Module 4 - Data Manipulation

Rsquared Academy

2021-03-05

Import Data & Load Libraries

```
analytics <- readRDS("analytics.rds")
library(forcats)
library(magrittr)
```

1. Display the count/frequency of the following variables in the descending order

   - device
   - landing_page
   - exit_page

```
# device
analytics %>%
    use_series(device) %>%
    fct_count(sort = TRUE)
```

```
## # A tibble: 3 x 2
##    f              n
##    <fct>     <int>
## 1 Desktop 177282
## 2 Mobile    63482
## 3 Tablet     3634
```

```
# landing page
analytics %>%
    use_series(landing_page) %>%
    fct_count(sort = TRUE)
```

```
## # A tibble: 33 x 2
##     f                  n
##     <fct>          <int>
##  1 Home          135485
##  2 Apparel        32234
##  3 Shop by Brand  26826
##  4 Store           8979
##  5 New             6340
##  6 Lifestyle       5695
##  7 Accessories     4529
```

```
##  8 Basket          3920
##  9 Bags            3462
## 10 Sign In         3184
## # ... with 23 more rows
```

```
# exit page
analytics %>%
    use_series(exit_page) %>%
    fct_count(sort = TRUE)
```

```
## # A tibble: 34 x 2
##    f                  n
##    <fct>          <int>
##  1 Home           88051
##  2 Apparel        38318
##  3 Shop by Brand  25231
##  4 Store          17016
##  5 Lifestyle      13996
##  6 Basket         12072
##  7 New            10426
##  8 Search          7185
##  9 Accessories     4856
## 10 Bags            4584
## # ... with 24 more rows
```

2. Check if `laptop` is a level in the `device` column.

```
fct_match(analytics$device, 'laptop')
```

3. Combine the following levels in `landing_page` into `Account`

   - My Account
   - Register
   - Sign In
   - Your Info

```
# using fct_collapse
analytics %>%
    use_series(landing_page) %>%
    fct_collapse(Account = c("My Account", "Register", "Sign In", "Your Info")) %>%
    fct_count()
```

```
## # A tibble: 30 x 2
##    f                 n
##    <fct>         <int>
##  1 Accessories    4529
##  2 Apparel       32234
##  3 Bags           3462
##  4 Basket         3920
##  5 Brands          128
##  6 Drinkware      2806
##  7 Eco              75
```

```
##  8 Electronics   486
##  9 FAQ           2626
## 10 Fun              8
## # ... with 20 more rows
```

```
# using fct_recode
analytics %>%
    use_series(landing_page) %>%
    fct_recode(Account = "My Account",
               Account = "Register",
               Account = "Sign In",
               Account = "Your Info") %>%
    fct_count()
```

```
## # A tibble: 30 x 2
##    f               n
##    <fct>        <int>
##  1 Accessories   4529
##  2 Apparel      32234
##  3 Bags          3462
##  4 Basket        3920
##  5 Brands         128
##  6 Drinkware     2806
##  7 Eco             75
##  8 Electronics    486
##  9 FAQ           2626
## 10 Fun              8
## # ... with 20 more rows
```

4. Combine levels in `landing_page` that drive less than 1000 visits.

```
analytics %>%
    use_series(landing_page) %>%
    fct_lump_min(1000) %>%
    fct_count()
```

```
## # A tibble: 16 x 2
##    f                  n
##    <fct>          <int>
##  1 Accessories     4529
##  2 Apparel        32234
##  3 Bags            3462
##  4 Basket          3920
##  5 Drinkware       2806
##  6 FAQ             2626
##  7 Home          135485
##  8 Lifestyle       5695
##  9 My Account      1121
## 10 New             6340
## 11 Office          2454
## 12 Search          2066
## 13 Shop by Brand  26826
```

```
## 14 Sign In          3184
## 15 Store            8979
## 16 Other            2671
```

5. Get top 10 landing and exit pages.

```
# landing page
analytics %>%
    use_series(landing_page) %>%
    fct_lump_n(10) %>%
    fct_count()
```

```
## # A tibble: 11 x 2
##    f                 n
##    <fct>          <int>
##  1 Accessories     4529
##  2 Apparel        32234
##  3 Bags            3462
##  4 Basket          3920
##  5 Home          135485
##  6 Lifestyle       5695
##  7 New             6340
##  8 Shop by Brand  26826
##  9 Sign In         3184
## 10 Store           8979
## 11 Other          13744
```

```
# exit page
analytics %>%
    use_series(exit_page) %>%
    fct_lump_n(10) %>%
    fct_count()
```

```
## # A tibble: 12 x 2
##    f                 n
##    <fct>          <int>
##  1 Accessories     4856
##  2 Apparel        38318
##  3 Bags            4584
##  4 Basket         12072
##  5 Home           88051
##  6 Lifestyle      13996
##  7 New            10426
##  8 Search          7185
##  9 Shop by Brand  25231
## 10 Store          17016
## 11 Other          22095
## 12 <NA>            568
```

6. Get landing pages that drive at least 5% of the total traffic to the website.

```
analytics %>%
    use_series(landing_page) %>%
    fct_lump_prop(0.05) %>%
    fct_count()
```

```
## # A tibble: 4 x 2
##   f                 n
##   <fct>         <int>
## 1 Apparel       32234
## 2 Home         135485
## 3 Shop by Brand 26826
## 4 Other         49853
```

7. Retain only the following levels in the `browser` column:

   - Chrome
   - Firefox
   - Safari
   - Edge

```
analytics %>%
    use_series(browser) %>%
    fct_other(keep = c("Chrome", "Firefox", "Safari", "Edge")) %>%
    levels()
```

```
## [1] "Chrome"  "Edge"    "Firefox" "Safari"  "Other"
```

8. Anonymize landing and exit page levels.

```
# landing page
analytics %>%
    use_series(landing_page) %>%
    fct_anon(prefix = "lp_") %>%
    levels()
```

```
##  [1] "lp_01" "lp_02" "lp_03" "lp_04" "lp_05" "lp_06" "lp_07" "lp_08" "lp_09"
## [10] "lp_10" "lp_11" "lp_12" "lp_13" "lp_14" "lp_15" "lp_16" "lp_17" "lp_18"
## [19] "lp_19" "lp_20" "lp_21" "lp_22" "lp_23" "lp_24" "lp_25" "lp_26" "lp_27"
## [28] "lp_28" "lp_29" "lp_30" "lp_31" "lp_32" "lp_33"
```

```
# exit page
analytics %>%
    use_series(exit_page) %>%
    fct_anon(prefix = "ep_") %>%
    levels()
```

```
##  [1] "ep_01" "ep_02" "ep_03" "ep_04" "ep_05" "ep_06" "ep_07" "ep_08" "ep_09"
## [10] "ep_10" "ep_11" "ep_12" "ep_13" "ep_14" "ep_15" "ep_16" "ep_17" "ep_18"
## [19] "ep_19" "ep_20" "ep_21" "ep_22" "ep_23" "ep_24" "ep_25" "ep_26" "ep_27"
## [28] "ep_28" "ep_29" "ep_30" "ep_31" "ep_32" "ep_33"
```

9. Make `Home` first level in the `landing_page` column.

```
analytics %>%
    use_series(landing_page) %>%
    fct_relevel("Home") %>%
    levels()
```

```
##  [1] "Home"                "Accessories"          "Apparel"
##  [4] "Bags"                "Basket"               "Brands"
##  [7] "Drinkware"           "Eco"                  "Electronics"
## [10] "FAQ"                 "Fun"                  "Kids"
## [13] "Lifestyle"           "My Account"           "Nest"
## [16] "New"                 "Office"               "Others"
## [19] "Payment"             "Register"             "Return Policy"
## [22] "Search"              "Shipping Information" "Shop"
## [25] "Shop by Brand"       "Sign In"              "Specials"
## [28] "Store"               "Terms of Use"         "Top Sellers"
## [31] "Wearables"           "Wishlist"             "Your Info"
```

10. Make `Apparel` second level in the `landing_page` column.

```
analytics %>%
    use_series(landing_page) %>%
    fct_relevel("Apparel", after = 1) %>%
    levels()
```

```
##  [1] "Accessories"         "Apparel"              "Bags"
##  [4] "Basket"              "Brands"               "Drinkware"
##  [7] "Eco"                 "Electronics"          "FAQ"
## [10] "Fun"                 "Home"                 "Kids"
## [13] "Lifestyle"           "My Account"           "Nest"
## [16] "New"                 "Office"               "Others"
## [19] "Payment"             "Register"             "Return Policy"
## [22] "Search"              "Shipping Information" "Shop"
## [25] "Shop by Brand"       "Sign In"              "Specials"
## [28] "Store"               "Terms of Use"         "Top Sellers"
## [31] "Wearables"           "Wishlist"             "Your Info"
```

11. Make `Specials` last level in the `landing_page` column.

```
analytics %>%
    use_series(landing_page) %>%
    fct_relevel("Specials", after = Inf) %>%
    levels()
```

```
##  [1] "Accessories"         "Apparel"              "Bags"
##  [4] "Basket"              "Brands"               "Drinkware"
##  [7] "Eco"                 "Electronics"          "FAQ"
## [10] "Fun"                 "Home"                 "Kids"
## [13] "Lifestyle"           "My Account"           "Nest"
## [16] "New"                 "Office"               "Others"
## [19] "Payment"             "Register"             "Return Policy"
## [22] "Search"              "Shipping Information" "Shop"
```

```
## [25] "Shop by Brand"        "Sign In"            "Store"
## [28] "Terms of Use"         "Top Sellers"        "Wearables"
## [31] "Wishlist"             "Your Info"          "Specials"
```

12. Order the levels in `browser` by frequency

```
analytics %>%
    use_series(browser) %>%
    fct_infreq() %>%
    levels()
```

```
##  [1] "Chrome"                   "Safari"
##  [3] "Firefox"                  "Edge"
##  [5] "Samsung Internet"         "Android Webview"
##  [7] "Opera"                    "Internet Explorer"
##  [9] "Safari (in-app)"          "YaBrowser"
## [11] "UC Browser"               "Amazon Silk"
## [13] "Android Browser"          "Opera Mini"
## [15] "Coc Coc"                  "User-Agent:Mozilla"
## [17] "Puffin"                   "Playstation 4"
## [19] "Mozilla Compatible Agent" "BlackBerry"
## [21] "Maxthon"                  "SeaMonkey"
## [23] "APKPure"                  "MRCHROME"
## [25] "Playstation Vita Browser" "Seznam"
```

13. Order the levels in `landing_page` by order of appearance

```
analytics %>%
    use_series(landing_page) %>%
    fct_inorder() %>%
    levels()
```

```
##  [1] "Home"        "Accessories"   "Sign In"
##  [4] "Apparel"     "Shop by Brand" "Lifestyle"
##  [7] "New"         "Bags"          "Store"
## [10] "Drinkware"   "Basket"        "My Account"
## [13] "Office"      "Return Policy" "Search"
## [16] "FAQ"         "Electronics"   "Shipping Information"
## [19] "Brands"      "Nest"          "Shop"
## [22] "Payment"     "Eco"           "Others"
## [25] "Wearables"   "Terms of Use"  "Kids"
## [28] "Fun"         "Register"      "Your Info"
## [31] "Top Sellers" "Wishlist"      "Specials"
```

14. Shuffle the levels in `os`

```
analytics %>%
    use_series(os) %>%
    fct_shuffle() %>%
    levels()
```

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

15. Reverse the levels in `browser`

```
analytics %>%
    use_series(browser) %>%
    fct_rev() %>%
    levels()
```

```
##  [1] "YaBrowser"                "User-Agent:Mozilla"
##  [3] "UC Browser"               "Seznam"
##  [5] "SeaMonkey"                "Samsung Internet"
##  [7] "Safari (in-app)"          "Safari"
##  [9] "Puffin"                   "Playstation Vita Browser"
## [11] "Playstation 4"            "Opera Mini"
## [13] "Opera"                    "MRCHROME"
## [15] "Mozilla Compatible Agent" "Maxthon"
## [17] "Internet Explorer"        "Firefox"
## [19] "Edge"                     "Coc Coc"
## [21] "Chrome"                   "BlackBerry"
## [23] "APKPure"                  "Android Webview"
## [25] "Android Browser"          "Amazon Silk"
```