# Defining the question

identifying the individuals who are most likely going to click on the Kenyan entrepreneur’s ads

# Metric for success

visualisations

# Context

a Kenyan entreneur has started an online cryptography course and would want to advertise it on her blog. she targets audiences from various countries.

# Experimental design

Data wrangling Feature engineering Data visualisation

#### appropriateness of the data

The Kenyan Entrepreneur has already collected the data that will enable me to work on.

# Loading the dataset  
advertising <- read.csv("C:/Users/I/Downloads/advertising.csv")

# Checking the dataset

# Previwing the top of the dataset  
head(advertising)

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

# Previewing the bottom of the dataset  
tail(advertising)

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

# Checking the columns of the dataset  
colnames(advertising)

## [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 rows of the dataset  
rownames(advertising)

## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10"   
## [11] "11" "12" "13" "14" "15" "16" "17" "18" "19" "20"   
## [21] "21" "22" "23" "24" "25" "26" "27" "28" "29" "30"   
## [31] "31" "32" "33" "34" "35" "36" "37" "38" "39" "40"   
## [41] "41" "42" "43" "44" "45" "46" "47" "48" "49" "50"   
## [51] "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"   
## [61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70"   
## [71] "71" "72" "73" "74" "75" "76" "77" "78" "79" "80"   
## [81] "81" "82" "83" "84" "85" "86" "87" "88" "89" "90"   
## [91] "91" "92" "93" "94" "95" "96" "97" "98" "99" "100"   
## [101] "101" "102" "103" "104" "105" "106" "107" "108" "109" "110"   
## [111] "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"   
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130"   
## [131] "131" "132" "133" "134" "135" "136" "137" "138" "139" "140"   
## [141] "141" "142" "143" "144" "145" "146" "147" "148" "149" "150"   
## [151] "151" "152" "153" "154" "155" "156" "157" "158" "159" "160"   
## [161] "161" "162" "163" "164" "165" "166" "167" "168" "169" "170"   
## [171] "171" "172" "173" "174" "175" "176" "177" "178" "179" "180"   
## [181] "181" "182" "183" "184" "185" "186" "187" "188" "189" "190"   
## [191] "191" "192" "193" "194" "195" "196" "197" "198" "199" "200"   
## [201] "201" "202" "203" "204" "205" "206" "207" "208" "209" "210"   
## [211] "211" "212" "213" "214" "215" "216" "217" "218" "219" "220"   
## [221] "221" "222" "223" "224" "225" "226" "227" "228" "229" "230"   
## [231] "231" "232" "233" "234" "235" "236" "237" "238" "239" "240"   
## [241] "241" "242" "243" "244" "245" "246" "247" "248" "249" "250"   
## [251] "251" "252" "253" "254" "255" "256" "257" "258" "259" "260"   
## [261] "261" "262" "263" "264" "265" "266" "267" "268" "269" "270"   
## [271] "271" "272" "273" "274" "275" "276" "277" "278" "279" "280"   
## [281] "281" "282" "283" "284" "285" "286" "287" "288" "289" "290"   
## [291] "291" "292" "293" "294" "295" "296" "297" "298" "299" "300"   
## [301] "301" "302" "303" "304" "305" "306" "307" "308" "309" "310"   
## [311] "311" "312" "313" "314" "315" "316" "317" "318" "319" "320"   
## [321] "321" "322" "323" "324" "325" "326" "327" "328" "329" "330"   
## [331] "331" "332" "333" "334" "335" "336" "337" "338" "339" "340"   
## [341] "341" "342" "343" "344" "345" "346" "347" "348" "349" "350"   
## [351] "351" "352" "353" "354" "355" "356" "357" "358" "359" "360"   
## [361] "361" "362" "363" "364" "365" "366" "367" "368" "369" "370"   
## [371] "371" "372" "373" "374" "375" "376" "377" "378" "379" "380"   
## [381] "381" "382" "383" "384" "385" "386" "387" "388" "389" "390"   
## [391] "391" "392" "393" "394" "395" "396" "397" "398" "399" "400"   
## [401] "401" "402" "403" "404" "405" "406" "407" "408" "409" "410"   
## [411] "411" "412" "413" "414" "415" "416" "417" "418" "419" "420"   
## [421] "421" "422" "423" "424" "425" "426" "427" "428" "429" "430"   
## [431] "431" "432" "433" "434" "435" "436" "437" "438" "439" "440"   
## [441] "441" "442" "443" "444" "445" "446" "447" "448" "449" "450"   
## [451] "451" "452" "453" "454" "455" "456" "457" "458" "459" "460"   
## [461] "461" "462" "463" "464" "465" "466" "467" "468" "469" "470"   
## [471] "471" "472" "473" "474" "475" "476" "477" "478" "479" "480"   
## [481] "481" "482" "483" "484" "485" "486" "487" "488" "489" "490"   
## [491] "491" "492" "493" "494" "495" "496" "497" "498" "499" "500"   
## [501] "501" "502" "503" "504" "505" "506" "507" "508" "509" "510"   
## [511] "511" "512" "513" "514" "515" "516" "517" "518" "519" "520"   
## [521] "521" "522" "523" "524" "525" "526" "527" "528" "529" "530"   
## [531] "531" "532" "533" "534" "535" "536" "537" "538" "539" "540"   
## [541] "541" "542" "543" "544" "545" "546" "547" "548" "549" "550"   
## [551] "551" "552" "553" "554" "555" "556" "557" "558" "559" "560"   
## [561] "561" "562" "563" "564" "565" "566" "567" "568" "569" "570"   
## [571] "571" "572" "573" "574" "575" "576" "577" "578" "579" "580"   
## [581] "581" "582" "583" "584" "585" "586" "587" "588" "589" "590"   
## [591] "591" "592" "593" "594" "595" "596" "597" "598" "599" "600"   
## [601] "601" "602" "603" "604" "605" "606" "607" "608" "609" "610"   
## [611] "611" "612" "613" "614" "615" "616" "617" "618" "619" "620"   
## [621] "621" "622" "623" "624" "625" "626" "627" "628" "629" "630"   
## [631] "631" "632" "633" "634" "635" "636" "637" "638" "639" "640"   
## [641] "641" "642" "643" "644" "645" "646" "647" "648" "649" "650"   
## [651] "651" "652" "653" "654" "655" "656" "657" "658" "659" "660"   
## [661] "661" "662" "663" "664" "665" "666" "667" "668" "669" "670"   
## [671] "671" "672" "673" "674" "675" "676" "677" "678" "679" "680"   
## [681] "681" "682" "683" "684" "685" "686" "687" "688" "689" "690"   
## [691] "691" "692" "693" "694" "695" "696" "697" "698" "699" "700"   
## [701] "701" "702" "703" "704" "705" "706" "707" "708" "709" "710"   
## [711] "711" "712" "713" "714" "715" "716" "717" "718" "719" "720"   
## [721] "721" "722" "723" "724" "725" "726" "727" "728" "729" "730"   
## [731] "731" "732" "733" "734" "735" "736" "737" "738" "739" "740"   
## [741] "741" "742" "743" "744" "745" "746" "747" "748" "749" "750"   
## [751] "751" "752" "753" "754" "755" "756" "757" "758" "759" "760"   
## [761] "761" "762" "763" "764" "765" "766" "767" "768" "769" "770"   
## [771] "771" "772" "773" "774" "775" "776" "777" "778" "779" "780"   
## [781] "781" "782" "783" "784" "785" "786" "787" "788" "789" "790"   
## [791] "791" "792" "793" "794" "795" "796" "797" "798" "799" "800"   
## [801] "801" "802" "803" "804" "805" "806" "807" "808" "809" "810"   
## [811] "811" "812" "813" "814" "815" "816" "817" "818" "819" "820"   
## [821] "821" "822" "823" "824" "825" "826" "827" "828" "829" "830"   
## [831] "831" "832" "833" "834" "835" "836" "837" "838" "839" "840"   
## [841] "841" "842" "843" "844" "845" "846" "847" "848" "849" "850"   
## [851] "851" "852" "853" "854" "855" "856" "857" "858" "859" "860"   
## [861] "861" "862" "863" "864" "865" "866" "867" "868" "869" "870"   
## [871] "871" "872" "873" "874" "875" "876" "877" "878" "879" "880"   
## [881] "881" "882" "883" "884" "885" "886" "887" "888" "889" "890"   
## [891] "891" "892" "893" "894" "895" "896" "897" "898" "899" "900"   
## [901] "901" "902" "903" "904" "905" "906" "907" "908" "909" "910"   
## [911] "911" "912" "913" "914" "915" "916" "917" "918" "919" "920"   
## [921] "921" "922" "923" "924" "925" "926" "927" "928" "929" "930"   
## [931] "931" "932" "933" "934" "935" "936" "937" "938" "939" "940"   
## [941] "941" "942" "943" "944" "945" "946" "947" "948" "949" "950"   
## [951] "951" "952" "953" "954" "955" "956" "957" "958" "959" "960"   
## [961] "961" "962" "963" "964" "965" "966" "967" "968" "969" "970"   
## [971] "971" "972" "973" "974" "975" "976" "977" "978" "979" "980"   
## [981] "981" "982" "983" "984" "985" "986" "987" "988" "989" "990"   
## [991] "991" "992" "993" "994" "995" "996" "997" "998" "999" "1000"

# Checking for the data types in the dataset  
dim(advertising)

## [1] 1000 10

there are 1000 rows and 10 columns

# Tidying the dataset

# Checking for missing data  
colSums(is.na(advertising))

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

there are no missing data in the dataset

# Checking for duplicates  
duplicated(advertising)

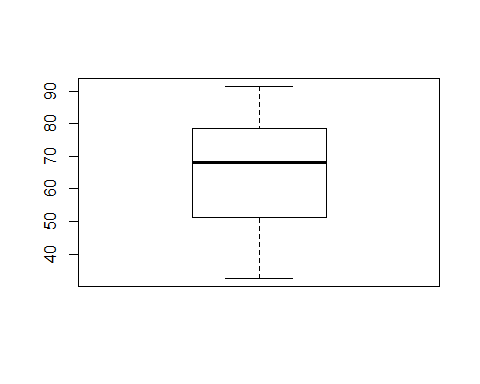
## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [217] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [229] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [277] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [301] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [325] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [337] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [349] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [373] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [385] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [433] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [445] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [457] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [469] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [481] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [493] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [505] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [517] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [529] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [541] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [553] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [565] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [577] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [589] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [601] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [613] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [625] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [637] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [649] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [661] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [673] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [685] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [697] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [709] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [721] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [733] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [745] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [757] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [769] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [781] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [793] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [805] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [817] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [829] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [841] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [853] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [865] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [877] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [889] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [901] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [913] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [925] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [937] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [949] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [961] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [973] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [985] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [997] FALSE FALSE FALSE FALSE

# Showing whether there are repeat entries  
advertising[duplicated(advertising), ]

## [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)

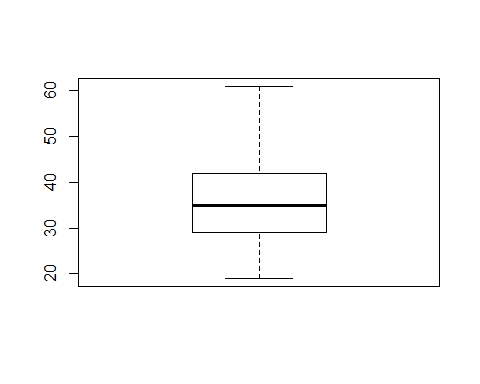
there are no duplicated data in the dataset # Checking for outliers

# In Daily.Time.Spent.on.Site  
boxplot(advertising$Daily.Time.Spent.on.Site)



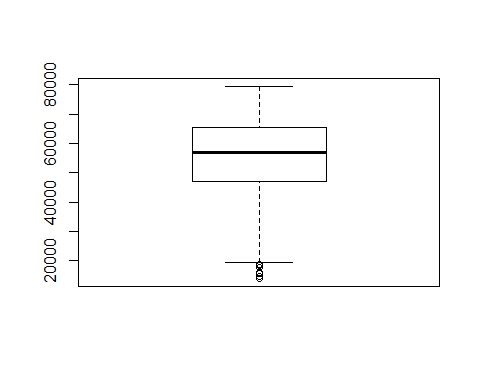
there are no outliers in the time spent on site columns

# In Age   
boxplot(advertising$Age)



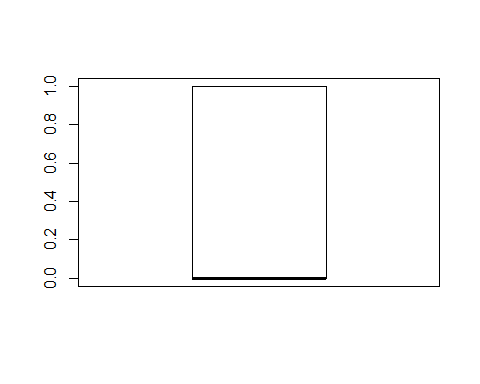
there are no outliers in the Age column

# In Area.Income  
boxplot(advertising$Area.Income)



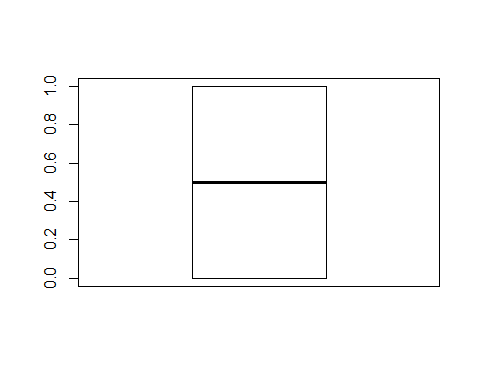
there are outliers that are less than 20,000. The outliers look valid since they are from a dataset that has been collected by the entrepreneur over a period of time

# Male  
boxplot(advertising$Male)



there are no outliers in the column

# Clicked.on.Ad  
boxplot(advertising$Clicked.on.Ad)



there are no clicks on ad

# Feature engineering

# Making the column names uniform  
names(advertising)[names(advertising)=="Daily.Time.Spent.on.Site"] <- "daily\_time\_spent\_on\_site"  
names(advertising)[names(advertising)=="Age"] <- "age"  
names(advertising)[names(advertising)=="Area.Income"] <- "area\_income"  
names(advertising)[names(advertising)=="Daily.Internet.Usage"] <- "daily\_internet\_usage"  
names(advertising)[names(advertising)=="Ad.Topic.Line"] <- "ad\_topic\_line"  
names(advertising)[names(advertising)=="City"] <- "city"  
names(advertising)[names(advertising)=="Male"] <- "sex" #changing the male name to sex because is representing males and female  
names(advertising)[names(advertising)=="Country"] <- "country"  
names(advertising)[names(advertising)=="Timestamp"] <- "timestamp"  
names(advertising)[names(advertising)=="Clicked.on.Ad"] <- "clicked\_on\_ad"

colnames(advertising)

## [1] "daily\_time\_spent\_on\_site" "age"   
## [3] "area\_income" "daily\_internet\_usage"   
## [5] "ad\_topic\_line" "city"   
## [7] "sex" "country"   
## [9] "timestamp" "clicked\_on\_ad"

verifying that the suggested changes has taken place

# Removing the ad\_topic\_line column because i do not see its relevancy  
advertising$ad\_topic\_line <- NULL

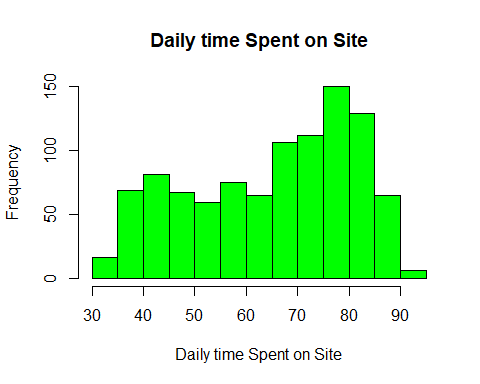
colnames(advertising)

## [1] "daily\_time\_spent\_on\_site" "age"   
## [3] "area\_income" "daily\_internet\_usage"   
## [5] "city" "sex"   
## [7] "country" "timestamp"   
## [9] "clicked\_on\_ad"

we have now verified that the ad topic line has been removed from the dataset

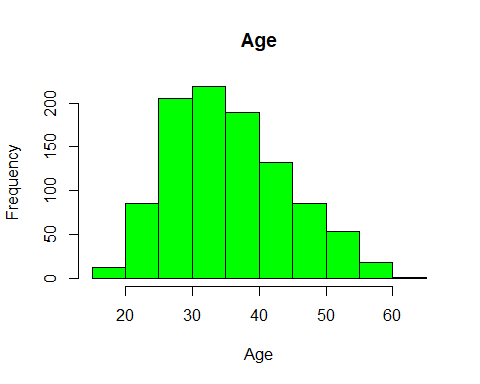
# Univariate data analysis

# Daily time spent on site  
x = hist(advertising$daily\_time\_spent\_on\_site,  
 main = "Daily time Spent on Site",  
 xlab = "Daily time Spent on Site",  
 col ='green')

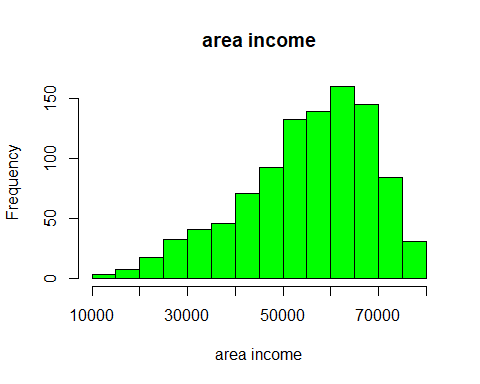


the data is skewed to the left and most of the data is distributed between 65 and 85

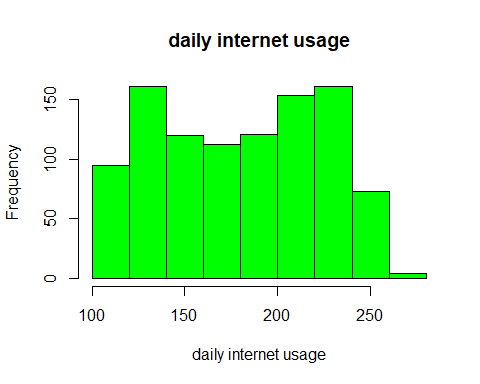
# Age  
x = hist(advertising$age,  
 main = "Age",  
 xlab = "Age",  
 col ='green')

 data is skewed to the right and most of the data is distributed between 25 and 40

# Area\_income  
x = hist(advertising$area\_income,  
 main = "area income",  
 xlab = "area income",  
 col ='green')

 data is skewed to the left and most of the data is distributed between 50,000 and 70,000

# Daily\_internet\_usage  
x = hist(advertising$daily\_internet\_usage,  
 main = "daily internet usage",  
 xlab = "daily internet usage",  
 col ='green')

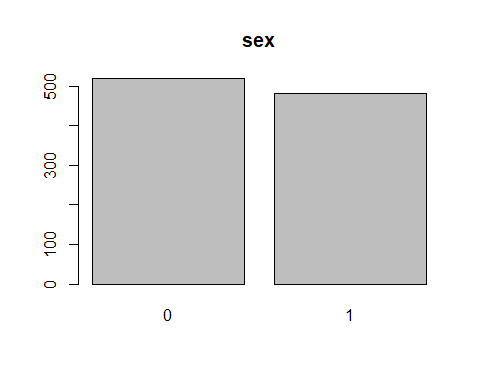


data is used a lot and most of it is distributed between 200 and 225

# Sex  
table(advertising$sex) # there are 519 males and 481 females

##   
## 0 1   
## 519 481

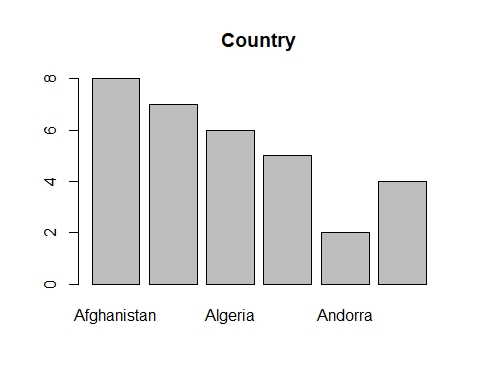
barplot(table(advertising$sex),  
 main="sex")



# Country  
table(advertising$country)

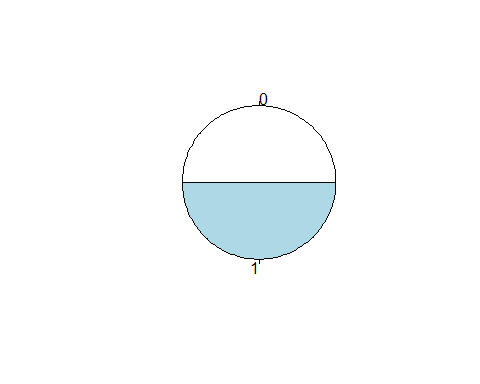
##   
## Afghanistan   
## 8   
## Albania   
## 7   
## Algeria   
## 6   
## American Samoa   
## 5   
## Andorra   
## 2   
## Angola   
## 4   
## Anguilla   
## 6   
## Antarctica (the territory South of 60 deg S)   
## 3   
## Antigua and Barbuda   
## 5   
## Argentina   
## 2   
## Armenia   
## 3   
## Aruba   
## 1   
## Australia   
## 8   
## Austria   
## 5   
## Azerbaijan   
## 3   
## Bahamas   
## 7   
## Bahrain   
## 5   
## Bangladesh   
## 4   
## Barbados   
## 5   
## Belarus   
## 6   
## Belgium   
## 5   
## Belize   
## 5   
## Benin   
## 2   
## Bermuda   
## 1   
## Bhutan   
## 2   
## Bolivia   
## 6   
## Bosnia and Herzegovina   
## 7   
## Bouvet Island (Bouvetoya)   
## 5   
## Brazil   
## 5   
## British Indian Ocean Territory (Chagos Archipelago)   
## 1   
## British Virgin Islands   
## 3   
## Brunei Darussalam   
## 5   
## Bulgaria   
## 6   
## Burkina Faso   
## 4   
## Burundi   
## 7   
## Cambodia   
## 7   
## Cameroon   
## 5   
## Canada   
## 5   
## Cape Verde   
## 1   
## Cayman Islands   
## 5   
## Central African Republic   
## 2   
## Chad   
## 4   
## Chile   
## 4   
## China   
## 6   
## Christmas Island   
## 6   
## Colombia   
## 2   
## Comoros   
## 2   
## Congo   
## 4   
## Cook Islands   
## 3   
## Costa Rica   
## 6   
## Cote d'Ivoire   
## 4   
## Croatia   
## 6   
## Cuba   
## 5   
## Cyprus   
## 8   
## Czech Republic   
## 9   
## Denmark   
## 3   
## Djibouti   
## 2   
## Dominica   
## 5   
## Dominican Republic   
## 4   
## Ecuador   
## 5   
## Egypt   
## 5   
## El Salvador   
## 6   
## Equatorial Guinea   
## 4   
## Eritrea   
## 7   
## Estonia   
## 3   
## Ethiopia   
## 7   
## Falkland Islands (Malvinas)   
## 4   
## Faroe Islands   
## 3   
## Fiji   
## 7   
## Finland   
## 5   
## France   
## 9   
## French Guiana   
## 4   
## French Polynesia   
## 5   
## French Southern Territories   
## 5   
## Gabon   
## 6   
## Gambia   
## 2   
## Georgia   
## 4   
## Germany   
## 1   
## Ghana   
## 4   
## Gibraltar   
## 3   
## Greece   
## 8   
## Greenland   
## 5   
## Grenada   
## 4   
## Guadeloupe   
## 2   
## Guam   
## 4   
## Guatemala   
## 4   
## Guernsey   
## 3   
## Guinea   
## 3   
## Guinea-Bissau   
## 2   
## Guyana   
## 5   
## Haiti   
## 2   
## Heard Island and McDonald Islands   
## 3   
## Holy See (Vatican City State)   
## 3   
## Honduras   
## 5   
## Hong Kong   
## 6   
## Hungary   
## 6   
## Iceland   
## 3   
## India   
## 2   
## Indonesia   
## 6   
## Iran   
## 5   
## Ireland   
## 3   
## Isle of Man   
## 3   
## Israel   
## 4   
## Italy   
## 5   
## Jamaica   
## 5   
## Japan   
## 4   
## Jersey   
## 6   
## Jordan   
## 1   
## Kazakhstan   
## 4   
## Kenya   
## 4   
## Kiribati   
## 1   
## Korea   
## 5   
## Kuwait   
## 2   
## Kyrgyz Republic   
## 6   
## Lao People's Democratic Republic   
## 4   
## Latvia   
## 4   
## Lebanon   
## 6   
## Lesotho   
## 1   
## Liberia   
## 8   
## Libyan Arab Jamahiriya   
## 4   
## Liechtenstein   
## 6   
## Lithuania   
## 3   
## Luxembourg   
## 7   
## Macao   
## 3   
## Macedonia   
## 2   
## Madagascar   
## 6   
## Malawi   
## 4   
## Malaysia   
## 3   
## Maldives   
## 4   
## Mali   
## 4   
## Malta   
## 6   
## Marshall Islands   
## 1   
## Martinique   
## 4   
## Mauritania   
## 2   
## Mauritius   
## 4   
## Mayotte   
## 6   
## Mexico   
## 6   
## Micronesia   
## 8   
## Moldova   
## 6   
## Monaco   
## 3   
## Mongolia   
## 6   
## Montenegro   
## 2   
## Montserrat   
## 1   
## Morocco   
## 3   
## Mozambique   
## 1   
## Myanmar   
## 5   
## Namibia   
## 2   
## Nauru   
## 3   
## Nepal   
## 3   
## Netherlands   
## 4   
## Netherlands Antilles   
## 6   
## New Caledonia   
## 2   
## New Zealand   
## 4   
## Nicaragua   
## 3   
## Niger   
## 3   
## Niue   
## 3   
## Norfolk Island   
## 5   
## Northern Mariana Islands   
## 3   
## Norway   
## 2   
## Pakistan   
## 5   
## Palau   
## 4   
## Palestinian Territory   
## 3   
## Panama   
## 2   
## Papua New Guinea   
## 5   
## Paraguay   
## 3   
## Peru   
## 8   
## Philippines   
## 6   
## Pitcairn Islands   
## 2   
## Poland   
## 6   
## Portugal   
## 3   
## Puerto Rico   
## 6   
## Qatar   
## 6   
## Reunion   
## 2   
## Romania   
## 1   
## Russian Federation   
## 3   
## Rwanda   
## 5   
## Saint Barthelemy   
## 2   
## Saint Helena   
## 5   
## Saint Kitts and Nevis   
## 1   
## Saint Lucia   
## 2   
## Saint Martin   
## 4   
## Saint Pierre and Miquelon   
## 5   
## Saint Vincent and the Grenadines   
## 6   
## Samoa   
## 6   
## San Marino   
## 3   
## Sao Tome and Principe   
## 2   
## Saudi Arabia   
## 4   
## Senegal   
## 8   
## Serbia   
## 5   
## Seychelles   
## 3   
## Sierra Leone   
## 2   
## Singapore   
## 6   
## Slovakia (Slovak Republic)   
## 2   
## Slovenia   
## 1   
## Somalia   
## 5   
## South Africa   
## 8   
## South Georgia and the South Sandwich Islands   
## 2   
## Spain   
## 3   
## Sri Lanka   
## 4   
## Sudan   
## 2   
## Suriname   
## 2   
## Svalbard & Jan Mayen Islands   
## 6   
## Swaziland   
## 2   
## Sweden   
## 4   
## Switzerland   
## 4   
## Syrian Arab Republic   
## 3   
## Taiwan   
## 7   
## Tajikistan   
## 3   
## Tanzania   
## 3   
## Thailand   
## 4   
## Timor-Leste   
## 5   
## Togo   
## 3   
## Tokelau   
## 4   
## Tonga   
## 5   
## Trinidad and Tobago   
## 3   
## Tunisia   
## 4   
## Turkey   
## 8   
## Turkmenistan   
## 6   
## Turks and Caicos Islands   
## 5   
## Tuvalu   
## 4   
## Uganda   
## 4   
## Ukraine   
## 5   
## United Arab Emirates   
## 6   
## United Kingdom   
## 3   
## United States Minor Outlying Islands   
## 4   
## United States of America   
## 5   
## United States Virgin Islands   
## 4   
## Uruguay   
## 5   
## Uzbekistan   
## 2   
## Vanuatu   
## 6   
## Venezuela   
## 7   
## Vietnam   
## 3   
## Wallis and Futuna   
## 4   
## Western Sahara   
## 7   
## Yemen   
## 3   
## Zambia   
## 4   
## Zimbabwe   
## 6

sample <- head(table(advertising$country))  
barplot(sample, main="Country")



we see that afghanistan has the most appearance.

# Clicked\_on\_ad  
  
pie(table(advertising$clicked\_on\_ad))



there are 500 no clicks and 500 click on ads

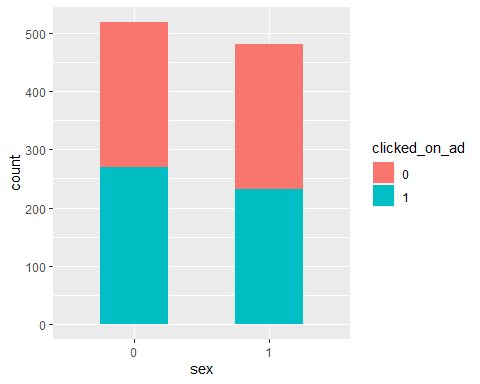
# Bivariate Analysis

# Converting variables into string so as to enable to make good visualisations.  
a = c('city', 'country')  
for (i in a) {  
 advertising[,i] = as.character(advertising[,i])  
}

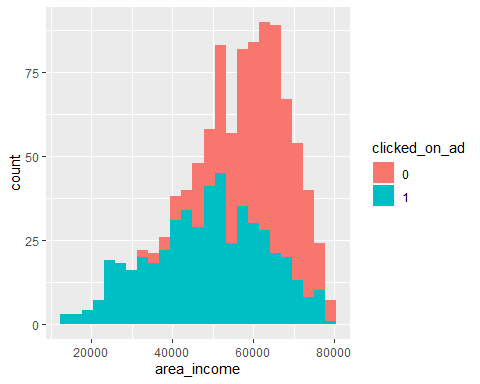
# Converting variables into factors so as to enable to make good visualisations.  
b = c('sex', 'clicked\_on\_ad')  
for (i in b) {  
advertising[,i] = as.factor(advertising[,i])  
}

# Importing the ggplot library  
library(ggplot2)

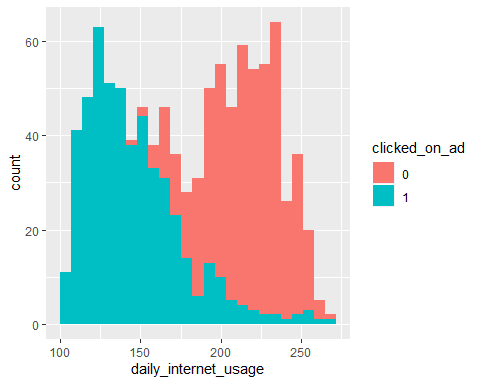
# Sex vs clicked on ad   
sex\_vs\_ad = ggplot(data = advertising, aes(x = sex, fill = clicked\_on\_ad))+  
geom\_bar(width = 0.5)  
sex\_vs\_ad



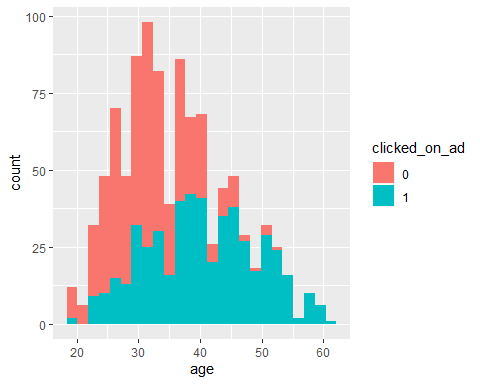
# Area\_income vs clicked on ad  
area\_vs\_ad = ggplot(data = advertising, aes(x = area\_income, fill = clicked\_on\_ad))+   
geom\_histogram(bins = 25)  
area\_vs\_ad

 many people who had an income between 50,000 and 70,000 did not click on the ad compared to those who did

# Daily\_internet\_usage vs clicked on ad  
daily\_vs\_ad = ggplot(data = advertising, aes(x = daily\_internet\_usage, fill = clicked\_on\_ad))+   
geom\_histogram(bins = 25)  
daily\_vs\_ad

 between 100 and 150 there are more people who clicked the ad than those who did not.

# Age vs clicked on ad  
age\_vs\_ad = ggplot(data = advertising, aes(x = age, fill = clicked\_on\_ad))+   
geom\_histogram(bins = 25)  
age\_vs\_ad

 generally many people do not click ads

# Recommendation

#### The entrepreneur should focus on people with an income between 10,000 and 50,000

#### The entrepreneur should focus on people with the age between 45 and 60

#### The entrepreneur should focus on people with a daily internet usage between 100 and 175