E-Commerce analysis with clustering

Jenipher Mawia

11/5/2020

1. Problem Definition

1.1 Defining the question

Perform clustering on the following data link stating insights drawn from your analysis and visualizations drawn towards learning the characteristics of customer groups.

1.2 Specifying the question

Implement the solution using unsupervised learning techniques such as **K-means clustering** and **hierarchical clustering**.

2. Defining the metrics for success

This project will be considered a success if the following are achieved: - Unsupervised learning models are built (K-means and hierarchical clustering). - Insights are drawn from the EDA process and modeling.

3. The Context

Kira Plastinina is a Russian brand that is sold through a defunct chain of retail stores in Russia, Ukraine, Kazakhstan, Belarus, China, Philippines, and Armenia. The brand's Sales and Marketing team would like to understand their customer's behavior from data that they have collected over the past year. More specifically, they would like to learn the characteristics of customer groups.

4. Experimental Design taken

The following is the order in which I went about to achieve the objectives of this project:

- Data Sourcing and Understanding
- Checking the data (head and tail, shape(number of records), datatypes)
- Data cleaning procedures (handling null values,outliers, anomalies)
- Exploratory data analysis (Univariate, Bivariate and Multivariate analyses)
- Implementing the solution(Clustering)

- Conclusion and recommendation
- Challenging the solution
- Follow-up questions

5. Data Sourcing

The data used for this project was sourced from the following website and can be downloaded here

Reading the data

```
ecommerce <- read.csv("http://bit.ly/EcommerceCustomersDataset")
```

6. Checking the Data

checking the top of the data

```
# checking the first 6 rows in the data
head(ecommerce)
```

```
##   Administrative Administrative_Duration Informational
Informational_Duration
## 1              0                       0             0
0
## 2              0                       0             0
0
## 3              0                      -1             0
-1
## 4              0                       0             0
0
## 5              0                       0             0
0
## 6              0                       0             0
0
##   ProductRelated ProductRelated_Duration BounceRates ExitRates PageValues
## 1              1                 0.000000  0.20000000 0.2000000          0
## 2              2                64.000000  0.00000000 0.1000000          0
## 3              1                -1.000000  0.20000000 0.2000000          0
## 4              2                 2.666667  0.05000000 0.1400000          0
## 5             10               627.500000  0.02000000 0.0500000          0
## 6             19               154.216667  0.01578947 0.0245614          0
##   SpecialDay Month OperatingSystems Browser Region TrafficType
## 1          0   Feb                1       1      1           1
## 2          0   Feb                2       2      1           2
## 3          0   Feb                4       1      9           3
## 4          0   Feb                3       2      2           4
## 5          0   Feb                3       3      1           4
## 6          0   Feb                2       2      1           3
##         VisitorType Weekend Revenue
```

```
## 1 Returning_Visitor   FALSE   FALSE
## 2 Returning_Visitor   FALSE   FALSE
## 3 Returning_Visitor   FALSE   FALSE
## 4 Returning_Visitor   FALSE   FALSE
## 5 Returning_Visitor    TRUE   FALSE
## 6 Returning_Visitor   FALSE   FALSE
```

checking the bottom of the data

```
# checking the last 6 rows in the data
tail(ecommerce)
```

```
##       Administrative Administrative_Duration Informational
## 12325              0                       0             1
## 12326              3                     145             0
## 12327              0                       0             0
## 12328              0                       0             0
## 12329              4                      75             0
## 12330              0                       0             0
##       Informational_Duration ProductRelated ProductRelated_Duration
BounceRates
## 12325                      0             16                 503.000
0.000000000
## 12326                      0             53                1783.792
0.007142857
## 12327                      0              5                 465.750
0.000000000
## 12328                      0              6                 184.250
0.083333333
## 12329                      0             15                 346.000
0.000000000
## 12330                      0              3                  21.250
0.000000000
##        ExitRates PageValues SpecialDay Month OperatingSystems Browser
Region
## 12325 0.03764706    0.00000          0   Nov                2       2
1
## 12326 0.02903061   12.24172          0   Dec                4       6
1
## 12327 0.02133333    0.00000          0   Nov                3       2
1
## 12328 0.08666667    0.00000          0   Nov                3       2
1
## 12329 0.02105263    0.00000          0   Nov                2       2
3
## 12330 0.06666667    0.00000          0   Nov                3       2
1
##       TrafficType       VisitorType Weekend Revenue
## 12325           1 Returning_Visitor   FALSE   FALSE
## 12326           1 Returning_Visitor    TRUE   FALSE
## 12327           8 Returning_Visitor    TRUE   FALSE
```

```
## 12328          13 Returning_Visitor    TRUE   FALSE
## 12329          11 Returning_Visitor   FALSE   FALSE
## 12330           2       New_Visitor    TRUE   FALSE
```

checking the shape of the data

```
# checking the dimensions of the data (number of entries and fields)
dim(ecommerce)

## [1] 12330    18
```

The data has 12330 entries and 18 columns.

checking the datatypes of the column

```
# getting the datatypes of each column
str(ecommerce)

## 'data.frame':    12330 obs. of  18 variables:
##  $ Administrative         : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ Administrative_Duration: num  0 0 -1 0 0 0 -1 -1 0 0 ...
##  $ Informational          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Informational_Duration : num  0 0 -1 0 0 0 -1 -1 0 0 ...
##  $ ProductRelated         : int  1 2 1 2 10 19 1 1 2 3 ...
##  $ ProductRelated_Duration: num  0 64 -1 2.67 627.5 ...
##  $ BounceRates            : num  0.2 0 0.2 0.05 0.02 ...
##  $ ExitRates              : num  0.2 0.1 0.2 0.14 0.05 ...
##  $ PageValues             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ SpecialDay             : num  0 0 0 0 0 0 0.4 0 0.8 0.4 ...
##  $ Month                  : chr  "Feb" "Feb" "Feb" "Feb" ...
##  $ OperatingSystems       : int  1 2 4 3 3 2 2 1 2 2 ...
##  $ Browser                : int  1 2 1 2 3 2 4 2 2 4 ...
##  $ Region                 : int  1 1 9 2 1 1 3 1 2 1 ...
##  $ TrafficType            : int  1 2 3 4 4 3 3 5 3 2 ...
##  $ VisitorType            : chr  "Returning_Visitor" "Returning_Visitor"
"Returning_Visitor" "Returning_Visitor" "Returning_Visitor" ...
##  $ Weekend                : logi  FALSE FALSE FALSE FALSE TRUE FALSE ...
##  $ Revenue                : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

The data consists of integer, numeric, character and logical datatypes. The first 10 columns are numerical attributes while the last 8 columns are categorical attributes.

7. Appropriateness of the available data to answer the given question

The data contains columns such as:

- “Administrative”, “Administrative Duration”, “Informational”, “Informational Duration”, “Product Related” and “Product Related Duration” represents the number of different types of pages visited by the visitor in that session and total time spent in each of these page categories. The values of these features are derived from the URL

information of the pages visited by the user and updated in real-time when a user takes an action, e.g. moving from one page to another.

- The "Bounce Rate", "Exit Rate" and "Page Value" features represent the metrics measured by "Google Analytics" for each page in the e-commerce site.
- The value of the "Bounce Rate" feature for a web page refers to the percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session.
- The value of the "Exit Rate" feature for a specific web page is calculated as for all pageviews to the page, the percentage that was the last in the session.
- The "Page Value" feature represents the average value for a web page that a user visited before completing an e-commerce transaction.
- The "Special Day" feature indicates the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with the transaction. The value of this attribute is determined by considering the dynamics of e-commerce such as the duration between the order date and delivery date. For example, for Valentina's day, this value takes a nonzero value between February 2 and February 12, zero before and after this date unless it is close to another special day, and its maximum value of 1 on February 8.
- The dataset also includes the operating system, browser, region, traffic type, visitor type as returning or new visitor, a Boolean value indicating whether the date of the visit is weekend, and month of the year.

All these fields are useful in learning the characteristics of customer groups and will play a great role in answering our research question.

Therefore, it can be concluded that the data available is appropriate and relevant to answer the given question.

8. Data Cleaning

8.1 Standardizing the column names to similar formatting

From the above outputs, we can see that the column names are not in a standard/similar format. This needs to be addressed:

```
# get the column names
colnames(ecommerce)
```

```
##  [1] "Administrative"          "Administrative_Duration"
##  [3] "Informational"           "Informational_Duration"
##  [5] "ProductRelated"          "ProductRelated_Duration"
##  [7] "BounceRates"             "ExitRates"
##  [9] "PageValues"              "SpecialDay"
## [11] "Month"                   "OperatingSystems"
## [13] "Browser"                 "Region"
```

```
## [15] "TrafficType"             "VisitorType"
## [17] "Weekend"                 "Revenue"

# changing the column names to a standard format
names(ecommerce)[names(ecommerce) == "Administrative"] <- "administrative"
names(ecommerce)[names(ecommerce) == "Administrative_Duration"] <-
"administrative_duration"
names(ecommerce)[names(ecommerce) == "Informational"] <- "informational"
names(ecommerce)[names(ecommerce) == "Informational_Duration"] <-
"informational_duration"
names(ecommerce)[names(ecommerce) == "ProductRelated"] <- "product_related"
names(ecommerce)[names(ecommerce) == "ProductRelated_Duration"] <-
"product_related_duration"
names(ecommerce)[names(ecommerce) == "BounceRates"] <- "bounce_rates"
names(ecommerce)[names(ecommerce) == "ExitRates"] <- "exit_rates"
names(ecommerce)[names(ecommerce) == "PageValues"] <- "page_values"
names(ecommerce)[names(ecommerce) == "SpecialDay"] <- "special_day"
names(ecommerce)[names(ecommerce) == "Month"] <- "month"
names(ecommerce)[names(ecommerce) == "OperatingSystems"] <-
"operating_systems"
names(ecommerce)[names(ecommerce) == "Browser"] <- "browser"
names(ecommerce)[names(ecommerce) == "Region"] <- "region"
names(ecommerce)[names(ecommerce) == "TrafficType"] <- "traffic_type"
names(ecommerce)[names(ecommerce) == "VisitorType"] <- "visitor_type"
names(ecommerce)[names(ecommerce) == "Weekend"] <- "weekend"
names(ecommerce)[names(ecommerce) == "Revenue"] <- "revenue"

#confirm changes made
colnames(ecommerce)

##  [1] "administrative"          "administrative_duration"
##  [3] "informational"           "informational_duration"
##  [5] "product_related"         "product_related_duration"
##  [7] "bounce_rates"            "exit_rates"
##  [9] "page_values"             "special_day"
## [11] "month"                   "operating_systems"
## [13] "browser"                 "region"
## [15] "traffic_type"            "visitor_type"
## [17] "weekend"                 "revenue"
```

8.2 Duplicated entries

```
#checking for duplicated entries
duplicates <- ecommerce[duplicated(ecommerce),]

library(plyr)
# get the number of duplicated entries in the dataframe
count(duplicates)

##    administrative administrative_duration informational
informational_duration
## 1               0                       0             0
```

```
0
## 2                0                    0              0
0
## 3                0                    0              0
0
## 4                0                    0              0
0
## 5                0                    0              0
0
## 6                0                    0              0
0
## 7                0                    0              0
0
## 8                0                    0              0
0
## 9                0                    0              0
0
## 10               0                    0              0
0
## 11               0                    0              0
0
## 12               0                    0              0
0
## 13               0                    0              0
0
## 14               0                    0              0
0
## 15               0                    0              0
0
## 16               0                    0              0
0
## 17               0                    0              0
0
## 18               0                    0              0
0
## 19               0                    0              0
0
## 20               0                    0              0
0
## 21               0                    0              0
0
## 22               0                    0              0
0
## 23               0                    0              0
0
## 24               0                    0              0
0
## 25               0                    0              0
0
## 26               0                    0              0
```

```
0
## 27              0                   0             0
0
## 28              0                   0             0
0
## 29              0                   0             0
0
## 30              0                   0             0
0
## 31              0                   0             0
0
## 32              0                   0             0
0
## 33              0                   0             0
0
## 34              0                   0             0
0
## 35              0                   0             0
0
## 36              0                   0             0
0
## 37              0                   0             0
0
## 38              0                   0             0
0
## 39              0                   0             0
0
## 40              0                   0             0
0
## 41              0                   0             0
0
## 42              0                   0             0
0
## 43              0                   0             0
0
## 44              0                   0             0
0
## 45              0                   0             0
0
## 46              0                   0             0
0
## 47              0                   0             0
0
## 48              0                   0             0
0
## 49              0                   0             0
0
## 50              0                   0             0
0
## 51              0                   0             0
```

```
0
## 52              0                         0             0
0
## 53              0                         0             0
0
## 54              0                         0             0
0
## 55              0                         0             0
0
## 56              0                         0             0
0
## 57              0                         0             0
0
## 58              0                         0             0
0
## 59              0                         0             0
0
## 60              0                         0             0
0
## 61              0                         0             0
0
## 62              0                         0             0
0
## 63              0                         0             0
0
## 64              0                         0             0
0
## 65              0                         0             0
0
## 66              0                         0             0
0
## 67              0                         0             0
0
## 68              0                         0             0
0
## 69              0                         0             0
0
## 70              0                         0             0
0
## 71              0                         0             0
0
## 72              0                         0             0
0
## 73             NA                        NA            NA
NA
##    product_related product_related_duration bounce_rates exit_rates
page_values
## 1                1                         0          0.2        0.2
0
## 2                1                         0          0.2        0.2
```

```
0
## 3                1                    0          0.2        0.2
0
## 4                1                    0          0.2        0.2
0
## 5                1                    0          0.2        0.2
0
## 6                1                    0          0.2        0.2
0
## 7                1                    0          0.2        0.2
0
## 8                1                    0          0.2        0.2
0
## 9                1                    0          0.2        0.2
0
## 10               1                    0          0.2        0.2
0
## 11               1                    0          0.2        0.2
0
## 12               1                    0          0.2        0.2
0
## 13               1                    0          0.2        0.2
0
## 14               1                    0          0.2        0.2
0
## 15               1                    0          0.2        0.2
0
## 16               1                    0          0.2        0.2
0
## 17               1                    0          0.2        0.2
0
## 18               1                    0          0.2        0.2
0
## 19               1                    0          0.2        0.2
0
## 20               1                    0          0.2        0.2
0
## 21               1                    0          0.2        0.2
0
## 22               1                    0          0.2        0.2
0
## 23               1                    0          0.2        0.2
0
## 24               1                    0          0.2        0.2
0
## 25               1                    0          0.2        0.2
0
## 26               1                    0          0.2        0.2
0
## 27               1                    0          0.2        0.2
```

```
0
## 28                1                     0          0.2        0.2
0
## 29                1                     0          0.2        0.2
0
## 30                1                     0          0.2        0.2
0
## 31                1                     0          0.2        0.2
0
## 32                1                     0          0.2        0.2
0
## 33                1                     0          0.2        0.2
0
## 34                1                     0          0.2        0.2
0
## 35                1                     0          0.2        0.2
0
## 36                1                     0          0.2        0.2
0
## 37                1                     0          0.2        0.2
0
## 38                1                     0          0.2        0.2
0
## 39                1                     0          0.2        0.2
0
## 40                1                     0          0.2        0.2
0
## 41                1                     0          0.2        0.2
0
## 42                1                     0          0.2        0.2
0
## 43                1                     0          0.2        0.2
0
## 44                1                     0          0.2        0.2
0
## 45                1                     0          0.2        0.2
0
## 46                1                     0          0.2        0.2
0
## 47                1                     0          0.2        0.2
0
## 48                1                     0          0.2        0.2
0
## 49                1                     0          0.2        0.2
0
## 50                1                     0          0.2        0.2
0
## 51                1                     0          0.2        0.2
0
## 52                1                     0          0.2        0.2
```

```
0
## 53               1                         0          0.2        0.2
0
## 54               1                         0          0.2        0.2
0
## 55               1                         0          0.2        0.2
0
## 56               1                         0          0.2        0.2
0
## 57               1                         0          0.2        0.2
0
## 58               1                         0          0.2        0.2
0
## 59               1                         0          0.2        0.2
0
## 60               1                         0          0.2        0.2
0
## 61               1                         0          0.2        0.2
0
## 62               1                         0          0.2        0.2
0
## 63               1                         0          0.2        0.2
0
## 64               1                         0          0.2        0.2
0
## 65               1                         0          0.2        0.2
0
## 66               1                         0          0.2        0.2
0
## 67               1                         0          0.2        0.2
0
## 68               2                         0          0.2        0.2
0
## 69               2                         0          0.2        0.2
0
## 70               2                         0          0.2        0.2
0
## 71               2                         0          0.2        0.2
0
## 72               2                         0          0.2        0.2
0
## 73              NA                        NA           NA         NA
0
##    special_day month operating_systems browser region traffic_type
## 1          0.0   Dec                 1       1      1            1
## 2          0.0   Dec                 1       1      1            2
## 3          0.0   Dec                 1       1      1            3
## 4          0.0   Dec                 1       1      4            1
## 5          0.0   Dec                 1      13      9           20
## 6          0.0   Dec                 2       2      1            1
```

```
## 7          0.0   Dec                2        2      1              3
## 8          0.0   Dec                2        2      1             13
## 9          0.0   Dec                2        2      2              1
## 10         0.0   Dec                2        2      6             13
## 11         0.0   Dec                2        2      8              1
## 12         0.0   Dec                3        2      3              1
## 13         0.0   Dec                3        2      6              1
## 14         0.0   Dec                8       13      9             20
## 15         0.0   Feb                1        1      1              3
## 16         0.0   Feb                3        2      3              3
## 17         0.0  June                2        2      1              1
## 18         0.0  June                3        2      3             13
## 19         0.0   Mar                1        1      1              1
## 20         0.0   Mar                1        1      1              3
## 21         0.0   Mar                1        1      1              3
## 22         0.0   Mar                1        1      1              9
## 23         0.0   Mar                1        1      2              1
## 24         0.0   Mar                1        1      3              3
## 25         0.0   Mar                1        1      4              1
## 26         0.0   Mar                1        1      8              1
## 27         0.0   Mar                2        2      1              1
## 28         0.0   Mar                2        2      1              3
## 29         0.0   Mar                2        2      2              1
## 30         0.0   Mar                2        2      4              1
## 31         0.0   Mar                2        2      7              1
## 32         0.0   Mar                2        4      1              1
## 33         0.0   Mar                3        2      1              1
## 34         0.0   Mar                3        2      2              1
## 35         0.0   Mar                3        2      3              1
## 36         0.0   May                1        1      1              3
## 37         0.0   May                1        1      3              3
## 38         0.0   May                1        1      3             15
## 39         0.0   May                1        1      4              3
## 40         0.0   May                1        1      6              4
## 41         0.0   May                2        2      1              1
## 42         0.0   May                2        2      1              3
## 43         0.0   May                2        2      1              4
## 44         0.0   May                2        2      1             13
## 45         0.0   May                2        2      2              1
## 46         0.0   May                2        2      4              1
## 47         0.0   May                2        2      6              3
## 48         0.0   May                2        2      7              4
## 49         0.0   May                2        4      1              3
## 50         0.0   May                2        4      1              6
## 51         0.0   May                3        2      1             13
## 52         0.0   May                3        2      3              3
## 53         0.0   May                3        2      3             13
## 54         0.0   May                3        2      9              3
## 55         0.0   Nov                1        1      3              2
## 56         0.0   Nov                1        1      3              3
```

```
## 57         0.0   Nov                 1       1      4             1
## 58         0.0   Nov                 2       2      1             1
## 59         0.0   Nov                 2       2      3             1
## 60         0.0   Nov                 2       4      3             3
## 61         0.0   Nov                 3       2      1             1
## 62         0.0   Nov                 3       2      1            13
## 63         0.0   Nov                 3       2      3             3
## 64         0.0   Nov                 3       2      4             3
## 65         0.0   Nov                 3       2      7            13
## 66         0.6   May                 2       2      1             1
## 67         0.8   May                 2       2      1             1
## 68         0.0   Mar                 1       1      1             1
## 69         0.0   Mar                 2       2      1             1
## 70         0.0   Mar                 2       5      1             1
## 71         0.0   May                 2       2      2             3
## 72         0.0   Nov                 1       1      1             1
## 73         0.0   Mar                 2       2      1             2
##         visitor_type weekend revenue freq
## 1  Returning_Visitor    TRUE   FALSE    2
## 2        New_Visitor   FALSE   FALSE    1
## 3  Returning_Visitor   FALSE   FALSE    1
## 4  Returning_Visitor    TRUE   FALSE    2
## 5  Returning_Visitor   FALSE   FALSE    1
## 6  Returning_Visitor   FALSE   FALSE    2
## 7  Returning_Visitor   FALSE   FALSE    1
## 8  Returning_Visitor   FALSE   FALSE    2
## 9  Returning_Visitor   FALSE   FALSE    1
## 10 Returning_Visitor   FALSE   FALSE    1
## 11 Returning_Visitor   FALSE   FALSE    1
## 12 Returning_Visitor    TRUE   FALSE    1
## 13 Returning_Visitor   FALSE   FALSE    1
## 14             Other   FALSE   FALSE    4
## 15 Returning_Visitor   FALSE   FALSE    1
## 16 Returning_Visitor   FALSE   FALSE    1
## 17 Returning_Visitor   FALSE   FALSE    2
## 18 Returning_Visitor   FALSE   FALSE    1
## 19 Returning_Visitor    TRUE   FALSE    1
## 20 Returning_Visitor   FALSE   FALSE    1
## 21 Returning_Visitor    TRUE   FALSE    1
## 22 Returning_Visitor    TRUE   FALSE    1
## 23 Returning_Visitor   FALSE   FALSE    1
## 24 Returning_Visitor   FALSE   FALSE    1
## 25 Returning_Visitor   FALSE   FALSE    1
## 26 Returning_Visitor   FALSE   FALSE    2
## 27 Returning_Visitor   FALSE   FALSE   13
## 28 Returning_Visitor   FALSE   FALSE    1
## 29 Returning_Visitor   FALSE   FALSE    1
## 30 Returning_Visitor   FALSE   FALSE    2
## 31 Returning_Visitor   FALSE   FALSE    1
## 32 Returning_Visitor   FALSE   FALSE    2
```

```
## 33 Returning_Visitor   FALSE   FALSE    3
## 34 Returning_Visitor   FALSE   FALSE    1
## 35 Returning_Visitor   FALSE   FALSE    5
## 36 Returning_Visitor   FALSE   FALSE    5
## 37 Returning_Visitor   FALSE   FALSE    2
## 38 Returning_Visitor   FALSE   FALSE    1
## 39 Returning_Visitor   FALSE   FALSE    3
## 40 Returning_Visitor    TRUE   FALSE    1
## 41 Returning_Visitor   FALSE   FALSE    1
## 42 Returning_Visitor   FALSE   FALSE    6
## 43 Returning_Visitor   FALSE   FALSE    1
## 44 Returning_Visitor   FALSE   FALSE    1
## 45 Returning_Visitor   FALSE   FALSE    1
## 46 Returning_Visitor   FALSE   FALSE    1
## 47 Returning_Visitor   FALSE   FALSE    1
## 48 Returning_Visitor   FALSE   FALSE    1
## 49 Returning_Visitor   FALSE   FALSE    1
## 50 Returning_Visitor   FALSE   FALSE    1
## 51 Returning_Visitor   FALSE   FALSE    1
## 52 Returning_Visitor   FALSE   FALSE    1
## 53 Returning_Visitor   FALSE   FALSE    1
## 54 Returning_Visitor   FALSE   FALSE    1
## 55 Returning_Visitor   FALSE   FALSE    1
## 56 Returning_Visitor   FALSE   FALSE    1
## 57 Returning_Visitor   FALSE   FALSE    1
## 58 Returning_Visitor   FALSE   FALSE    3
## 59 Returning_Visitor   FALSE   FALSE    2
## 60 Returning_Visitor   FALSE   FALSE    1
## 61 Returning_Visitor   FALSE   FALSE    1
## 62 Returning_Visitor   FALSE   FALSE    1
## 63 Returning_Visitor   FALSE   FALSE    1
## 64 Returning_Visitor   FALSE   FALSE    1
## 65 Returning_Visitor   FALSE   FALSE    1
## 66 Returning_Visitor   FALSE   FALSE    1
## 67 Returning_Visitor   FALSE   FALSE    1
## 68 Returning_Visitor   FALSE   FALSE    1
## 69 Returning_Visitor   FALSE   FALSE    1
## 70 Returning_Visitor   FALSE   FALSE    1
## 71 Returning_Visitor   FALSE   FALSE    2
## 72 Returning_Visitor   FALSE   FALSE    1
## 73 Returning_Visitor   FALSE   FALSE    2
```

There is a total of 119 duplicated records in the data. We will remove them since duplicated data may imply inaccurate reporting and thus lead to less informed decisions.

```
# removing duplicated data
ecommerce_unique <- unique(ecommerce)

# confirming from the data for any duplicated records
anyDuplicated(ecommerce_unique)
```

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

8.3 Missing data

```
#check for missing data per column
colSums(is.na(ecommerce_unique))
```

```
##           administrative  administrative_duration           informational 
##                       12                       12                      12 
##   informational_duration          product_related product_related_duration 
##                       12                       12                      12 
##             bounce_rates             exit_rates              page_values 
##                       12                       12                       0 
##              special_day                    month       operating_systems 
##                        0                        0                       0 
##                  browser                   region             traffic_type 
##                        0                        0                       0 
##             visitor_type                  weekend                  revenue 
##                        0                        0                       0
```

Since the missing values occur in the numerical columns only, they can be filled with the means of the columns

```
# recode missing values with the means
ecommerce_unique$administrative[is.na(ecommerce_unique$administrative)] <-
mean(ecommerce_unique$administrative, na.rm = TRUE)

ecommerce_unique$administrative_duration[is.na(ecommerce_unique$administrativ
e_duration)] <- mean(ecommerce_unique$administrative_duration, na.rm = TRUE)

ecommerce_unique$informational[is.na(ecommerce_unique$informational)] <-
mean(ecommerce_unique$informational, na.rm = TRUE)

ecommerce_unique$informational_duration[is.na(ecommerce_unique$informational_
duration)] <- mean(ecommerce_unique$informational_duration, na.rm = TRUE)

ecommerce_unique$product_related[is.na(ecommerce_unique$product_related)] <-
mean(ecommerce_unique$product_related, na.rm = TRUE)

ecommerce_unique$product_related_duration[is.na(ecommerce_unique$product_rela
ted_duration)] <- mean(ecommerce_unique$product_related_duration, na.rm =
TRUE)

ecommerce_unique$bounce_rates[is.na(ecommerce_unique$bounce_rates)] <-
mean(ecommerce_unique$bounce_rates, na.rm = TRUE)

ecommerce_unique$exit_rates[is.na(ecommerce_unique$exit_rates)] <-
mean(ecommerce_unique$exit_rates, na.rm = TRUE)

# confirm from the data for any more missing values
colSums(is.na(ecommerce_unique))
```

```
##           administrative  administrative_duration            informational
##                        0                       0                        0
##   informational_duration         product_related product_related_duration
##                        0                       0                        0
##             bounce_rates             exit_rates              page_values
##                        0                       0                        0
##              special_day                   month        operating_systems
##                        0                       0                        0
##                  browser                  region             traffic_type
##                        0                       0                        0
##             visitor_type                 weekend                  revenue
##                        0                       0                        0
```

8.4 Outliers

These are data points that occur far away from the other points in the data. They could cause inconsistencies by distorting summaries of the distribution of values. We can screen for outliers by plotting boxplots of numerical columns in the data.

```
# get numerical columns in the data
nums <- unlist(lapply(ecommerce_unique, is.numeric))
nums
```

```
##           administrative  administrative_duration            informational
##                     TRUE                    TRUE                     TRUE
##   informational_duration         product_related product_related_duration
##                     TRUE                    TRUE                     TRUE
##             bounce_rates             exit_rates              page_values
##                     TRUE                    TRUE                     TRUE
##              special_day                   month        operating_systems
##                     TRUE                   FALSE                     TRUE
##                  browser                  region             traffic_type
##                     TRUE                    TRUE                     TRUE
##             visitor_type                 weekend                  revenue
##                    FALSE                   FALSE                    FALSE
```

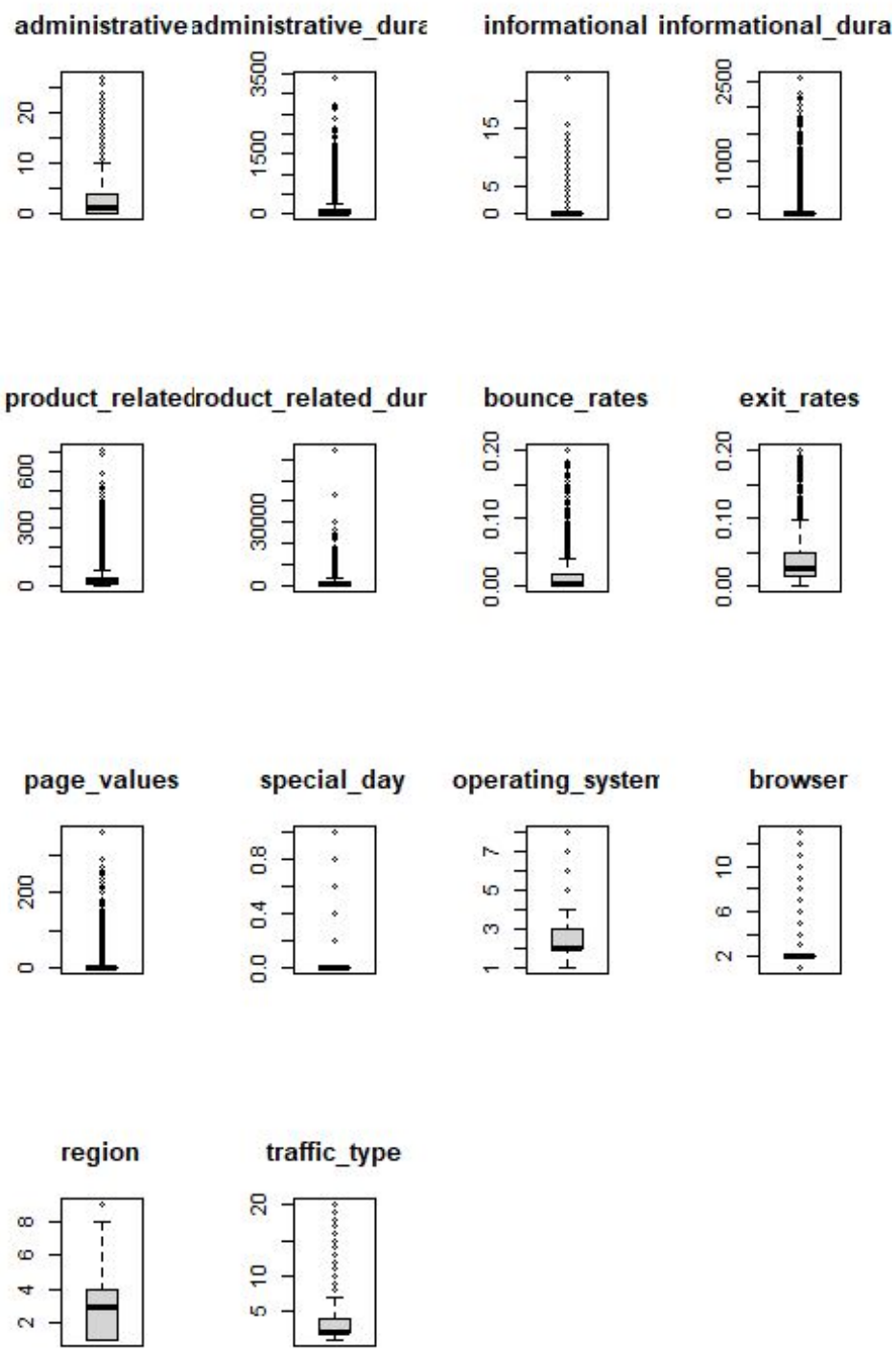
Out of the total 18 columns, there are 14 columns that contain numerical data. However, some of these columns are categorical in nature but do contain numerical data.

```
# output the numeric columns in form of a dataframe and check the top of the
resulting dataframe
numeric_cols <- ecommerce_unique[ , nums]
head(numeric_cols)
```

```
##   administrative administrative_duration informational
informational_duration
## 1              0                       0             0
0
## 2              0                       0             0
0
## 3              0                      -1             0
```

```
-1
## 4              0                        0             0
0
## 5              0                        0             0
0
## 6              0                        0             0
0
##   product_related product_related_duration bounce_rates exit_rates
page_values
## 1               1                  0.000000   0.20000000  0.2000000
0
## 2               2                 64.000000   0.00000000  0.1000000
0
## 3               1                 -1.000000   0.20000000  0.2000000
0
## 4               2                  2.666667   0.05000000  0.1400000
0
## 5              10                627.500000   0.02000000  0.0500000
0
## 6              19                154.216667   0.01578947  0.0245614
0
##   special_day operating_systems browser region traffic_type
## 1           0                 1       1      1            1
## 2           0                 2       2      1            2
## 3           0                 4       1      9            3
## 4           0                 3       2      2            4
## 5           0                 3       3      1            4
## 6           0                 2       2      1            3
```

```r
# make multiple boxplots of the numerical columns to check for any outliers
present
par ( mfrow= c (  2, 4 ))
for (i in 1 : length (numeric_cols)) {
boxplot (numeric_cols[,i], main= names (numeric_cols[i]), type= "l" )
}
```



From the boxplots above, there are a couple number of outliers in the numeric columns

```
colnames(ecommerce_unique)
```

```
##  [1] "administrative"          "administrative_duration"
##  [3] "informational"           "informational_duration"
##  [5] "product_related"         "product_related_duration"
##  [7] "bounce_rates"            "exit_rates"
##  [9] "page_values"             "special_day"
## [11] "month"                   "operating_systems"
## [13] "browser"                 "region"
## [15] "traffic_type"            "visitor_type"
## [17] "weekend"                 "revenue"
```

```
# checking to see the unique values of the categorical columns
unique(ecommerce_unique$operating_systems)
```

```
## [1] 1 2 4 3 7 6 8 5
```

```
unique(ecommerce_unique$browser)
```

```
##  [1]  1  2  3  4  5  6  7 10  8  9 12 13 11
```

```
unique(ecommerce_unique$region)
```

```
## [1] 1 9 2 3 4 5 6 7 8
```

```
unique(ecommerce_unique$traffic_type)
```

```
##  [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 18 19 16 17 20
```

```
unique(ecommerce_unique$special_day)
```

```
## [1] 0.0 0.4 0.8 1.0 0.2 0.6
```

Since values in the categorical columns are discrete in nature, removing outliers will make a poor representation of the original data. Hence, we will only deal with the outliers in the numerical columns with continuos data.

8.4.1 Dealing with outliers

Capping

```
# capping
# administrative
qnt <- quantile (ecommerce_unique$administrative, probs= c (.25 , .75 ),
na.rm = T)
caps <- quantile (ecommerce_unique$administrative, probs= c (.05 , .95 ),
na.rm = T)
H <- 1.5 * IQR (ecommerce_unique$administrative, na.rm = T)
ecommerce_unique$administrative[ecommerce_unique$administrative < (qnt[ 1 ] -
H)] <- caps[ 1 ]
ecommerce_unique$administrative[ecommerce_unique$administrative > (qnt[ 2 ] +
H)] <- caps[ 2 ]

# administrative_duration
```

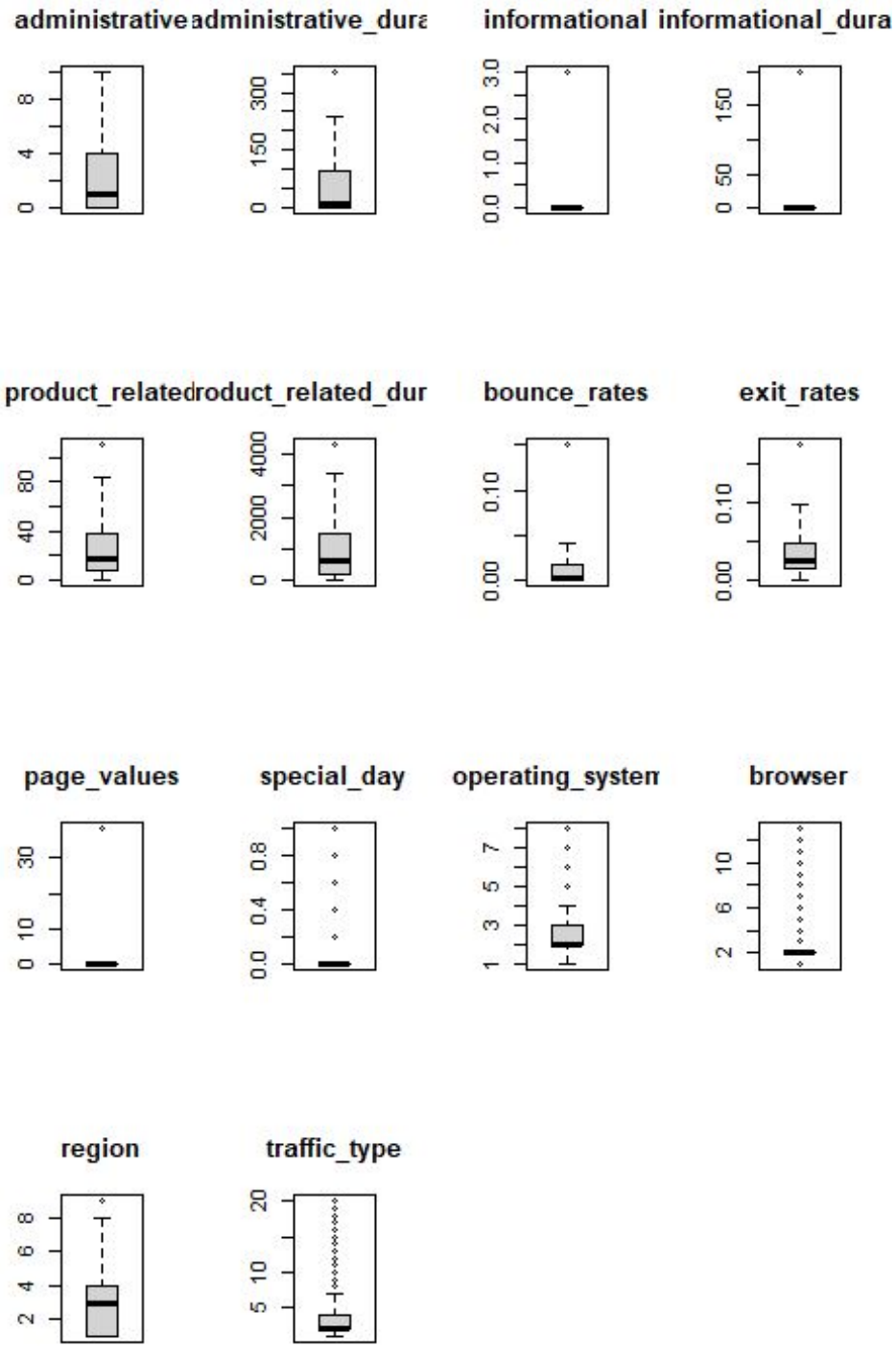
```r
qnt1 <- quantile (ecommerce_unique$administrative_duration, probs= c (.25 ,
.75 ), na.rm = T)
caps1 <- quantile (ecommerce_unique$administrative_duration, probs= c (.05 ,
.95 ), na.rm = T)
H <- 1.5 * IQR (ecommerce_unique$administrative_duration, na.rm = T)
ecommerce_unique$administrative_duration[ecommerce_unique$administrative_dura
tion < (qnt1[ 1 ] - H)] <- caps1[ 1 ]
ecommerce_unique$administrative_duration[ecommerce_unique$administrative_dura
tion > (qnt1[ 2 ] + H)] <- caps1[ 2 ]

# informational
qnt2 <- quantile (ecommerce_unique$informational, probs= c (.25 , .75 ),
na.rm = T)
caps2 <- quantile (ecommerce_unique$informational, probs= c (.05 , .95 ),
na.rm = T)
H <- 1.5 * IQR (ecommerce_unique$informational, na.rm = T)
ecommerce_unique$informational[ecommerce_unique$informational < (qnt2[ 1 ] -
H)] <- caps2[ 1 ]
ecommerce_unique$informational[ecommerce_unique$informational > (qnt2[ 2 ] +
H)] <- caps2[ 2 ]

# informational_duration
qnt3 <- quantile (ecommerce_unique$informational_duration, probs= c (.25 ,
.75 ), na.rm = T)
caps3 <- quantile (ecommerce_unique$informational_duration, probs= c (.05 ,
.95 ), na.rm = T)
H <- 1.5 * IQR (ecommerce_unique$informational_duration, na.rm = T)
ecommerce_unique$informational_duration[ecommerce_unique$informational_durati
on < (qnt3[ 1 ] - H)] <- caps3[ 1 ]
ecommerce_unique$informational_duration[ecommerce_unique$informational_durati
on > (qnt3[ 2 ] + H)] <- caps3[ 2 ]

# product_related
qnt4 <- quantile (ecommerce_unique$product_related, probs= c (.25 , .75 ),
na.rm = T)
caps4 <- quantile (ecommerce_unique$product_related, probs= c (.05 , .95 ),
na.rm = T)
H <- 1.5 * IQR (ecommerce_unique$product_related, na.rm = T)
ecommerce_unique$product_related[ecommerce_unique$product_related < (qnt4[ 1
] - H)] <- caps4[ 1 ]
ecommerce_unique$product_related[ecommerce_unique$product_related > (qnt4[ 2
] + H)] <- caps4[ 2 ]

# product_related_duration
qnt5 <- quantile (ecommerce_unique$product_related_duration, probs= c (.25 ,
.75 ), na.rm = T)
caps5 <- quantile (ecommerce_unique$product_related_duration, probs= c (.05 ,
.95 ), na.rm = T)
H <- 1.5 * IQR (ecommerce_unique$product_related_duration, na.rm = T)
```

```
ecommerce_unique$product_related_duration[ecommerce_unique$product_related_du
ration < (qnt5[ 1 ] - H)] <- caps5[ 1 ]
ecommerce_unique$product_related_duration[ecommerce_unique$product_related_du
ration > (qnt5[ 2 ] + H)] <- caps5[ 2 ]

# bounce_rates
qnt6 <- quantile (ecommerce_unique$bounce_rates, probs= c (.25 , .75 ), na.rm
= T)
caps6 <- quantile (ecommerce_unique$bounce_rates, probs= c (.05 , .95 ),
na.rm = T)
H <- 1.5 * IQR (ecommerce_unique$bounce_rates, na.rm = T)
ecommerce_unique$bounce_rates[ecommerce_unique$bounce_rates < (qnt6[ 1 ] -
H)] <- caps6[ 1 ]
ecommerce_unique$bounce_rates[ecommerce_unique$bounce_rates > (qnt6[ 2 ] +
H)] <- caps6[ 2 ]

# exit_rates
qnt7 <- quantile (ecommerce_unique$exit_rates, probs= c (.25 , .75 ), na.rm =
T)
caps7 <- quantile (ecommerce_unique$exit_rates, probs= c (.05 , .95 ), na.rm
= T)
H <- 1.5 * IQR (ecommerce_unique$exit_rates, na.rm = T)
ecommerce_unique$exit_rates[ecommerce_unique$exit_rates < (qnt7[ 1 ] - H)] <-
caps7[ 1 ]
ecommerce_unique$exit_rates[ecommerce_unique$exit_rates > (qnt7[ 2 ] + H)] <-
caps7[ 2 ]

# page_values
qnt8 <- quantile (ecommerce_unique$page_values, probs= c (.25 , .75 ), na.rm
= T)
caps8 <- quantile (ecommerce_unique$page_values, probs= c (.05 , .95 ), na.rm
= T)
H <- 1.5 * IQR (ecommerce_unique$page_values, na.rm = T)
ecommerce_unique$page_values[ecommerce_unique$page_values < (qnt8[ 1 ] - H)]
<- caps8[ 1 ]
ecommerce_unique$page_values[ecommerce_unique$page_values > (qnt8[ 2 ] + H)]
<- caps8[ 2 ]
```

To see the effects of the changes made, we will make another plot of boxplots to check or outliers present

```
# get numerical columns in the data
nums <- unlist(lapply(ecommerce_unique, is.numeric))

# output the numeric columns in form of a dataframe
numeric_cols <- ecommerce_unique[ , nums]

# make boxplots
par ( mfrow= c (  2, 4 ))
for (i in 1 : length (numeric_cols)) {
```

```
boxplot (numeric_cols[,i], main= names (numeric_cols[i]), type= "l" )
}
```



From the plots, we can see there are few to no outliers in the columns containing continuos numerical data. We will not deal with outliers in categorical data to avoid causing a lot of inconsistencies.

8.5 Anomalies

Anomalies are inconsistencies in the data and this can be checked for in many ways. These are rare items, events or observations which raise suspicions by differing significantly from the majority of the data.

9. Exploratory Data Analysis

9.1 Univariate Data Analysis

9.1.1 Measures of Central Tendency

Mean

Get the mean of all numerical columns

```
# get numerical columns and save them on a new dataframe that will be used
for analysis
numerical <- ecommerce_unique[c(1,2,3,4,5,6,7,8,9,10)]
# get the means of each column
colMeans(numerical)
```

```
##           administrative  administrative_duration            informational
##             2.189426e+00             6.881694e+01             6.490869e-01
##   informational_duration         product_related product_related_duration
##             3.937302e+01             2.907786e+01             1.073000e+03
##             bounce_rates              exit_rates              page_values
##             2.328499e-02             4.363287e-02             8.560655e+00
##              special_day
##             6.191139e-02
```

Median

Get the median of all numerical columns

```
apply (numerical, 2 ,median)
```

```
##           administrative  administrative_duration            informational
##             1.000000e+00             9.000000e+00             0.000000e+00
##   informational_duration         product_related product_related_duration
##             0.000000e+00             1.800000e+01             6.110000e+02
##             bounce_rates              exit_rates              page_values
##             2.941176e-03             2.500000e-02             0.000000e+00
##              special_day
##             0.000000e+00
```

*Mode

Administrative

```
# Create the function.
getmode <- function(v) {
uniqv <- unique(v)
uniqv[which.max(tabulate(match(v, uniqv)))]
}
# Calculate the mode using the user function.
getmode (ecommerce_unique$administrative)
```

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

Most users visiting the site did not visit administrative types of pages.

administrative_duration

```
getmode(ecommerce_unique$administrative_duration)
```

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

Since most users did not visit administrative page types, it is expected that the total time spent in this page category will be zero, which is true from the computed mode.

Informational

```
getmode(ecommerce_unique$informational)
```

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

Here again, most users visiting the site did not visit a page related to informational category.

Informational duration

```
getmode(ecommerce_unique$informational_duration)
```

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

As expected, the mode of the duration of time users take on pages related to the category "Informational" is zero.

Product Related

```
getmode(ecommerce_unique$product_related)
```

```
## [1] 110
```

The modal number of pages that a user visits related to the category "Product" is 110.

Product related duration

```
getmode(ecommerce_unique$product_related_duration)
```

```
## [1] 4312.682
```

The most occuring total time spent by a user on Product related page categories is 4312.682

Bounce Rates

```
getmode(ecommerce_unique$bounce_rates)
```

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

The modal percentage of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests to the analytics server during that session is 0%.

Exit rates

```
getmode(ecommerce_unique$exit_rates)
```

```
## [1] 0.175
```

For all pageviews to the page, the modal percentage that was the last in the session is 0.175

Page Values

```
getmode(ecommerce_unique$page_values)
```

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

The modal average value for a web page that a user visited before completing an e-commerce transaction is 0.

Special Day

```
getmode(ecommerce_unique$special_day)
```

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

The modal value of closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with the transaction is 0. This shows that special days do not have a significant effect on determining whether a user will visit the site to make a purchase or not.

Month

```
getmode(ecommerce_unique$month)
```

```
## [1] "May"
```

The modal month is May. It appears that most users visited the site during this month. Could it be because of a possible mid-year sale offered by the e-commerce site?

Visitor Type

```
getmode(ecommerce_unique$visitor_type)
```

```
## [1] "Returning_Visitor"
```

Most users visiting the site are “Returning Visitors”. These are visitors who are not new to the site.

Weekend

```
getmode(ecommerce_unique$weekend)
```

```
## [1] FALSE
```

Most users visit the site during weekdays.

Revenue

```
getmode(ecommerce_unique$revenue)
```

```
## [1] FALSE
```

Most users visiting the site did not purchase and hence revenue was not made from most visits.

9.1.2 Measures of Dispersion

*Find the **minimum, maximum and quantiles** of the columns in the data.*

```
summary(ecommerce_unique)
```

```
##  administrative   administrative_duration informational
##  Min.   : 0.000   Min.   : -1.00          Min.   :0.0000
##  1st Qu.: 0.000   1st Qu.:  0.00          1st Qu.:0.0000
##  Median : 1.000   Median :  9.00          Median :0.0000
##  Mean   : 2.189   Mean   : 68.82          Mean   :0.6491
##  3rd Qu.: 4.000   3rd Qu.: 94.60          3rd Qu.:0.0000
##  Max.   :10.000   Max.   :352.17          Max.   :3.0000
##  informational_duration product_related  product_related_duration
##  Min.   :  0.00         Min.   :  0.00   Min.   :  -1
##  1st Qu.:  0.00         1st Qu.:  8.00   1st Qu.: 194
##  Median :  0.00         Median : 18.00   Median : 611
##  Mean   : 39.37         Mean   : 29.08   Mean   :1073
##  3rd Qu.:  0.00         3rd Qu.: 38.00   3rd Qu.:1476
##  Max.   :199.00         Max.   :110.00   Max.   :4313
##   bounce_rates        exit_rates       page_values      special_day
##  Min.   :0.000000   Min.   :0.00000   Min.   : 0.000   Min.   :0.00000
##  1st Qu.:0.000000   1st Qu.:0.01426   1st Qu.: 0.000   1st Qu.:0.00000
##  Median :0.002941   Median :0.02500   Median : 0.000   Median :0.00000
##  Mean   :0.023285   Mean   :0.04363   Mean   : 8.561   Mean   :0.06191
##  3rd Qu.:0.016667   3rd Qu.:0.04847   3rd Qu.: 0.000   3rd Qu.:0.00000
##  Max.   :0.150000   Max.   :0.17500   Max.   :38.291   Max.   :1.00000
##     month           operating_systems    browser          region
##  Length:12211       Min.   :1.000     Min.   : 1.000   Min.   :1.000
```

```
##  Class :character   1st Qu.:2.000     1st Qu.: 2.000   1st Qu.:1.000
##  Mode  :character   Median :2.000     Median : 2.000   Median :3.000
##                     Mean   :2.124     Mean   : 2.358   Mean   :3.153
##                     3rd Qu.:3.000     3rd Qu.: 2.000   3rd Qu.:4.000
##                     Max.   :8.000     Max.   :13.000   Max.   :9.000
##   traffic_type    visitor_type        weekend         revenue
##  Min.   : 1.000   Length:12211       Mode :logical   Mode :logical
##  1st Qu.: 2.000   Class :character   FALSE:9352      FALSE:10303
##  Median : 2.000   Mode  :character   TRUE :2859      TRUE :1908
##  Mean   : 4.074
##  3rd Qu.: 4.000
##  Max.   :20.000
```

Range

Range is the difference between the maximum point and the minimum point in a set of data.

Administrative

```
# Get the range of each numerical column
range(ecommerce_unique$administrative)
```

```
## [1]  0 10
```

Pages of the administrative category range from 0-10

Administrative duration

```
range(ecommerce_unique$administrative_duration)
```

```
## [1]  -1.0000 352.1702
```

The range of total time spent by a user on administrative pages is -1 to 352.1702.

Informational

```
range(ecommerce_unique$informational)
```

```
## [1] 0 3
```

The range of informational pages is 0-3.

Informational Duration

```
range(ecommerce_unique$informational_duration)
```

```
## [1]   0 199
```

The total time a user spends on informational pages ranges between 0-199 minutes.

Product related

```
range(ecommerce_unique$product_related)
```

```
## [1]   0 110
```

Product related pages range between 0-110.

Product related duration

```
range(ecommerce_unique$product_related_duration)
```

```
## [1]   -1.000 4312.682
```

The total time a user spends time on product related pages ranges from -1 to 4312.682 minutes

Bounce rates

```
range(ecommerce_unique$bounce_rates)
```

```
## [1] 0.00 0.15
```

The percentage range of bounce rates is between 0-0.15

Exit rates

```
range(ecommerce_unique$exit_rates)
```

```
## [1] 0.000 0.175
```

The percentage range of exit rates is between 0-0.175

Page Values

```
range(ecommerce_unique$page_values)
```

```
## [1]  0.0000 38.2909
```

The average value of a web page that a user visited before completing an e-commerce transaction ranges between 0-38.2909.

Special day

```
range(ecommerce_unique$special_day)
```

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

Special days range between 0 and 1. 0 being- not close to a special day and 1 means a user visited the site on a day closer to a special day such as Mother's day.

Interquartile Range

The interquartile range also commonly known as IQR is the range between the 1st and 3rd quantiles. It is the difference between the two quantiles.

Administrative

```
IQR(ecommerce_unique$administrative)
```

```
## [1] 4
```

Administrative duration

```
IQR(ecommerce_unique$administrative_duration)
```

```
## [1] 94.6
```

Informational

```
IQR(ecommerce_unique$informational)
```

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

Informational duration

```
IQR(ecommerce_unique$informational_duration)
```

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

Product related

```
IQR(ecommerce_unique$product_related)
```

```
## [1] 30
```

Product related duration

```
IQR(ecommerce_unique$product_related_duration)
```

```
## [1] 1282.4
```

Bounce rates

```
IQR(ecommerce_unique$bounce_rates)
```

```
## [1] 0.01666667
```

Exit rates

```
IQR(ecommerce_unique$exit_rates)
```

```
## [1] 0.03421079
```

Page values

```
IQR(ecommerce_unique$page_values)
```

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

Special day

```
IQR(ecommerce_unique$special_day)
```

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

Standard Deviation

Find the standard deviation of the various columns in the data

```
apply (numerical, 2 ,sd)
```

```
##           administrative  administrative_duration            informational
##             2.846355e+00             1.070202e+02             1.235343e+00
##   informational_duration          product_related product_related_duration
##             7.928122e+01             3.038522e+01             1.214605e+03
##             bounce_rates               exit_rates              page_values
##             4.705701e-02             4.952400e-02             1.595404e+01
##              special_day
##             1.996219e-01
```

Variance

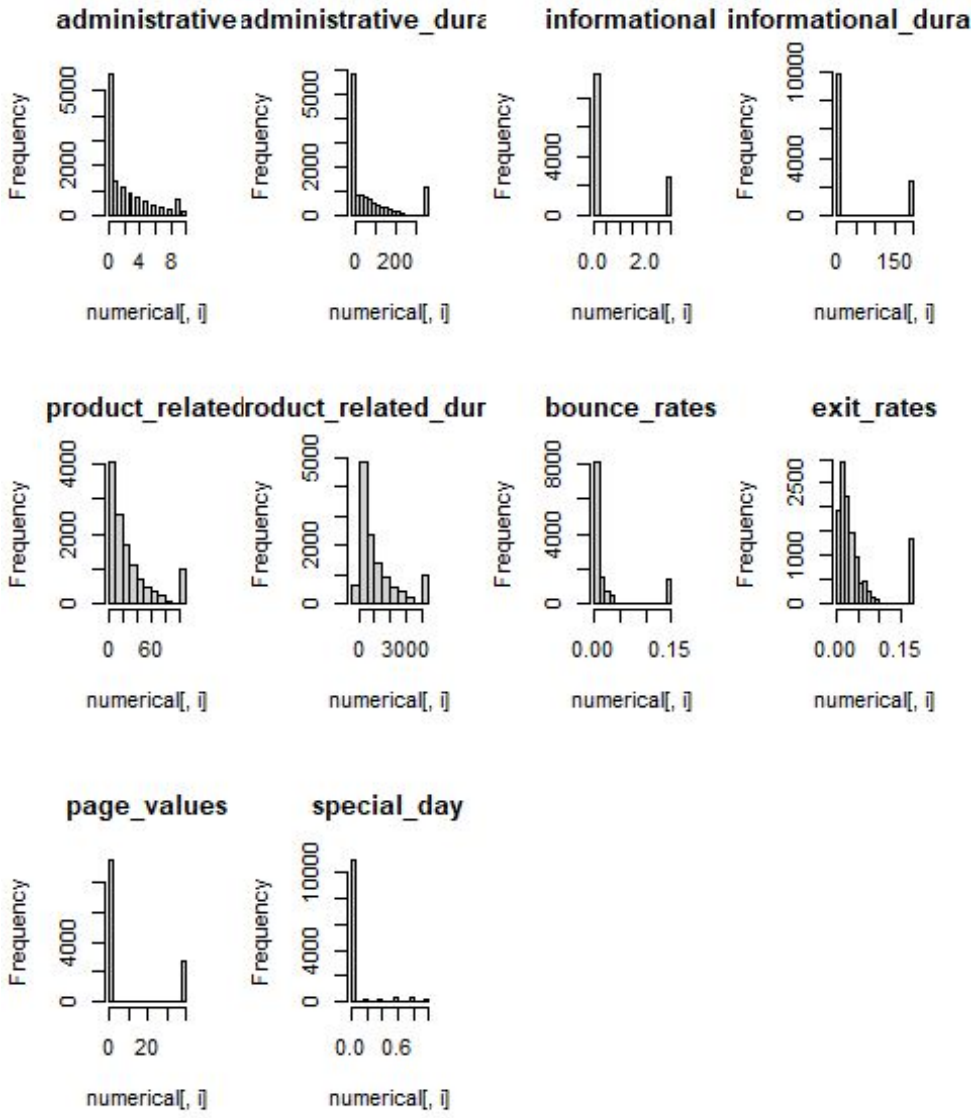
Find the variance of the numerical columns

```
sapply (numerical, var)
```

```
##           administrative  administrative_duration            informational
##             8.101739e+00             1.145333e+04             1.526072e+00
##   informational_duration          product_related product_related_duration
##             6.285512e+03             9.232614e+02             1.475267e+06
##             bounce_rates               exit_rates              page_values
##             2.214362e-03             2.452627e-03             2.545312e+02
##              special_day
##             3.984889e-02
```

Histograms

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



Most of the data is skewed to the right.

9.2 Bivariate and Multivariate analysis

Since our target variable is Revenue, we will investigate its relationship with the other variables.

```
# how often does a user make a purchase on the site if he/she visits
administrative pages
adm_revenue <- table(ecommerce_unique$administrative,
ecommerce_unique$revenue)
names(dimnames(adm_revenue)) <- c("admin" , "revenue" )
adm_revenue
```

```
##                    revenue
## admin               FALSE TRUE
##   0                  5123  514
##   1                  1063  291
##   2                   909  205
##   2.34002787113698     12    0
##   3                   741  174
##   4                   612  153
##   5                   457  118
##   6                   321  111
##   7                   272   66
##   8                   214   73
##   9                   458  171
##   10                  121   32
```

Most site visits that did not result in revenue had most users visiting administrative pages.

```
# does time spent on administrative pages result in purchase
adm_duration_revenue <- table(ecommerce_unique$administrative_duration,
ecommerce_unique$revenue)
names(dimnames(adm_duration_revenue)) <- c ("admin duration" , "revenue")

head(adm_duration_revenue)
```

```
##                revenue
## admin duration FALSE TRUE
##    -1             33    0
##    0            5192  548
##    1.333333333     1    0
##    2              13    2
##    3              22    4
##    3.5             4    0
```

```
tail(adm_duration_revenue)
```

```
##                revenue
## admin duration FALSE TRUE
##    236             1    0
##    236.0795455     1    0
```

```
##    236.25          0    1
##    236.4           1    0
##    236.4083333     1    0
##    352.1702381   866  284
```

Most users that did not spend time on administrative pages made a purchase on the ecommerce site and resulted in revenue for the ecommerce company.

```
# how often does a user make a purchase on the site if he/she visits
informational pages
info_revenue <- table(ecommerce_unique$informational,
ecommerce_unique$revenue)
names(dimnames(info_revenue)) <- c ( "informational" , "revenue" )
info_revenue
```

```
##              revenue
## informational FALSE TRUE
##             0  8274 1295
##             3  2029  613
```

Most site visitors who did not visit informational pages made a purchase and resulted in revenue for the ecommerce company.

```
# does time spent on informational pages result in revenue?
info_duration_revenue <- table(ecommerce_unique$informational_duration,
ecommerce_unique$revenue)
names(dimnames(info_duration_revenue)) <- c ( "info duration" , "revenue" )
info_duration_revenue
```

```
##              revenue
## info duration FALSE TRUE
##           0    8452 1343
##           199  1851  565
```

Most users who spent a lot of time on informational pages did not result in revenue for the ecommerce company while those who did not visit this category of pages made a purchase from the ecommerce site.

```
# how often does a user make a purchase on the site if he/she visits product
related pages
prod_revenue <- table(ecommerce_unique$product_related,
ecommerce_unique$revenue)
names(dimnames(prod_revenue)) <- c ( "Prod_related" , "revenue" )
prod_revenue
```

```
##                   revenue
## Prod_related       FALSE TRUE
##   0                   29    6
##   1                  501   13
##   2                  434   20
##   3                  431   25
```

```
##   4                  382   18
##   5                  360   20
##   6                  371   24
##   7                  356   35
##   8                  328   42
##   9                  283   34
##   10                 280   50
##   11                 272   36
##   12                 273   40
##   13                 244   45
##   14                 208   43
##   15                 230   40
##   16                 222   38
##   17                 186   40
##   18                 171   29
##   19                 176   42
##   20                 186   39
##   21                 157   42
##   22                 169   44
##   23                 147   33
##   24                 157   35
##   25                 132   22
##   26                 126   29
##   27                 141   36
##   28                 115   29
##   29                 107   28
##   30                 105   37
##   31                 102   26
##   32                  84   35
##   32.0584474137224    12    0
##   33                  97   24
##   34                  82   21
##   35                  82   19
##   36                  97   13
##   37                  92   26
##   38                  72   17
##   39                  90   19
##   40                  61   10
##   41                  69   16
##   42                  62   15
##   43                  59   14
##   44                  54   15
##   45                  56   15
##   46                  56   12
##   47                  46   10
##   48                  48   14
##   49                  48   12
##   50                  47   17
##   51                  41   11
##   52                  42    8
```

```
##   53                   46   14
##   54                   36   10
##   55                   39    6
##   56                   33   11
##   57                   34   13
##   58                   31   10
##   59                   37   11
##   60                   32   10
##   61                   35    5
##   62                   37    9
##   63                   29   10
##   64                   26   10
##   65                   29    4
##   66                   30    8
##   67                   24    6
##   68                   28    4
##   69                   22    8
##   70                   21    8
##   71                   27    6
##   72                   23    4
##   73                   19    5
##   74                   21    4
##   75                   10    7
##   76                   14    4
##   77                   18    6
##   78                   11    1
##   79                   25    6
##   80                   17    8
##   81                   25   12
##   82                   15    8
##   83                   13    8
##   110                 718  289
```

Product related pages 1-31 had the most number of visits in general and with the higher number of revenue returned. Page 110 also had a higher number of visits and more revenue from it was generated.

```
# does the duration of time spent on product related pages result in revenue?
prod_duration_revenue <- table(ecommerce_unique$product_related_duration,
ecommerce_unique$revenue)
names(dimnames(prod_duration_revenue)) <- c ( "prod_duration" , "revenue" )
# check the top of the dataframe
head(prod_duration_revenue)
```

```
##              revenue
## prod_duration FALSE TRUE
##   -1             33    0
##   0             589   13
##   0.5             1    0
##   1               2    0
```

```
##   2.333333333     1    0
##   2.666666667     1    0
```

```
# check the bottom of the dataframe
tail(prod_duration_revenue)
```

```
##               revenue
## prod_duration  FALSE TRUE
##   3391.68588       1    0
##   3393.903571      1    0
##   3394.130159      1    0
##   3395.729484      1    0
##   3397.957955      1    0
##   4312.6820515   665  289
```

Users who spent a lot of time on the ecommerce site on product related pages ended up bringing in revenue to the company. Users who did not spend time in these pages also brought revenue though not as much as those who spent a lot of time

```
# how does bounce rate affect revenue?
bounce_revenue <- table(ecommerce_unique$bounce_rates,
ecommerce_unique$revenue)
names(dimnames(bounce_revenue)) <- c ( "bounce_rates" , "revenue" )
# check the top of the dataframe
head(bounce_revenue)
```

```
##             revenue
## bounce_rates FALSE TRUE
##     0         4474 1036
##     2.73e-05     1    0
##     3.35e-05     1    0
##     3.83e-05     1    0
##     3.94e-05     0    1
##     7.09e-05     1    0
```

```
# check the bottom of the dataframe
tail(bounce_revenue)
```

```
##              revenue
## bounce_rates  FALSE TRUE
##   0.041176471     3    0
##   0.041269841     1    0
##   0.041333333     1    0
##   0.041463415     1    0
##   0.041666667     9    1
##   0.15         1406   25
```

0% bounce rate resulted in more revenue for the ecommerce company. The highest percentage of 0.15% bounce rate also resulted in more revenue but not as much as the zero percentage.

```
# how does exit rate affect revenue?
exit_revenue <- table(ecommerce_unique$exit_rates, ecommerce_unique$revenue)
names(dimnames(exit_revenue)) <- c ( "exit_rates" , "revenue" )
# check the top of the dataframe
head(exit_revenue)
```

```
##              revenue
## exit_rates    FALSE TRUE
##   0              42   34
##   0.000175593     1    0
##   0.000250438     1    0
##   0.000262123     1    0
##   0.000263158     1    0
##   0.000292398     1    0
```

```
# check the bottom of the dataframe
tail(exit_revenue)
```

```
##              revenue
## exit_rates    FALSE TRUE
##   0.096969697     1    0
##   0.097142857     1    0
##   0.097619048     1    0
##   0.097777778     2    0
##   0.098039216     1    0
##   0.175        1319    8
```

0% exit rate resulted in a higher revenue return for the ecommerce company.

```
# how does page value affect revenue?
page_revenue <- table(ecommerce_unique$page_values, ecommerce_unique$revenue)
names(dimnames(page_revenue)) <- c ( "page_values" , "revenue" )
# check the top of the dataframe
page_revenue
```

```
##              revenue
## page_values   FALSE TRUE
##   0            9111  370
##   38.29090231  1192 1538
```

Pages with a higher page value resulted in higher revenue returns for the ecommerce company.

```
# how do special days affect revenue?
special_revenue <- table(ecommerce_unique$special_day,
ecommerce_unique$revenue)
names(dimnames(special_revenue)) <- c ( "special_day" , "revenue" )
# check the top of the dataframe
special_revenue
```

```
##             revenue
## special_day FALSE TRUE
```

```
##          0    9131 1831
##          0.2   164   14
##          0.4   230   13
##          0.6   321   29
##          0.8   313   11
##          1     144   10
```

There are fewer visits to the site on Special days, which results in less revenue for the ecommerce company. More revenue is generated on non special days indicated by 0. This is expected as special days occur very few times in a year.

```
# Which months bring highest revenue?
month_revenue <- table(ecommerce_unique$month, ecommerce_unique$revenue)
names(dimnames(month_revenue)) <- c ( "month" , "revenue" )
# check the top of the dataframe
month_revenue
```

```
##       revenue
## month  FALSE TRUE
##   Aug    357   76
##   Dec   1490  216
##   Feb    179    3
##   Jul    366   66
##   June   256   29
##   Mar   1672  192
##   May   2964  365
##   Nov   2223  760
##   Oct    434  115
##   Sep    362   86
```

The month of May has the highest number of visits to the site, followed by November while February had the least number of visits. May however, does not result in a higher revenue as November does. It could be that more users make a lot of purchase to gift their loved ones over festivities such as Thanksgiving holiday, Christmas holiday, New year's holiday.

```
# how does visitor type affect revenue?
visitor_revenue <- table(ecommerce_unique$visitor_type,
ecommerce_unique$revenue)
names(dimnames(visitor_revenue)) <- c ( "visitor_type" , "revenue" )
# check the top of the dataframe
visitor_revenue
```

```
##                    revenue
## visitor_type        FALSE TRUE
##   New_Visitor        1271  422
##   Other                65   16
##   Returning_Visitor  8967 1470
```

Most users who are not new to the ecommerce site make purchases which result in more revenue generation by the ecommerce company.

```
# how does bounce rate affect revenue?
weekend_revenue <- table(ecommerce_unique$weekend, ecommerce_unique$revenue)
names(dimnames(weekend_revenue)) <- c ( "weekend" , "revenue" )
# check the top of the dataframe
weekend_revenue
```

```
##         revenue
## weekend FALSE TRUE
##   FALSE  7943 1409
##   TRUE   2360  499
```

More revenue is generated for site visits made during weekdays

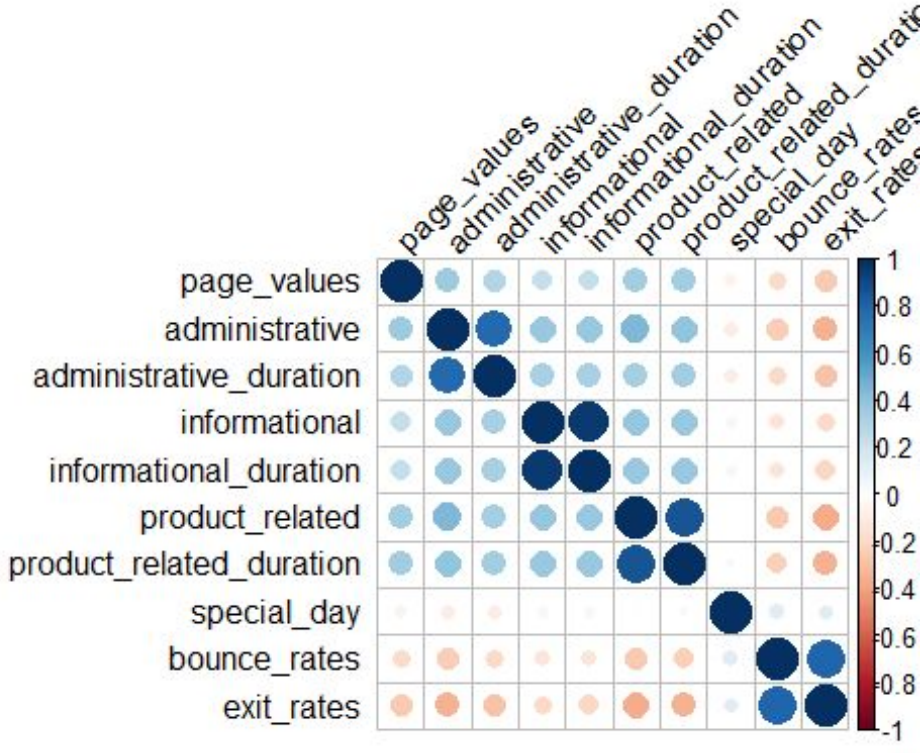
Correlation Matrix

Find the correlations of the numerical columns and make a correlation matrix plot

```
# find the correlations and round them off to 2 decimal places
res <- round(cor(numerical), 2 )
library (corrplot)
```

```
## corrplot 0.84 loaded
```

```
corrplot(res, type = "full", order = "hclust", tl.col = "black", tl.srt = 45
)
```



There is a high correlation between the following fields: administrative and administrative duration, informational and informational duration, product related and product related duration, bounce rates and exit rates.

10. Implementing the solution

10.1 Data Pre-processing

Before we begin modelling, we must ensure that the datatypes in the data we will use are in the appropriate mode i.e. numeric.

```
# check the datat types of the columns in the data
str(ecommerce_unique)
```

```
## 'data.frame':    12211 obs. of  18 variables:
##  $ administrative          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ administrative_duration : num  0 0 -1 0 0 0 -1 -1 0 0 ...
##  $ informational           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ informational_duration  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ product_related         : num  1 2 1 2 10 19 1 1 2 3 ...
##  $ product_related_duration: num  0 64 -1 2.67 627.5 ...
##  $ bounce_rates            : num  0.15 0 0.15 0.15 0.02 ...
##  $ exit_rates              : num  0.175 0.175 0.175 0.175 0.05 ...
##  $ page_values             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ special_day             : num  0 0 0 0 0 0 0.4 0 0.8 0.4 ...
##  $ month                   : chr  "Feb" "Feb" "Feb" "Feb" ...
##  $ operating_systems       : int  1 2 4 3 3 2 2 1 2 2 ...
##  $ browser                 : int  1 2 1 2 3 2 4 2 2 4 ...
##  $ region                  : int  1 1 9 2 1 1 3 1 2 1 ...
##  $ traffic_type            : int  1 2 3 4 4 3 3 5 3 2 ...
##  $ visitor_type            : chr  "Returning_Visitor" "Returning_Visitor"
"Returning_Visitor" "Returning_Visitor" ...
##  $ weekend                 : logi  FALSE FALSE FALSE FALSE TRUE FALSE ...
##  $ revenue                 : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

From the output, we can see that some of the fields we observed to be important during the EDA process are of character and logical types. The last 8 columns are categorical and can be converted to factor types for label encoding.

Encoding categorical variables

The easiest way to do this is to convert the variables to factor datatypes and then to numeric datatypes.

```
# encoding categorical variables
ecommerce_unique$month <- as.numeric(as.factor(ecommerce_unique$month))

ecommerce_unique$operating_systems <-
as.numeric(as.factor(ecommerce_unique$operating_systems))
```

```
ecommerce_unique$browser <- as.numeric(as.factor(ecommerce_unique$browser))

ecommerce_unique$region <- as.numeric(as.factor(ecommerce_unique$region))

ecommerce_unique$traffic_type <-
as.numeric(as.factor(ecommerce_unique$traffic_type))

ecommerce_unique$visitor_type <-
as.numeric(as.factor(ecommerce_unique$visitor_type))

ecommerce_unique$weekend <- as.numeric(as.factor(ecommerce_unique$weekend))

ecommerce_unique$revenue <- as.numeric(as.factor(ecommerce_unique$revenue))
```

Check the effect of the changes made

```
# check the datatypes
str(ecommerce_unique)
```

```
## 'data.frame':    12211 obs. of  18 variables:
##  $ administrative          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ administrative_duration : num  0 0 -1 0 0 0 -1 -1 0 0 ...
##  $ informational           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ informational_duration  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ product_related         : num  1 2 1 2 10 19 1 1 2 3 ...
##  $ product_related_duration: num  0 64 -1 2.67 627.5 ...
##  $ bounce_rates            : num  0.15 0 0.15 0.15 0.02 ...
##  $ exit_rates              : num  0.175 0.175 0.175 0.175 0.05 ...
##  $ page_values             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ special_day             : num  0 0 0 0 0 0 0.4 0 0.8 0.4 ...
##  $ month                   : num  3 3 3 3 3 3 3 3 3 3 ...
##  $ operating_systems       : num  1 2 4 3 3 2 2 1 2 2 ...
##  $ browser                 : num  1 2 1 2 3 2 4 2 2 4 ...
##  $ region                  : num  1 1 9 2 1 1 3 1 2 1 ...
##  $ traffic_type            : num  1 2 3 4 4 3 3 5 3 2 ...
##  $ visitor_type            : num  3 3 3 3 3 3 3 3 3 3 ...
##  $ weekend                 : num  1 1 1 1 2 1 1 2 1 1 ...
##  $ revenue                 : num  1 1 1 1 1 1 1 1 1 1 ...
```

All the variables are now in numeric type.

Feature Selection

We exclude our target variable from the features

```
# remove the target variable from the features
ecommerce_new <- ecommerce_unique[,
c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17)]

# save the target variable on a new dataframe
```

```r
ecommerce_label <- ecommerce_unique[, "revenue"]

#check the top of the two dataframes
head(ecommerce_new)
```

```
##   administrative administrative_duration informational
informational_duration
## 1              0                       0             0
0
## 2              0                       0             0
0
## 3              0                      -1             0
0
## 4              0                       0             0
0
## 5              0                       0             0
0
## 6              0                       0             0
0
##   product_related product_related_duration bounce_rates exit_rates
page_values
## 1               1                 0.000000   0.15000000  0.1750000
0
## 2               2                64.000000   0.00000000  0.1750000
0
## 3               1                -1.000000   0.15000000  0.1750000
0
## 4               2                 2.666667   0.15000000  0.1750000
0
## 5              10               627.500000   0.02000000  0.0500000
0
## 6              19               154.216667   0.01578947  0.0245614
0
##   special_day month operating_systems browser region traffic_type
visitor_type
## 1           0     3                 1       1      1            1
3
## 2           0     3                 2       2      1            2
3
## 3           0     3                 4       1      9            3
3
## 4           0     3                 3       2      2            4
3
## 5           0     3                 3       3      1            4
3
## 6           0     3                 2       2      1            3
3
##   weekend
## 1       1
## 2       1
```

```
## 3       1
## 4       1
## 5       2
## 6       1
```

```
head(ecommerce_label)
```

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

Normalization

Normalizing the variables in the dataset is done so that no particular attribute has more impact on the clustering algorithm than others

```
#normalize function
normalize <- function(x){
  return ((x-min(x)) / (max(x)-min(x)))
}

#apply the function on the features and check the top of the dataset
ecomm_new <- as.data.frame(lapply(ecommerce_new, normalize))
```

```
head(ecomm_new)
```

```
##   administrative administrative_duration informational
informational_duration
## 1              0             0.002831496             0
0
## 2              0             0.002831496             0
0
## 3              0             0.000000000             0
0
## 4              0             0.002831496             0
0
## 5              0             0.002831496             0
0
## 6              0             0.002831496             0
0
##   product_related product_related_duration bounce_rates exit_rates
page_values
## 1     0.009090909              0.0002318205    1.0000000  1.0000000
0
## 2     0.018181818              0.0150683336    0.0000000  1.0000000
0
## 3     0.009090909              0.0000000000    1.0000000  1.0000000
0
## 4     0.018181818              0.0008500086    1.0000000  1.0000000
0
## 5     0.090909091              0.1456991944    0.1333333  0.2857143
0
## 6     0.172727273              0.0359824078    0.1052632  0.1403509
```

```
0
##   special_day     month operating_systems    browser region traffic_type
## 1           0 0.2222222         0.0000000 0.00000000  0.000   0.00000000
## 2           0 0.2222222         0.1428571 0.08333333  0.000   0.05263158
## 3           0 0.2222222         0.4285714 0.00000000  1.000   0.10526316
## 4           0 0.2222222         0.2857143 0.08333333  0.125   0.15789474
## 5           0 0.2222222         0.2857143 0.16666667  0.000   0.15789474
## 6           0 0.2222222         0.1428571 0.08333333  0.000   0.10526316
##   visitor_type weekend
## 1            1       0
## 2            1       0
## 3            1       0
## 4            1       0
## 5            1       1
## 6            1       0
```

10.2 Modelling with K-Means Clustering

We can now build a clustering model with kmeans. Since we already know the expected number of clusters in our target field, we can specify the number of centroids,k=2(target revenue has two values TRUE and FALSE equal to two clusters)

```
#modeling using kmeans, k=2
kmeans_model<- kmeans(ecomm_new,2)
```

Previewing the number of records in each cluster

```
# number of records in each cluster
kmeans_model$size

## [1] 9683 2528
```

There are 2528 records in the first cluster and 9683 records in the second cluster, this sums to 12211 records which is the size of our initial data.

We can also get the value of cluster center datapoint for each variable.

```
# getting cluster center datapoints
kmeans_model$centers

##   administrative administrative_duration informational
informational_duration
## 1      0.1635340               0.1458120    0.01187648
0.0000000
## 2      0.4311741               0.3963806    0.99960443
0.9556962
##   product_related product_related_duration bounce_rates exit_rates
page_values
## 1       0.2095228                0.1939163   0.17883633   0.280029
0.1711247
## 2       0.4743265                0.4598673   0.06482646   0.131747
```

```
0.4244462
##   special_day     month operating_systems   browser    region traffic_type
## 1  0.06708665 0.5698303         0.1611514 0.1157441 0.2717133    0.1655261
## 2  0.04208861 0.5909810         0.1585104 0.1031777 0.2591970    0.1474850
##   visitor_type   weekend
## 1    0.8480326 0.2242074
## 2    0.8963608 0.2721519
```

We can also get the cluster vector that shows the cluster where each record falls

```
# getting cluster vector
kmeans_model$cluster
```

```
##     [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [37] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##    [73] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
##   [109] 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##   [145] 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##   [181] 1 1 2 1 1 2 2 2 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 1
##   [217] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 2 2 1
##   [253] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2
##   [289] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1
##   [325] 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1
##   [361] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1
##   [397] 1 2 1 1 2 1 1 1 1 1 2 1 1 2 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
##   [433] 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2
##   [469] 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2
##   [505] 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
##   [541] 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
##   [577] 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 1 1 1 1
##   [613] 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1
##   [649] 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2
```

```
1 1 1
##   [685] 1 2 1 2 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1
##   [721] 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 2 1
1 1 1
##   [757] 1 1 1 1 2 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1
2 1 1
##   [793] 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1
1 1 1
##   [829] 1 1 1 1 1 1 2 2 1 1 2 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 1 1
##   [865] 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 2 2 1 1
2 1 1
##   [901] 1 1 1 1 1 1 1 2 2 1 2 2 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1
1 1 1
##   [937] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1
##   [973] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
1 1 1
##  [1009] 1 1 1 2 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1
1 1 1
##  [1045] 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1
1 1 1
##  [1081] 2 2 1 1 2 2 2 1 1 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
1 2 2
##  [1117] 2 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1
2 1 1
##  [1153] 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1
1 1 1
##  [1189] 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [1225] 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 2 2 1 1 2 1
1 1 2
##  [1261] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2
1 1 1
##  [1297] 1 2 1 1 2 2 2 1 1 2 1 1 1 1 2 2 1 1 1 2 1 2 2 1 1 1 2 1 1 1 1 2 1
1 1 1
##  [1333] 1 1 2 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
1 1 1
##  [1369] 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1
1 1 1
##  [1405] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 2 1
2 1 1
##  [1441] 2 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1
1 1 1
##  [1477] 1 1 2 1 1 1 1 1 2 2 1 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [1513] 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1
1 2 1
##  [1549] 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1
```

```
1 1 1
##  [1585] 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 2 1
1 2 1
##  [1621] 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1
1 1 1
##  [1657] 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1
1 1 1
##  [1693] 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 2 1
##  [1729] 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 2 1 1 2
1 1 1
##  [1765] 2 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 2
1 1 1
##  [1801] 2 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 2 1 1 1 2 2 1 1 1 1
1 1 2
##  [1837] 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 2 1 2 2 1 1 1
1 1 1
##  [1873] 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1
1 1 1
##  [1909] 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1
1 1 1
##  [1945] 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 2 1 2 2 2 1 1 2 2 1 1
1 1 1
##  [1981] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [2017] 1 1 1 2 1 2 1 1 1 2 2 1 2 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2
1 1 1
##  [2053] 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 2
##  [2089] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [2125] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 2 2 1 1 1 1 1 1 2
2 2 1
##  [2161] 1 1 1 1 2 1 1 1 2 2 1 2 1 1 2 1 1 1 1 1 2 1 2 1 1 2 1 1 1 2 1 1 2
1 1 1
##  [2197] 1 1 1 2 1 1 1 2 1 2 1 1 2 2 1 2 2 2 2 1 2 1 1 1 1 1 1 1 1 1 2 1
2 1 1
##  [2233] 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1
1 2 2
##  [2269] 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1
1 2 1
##  [2305] 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1
##  [2341] 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 1 1 1 1 1 2 1 1 1 1 2 1 1 2 2 1 1 1
2 1 1
##  [2377] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1 1 1 2 1 1 1
1 1 1
##  [2413] 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 2 2 1 1 2 1 2 1 1 1 1 1 2 1 2 1 1 2 1
1 1 1
##  [2449] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 2 2
```

```
1 1 2
##  [2485] 2 1 1 1 1 1 1 2 2 2 1 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 2 2 1
1 2 1
##  [2521] 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1
2 1 1
##  [2557] 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 2 2 1 1
2 1 1
##  [2593] 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1
1 1 1
##  [2629] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 1 2
##  [2665] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
2 2 1
##  [2701] 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 2 1 1 2 2 1 1 1
1 2 2
##  [2737] 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1
##  [2773] 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1
1 2 1
##  [2809] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 2 1
##  [2845] 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1
1 1 2
##  [2881] 1 1 1 1 2 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [2917] 2 1 1 2 2 2 2 1 1 2 2 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 2
1 1 1
##  [2953] 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1
1 1 1
##  [2989] 1 1 2 2 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 2 2 2 1 1 2 1
1 1 1
##  [3025] 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1
1 1 1
##  [3061] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1
1 1 1
##  [3097] 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1
1 2 2
##  [3133] 1 1 1 1 1 2 2 1 1 2 1 1 1 2 1 2 1 2 1 1 2 1 2 2 1 2 1 1 1 1 1 2 1
2 1 1
##  [3169] 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
1 1 1
##  [3205] 2 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1
1 1 1
##  [3241] 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1
1 1 1
##  [3277] 1 2 2 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [3313] 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1
1 2 1
##  [3349] 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 2 1 1
```

```
1 1 1
##  [3385] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
2 1 1
##  [3421] 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1
1 1 2
##  [3457] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 2 1 1 1 1 1
1 2 2
##  [3493] 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1
1 1 1
##  [3529] 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1
1 1 1
##  [3565] 1 1 1 2 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 1
1 1 1
##  [3601] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2
1 1 2
##  [3637] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1
1 2 1
##  [3673] 2 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 2 2 2 1
1 1 1
##  [3709] 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1
1 1 1
##  [3745] 1 1 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1
2 1 2
##  [3781] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 2 2 1 1 1
1 1 1
##  [3817] 1 1 1 1 2 2 1 2 1 2 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [3853] 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1
1 1 1
##  [3889] 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 2 2 1 2 1 1 1 1 1 1 1
1 1 1
##  [3925] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1
2 1 1
##  [3961] 1 1 1 1 1 1 1 1 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1
2 1 1
##  [3997] 2 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 2 1 1
1 1 1
##  [4033] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 2 1 2 1 1 2 1
1 1 1
##  [4069] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1
2 1 1
##  [4105] 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1
1 1 1
##  [4141] 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 2
1 1 2
##  [4177] 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1
1 2 1
##  [4213] 2 1 1 1 1 1 1 2 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1
1 1 1
##  [4249] 1 1 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1
```

```
1 1 1
##  [4285] 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 1 1
##  [4321] 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1
##  [4357] 1 1 1 2 1 2 1 1 1 1 1 1 2 2 2 1 2 1 2 1 1 1 1 1 2 2 1 1 1 1 2 2 2
1 1 1
##  [4393] 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 1 1
##  [4429] 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 2 2 1 1 1 2 2 1 1 1 2 2 1 1 1 1
1 2 1
##  [4465] 1 2 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 2 2 1
2 1 1
##  [4501] 1 1 2 1 1 1 1 2 1 1 2 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2
1 1 2
##  [4537] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1
1 1 1
##  [4573] 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2
1 1 1
##  [4609] 1 1 1 1 1 1 2 2 1 1 1 1 1 1 2 2 1 1 1 2 1 2 1 2 1 1 2 1 2 1 1 1 1
1 1 1
##  [4645] 2 1 2 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [4681] 1 1 1 2 1 2 1 1 1 1 1 1 2 2 2 2 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2
1 2 1
##  [4717] 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
1 2 1
##  [4753] 1 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1
1 1 2
##  [4789] 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1
2 1 1
##  [4825] 1 1 2 1 1 2 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
1 1 1
##  [4861] 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 2 1 2 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1
1 2 1
##  [4897] 1 2 1 2 1 1 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 2 1 1 1 1
2 1 1
##  [4933] 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 1 2 1 1 2 1 1 1 2 1 1
1 2 1
##  [4969] 1 2 2 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1
1 1 2
##  [5005] 1 1 1 2 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1
1 1 1
##  [5041] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 1 2 1 1 1 1
2 1 1
##  [5077] 1 1 2 2 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 2
##  [5113] 1 2 1 2 2 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 2 1 2 1 2 1 1
2 2 2
##  [5149] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 2 1 1 1 1 1 1 2 2 1 1 1 1
```

```
1 1 1
##  [5185] 1 2 1 2 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
2 1 1
##  [5221] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1
1 1 2
##  [5257] 2 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1
1 2 2
##  [5293] 2 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1
2 1 1
##  [5329] 1 1 1 1 1 2 1 1 1 1 1 2 2 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 1 2 1 1 2
1 1 1
##  [5365] 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1
1 1 1
##  [5401] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2
1 1 1
##  [5437] 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
1 1 1
##  [5473] 2 1 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2
2 1 1
##  [5509] 1 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1
1 1 1
##  [5545] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1
1 2 2
##  [5581] 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 2 1 1 1 1 2 2 1 1
1 1 1
##  [5617] 1 1 2 1 2 1 1 2 1 1 1 2 2 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
1 2 1
##  [5653] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2
1 1 1
##  [5689] 1 2 2 1 2 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1
1 2 2
##  [5725] 1 1 2 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 2 2 1 2
1 1 2
##  [5761] 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1
2 1 2
##  [5797] 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 2 2 2 1 1 1 2 1 2
1 1 1
##  [5833] 2 1 1 1 2 1 2 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 2 1 1 2 2 1 1 1 1 1
1 1 1
##  [5869] 2 1 1 1 2 1 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
2 1 1
##  [5905] 1 1 1 1 1 2 2 1 1 2 1 1 2 1 1 1 1 1 2 1 2 2 1 1 1 1 1 2 1 1 1 1 1
1 1 1
##  [5941] 2 1 2 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1
1 1 1
##  [5977] 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 2 1 1 1 2 1 1
1 1 1
##  [6013] 2 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2
1 1 2
##  [6049] 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1
```

```
1 2 1
##  [6085] 1 2 2 2 1 1 2 1 1 2 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1
1 1 2
##  [6121] 1 2 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1 1 1
1 2 1
##  [6157] 1 1 1 1 2 2 1 2 1 1 1 1 2 2 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
1 1 1
##  [6193] 1 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1 1 1 2 2 1 1 1 1 2 1 1
1 1 1
##  [6229] 1 1 1 2 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1 2 1 1 2 2 1 1 2 2 1 1 1 2
1 1 1
##  [6265] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 2 2 1 1 1 1
1 1 2
##  [6301] 2 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 1 1
##  [6337] 1 1 1 1 1 2 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1 2 1
1 1 1
##  [6373] 1 1 1 1 1 2 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 2 1 1 2 2 1 2 1 1 1 1 1
1 1 2
##  [6409] 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 1 2
1 1 1
##  [6445] 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
1 1 2
##  [6481] 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 2 2 1 1
2 1 2
##  [6517] 1 1 2 1 1 2 1 2 1 2 1 2 1 1 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1 2 1
1 1 1
##  [6553] 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
2 2 1
##  [6589] 1 1 1 2 2 1 2 1 1 1 2 1 1 2 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1
1 1 1
##  [6625] 1 1 1 1 2 2 2 2 2 1 2 2 1 1 1 2 1 1 1 1 1 1 2 1 2 1 2 1 1 2 1 1 1
1 1 1
##  [6661] 1 1 2 1 1 2 1 1 1 2 1 1 1 1 2 1 2 2 1 1 1 1 1 2 1 1 1 1 1 2 2 1 1
1 1 1
##  [6697] 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1
2 1 2
##  [6733] 2 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 2 2 1 2 2 2 1 2 1 1 1 1
2 1 1
##  [6769] 2 1 1 1 2 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1
1 2 1
##  [6805] 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 2 2 1 2 2 1 1 1 1 2 2 1
1 1 1
##  [6841] 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1
1 1 2
##  [6877] 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1
1 1 1
##  [6913] 2 1 1 1 2 2 2 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1
1 2 1
##  [6949] 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 2 2 1 2 2 1 1 1 1
```

```
1 1 1
##  [6985] 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 2 2 1 2 2 1 1 1
2 1 1
##  [7021] 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 2 1 1 1 1 1 2 1 1 1
2 1 1
##  [7057] 1 1 1 2 1 2 1 2 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1
1 1 1
##  [7093] 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2
1 1 2
##  [7129] 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1
1 2 1
##  [7165] 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1
1 2 2
##  [7201] 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 2 1 1
2 1 1
##  [7237] 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 1 2 1 2 1
1 1 2
##  [7273] 1 1 1 2 2 1 1 1 2 2 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1
2 1 1
##  [7309] 2 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1
1 2 1
##  [7345] 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 2 2 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1
1 1 1
##  [7381] 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 2 1 2 1 2 1 2 1 1 1 1 1 2 1 2 1 2
1 1 1
##  [7417] 2 1 1 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1
1 2 2
##  [7453] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 2 1 1 2 2 1 1 1 1 1 1
1 1 1
##  [7489] 1 1 2 1 2 2 2 2 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1
2 1 1
##  [7525] 1 1 1 1 2 2 1 1 1 1 2 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
1 1 1
##  [7561] 2 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 1 1 1 2 1 2 1 2 1 1 2 2 2 1 1 1 1
1 2 1
##  [7597] 2 1 1 1 1 2 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 2
1 1 2
##  [7633] 2 1 1 2 1 1 1 2 2 1 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 2 2
1 1 1
##  [7669] 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 2 1
##  [7705] 1 1 1 1 2 1 1 1 1 1 2 2 1 1 1 1 2 2 1 1 2 1 1 1 2 2 1 1 1 1 1 1 2
1 2 1
##  [7741] 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 2 1
2 1 2
##  [7777] 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 2 2 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1
2 2 1
##  [7813] 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 2 2 2
1 1 1
##  [7849] 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1
```

```
2 2 1
##  [7885] 2 1 2 2 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 2 2 1 1 1 1
1 1 1
##  [7921] 2 1 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1
1 1 1
##  [7957] 1 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2
1 1 1
##  [7993] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1
1 2 1
##  [8029] 1 1 1 1 1 1 2 1 2 2 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1
1 1 1
##  [8065] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1
1 1 1
##  [8101] 1 1 1 2 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 1 1
##  [8137] 1 1 1 1 2 1 1 1 1 2 1 2 2 1 1 2 1 1 1 2 1 1 1 1 1 1 2 2 1 1 1 1 1
2 1 1
##  [8173] 1 1 1 1 1 2 1 1 1 2 1 1 1 1 2 2 1 2 1 1 1 1 2 1 2 1 2 2 1 2 1 1 1
1 1 1
##  [8209] 2 1 1 2 1 1 1 1 1 2 2 1 2 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [8245] 1 1 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 2 1 2 1 1
1 1 1
##  [8281] 1 1 1 1 1 1 1 1 2 2 1 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1
1 2 1
##  [8317] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [8353] 2 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2
2 1 1
##  [8389] 1 1 1 1 1 2 1 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 2
1 2 1
##  [8425] 1 1 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 2 1 1 1 1 2 1 1 1 1
1 2 2
##  [8461] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
##  [8497] 1 1 1 1 2 1 1 2 1 1 1 1 1 2 2 1 2 1 1 1 2 1 2 1 1 1 1 2 2 1 1 1 1
2 1 1
##  [8533] 1 1 2 1 1 1 2 2 1 2 1 1 1 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 2 2 1 2
1 1 1
##  [8569] 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1
1 1 2
##  [8605] 2 2 1 1 2 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1
1 1 2
##  [8641] 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2
1 1 1
##  [8677] 1 1 1 2 2 1 1 1 2 2 2 1 2 1 1 1 1 2 1 2 1 1 1 1 2 1 1 2 1 1 1 1 2
1 1 1
##  [8713] 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 2 1 1 2 1
1 2 2
##  [8749] 1 2 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1
```

```
1 1 2
##  [8785] 1 2 2 1 2 1 2 1 1 2 2 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1
2 2 1
##  [8821] 2 1 2 2 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1
2 1 1
##  [8857] 1 1 1 1 1 1 1 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2
1 2 2
##  [8893] 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 2 1 1 1
1 1 1
##  [8929] 1 1 1 2 1 1 2 2 1 1 2 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1 1 2 2 1 2 2
1 1 1
##  [8965] 1 1 2 1 2 1 2 1 2 1 2 2 1 1 1 1 1 1 2 2 1 2 2 1 2 1 1 1 1 1 1 1 1
1 2 1
##  [9001] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1 2 1 1 1 1 2 1 1 1
1 2 1
##  [9037] 1 1 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 2 1
1 2 1
##  [9073] 1 1 1 2 2 1 1 2 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1
1 1 1
##  [9109] 1 1 1 2 1 2 1 1 1 1 2 1 2 1 1 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1
1 1 1
##  [9145] 1 2 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 2 1 1 2 2 1 2 2 1 1 1 1 2 1 1 2
1 1 1
##  [9181] 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 1 2 1 2 1 1 1
1 2 1
##  [9217] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1
2 1 2
##  [9253] 1 1 1 1 2 1 1 2 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 2 1 1 1 2 1
2 2 1
##  [9289] 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2 2 1 1 2 1 1
1 1 1
##  [9325] 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2
1 1 1
##  [9361] 1 2 1 1 1 1 1 1 1 1 2 2 2 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 2 1
1 1 1
##  [9397] 2 1 2 2 1 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2
1 1 1
##  [9433] 1 1 2 1 1 1 2 1 2 2 2 1 1 1 1 1 1 1 2 2 1 1 1 2 1 1 1 2 1 1 2 1 1
1 1 1
##  [9469] 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1
1 1 2
##  [9505] 2 2 2 1 2 2 1 1 1 1 1 2 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1 1
1 1 1
##  [9541] 1 1 1 1 2 2 2 1 1 1 2 2 1 1 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 2 1
1 1 2
##  [9577] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 2 2 1 1 1 1 2 2 1
1 1 1
##  [9613] 1 1 1 2 2 1 1 1 2 1 1 2 2 2 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1
1 2 1
##  [9649] 1 2 1 1 1 1 1 1 2 1 2 2 1 1 1 1 1 2 1 1 1 2 1 1 1 2 2 1 1 1 1 1 2
```

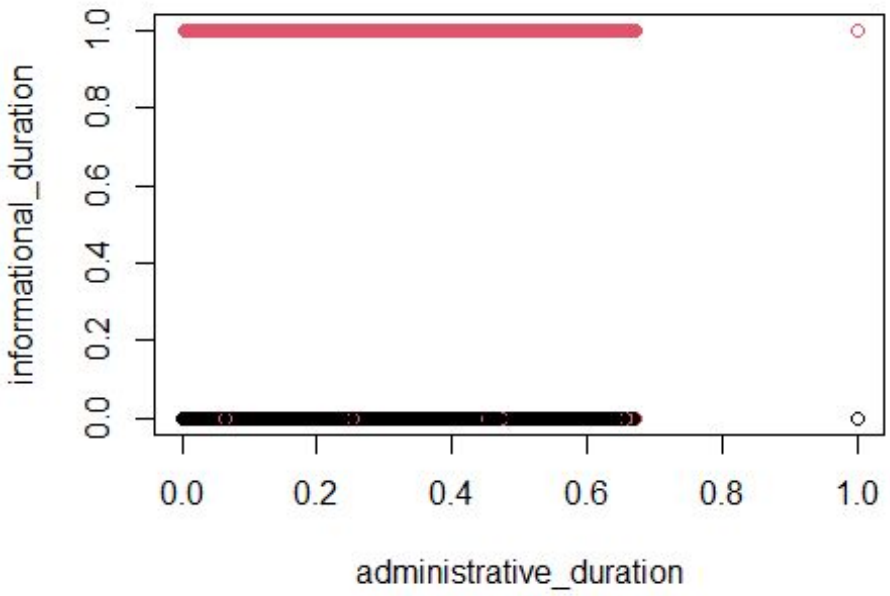
```
1 1 1
##  [9685] 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1
1 1 2
##  [9721] 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 2 2 1 1 1 1
1 2 2
##  [9757] 1 1 2 1 2 2 1 1 2 1 2 1 2 2 1 1 2 1 1 1 1 1 1 1 2 1 2 1 2 1 2 1 2
2 1 1
##  [9793] 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1
2 1 1
##  [9829] 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 2
##  [9865] 1 2 2 1 2 2 2 1 1 1 1 1 2 1 2 1 2 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1
1 1 2
##  [9901] 1 2 2 1 1 2 1 1 2 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
2 1 1
##  [9937] 1 1 1 2 2 1 1 2 1 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1
1 1 1
##  [9973] 1 2 2 1 1 1 1 2 2 1 1 2 2 1 2 2 1 2 2 1 2 1 2 1 2 2 1 1 1 1 1 1 1
1 1 1
## [10009] 1 2 2 2 2 1 1 2 1 2 1 1 2 1 1 1 2 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1
2 2 1
## [10045] 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2
2 2 2
## [10081] 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 1 1 1 1 2 1 2 1 2 2 2 1 1 2 2 1 1 1
2 1 1
## [10117] 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 1 1
## [10153] 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 1 1 1 2 2
2 1 1
## [10189] 2 1 1 1 2 1 2 1 1 2 2 1 2 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1
2 1 1
## [10225] 1 2 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 2
1 1 1
## [10261] 1 2 2 1 2 1 1 1 1 2 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 2 2 2 1 2 1
1 1 1
## [10297] 2 1 1 1 1 1 1 2 1 2 2 1 2 2 1 1 1 2 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1
1 1 1
## [10333] 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 2
1 2 1
## [10369] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1
1 1 1
## [10405] 1 1 2 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1
2 1 1
## [10441] 1 1 2 1 1 1 1 2 1 1 2 1 1 1 2 2 2 1 2 2 1 1 2 1 1 1 1 1 1 1 1 2 1
1 1 1
## [10477] 1 1 1 2 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 2 1
2 1 2
## [10513] 1 1 1 1 1 1 1 1 2 1 1 2 2 2 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 2 1 2
1 2 1
## [10549] 1 1 2 1 1 1 1 1 1 1 1 2 2 1 2 1 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
```

```
1 1 1
## [10585] 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 2 1 2
2 1 1
## [10621] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 2 1 2 1 1
1 2 1
## [10657] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1
2 1 1
## [10693] 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 2 1 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1
2 1 2
## [10729] 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1
2 1 1
## [10765] 1 1 2 1 2 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2
1 1 1
## [10801] 1 2 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1 1
1 1 1
## [10837] 1 1 2 1 2 1 1 2 1 1 1 2 1 1 1 2 1 2 1 2 1 1 1 1 1 1 1 2 1 1 2 2 2
1 1 1
## [10873] 2 1 1 1 1 1 2 1 1 2 1 2 2 2 2 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 2 1 1
1 1 1
## [10909] 1 1 1 2 2 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1
1 1 1
## [10945] 1 1 1 2 1 1 1 1 1 1 2 1 2 2 1 1 2 1 1 2 1 2 2 1 1 1 1 1 1 1 2 2 1 1
1 1 1
## [10981] 1 1 2 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 1 1 1 2 2 1 1
2 1 2
## [11017] 1 2 1 1 1 2 1 1 1 2 1 2 1 1 1 2 1 1 2 2 1 1 1 1 2 1 1 2 1 1 1 1 1
2 1 1
## [11053] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1
1 1 1
## [11089] 1 1 2 2 1 2 1 1 1 1 2 1 1 2 1 2 2 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 1
1 1 2
## [11125] 1 2 1 1 1 2 1 1 2 1 2 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 2 2 2 2 1 1
1 1 1
## [11161] 1 1 1 1 1 2 2 1 1 1 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 2 1 1 1
1 2 2
## [11197] 1 2 1 1 1 2 2 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2
2 1 1
## [11233] 1 1 1 1 1 1 1 1 1 2 2 2 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1
1 1 2
## [11269] 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1
2 2 2
## [11305] 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1
2 1 1
## [11341] 1 1 2 2 1 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 1 1
1 1 1
## [11377] 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1
1 1 1
## [11413] 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 2 1 2 1 1 1 2 1 1 1 2 1 1 1 2 2
2 1 1
## [11449] 1 1 1 2 2 1 1 2 1 1 1 1 1 2 1 1 1 2 1 1 2 1 2 1 2 1 1 1 1 1 1 2 2
```

```
1 1 1
## [11485] 1 1 1 2 1 1 1 2 1 2 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1
1 1 1
## [11521] 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2 2 1 1 1 1 2 1 1 1 1 1 1 1 1
1 2 2
## [11557] 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
1 2 2
## [11593] 2 1 1 1 1 1 1 1 2 1 1 1 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1
2 2 1
## [11629] 1 2 1 2 1 2 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 2
1 2 1
## [11665] 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 2 1 1
1 1 1
## [11701] 1 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 2 2 1 2 2 1 1 1 1 1 1 1 1 1 2 2 1
1 2 1
## [11737] 2 1 1 1 2 2 1 1 2 1 1 2 1 2 1 1 2 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1
1 1 1
## [11773] 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 2 1 1 2 2 2 2 1 1 1 1 1 1 1 1 1 2 1
1 2 1
## [11809] 1 1 1 1 1 2 1 1 2 2 1 2 1 1 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1
1 1 1
## [11845] 1 2 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1
1 1 1
## [11881] 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1
1 1 1
## [11917] 1 2 1 2 1 1 1 1 1 1 2 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1
2 1 2
## [11953] 1 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1
2 1 2
## [11989] 1 1 2 1 1 1 2 2 1 2 2 2 1 1 1 1 2 1 2 1 1 1 1 1 1 2 2 2 2 2 1 1 1
1 1 1
## [12025] 1 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1
1 1 1
## [12061] 1 2 1 1 2 1 2 1 1 2 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
## [12097] 2 2 1 2 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 2 1 2 1 1 1 1 1 2 1 1 1
1 1 1
## [12133] 2 1 1 1 2 1 1 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1
1 2 1
## [12169] 2 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 1 1
1 1 1
## [12205] 1 1 1 1 1 1 1
```

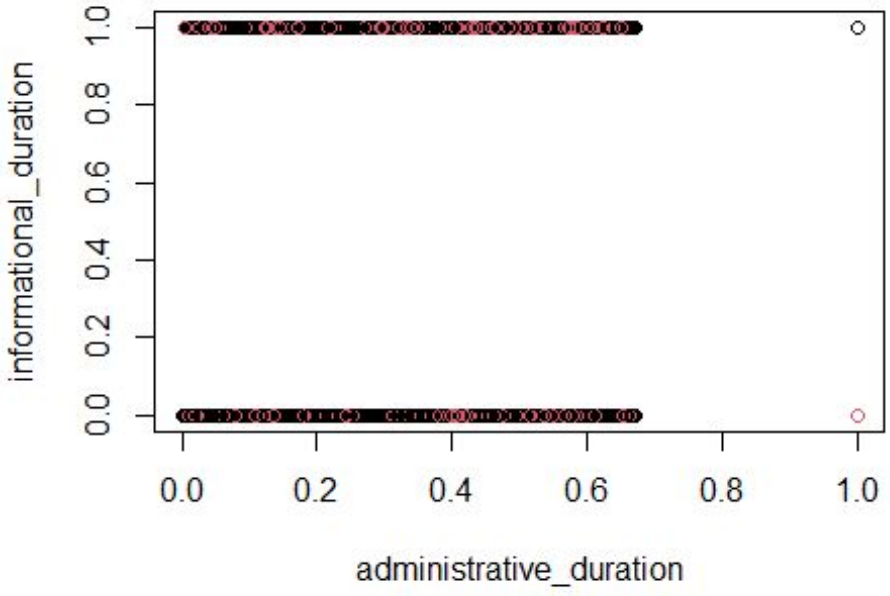
We can also **plot to see how various fields were distributed in the clusters** and also compare the plot with the original dataset

```
# plot to check administrative duration and informational duration in the
clusters
plot(ecomm_new[c(2,4)], col = kmeans_model$cluster)
```



Compare the plot with the initial distribution in the data

```
plot(ecomm_new[c(2,4)], col = ecommerce_label)
```



The distribution on the model is similar to the distribution in the dataset.

We can **check the distribution on a table**

```
table(kmeans_model$cluster, ecommerce_label)
```

```
##    ecommerce_label
##        1    2
##   1 8381 1302
##   2 1922  606
```

The result of the table shows that cluster 1 corresponds to the above table, cluster 1 corresponds to revenue = FALSE and cluster 2 corresponds to revenue = TRUE

We can see that the model did not perform so well since there are a couple number of misclassifications in the matrix. We can build a hierarchical model to observe if it will perform better than kmeans model.

10.3 Hierarchical modeling

Compute the Euclidean distance between observations using the dist function

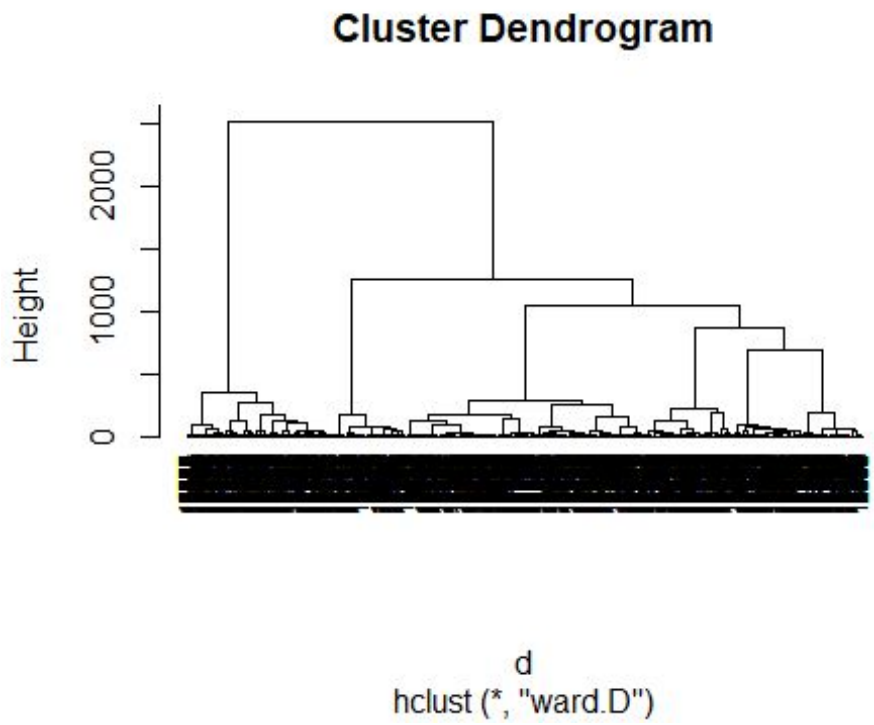
```
#compute euclidean distance
d <- dist(ecomm_new, method = "euclidean")
```

We then build a hierachical model with the distance function and method average

```
# build hierarchical model
hier <- hclust(d, method = "ward.D" )
```

Plotting the dendrogram

```
# plot the obtained dendrogram
plot(hier, cex = 0.6, hang = -1)
```



It looks like the first smaller cluster corresponds to revenue = FALSE, which was the cluster with a few values, while the second large cluster with many other clusters corresponds to revenue = TRUE

```
# install the package and load it: install.packages('ape')
#library(ape)
# plot basic tree
#plot(as.phylo(hier), cex = 0.9, label.offset = 1)
```

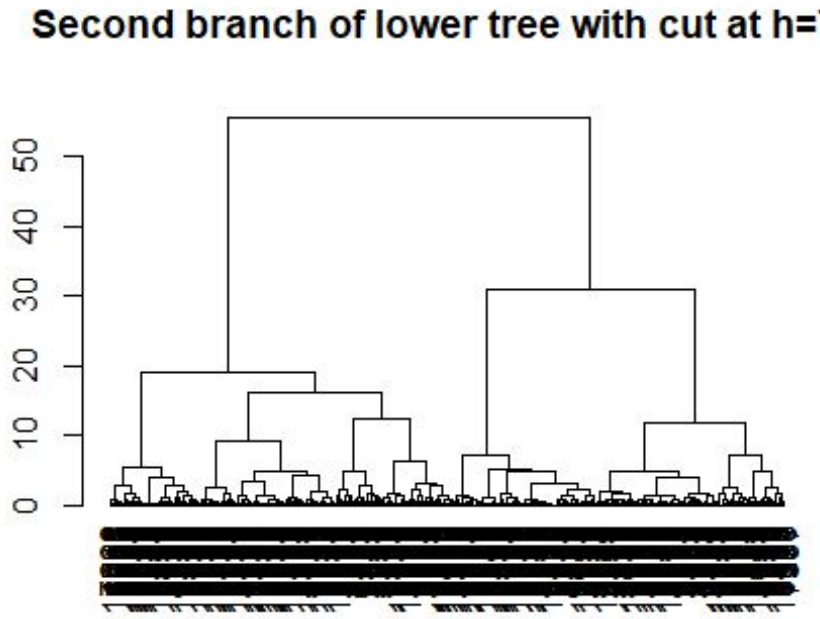
An alternative way to produce dendrograms is to specifically convert hclust objects into dendrograms objects. This makes it easier to truncate the dendrogram at specific points for easy interpretation

```
# convert the hierarchical clustering object to dendrogram object
hcd = as.dendrogram(hier)
```

Now we can truncate the original dendrogram for easy interpretation

```
plot(cut(hcd, h = 75)$upper, main = "Upper tree of cut at h=75")
```


```
plot(cut(hcd, h = 75)$lower[[2]], main = "Second branch of lower tree with
cut at h=75")
```



11. Challenging the solution

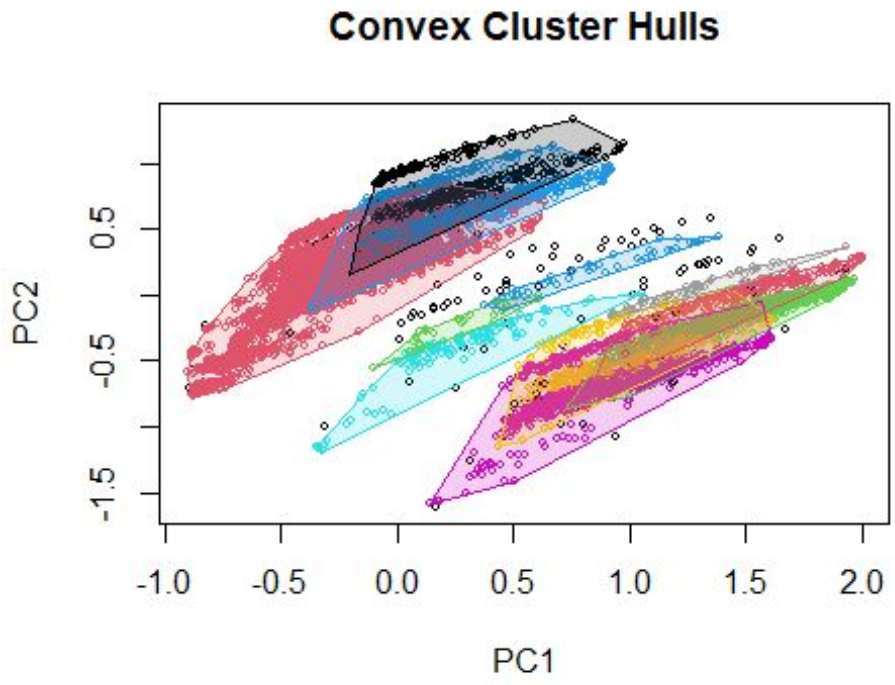
We can challenge the solution by using the DBSCAN Clustering method and checking to see ow well it performs.

```
library("dbscan")

db<-dbscan(ecomm_new,eps=1,MinPts = 14)

## Warning in dbscan(ecomm_new, eps = 1, MinPts = 14): converting argument
MinPts
## (fpc) to minPts (dbscan)!

print(db)

## DBSCAN clustering for 12211 objects.
## Parameters: eps = 1, minPts = 14
## The clustering contains 11 cluster(s) and 139 noise points.
##
##    0    1    2    3    4    5    6    7    8    9   10   11
##  139 7903  694 1235  114 1003  373   50  384  246   26   44
##
## Available fields: cluster, eps, minPts
```

After trying out different values for the parameters, the minimum number of clusters obtained is 11 clusters at minimum points=14 and eps=1.

We can make a **plot of the clusters**

```
hullplot(ecomm_new,db$cluster)

## Warning in hullplot(ecomm_new, db$cluster): Not enough colors. Some colors
will
## be reused.
```



11. Follow up questions

a). Did we have the right data?

Yes we had the right data to answer the research question at hand.

b). Do we need other data to answer our question?

It would be desirable to have more variables included that can help improve the predictive power of the model. This could be variables such as the types of products that lead to revenue earnings for the ecommerce company.

c). Did we have the right question?

The research question was specific, appropriate and in line with our needs and the data available.