

Module 3 - Summarize

Rsquared Academy

2021-03-05

Import Data & Load Libraries

```
analytics <- readRDS("analytics.rds")
library(forcats)
library(dplyr)
library(gmodels)
library(descriptr)
```

1. Display the number of levels in

- browser
- channel
- landing_page
- exit_page

```
nlevels(analytics$browser)
```

```
## [1] 26
```

```
nlevels(analytics$channel)
```

```
## [1] 8
```

```
nlevels(analytics$landing_page)
```

```
## [1] 33
```

```
nlevels(analytics$exit_page)
```

```
## [1] 33
```

2. Display the categories in

- os
- channel
- browser
- gender
- user_type

```
levels(analytics$os)
```

```
## [1] "(not set)"      "Android"        "BlackBerry"     "Chrome OS"
## [5] "Firefox OS"     "iOS"            "Linux"          "Macintosh"
## [9] "OS/2"           "Playstation 4"  "Playstation Vita" "Samsung"
## [13] "Tizen"          "Windows"        "Windows Phone"  "Xbox"
```

```
levels(analytics$channel)
```

```
## [1] "(Other)"      "Affiliates"    "Direct"        "Display"
## [5] "Organic Search" "Paid Search"   "Referral"      "Social"
```

```
unique(analytics$browser)
```

```
## [1] Chrome           Safari            Firefox
## [4] Samsung Internet Opera            Edge
## [7] Safari (in-app)   User-Agent:Mozilla YaBrowser
## [10] UC Browser        Android Webview   Internet Explorer
## [13] Mozilla Compatible Agent Opera Mini        Android Browser
## [16] Amazon Silk       Playstation 4     Coc Coc
## [19] Puffin            Maxthon           BlackBerry
## [22] Seznam            MRCHROME          SeaMonkey
## [25] APKPure           Playstation Vita Browser
## 26 Levels: Amazon Silk Android Browser Android Webview APKPure ... YaBrowser
```

```
unique(analytics$gender)
```

```
## [1] female <NA>    male
## Levels: female male <NA>
```

```
forcats::fct_unique(analytics$user_type)
```

```
## [1] New Visitor      Returning Visitor
## Levels: New Visitor Returning Visitor
```

3. Display the count/frequency of

- channel
- user_type

```
table(analytics$channel)
```

```
##
##      (Other)    Affiliates    Direct    Display Organic Search
##      6073      7388      39853      3375      139668
##      Paid Search    Referral    Social
##      4395      35615      8031
```

```
summary(analytics$user_type)
```

```
##      New Visitor Returning Visitor  
##      179045      65353
```

```
# using forcats
```

```
forcats::fct_count(analytics$channel)
```

```
## # A tibble: 8 x 2  
##   f              n  
##   <fct>         <int>  
## 1 (Other)      6073  
## 2 Affiliates   7388  
## 3 Direct      39853  
## 4 Display     3375  
## 5 Organic Search 139668  
## 6 Paid Search  4395  
## 7 Referral    35615  
## 8 Social      8031
```

```
forcats::fct_count(analytics$user_type)
```

```
## # A tibble: 2 x 2  
##   f              n  
##   <fct>         <int>  
## 1 New Visitor   179045  
## 2 Returning Visitor 65353
```

```
# using dplyr package
```

```
# count
```

```
count(analytics, channel)
```

```
## # A tibble: 8 x 2  
##   channel              n  
##   <fct>         <int>  
## 1 (Other)      6073  
## 2 Affiliates   7388  
## 3 Direct      39853  
## 4 Display     3375  
## 5 Organic Search 139668  
## 6 Paid Search  4395  
## 7 Referral    35615  
## 8 Social      8031
```

```
count(analytics, user_type)
```

```
## # A tibble: 2 x 2  
##   user_type              n  
##   <fct>         <int>  
## 1 New Visitor   179045  
## 2 Returning Visitor 65353
```

```
# tally
analytics %>%
  group_by(channel) %>%
  tally()
```

```
## # A tibble: 8 x 2
##   channel      n
##   <fct>      <int>
## 1 (Other)    6073
## 2 Affiliates 7388
## 3 Direct    39853
## 4 Display    3375
## 5 Organic Search 139668
## 6 Paid Search 4395
## 7 Referral   35615
## 8 Social     8031
```

```
analytics %>%
  group_by(user_type) %>%
  tally()
```

```
## # A tibble: 2 x 2
##   user_type      n
##   <fct>      <int>
## 1 New Visitor 179045
## 2 Returning Visitor 65353
```

4. Examine the distribution of the following and summarize your observations:

- channel by user_type
- device by purchase_flag
- channel by device
- channel by purchase_flag
- user_type by purchase_flag

```
table(analytics$channel, analytics$user_type)
```

```
##
##           New Visitor Returning Visitor
## (Other)           3718           2355
## Affiliates         5785           1603
## Direct            31576           8277
## Display            2631            744
## Organic Search    106197          33471
## Paid Search        2477            1918
## Referral           19511          16104
## Social             7150            881
```

```
# dplyr count
count(analytics, device, purchase_flag)
```

```
## # A tibble: 6 x 3
##   device purchase_flag     n
##   <fct>   <lgl>       <int>
## 1 Desktop FALSE      177202
## 2 Desktop TRUE        80
## 3 Mobile  FALSE     63294
## 4 Mobile  TRUE       188
## 5 Tablet  FALSE     3616
## 6 Tablet  TRUE        18
```

```
# dplyr tally
analytics %>%
  group_by(channel, device) %>%
  tally()
```

```
## # A tibble: 24 x 3
## # Groups:   channel [8]
##   channel     device     n
##   <fct>     <fct>   <int>
## 1 (Other)   Desktop  4022
## 2 (Other)   Mobile   1951
## 3 (Other)   Tablet    100
## 4 Affiliates Desktop  6546
## 5 Affiliates Mobile    777
## 6 Affiliates Tablet    65
## 7 Direct    Desktop 27619
## 8 Direct    Mobile 11650
## 9 Direct    Tablet   584
## 10 Display   Desktop   887
## # ... with 14 more rows
```

```
# gmodels
gmodels::CrossTable(analytics$channel, analytics$device)
```

```
##
##
##   Cell Contents
## |-----|
## |                      N |
## | Chi-square contribution |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  244398
##
##
##               | analytics$device
## analytics$channel |   Desktop |   Mobile |   Tablet | Row Total |
## -----|-----|-----|-----|-----|
##           (Other) |     4022 |     1951 |       100 |       6073 |
```

##		33.342	88.458	1.042	
##		0.662	0.321	0.016	0.025
##		0.023	0.031	0.028	
##		0.016	0.008	0.000	
##	-----	-----	-----	-----	-----
##	Affiliates	6546	777	65	7388
##		262.855	679.624	18.314	
##		0.886	0.105	0.009	0.030
##		0.037	0.012	0.018	
##		0.027	0.003	0.000	
##	-----	-----	-----	-----	-----
##	Direct	27619	11650	584	39853
##		57.534	162.817	0.124	
##		0.693	0.292	0.015	0.163
##		0.156	0.184	0.161	
##		0.113	0.048	0.002	
##	-----	-----	-----	-----	-----
##	Display	887	2155	333	3375
##		995.536	1864.113	1593.853	
##		0.263	0.639	0.099	0.014
##		0.005	0.034	0.092	
##		0.004	0.009	0.001	
##	-----	-----	-----	-----	-----
##	Organic Search	96196	41266	2206	139668
##		258.414	685.659	8.044	
##		0.689	0.295	0.016	0.571
##		0.543	0.650	0.607	
##		0.394	0.169	0.009	
##	-----	-----	-----	-----	-----
##	Paid Search	2806	1490	99	4395
##		45.785	106.331	17.327	
##		0.638	0.339	0.023	0.018
##		0.016	0.023	0.027	
##		0.011	0.006	0.000	
##	-----	-----	-----	-----	-----
##	Referral	34690	874	51	35615
##		3035.477	7585.514	432.478	
##		0.974	0.025	0.001	0.146
##		0.196	0.014	0.014	
##		0.142	0.004	0.000	
##	-----	-----	-----	-----	-----
##	Social	4516	3319	196	8031
##		294.378	728.745	49.118	
##		0.562	0.413	0.024	0.033
##		0.025	0.052	0.054	
##		0.018	0.014	0.001	
##	-----	-----	-----	-----	-----
##	Column Total	177282	63482	3634	244398
##		0.725	0.260	0.015	
##	-----	-----	-----	-----	-----
##					
##					

```
# descriptr
descriptr::ds_cross_table(analytics, channel, device)
```

```
##      Cell Contents
## |-----|
## |      Frequency |
## |      Percent   |
## |      Row Pct   |
## |      Col Pct   |
## |-----|
##
## Total Observations:  244398
##
## -----
## |              |              device              |
## -----
## |      channel |      Desktop |      Mobile |      Tablet |      Row Total |
## -----
## |      (Other) |      4022 |      1951 |      100 |      6073 |
## |              |      0.016 |      0.008 |      0 |              |
## |              |      0.66 |      0.32 |      0.02 |      0.02 |
## |              |      0.02 |      0.03 |      0.03 |              |
## -----
## |      Affiliates |      6546 |      777 |      65 |      7388 |
## |              |      0.027 |      0.003 |      0 |              |
## |              |      0.89 |      0.11 |      0.01 |      0.03 |
## |              |      0.04 |      0.01 |      0.02 |              |
## -----
## |      Direct |      27619 |      11650 |      584 |      39853 |
## |              |      0.113 |      0.048 |      0.002 |              |
## |              |      0.69 |      0.29 |      0.01 |      0.16 |
## |              |      0.16 |      0.18 |      0.16 |              |
## -----
## |      Display |      887 |      2155 |      333 |      3375 |
## |              |      0.004 |      0.009 |      0.001 |              |
## |              |      0.26 |      0.64 |      0.1 |      0.01 |
## |              |      0.01 |      0.03 |      0.09 |              |
## -----
## | Organic Search |      96196 |      41266 |      2206 |      139668 |
## |              |      0.394 |      0.169 |      0.009 |              |
## |              |      0.69 |      0.3 |      0.02 |      0.57 |
## |              |      0.54 |      0.65 |      0.61 |              |
## -----
## | Paid Search |      2806 |      1490 |      99 |      4395 |
## |              |      0.011 |      0.006 |      0 |              |
## |              |      0.64 |      0.34 |      0.02 |      0.02 |
## |              |      0.02 |      0.02 |      0.03 |              |
## -----
## |      Referral |      34690 |      874 |      51 |      35615 |
## |              |      0.142 |      0.004 |      0 |              |
## |              |      0.97 |      0.02 |      0 |      0.15 |
## |              |      0.2 |      0.01 |      0.01 |              |
## -----
```

##		Social		4516		3319		196		8031	
##				0.018		0.014		0.001			
##				0.56		0.41		0.02		0.03	
##				0.03		0.05		0.05			
##	-----										
##		Column Total		177282		63482		3634		244398	
##				0.725		0.261		0.013			
##	-----										

```
# xtabs
mytable <- xtabs(~user_type+purchase_flag, data = analytics)
ftable(mytable)
```

##		purchase_flag	FALSE	TRUE
##	user_type			
##	New Visitor		178862	183
##	Returning Visitor		65250	103