

Almost Famous: Analyse campaign query combinations

Cindy Lamm

23:52, Tuesday 20th January, 2015

Load variable names and types:

```
nameTypeDataFile <- "../data/raw_variables.csv"
variableNames <- read.csv(nameTypeDataFile, header=TRUE, stringsAsFactors=FALSE)
variableNames

##      name      type
## 1 visit_id  factor
## 2      uid  factor
## 3 campaign  factor
## 4    tstamp character
## 5 experiments factor
## 6   action  factor
## 7    query  factor

factorIdx <- which(variableNames$type=="factor")
factorNames <- variableNames$name[factorIdx]
```

Read the per visit aggregated web log data:

```
webFile <- "../data/web_visits.csv"
visitData <- read.csv(webFile, stringsAsFactors=FALSE, col.names=variableNames$name,
                      colClasses=variableNames$type, na.strings=c("NA",""))
visitData$tstamp <- as.POSIXct(visitData$tstamp)
str(visitData)

## 'data.frame': 1482602 obs. of 7 variables:
## $ visit_id : Factor w/ 1482602 levels "10000024498",...: 1252062 128641 583195 349394 830165 690964 ...
## $ uid      : Factor w/ 1064214 levels "100000493","100000682",...: 858988 92339 95584 929716 656934 ...
## $ campaign : Factor w/ 10 levels "103","127","14",...: 7 1 7 1 4 8 1 1 1 2 ...
## $ tstamp   : POSIXct, format: "2014-09-18 05:43:18" "2014-09-16 21:24:08" ...
## $ experiments: Factor w/ 4 levels "[1 3]","[1 4]",...: 2 1 4 1 3 2 1 3 2 1 ...
## $ action    : Factor w/ 8 levels "[landed adclick adclick adclick]",...: 8 8 8 8 8 8 8 8 8 8 ...
## $ query     : Factor w/ 5 levels "advanced analytics",...: 1 5 1 5 3 5 5 5 4 ...
```

```
summary(visitData)
```

	visit_id	uid	campaign	tstamp
## 10000024498:	1	102486699:	7 558 :324872	Min. :2014-09-15 00:00:01
## 10000032484:	1	123618732:	7 103 :324027	1st Qu.:2014-09-18 16:32:04

```
## 10000079220:      1  143588980:      7  59      :232002  Median :2014-09-22 16:55:36
## 10000092303:      1  159226004:      7  31      :231685  Mean   :2014-09-22 20:33:11
## 10000132469:      1  168873739:      7  127     : 92681  3rd Qu.:2014-09-26 19:41:15
## 10000206890:      1  171898393:      7  94      : 92436  Max.   :2014-09-30 23:53:20
## (Other)      :1482596  (Other)   :1482560  (Other):184899
## experiments                                     action
## [1 3]:370018   landed                               :1291256
## [1 4]:371852   [landed signup]                       : 84889
## [2 3]:370082   [landed order]                         : 43930
## [2 4]:370650   [landed adclick]                       : 28233
##               [landed adclick adclick adclick]: 14956
##               [landed adclick adclick]           : 14875
##               (Other)                             : 4463
##               query
## advanced analytics                               :463687
## building predictive models: 92454
## data science                                     : 92445
## data science training                           :185117
## predictive modeling                             :648899
##
##
```

How many visits are there per campaign-query combination?

```
combinations <- expand.grid(queries=levels(visitData$query), campaigns=levels(visitData$campaign))
webAggCampaignQuery <- aggregatePerCQ(visitData)
webAggCampaignQuery
```

	campaign	query	nb_visits	nb_uids
## 1	103	predictive modeling	324027	300394
## 2	127	data science training	92681	90761
## 3	14	building predictive models	46252	45738
## 4	17	data science	46308	45814
## 5	203	building predictive models	46202	45711
## 6	23	data science	46137	45630
## 7	31	advanced analytics	231685	219610
## 8	558	predictive modeling	324872	301174
## 9	59	advanced analytics	232002	219833
## 10	94	data science training	92436	90510