

Part 2 Kira Plastinina Project

Margaret Gathoni

Introduction

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.

Problem Statement

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.

Metrics of Success

- Find and deal with outliers, anomalies, and missing data within the data set.
- Perform uni variate and bivariate analysis.
- Perform clustering stating insights drawn from your analysis and visualizations.
- Provide comparisons between different approaches i.e. K-Means clustering vs Hierarchical clustering highlighting the strengths and limitations of each approach in the context of your analysis.

Data Understanding

Will load the data from the following source <http://bit.ly/EcommerceCustomersDataset>

```
library(data.table)
## Warning: package 'data.table' was built under R version 4.1.2
data <- fread('http://bit.ly/EcommerceCustomersDataset')
```

Will review the first six rows

```
head(data)
##      Administrative Administrative_Duration Informational
##      Informational_Duration
## 1:              0              0              0
## 2:              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
```

Number of records

```
dim(data)
## [1] 12330 18
```

Attributes

```
str(data)
## Classes 'data.table' and '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" ...
## $ Weekend         : logi  FALSE FALSE FALSE FALSE TRUE FALSE ...
## $ Revenue          : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
## - attr(*, ".internal.selfref")=<externalptr>
```

Summary Statistics

```
summary(data)
```

```
## Administrative      Administrative_Duration Informational
## Min.   : 0.000      Min.   : -1.00      Min.   : 0.000
## 1st Qu.: 0.000      1st Qu.:  0.00      1st Qu.: 0.000
## Median : 1.000      Median :  8.00      Median : 0.000
## Mean   : 2.318      Mean   : 80.91      Mean   : 0.504
## 3rd Qu.: 4.000      3rd Qu.: 93.50      3rd Qu.: 0.000
## Max.   :27.000      Max.   :3398.75      Max.   :24.000
## NA's   :14          NA's   :14          NA's   :14
## Informational_Duration ProductRelated      ProductRelated_Duration
## Min.   : -1.00      Min.   :  0.00      Min.   : -1.0
## 1st Qu.:  0.00      1st Qu.:  7.00      1st Qu.: 185.0
## Median :  0.00      Median : 18.00      Median : 599.8
## Mean   : 34.51      Mean   : 31.76      Mean   : 1196.0
## 3rd Qu.:  0.00      3rd Qu.: 38.00      3rd Qu.: 1466.5
## Max.   :2549.38      Max.   :705.00      Max.   :63973.5
## NA's   :14          NA's   :14          NA's   :14
## BounceRates          ExitRates          PageValues          SpecialDay
## Min.   :0.000000      Min.   :0.00000      Min.   : 0.000      Min.   :0.00000
## 1st Qu.:0.000000      1st Qu.:0.01429      1st Qu.: 0.000      1st Qu.:0.00000
## Median :0.003119      Median :0.02512      Median : 0.000      Median :0.00000
## Mean   :0.022152      Mean   :0.04300      Mean   : 5.889      Mean   :0.06143
## 3rd Qu.:0.016684      3rd Qu.:0.05000      3rd Qu.: 0.000      3rd Qu.:0.00000
## Max.   :0.200000      Max.   :0.20000      Max.   :361.764      Max.   :1.00000
## NA's   :14          NA's   :14
## Month                OperatingSystems      Browser                Region
## Length:12330          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   : 2.357      Mean   :3.147
##                        3rd Qu.:3.000      3rd Qu.: 2.000      3rd Qu.:4.000
##                        Max.   :8.000      Max.   :13.000      Max.   :9.000
##
## TrafficType          VisitorType          Weekend              Revenue
## Min.   : 1.00      Length:12330      Mode :logical      Mode :logical
## 1st Qu.: 2.00      Class :character  FALSE:9462         FALSE:10422
```

```
## Median : 2.00   Mode  :character   TRUE :2868       TRUE :1908
## Mean    : 4.07
## 3rd Qu.: 4.00
## Max.    :20.00
##
```

Data Cleaning and Data Preparation

a. Checking for missing values

```
library(tidyverse)

## -- Attaching packages ----- tidyverse
1.3.2 --
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.7      v dplyr  1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.1.3
## Warning: package 'tibble' was built under R version 4.1.3
## Warning: package 'tidyr' was built under R version 4.1.3
## Warning: package 'readr' was built under R version 4.1.2
## Warning: package 'purrr' was built under R version 4.1.2
## Warning: package 'dplyr' was built under R version 4.1.3
## Warning: package 'stringr' was built under R version 4.1.1
## Warning: package 'forcats' was built under R version 4.1.2

## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::between() masks data.table::between()
## x dplyr::filter()  masks stats::filter()
## x dplyr::first()   masks data.table::first()
## x dplyr::lag()     masks stats::lag()
## x dplyr::last()    masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

cbind(lapply(lapply(data, is.na), sum))

##           [,1]
## Administrative 14
## Administrative_Duration 14
## Informational 14
## Informational_Duration 14
## ProductRelated 14
## ProductRelated_Duration 14
## BounceRates 14
```

```
## ExitRates          14
## PageValues         0
## SpecialDay         0
## Month              0
## OperatingSystems   0
## Browser            0
## Region             0
## TrafficType        0
## VisitorType        0
## Weekend            0
## Revenue            0
```

Out of the 12330 rows we only have 14 that have missing values meaning they won't have a great significance so will drop them

```
df <- data%>% drop_na()
df
```

```
##      Administrative Administrative_Duration Informational
##      1:              0                      0              0
##      2:              0                      0              0
##      3:              0                     -1              0
##      4:              0                      0              0
##      5:              0                      0              0
##      ---
## 12312:              3                     145              0
## 12313:              0                      0              0
## 12314:              0                      0              0
## 12315:              4                      75              0
## 12316:              0                      0              0
##      Informational_Duration ProductRelated ProductRelated_Duration
##      1:                    0                  1          0.000000
##      2:                    0                  2          64.000000
##      3:                   -1                  1          -1.000000
##      4:                    0                  2           2.666667
##      5:                    0                 10          627.500000
##      ---
## 12312:                    0                 53          1783.791667
## 12313:                    0                  5           465.750000
## 12314:                    0                  6           184.250000
## 12315:                    0                 15           346.000000
## 12316:                    0                  3           21.250000
##      BounceRates ExitRates PageValues SpecialDay Month OperatingSystems
##      1: 0.200000000 0.20000000 0.00000         0   Feb              1
##      2: 0.000000000 0.10000000 0.00000         0   Feb              2
##      3: 0.200000000 0.20000000 0.00000         0   Feb              4
##      4: 0.050000000 0.14000000 0.00000         0   Feb              3
##      5: 0.020000000 0.05000000 0.00000         0   Feb              3
##      ---
## 12312: 0.007142857 0.02903061 12.24172         0   Dec              4
```

```
## 12313: 0.000000000 0.02133333 0.00000 0 Nov 3
## 12314: 0.083333333 0.08666667 0.00000 0 Nov 3
## 12315: 0.000000000 0.02105263 0.00000 0 Nov 2
## 12316: 0.000000000 0.06666667 0.00000 0 Nov 3
##      Browser Region TrafficType VisitorType Weekend Revenue
##      1:      1      1      1 Returning_Visitor FALSE FALSE
##      2:      2      1      2 Returning_Visitor FALSE FALSE
##      3:      1      9      3 Returning_Visitor FALSE FALSE
##      4:      2      2      4 Returning_Visitor FALSE FALSE
##      5:      3      1      4 Returning_Visitor TRUE  FALSE
##      ---
## 12312:      6      1      1 Returning_Visitor TRUE  FALSE
## 12313:      2      1      8 Returning_Visitor TRUE  FALSE
## 12314:      2      1     13 Returning_Visitor TRUE  FALSE
## 12315:      2      3     11 Returning_Visitor FALSE FALSE
## 12316:      2      1      2      New_Visitor TRUE  FALSE
```

```
cbind(lapply(lapply(df, is.na), sum))
```

```
##      [,1]
## Administrative      0
## Administrative_Duration 0
## Informational      0
## Informational_Duration 0
## ProductRelated      0
## ProductRelated_Duration 0
## BounceRates      0
## ExitRates      0
## PageValues      0
## SpecialDay      0
## Month      0
## OperatingSystems      0
## Browser      0
## Region      0
## TrafficType      0
## VisitorType      0
## Weekend      0
## Revenue      0
```

we now have no null values.

b. Will check for duplicates

```
data_duplicated <- df[duplicated(df),]
data_duplicated
```

```
##      Administrative Administrative_Duration Informational
##      1:      0      0      0
##      2:      0      0      0
##      3:      0      0      0
##      4:      0      0      0
##      5:      0      0      0
```

```

## ---
## 113:          0          0          0
## 114:          0          0          0
## 115:          0          0          0
## 116:          0          0          0
## 117:          0          0          0
##      Informational_Duration ProductRelated ProductRelated_Duration
BounceRates
## 1:          0          1          0
0.2
## 2:          0          1          0
0.2
## 3:          0          1          0
0.2
## 4:          0          1          0
0.2
## 5:          0          1          0
0.2
## ---
## 113:          0          1          0
0.2
## 114:          0          1          0
0.2
## 115:          0          1          0
0.2
## 116:          0          1          0
0.2
## 117:          0          1          0
0.2
##      ExitRates PageValues SpecialDay Month OperatingSystems Browser Region
## 1:      0.2          0          0 Feb              1          1          1
## 2:      0.2          0          0 Feb              3          2          3
## 3:      0.2          0          0 Mar              1          1          1
## 4:      0.2          0          0 Mar              2          2          4
## 5:      0.2          0          0 Mar              3          2          3
## ---
## 113:      0.2          0          0 Dec              1          1          1
## 114:      0.2          0          0 Dec              1          1          4
## 115:      0.2          0          0 Dec              1          1          1
## 116:      0.2          0          0 Dec              1         13          9
## 117:      0.2          0          0 Dec              8         13          9
##      TrafficType      VisitorType Weekend Revenue
## 1:          3 Returning_Visitor  FALSE  FALSE
## 2:          3 Returning_Visitor  FALSE  FALSE
## 3:          1 Returning_Visitor   TRUE  FALSE
## 4:          1 Returning_Visitor  FALSE  FALSE
## 5:          1 Returning_Visitor  FALSE  FALSE
## ---
## 113:          2      New_Visitor  FALSE  FALSE
## 114:          1 Returning_Visitor   TRUE  FALSE

```

```
## 115:      3 Returning_Visitor  FALSE  FALSE
## 116:     20 Returning_Visitor  FALSE  FALSE
## 117:     20                Other  FALSE  FALSE
```

we have 119 rows duplicate rows, we are going to drop them

```
unique(df)
```

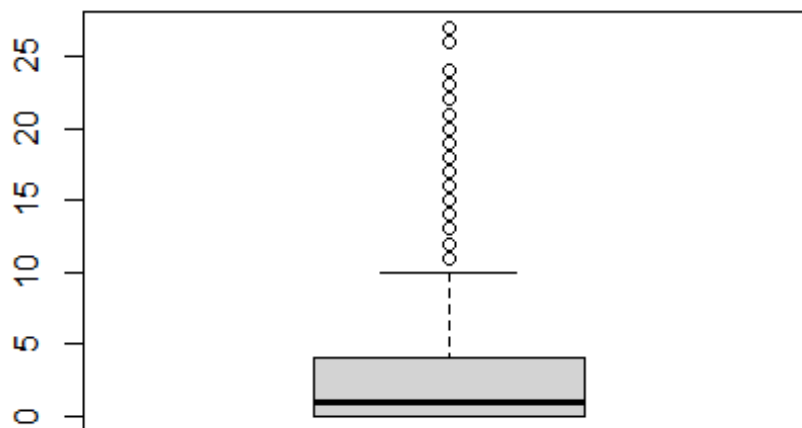
```
##      Administrative Administrative_Duration Informational
##      1:              0                      0            0
##      2:              0                      0            0
##      3:              0                     -1            0
##      4:              0                      0            0
##      5:              0                      0            0
##      ---
## 12195:              3                     145            0
## 12196:              0                      0            0
## 12197:              0                      0            0
## 12198:              4                      75            0
## 12199:              0                      0            0
##      Informational_Duration ProductRelated ProductRelated_Duration
##      1:                  0                1          0.000000
##      2:                  0                2          64.000000
##      3:                 -1                1          -1.000000
##      4:                  0                2           2.666667
##      5:                  0               10          627.500000
##      ---
## 12195:                  0                53          1783.791667
## 12196:                  0                 5           465.750000
## 12197:                  0                 6           184.250000
## 12198:                  0                15           346.000000
## 12199:                  0                 3            21.250000
##      BounceRates  ExitRates PageValues SpecialDay Month OperatingSystems
##      1: 0.200000000 0.20000000 0.00000 0 Feb 1
##      2: 0.000000000 0.10000000 0.00000 0 Feb 2
##      3: 0.200000000 0.20000000 0.00000 0 Feb 4
##      4: 0.050000000 0.14000000 0.00000 0 Feb 3
##      5: 0.020000000 0.05000000 0.00000 0 Feb 3
##      ---
## 12195: 0.007142857 0.02903061 12.24172 0 Dec 4
## 12196: 0.000000000 0.02133333 0.00000 0 Nov 3
## 12197: 0.083333333 0.08666667 0.00000 0 Nov 3
## 12198: 0.000000000 0.02105263 0.00000 0 Nov 2
## 12199: 0.000000000 0.06666667 0.00000 0 Nov 3
##      Browser Region TrafficType VisitorType Weekend Revenue
##      1:      1      1          1 Returning_Visitor  FALSE  FALSE
##      2:      2      1          2 Returning_Visitor  FALSE  FALSE
##      3:      1      9          3 Returning_Visitor  FALSE  FALSE
##      4:      2      2          4 Returning_Visitor  FALSE  FALSE
##      5:      3      1          4 Returning_Visitor   TRUE  FALSE
```



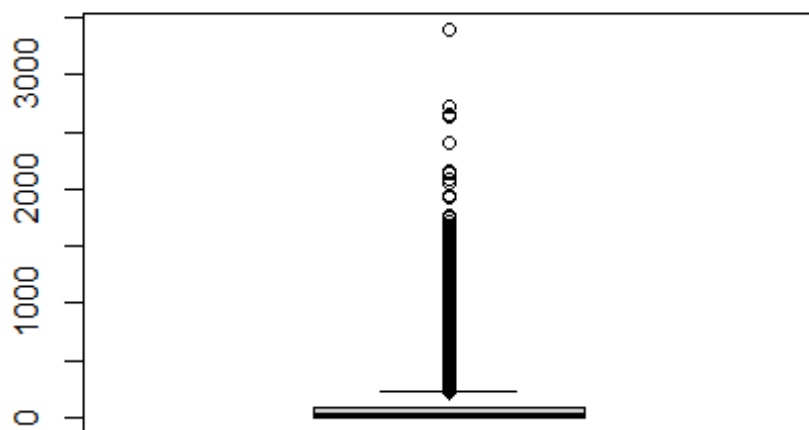
```
##      ---
## 12195:      6      1      1 Returning_Visitor      TRUE      FALSE
## 12196:      2      1      8 Returning_Visitor      TRUE      FALSE
## 12197:      2      1     13 Returning_Visitor      TRUE      FALSE
## 12198:      2      3     11 Returning_Visitor     FALSE      FALSE
## 12199:      2      1      2      New_Visitor      TRUE      FALSE
```

c. Will check for outliers

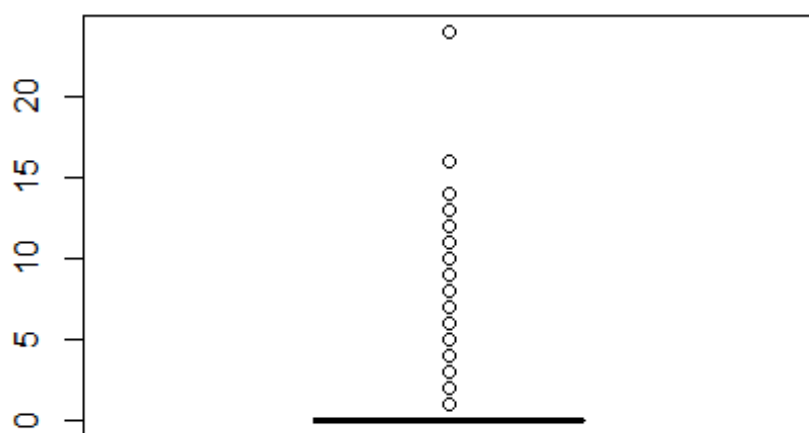
```
boxplot(df$Administrative)
```



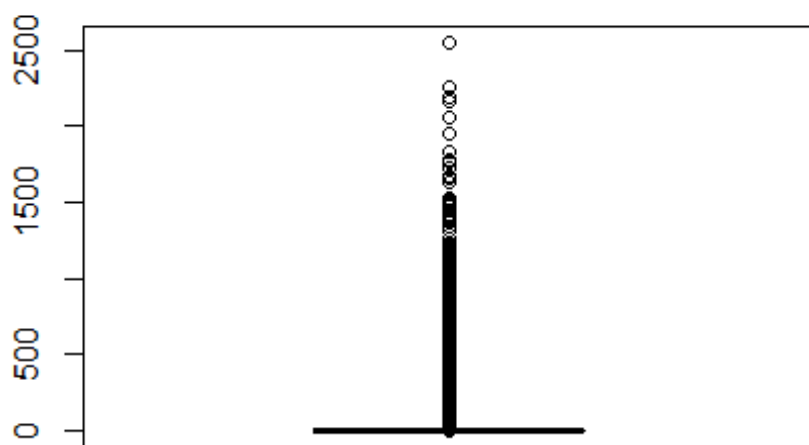
```
boxplot(df$Administrative_Duration)
```



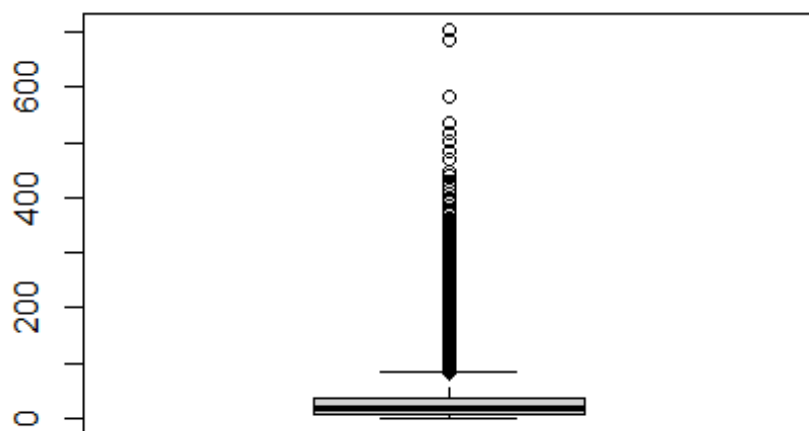
```
boxplot(df$Informational)
```



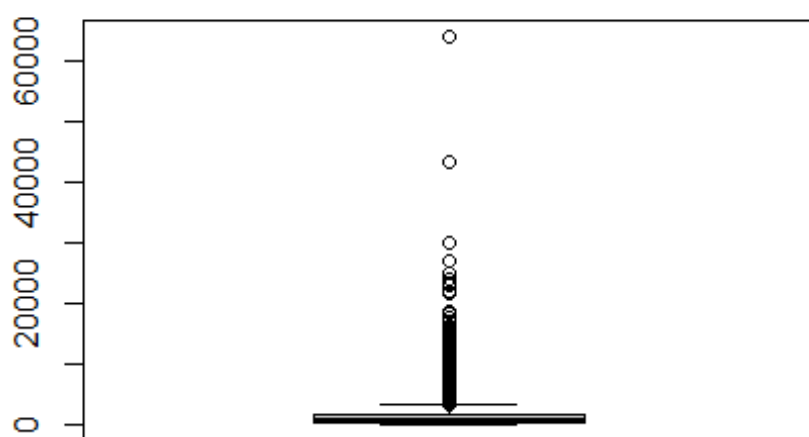
```
boxplot(df$Informational_Duration)
```



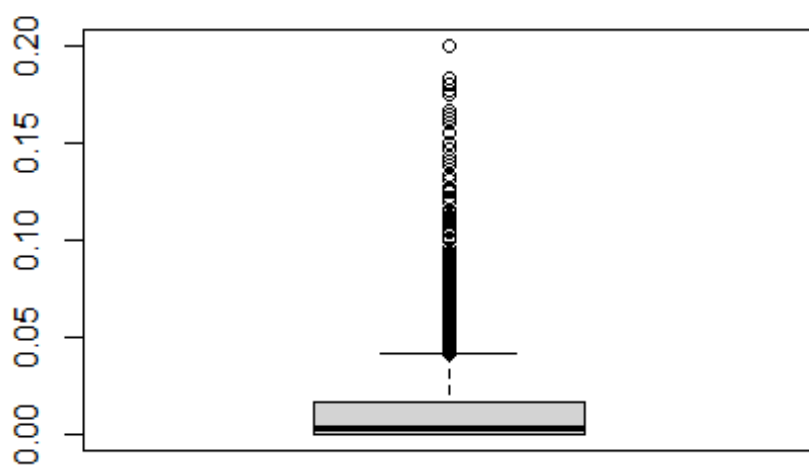
```
boxplot(df$ProductRelated)
```



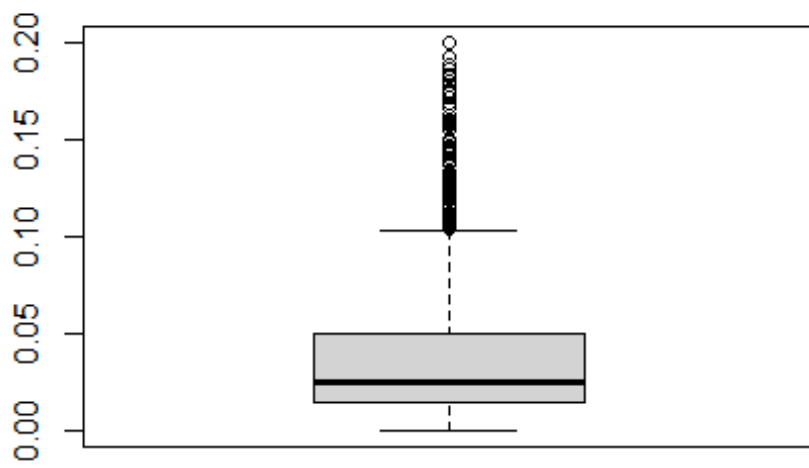
```
boxplot(df$ProductRelated_Duration)
```



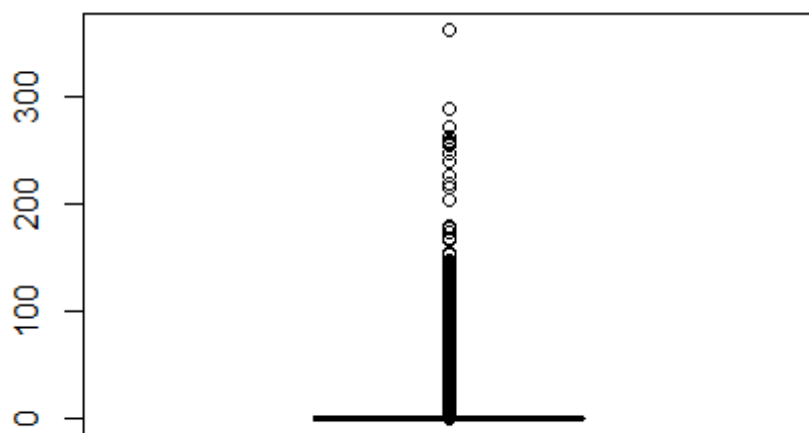
```
boxplot(df$BounceRates)
```



```
boxplot(df$ExitRates)
```



```
boxplot(df$PageValues)
```

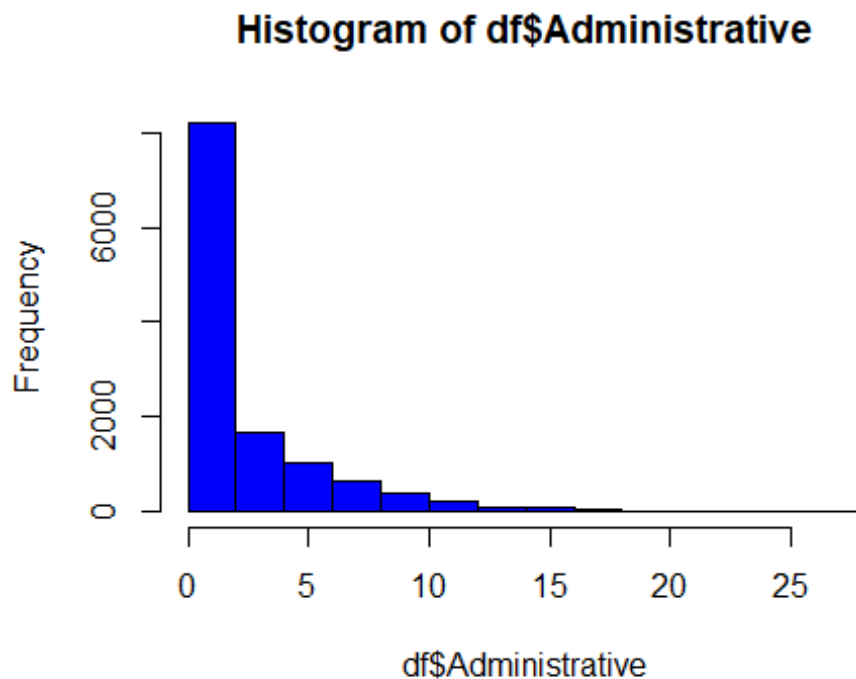


From the above box plot we have quite a number of outliers in the data set, however due to the details entailed in every attribute of the data these outliers are important for the project and will therefore not drop them.

EXPLORATORY DATA ANALYSIS

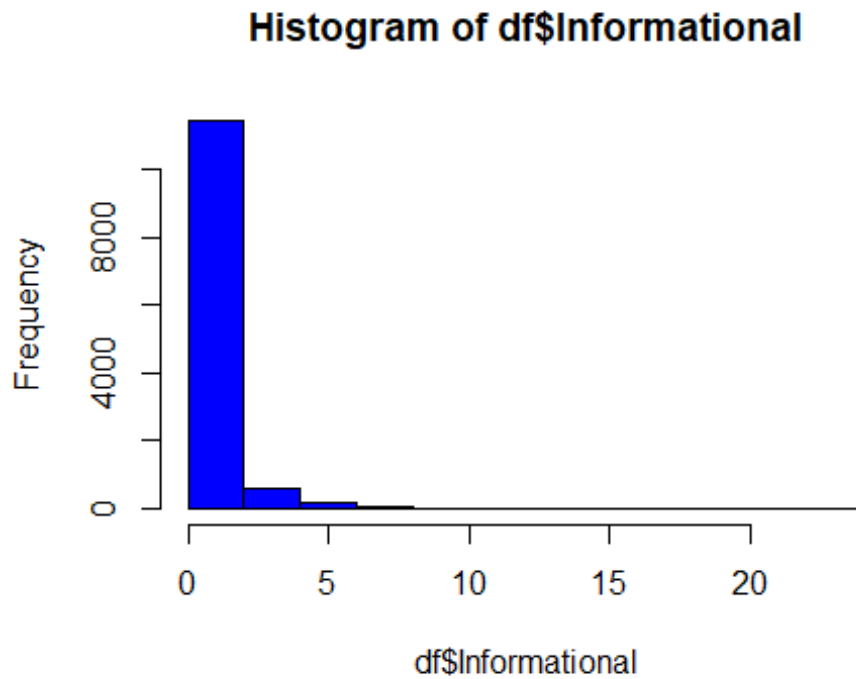
1. Univariate Analysis

```
hist(df$Administrative, col="blue")
```



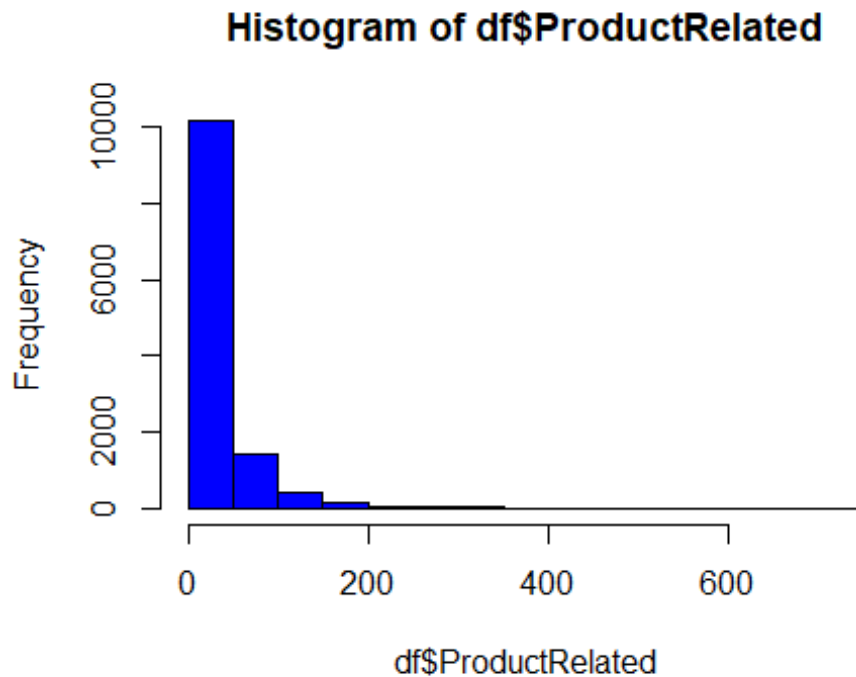
The histogram above shows the distribution of visitor in the administrative page and total time spent. We can see that most of the visitors in this page spent around 5 minutes with a sparse of the visitor spending up to 25 minutes

```
hist(df$Informational, col="blue")
```



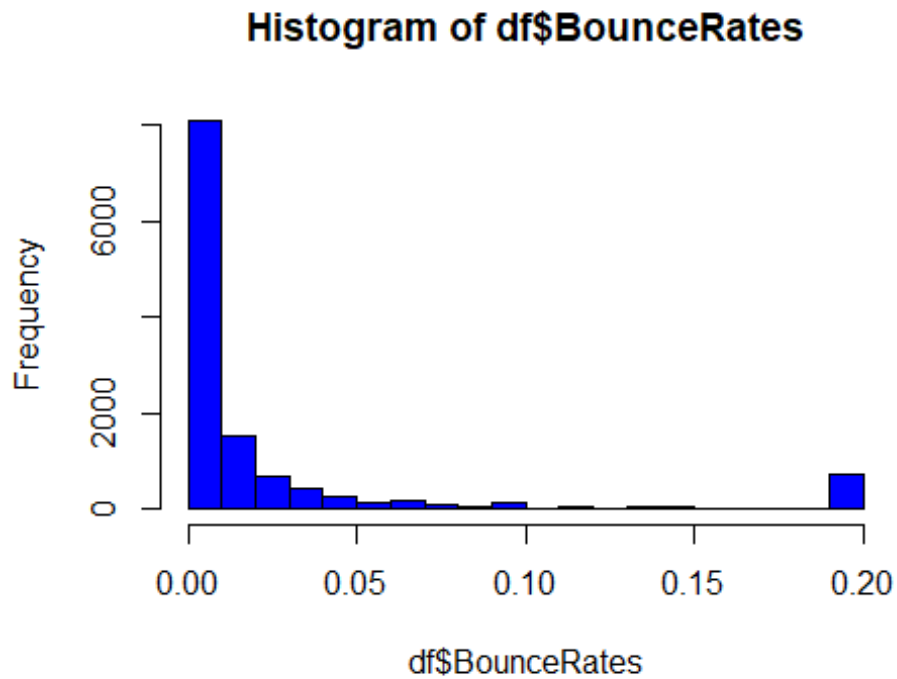
The histogram above shows the distribution of visitor in the Informational page and total time spent. We can see that most of the visitors in this page spent around 5 minutes with a sparse of the visitor spending up to 10 minutes

```
hist(df$ProductRelated, col="blue")
```



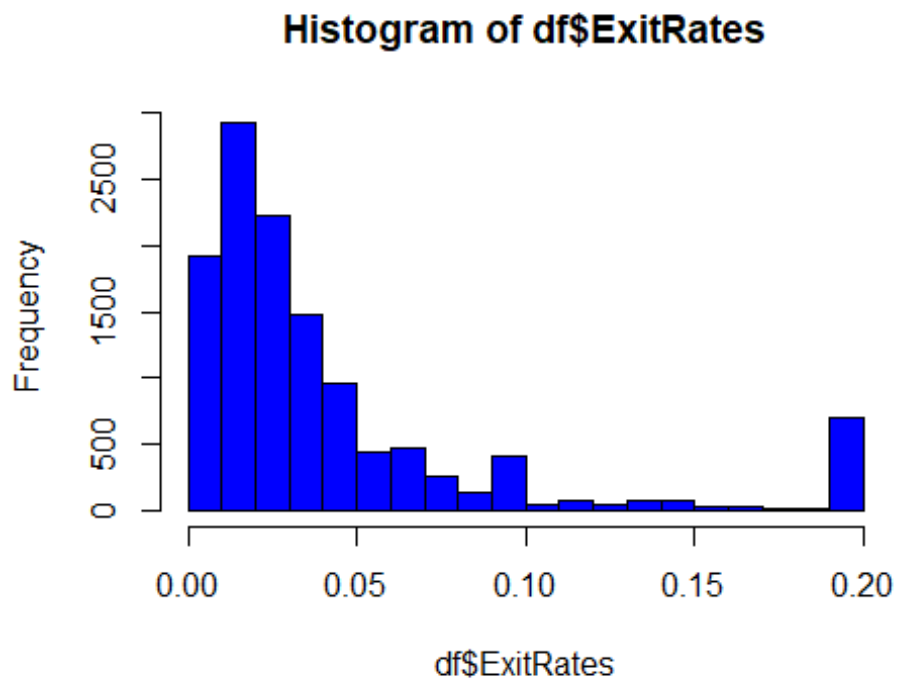
The histogram above shows the distribution of visitor in the Product Related page and total time spent. We can see that most of the visitors in this page spent around 200 minutes with a sparse of the visitor spending up to 400 minutes

```
hist(df$BounceRates, col="blue")
```

The histogram shows the percentage of visitors who enter the site from that page and then leave without triggering any other requests to the analytics server during that session

```
hist(df$ExitRates, col = "blue")
```

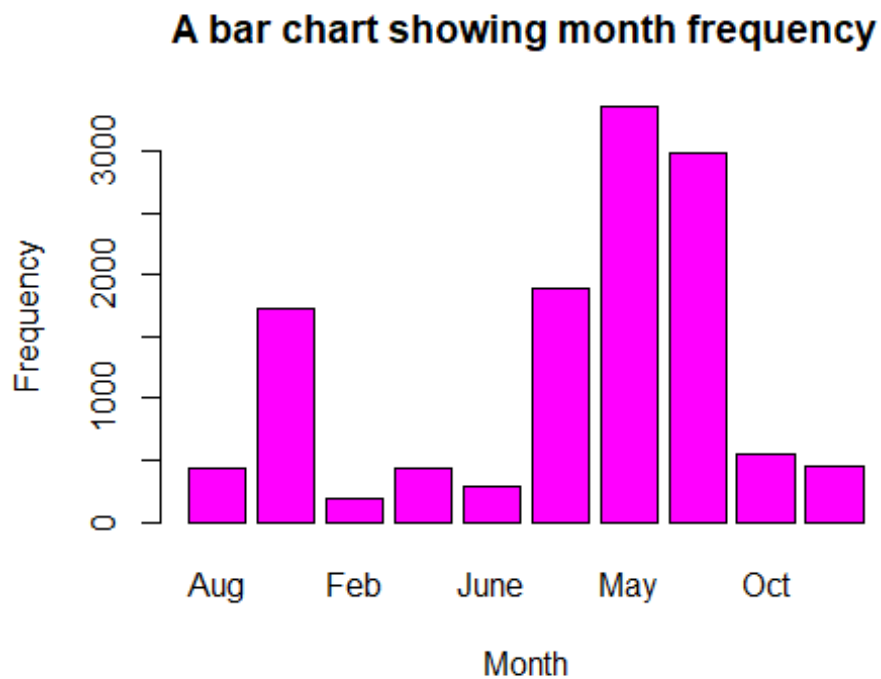


Let's Visualize visitor's distribution through out the year

```
month <- (df$Month)
month.frequency <- table(month)
month.frequency

## month
## Aug Dec Feb Jul June Mar May Nov Oct Sep
## 433 1727 184 432 288 1894 3363 2998 549 448

barplot(month.frequency,
  main="A bar chart showing month frequency",
  xlab="Month",
  ylab = "Frequency",
  col=c("magenta")
)
```



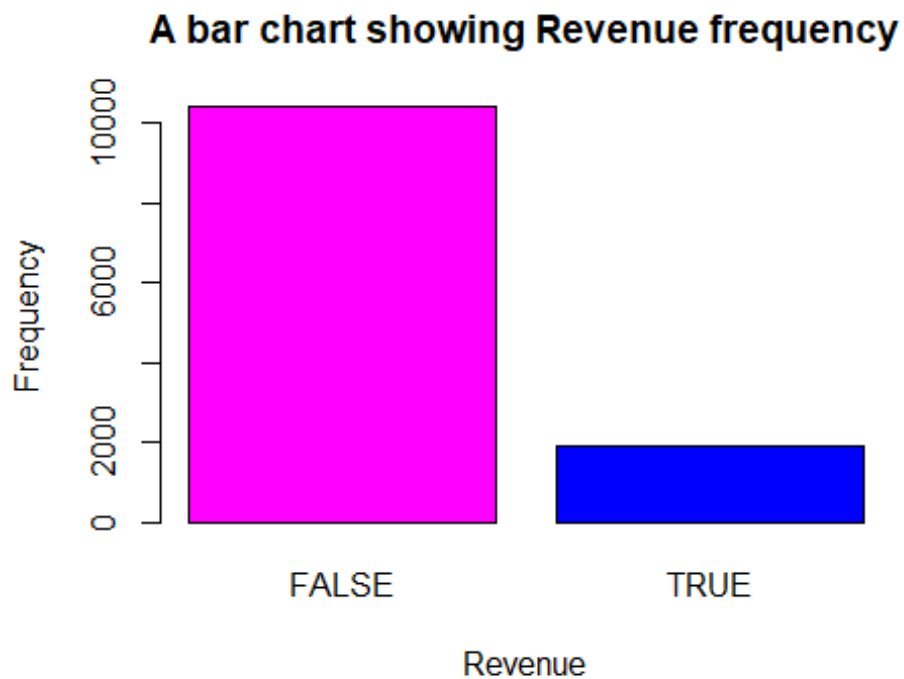
We can see the month with most visitors was may and the least with February.

Revenue Frequency

```
revenue <- (df$Revenue)
revenue.frequency <- table(revenue)
revenue.frequency

## revenue
## FALSE TRUE
## 10408 1908

barplot(revenue.frequency,
  main="A bar chart showing Revenue frequency",
  xlab="Revenue",
  ylab = "Frequency",
  col=c("magenta", "blue")
)
```



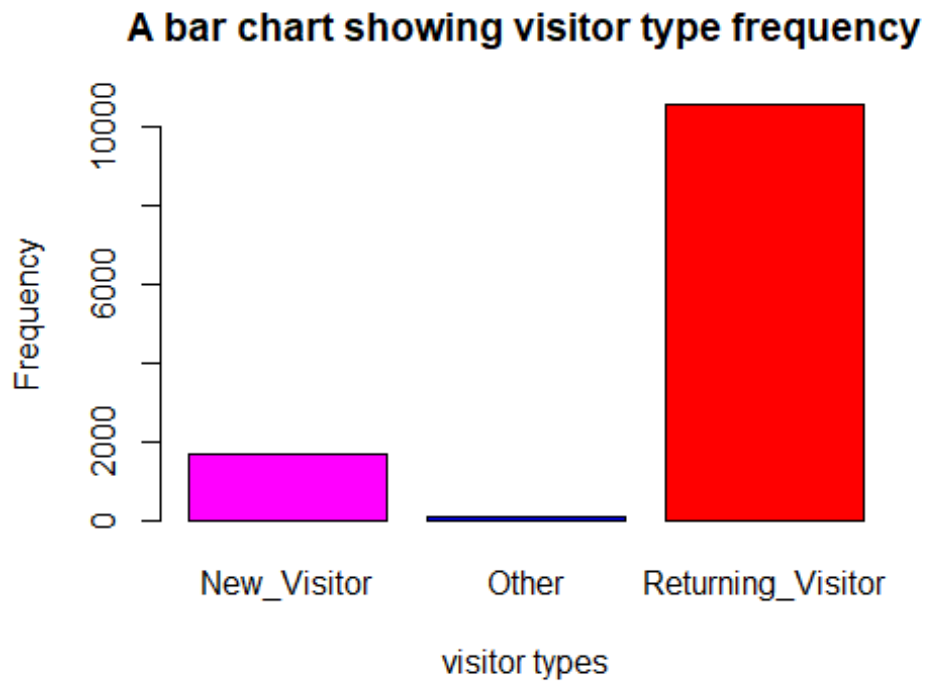
We can see that most of the visit were not revenue generating compared to those generating revenue

Visitor type distribution

```
visitor <- (df$VisitorType)
visitor.frequency <- table(visitor)
visitor.frequency

## visitor
##      New_Visitor      Other Returning_Visitor
##           1694             85           10537

barplot(visitor.frequency,
  main="A bar chart showing visitor type frequency",
  xlab="visitor types",
  ylab = "Frequency",
  col=c("magenta", "blue", "red")
)
```



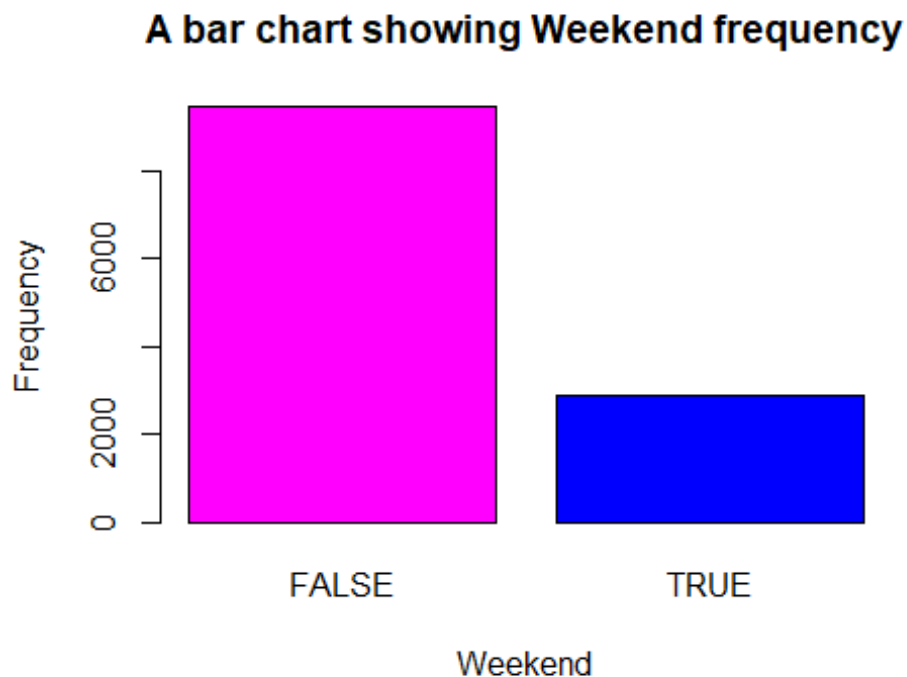
Returning visitors were more compared to new visitors and other being the least.

Week day Distribution

```
weekend <- (df$Weekend)
weekend.frequency <- table(weekend)
weekend.frequency

## weekend
## FALSE  TRUE
##  9451  2865

barplot(weekend.frequency,
  main="A bar chart showing Weekend frequency",
  xlab="Weekend",
  ylab = "Frequency",
  col=c("magenta", "blue")
)
```



Visits on the weekend were less compared to other days.

Region visitors distribution

```
library(janitor)

## Warning: package 'janitor' was built under R version 4.1.3

##
## Attaching package: 'janitor'

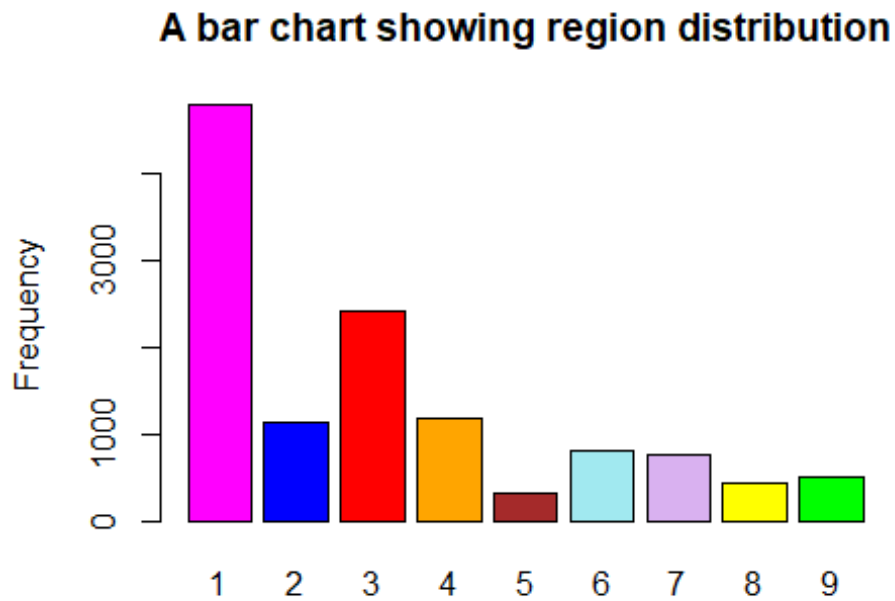
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test

region <- (df$Region)
region.frequency <- table(region)
region.frequency

## region
##   1    2    3    4    5    6    7    8    9
## 4774 1134 2402 1179  317  804  761  434  511

barplot(region.frequency,
  main="A bar chart showing region distribution",
  xlab="Region(1 = Russian, 2 = Ukraine, 3 = Kazakhstan, 4 = Belarus, 5 =
China, 6 = Phillipines, 7 = Armenia, 8 = Rest os asia, 9 = Rest of globe)",
  ylab = "Frequency",
```

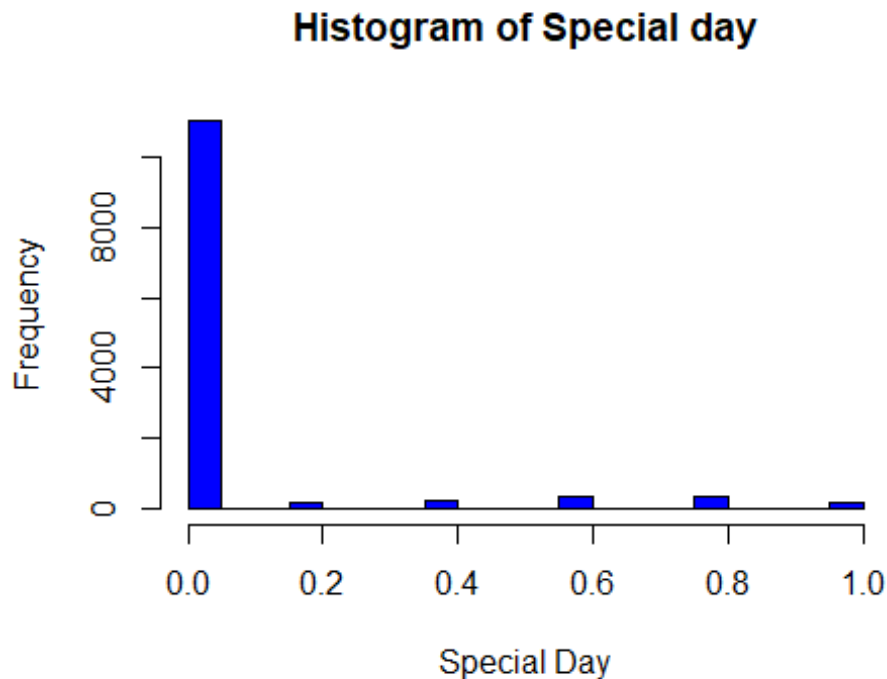
```
col=c("magenta", "blue", "red", "orange", "brown", "#a1e9f0", "#d9b1f0",
"yellow", "green")
)
```



e, 3 = Kazakhstan, 4 = Belarus, 5 = China, 6 = Phillipines, 7 = Armenia

Clearly shows most of the visitors come from Russia and the least from China.

```
hist((df$SpecialDay),
main = "Histogram of Special day",
xlab = 'Special Day',
ylab = 'Frequency',
col = "blue")
```



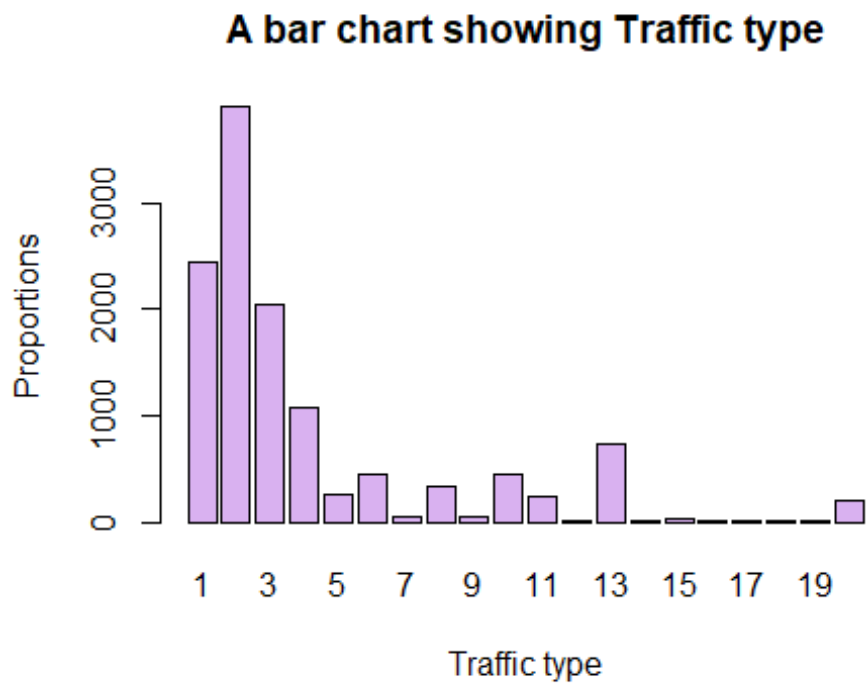
The above graph shows the closeness of the site visiting time to a specific special day (e.g. Mother's Day, Valentine's Day) in which the sessions are more likely to be finalized with the transaction. when close to a special day this value takes a nonzero value meaning most of the transaction happened when there was no special day close since the highest count is around zero.

Traffic Type Proportions

```
traffic <- (df$TrafficType)
traffic.frequency <- table(traffic)
traffic.frequency

## traffic
##      1      2      3      4      5      6      7      8      9     10     11     12     13     14     15
## 2444 3909 2051 1069  260  444   40  343   42  450  247    1   737   13   37
##
##      17     18     19     20
##      1     10     17    198

barplot(traffic.frequency,
  main="A bar chart showing Traffic type",
  xlab="Traffic type",
  ylab = "Proportions",
  col=c("#d9b1f0")
)
```

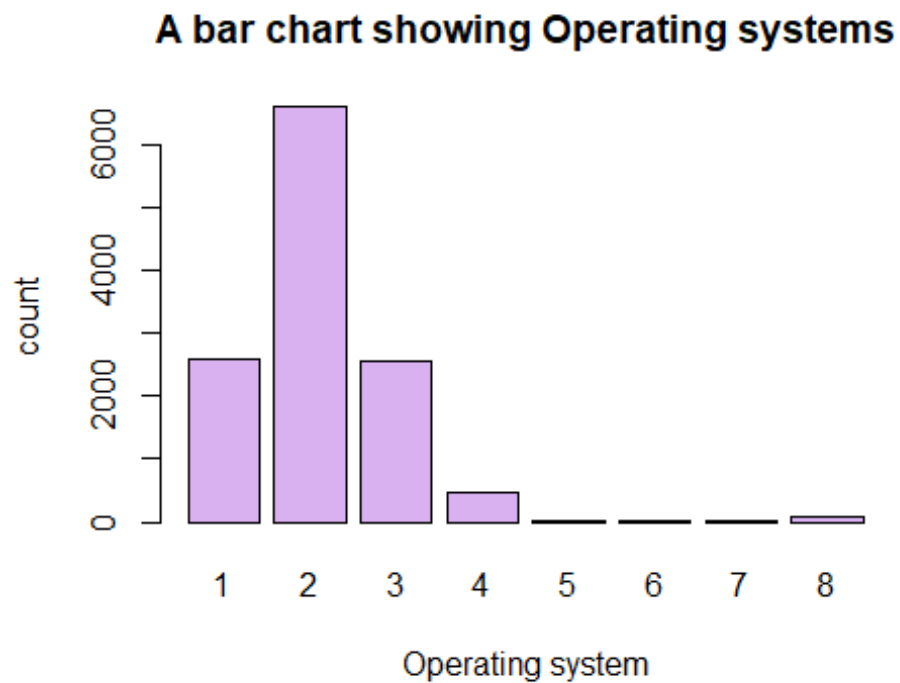
Traffic type 2 had the highest proportion.

Operating Systems Proportion

```
os <- (df$OperatingSystems)
os.frequency <- table(os)
os.frequency

## os
##   1   2   3   4   5   6   7   8
## 2582 6593 2552 478   6  19   7  79

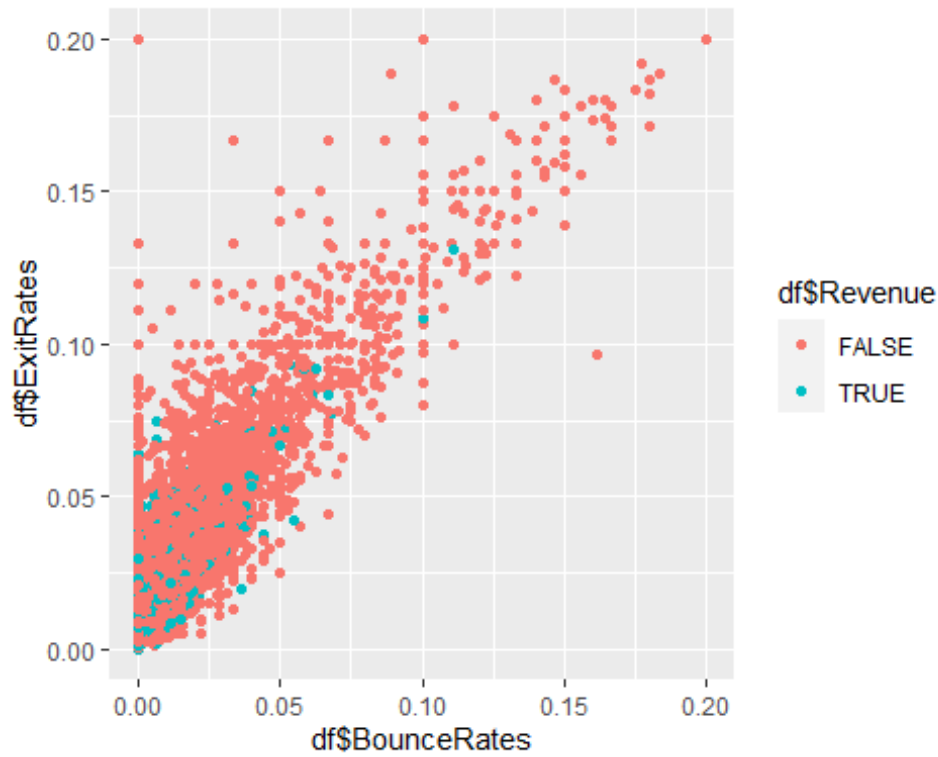
barplot(os.frequency,
  main="A bar chart showing Operating systems",
  xlab="Operating system",
  ylab = "count",
  col=c("#d9b1f0")
)
```



2. Bivariate Analysis

Bounce rate Vs Exit Rates in respect to Revenue

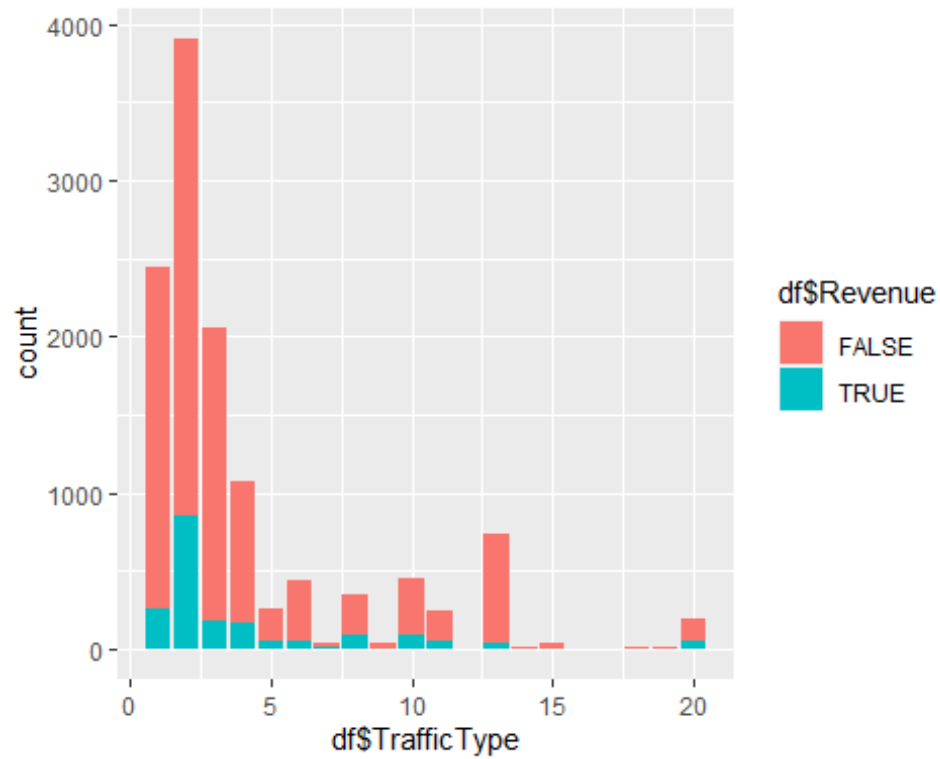
```
ggplot(df, aes(x=df$BounceRates, y=df$ExitRates, col=df$Revenue))+geom_point(aes(
  color=df$Revenue))
```



The graph shows revenue generation in respect to Bounce and Exit rates. There is revenue generation when the rates are low but this decreases with increase in the rates.

Traffic Types vs Revenue

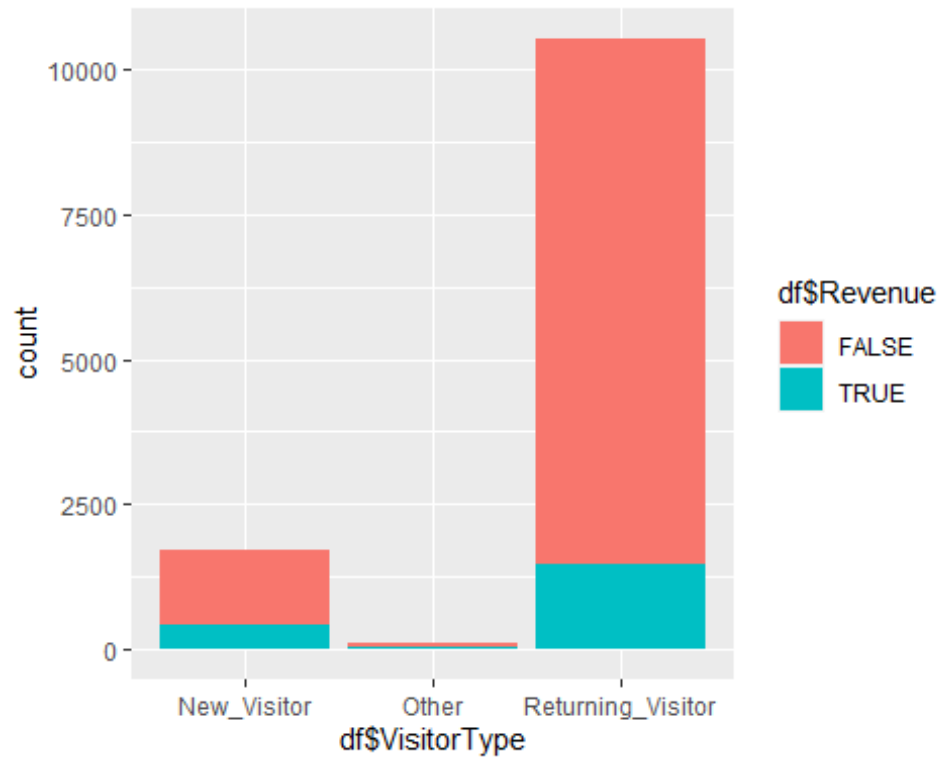
```
ggplot(df, aes(x = df$TrafficType, fill = df$Revenue)) +  
  geom_bar()
```



The graph above shows most of the traffic wasn't generating revenue.

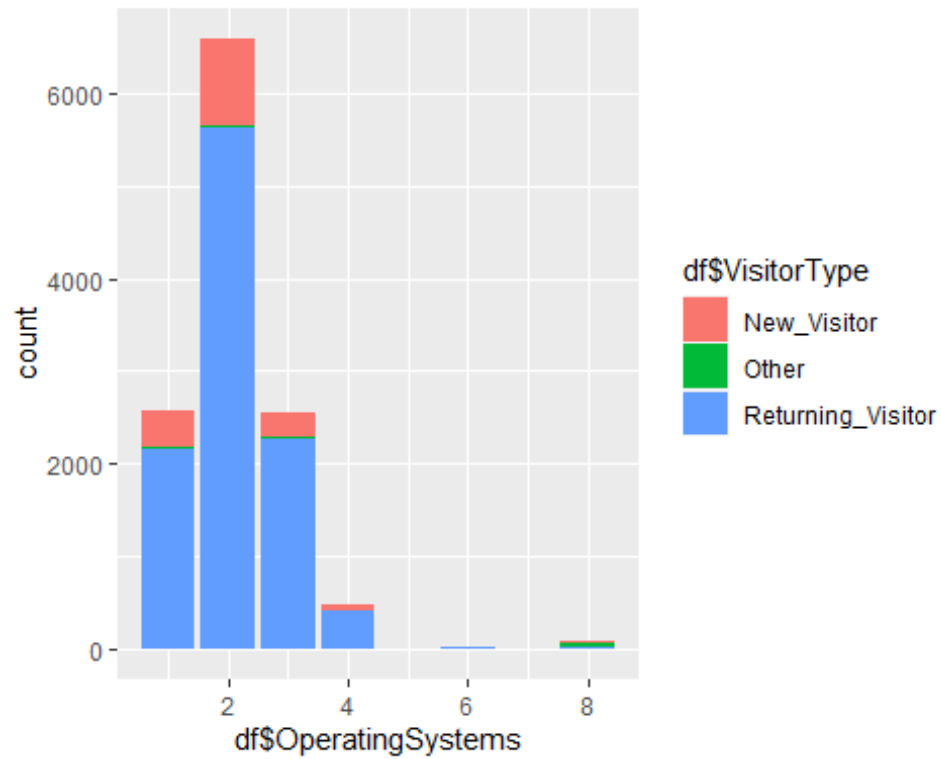
Visitor Type vs Revenue

```
ggplot(df, aes(x = df$VisitorType, fill = df$Revenue)) +  
  geom_bar()
```



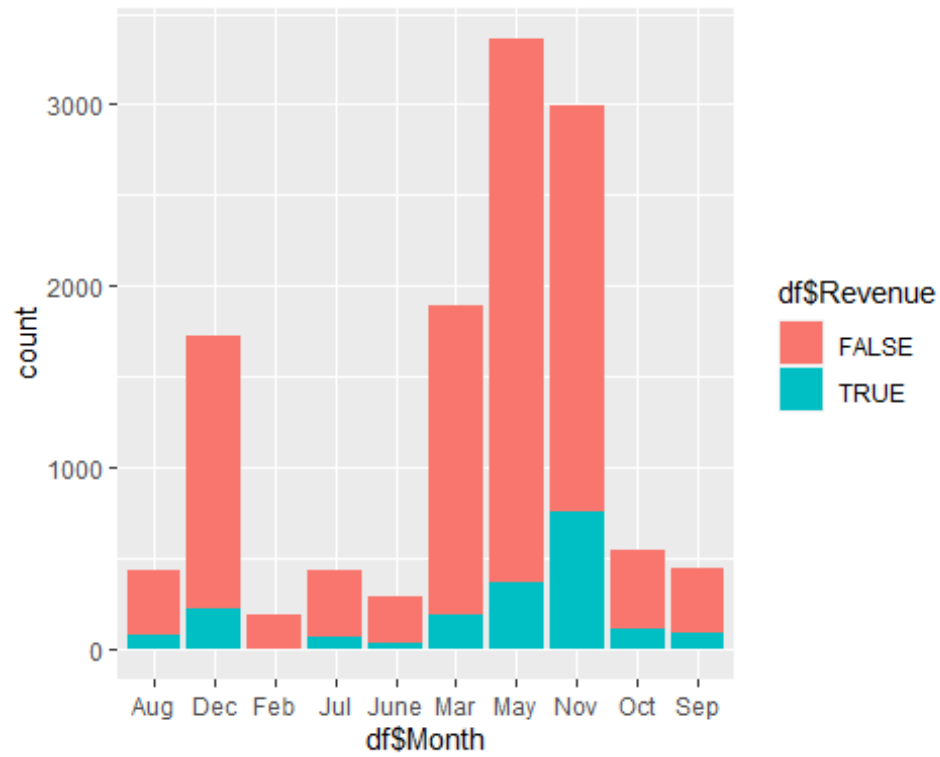
The number of visitor generating revenue were more on returning visitor compared to new visitors and other visitor types.

```
ggplot(df, aes(x = df$OperatingSystems, fill = df$VisitorType)) +  
  geom_bar()
```

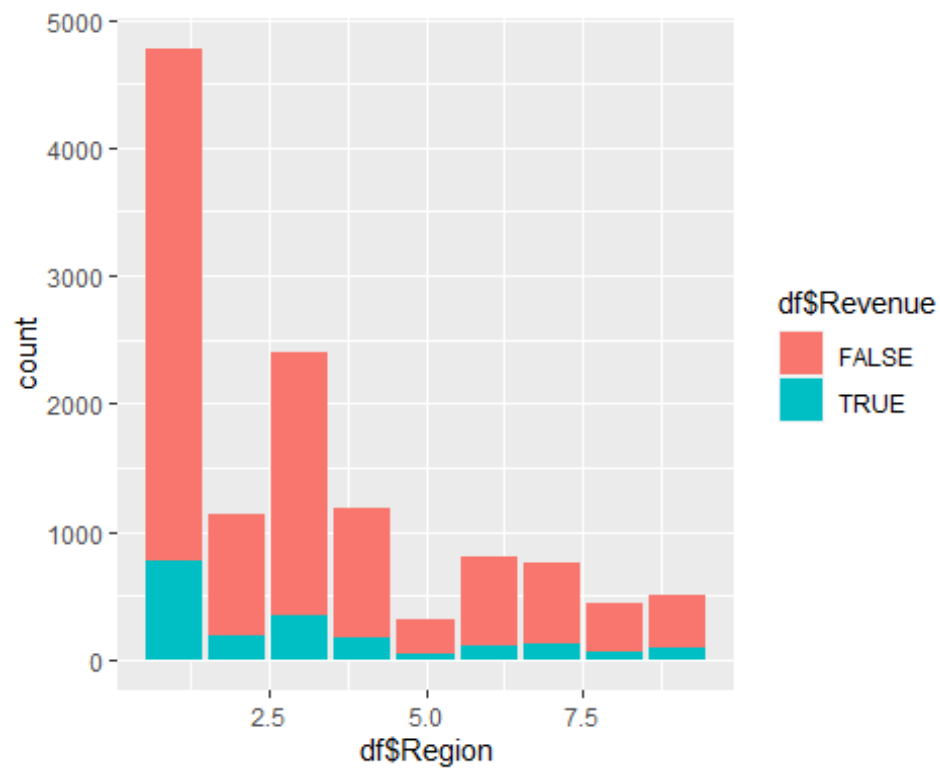


Most of the visitors were using operating system 2 with an outlier of other visitor using operating system 8.

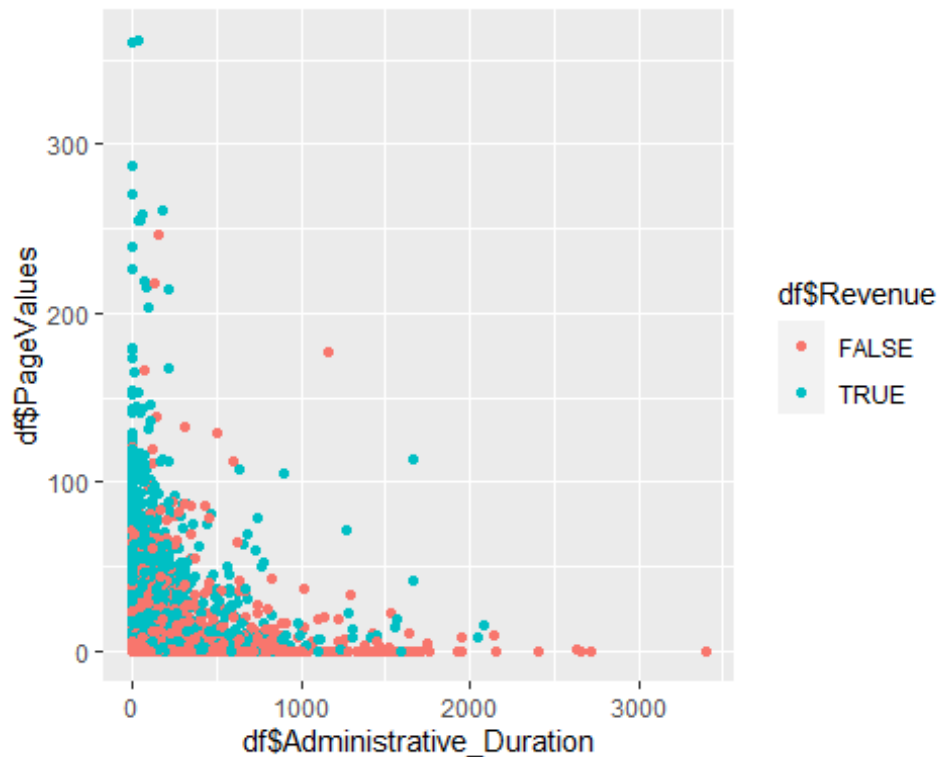
```
ggplot(df, aes(x = df$Month, fill = df$Revenue)) +  
  geom_bar()
```



```
ggplot(df, aes(x = df$Region, fill = df$Revenue)) +  
  geom_bar()
```



```
ggplot(df,aes(x=df$Administrative_Duration,y=df$PageValues,col=df$Revenue))+geom_point(aes(color=df$Revenue))
```



The higher the page value the less the administrative duration

3. Multivariate Analysis

```
head(df)
```

	Administrative_Duration	Informational_Duration	ProductRelated	ProductRelated_Duration	BounceRates	ExitRates	PageValues
## 1:	0	0	1	0.000000	0.20000000	0.20000000	0
## 2:	0	0	2	64.000000	0.00000000	0.10000000	0
## 3:	0	-1	1	-1.000000	0.20000000	0.20000000	0
## 4:	0	0	2	2.666667	0.05000000	0.14000000	0


```
## 5:          10          627.500000 0.02000000 0.05000000      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
```

Will check for correlation among numerical attributes

```
Monthly_statistics <- df %>% select(Month, Administrative_Duration,
                                   Informational_Duration,
                                   ProductRelated_Duration, PageValues,
                                   ExitRates, BounceRates) %>%
  group_by(Month)%>%summarise_all(mean)
Monthly_statistics

## # A tibble: 10 x 7
##   Month Administrative_Duration Informational~1 Produ~2 PageV~3 ExitR~4
##   Bounc~5
##   <chr>          <dbl>          <dbl>    <dbl>    <dbl>    <dbl>
##   <dbl>
## 1 Aug          107.          35.5    1273.    5.94    0.0377
##   0.0182
## 2 Dec           78.6          38.1    1111.    6.83    0.0413
##   0.0201
## 3 Feb           16.8           2.32     471.    0.890    0.0741
##   0.0470
## 4 Jul           78.9          45.5    1218.    4.10    0.0453
##   0.0247
## 5 June          59.1          20.5    1213.    3.39    0.0582
##   0.0351
## 6 Mar           71.7          30.9     817.    3.99    0.0442
##   0.0215
## 7 May           69.5          27.2     982.    5.43    0.0488
##   0.0268
## 8 Nov           90.9          43.6    1758.    7.13    0.0382
##   0.0193
## 9 Oct          126.          38.7    1117.    8.65    0.0290
##   0.0118
## 10 Sep         109.          35.7    1253.    7.56    0.0303
```

0.0122

```
## # ... with abbreviated variable names 1: Informational_Duration,  
## # 2: ProductRelated_Duration, 3: PageValues, 4: ExitRates, 5:  
BounceRates
```

- Month with highest administrative duration was October while February had the least.
- Month with highest Informational duration was July while February had the least.
- Month with highest Product related duration was November while February had the least.
- Month with highest Page value was October while February had the least.
- Month with highest Exit rates was February while October had the least.
- Month with highest Bounce rates was February while October had the least.

```
Region_statistics <- df %>% select(Region, Administrative_Duration,  
                                Informational_Duration,  
                                ProductRelated_Duration, PageValues,  
                                ExitRates, BounceRates) %>%  
  group_by(Region)%>%summarise_all(mean)  
Region_statistics
```

```
## # A tibble: 9 x 7
```

```
##   Region Administrative_Duration Informational~1 Produ~2 PageV~3 ExitR~4  
Bounc~5
```

```
##   <int>          <dbl>          <dbl>    <dbl>    <dbl>    <dbl>  
<dbl>
```

```
## 1      1      79.7      37.3  1278.    5.91  0.0431  
0.0221
```

```
## 2      2      87.0      36.1  1184.    5.92  0.0430  
0.0229
```

```
## 3      3      83.1      35.2  1189.    5.35  0.0438  
0.0224
```

```
## 4      4      83.3      36.4  1116.    5.82  0.0439  
0.0231
```

```
## 5      5      88.5      31.4  1151.    9.28  0.0402  
0.0201
```

```
## 6      6      68.7      29.9  1036.    4.81  0.0440  
0.0237
```

```
## 7      7      78.8      26.6  1142.    6.26  0.0398  
0.0204
```

```
## 8      8      89.6      32.8  1096.    4.26  0.0400  
0.0192
```

```
## 9      9      73.0      19.4  1121.    8.94  0.0438  
0.0217
```

```
## # ... with abbreviated variable names 1: Informational_Duration,  
## # 2: ProductRelated_Duration, 3: PageValues, 4: ExitRates, 5:  
BounceRates
```

Modeling

```
head(df)

##      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
```

K-MEAN Clustering

Will start by labeling our categorical attribute from categorical to numerical labels and also remove the class label (Revenue)

```
df1 <- df[, 1:17]

# Change the 'weekend' column's data type to 'factor'
df1$Weekend <- as.factor(df$Weekend)
```

```

library(caret)

## Warning: package 'caret' was built under R version 4.1.3

## Loading required package: lattice

## Warning: package 'lattice' was built under R version 4.1.3

##
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':
##
## lift

dummy <- dummyVars("~ .", "Month + OperatingSystems + Browser + Region +
                      TrafficType + VisitorType + Weekend", data=df1)
encoded <- data.frame(predict(dummy, newdata = df1))
df1 <- cbind(df1[, 1:10], encoded)

# Normalize the values
normal <-function(x) { (x -min(x))/(max(x)-min(x))}
df_norm <- as.data.frame(lapply(df1, normal))

# Preview the top six records
head(df_norm)

## Administrative Administrative_Duration Informational
Informational_Duration
## 1 0 0.0002941393 0
0.0003920992
## 2 0 0.0002941393 0
0.0003920992
## 3 0 0.0000000000 0
0.0000000000
## 4 0 0.0002941393 0
0.0003920992
## 5 0 0.0002941393 0
0.0003920992
## 6 0 0.0002941393 0
0.0003920992
## ProductRelated ProductRelated_Duration BounceRates ExitRates PageValues
## 1 0.001418440 1.563122e-05 1.00000000 1.000000 0
## 2 0.002836879 1.016029e-03 0.00000000 0.500000 0
## 3 0.001418440 0.000000e+00 1.00000000 1.000000 0
## 4 0.002836879 5.731448e-05 0.25000000 0.700000 0
## 5 0.014184397 9.824223e-03 0.10000000 0.250000 0
## 6 0.026950355 2.426226e-03 0.07894737 0.122807 0
## SpecialDay Administrative.1 Administrative_Duration.1 Informational.1
## 1 0 0 0.0002941393 0

```

```

## 2          0          0          0.0002941393          0
## 3          0          0          0.0000000000          0
## 4          0          0          0.0002941393          0
## 5          0          0          0.0002941393          0
## 6          0          0          0.0002941393          0
## Informational_Duration.1 ProductRelated.1 ProductRelated_Duration.1
## 1          0.0003920992          0.001418440          1.563122e-05
## 2          0.0003920992          0.002836879          1.016029e-03
## 3          0.0000000000          0.001418440          0.000000e+00
## 4          0.0003920992          0.002836879          5.731448e-05
## 5          0.0003920992          0.014184397          9.824223e-03
## 6          0.0003920992          0.026950355          2.426226e-03
## BounceRates.1 ExitRates.1 PageValues.1 SpecialDay.1 MonthAug MonthDec
## 1          1.00000000          1.000000          0          0          0          0
## 2          0.00000000          0.500000          0          0          0          0
## 3          1.00000000          1.000000          0          0          0          0
## 4          0.25000000          0.700000          0          0          0          0
## 5          0.10000000          0.250000          0          0          0          0
## 6          0.07894737          0.122807          0          0          0          0
## MonthFeb MonthJul MonthJune MonthMar MonthMay MonthNov MonthOct MonthSep
## 1          1          0          0          0          0          0          0          0
## 2          1          0          0          0          0          0          0          0
## 3          1          0          0          0          0          0          0          0
## 4          1          0          0          0          0          0          0          0
## 5          1          0          0          0          0          0          0          0
## 6          1          0          0          0          0          0          0          0
## OperatingSystems Browser Region TrafficType VisitorTypeNew_Visitor
## 1          0.00000000 0.00000000 0.000 0.00000000          0
## 2          0.1428571 0.08333333 0.000 0.05263158          0
## 3          0.4285714 0.00000000 1.000 0.10526316          0
## 4          0.2857143 0.08333333 0.125 0.15789474          0
## 5          0.2857143 0.16666667 0.000 0.15789474          0
## 6          0.1428571 0.08333333 0.000 0.10526316          0
## VisitorTypeOther VisitorTypeReturning_Visitor
## 1          0          1
## 2          0          1
## 3          0          1
## 4          0          1
## 5          0          1
## 6          0          1
##
WeekendMonth...OperatingSystems...Browser...Region.....Traf
ficType...VisitorType...WeekendFALSE
## 1
1
## 2
1
## 3
1
## 4
1

```

```

1
## 5
0
## 6
1
##
WeekendMonth...OperatingSystems...Browser...Region.....Traf
ficType...VisitorType...WeekendTRUE
## 1
0
## 2
0
## 3
0
## 4
0
## 5
1
## 6
0

library(dplyr)

```

Applying the K-means clustering algorithm with no. of centroids (k)= 4

```

Result<- kmeans(df_norm, 2, nstart = 25)

# check the number of records in each cluster
Result$size

## [1] 2865 9451

Result$cluster

##      [1] 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 1 2
2 2 2
##      [37] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2
2 2 2
##      [73] 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2
2 2 2
##      [109] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2
2 2 2
##      [145] 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 1 1 2 2 2 2 2 2 1 1 2 2 1 1 2
2 2 2
##      [181] 1 2 2 1 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 1 1 1 2 2 2 2 2 1 2 2 2
2 2
##      [217] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 1 1 1 2 1 1 1 2 2 2 2 1 2 2 2
2 2
##      [253] 2 2 1 2 2 2 2 2 1 2 1 2 2 1 2 2 2 1 1 2 2 2 1 2 1 2 2 2 1 1 2
2 2
##      [289] 2 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 1 2 2 2 2 2 2 2
2

```

2 2 2
[325] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2
2 2 2
[361] 2 1 2 2 1 2 1 2 2 1 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1
2 1 2
[397] 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 1 1 2 1 1 1 2 1 2 2 1 2 1 1 2 2
2 2 1
[433] 2 1 1 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 1 1 2 1 2 2 1 2 1 2 2
1 2 2
[469] 2 1 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 1 1 1 1 1 2 2
2 2 2
[505] 2 1 2 2 2 2 2 2 2 1 2 2 2 1 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 1 1 2 2
2 2 2
[541] 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 1 2 2 1 1 2 2
1 2 1
[577] 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 1
2 1 2
[613] 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2
2 2 1
[649] 2 2 1 2 2 1 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 1 2 2 2 2 1 2 2 2 1 1 2 2
2 1 2
[685] 2 1 2 1 2 2 2 2 1 2 2 1 1 1 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
2 2 2
[721] 2 2 1 1 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2
1 1 2
[757] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 1 2 1 2 2 2 1 1 2 2 1 2 2 2 2 2 1 1
2 2 2
[793] 1 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 1 1 1 2 2 2 1 2 2 2 2 2 2 2 2
2 2 2
[829] 1 1 2 2 2 1 1 1 1 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 1 2 1 2
2 2 1
[865] 2 2 1 2 2 2 1 2 2 1 1 2 1 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1 2 1
2 2 2
[901] 2 2 2 1 2 1 2 1 1 2 2 1 2 2 1 1 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
2 2 1
[937] 2 2 2 2 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 2 2 2 1 1
2 2 2
[973] 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 1 2 2 1 1 2 1 2 2 2 2 2 2 2 2 1 2 2
2 1 2
[1009] 2 1 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 1 2 2 1
1 1 2
[1045] 2 2 2 2 2 1 1 2 2 1 1 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2
2 2 2
[1081] 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 1 2 1 1 2 1 1 1 2 2 2 2 2 2 2
2 1 2
[1117] 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 1 1 1
2 2 2
[1153] 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2 2 1 2
2 2 2
[1189] 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 2 1 1 2 1 2 2

```
2 2 2
## [1225] 2 2 1 1 1 1 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2
2 2 2
## [1261] 2 2 1 1 2 2 2 2 2 2 2 2 1 1 1 2 2 1 2 2 2 1 2 2 2 2 2 2 1 2 2 2 1
2 2 1
## [1297] 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 1 1 2 1 2 2 1 2 2 2 2 1 2 2
2 2 2
## [1333] 1 1 1 2 2 1 1 1 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2
2 1 2
## [1369] 2 2 1 1 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 1 2 1 1 2 2 2 2 1 2 1 2 2 1
2 2 1
## [1405] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2
2 2 2
## [1441] 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 1 1 2 2 2 1 2 2
2 2 2
## [1477] 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2 1 2 2 2
2 2 1
## [1513] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 1 1 2 1 2 2 2 2 2 1 1
2 1 2
## [1549] 1 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2
2 1 2
## [1585] 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2 2 2 1 2 2 1 2 1
2 2 2
## [1621] 2 1 2 2 2 2 1 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 1 2 2 2
1 1 1
## [1657] 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 1 2 1 1 2 1 2 2 2 2 1
1 2 2
## [1693] 2 1 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2
2 1 1
## [1729] 2 2 2 2 2 2 1 1 2 1 2 2 2 2 1 2 1 2 2 1 2 1 2 2 2 2 2 1 1 2 2 2 2
2 2 2
## [1765] 1 2 2 1 2 2 1 1 1 1 2 2 1 2 2 1 2 1 2 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2
2 2 1
## [1801] 2 2 2 1 1 2 2 1 2 2 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2
2 2 1
## [1837] 2 2 2 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1
2 2 2
## [1873] 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 2 1 2 1 2 2 1 2 2 2 1 2 2 1 1 2 2
2 1 2
## [1909] 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 1 2 2 1 2 2 2 2 1 2 2 2 2 2
2 2 2
## [1945] 1 2 1 1 1 2 2 1 2 1 2 2 2 2 2 2 2 1 2 1 1 2 2 2 1 2 2 2 1 2 2 2 2 2
2 2 2
## [1981] 2 2 2 1 2 2 2 1 2 2 1 1 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 1 1 2 2 1
1 2 2
## [2017] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 2 2
## [2053] 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1 1 2 1 2 1 1 2 2 2 2 2 2
2 2 2
## [2089] 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 1 1 1
```


2 2 2
[2125] 2 2 2 1 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 1 2 1 1 1
2 2 2
[2161] 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 1 2 2 2 1 2
2 2 2
[2197] 1 2 2 1 1 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2
2 1 2
[2233] 1 1 1 1 2 2 2 2 2 2 1 2 1 2 1 2 2 1 2 1 2 2 1 2 2 2 2
2 2 2
[2269] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 1
2 1 1
[2305] 2 2 2 1 2 1 2 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1 1 2 2 2
2 2 2
[2341] 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 1 2 2 2 2
2 2 2
[2377] 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2 1 1 1 2 2 2 2 1 2 2 1 1 2 2
2 2 2
[2413] 2 1 2 1 2 2 2 1 2 2
2 1 2
[2449] 1 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 1 1 1 1 2 1 2 2 2
2 2 2
[2485] 1 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2
2 2 1
[2521] 2 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2
2 2 1
[2557] 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2
1 1 1
[2593] 2 2 2 2 1 1 2 1 2 2 2 1 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2
2 1 1
[2629] 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 1 2 2 1 1 2
1 2 2
[2665] 2 1 2 2 2 2 1 2 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2
2 2 2
[2701] 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 1 2 2 1 2
2 2 1
[2737] 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 1 1 1 2 1 1 2 1 2 2 2 2 1 2
1 2 2
[2773] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2
2 2 2
[2809] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2
2 1 2
[2845] 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 1 2 1 1 1 2 1 2 2 2 2 2 1 2 2
2 1 2
[2881] 1 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 1 2 1 2 2 2 2 2 2 2 2 1 1 2
2 2 1
[2917] 2 1 2 2 1 2 2 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1
2 1 1
[2953] 1 1 2 1 1 2 2 1 2 2 1 2 2 1 2 2 1 1 1 1 2 2 2 2 2 2 1 1 2 2
1 2 2
[2989] 1 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 1 2 2 1 2 2 2 2 2 1 2 2 2 2

2 2 2
[3025] 2 2 2 2 2 2 1 2 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2
2 2 2
[3061] 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1
2 2 2
[3097] 2 1 2 1 1 2 2 1 2 2 2 2 2 1 2 2 2 1 2 1 1 1 2 2 2 2 2 2 1 1 1 2 2
2 2 2
[3133] 2 2 2 1 1 2 2 2 1 2 2 1 2 1 1 2 2 1 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2
2 2 2
[3169] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2
2 2 2
[3205] 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 1 2 2 2 2 2 1 1
2 1 1
[3241] 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2
2 2 1
[3277] 2 2 2 2 2 2 2 2 1 1 2 2 2 1 1 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 1 1 1
2 1 2
[3313] 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1
1 2 2
[3349] 2 1 2 2 2 2 2 2 1 1 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2
2 2 2
[3385] 1 1 2 2 2 2 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2
2 2 2
[3421] 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 2
2 2 1
[3457] 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 1 1 2 2
1 1 2
[3493] 1 1 2 2 1 2 2 2 2 2 1 2 1 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2 1 2 1
2 2 1
[3529] 2 2 2 2 2 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2
2 1 2
[3565] 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2 1 1 1 2 2 2 1 2 2
2 2 2
[3601] 2 2 2 2 1 2 2 2 1 1 2 1 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2
2 2 2
[3637] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 2 1 2 2 1 2 2 2 1 2 2 2 2 1
2 2 2
[3673] 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2
1 2 2
[3709] 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 1 2
2 2 1
[3745] 1 1 2 2 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 2 1 2 1 1 2 1 2 1 2 2 2 2
1 2 1
[3781] 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 1 1 2 2 2 2 2
2 2 2
[3817] 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2
2 2 1
[3853] 2 1 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2 1 1 2 2 1 2 2 2 2 2 1 2 2 1 2 2
1 1 2
[3889] 2 2 2 1 2 2 2 1 2 2 1 1 2 2 1 2 1 1 2 2 2 1 2 1 1 2 1 1 2 2 1 1 2

2 2 2
[3925] 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2
2 2 2
[3961] 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 1 1
2 2 2
[3997] 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 1 1 1 1 2 2 1 1 2 2 2 2
2 1 2
[4033] 1 2 1 2 1 2 2 1 1 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 1 2 2 2 2 2 2 1
1 2 2
[4069] 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 1 2
[4105] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1 2 1 1 2
2 2 2
[4141] 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 1 2 1 2 2 1 1 1 1
2 2 2
[4177] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1
2 2 1
[4213] 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1
2 2 1
[4249] 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 1 2 2 1 2 1 2 2 2 1 2 1 1 2 2 1 2 2
2 2 2
[4285] 2 2 2 2 2 1 1 2 1 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2
2 1 2
[4321] 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 1 2 2 2 2 1 2 2 2 2 2 2 2 1
2 2 2
[4357] 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2
2 1 1
[4393] 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2
2 1 2
[4429] 1 2 2 2 1 2 2 2 1 2 2 2 2 1 1 2 2 2 2 1 2 2 2 1 2 1 1 2 1 2 1 1 2
2 2 2
[4465] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 1 1 2 2 2 2 2 1 2 2
2 2 2
[4501] 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2
2 2 2
[4537] 2 2 2 2 2 2 1 1 2 2 1 2 2 1 1 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[4573] 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2
2 2 2
[4609] 1 1 2 2 2 2 2 1 2 2 2 1 2 1 2 2 2 1 1 2 2 1 2 2 1 1 2 2 1 2 1 2 2
2 1 1
[4645] 2 1 2 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2
2 2 2
[4681] 1 2 2 2 2 2 1 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2
2 2 2
[4717] 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2
1 2 2
[4753] 2 1 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 2 2 2
2 2 1
[4789] 2 2 2 1 2 1 1 2 2 2 1 1 1 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2

2 1 2
[4825] 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2
1 2 2
[4861] 1 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 1 2 2 2 2 1 1 2 1 2 2 1 2 2
2 2 2
[4897] 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2
2 1 2
[4933] 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2
2 2 2
[4969] 2 1 2 2 1 1 2 2 2 2 2 2 2
2 2 2
[5005] 2 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2
2 1 1
[5041] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2
2 1 2
[5077] 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 1 2
2 2 2
[5113] 2 2 1 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 1 2 2 1 2 2 2
2 2 2
[5149] 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 1 1 1 2 1 1 2 2 2 2 1 1 2 1 2
1 2 2
[5185] 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2
2 2 2
[5221] 2 2 1 2 1 2 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1 1 1 2 1 2
2 2 2
[5257] 2 1 1 2 2 2 2 1 2 2 1 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 1 2
2 2 1
[5293] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 1 2 2 2 2 2 1 1
2 2 2
[5329] 1 1 2 2 2 2 1 2 2 1 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2
2 2 2
[5365] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2
1 2 2
[5401] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 1 2 2 2 1
2 2 2
[5437] 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1
2 1 2
[5473] 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 1 2 2 1 1 1 2 2 2 2 1 1
2 2 2
[5509] 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 2 1 2 2 2 1 2 1 2 2 1 2 2 1 2 2 2
2 2 2
[5545] 2 2 2 1 2 1 2 2 2 1 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1
2 1 2
[5581] 2 2 2 2 1 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 1 2 2 1 2
2 2 1
[5617] 1 1 2 2 1 1 2 2 1 2 1 2 2 1 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2
2 2 2
[5653] 2 2 2 2 2 2 2 1 1 2 2 1 2 2 1 2 2 1 1 2 2 2 2 2 2 2 2 1 1 2 2 2 1
2 1 2
[5689] 2 2 1 2 1 2 1 2 2 1 2 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1

2 2 1
[5725] 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2 2
2 1 2
[5761] 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 1 1
2 1 1
[5797] 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 2
2 2 2
[5833] 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2
2 1 2
[5869] 2 1 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 1 1 2 2 1 1 2 1 2 2 2 2 2 2
2 2 1
[5905] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 1 1 2 1 2 2 2 2 2 2 2 2 2 1
1 2 2
[5941] 2 2 2 1 2 2 2 2 2 1 2 2 2 2 1 2 2 1 1 2 1 2 2 2 2 1 2 2 2 1 1 2 2
2 2 1
[5977] 2 1 2 2 2 1 2 2 2 1 2 1 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 1 1 1 2
2 2 2
[6013] 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 1 2 2 2 1
2 1 2
[6049] 2 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2
2 2 2
[6085] 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2 1 2 1 2 1 2 2 2 2 2
2 2 2
[6121] 1 2 2 1 1 2 1 2 1 2 1 2 2 2 1 2 2 1 2 1 1 2 1 2 1 2 1 2 2 1 2 2 2
1 2 2
[6157] 1 2 2 2 2 2 2 2 1 1 1 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2
1 2 2
[6193] 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 1 1 1 2 2 2 1 2 2 2 2 2 2 2 2 2
1 2 2
[6229] 1 2 2 1 2 1 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 1 2
2 1 1
[6265] 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2
2 2 2
[6301] 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 1 2 1 1 2 2 2 2
2 2 2
[6337] 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 1 2 2 1 1 1 1
2 2 2
[6373] 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 1 2
1 2 2
[6409] 1 1 2 2 2 1 2 1 1 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2
2 2 2
[6445] 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 1 2 2 2 2
2 2 1
[6481] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 1 2 2 2 1
2 1 2
[6517] 1 2 1 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2
2 1 2
[6553] 2 2 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2
2 2 1
[6589] 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1

```
2 2 2
## [6625] 2 1 2 2 1 1 2 2 1 1 2 1 2 2 2 2 1 2 2 2 2 1 2 2 1 2 2 2 2 2 1 2
1 2 2
## [6661] 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
2 2 2
## [6697] 2 2 2 2 2 2 1 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
2 2 2
## [6733] 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2
2 2 2
## [6769] 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 2 2 2 1 2 2 2 1 1 1 2 2 2
2 1 1
## [6805] 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2
2 2 2
## [6841] 2 2 2 2 2 2 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2 1 1 2 2 1 1 2 2 2 1 1 2 2
1 2 2
## [6877] 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 2 2
2 1 2
## [6913] 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 1
1 2 2
## [6949] 1 2 2 1 2 2 2 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 1 2 2 1 2
2 2 2
## [6985] 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 1 1 1 1 2 2 2
2 1 2
## [7021] 2 2 2 2 2 1 2 2 2 1 2 2 1 1 2 1 1 2 1 1 2 2 1 2 2 2 1 2 2 2 2 1 1
1 2 2
## [7057] 1 2 2 2 1 2 2 2 2 2 2 1 1 2 1 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2
2 1 2
## [7093] 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2
2 2 2
## [7129] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 2 2 1 2 1 2 2 2 2 1
1 2 2
## [7165] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 1 2 2 2 2 2
2 2 2
## [7201] 1 1 2 1 1 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 1 2 1 2
2 2 2
## [7237] 2 1 2 2 1 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 1 2 1 2 2 2 2
2 2 2
## [7273] 1 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 1 2 2 1 2 2
2 2 2
## [7309] 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1 1 2 2 2 2 1 2 1 1 2 2 2 2 2 2
2 1 1
## [7345] 2 2 2 1 2 1 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2
2 1 1
## [7381] 1 2 2 1 2 2 2 1 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2
2 1 2
## [7417] 2 2 2 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2
1 2 2
## [7453] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2
2 2 2
## [7489] 2 1 1 1 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 1
```

2 2 2
[7525] 2 2 1 2 2 1 1 2 2 2 2 1 2 2 2 1 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2
2 2 2
[7561] 2 2 1 1 2 2 2 1 2 1 2 2 2 2 2 1 1 2 2 2 1 2 1 2 1 2 2 1 2 2 1 2 2
2 2 2
[7597] 1 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
2 2 1
[7633] 2 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
2 1 2
[7669] 2 1 2 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1
2 2 2
[7705] 2 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 1 2 2
1 1 2
[7741] 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2
2 1 2
[7777] 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2
2 2 2
[7813] 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1
2 1 2
[7849] 2 2 1 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 2 1 1 1 2 2 2 1 2 2 1 1 2 2 1
2 2 2
[7885] 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2 2 1 1 2
2 2 2
[7921] 2 1 1 2 2 1 2 2 2 1 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 1 1
2 2 2
[7957] 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 1 2 2 2 2 1 1 2
2 2 2
[7993] 2 1 2 2 2 1 1 2 2 1 1 2 1 2 1 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2
2 2 2
[8029] 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
1 1 2
[8065] 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2
2 2 2
[8101] 2 2 1 2 2 2 2 1 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 1 2 1 1 2 2 2 2 2 2
2 2 2
[8137] 1 2 1 2 2 1 2 1 2 1 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[8173] 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2
2 2 2
[8209] 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 2 1 1 1 2 2 2 2 2 1 2
2 1 2
[8245] 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 1 1
[8281] 2 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2
1 1 2
[8317] 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2
2 1 2
[8353] 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2 2 2 1
2 2 2
[8389] 1 1 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 1 2 1

2 2 2
[8425] 2 2 1 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2
2 1 1
[8461] 1 2 1 1 2 2 1 2 1 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 1 1 2 2 2
2 1 2
[8497] 2 1 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1
2 2 2
[8533] 2 2 2 2 1 2 2 2 2 2 1 2 2 2 1 2 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2
2 2 2
[8569] 1 1 2 1 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2 2 2
2 2 2
[8605] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2
2 2 2
[8641] 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 1 1 2 2 1
2 2 1
[8677] 1 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2
1 2 2
[8713] 2 1 2 1 1 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 2 2 2 1 2 2 2 2 2 1 2 2
2 2 2
[8749] 2 1 1 2 2 2 2 1 2 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2
1 2 2
[8785] 1 1 2 2 2 1 2 2 2 1 2 2 2 1 2 1 1 2 2 2 1 1 2 2 2 2 2 2 1 2 2 2 2
2 1 1
[8821] 2 2 1 1 2 1 2 1 2 2 1 1 1 2 2 2 2 2 2 1 1 1 2 2 2 2 1 2 1 2 1 2 2
2 1 1
[8857] 1 2 1 2 2 2 2 1 1 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 1
1 1 2
[8893] 2 2 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2 2 2 2 2
1 2 2
[8929] 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 2 1 2 1 2 2 1
2 1 2
[8965] 2 2 2 2 2 2 1 1 2 1 2 2 2 1 2 2 2 2 1 2 2 1 2 1 1 1 2 2 2 2 2 1 2
2 1 2
[9001] 2 2 2 2 1 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 2 2
[9037] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 1 2 2 2 2 2 1 2 2 1 2
1 2 1
[9073] 1 2 2 2 1 1 1 2 2 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 1 2 1 1 2 2
1 2 2
[9109] 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 1 2 2 1
2 2 2
[9145] 2 1 2 1 2 1 2 1 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 2 2 2 2 1 1 2 2 1
1 2 2
[9181] 2 1 2 2 2 1 1 2 2 1 1 2 1 2 1 2 2 2 1 2 2 2 2 2 1 2 2 1 1 2 1 2 1
2 1 2
[9217] 1 2 1 2 2 1 2 1 2 2 1 1 2 2 2 1 2 2 1 2 2 2 1 2 1 2 2 1 1 2 2 2 2
1 2 2
[9253] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2
2 1 2
[9289] 2 2 2 2 1 2 1 2 2 2 2 2 2 1 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2 1 2 1 2

1 2 2
[9325] 2 1 1 2 1 2 1 2 2 2 1 1 2 2 2 1 2 1 2 1 1 1 2 2 1 2 2 2 1 2 2 2 2
2 2 1
[9361] 1 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2
1 2 2
[9397] 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 1 2 2 2 2 2 1 2 1 2
2 1 1
[9433] 2 1 2 2 1 2 2 2 2 1 1 2 2 1 2 2 2 2 1 2 2 2 2 2 2 1 2 1 2 1 2 1 1
2 2 2
[9469] 2 2 2 2 1 1 1 1 2 1 2 2 2 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 1 2 1 2 2
1 1 2
[9505] 2 2 2 1 1 2 2 2 2 2 1 2 2 1 2 2 1 2 1 2 1 2 2 1 2 2 2 2 2 2 1 1 2
2 2 2
[9541] 2 2 2 2 1 2 2 1 2 2 1 2 1 2 2 2 1 2 1 2 2 1 2 2 2 1 2 2 2 2 1 2 2
2 2 2
[9577] 2 1 1 1 1 2 1 2 2 2 2 2 2 2 1 2 2 1 1 2 2 1 1 2 2 1 1 1 2 1 2 2 2
2 1 2
[9613] 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 2 2 1 2 1 2 2 1 2 1 2 2 2 1 2 2 2 1 1
1 2 1
[9649] 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 1 1 2 1 1 2 2 1 1 2 2 2 2 2
1 1 1
[9685] 2 1 2 2 1 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 1
1 2 1
[9721] 1 2 2 2 2 2 1 2 2 2 1 2 1 1 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
2 2 2
[9757] 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2
1 2 2
[9793] 2 2 2 1 2 2 2 2 2 2 1 2 2 1 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2
1 2 2
[9829] 2 1 1 1 1 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 1 1
2 2 2
[9865] 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 1 2 1 2 1 2 2 2 2 2
2 1 2
[9901] 1 2 1 2 1 2 2 2 1 2 2 1 2 2 1 2 1 2 2 1 2 1 2 2 1 1 2 2 2 2 2 1 2
1 2 2
[9937] 2 2 1 1 2 1 2 2 1 2 2 1 1 1 2 2 2 1 2 2 2 2 2 1 2 1 1 2 2 2 1 2 2
2 1 2
[9973] 2 2 2 2 2 2 2 2 2 1 2 2 1 2 1 2 1 2 2 2 2 2 1 2 1 1 2 1 1 2 1 2 2
2 2 2
[10009] 2 2 2 2 1 2 1 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2
2 2 2
[10045] 2 2 1 2 2 1 2 2 2 2 2 1 2 1 2 1 2 2 2 2 1 2 2 2 1 1 2 2 1 2 1 1 2
1 2 2
[10081] 2 1 2 2 2 2 2 2 1 2 2 2 2 1 1 2 1 1 2 1 2 2 1 1 2 2 2 2 2 2 2 2 2
2 2 2
[10117] 2 2 1 2 1 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 1 2 1 2 2 1 1 2
2 2 2
[10153] 2 1 2 1 2 1 2 2 2 2 2 2 2 1 2 2 1 1 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2
2 2 2
[10189] 2 2 2 2 2 1 1 2 2 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2

2 2 2
[10225] 1 2 2 1 2 1 1 2 2 2 2 1 2 2 2 1 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2
2 1 2
[10261] 1 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2
2 2 1
[10297] 2 1 2 2 2 1 2 1 1 2 1 2 2 1 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 2 1 1
1 2 2
[10333] 1 1 2 2 1 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2
2 1 2
[10369] 2 2 2 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 1 2 1
2 2 2
[10405] 2 1 2 1 2 2 2 2 2 2 1 2 1 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2
2 1 2
[10441] 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2
2 1 2
[10477] 2 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 1 2 2 2 2 2
2 1 2
[10513] 2 2 2 2 1 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2
1 2 1
[10549] 1 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 1 1 2 2 2 2
2 2 1
[10585] 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 2 2 1 2 1 2 2 2 1 2 2 2 1 2 2 2 1
2 2 2
[10621] 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 2 2 1 1 2 2 2
2 2 2
[10657] 2 2 2 2 1 2 2 2 2 1 1 2 1 2 2 2 2 2 2 1 1 2 2 1 2 2 1 2 2 2 2 2 2
2 2 2
[10693] 2 2 2 2 1 2 2 2 2 2 2 2 1 1 2 2 1 2 1 2 1 2 2 2 1 2 2 2 2 2 2 2 2
1 2 2
[10729] 2 1 1 1 2 2 2 2 2 1 1 2 1 2 1 1 2 2 2 2 2 1 2 1 1 1 1 2 2 2 2 2 2
2 2 2
[10765] 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2
2 2 1
[10801] 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1
2 1 2
[10837] 1 1 2 2 1 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 1 1 2 2 2 2 2 1 2
2 2 2
[10873] 2 2 1 2 2 2 2 1 1 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 1 1 2 2 2 1 2 2 2
2 2 2
[10909] 2 2 2 2 1 2 2 2 2 2 1 2 2 2 1 2 2 1 1 2 2 2 1 2 2 2 2 1 2 1 2 2 1
2 2 2
[10945] 2 1 2 2 2 2 1 1 2 1 2 2 2 2 2 2 2 1 1 1 1 2 2 2 1 2 2 2 2 2 2 2 2
2 2 1
[10981] 1 2 2 2 2 2 2 1 2 1 2 1 1 2 2 2 2 1 1 2 2 1 2 2 2 2 1 1 2 2 1 1 2
1 2 2
[11017] 2 2 2 2 2 1 2 1 2 1 2 2 2 2 1 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2
2 1 2
[11053] 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 2 1 2 2
2 1 1
[11089] 2 2 2 2 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 1 2 1 2 2 1 2 2 2 2 2 2 2

1 2 2
[11125] 2 2 2 2 1 1 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 2
2 2 2
[11161] 2 1 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2
2 1 2
[11197] 2 1 2 2 1 1 2 2 2 1 2 2 2 1 2 1 2 2 2 1 1 2 2 1 2 2 2 2 1 1 2 2 2
2 2 2
[11233] 1 2 2 2 2 2 1 2 2 1 2 2 1 2 1 2 2 2 2 2 2 2 2 2 1 1 2 2 2 1 2 2 2
1 1 2
[11269] 1 1 1 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 2
2 2 2
[11305] 2 2 1 1 2 1 2 1 1 2 2 2 1 1 2 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 2
2 2 2
[11341] 2 2 2 1 1 1 1 2 2 2 2 1 2 2 1 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 2 1
2 2 1
[11377] 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 2 1 2 2 2 2 1 2 1 2 2 2 2 2 2 2
1 2 2
[11413] 2 2 2 2 1 2 2 1 2 2 2 1 2 1 2 2 2 2 2 2 1 2 1 2 1 2 2 1 2 1 2 2 1
1 1 1
[11449] 2 1 2 1 2 2 2 1 2 2 1 1 2 2 2 2 2 1 2 1 2 1 2 1 1 2 2 1 2 2 2 2 2
2 2 2
[11485] 1 2 2 2 2 1 2 2 2 2 2 2 2 2 1 2 1 1 1 2 2 2 1 2 2 2 2 2 2 2 1 2 1
2 2 1
[11521] 2 2 2 2 2 2 2 2 1 2 1 1 2 2 1 2 1 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2
2 1 2
[11557] 2 2 2 2 2 1 2 1 2 2 2 2 1 2 1 2 2 1 2 1 2 1 2 2 2 1 1 2 1 2 2 2 2
2 2 2
[11593] 2 2 2 1 2 2 1 1 1 1 1 1 2 1 2 2 2 1 2 2 2 2 1 1 2 1 2 2 1 2 2 2 1
2 2 2
[11629] 2 2 2 2 2 2 1 2 1 1 1 1 1 2 2 2 1 1 1 2 1 2 2 2 2 2 2 1 2 2 2 2 2
2 2 2
[11665] 2 2 1 2 2 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2
1 2 2
[11701] 2 2 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1
2 1 1
[11737] 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 1 2 2 2
1 2 2
[11773] 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 1 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 2 1
2 1 1
[11809] 2 2 2 2 1 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 2 1 1 1 2 2 2 2 2 2 1 1 1
2 2 2
[11845] 2 2 1 2 2 2 1 2 2 2 2 1 1 2 2 2 2 2 1 2 1 2 2 2 2 2 1 1 2 2 2 2 2
1 2 2
[11881] 1 2 2 2 1 2 1 2 1 1 2 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 2 2 2 2 2 1
1 1 2
[11917] 2 2 2 2 2 1 2 1 1 2 2 2 2 2 1 2 1 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2
2 2 2
[11953] 1 2 2 2 2 2 2 2 1 2 2 2 2 1 2 2 2 2 1 2 2 1 2 2 2 2 1 2 1 2 2 2 2
1 2 2
[11989] 1 2 2 2 2 2 2 2 2 1 1 2 1 1 1 2 2 1 2 2 1 1 2 1 1 2 2 2 2 2 2 1 2

```

1 2 2
## [12025] 1 1 2 2 1 2 1 2 2 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 1
2 2 2
## [12061] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 2 2 2 2
2 2 2
## [12097] 2 2 2 2 2 2 2 2 2 2 1 2 1 1 2 2 2 2 2 2 1 1 1 2 1 2 2 2 1 2 1 2 1
1 2 2
## [12133] 1 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 2 2 1 2 1 1 2
2 2 2
## [12169] 2 1 2 2 1 2 2 1 2 1 1 2 1 2 1 2 2 2 1 2 1 1 2 2 2 2 2 2 2 1 2 1 2
2 1 2
## [12205] 1 2 2 1 1 2 1 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2
1 2 2
## [12241] 1 1 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2 1 1 2 2 2 2 2 1 2 2 2 2 1 2 2
2 2 1
## [12277] 1 2 2 2 2 2 2 2 2 2 2 2 1 2 1 2 2 2 1 2 2 2 1 2 1 1 2 2 1 2 2 2 2
2 2 1
## [12313] 1 1 2 1

table(Result$cluster, df$Revenue)

##
##      FALSE TRUE
##    1  2366  499
##    2  8042 1409

```

Hierarchical Clustering

Scaling

```

df <- scale(df_norm)
head(df)

##      Administrative Administrative_Duration Informational
## [1,]      -0.6975533           -0.4574578      -0.3966145
## [2,]      -0.6975533           -0.4574578      -0.3966145
## [3,]      -0.6975533           -0.4631119      -0.3966145
## [4,]      -0.6975533           -0.4574578      -0.3966145
## [5,]      -0.6975533           -0.4574578      -0.3966145
## [6,]      -0.6975533           -0.4574578      -0.3966145
##      Informational_Duration ProductRelated ProductRelated_Duration
BounceRates
## [1,]      -0.2450294      -0.6914734           -0.6247671
3.67247746
## [2,]      -0.2450294      -0.6689966           -0.5913358 -
0.45743910
## [3,]      -0.2521304      -0.6914734           -0.6252895
3.67247746
## [4,]      -0.2450294      -0.6689966           -0.6233742
0.57504004
## [5,]      -0.2450294      -0.4891823           -0.2969835 -

```

```

0.04444744
## [6,]          -0.2450294      -0.2868911          -0.5442099 -
0.13139305
##      ExitRates PageValues SpecialDay Administrative.1
## [1,] 3.2352400 -0.3173633 -0.309001      -0.6975533
## [2,] 1.1745443 -0.3173633 -0.309001      -0.6975533
## [3,] 3.2352400 -0.3173633 -0.309001      -0.6975533
## [4,] 1.9988226 -0.3173633 -0.309001      -0.6975533
## [5,] 0.1441964 -0.3173633 -0.309001      -0.6975533
## [6,] -0.3800157 -0.3173633 -0.309001      -0.6975533
##      Administrative_Duration.1 Informational.1 Informational_Duration.1
## [1,]          -0.4574578      -0.3966145          -0.2450294
## [2,]          -0.4574578      -0.3966145          -0.2450294
## [3,]          -0.4631119      -0.3966145          -0.2521304
## [4,]          -0.4574578      -0.3966145          -0.2450294
## [5,]          -0.4574578      -0.3966145          -0.2450294
## [6,]          -0.4574578      -0.3966145          -0.2450294
##      ProductRelated.1 ProductRelated_Duration.1 BounceRates.1 ExitRates.1
## [1,]          -0.6914734          -0.6247671      3.67247746      3.2352400
## [2,]          -0.6689966          -0.5913358      -0.45743910      1.1745443
## [3,]          -0.6914734          -0.6252895      3.67247746      3.2352400
## [4,]          -0.6689966          -0.6233742      0.57504004      1.9988226
## [5,]          -0.4891823          -0.2969835      -0.04444744      0.1441964
## [6,]          -0.2868911          -0.5442099      -0.13139305      -0.3800157
##      PageValues.1 SpecialDay.1      MonthAug      MonthDec MonthFeb      MonthJul
## [1,] -0.3173633      -0.309001 -0.1908812 -0.4038323 8.119694 -0.1906527
## [2,] -0.3173633      -0.309001 -0.1908812 -0.4038323 8.119694 -0.1906527
## [3,] -0.3173633      -0.309001 -0.1908812 -0.4038323 8.119694 -0.1906527
## [4,] -0.3173633      -0.309001 -0.1908812 -0.4038323 8.119694 -0.1906527
## [5,] -0.3173633      -0.309001 -0.1908812 -0.4038323 8.119694 -0.1906527
## [6,] -0.3173633      -0.309001 -0.1908812 -0.4038323 8.119694 -0.1906527
##      MonthJune      MonthMar      MonthMay      MonthNov      MonthOct      MonthSep
## [1,] -0.1547326 -0.4262818 -0.6128603 -0.5672008 -0.215991 -0.194282
## [2,] -0.1547326 -0.4262818 -0.6128603 -0.5672008 -0.215991 -0.194282
## [3,] -0.1547326 -0.4262818 -0.6128603 -0.5672008 -0.215991 -0.194282
## [4,] -0.1547326 -0.4262818 -0.6128603 -0.5672008 -0.215991 -0.194282
## [5,] -0.1547326 -0.4262818 -0.6128603 -0.5672008 -0.215991 -0.194282
## [6,] -0.1547326 -0.4262818 -0.6128603 -0.5672008 -0.215991 -0.194282
##      OperatingSystems      Browser      Region TrafficType
VisitorTypeNew_Visitor
## [1,] -1.2332048 -0.7901988 -0.8941841 -0.76292777      -
0.3993337
## [2,] -0.1361914 -0.2081361 -0.8941841 -0.51445574      -
0.3993337
## [3,] 2.0578354 -0.7901988 2.4360812 -0.26598370      -
0.3993337
## [4,] 0.9608220 -0.2081361 -0.4779009 -0.01751167      -
0.3993337
## [5,] 0.9608220 0.3739266 -0.8941841 -0.01751167      -
0.3993337

```

```
## [6,]          -0.1361914 -0.2081361 -0.8941841 -0.26598370      -
0.3993337
##      VisitorTypeOther VisitorTypeReturning_Visitor
## [1,]          -0.0833606              0.410877
## [2,]          -0.0833606              0.410877
## [3,]          -0.0833606              0.410877
## [4,]          -0.0833606              0.410877
## [5,]          -0.0833606              0.410877
## [6,]          -0.0833606              0.410877
##
WeekendMonth...OperatingSystems...Browser...Region.....Traf
ficType...VisitorType...WeekendFALSE
## [1,]
0.5505615
## [2,]
0.5505615
## [3,]
0.5505615
## [4,]
0.5505615
## [5,]
-1.8161802
## [6,]
0.5505615
##
WeekendMonth...OperatingSystems...Browser...Region.....Traf
ficType...VisitorType...WeekendTRUE
## [1,]
-0.5505615
## [2,]
-0.5505615
## [3,]
-0.5505615
## [4,]
-0.5505615
## [5,]
1.8161802
## [6,]
-0.5505615
```

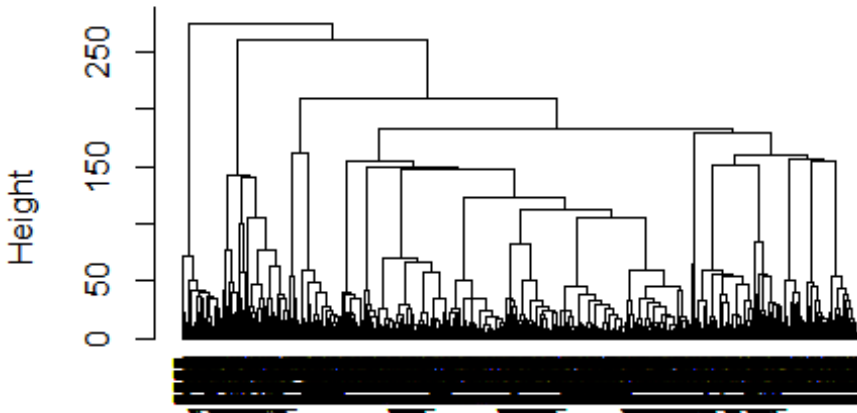
computing the Euclidean distance between observations,

```
d <- dist(df, method = "euclidean")
```

We then hierarchical clustering using the Ward's method

```
res.hc <- hclust(d, method = "ward.D2" )
plot(res.hc, cex = 0.6, hang = -1)
```

Cluster Dendrogram



```
hclust (*, "ward.D2")
```

```
cut <- cutree(res.hc, k = 2)
```

cut

[illegible]

[illegible]

[illegible]

[illegible]

```
## [3133] 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 1
1 1 2
## [3169] 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 1
1 1 1
## [3205] 1 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 1
## [3241] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 2 1
1 2 1
## [3277] 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
## [3313] 1 1 1 1 2 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 1
1 1 2
## [3349] 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 1
1 1 1
## [3385] 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 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
1 1 1
## [3457] 1 1 1 2 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 1 1
## [3493] 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
1 1 1
## [3529] 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 1 1 1 1 1
1 1 2
## [3565] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 2 2 1 2 1 1
1 1 1
## [3601] 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 1 1 1 1 1 1 1
1 1 2
## [3637] 2 1 1 1 1 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 1 1 1
1 1 1
## [3673] 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 1 1 1 1 1 1
1 1 2
## [3709] 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
2 1 1
## [3745] 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 1 2 1 1 1 1 1
1 1 1
## [3781] 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
1 1 1
## [3817] 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 1 1 1 1 1 1
1 1 1
## [3853] 1 1 1 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
## [3889] 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 1 1 1 1 2 1 1
1 1 1
## [3925] 1 1 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
## [3961] 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1
## [3997] 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 2 1 1 1
1 1 1
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

##	[9433]	1	2	1	1	1	1	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																													
##	[9469]	1	1	1	1	1	1	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	1	1																													
##	[9505]	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	1	2	1	1
2	2	1																													
##	[9541]	1	1	1	2	1	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	2	1	2	1	2	1	1	1
1	1	1																													
##	[9577]	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	1	1																													
##	[9613]	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
1	1	1																													
##	[9649]	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	1	1																													
##	[9685]	1	1	1	1	1	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																													
##	[9721]	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	2																													
##	[9757]	1	1	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	1	1	1
1	1	1																													
##	[9793]	1	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	2	1	1	1
1	1	1																													
##	[9829]	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																													
##	[9865]	2	1	1	1	1	1																								

[illegible]

## [11233]	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	1	1
1 1 1																																					
## [11269]	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	1	1	1	1	1	1	1	1
1 1 1																																					
## [11305]	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	1	1
1 1 1																																					
## [11341]	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	1	1
1 1 1																																					
## [11377]	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	1	1	1	1	1
1 1 1																																					
## [11413]	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	1	1	1	1	1	1	1
1 1 1																																					
## [11449]	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	1	1	1	1	1	1	1	1	1	1
1 1 1																																					
## [11485]	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	1	1	1	1	1	1	1	1	1
1 1 1																																					
## [11521]	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	1	1	1	1	1	1
1 1 1																																					
## [11557]	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	1	1
1 1 1																																					
## [11593]	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	2	1	1																	

```

## [12133] 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 2 1 1 1
1 2 1
## [12169] 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 1 1
1 1 1
## [12205] 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 1 1 1 1 1
1 1 1
## [12241] 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
## [12277] 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 1 1 1 1
1 1 1
## [12313] 1 1 1 1

table(cut)

## cut
##      1      2
## 11563   753

```

Conclusion Both K-mean and Hierarchical Clustering were unable to predict correctly if Revenue will be true or false. DBSCAN is not stable for this data since it has high dimensional.

Recommendation: Dimensional reduction is the best option to get better results