8.785562      13414.634022
##   Daily.Internet.Usage
##           43.902339
```

Variance

```
sapply(advt[,c(1:4)], var)
```

```
## Daily.Time.Spent.on.Site      Age      Area.Income
##           2.513371e+02           7.718611e+01      1.799524e+08
##   Daily.Internet.Usage
##           1.927415e+03
```

Daily.Time.Spent.on.Site: The mean amount of time spent on the site was 65.002 with a standard deviation of 15.8536 and a variance of 2.513371e+02.

Age: The mean age of person was 36years with a standard deviation of 8 years and a variance of 7.718611e+01.

Area.Income:The mean Area.Income was 55,000 with a standard deviation of 13414.634 and a variance of 1.799524e+08.

Daily.Internet.Usage: The mean internet usage was 180 with a standard deviation of 43.90 and a variance of 1.927415e+03.

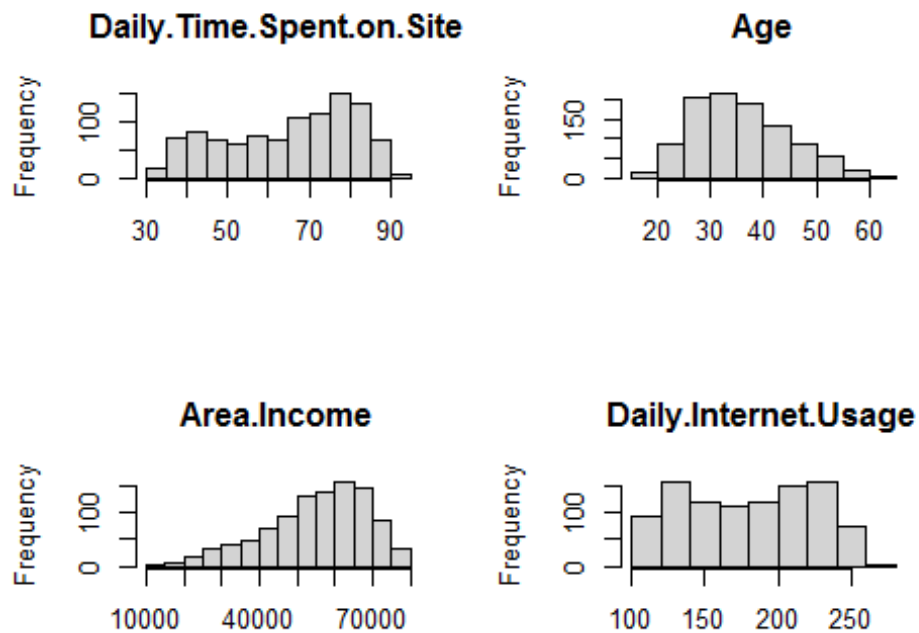
Range

```
sapply(advt[,c(1:4)], IQR)
```

##	Daily.Time.Spent.on.Site	Age	Area.Income
##	27.1875	13.0000	18438.8325
##	Daily.Internet.Usage		
##	79.9625		

Data distribution(Histogram)

```
par(mfrow = c(2,2))
for (i in 1:4){
  hist(advt[,i],main = names(advt)[i], xlab = NULL)
}
```



Skewness

```
skew <- apply(advt[,c(1:4)],2, skewness)
skew
```

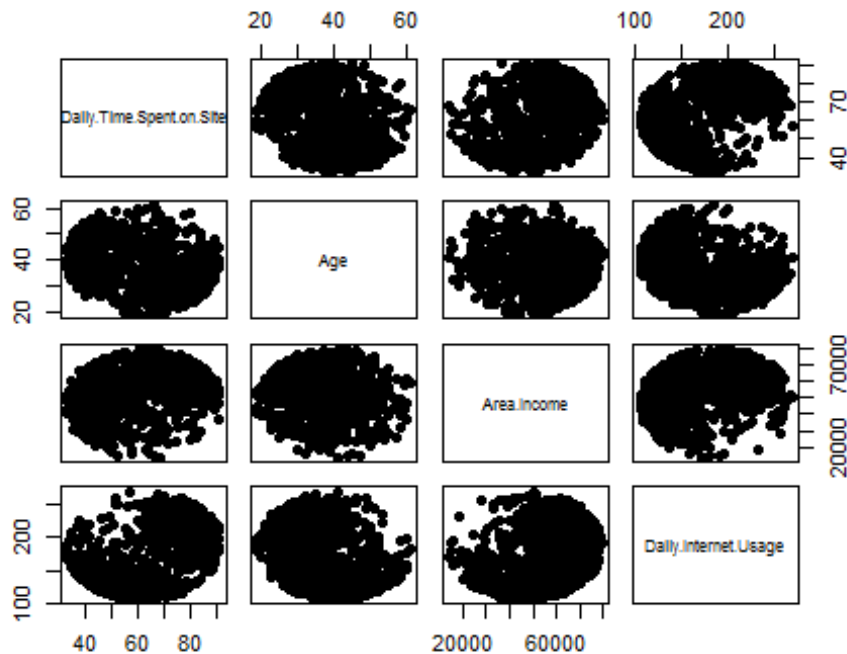
##	Daily.Time.Spent.on.Site	Age	Area.Income
##	-0.37064595	0.47770522	-0.64842285
##	Daily.Internet.Usage		
##	-0.03343681		

Daily.Time.Spent.on.Site, Age, Daily.Internet.Usage all have an approximately symmetrical distribution. Area.Income distribution is moderately skewed to the right.

Correlation

Scatter plots

```
num <- advt[,1:4]
pairs(num, pch = 19)
```



Correlation coefficients

```
num <- advt[,1:4]
cor(num)

##               Daily.Time.Spent.on.Site      Age Area.Income
## Daily.Time.Spent.on.Site      1.0000000 -0.3315133  0.3109544
## Age                          -0.3315133  1.0000000 -0.1826050
## Area.Income                   0.3109544 -0.1826050  1.0000000
## Daily.Internet.Usage          0.5186585 -0.3672086  0.3374955
##               Daily.Internet.Usage
## Daily.Time.Spent.on.Site      0.5186585
## Age                          -0.3672086
## Area.Income                   0.3374955
## Daily.Internet.Usage          1.0000000
```

Daily time spent on site, area.income and daily internet usage are all positively correlated to each other.

Age has a negative correlation to each of these three variable.

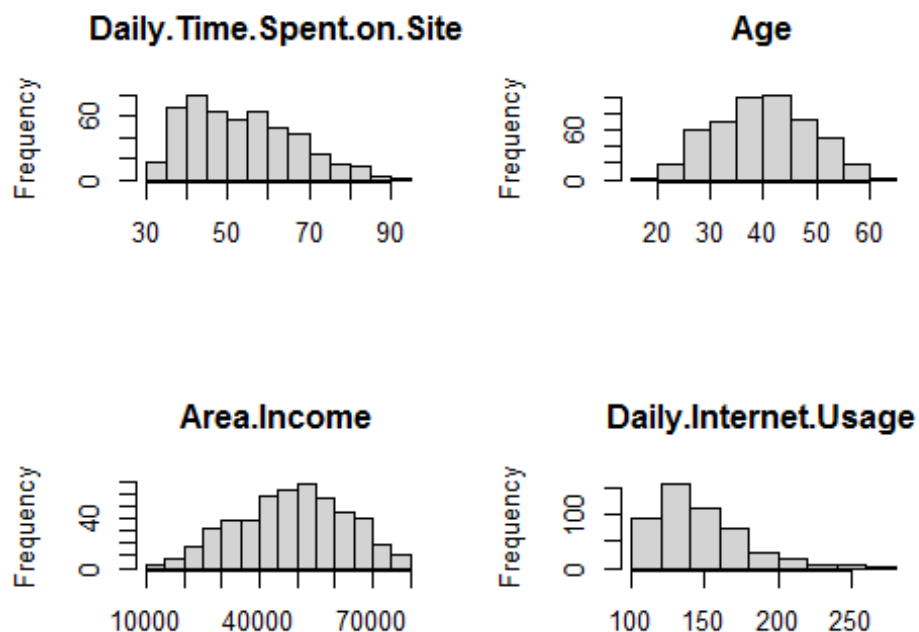
The correlation between all variables is weak with the highest correlation being between time spent on site and daily internet usage, with a correlation on 0.519.

Clicked on ads

Separate between ads that were clicked on(1) and those that were not.

created a subset data of only ads that were clicked on to analyze the target customers that click on the ads.

```
clicked <- advt[advt$Clicked.on.Ad == 1,]
par(mfrow = c(2,2))
for (i in 1:4){
  hist(clicked[,i],main = names(clicked)[i], xlab = NULL)
}
```



There is a difference in the histograms. Looking at the clicked on ads, the age peaked between ages 35 to 45 unlike in the general population that peaked only between ages 30 to 35.

In the daily internet usage, we see that most of the ads clicked were by people who spent less on internet with the data now clearly skewed to the left.

Looking at the time spent on the site. We can see a difference in the frequencies where in the general population the amount of time spent on site peaked between 75 and 80 but looking at those that clicked on the ad, the time peaked at 40 to 45.

```
summary(clicked[,c(5:10)])
```

```
##                               Ad.Topic.Line           City      Male
## Adaptive asynchronous attitude      : 1  Lake David    : 2    0:269
## Adaptive context-sensitive application : 1  Lake James    : 2    1:231
## Adaptive contextually-based methodology: 1  Lisamouth      : 2
## Adaptive demand-driven knowledgebase   : 1  Michelleside   : 2
## Adaptive uniform capability            : 1  Millerbury     : 2
## Advanced 24/7 productivity             : 1  Robertfurt     : 2
## (Other)                               :494  (Other)        :488
##           Country                    Timestamp Clicked.on.Ad
## Australia      : 7  2016-01-01 15:14:24: 1    0: 0
## Ethiopia       : 7  2016-01-01 20:17:49: 1    1:500
## Turkey         : 7  2016-01-02 12:25:36: 1
## Liberia        : 6  2016-01-03 03:22:15: 1
## Liechtenstein: 6  2016-01-03 04:39:47: 1
## South Africa   : 6  2016-01-03 05:34:33: 1
## (Other)        :461  (Other)           :494
```

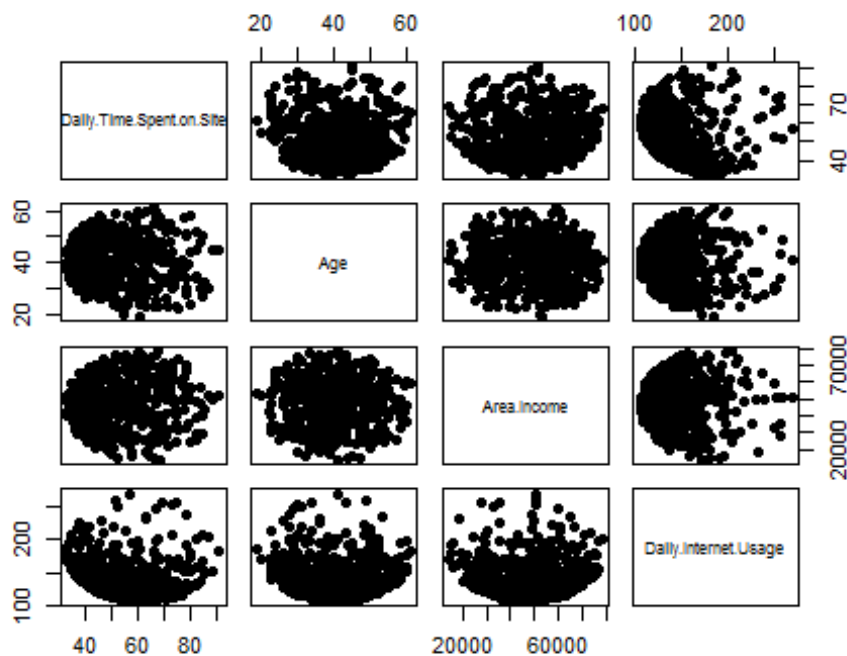
The above summary shows that the top cities that clicked on the ad were Lake David and Lake James and the top countries are Australia Ethiopia and Turkey.(None of these locations were among the top locations in the general population analysis.)

231 of the people that clicked on the ads were male and 269 were female.

Correlation

Scatter plots

```
num_c <- clicked[,1:4]
pairs(num_c, pch = 19)
```



#Correlationcoefficients

```
cor(num_c)
```

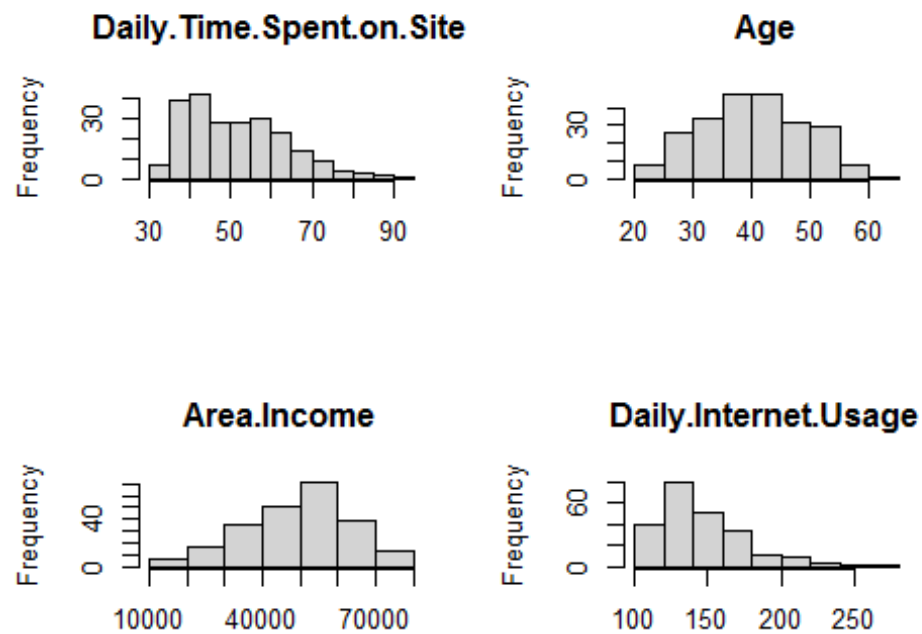
```
##           Daily.Time.Spent.on.Site      Age  Area.Income
## Daily.Time.Spent.on.Site      1.000000000 -0.01280025  0.007982346
## Age                          -0.012800250  1.000000000 -0.023701770
## Area.Income                  0.007982346 -0.02370177  1.000000000
## Daily.Internet.Usage         -0.170916216 -0.05693449 -0.010679858
##           Daily.Internet.Usage
## Daily.Time.Spent.on.Site      -0.17091622
## Age                          -0.05693449
## Area.Income                  -0.01067986
## Daily.Internet.Usage          1.00000000
```

Among the ads clicked, all the above variables had a weak negative correlation to each other except for Area.Income and Daily.Time.Spent.on.Site which maintained a weak positive correlation.

Gender analysis on the ads clicked on

Males who clicked on the add

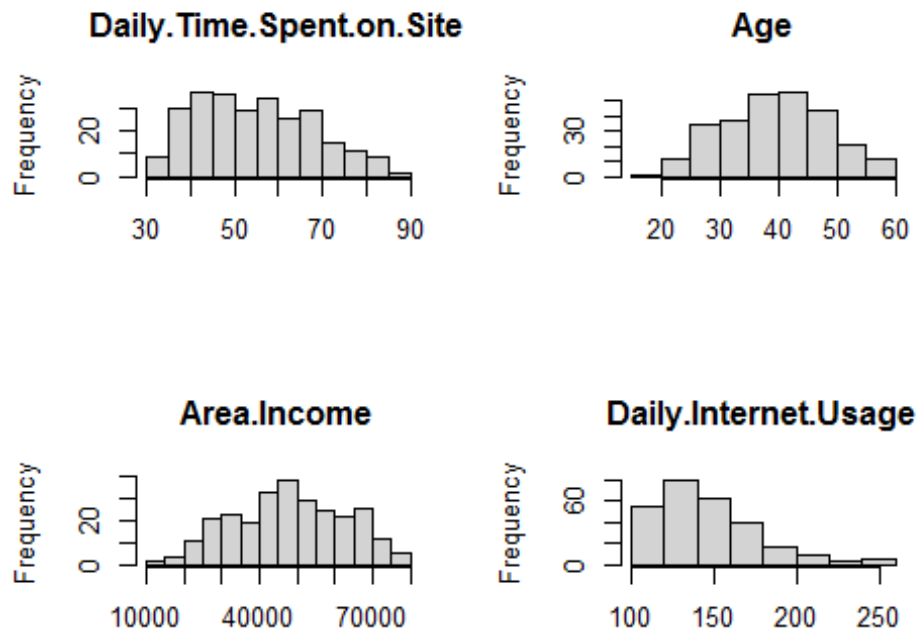
```
Male <- clicked[clicked$Male == 1,]
par(mfrow = c(2,2))
for (i in 1:4){
  hist(Male[,i],main = names(Male)[i], xlab = NULL)
}
```

Females

that clicked on the ad

```
Female <- clicked[clicked$Male == 0,]
par(mfrow = c(2,2))
for (i in 1:4){
  hist(Female[,i],main = names(Female)[i], xlab = NULL)
}
```



There is a lot more variance in female Area.Income than for male, the peak for Male salaries was between 50,000 to 55,000 while for female it was between 45,000 to 50,000.

On average Male who clicked on the ad spent less time on the site than female.

Time Spent on Site

```
#Sort dataframe based on time spent on site variable
time_spent <- clicked[order(-clicked$Daily.Time.Spent.on.Site),]
head(time_spent$Ad.Topic.Line)

## [1] Re-engineered composite moratorium
## [2] Advanced web-enabled standardization
## [3] Fully-configurable 5thgeneration circuit
## [4] Stand-alone radical throughput
## [5] Synchronized leadingedge help-desk
## [6] Stand-alone tangible moderator
## 1000 Levels: Adaptive 24hour Graphic Interface ... Visionary reciprocal
## circuit
```

These are the ad topic line clicked among the people that spent the most amount of time on the site

```
tail(time_spent$Ad.Topic.Line)

## [1] Polarized clear-thinking budgetary management
## [2] Phased full-range hardware
## [3] Future-proofed fresh-thinking conglomeration
```

```
## [4] Triple-buffered 3rdgeneration migration
## [5] Multi-tiered interactive neural-net
## [6] Customizable homogeneous contingency
## 1000 Levels: Adaptive 24hour Graphic Interface ... Visionary reciprocal
circuit
```

These are the ad topics clicked by persons that spent the least time on the site.

Recommendations

- Target the ads to persons between the ages of 35 years to 45 years.
- Ads should be tailored towards both male and female equally.
- Target persons earn an income between 45,000 to 55,000.
- The top countries to target are Australia, Ethiopia, Turkey, Liberia and Liechtenstein.
- The top cities to target are Lake David lake James, Lisamouth, Michelleside, Millerbury and Robertfurt.
- Men take 35 to 45 minutes before clicking on an ad while women spend 40 to 60 minutes on the site before clicking on an add.

Conclusion

The data provided did have the necessary information needed to analyze the site visitors. However there were 1000 unique ad topics for the 1000 records provided. Providing more data per topic would have been useful in determining the kinds of topic that were more popular than others.