

Online Shoppers Purchasing Intention

Tatiksha Singh

BANA 200

Goal

Identifying customer online purchasing behavior based on various attributes

Determining types of customers based on pages visited and bounce rates – Are customers less likely to have a high bounce rate if they have spent a considerable amount of time on the site?

Dependent Variable: Page Value (\$)

Independent Variables:

- Administrative Pages Visited
 - Administrative Duration
- Informational Pages Visited
 - Informational Duration
- Product Related Pages Visited
 - Product Related Duration
 - Bounce Rates
 - Exit Rates

New Independent Variables:

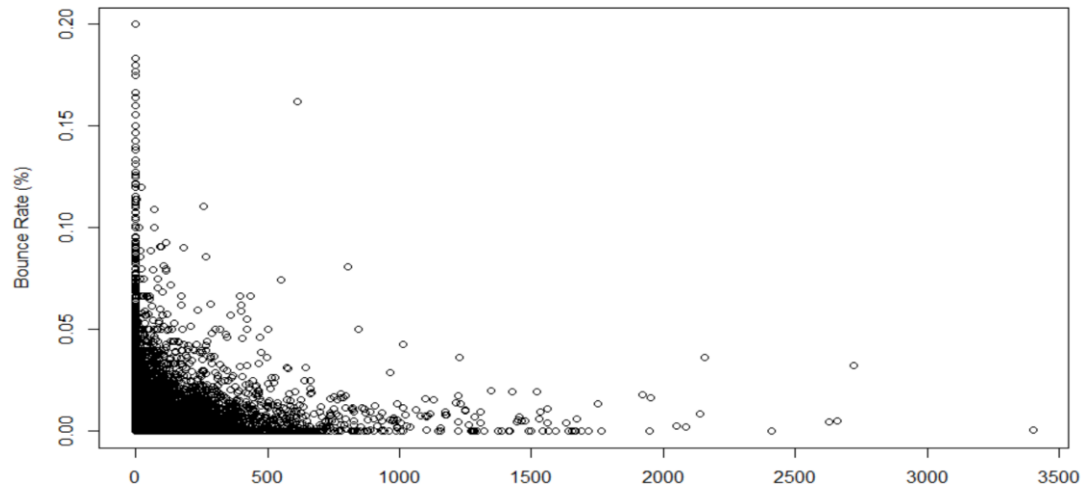
- Average Session Duration
- Average Different Page Visited
- Minutes Spent on Product Related Page

Information about the Dataset

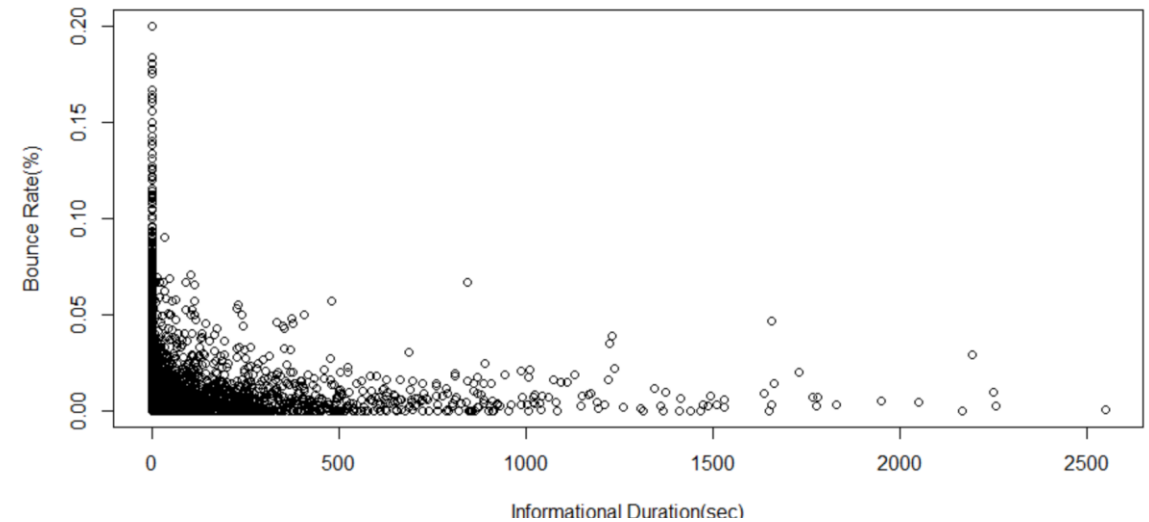
- The dataset consists of 10 numerical and 8 categorical attributes
- "Administrative", "Administrative Duration", "Informational", "Informational Duration", "Product Related" and "Product Related Duration"
 - These all represent the no of diff types of pages visited by the visitor in that session and total time spent in each of these page categories
- "Bounce Rate", "Exit Rate" and "Page Value"
 - These represent metrics measured by "Google Analytics" for each page in the e-commerce site
- Bounce rate= % of visitors who enter the site from that page and then leave ("bounce") without triggering any other requests
- "Page Value" = the avg value for a web page that a user visited before completing an e-commerce transaction

Pages Relationship with Bounce Rates

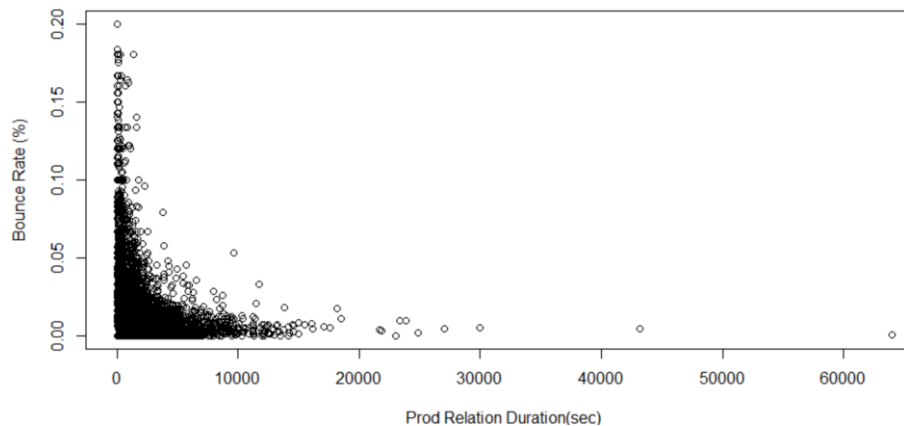
Relationship between Admin Duration and Bounce Rates



Relationship between Informational Duration and Bounce Rate

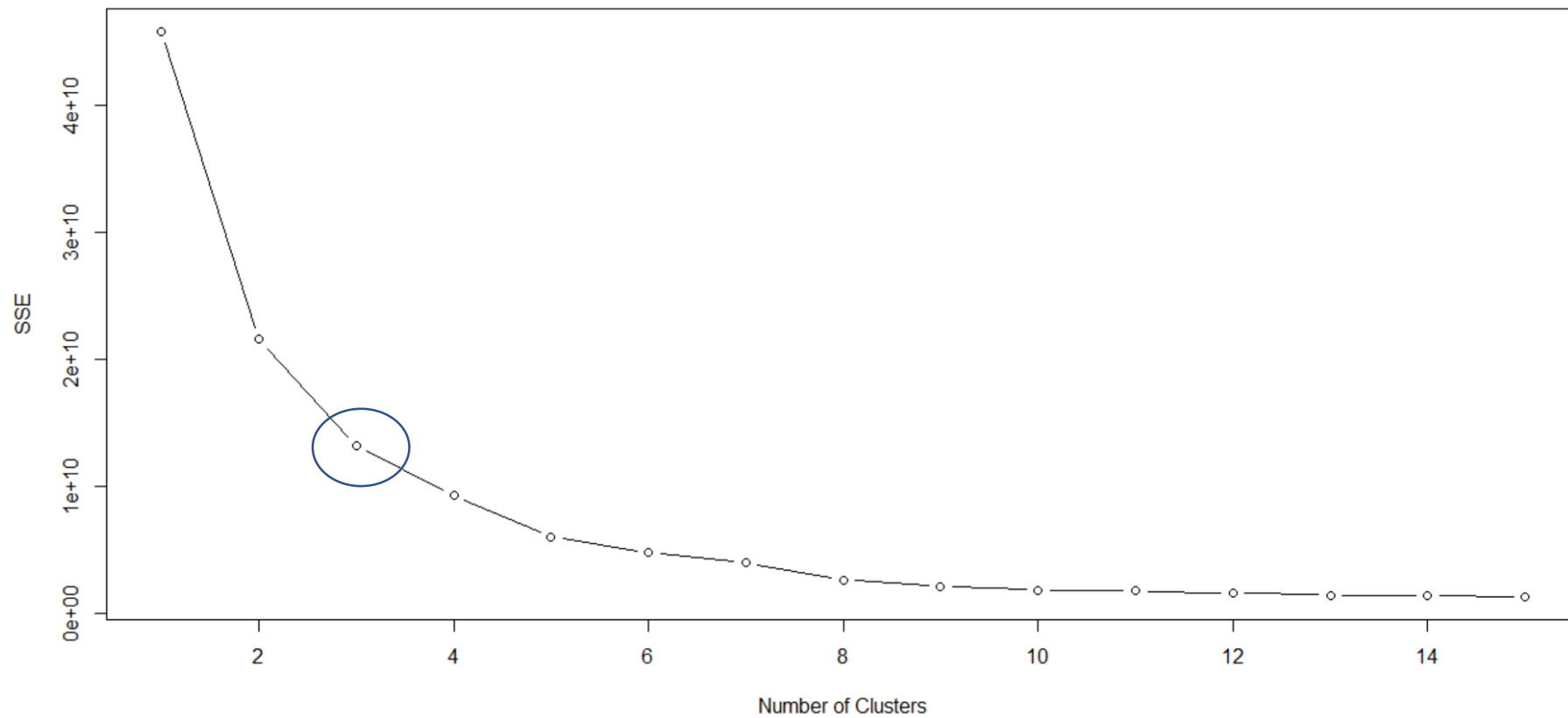


Relationship between Prod Related Duration and Bounce Rate

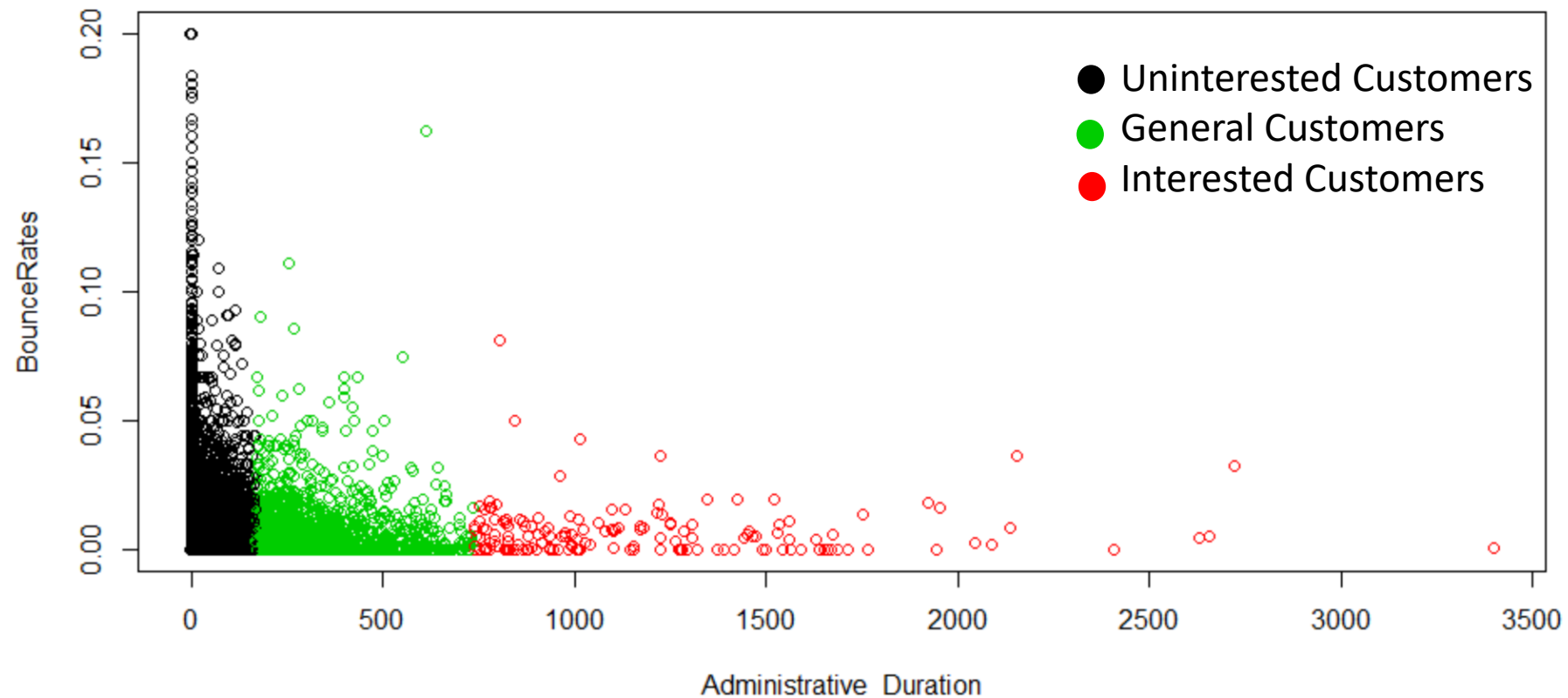


All three graphs demonstrate similar behavior where customers who generally spent more time on a page were less likely to leave the website from the landing page without browsing further

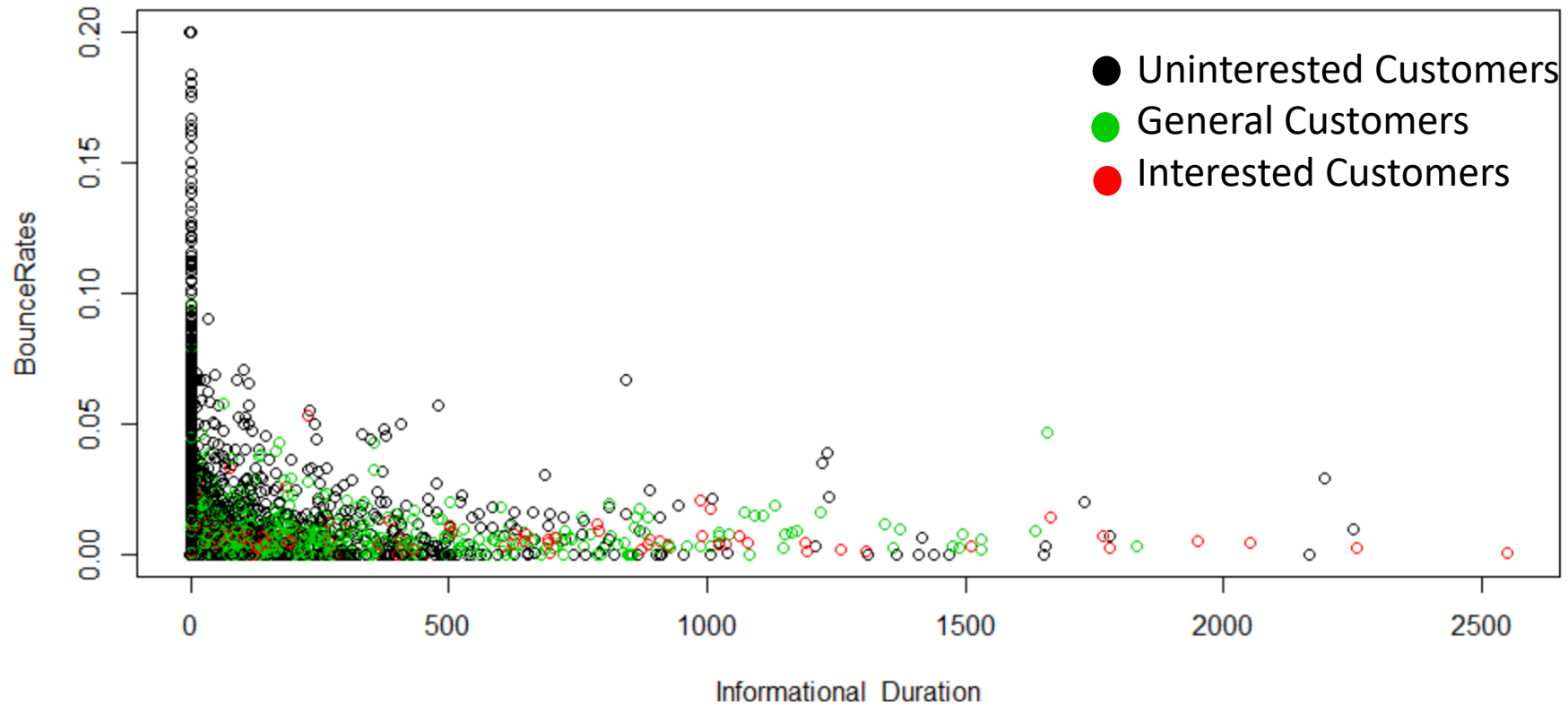
Determining Number of K Clusters (Elbow curve)



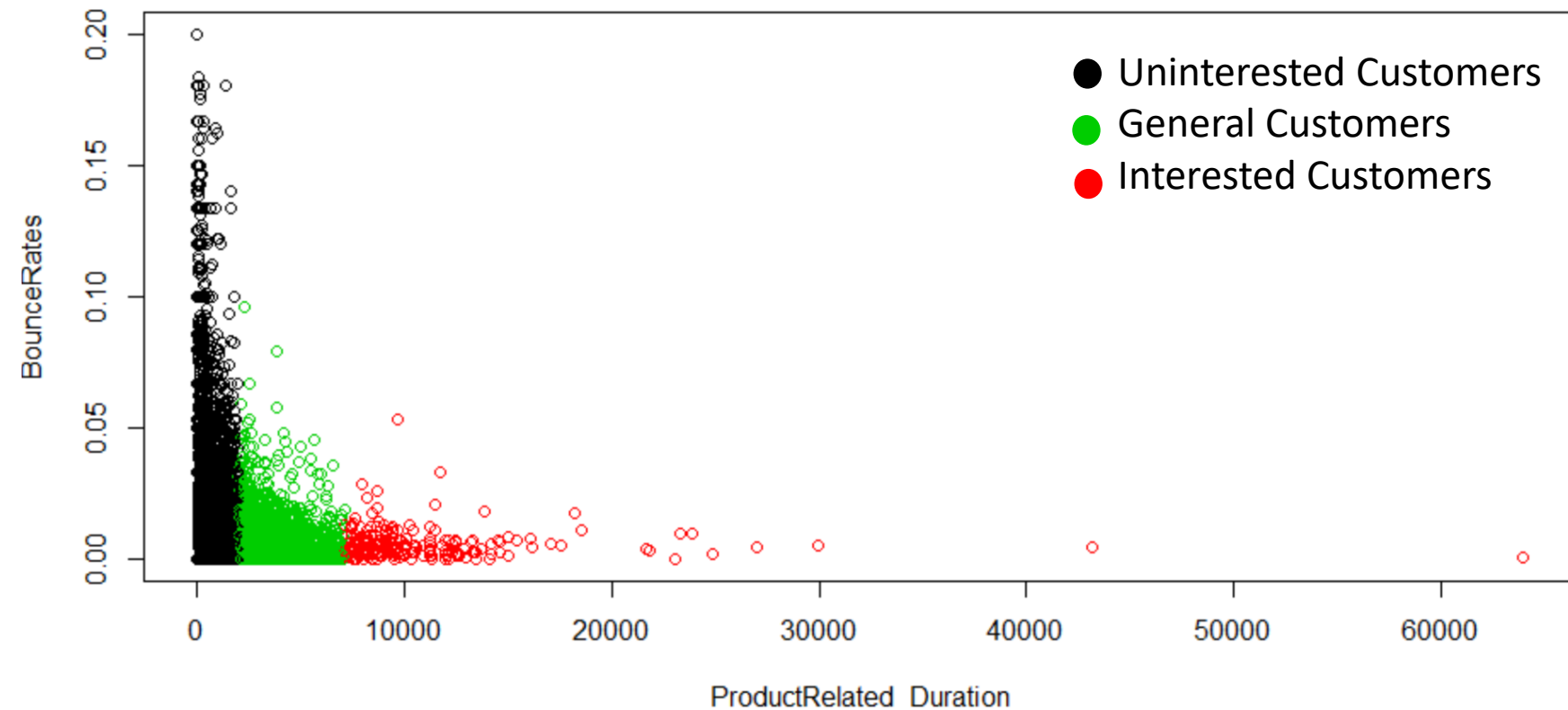
Types of Customers Clustering - Administrative Page



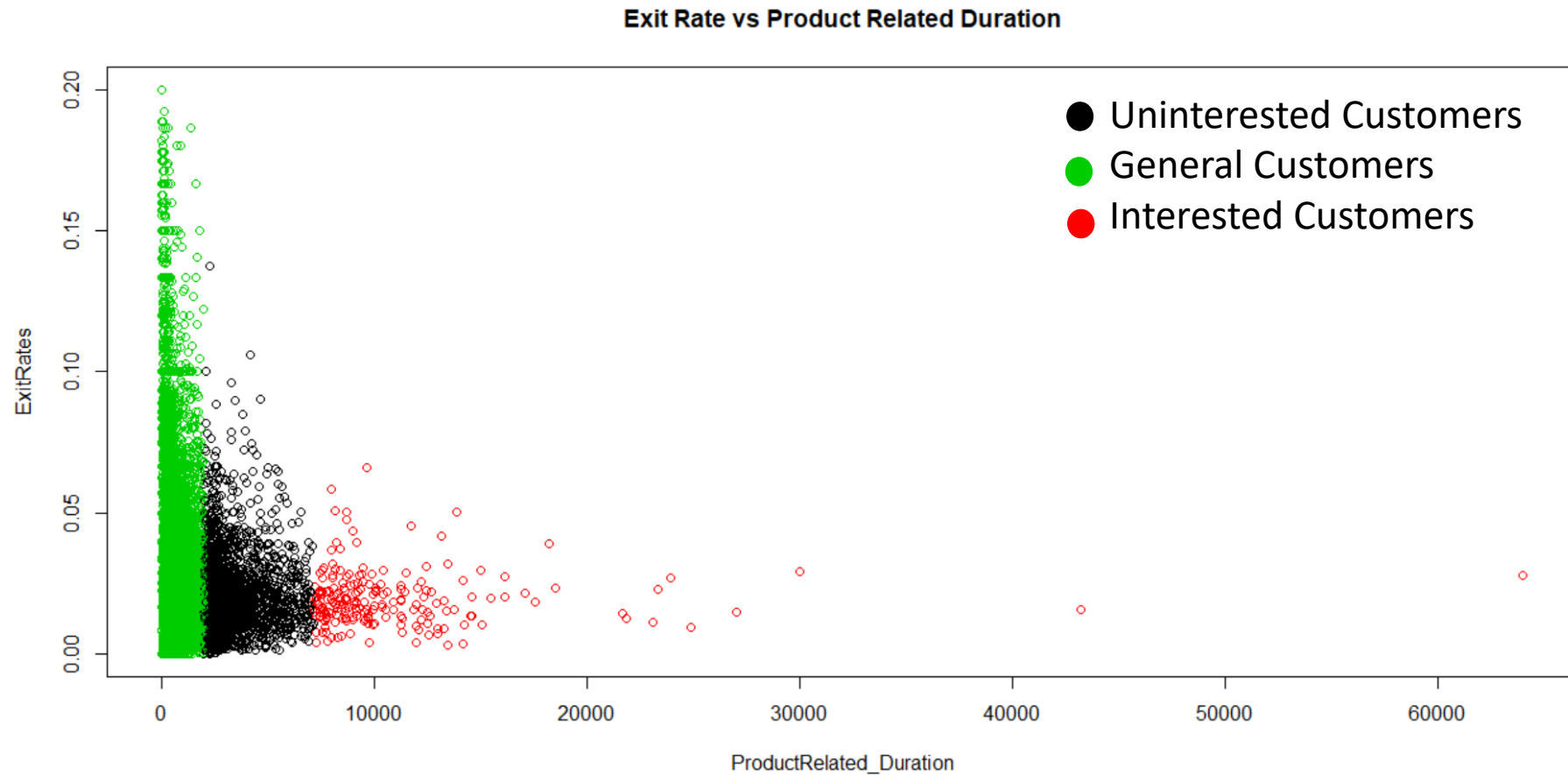
Types of Customers Clustering - Informational Page



Types of Customers Clustering- Product Related Page



Types of Customers Clustering -Exit Rate



1st New Independent Variable – Average Session Duration

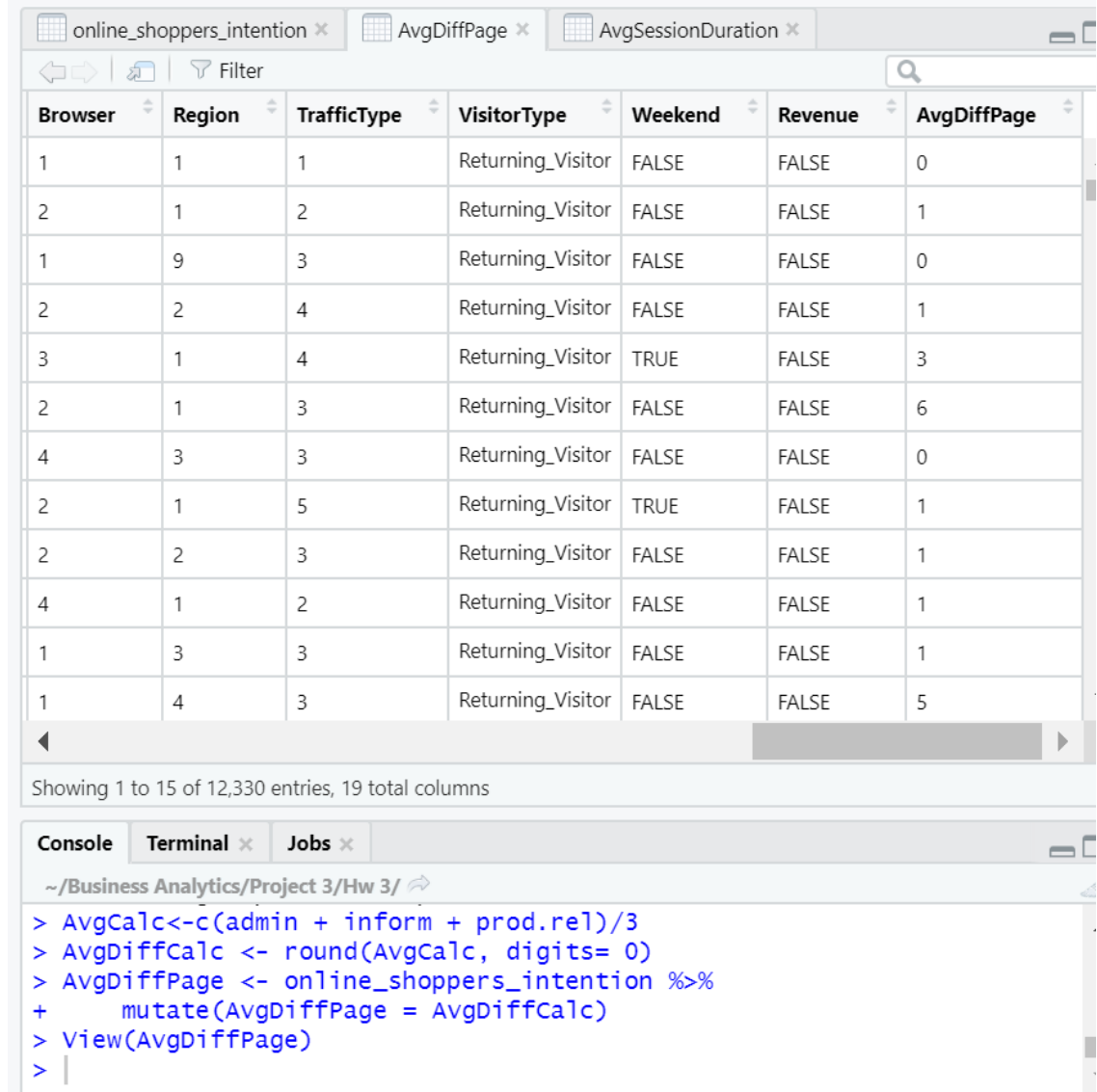
```
> Avg_dur_calc <- c(admin_dur + inform_dur + prod.rel_dur)/3
> Avg_dur_calc
 [1] 0.0000000 21.3333333 -1.0000000 0.8888889
 [5] 209.1666667 51.4055556 -1.0000000 -1.0000000
 [9] 12.3333333 246.0000000 131.6666667 135.9166667
[13] 93.5000000 32.6666667 22.6666667 573.7617063
[17] -1.0000000 111.6555556 10.6666667 993.7222223
[21] 45.3888889 -1.0000000 35.0000000 5.0000000
[25] -1.0000000 52.0000000 400.0148147 25.3333333
[29] 21.0000000 529.5833333 11.6666667 26.0000000
[33] 69.8333333 61.2222222 126.8333333 695.4142857
```

The screenshot shows the RStudio interface. At the top, there are two tabs: 'online_shoppers_intention' and 'AvgSessionDuration'. Below the tabs is a table with 7 columns: 'er', 'Region', 'TrafficType', 'VisitorType', 'Weekend', 'Revenue', and 'AvgSessionDuration'. The table displays 9 rows of data. Below the table, it says 'Showing 1 to 9 of 12,330 entries, 19 total columns'. At the bottom, there is a 'Console' tab with the following output:

```
~/Business Analytics/Project 3/Hw 3/
# Browser <dbl>, Region <dbl>, TrafficType <dbl>,
# VisitorType <chr>, Weekend <lgl>, Revenue <lgl>,
# AverageSessionDur <dbl>
> online_shoppers_intention %>%
+
> AvgSessionDuration <- online_shoppers_intention %>% mutate(AvgSessionDuration= Avg_dur_calc)
> AvgSessionDuration
# A tibble: 12,330 x 19
  Administrative Administrative_~ Informational Informational_D~
      <dbl>      <dbl>      <dbl>      <dbl>
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
```

2nd New Independent Variable – Average Different Page Visited

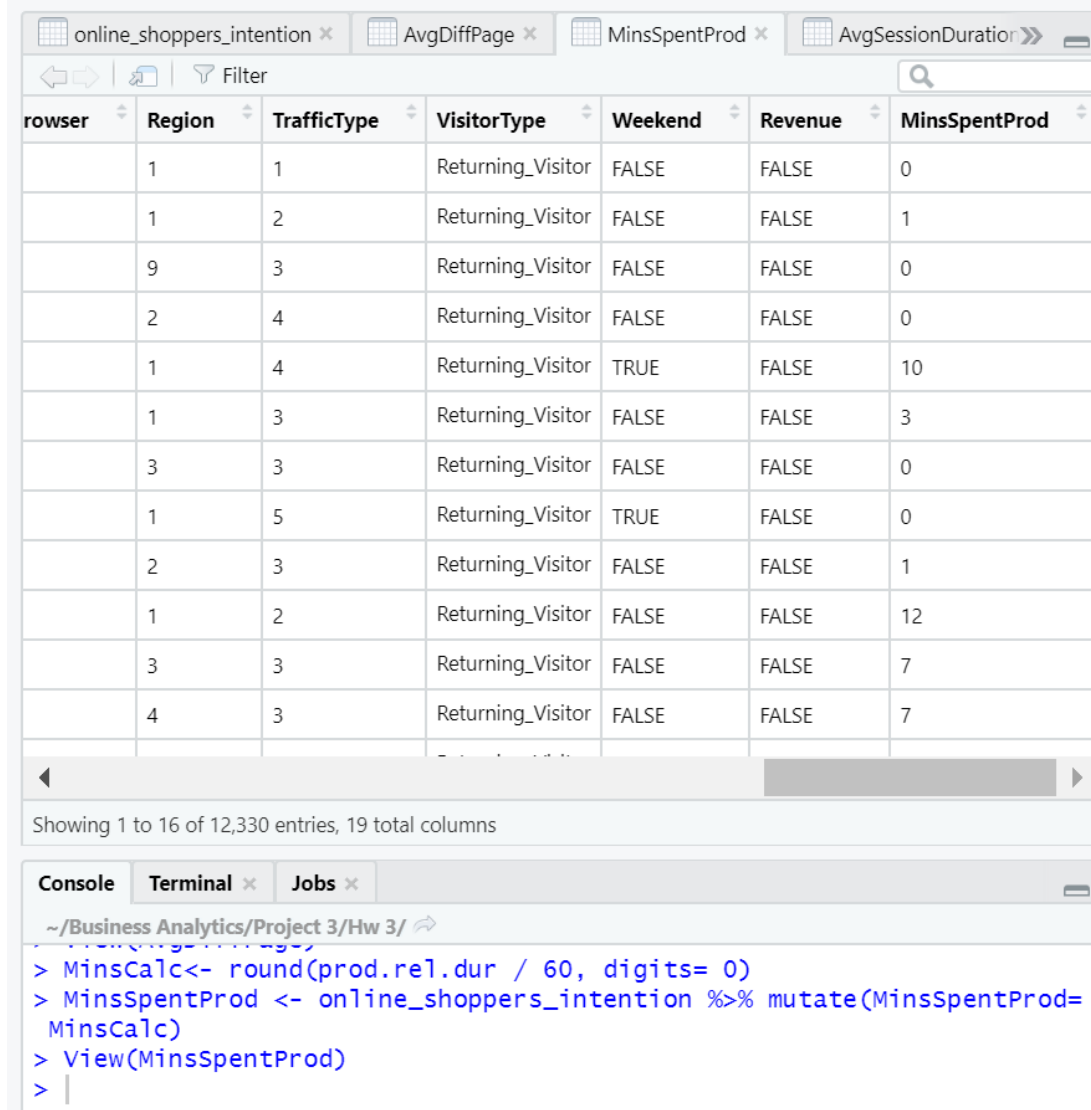
```
> AvgCalc<-c(admin + inform + prod.rel)/3  
> AvgDiffCalc <- round(AvgCalc, digits= 0)  
> AvgDiffPage <- online_shoppers_intention %>%  
+   mutate(AvgDiffPage = AvgDiffCalc)  
> |
```



The screenshot displays the RStudio interface. At the top, three tabs are open: 'online_shoppers_intention', 'AvgDiffPage', and 'AvgSessionDuration'. Below the tabs is a table with 7 columns: Browser, Region, TrafficType, VisitorType, Weekend, Revenue, and AvgDiffPage. The table contains 15 rows of data. At the bottom, the 'Console' tab is active, showing the R code that was executed, which matches the code in the left block. The console path is '~ /Business Analytics/Project 3/Hw 3/'. The table data is as follows:

Browser	Region	TrafficType	VisitorType	Weekend	Revenue	AvgDiffPage
1	1	1	Returning_Visitor	FALSE	FALSE	0
2	1	2	Returning_Visitor	FALSE	FALSE	1
1	9	3	Returning_Visitor	FALSE	FALSE	0
2	2	4	Returning_Visitor	FALSE	FALSE	1
3	1	4	Returning_Visitor	TRUE	FALSE	3
2	1	3	Returning_Visitor	FALSE	FALSE	6
4	3	3	Returning_Visitor	FALSE	FALSE	0
2	1	5	Returning_Visitor	TRUE	FALSE	1
2	2	3	Returning_Visitor	FALSE	FALSE	1
4	1	2	Returning_Visitor	FALSE	FALSE	1
1	3	3	Returning_Visitor	FALSE	FALSE	1
1	4	3	Returning_Visitor	FALSE	FALSE	5

3rd New Independent Variable – Minutes Spent on Product Duration Page



The screenshot displays a data table with 19 columns and 12,330 entries. The columns are: rowser, Region, TrafficType, VisitorType, Weekend, Revenue, and MinsSpentProd. The table shows data for various regions and traffic types, with 'MinsSpentProd' values ranging from 0 to 12. Below the table, the R console shows the following commands:

```
> MinsCalc<- round(prod.rel.dur / 60, digits= 0)
> MinsSpentProd <- online_shoppers_intention %>% mutate(MinsSpentProd=
  MinsCalc)
> View(MinsSpentProd)
>
```

rowser	Region	TrafficType	VisitorType	Weekend	Revenue	MinsSpentProd
	1	1	Returning_Visitor	FALSE	FALSE	0
	1	2	Returning_Visitor	FALSE	FALSE	1
	9	3	Returning_Visitor	FALSE	FALSE	0
	2	4	Returning_Visitor	FALSE	FALSE	0
	1	4	Returning_Visitor	TRUE	FALSE	10
	1	3	Returning_Visitor	FALSE	FALSE	3
	3	3	Returning_Visitor	FALSE	FALSE	0
	1	5	Returning_Visitor	TRUE	FALSE	0
	2	3	Returning_Visitor	FALSE	FALSE	1
	1	2	Returning_Visitor	FALSE	FALSE	12
	3	3	Returning_Visitor	FALSE	FALSE	7
	4	3	Returning_Visitor	FALSE	FALSE	7

Cluster Analysis

```
[1] "cluster"      "centers"      "totss"        "withinss"
[5] "tot.withinss" "betweenss"    "size"         "iter"
[9] "ifault"
```

```
> Cluster <- subset(ecommerce.data, select=c(1:9))
> kmeans(Cluster, 3)
```

K-means clustering with 3 clusters of sizes 10134, 207, 1975

Cluster means:

	Administrative	Administrative_Duration	Informational
1	1.821985	62.10943	0.3275113
2	7.439614	295.36489	2.7487923
3	4.325063	154.87748	1.1741772

	Informational_Duration	ProductRelated	ProductRelated_Duration
1	19.98624	18.63173	577.3255
2	266.72309	236.10628	10886.1744
3	84.67262	77.72962	3355.1075

	BounceRates	ExitRates	PageValues
1	0.025411677	0.04763612	5.511318
2	0.005939756	0.01968831	4.521370
3	0.007128234	0.02167056	8.013634

Clustering vector:

```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1
[32] 1 1 1 1 3 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[63] 1 1 1 1 3 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[94] 1 1 1 1 1 1 1 1 3 1 1 3 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[125] 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[156] 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
[187] 1 3 1 3 1 3 1 1 1 3 3 3 1 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[218] 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1
[249] 1 3 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 3 1 1
```

In storage.mode(x) <- "double" : NAs introduced by coercion

```
> Cluster <- subset(ecommerce.data, select=c(1:9))
```

```
> Results <- kmeans(Cluster, 3)
```

```
> Results
```

```
[1] "cluster"      "centers"      "totss"        "withinss"    "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"       "
```

```
> Results$size
```

```
[1] 10134 207 1975
```

```
> Results$cluster
```

Cluster Analysis

ion	ProductRelated	ProductRelated_Duration	BounceRates	ExitRates	PageValues	segment
1	1	0.000000	0.200000000	0.200000000	0.00000	1
2	2	64.000000	0.000000000	0.100000000	0.00000	1
1	1	-1.000000	0.200000000	0.200000000	0.00000	1
2	2	2.666667	0.050000000	0.140000000	0.00000	1
10	10	627.500000	0.020000000	0.050000000	0.00000	1
19	19	154.216667	0.015789474	0.024561404	0.00000	1
1	1	-1.000000	0.200000000	0.200000000	0.00000	1
1	1	-1.000000	0.200000000	0.200000000	0.00000	1

Showing 1 to 11 of 12,316 entries, 11 total columns

Console

Terminal

Jobs

```
~/Business Analytics/Project 3/Hw 3/
Unknown or uninitialised column: 'class'.
> segment<-Results$cluster
> Cluster <- cbind(Cluster, segment)
> Cluster <- cbind(Cluster, Clustersegment)
Error in cbind(Cluster, Clustersegment) :
  object 'Clustersegment' not found
> Cluster <- cbind(Cluster, segment)
> -c(Cluster)
Error in -c(Cluster) : invalid argument to unary operator
> df = subset(Cluster, select = -c(11) )
> Cluster
  Administrative Administrative_Duration Informational Informational_Duration
1              0              0.00000              0              0
2              0              0.00000              0              0
```

Creating New Data Frames based on Clusters

```
> Seg1<- Cluster %>% filter(Cluster$segment ==0)
> Seg1<- Cluster %>% filter(Cluster$segment ==1)
> View(Seg1)
> df1 <- data.frame(Seg1)
> View(df1)
> Seg2 <- Cluster %>% filter(Cluster$segment ==2)
> df2 <- data.frame(Seg2)
> Seg3 <- Cluster %>% filter(Cluster$segment ==3)
> df3 <- data.frame(Seg3)
Error in data.frame(Seg3) : could not find function "data.frame"
> df3 <- data.frame(Seg3)
```

ditional_Duration	ProductRelated	ProductRelated_Duration	BounceRates	ExitRates	PageValues	segment
	1	0.000000	0.200000000	0.200000000	0.00000	1
	2	64.000000	0.000000000	0.100000000	0.00000	1
	1	-1.000000	0.200000000	0.200000000	0.00000	1
	2	2.666667	0.050000000	0.140000000	0.00000	1
	10	627.500000	0.020000000	0.050000000	0.00000	1
	19	154.216667	0.015789474	0.024561404	0.00000	1
	1	-1.000000	0.200000000	0.200000000	0.00000	1
	1	-1.000000	0.200000000	0.200000000	0.00000	1

Data Frame 1- Multivariate Linear Regression Analysis

Call:

```
lm(formula = df1$PageValues ~ df1$Administrative + df1$Administrative_Duration +  
    df1$Informational + df1$Informational_Duration + df1$ProductRelated_Duration +  
    df1$ProductRelated + df1$BounceRates + df1$ExitRates)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.31	-7.41	-4.79	1.21	353.46

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	7.436e+00	4.931e-01	15.080	< 2e-16	***
df1\$Administrative	9.431e-02	8.953e-02	1.053	0.2922	
df1\$Administrative_Duration	3.914e-04	1.633e-03	0.240	0.8107	
df1\$Informational	-3.100e-01	2.449e-01	-1.266	0.2057	
df1\$Informational_Duration	1.068e-03	2.269e-03	0.471	0.6380	
df1\$ProductRelated_Duration	4.019e-03	5.320e-04	7.555	4.54e-14	***
df1\$ProductRelated	-3.545e-02	1.622e-02	-2.186	0.0288	*
df1\$BounceRates	7.319e+01	8.833e+00	8.285	< 2e-16	***
df1\$ExitRates	-1.167e+02	9.587e+00	-12.175	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.28 on 10125 degrees of freedom

Multiple R-squared: 0.04697, Adjusted R-squared: 0.04622

F-statistic: 62.38 on 8 and 10125 DF, p-value: < 2.2e-16

Multiple Linear Regression Equation:

$$\text{Page Value (\$)} = 7.436e+00 + 9.431e-02(x_1) + 3.914e-04(x_2) + (-3.100e-01(x_3)) + 1.068e-03(x_4) + 4.019e-03(x_5) + (-3.545e-02(x_6)) + 7.319e+01(x_7) + (-1.167e+02(x_8))$$

Data Frame 2- Multivariate Linear Regression Analysis

Call:

```
lm(formula = df2$PageValues ~ df2$Administrative + df2$Administrative_Duration +  
  df2$Informational + df2$Informational_Duration + df2$ProductRelated_Duration +  
  df2$ProductRelated + df2$BounceRates + df2$ExitRates)
```

Residuals:

Min	1Q	Median	3Q	Max
-9.745	-4.294	-1.988	1.485	50.509

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.021e+01	2.311e+00	4.421	1.62e-05	***
df2\$Administrative	2.955e-01	1.287e-01	2.296	0.02271	*
df2\$Administrative_Duration	-3.674e-04	2.277e-03	-0.161	0.87198	
df2\$Informational	5.233e-02	2.444e-01	0.214	0.83065	
df2\$Informational_Duration	-8.012e-04	1.698e-03	-0.472	0.63745	
df2\$ProductRelated_Duration	1.116e-04	2.049e-04	0.544	0.58681	
df2\$ProductRelated	-1.569e-02	7.808e-03	-2.009	0.04589	*
df2\$BounceRates	8.964e+01	1.610e+02	0.557	0.57829	
df2\$ExitRates	-2.924e+02	1.057e+02	-2.767	0.00619	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.321 on 198 degrees of freedom

Multiple R-squared: 0.1082, Adjusted R-squared: 0.07212

F-statistic: 3.001 on 8 and 198 DF, p-value: 0.003362

Data Frame 3- Multivariate Linear Regression Analysis

Call:

```
lm(formula = df3$pageValues ~ df3$Administrative + df3$Administrative_Duration +  
    df3$Informational + df3$Informational_Duration + df3$ProductRelated_Duration +  
    df3$ProductRelated + df3$BounceRates + df3$ExitRates)
```

Residuals:

Min	1Q	Median	3Q	Max
-18.059	-8.831	-4.588	1.476	232.038

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.633e+01	1.422e+00	11.481	< 2e-16	***
df3\$Administrative	3.042e-01	1.200e-01	2.535	0.011312	*
df3\$Administrative_Duration	-2.734e-04	1.988e-03	-0.138	0.890616	
df3\$Informational	6.758e-01	2.850e-01	2.371	0.017830	*
df3\$Informational_Duration	-1.836e-03	2.421e-03	-0.758	0.448332	
df3\$ProductRelated_Duration	1.036e-04	4.021e-04	0.258	0.796615	
df3\$ProductRelated	-4.608e-02	1.184e-02	-3.893	0.000102	***
df3\$BounceRates	1.234e+02	7.525e+01	1.640	0.101074	
df3\$ExitRates	-3.632e+02	4.829e+01	-7.520	8.29e-14	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.91 on 1966 degrees of freedom

Multiple R-squared: 0.06636, Adjusted R-squared: 0.06256

F-statistic: 17.47 on 8 and 1966 DF, p-value: < 2.2e-16

Key Insights

- All three graphs on Page 5 demonstrate similar behavior where customers who generally spent more time on a page were less likely to leave the website from the landing page without browsing further
 - There seems to be a slight negative relationship
- Based on the clustering graphs on Pages 6-8, customers can be categorized into 3 groups:
 - Interested Customers → these customers stayed on the administrative page and product related pages for a considerable amount of time. Moreover, these customers are less likely to bounce from a page, thus they should be targeted more heavily
 - General Customers
 - Unsatisfied Customers → these customers showed little to no interest as they left the site early on. These types of customers require more analysis as to why they were dissatisfied with their visit
- The cluster analysis revealed the total amount of observations to be categorized into 3 clusters. The first cluster had a considerably larger group of 10,134 observations, whereas the second cluster had 207 and the third cluster has 1975
 - The multivariate linear regression also revealed significant p-values with certain variables which could ascertain that users completed an eCommerce transaction based on spending considerable time on product related and administrative pages.
 - On Google Analytics, the company can assign a monetary goal value (e.g. \$10) to a certain page, this allows one to identify which pages are most valuable for the business
 - One can then promote low traffic pages and high page values to increase conversion rates

Appendix

```
> ecommerce.data <- na.omit(online_shoppers_intention)
```

```
> complete.cases(ecommerce.data)
```

```
 [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[16] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[31] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[46] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[61] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[76] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[91] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[106] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[121] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[136] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
[946] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[961] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[976] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[991] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[ reached getOption("max.print") -- omitted 11316 entries ]
```

Showing 1 to 9 of 12,330 entries, 18 total columns

Console Terminal x Jobs x

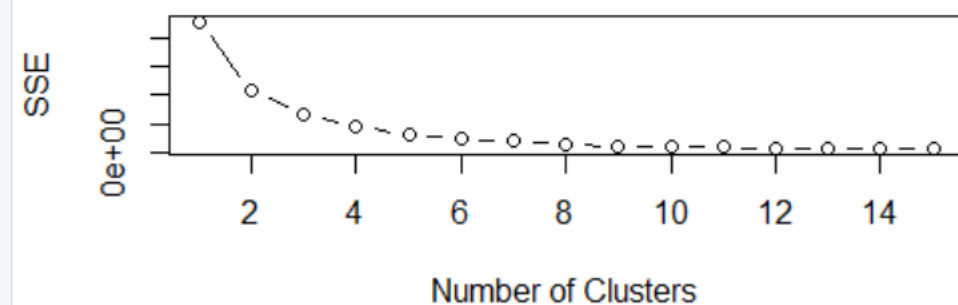
~/Business Analytics/Project 3/Hw 3/ ↗

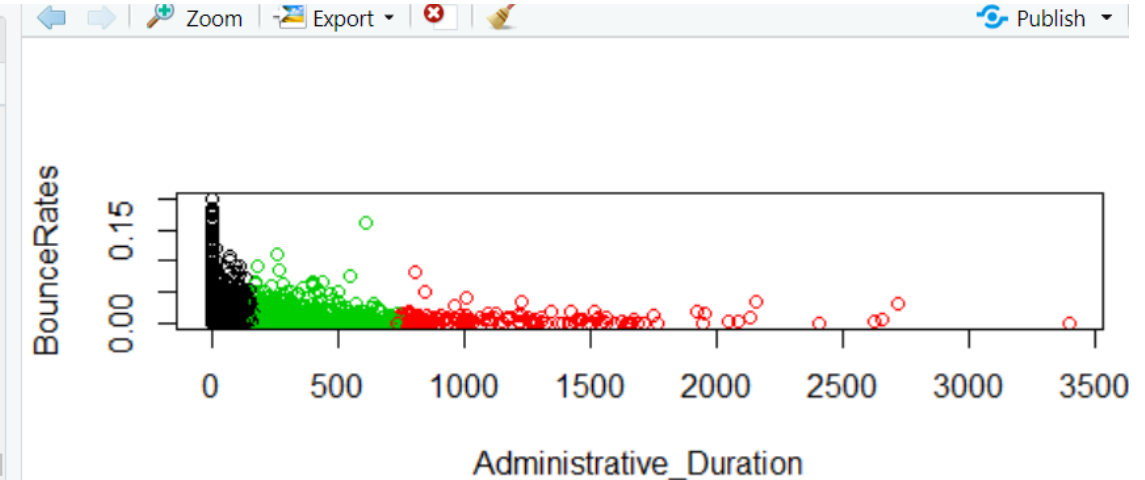
```
> SSE_curve <- c()
> for (n in 1:15) {
+   kcluster <- kmeans(clustering_data, n)
+   sse <- sum(kcluster$withinss)
+   SSE_curve[n] <- sse
+ }
Warning message:
Quick-TRANSfer stage steps exceeded maximum (= 615800)
> SSE_curve
[1] 45766279347 21527901668 13201082032 9258262888 6018976135
[6] 4754384942 3968717617 2597257590 2129735719 1817149065
[11] 1739057254 1559089806 1419443381 1399143085 1303039687
> print("SSE curve for the ideal k value")
[1] "SSE curve for the ideal k value"
> plot(1:15, SSE_curve, type="b", xlab="Number of Clusters", ylab="SSE")
> |
```

```
print("SSE curve for the ideal k value")
plot(1:15, SSE_curve, type="b", xlab="Number of Clusters", ylab="SSE")
```

Files Plots Packages Help Viewer

Zoom Export Publish





```
+ df1$Informational + df1$Informational_Duration + df1$ProductRelated_Duration + df1$Pr  
oductRelated + df1$BounceRates + df1$ExitRates)  
> summary(model)
```

Call:

```
lm(formula = df1$PageValues ~ df1$Administrative + df1$Administrative_Duration +  
    df1$Informational + df1$Informational_Duration + df1$ProductRelated_Duration +  
    df1$ProductRelated + df1$BounceRates + df1$ExitRates)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.31	-7.41	-4.79	1.21	353.46

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	7.436e+00	4.931e-01	15.080	< 2e-16	***
df1\$Administrative	9.431e-02	8.953e-02	1.053	0.2922	
df1\$Administrative_Duration	3.914e-04	1.633e-03	0.240	0.8107	
df1\$Informational	-3.100e-01	2.449e-01	-1.266	0.2057	
df1\$Informational_Duration	1.068e-03	2.269e-03	0.471	0.6380	
df1\$ProductRelated_Duration	4.019e-03	5.320e-04	7.555	4.54e-14	***
df1\$ProductRelated	-3.545e-02	1.622e-02	-2.186	0.0288	*
df1\$BounceRates	7.319e+01	8.833e+00	8.285	< 2e-16	***
df1\$ExitRates	-1.167e+02	9.587e+00	-12.175	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

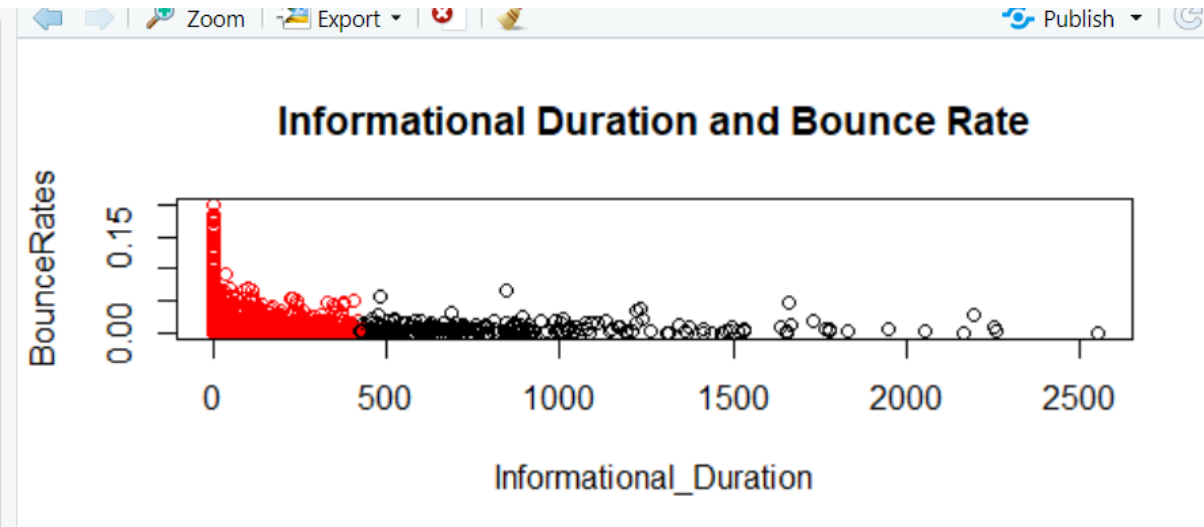
Residual standard error: 18.28 on 10125 degrees of freedom

Multiple R-squared: 0.04697, Adjusted R-squared: 0.04622

F-statistic: 62.38 on 8 and 10125 DF, p-value: < 2.2e-16

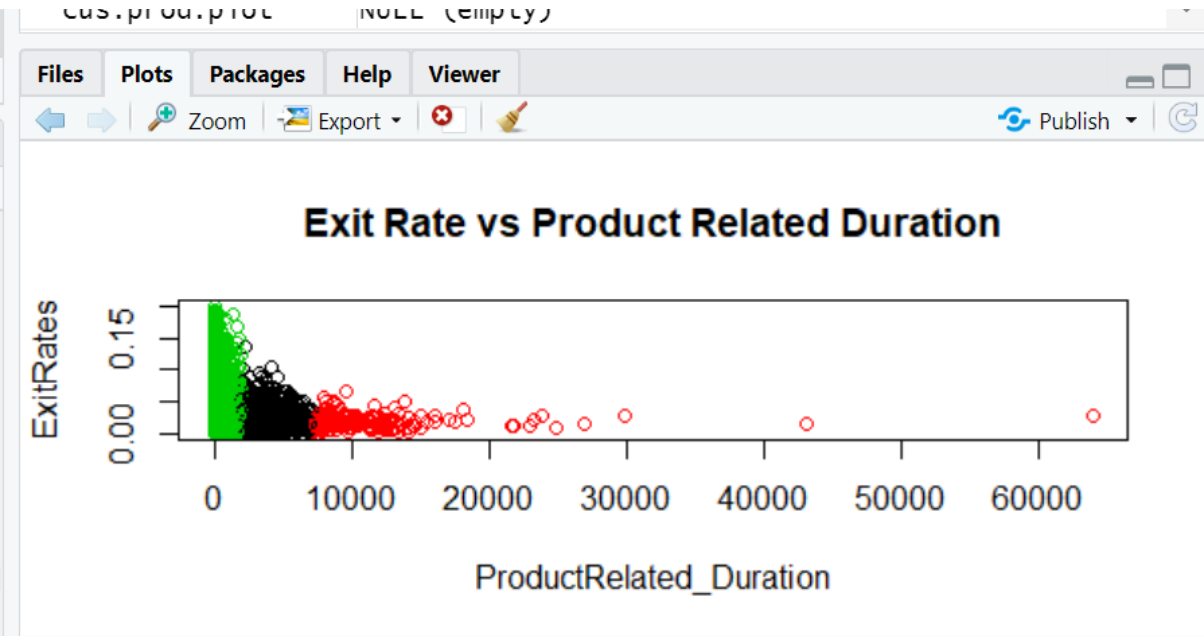
> |

```
Console Terminal x Jobs x
~/Business Analytics/Project 3/Hw 3/
> plot(ecommerce.data[c(2, 7)], col= practice.results$cluster)
Error: unexpected ')' in "plot(ecommerce.data[c(2, 7)], col= practice.results$cluster)"
> plot(ecommerce.data[c(2, 7)], col= practice.results$cluster)
> plot(ecommerce.data[c(2, 7)], col= cus$cluster)
Error in plot.xy(xy, type, ...) : object 'cus' not found
> plot(ecommerce.data[c(2, 7)], col= cus.result$cluster)
> plot(ecommerce.data[c(6, 7)], col= cus.result$cluster)
> customer.cluster <- subset(ecommerce.data, select= c(4,7))
> cus.result <- kmeans(customer.cluster, 2)
> plot(ecommerce.data[c(4,7)], col= cus.result$cluster, main= "Informational Duration and Bounce Rate")
> |
```



```
Showing 1 to 11 of 12,330 entries, 18 total columns

Console Terminal x Jobs x
~/Business Analytics/Project 3/Hw 3/
> plot(ecommerce.data[c(6, 7)], col= cus.result$cluster)
> customer.cluster <- subset(ecommerce.data, select= c(4,7))
> cus.result <- kmeans(customer.cluster, 2)
> plot(ecommerce.data[c(4,7)], col= cus.result$cluster, main= "Informational Duration and Bounce Rate")
> exit.cus <- subset(ecommerce.data, select= c(6,8))
> exit.result <- kmeans(exit.cus, 2)
> plot(ecommerce.data[c(6,8)], col= exit.result$cluster, main= "Exit Rate vs Product Related Duration")
> exit.result <- kmeans(exit.cus, 3)
> plot(ecommerce.data[c(6,8)], col= exit.result$cluster, main= "Exit Rate vs Product Related Duration")
> |
```




```
Console Terminal x Jobs x
~/Business Analytics/Project 3/Cluster/
> library(corrplot)
corrplot 0.84 loaded
> corrplot(Cluster, method="circle", type="upper")
Error in matrix(if (is.null(value)) logical() else value, nrow = nr, dimnames = list(rn,
:
length of 'dimnames' [2] not equal to array extent
> corrplot(Cluster, method="circle", type="upper", tl.cex=0.3)
Error in matrix(if (is.null(value)) logical() else value, nrow = nr, dimnames = list(rn,
:
length of 'dimnames' [2] not equal to array extent
> corrplot(df1, method="circle", tl.cex = 0.3)
Error in matrix(if (is.null(value)) logical() else value, nrow = nr, dimnames = list(rn,
:
length of 'dimnames' [2] not equal to array extent
> plot(Cluster$Administrative_Duration, Cluster$BounceRates)
> plot(Cluster$Administrative_Duration, Cluster$BounceRates, main="Relationship between
Admin Duration and Bounce Rates", xlab="Admin Duration", ylab="Bounce Rate (%)")
> plot(Cluster$Informational_Duration, Cluster$BounceRates, main="Relationship between I
nformational Duration and Bounce Rate", xlab="Informational Duration(sec)", ylab="Bounce
Rate(%)"")
> > plot(Cluster$ProductRelated_Duration, Cluster$BounceRates, main="Relationship between
n Product Related Duration and Bounce Rate", xlab="Product Related Duration(sec)", ylab
="Bounce Rate(%)", col="blue")
Error: unexpected '>' in ">"
> > plot(Cluster$ProductRelated_Duration, Cluster$BounceRates, main="Relationship between
n Product Related Duration and Bounce Rate", xlab="Product Related Duration(sec)", ylab
="Bounce Rate(%)"")
Error: unexpected '>' in ">"
> plot(Cluster$ProductRelated_Duration, Cluster$BounceRates, main="Relationship between
Prod Related Duration and Bounce Rate", xlab="Prod Relation Duration(sec)", ylab="Bounc
e Rate (%)")
> |
```

