process_data.R

Guillermo

Tue Apr 07 10:44:36 2015

```r
library(printr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(lubridate)

## My own directory where I downloaded all the original files
dir <- "../DMC_2015/"

## List all the txt files in that directory
files <-
  list.files(dir, pattern = "*txt")

## Read the files appropriately
train <-
  read.table(paste0(dir, files[2]), sep = "|", head = T) %>%
  tbl_df() %>%
  mutate(orderTime = ymd_hms(orderTime),
         couponsReceived = ymd_hms(couponsReceived))

## Read the files appropriately
test <-
  read.table(paste0(dir, files[1]), sep = "|", head = T) %>%
  tbl_df() %>%
  mutate(orderTime = ymd_hms(orderTime),
         couponsReceived = ymd_hms(couponsReceived))
```

```r
train[, 1:4 + 0*4] %>% tail()
```

orderID	orderTime	userID	couponsReceived
6048	2015-03-09 21:55:52	2c5e9d0d342a3df2c73b9a1ecd62b76f	2015-03-04 23:03:18
6049	2015-03-09 22:09:43	429854cd82ef2427c97b35181c1a326e	2015-03-03 22:20:21
6050	2015-03-09 22:10:30	682d6de07accc4a98958d09e7ed9e993	2015-03-07 23:49:29
6051	2015-03-09 22:19:05	91e7eb87e6625f300234186ddae97515	2015-03-03 22:44:13

orderID	orderTime	userID	couponsReceived
6052	2015-03-09 23:25:12	6fb117c077008de1fcb76551909fbe02	2015-03-03 19:39:36
6053	2015-03-09 23:26:57	4c152a0a4ee7d9541b6af77d297aba9c	2015-03-03 13:49:13

```
train[, 1:4 + 1*4] %>% tail()
```

couponID1	price1	basePrice1	reward1
4b7f8cd394ebc65414c182e7261127f4	2.73	2.09	1.57
64a3f47f069fd3f1cb5ac46288c649e6	6.44	0.00	0.94
6f4bf39b8f0457229f7443f6f3b9ae74	4.31	1.72	1.57
173cac356b89eee4fcfa9eeb2a8a8f2c	6.11	2.04	1.26
fd95fcb38dd6dbd3a56d9769e9d9871f	4.91	4.91	1.26
88258bcc0389d4a1825ebafeda054f72	3.89	1.95	0.94

```
train[, 1:4 + 2*4] %>% tail()
```

premiumProduct1	brand1	productGroup1	categoryIDs1
0	a38b030516e52b6f126fb96c159aba25	e07e0204e8a62673afc9ad29b4f79f4a	b7e3b0110f96213f508703⋯
1	4e03e74099990e8529eb7e7328220af1	dc140663301ed170b7e065adbe4ab044	b78123c345b9f8a300a7fa⋯
0		f0b887f83f4eee519485cd646640e308	b78123c345b9f8a300a7fa⋯
0		ca7734a09ece2e6d5f3535b2b7e31fe0	b7e3b0110f96213f508703⋯
0		349e7ceb68b2f9fdfcc0d27118406b6e	e73355e7a58b57d741111e⋯
1	4e03e74099990e8529eb7e7328220af1	86400818cd399e401fef12f5e220240d	b78123c345b9f8a300a7fa⋯

```
train[, 1:4 + 3*4] %>% tail()
```

couponID2	price2	basePrice2	reward2
0e0acfd24631380ef26b64a468a059a6	5.00	2.27	1.26
2c9d9667cc58220a1be35fc8245e1dfc	2.73	0.45	0.63
4cdf5b51ec0f362fce51d52b10969dca	8.10	0.90	1.26
a8aeb37695d7d5a6bc2fb7cde6e7a50b	12.60	1.26	0.94
ff6b305ca83500b1c74b721604723008	5.93	0.79	1.57
6f4e6f514c8d01d1af8dd9b226420b21	9.08	2.27	1.57

```
train[, 1:4 + 4*4] %>% tail()
```

premiumProduct2	brand2	productGroup2	categoryIDs2
0	88566ad89b50ab4d0b2d6bd785b2e46c	daf40da25b745f9c176c910885d8c1d3	f12c4112c9ea685dc2d270⋯
0	f6601412b868a05c118961e9d7869bc8	008b99d9c2fd9d58dd1483e364f8e8d3	f12c4112c9ea685dc2d270⋯
0	370c2d67b4ad305fba6dfa8a0942a427	041a02d031057ba1c2e210fe6b0db0f4	4c65852ba87d9e7d2cc82⋯
0	f6601412b868a05c118961e9d7869bc8	37d4e26777956be4bae6409f8a1c4f2b	f12c4112c9ea685dc2d270⋯
0		e54573aaf3c08c40545b97cc5ddf1732	f12c4112c9ea685dc2d270⋯
1	4e03e74099990e8529eb7e7328220af1	c062480d27f1c080c666718888417818	f12c4112c9ea685dc2d270⋯

```
train[, 1:4 + 5*4] %>% tail()
```

couponID3	price3	basePrice3	reward3
3e4bfce4c3c7ae98db3d9c859a5dcfe8	3.43	0.48	2.20
2db66e8fc33b5e4699cfd8c985fbb2bb	6.02	2.41	1.57
4270374b0780f856a6afdf98bf759f1a	2.73	0.88	1.26
4d44f617fb70d50519502374dbb066c5	4.58	0.92	1.26
9622f0bbc23997b902cf1445b5196fdb	3.01	0.00	1.57
ee833d6dd1c508a53c1674497a80c753	3.66	1.46	1.57

```
train[, 1:4 + 6*4] %>% tail()
```

premiumProduct3	brand3	productGroup3	categoryIDs3
1	4e03e74099990e8529eb7e7328220af1	1ba3059382b4d1db6dfe2e3d484a6d98	796c5815752a170021838a
1	4e03e74099990e8529eb7e7328220af1	94a7d11a8972ed94819c748a1d1b42f8	796c5815752a170021838a
0	304979e999d95fc59d85cba5f1633595	e006513b5c7be290e81c8bd614362d07	796c5815752a170021838a
0	304979e999d95fc59d85cba5f1633595	5b5714af4a2acb5ea5a65c7e0466e9e4	796c5815752a170021838a
0	f6601412b868a05c118961e9d7869bc8	0949b1d7fc3a85d158b7bad3b1823afa	f74dc92e8e6835ab41d6cb
0		443cdbe7628a33457d2822ed1ed36a54	796c5815752a170021838a

```
train[, 1:4 + 7*4] %>% tail()
```

coupon1Used	coupon2Used	coupon3Used	basketValue
0	1	0	271.56
0	0	0	266.70
0	0	0	913.19
0	0	0	95.58
0	0	0	115.91
0	0	0	144.77