Online Shopping, Clothing, Analysis

Naomi Chebet

9/10/2020

1. Business Understanding

1.1 Defining the Question

Main Research Question: Kira Plastina's brand sales and marketing is interested in understanding customer behavior. They have collected data for a year. With this data the team wants to know — what are the characteristics of customer groups that shop at Kira Platina?

1.2 Providing the Context

Kira Plastina is an online clothing shop that sells ready-to-wear designer clothes. According to it's website, Kira Platina brand targets young girls from the age of 18-25years and older. The brand first appeared in Russia in 2007 with Kira being the youngest of the Russia high fashion. In 2008, the brand received a "Successful Debut at Milan, Italy, Fashion Week". The successful debut was awarded by Milan Boselli, the President of the Italian National Fashion Chamber.

Some of Kira Platina's collection include: dresses, coats, trousers, tops, and blouses, which are all produced in Russia.

The objective of this project is to analyze the given data and provide the results to the brand and marketing team on features/characteristics of customers that shop at Kira Plastina.

Understanding the traits of your customers is important for any business. In the case of Kira Plastina, their advertisement and all the marketing efforts will be better targeted to the right customer.

1.1 Metric for Success

a.) Successfully implementing K-Means with optimal number of clusters. There after performing EDA on the clusters to get insights on the traits of customers that shop at Kira Plastina.

b.) Successfully implementing Hierachical Clustering. After implementation, interpret the data accordingly to get insights on the kinds of customers that shop at Kira plastina.

c.) Compile a list of the kinds of customers that shop at Kira Plastina, the list will be presented to the brand and marketing team of Kira Plastina.

1.3 Experimental Design

The approach for the project will include:

1. Business Understanding
2. Loading and Checking the Data
3. Cleaning the Data
4. EDA using Univariate and Bivariate Analysis
5. Implementing the Solution with K-Means and Hierachical Clustering
6. Challenging the Solution

1.4 Data Relevance

Here are some of the information provided about the dataset:

a.) The dataset consists of 10 numerical and 8 categorical attributes. The 'Revenue' attribute can be used as the class label.

b.) "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.

c.) The "Bounce Rate", "Exit Rate" and "Page Value" features represent the metrics measured by "Google Analytics" for each page in the e-commerce site.

Note: it's likely that we won't be using these columns in our analysis as it's specific to what Google Metric when someone visits a page online.

d.) 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.

e.) The value of the "Exit Rate" feature for a specific web page is calculated as for all page views to the page, the percentage that was the last in the session.

Note: this seems to be Google specific as well

f.) The "Page Value" feature represents the average value for a web page that a user visited before completing an e-commerce transaction.

g.) 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 Valentine'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.

h.) 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.

2. Importing Libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(ggcorrplot)
```

3. Loading and Checking the Data

```
#loading the data
shoppers <- read.csv("~/Moringa School/R Programming/R datasets/online_shoppers_intention.csv")
```

```
#previewing the top of the data
head(shoppers)
```

```
##   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
```

```
#previewing the bottom of the data
tail(shoppers)
```

```
##       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 how the data structure looks like using glimpse

glimpse(shoppers)
```

```
## Rows: 12,330
## Columns: 18
## $ Administrative          <int> 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0...
## $ Administrative_Duration <dbl> 0, 0, -1, 0, 0, 0, -1, -1, 0, 0, 0, 0, 0, 0...
## $ Informational           <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ Informational_Duration  <dbl> 0, 0, -1, 0, 0, 0, -1, -1, 0, 0, 0, 0, 0, 0...
## $ ProductRelated          <int> 1, 2, 1, 2, 10, 19, 1, 1, 2, 3, 3, 16, 7, 6...
## $ ProductRelated_Duration <dbl> 0.000000, 64.000000, -1.000000, 2.666667, 6...
## $ BounceRates             <dbl> 0.200000000, 0.000000000, 0.200000000, 0.05...
## $ ExitRates               <dbl> 0.200000000, 0.100000000, 0.200000000, 0.14...
## $ PageValues              <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ SpecialDay              <dbl> 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.4, 0.0, 0.8...
## $ Month                   <chr> "Feb", "Feb", "Feb", "Feb", "Feb", "Feb", "...
## $ OperatingSystems        <int> 1, 2, 4, 3, 3, 2, 2, 1, 2, 2, 1, 1, 1, 2, 3...
## $ Browser                 <int> 1, 2, 1, 2, 3, 2, 4, 2, 2, 4, 1, 1, 1, 5, 2...
## $ Region                  <int> 1, 1, 9, 2, 1, 1, 3, 1, 2, 1, 3, 4, 1, 1, 3...
## $ TrafficType             <int> 1, 2, 3, 4, 4, 3, 3, 5, 3, 2, 3, 3, 3, 3, 3...
## $ VisitorType             <chr> "Returning_Visitor", "Returning_Visitor", "...
## $ Weekend                 <lgl> FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FA...
## $ Revenue                 <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, F...
```

The dataset has 12,330 records and 18 columns. The columns have numerical and categorical data types. We will further explore on the unique values in each column below.

```
#checking for unique values in all the columns
sapply(shoppers, function(x)length(unique(x)))
```

```
##          Administrative Administrative_Duration           Informational
##                      28                    3337                      18
##  Informational_Duration          ProductRelated ProductRelated_Duration
##                    1260                     312                    9553
##             BounceRates               ExitRates              PageValues
##                    1873                    4778                    2704
##              SpecialDay                   Month        OperatingSystems
##                       6                      10                       8
##                 Browser                  Region             TrafficType
##                      13                       9                      20
##             VisitorType                 Weekend                 Revenue
```

```
##                         3                         2                         2
```

```
#checking for unique values in columns of interest

unique(shoppers$Administrative)
```

```
##  [1]  0  1  2  4 12  3 10  6  5  9  8 16 13 11  7 18 14 17 19 15 NA 24 22 21 20
## [26] 23 27 26
```

```
unique(shoppers$Informational)
```

```
##  [1]  0  1  2  4 16  5  3 14  6 12  7 NA  9 10  8 11 24 13
```

```
unique (shoppers$ProductRelated)
```

```
##   [1]   1   2  10  19   3  16   7   6  23  13  20   8   5  32   4  45  14  52
##  [19]   9  46  15  22  11  12  36  42  27  90  18  38  17 128  25  30  21  51
##  [37]  26  28  31  24  50  96  49  68  98  67  55  35  37  29  34  71  63  87
##  [55]  40  33  54  64  75  39 111  81  61  47  44  88 149  41   0  79  66  43
##  [73] 258  80  62  83 173  48  58  57  56  69  82  59 109 287  53  84  78 137
##  [91] 113  89  65  60  NA 104 129  77  74  93  76  72 194 140 110 132 115  73
## [109] 328 160  86 150  95 130 151 117 124 127 125 116 105  92 157 154 220 187
## [127] 112 131 159  94 204 142 206 102 313 145  85  97 198 181 126 106 101 108
## [145] 119  70 122  91 276 100 291 114 172 217 141 133 156 136 180 135 195  99
## [163] 362 179 118 175 148 440 103 178 184 705 134 176 146 189 120 193 222 121
## [181] 107 305 199 439 223 230 280 377 310 158 486 153 139 182 221 229 216 170
## [199] 202 346 274 240 162 123 211 227 168 161 429 686 167 518 256 255 358 213
## [217] 191 282 155 138 246 237 271 171 414 219 262 409 243 241 197 449 143 188
## [235] 391 238 152 165 293 174 584 164 311 340 250 200 385 292 232 251 517 225
## [253] 169 309 235 501 224 275 318 144 397 343 245 186 337 351 166 349 423 359
## [271] 163 147 264 312 226 324 266 260 338 272 534 470 207 218 326 190 304 205
## [289] 233 401 177 330 286 247 357 315 231 339 283 374 248 279 281 234 261 290
## [307] 336 378 254 183 210 192
```

```
unique(shoppers$SpecialDay)
```

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

```
unique(shoppers$Month)
```

```
##  [1] "Feb"  "Mar"  "May"  "Oct"  "June" "Jul"  "Aug"  "Nov"  "Sep"  "Dec"
```

```
unique(shoppers$OperatingSystems)
```

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

```
unique(shoppers$Browser)
```

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

```
unique(shoppers$Region)
```

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

```
unique(shoppers$TrafficType)
```

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

```
unique(shoppers$VisitorType)
```

```
## [1] "Returning_Visitor" "New_Visitor"       "Other"
```

```r
unique(shoppers$Weekend)
```

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

```r
unique(shoppers$Revenue)
```

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

a.) In the administrative column we see the total number of administrative pages visited per visitor. The total number ranges from 0 to 19 with some missing values available. We will take care of the missing values during cleaning

b.) The total number of informational pages visited by a visitor ranges from 0-16. With some missing values present, we will take care of this during cleaning

c.) Product related pages visited ranges from 1 - 486, with missing values present

d.) The closeness to a special day ranged from 0-1.

e.) We have 10 unique months when the visit to the site happened. No January and April.

f.) The type of browser used ranged from 1-8

g.) The data also includes the region from which the visitor was visiting from. We have 10 unique regions

h.) The type of customer visiting Kira Plastina were: returning_visitor, New_visitor, and other.

i.) We have information to indicate if the page was visited on the weekend or not

j.) The revenue column gives us information on whether a visit to the site resulted in revenue or not.

Note: for the remaining columns with a lot of unique values, we will utilize summary statistics to get some insights on the column's data.

4. Data Cleaning

4.1 Missing values

```r
#checking for total number of missing values in all the columns

colSums(is.na(shoppers))
```

```
##          Administrative Administrative_Duration           Informational
##                      14                      14                      14
##  Informational_Duration          ProductRelated ProductRelated_Duration
##                      14                      14                      14
##             BounceRates               ExitRates              PageValues
##                      14                      14                       0
##              SpecialDay                   Month        OperatingSystems
##                       0                       0                       0
##                 Browser                  Region             TrafficType
##                       0                       0                       0
##             VisitorType                 Weekend                 Revenue
##                       0                       0                       0
```

There are about 14 missing values in column 1 to 8.

```r
#let's confirm the total number of missing values in all the columns
sum(is.na(shoppers))
```

```
## [1] 112
```

In total we have 112 missing values. The dataset has a total of 12,330 records. The missing values is about .908 percent of the total dataset, we will get rid of them so that they dont affect our analysis.

```
new.shoppers <- na.omit(shoppers) #creating a dataset with no missing values
colSums(is.na(new.shoppers)) #confirming that the missing values have been dropped
```

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

The missing values have been dropped

```
#checking for duplicates
new.shoppers.dup <- new.shoppers[duplicated(new.shoppers),]
new.shoppers.dup
```

```
##       Administrative Administrative_Duration Informational
## 159                0                       0             0
## 179                0                       0             0
## 419                0                       0             0
## 457                0                       0             0
## 484                0                       0             0
## 513                0                       0             0
## 555                0                       0             0
## 590                0                       0             0
## 660                0                       0             0
## 775                0                       0             0
## 873                0                       0             0
## 890                0                       0             0
## 923                0                       0             0
## 948                0                       0             0
## 975                0                       0             0
## 1035               0                       0             0
## 1120               0                       0             0
## 1171               0                       0             0
## 1177               0                       0             0
## 1214               0                       0             0
## 1215               0                       0             0
## 1292               0                       0             0
## 1326               0                       0             0
## 1357               0                       0             0
## 1367               0                       0             0
## 1382               0                       0             0
## 1391               0                       0             0
## 1395               0                       0             0
## 1437               0                       0             0
## 1454               0                       0             0
```

```
## 1516              0                   0             0
## 1574              0                   0             0
## 1609              0                   0             0
## 1698              0                   0             0
## 1776              0                   0             0
## 1805              0                   0             0
## 1840              0                   0             0
## 1867              0                   0             0
## 1926              0                   0             0
## 1934              0                   0             0
## 1950              0                   0             0
## 2057              0                   0             0
## 2058              0                   0             0
## 2236              0                   0             0
## 2622              0                   0             0
## 2740              0                   0             0
## 3232              0                   0             0
## 3273              0                   0             0
## 3282              0                   0             0
## 3578              0                   0             0
## 3651              0                   0             0
## 3664              0                   0             0
## 3722              0                   0             0
## 3892              0                   0             0
## 4164              0                   0             0
## 4183              0                   0             0
## 4232              0                   0             0
## 4344              0                   0             0
## 4375              0                   0             0
## 4404              0                   0             0
## 4427              0                   0             0
## 4464              0                   0             0
## 4490              0                   0             0
## 4553              0                   0             0
## 4818              0                   0             0
## 4884              0                   0             0
## 4914              0                   0             0
## 5039              0                   0             0
## 5044              0                   0             0
## 5057              0                   0             0
## 5119              0                   0             0
## 5199              0                   0             0
## 5200              0                   0             0
## 5255              0                   0             0
## 5277              0                   0             0
## 5287              0                   0             0
## 5356              0                   0             0
## 5408              0                   0             0
## 6930              0                   0             0
## 7152              0                   0             0
## 7636              0                   0             0
## 8545              0                   0             0
## 9307              0                   0             0
## 9495              0                   0             0
```

```
## 9552                0                       0              0
## 9569                0                       0              0
## 9582                0                       0              0
## 9719                0                       0              0
## 9770                0                       0              0
## 9879                0                       0              0
## 9908                0                       0              0
## 10147               0                       0              0
## 10223               0                       0              0
## 10270               0                       0              0
## 10573               0                       0              0
## 10632               0                       0              0
## 10752               0                       0              0
## 10796               0                       0              0
## 10842               0                       0              0
## 10989               0                       0              0
## 11044               0                       0              0
## 11206               0                       0              0
## 11405               0                       0              0
## 11524               0                       0              0
## 11582               0                       0              0
## 11625               0                       0              0
## 11659               0                       0              0
## 11734               0                       0              0
## 11748               0                       0              0
## 11802               0                       0              0
## 11814               0                       0              0
## 11828               0                       0              0
## 11935               0                       0              0
## 11939               0                       0              0
## 12160               0                       0              0
## 12181               0                       0              0
## 12186               0                       0              0
##       Informational_Duration ProductRelated ProductRelated_Duration BounceRates
## 159                        0              1                       0         0.2
## 179                        0              1                       0         0.2
## 419                        0              1                       0         0.2
## 457                        0              1                       0         0.2
## 484                        0              1                       0         0.2
## 513                        0              1                       0         0.2
## 555                        0              1                       0         0.2
## 590                        0              1                       0         0.2
## 660                        0              2                       0         0.2
## 775                        0              1                       0         0.2
## 873                        0              1                       0         0.2
## 890                        0              1                       0         0.2
## 923                        0              1                       0         0.2
## 948                        0              1                       0         0.2
## 975                        0              1                       0         0.2
## 1035                       0              1                       0         0.2
## 1120                       0              1                       0         0.2
## 1171                       0              1                       0         0.2
## 1177                       0              1                       0         0.2
## 1214                       0              1                       0         0.2
```

```
## 1215    0    1    0    0.2
## 1292    0    2    0    0.2
## 1326    0    1    0    0.2
## 1357    0    2    0    0.2
## 1367    0    1    0    0.2
## 1382    0    1    0    0.2
## 1391    0    1    0    0.2
## 1395    0    1    0    0.2
## 1437    0    1    0    0.2
## 1454    0    1    0    0.2
## 1516    0    1    0    0.2
## 1574    0    1    0    0.2
## 1609    0    1    0    0.2
## 1698    0    1    0    0.2
## 1776    0    1    0    0.2
## 1805    0    1    0    0.2
## 1840    0    1    0    0.2
## 1867    0    1    0    0.2
## 1926    0    1    0    0.2
## 1934    0    1    0    0.2
## 1950    0    1    0    0.2
## 2057    0    1    0    0.2
## 2058    0    1    0    0.2
## 2236    0    1    0    0.2
## 2622    0    1    0    0.2
## 2740    0    1    0    0.2
## 3232    0    1    0    0.2
## 3273    0    1    0    0.2
## 3282    0    1    0    0.2
## 3578    0    1    0    0.2
## 3651    0    1    0    0.2
## 3664    0    1    0    0.2
## 3722    0    1    0    0.2
## 3892    0    1    0    0.2
## 4164    0    1    0    0.2
## 4183    0    1    0    0.2
## 4232    0    1    0    0.2
## 4344    0    1    0    0.2
## 4375    0    1    0    0.2
## 4404    0    1    0    0.2
## 4427    0    1    0    0.2
## 4464    0    1    0    0.2
## 4490    0    1    0    0.2
## 4553    0    2    0    0.2
## 4818    0    1    0    0.2
## 4884    0    1    0    0.2
## 4914    0    1    0    0.2
## 5039    0    1    0    0.2
## 5044    0    1    0    0.2
## 5057    0    1    0    0.2
## 5119    0    1    0    0.2
## 5199    0    1    0    0.2
## 5200    0    2    0    0.2
## 5255    0    1    0    0.2
```

```
## 5277                     0             1                     0         0.2
## 5287                     0             1                     0         0.2
## 5356                     0             1                     0         0.2
## 5408                     0             1                     0         0.2
## 6930                     0             1                     0         0.2
## 7152                     0             1                     0         0.2
## 7636                     0             1                     0         0.2
## 8545                     0             1                     0         0.2
## 9307                     0             1                     0         0.2
## 9495                     0             1                     0         0.2
## 9552                     0             1                     0         0.2
## 9569                     0             1                     0         0.2
## 9582                     0             1                     0         0.2
## 9719                     0             1                     0         0.2
## 9770                     0             1                     0         0.2
## 9879                     0             1                     0         0.2
## 9908                     0             1                     0         0.2
## 10147                    0             1                     0         0.2
## 10223                    0             2                     0         0.2
## 10270                    0             1                     0         0.2
## 10573                    0             1                     0         0.2
## 10632                    0             1                     0         0.2
## 10752                    0             1                     0         0.2
## 10796                    0             1                     0         0.2
## 10842                    0             1                     0         0.2
## 10989                    0             1                     0         0.2
## 11044                    0             1                     0         0.2
## 11206                    0             1                     0         0.2
## 11405                    0             1                     0         0.2
## 11524                    0             1                     0         0.2
## 11582                    0             1                     0         0.2
## 11625                    0             1                     0         0.2
## 11659                    0             1                     0         0.2
## 11734                    0             1                     0         0.2
## 11748                    0             1                     0         0.2
## 11802                    0             1                     0         0.2
## 11814                    0             1                     0         0.2
## 11828                    0             1                     0         0.2
## 11935                    0             1                     0         0.2
## 11939                    0             1                     0         0.2
## 12160                    0             1                     0         0.2
## 12181                    0             1                     0         0.2
## 12186                    0             1                     0         0.2
##       ExitRates PageValues SpecialDay Month OperatingSystems Browser Region
## 159         0.2          0       0.0   Feb                 1       1      1
## 179         0.2          0       0.0   Feb                 3       2      3
## 419         0.2          0       0.0   Mar                 1       1      1
## 457         0.2          0       0.0   Mar                 2       2      4
## 484         0.2          0       0.0   Mar                 3       2      3
## 513         0.2          0       0.0   Mar                 2       2      1
## 555         0.2          0       0.0   Mar                 2       2      1
## 590         0.2          0       0.0   Mar                 2       2      1
## 660         0.2          0       0.0   Mar                 2       5      1
## 775         0.2          0       0.0   Mar                 2       2      4
```

```
## 873         0.2         0         0.0   Mar               3       2       3
## 890         0.2         0         0.0   Mar               1       1       2
## 923         0.2         0         0.0   Mar               3       2       2
## 948         0.2         0         0.0   Mar               2       2       1
## 975         0.2         0         0.0   Mar               2       2       1
## 1035        0.2         0         0.0   Mar               2       2       1
## 1120        0.2         0         0.0   Mar               2       2       1
## 1171        0.2         0         0.0   Mar               3       2       1
## 1177        0.2         0         0.0   Mar               2       4       1
## 1214        0.2         0         0.0   Mar               3       2       3
## 1215        0.2         0         0.0   Mar               1       1       1
## 1292        0.2         0         0.0   Mar               2       2       1
## 1326        0.2         0         0.0   Mar               1       1       3
## 1357        0.2         0         0.0   Mar               1       1       1
## 1367        0.2         0         0.0   Mar               1       1       8
## 1382        0.2         0         0.0   Mar               1       1       4
## 1391        0.2         0         0.0   Mar               2       2       1
## 1395        0.2         0         0.0   Mar               2       2       1
## 1437        0.2         0         0.0   Mar               3       2       3
## 1454        0.2         0         0.0   Mar               2       2       1
## 1516        0.2         0         0.0   Mar               1       1       1
## 1574        0.2         0         0.0   Mar               2       2       1
## 1609        0.2         0         0.0   Mar               2       2       7
## 1698        0.2         0         0.0   Mar               2       2       2
## 1776        0.2         0         0.0   Mar               3       2       1
## 1805        0.2         0         0.0   Mar               1       1       8
## 1840        0.2         0         0.0   Mar               2       2       1
## 1867        0.2         0         0.0   Mar               1       1       1
## 1926        0.2         0         0.0   Mar               3       2       1
## 1934        0.2         0         0.0   Mar               2       2       1
## 1950        0.2         0         0.0   Mar               2       2       1
## 2057        0.2         0         0.0   Mar               3       2       3
## 2058        0.2         0         0.0   Mar               2       4       1
## 2236        0.2         0         0.0   May               1       1       4
## 2622        0.2         0         0.0   May               1       1       1
## 2740        0.2         0         0.0   May               2       2       1
## 3232        0.2         0         0.0   May               2       4       1
## 3273        0.2         0         0.0   May               1       1       3
## 3282        0.2         0         0.0   May               1       1       1
## 3578        0.2         0         0.0   May               2       2       1
## 3651        0.2         0         0.0   May               2       2       4
## 3664        0.2         0         0.0   May               1       1       1
## 3722        0.2         0         0.0   May               1       1       4
## 3892        0.2         0         0.0   May               2       2       7
## 4164        0.2         0         0.0   May               1       1       4
## 4183        0.2         0         0.0   May               1       1       1
## 4232        0.2         0         0.0   May               2       2       2
## 4344        0.2         0         0.0   May               3       2       1
## 4375        0.2         0         0.0   May               2       2       1
## 4404        0.2         0         0.0   May               2       2       1
## 4427        0.2         0         0.0   May               2       2       1
## 4464        0.2         0         0.0   May               1       1       1
## 4490        0.2         0         0.0   May               3       2       9
## 4553        0.2         0         0.0   May               2       2       2
```

```
## 4818        0.2          0        0.0    May                2       2       1
## 4884        0.2          0        0.0    May                2       2       1
## 4914        0.2          0        0.8    May                2       2       1
## 5039        0.2          0        0.0    May                3       2       3
## 5044        0.2          0        0.0    May                2       2       1
## 5057        0.2          0        0.0    May                2       2       6
## 5119        0.2          0        0.0    May                1       1       6
## 5199        0.2          0        0.0    May                2       2       1
## 5200        0.2          0        0.0    May                2       2       2
## 5255        0.2          0        0.6    May                2       2       1
## 5277        0.2          0        0.0    May                3       2       3
## 5287        0.2          0        0.0    May                1       1       3
## 5356        0.2          0        0.0    May                1       1       3
## 5408        0.2          0        0.0    May                2       4       1
## 6930        0.2          0        0.0   June                2       2       1
## 7152        0.2          0        0.0   June                2       2       1
## 7636        0.2          0        0.0   June                3       2       3
## 8545        0.2          0        0.0    Nov                3       2       3
## 9307        0.2          0        0.0    Dec                3       2       3
## 9495        0.2          0        0.0    Dec                2       2       1
## 9552        0.2          0        0.0    Nov                3       2       4
## 9569        0.2          0        0.0    Dec                2       2       8
## 9582        0.2          0        0.0    Nov                2       2       1
## 9719        0.2          0        0.0    Nov                3       2       7
## 9770        0.2          0        0.0    Dec                2       2       2
## 9879        0.2          0        0.0    Dec                2       2       6
## 9908        0.2          0        0.0    Dec                2       2       1
## 10147       0.2          0        0.0    Dec                8      13       9
## 10223       0.2          0        0.0    Nov                1       1       1
## 10270       0.2          0        0.0    Nov                1       1       3
## 10573       0.2          0        0.0    Nov                2       2       3
## 10632       0.2          0        0.0    Nov                2       2       1
## 10752       0.2          0        0.0    Dec                1       1       1
## 10796       0.2          0        0.0    Nov                1       1       4
## 10842       0.2          0        0.0    Nov                2       2       3
## 10989       0.2          0        0.0    Nov                2       4       3
## 11044       0.2          0        0.0    Dec                3       2       6
## 11206       0.2          0        0.0    Dec                8      13       9
## 11405       0.2          0        0.0    Nov                3       2       1
## 11524       0.2          0        0.0    Dec                2       2       1
## 11582       0.2          0        0.0    Dec                8      13       9
## 11625       0.2          0        0.0    Nov                3       2       1
## 11659       0.2          0        0.0    Dec                1       1       1
## 11734       0.2          0        0.0    Nov                2       2       1
## 11748       0.2          0        0.0    Nov                1       1       3
## 11802       0.2          0        0.0    Dec                1       1       4
## 11814       0.2          0        0.0    Dec                2       2       1
## 11828       0.2          0        0.0    Dec                2       2       1
## 11935       0.2          0        0.0    Dec                1       1       1
## 11939       0.2          0        0.0    Dec                1       1       4
## 12160       0.2          0        0.0    Dec                1       1       1
## 12181       0.2          0        0.0    Dec                1      13       9
## 12186       0.2          0        0.0    Dec                8      13       9
##       TrafficType       VisitorType Weekend Revenue
```

```
## 159             3 Returning_Visitor   FALSE   FALSE
## 179             3 Returning_Visitor   FALSE   FALSE
## 419             1 Returning_Visitor    TRUE   FALSE
## 457             1 Returning_Visitor   FALSE   FALSE
## 484             1 Returning_Visitor   FALSE   FALSE
## 513             1 Returning_Visitor   FALSE   FALSE
## 555             1 Returning_Visitor   FALSE   FALSE
## 590             1 Returning_Visitor   FALSE   FALSE
## 660             1 Returning_Visitor   FALSE   FALSE
## 775             1 Returning_Visitor   FALSE   FALSE
## 873             1 Returning_Visitor   FALSE   FALSE
## 890             1 Returning_Visitor   FALSE   FALSE
## 923             1 Returning_Visitor   FALSE   FALSE
## 948             1 Returning_Visitor   FALSE   FALSE
## 975             1 Returning_Visitor   FALSE   FALSE
## 1035            1 Returning_Visitor   FALSE   FALSE
## 1120            1 Returning_Visitor   FALSE   FALSE
## 1171            1 Returning_Visitor   FALSE   FALSE
## 1177            1 Returning_Visitor   FALSE   FALSE
## 1214            1 Returning_Visitor   FALSE   FALSE
## 1215            3 Returning_Visitor   FALSE   FALSE
## 1292            1 Returning_Visitor   FALSE   FALSE
## 1326            3 Returning_Visitor   FALSE   FALSE
## 1357            1 Returning_Visitor   FALSE   FALSE
## 1367            1 Returning_Visitor   FALSE   FALSE
## 1382            1 Returning_Visitor   FALSE   FALSE
## 1391            1 Returning_Visitor   FALSE   FALSE
## 1395            1 Returning_Visitor   FALSE   FALSE
## 1437            1 Returning_Visitor   FALSE   FALSE
## 1454            1 Returning_Visitor   FALSE   FALSE
## 1516            3 Returning_Visitor    TRUE   FALSE
## 1574            1 Returning_Visitor   FALSE   FALSE
## 1609            1 Returning_Visitor   FALSE   FALSE
## 1698            1 Returning_Visitor   FALSE   FALSE
## 1776            1 Returning_Visitor   FALSE   FALSE
## 1805            1 Returning_Visitor   FALSE   FALSE
## 1840            3 Returning_Visitor   FALSE   FALSE
## 1867            9 Returning_Visitor    TRUE   FALSE
## 1926            1 Returning_Visitor   FALSE   FALSE
## 1934            1 Returning_Visitor   FALSE   FALSE
## 1950            1 Returning_Visitor   FALSE   FALSE
## 2057            1 Returning_Visitor   FALSE   FALSE
## 2058            1 Returning_Visitor   FALSE   FALSE
## 2236            3 Returning_Visitor   FALSE   FALSE
## 2622            3 Returning_Visitor   FALSE   FALSE
## 2740            1 Returning_Visitor   FALSE   FALSE
## 3232            3 Returning_Visitor   FALSE   FALSE
## 3273            3 Returning_Visitor   FALSE   FALSE
## 3282            3 Returning_Visitor   FALSE   FALSE
## 3578            4 Returning_Visitor   FALSE   FALSE
## 3651            1 Returning_Visitor   FALSE   FALSE
## 3664            3 Returning_Visitor   FALSE   FALSE
## 3722            3 Returning_Visitor   FALSE   FALSE
## 3892            4 Returning_Visitor   FALSE   FALSE
```

```
## 4164            3 Returning_Visitor   FALSE   FALSE
## 4183            3 Returning_Visitor   FALSE   FALSE
## 4232            1 Returning_Visitor   FALSE   FALSE
## 4344           13 Returning_Visitor   FALSE   FALSE
## 4375            3 Returning_Visitor   FALSE   FALSE
## 4404            3 Returning_Visitor   FALSE   FALSE
## 4427            3 Returning_Visitor   FALSE   FALSE
## 4464            3 Returning_Visitor   FALSE   FALSE
## 4490            3 Returning_Visitor   FALSE   FALSE
## 4553            3 Returning_Visitor   FALSE   FALSE
## 4818            3 Returning_Visitor   FALSE   FALSE
## 4884            3 Returning_Visitor   FALSE   FALSE
## 4914            1 Returning_Visitor   FALSE   FALSE
## 5039            3 Returning_Visitor   FALSE   FALSE
## 5044            3 Returning_Visitor   FALSE   FALSE
## 5057            3 Returning_Visitor   FALSE   FALSE
## 5119            4 Returning_Visitor    TRUE   FALSE
## 5199           13 Returning_Visitor   FALSE   FALSE
## 5200            3 Returning_Visitor   FALSE   FALSE
## 5255            1 Returning_Visitor   FALSE   FALSE
## 5277           13 Returning_Visitor   FALSE   FALSE
## 5287           15 Returning_Visitor   FALSE   FALSE
## 5356            3 Returning_Visitor   FALSE   FALSE
## 5408            6 Returning_Visitor   FALSE   FALSE
## 6930            1 Returning_Visitor   FALSE   FALSE
## 7152            1 Returning_Visitor   FALSE   FALSE
## 7636           13 Returning_Visitor   FALSE   FALSE
## 8545            3 Returning_Visitor   FALSE   FALSE
## 9307            1 Returning_Visitor    TRUE   FALSE
## 9495            3 Returning_Visitor   FALSE   FALSE
## 9552            3 Returning_Visitor   FALSE   FALSE
## 9569            1 Returning_Visitor   FALSE   FALSE
## 9582            1 Returning_Visitor   FALSE   FALSE
## 9719           13 Returning_Visitor   FALSE   FALSE
## 9770            1 Returning_Visitor   FALSE   FALSE
## 9879           13 Returning_Visitor   FALSE   FALSE
## 9908           13 Returning_Visitor   FALSE   FALSE
## 10147          20             Other   FALSE   FALSE
## 10223           1 Returning_Visitor   FALSE   FALSE
## 10270           2 Returning_Visitor   FALSE   FALSE
## 10573           1 Returning_Visitor   FALSE   FALSE
## 10632           1 Returning_Visitor   FALSE   FALSE
## 10752           1 Returning_Visitor    TRUE   FALSE
## 10796           1 Returning_Visitor   FALSE   FALSE
## 10842           1 Returning_Visitor   FALSE   FALSE
## 10989           3 Returning_Visitor   FALSE   FALSE
## 11044           1 Returning_Visitor   FALSE   FALSE
## 11206          20             Other   FALSE   FALSE
## 11405          13 Returning_Visitor   FALSE   FALSE
## 11524          13 Returning_Visitor   FALSE   FALSE
## 11582          20             Other   FALSE   FALSE
## 11625           1 Returning_Visitor   FALSE   FALSE
## 11659           1 Returning_Visitor    TRUE   FALSE
## 11734           1 Returning_Visitor   FALSE   FALSE
```

```
## 11748           3 Returning_Visitor   FALSE   FALSE
## 11802           1 Returning_Visitor    TRUE   FALSE
## 11814           1 Returning_Visitor   FALSE   FALSE
## 11828           1 Returning_Visitor   FALSE   FALSE
## 11935           2       New_Visitor   FALSE   FALSE
## 11939           1 Returning_Visitor    TRUE   FALSE
## 12160           3 Returning_Visitor   FALSE   FALSE
## 12181          20 Returning_Visitor   FALSE   FALSE
## 12186          20             Other   FALSE   FALSE
```

EDA

Implementing the Solution