ADVERTISEMENT ANALYSIS IN R

Grace Ndunda

5/20/2021

## 1. Defining the reserch question

* 1. **To perform analysis on the advertisement sector**

## 2. Defining the metrics of success

* 1. **Loading the dataset**
  2. **Performing data cleaning**
  3. **Performing Bivariate Analysis**
  4. **Performing univariarte Analysis**
  5. **Providing recomendations**

## **Loading the dataset**

advertising\_dataset <- read.csv("C:/Users/HP/Downloads/advertising.csv")

## **Previewing the dataset**

summary(advertising\_dataset)

## 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

## **printing the first rows of my dataset**

#print first 10 rows of mydata  
head(advertising\_dataset, n=10)

## 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  
## 7 88.91 33 53852.85 208.36  
## 8 66.00 48 24593.33 131.76  
## 9 74.53 30 68862.00 221.51  
## 10 69.88 20 55642.32 183.82  
## 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  
## 7 Enhanced dedicated support Brandonstad 0 Myanmar  
## 8 Reactive local challenge Port Jefferybury 1 Australia  
## 9 Configurable coherent function West Colin 1 Grenada  
## 10 Mandatory homogeneous architecture Ramirezton 1 Ghana  
## 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  
## 7 2016-01-28 20:59:32 0  
## 8 2016-03-07 01:40:15 1  
## 9 2016-04-18 09:33:42 0  
## 10 2016-07-11 01:42:51 0

## **printing the last rows of my dataset**

# print last 5 rows of mydata  
tail(advertising\_dataset, n=5)

## Daily.Time.Spent.on.Site Age Area.Income Daily.Internet.Usage  
## 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  
## 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  
## 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

## **performing str(),dim()and class()functions**

str(advertising\_dataset)

## '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 standardization" "Organic bottom-line service-desk" "Triple-buffered reciprocal time-frame" ...  
## $ 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" "2016-01-10 02:31:19" ...  
## $ Clicked.on.Ad : int 0 0 0 0 0 0 0 1 0 0 ...

dim(advertising\_dataset)

## [1] 1000 10

class(advertising\_dataset)

## [1] "data.frame"

### **Data Cleaning**

## ***checking for null value***

sum(is.na(advertising\_dataset))

## [1] 0

**Checking for duplicated values**

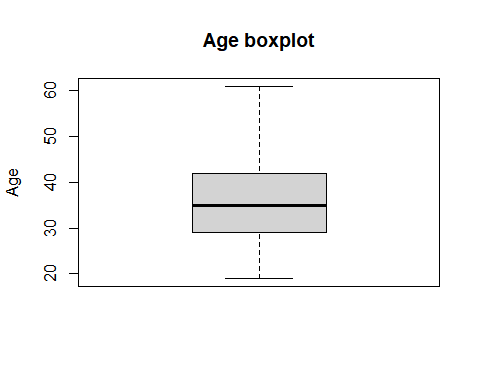
sum(duplicated(advertising\_dataset))

## [1] 0

**There are no duplicated values**

## **Checking for outliers using Age variable**

boxplot(advertising\_dataset$Age,  
 ylab = "Age",main="Age boxplot"  
)

 **Observation**

***most o0f the ages lie between 20 to 40 years . focusing the adverts on this age would lead to more clicks on the adds.***

## **using stats function to check for outliers**

boxplot.stats(advertising\_dataset$Age)$out

## integer(0)

boxplot.stats(advertising\_dataset$Daily.Internet.Usage)$out

## numeric(0)

boxplot.stats(advertising\_dataset$Daily.Time.Spent.on.Site)$out

## numeric(0)

**There are no outliers**

### **Univariate Analysis**

## **Measures of Central tendancy**

## **finding the mean of Age**

Age.mean <- mean(advertising\_dataset$Age)  
Age.mean

## [1] 36.009

## **Finding the median**

Time.median <- median(advertising\_dataset$Daily.Time.Spent.on.Site)  
Time.median

## [1] 68.215

## **Measures of dispersion**

## **Finding the minimum element of Area Income**

Income.min <- min(advertising\_dataset$Area.Income)  
Income.min

## [1] 13996.5

## **Finding the maximum element of Area Income**

Income.max <- max(advertising\_dataset$Area.Income)  
Income.max

## [1] 79484.8

## **Finding Age range**

Age.range <- range(advertising\_dataset$Age)  
Age.range

## [1] 19 61

## **checking the interquatile range of daily internet usage**

Internet.usage.quantile <- quantile(advertising\_dataset$Daily.Internet.Usage)  
Internet.usage.quantile

## 0% 25% 50% 75% 100%   
## 104.7800 138.8300 183.1300 218.7925 269.9600

## **Calculating Age variance**

Age.variance <- var(advertising\_dataset$Age)  
Age.variance

## [1] 77.18611

## **Standard Deviation**

Age.sd <- sd(advertising\_dataset$Age)  
Age.sd

## [1] 8.785562

## **Skewness**

library(moments)  
skewness(advertising\_dataset$Age)

## [1] 0.4784227

# this is a positive skew.

# **Kurtosis**

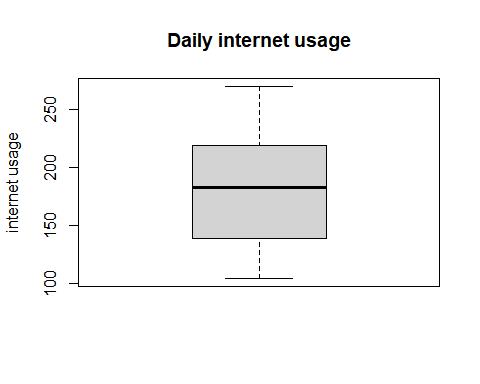
data <- advertising\_dataset$Age  
kurtosis(data)

## [1] 2.595482

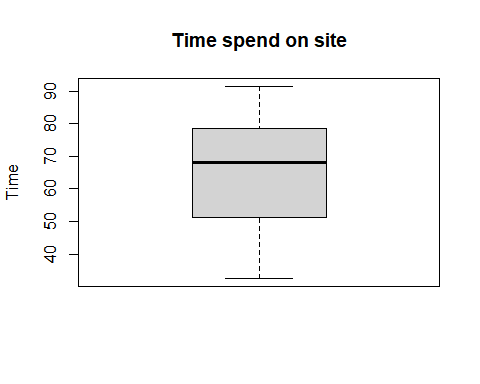
# **this is a positive kurtosis this indicates a heavy tail**

## **Boxplot of Daily Internet Usage**

boxplot(advertising\_dataset$Daily.Internet.Usage,  
 ylab = "internet usage",main="Daily internet usage"  
)

 ## boxplot of Time spend on site

boxplot(advertising\_dataset$Daily.Time.Spent.on.Site,  
 ylab = "Time",main="Time spend on site"  
)



## Observation

**internet usage lies between 150 to 250 per day this peple would mostly click on adds** **most people spend 50 to 80 minutes on internet daily this people would click o adds**

## **Age Frequency Bar Graph**

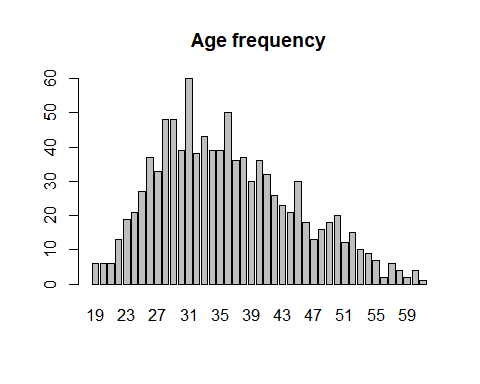
Age <- advertising\_dataset$Age  
#Applying the table() function will compute the frequency distribution of Ages  
Age\_frequency <- table(Age)  
Age\_frequency

## 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

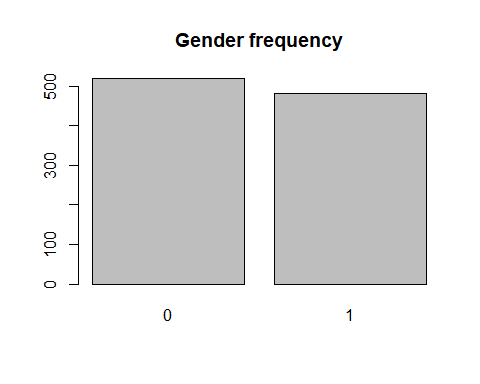
gender <- advertising\_dataset$Male  
gender\_frequency <- table(gender)  
gender\_frequency

## gender  
## 0 1   
## 519 481

barplot(Age\_frequency,gender\_frequency,main="Age frequency")



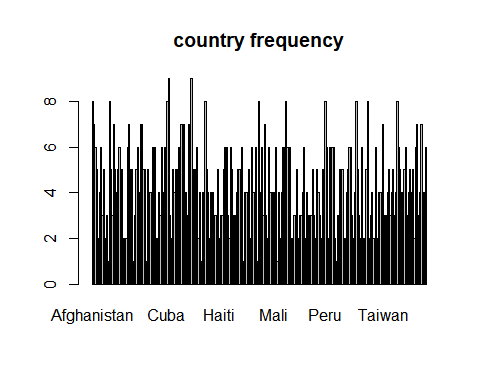
barplot(gender\_frequency,main="Gender frequency")

 ## **Observations** \* **Most of the people are of 30 to 40 years** \* **The male (0) gender has the highest frequency as compared to the ladies(1)**

country <- advertising\_dataset$Country  
#Applying the table() function will compute the frequency distribution of Ages  
country\_frequency <- table(country)  
head(country\_frequency)

## country  
## Afghanistan Albania Algeria American Samoa Andorra   
## 8 7 6 5 2   
## Angola   
## 4

barplot(country\_frequency,,main="country frequency")



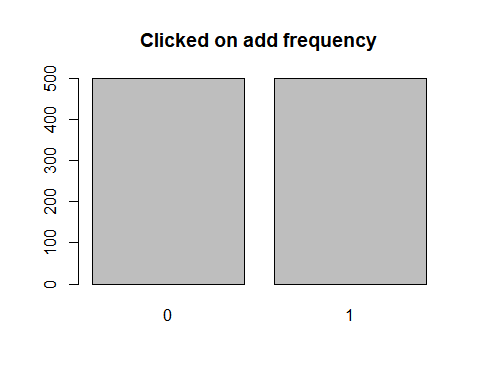
## \*\*Observations

**Afghanistan and Turkey have a high Frequency this means this countries can be a target for the advertisements**

clicked\_on\_add <- advertising\_dataset$Clicked.on.Ad  
#Applying the table() function will compute the frequency distribution of Ages  
clicked\_frequency <- table(clicked\_on\_add)  
clicked\_frequency

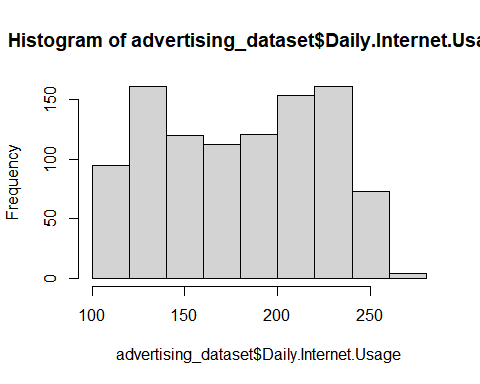
## clicked\_on\_add  
## 0 1   
## 500 500

barplot(clicked\_frequency,main="Clicked on add frequency")

 ## clicked on add and unclicked adds have the same frequency

## **Histogram**

hist(advertising\_dataset$Daily.Internet.Usage)



**observation** \* Most of the people use 150 to 200 internet daily.

### **Bivariate Analysis**

## **covariance**

Age <- advertising\_dataset$Age  
# Assigning the usage column to the variable Male  
# ---  
#   
usage <- advertising\_dataset$Daily.Internet.Usage  
cov(Age,usage)

## [1] -141.6348

# they have a negative covarience

**observation** \* The covariance of Age and Daily.Internet.Usage is about -141.6348 It indicates a negative non\_linear relationship between the two variables.

## **CoRRELATION**

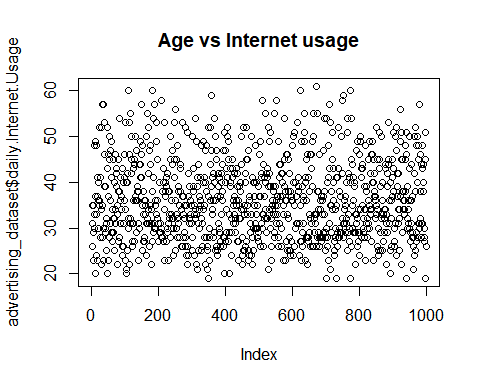
Age <- advertising\_dataset$Age  
# Assigning the usage column to the variable Male  
# ---  
#   
usage <- advertising\_dataset$Daily.Internet.Usage  
cor(Age,usage,method="pearson")

## [1] -0.3672086

**The correlation coefficient is -0.3672086** **Because it is close to 0, I can conclude that the variables are negatively linearly related**

## plotting age vs internet Usage

plot(advertising\_dataset$Age,advertising\_dataset$daily.Internet.Usage,main="Age vs Internet usage")

 **Observations** \* There is no relationship between the two variables

#myvars <- c("Age")  
#me <- ("Daily.Time.Spent.On.Site")  
cormat <- round(cor(x=advertising\_dataset$Daily.Time.Spent.on.Site,y=advertising\_dataset$Age),2)  
head(cormat)

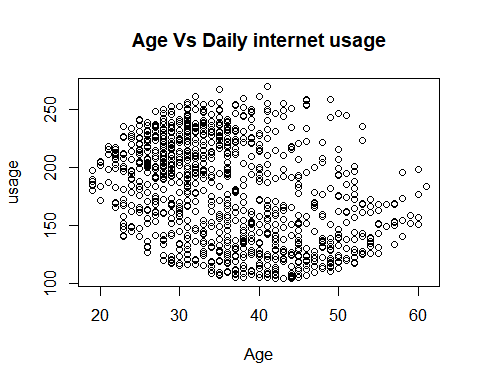
## [1] -0.33

library(reshape2)  
melted\_cormat <- melt(cormat)  
head(melted\_cormat)

## value  
## 1 -0.33

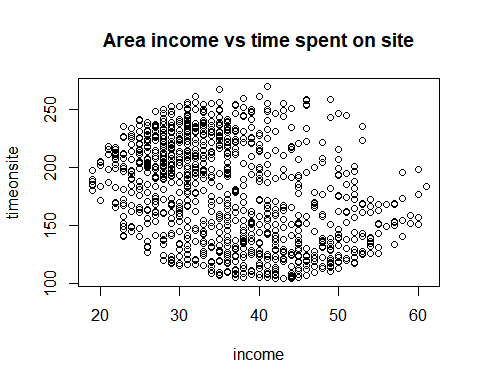
# **Scatterplots**

Age <- advertising\_dataset$Age   
usage <- advertising\_dataset$ Daily.Internet.Usage  
plot(Age, usage, xlab="Age", ylab="usage",main="Age Vs Daily internet usage")



**# Observations** *1. Age and daily internet usage do not relate.* 2. someones age cannot affect their internet usage

income <- advertising\_dataset$Area.Income   
timeonsite <- advertising\_dataset$Daily.Time.Spent.on.Site  
plot(Age, usage, xlab="income", ylab="timeonsite",main="Area income vs time spent on site")



**# Observations** \*1. Time spent on site does not relate to income.

# **Recommendations**

**1. Area Income does not affect internet usage**

**2. Focus on ages of 30 to 40 the frequency of this ages is high meaning they will be able to click on adds.**

**3. focus on people who have a internet usage of 150 to 250 this people would most probably click on adds.**

**4. Focusing on the male(0)gender which has the higher frequency as compared to ladies (1) could also lead to more viewing of the adds.**

**5. Afghanistan and Turkey have a high Frequency this means this countries can be a target for the advertisements**

**6. Most people spend 50 to 80 minutes on internet daily this people would click on adds**

**# conclusions**

**1. the clicked adds and the unclicked adds have the same frequency**

**2. Age and internet usage have no relationship**

## \*\*Follow up Questions

**1. Did I have the right data?**

**Yes the data was right for my analysis.**

**2. Did we need other data?**

**No**

**3. Did I the right objective**

**Yes, I was able to achieve my objective**