

5.3 Assignment: Create Optimal Hotel Recommendation

Shani Kumar

7/07/2020

5.3 Assignment: Create Optimal Hotel Recommendation

All online travel agencies are scrambling to meet the Artificial Intelligence driven personalization standard set by Amazon and Netflix. In addition, the world of online travel has become a highly competitive space where brands try to capture our attention (and wallet) with recommending, comparing, matching, and sharing. For this assignment, we aim to create the optimal hotel recommendations for Expedia's users that are searching for a hotel to book. For this assignment, you need to predict which "hotel cluster" the user is likely to book, given his (or her) search details. In doing so, you should be able to demonstrate your ability to use four different algorithms (of your choice). The data set can be found at Kaggle: Expedia Hotel Recommendations. To get you started, I would suggest you use `train.csv` which captured the logs of user behavior and `destinations.csv` which contains information related to hotel reviews made by users. You are also required to write a one page summary of your approach in getting to your prediction methods. I expect you to use a combination of R and Python in your answer.

Part 1 - Import and show a glimpse of data

```
## Warning: package 'ggplot2' was built under R version 3.6.2
## Warning: package 'lubridate' was built under R version 3.6.2
##
## Attaching package: 'lubridate'
##
## The following objects are masked from 'package:data.table':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
##
## corrrplot 0.84 loaded
```

Since the data is huge I am using first 100k records from **train.csv** dataset.

```
##           date_time site_name posa_continent user_location_country
## 1: 2014-08-11 07:46:59         2             3                    66
## 2: 2014-08-11 08:22:12         2             3                    66
## 3: 2014-08-11 08:24:33         2             3                    66
## 4: 2014-08-09 18:05:16         2             3                    66
## 5: 2014-08-09 18:08:18         2             3                    66
## 6: 2014-08-09 18:13:12         2             3                    66
## user_location_region user_location_city orig_destination_distance user_id
## 1:                  348             48862             2234.2641         12
## 2:                  348             48862             2234.2641         12
```

```

## 3:          348          48862          2234.2641          12
## 4:          442          35390          913.1932          93
## 5:          442          35390          913.6259          93
## 6:          442          35390          911.5142          93
##   is_mobile is_package channel   srch_ci   srch_co srch_adults_cnt
## 1:         0         1       9 2014-08-27 2014-08-31           2
## 2:         0         1       9 2014-08-29 2014-09-02           2
## 3:         0         0       9 2014-08-29 2014-09-02           2
## 4:         0         0       3 2014-11-23 2014-11-28           2
## 5:         0         0       3 2014-11-23 2014-11-28           2
## 6:         0         0       3 2014-11-23 2014-11-28           2
##   srch_children_cnt srch_rm_cnt srch_destination_id srch_destination_type_id
## 1:                 0         1             8250             1
## 2:                 0         1             8250             1
## 3:                 0         1             8250             1
## 4:                 0         1            14984             1
## 5:                 0         1            14984             1
## 6:                 0         1            14984             1
##   is_booking cnt hotel_continent hotel_country hotel_market hotel_cluster
## 1:         0   3             2             50             628             1
## 2:         1   1             2             50             628             1
## 3:         0   1             2             50             628             1
## 4:         0   1             2             50            1457             80
## 5:         0   1             2             50            1457             21
## 6:         0   1             2             50            1457             92

```

Data structure:

```

## Classes 'data.table' and 'data.frame':  100000 obs. of  24 variables:
## $ date_time      : chr  "2014-08-11 07:46:59" "2014-08-11 08:22:12" "2014-08-11 08:24:33"
## $ site_name      : int  2 2 2 2 2 2 2 2 2 2 ...
## $ posa_continent : int  3 3 3 3 3 3 3 3 3 3 ...
## $ user_location_country : int  66 66 66 66 66 66 66 66 66 66 ...
## $ user_location_region : int  348 348 348 442 442 442 189 189 189 189 ...
## $ user_location_city : int  48862 48862 48862 35390 35390 35390 10067 10067 10067 10067 ...
## $ orig_destination_distance: num  2234 2234 2234 913 914 ...
## $ user_id        : int  12 12 12 93 93 93 501 501 501 501 ...
## $ is_mobile      : int  0 0 0 0 0 0 0 0 0 0 ...
## $ is_package     : int  1 1 0 0 0 0 0 1 0 0 ...
## $ channel        : int  9 9 9 3 3 3 2 2 2 2 ...
## $ srch_ci        : chr  "2014-08-27" "2014-08-29" "2014-08-29" "2014-11-23" ...
## $ srch_co        : chr  "2014-08-31" "2014-09-02" "2014-09-02" "2014-11-28" ...
## $ srch_adults_cnt : int  2 2 2 2 2 2 2 2 2 2 ...
## $ srch_children_cnt : int  0 0 0 0 0 0 0 0 0 0 ...
## $ srch_rm_cnt    : int  1 1 1 1 1 1 1 1 1 1 ...
## $ srch_destination_id : int  8250 8250 8250 14984 14984 14984 8267 8267 8267 8267 ...
## $ srch_destination_type_id : int  1 1 1 1 1 1 1 1 1 1 ...
## $ is_booking     : int  0 1 0 0 0 0 0 0 0 0 ...
## $ cnt           : int  3 1 1 1 1 1 2 1 1 1 ...
## $ hotel_continent : int  2 2 2 2 2 2 2 2 2 2 ...
## $ hotel_country   : int  50 50 50 50 50 50 50 50 50 50 ...
## $ hotel_market    : int  628 628 628 1457 1457 1457 675 675 675 675 ...
## $ hotel_cluster   : int  1 1 1 80 21 92 41 41 69 70 ...
## - attr(*, ".internal.selfref")=<externalptr>

```

is_mobile, is_package, channel, posa_continent, hotel_continent are categorical variables. So converting them in factor. Also date_time is defined as character so I am converting it into date time format. After converting extracting year, month, day, weekday and hour from it.

```
# Convert into factor
trn[, `:=`(is_mobile = as.factor(is_mobile),
           is_package = as.factor(is_package),
           channel = as.factor(channel),
           posa_continent = as.factor(posa_continent),
           hotel_continent = as.factor(hotel_continent)))]

# Convert into date time format
trn[, date_time := parse_date_time(date_time, "%y-%m-%d %H:%M:%S")]
trn[, srch_ci := parse_date_time(srch_ci, "%y-%m-%d")]
trn[, srch_co := parse_date_time(srch_co, "%y-%m-%d")]

# Extract year, month, day, week day and hour
trn[, `:=`(date_year = as.factor(year(date_time)),
           date_month = as.factor(month(date_time)),
           date_day = as.factor(day(date_time)),
           date_wday = as.factor(wday(date_time, label = T)),
           date_hour = as.factor(hour(date_time)),
           ci_year = as.factor(year(srch_ci)),
           ci_month = as.factor(month(srch_ci)),
           ci_day = as.factor(day(srch_ci)),
           co_year = as.factor(year(srch_co)),
           co_month = as.factor(month(srch_co)),
           co_day = as.factor(day(srch_co)))]
```

Data Summary:

```
##      date_time      site_name      posa_continent
##  Min.   :2013-01-07 05:40:13  Min.   : 2.0      0: 351
##  1st Qu.:2013-11-19 20:15:43  1st Qu.: 2.0      1:14527
##  Median :2014-06-05 09:06:28  Median : 2.0      2: 7178
##  Mean   :2014-04-11 07:16:56  Mean   : 9.1      3:76874
##  3rd Qu.:2014-09-22 16:26:21  3rd Qu.:11.0     4: 1070
##  Max.   :2014-12-31 23:08:08  Max.   :53.0
##
##  user_location_country user_location_region user_location_city
##  Min.   : 0.00      Min.   : 0.0      Min.   : 3
##  1st Qu.: 66.00      1st Qu.: 174.0    1st Qu.:13914
##  Median : 66.00      Median : 311.0    Median :27733
##  Mean   : 84.53      Mean   : 311.6    Mean   :28465
##  3rd Qu.: 69.00      3rd Qu.: 385.0    3rd Qu.:43113
##  Max.   :239.00      Max.   :1025.0    Max.   :56495
##
##  orig_destination_distance  user_id      is_mobile is_package
##  Min.   : 0.01      Min.   : 12      0:86197  0:73953
##  1st Qu.: 290.53      1st Qu.:107548  1:13803  1:26047
##  Median : 1093.89      Median :181983
##  Mean   : 1897.61      Mean   :195701
##  3rd Qu.: 2518.18      3rd Qu.:301357
##  Max.   :11641.22      Max.   :391007
##  NA's   :36922
```

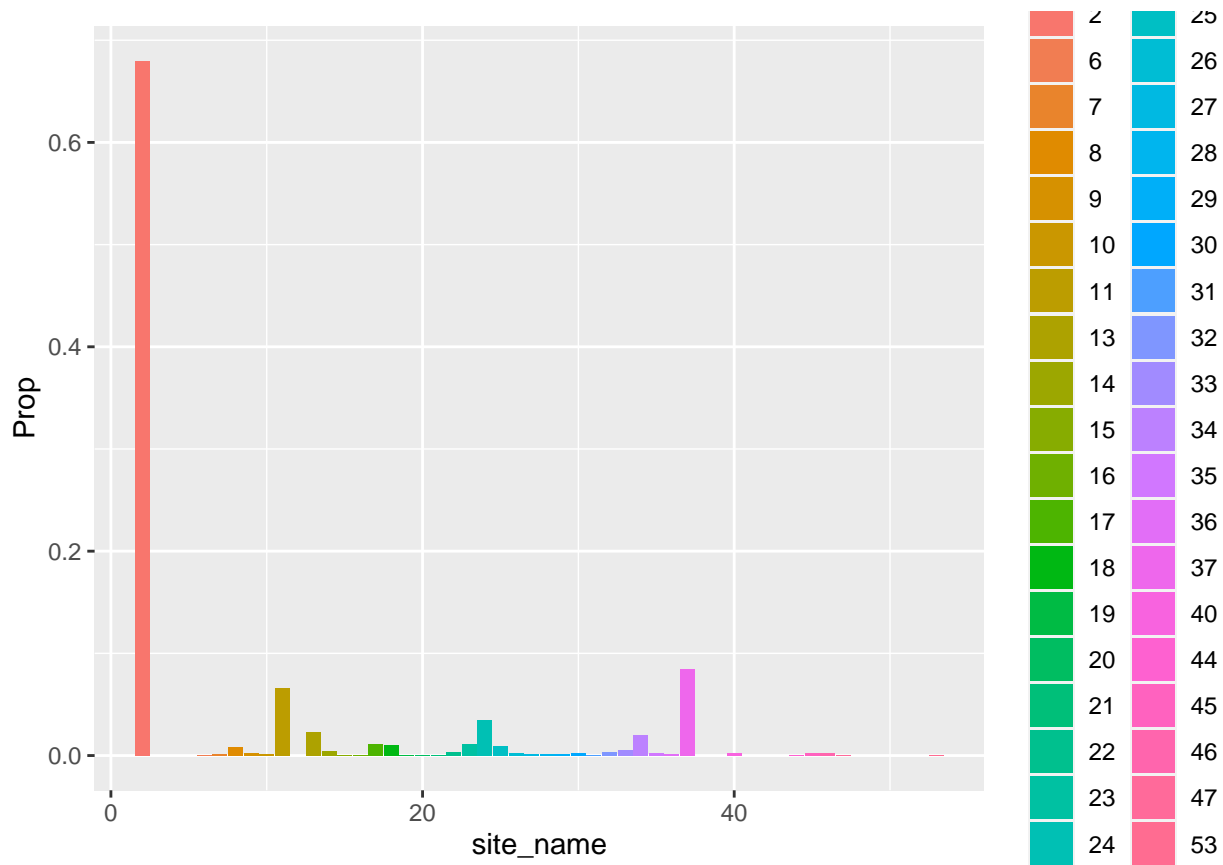
```

##      channel      srch_ci      srch_co
## 9      :54652  Min.   :2013-01-07 00:00:00  Min.   :2013-01-08 00:00:00
## 0      :13365  1st Qu.:2014-01-05 00:00:00  1st Qu.:2014-01-09 00:00:00
## 1      :10432  Median :2014-07-26 00:00:00  Median :2014-07-30 00:00:00
## 2      : 8615  Mean   :2014-06-04 15:21:40  Mean   :2014-06-08 00:38:15
## 5      : 4585  3rd Qu.:2014-11-15 00:00:00  3rd Qu.:2014-11-18 00:00:00
## 3      : 4326  Max.   :2016-03-08 00:00:00  Max.   :2016-03-15 00:00:00
## (Other): 4025  NA's   :71      NA's   :71
## srch_adults_cnt srch_children_cnt srch_rm_cnt srch_destination_id
## Min.   :0.00  Min.   :0.0000  Min.   :0.000  Min.   : 8
## 1st Qu.:2.00  1st Qu.:0.0000  1st Qu.:1.000  1st Qu.: 8267
## Median :2.00  Median :0.0000  Median :1.000  Median :11271
## Mean   :2.03  Mean   :0.3252  Mean   :1.113  Mean   :14511
## 3rd Qu.:2.00  3rd Qu.:0.0000  3rd Qu.:1.000  3rd Qu.:18881
## Max.   :9.00  Max.   :9.0000  Max.   :8.000  Max.   :65035
##
## srch_destination_type_id is_booking cnt hotel_continent
## Min.   :1.000  Min.   :0.0000  Min.   : 1.000  0: 1151
## 1st Qu.:1.000  1st Qu.:0.0000  1st Qu.: 1.000  2:52859
## Median :1.000  Median :0.0000  Median : 1.000  3:12020
## Mean   :2.591  Mean   :0.0827  Mean   : 1.478  4:11908
## 3rd Qu.:5.000  3rd Qu.:0.0000  3rd Qu.: 2.000  5: 1129
## Max.   :9.000  Max.   :1.0000  Max.   :59.000  6:20933
##
## hotel_country hotel_market hotel_cluster date_year date_month
## Min.   : 0.00  Min.   : 0.0  Min.   : 0.00  2013:27716 7 :10469
## 1st Qu.: 50.00  1st Qu.: 160.0  1st Qu.:25.00  2014:72284 8 :10386
## Median : 50.00  Median : 585.0  Median :49.00  12 :10119
## Mean   : 80.97  Mean   : 597.6  Mean   :49.75  10 :10113
## 3rd Qu.:106.00  3rd Qu.: 701.0  3rd Qu.:73.00  9 : 9320
## Max.   :212.00  Max.   :2117.0  Max.   :99.00  11 : 9254
## (Other):40339
## date_day date_wday date_hour ci_year ci_month
## 8 : 3674 Sun:13103 11 : 6337 2013:24586 12 :12359
## 9 : 3649 Mon:15924 18 : 6137 2014:63962 8 :11784
## 7 : 3590 Tue:16289 17 : 6129 2015:11380 7 : 9481
## 26 : 3553 Wed:16660 15 : 5992 2016: 1 10 : 9326
## 22 : 3482 Thu:14756 12 : 5902 NA's: 71 11 : 8800
## 17 : 3468 Fri:12491 16 : 5860 (Other):48179
## (Other):78584 Sat:10777 (Other):63643 NA's : 71
## ci_day co_year co_month co_day
## 26 : 3741 2013:24134 8 :11385 30 : 3865
## 27 : 3653 2014:62593 12 :10987 1 : 3672
## 1 : 3610 2015:13197 7 : 9409 28 : 3580
## 20 : 3597 2016: 5 10 : 9342 20 : 3521
## 28 : 3576 NA's: 71 9 : 8955 26 : 3489
## (Other):81752 (Other):49851 (Other):81802
## NA's : 71 NA's : 71 NA's : 71

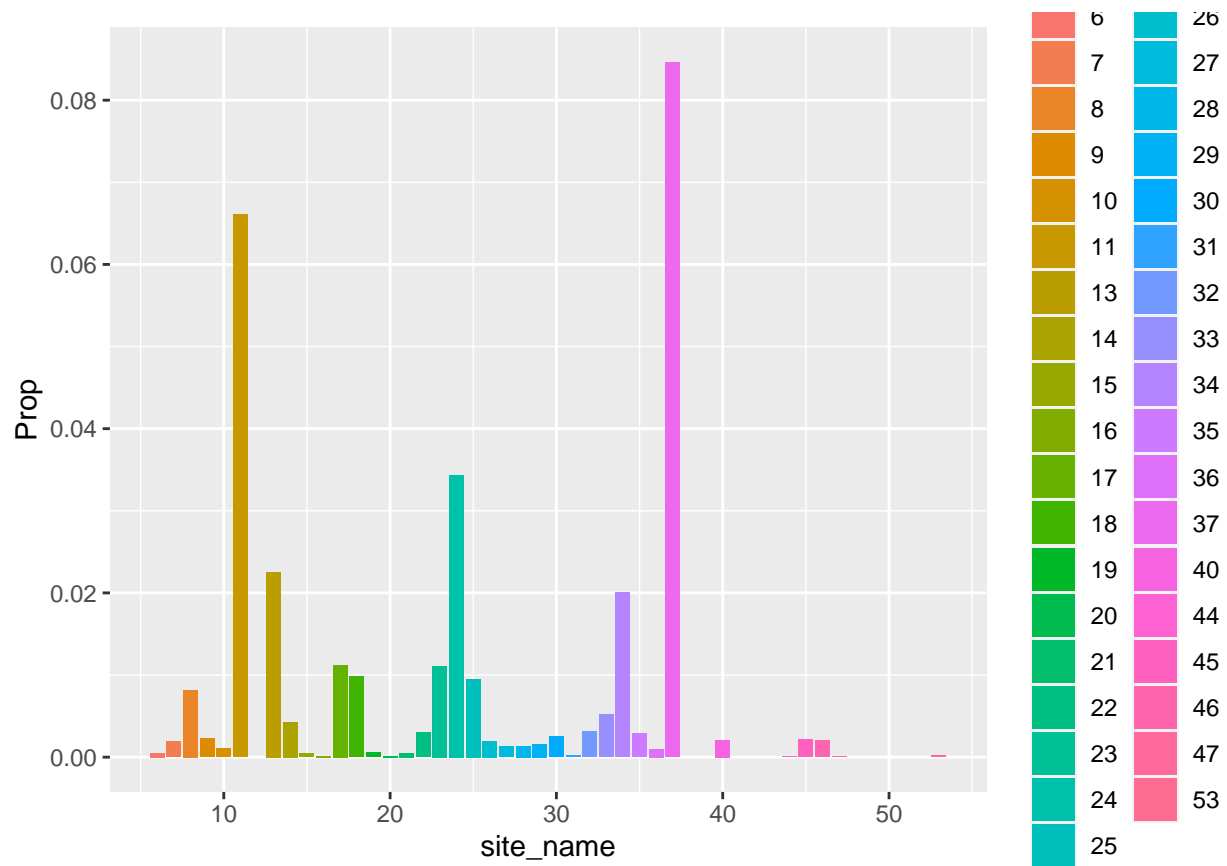
```

Explore Data:

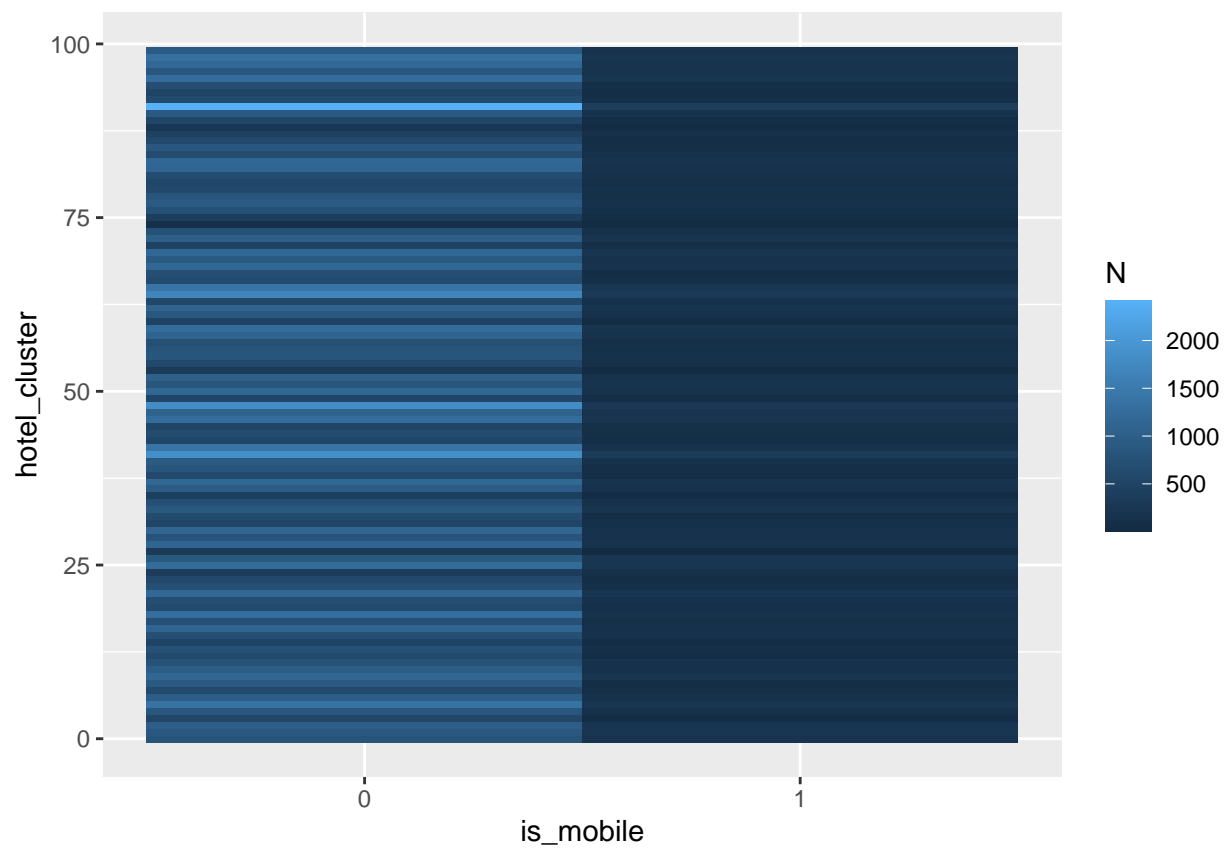
Site_name - Barchart



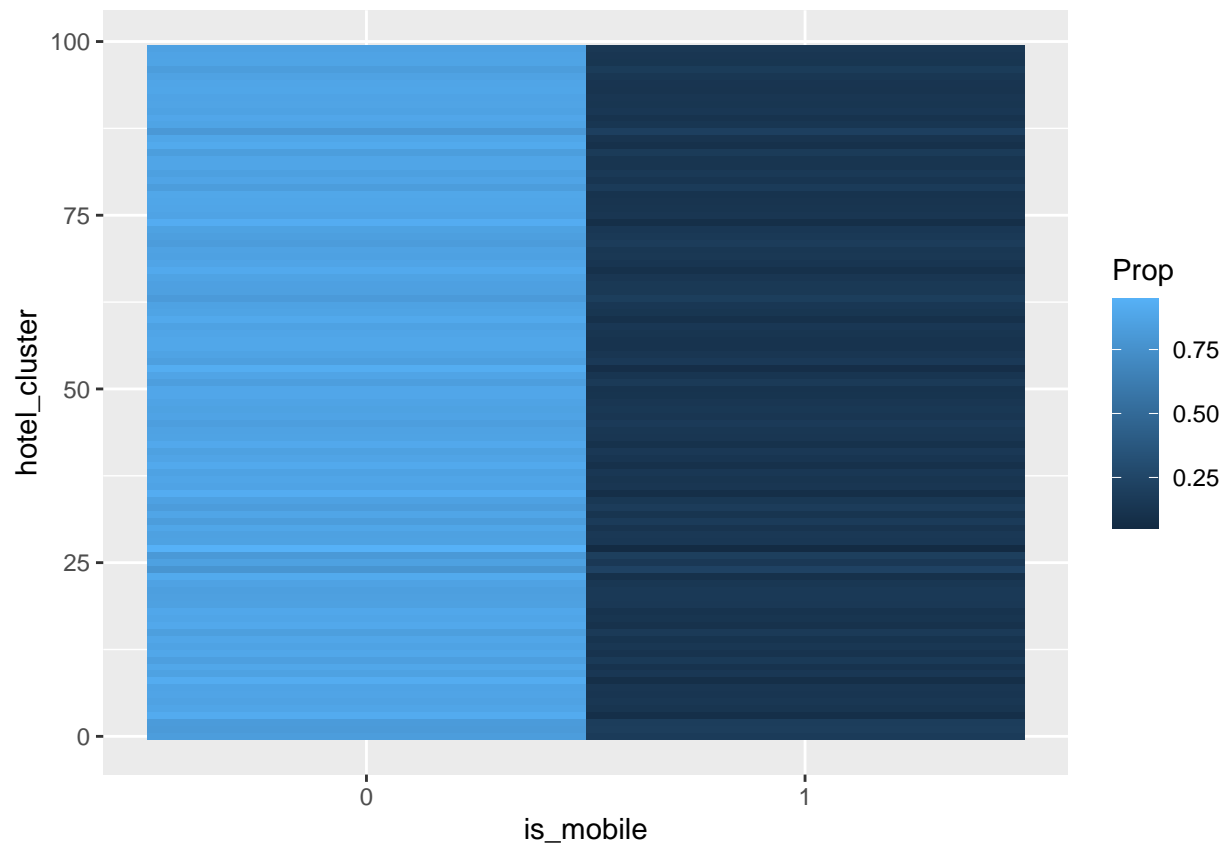
In the above chart we can see site name 2 is dominating set of other values so redrawing it after removing this value. We can see in below chart there are few site which gets more by user as compared to rest of the site names.



Mobile & Hotel cluster

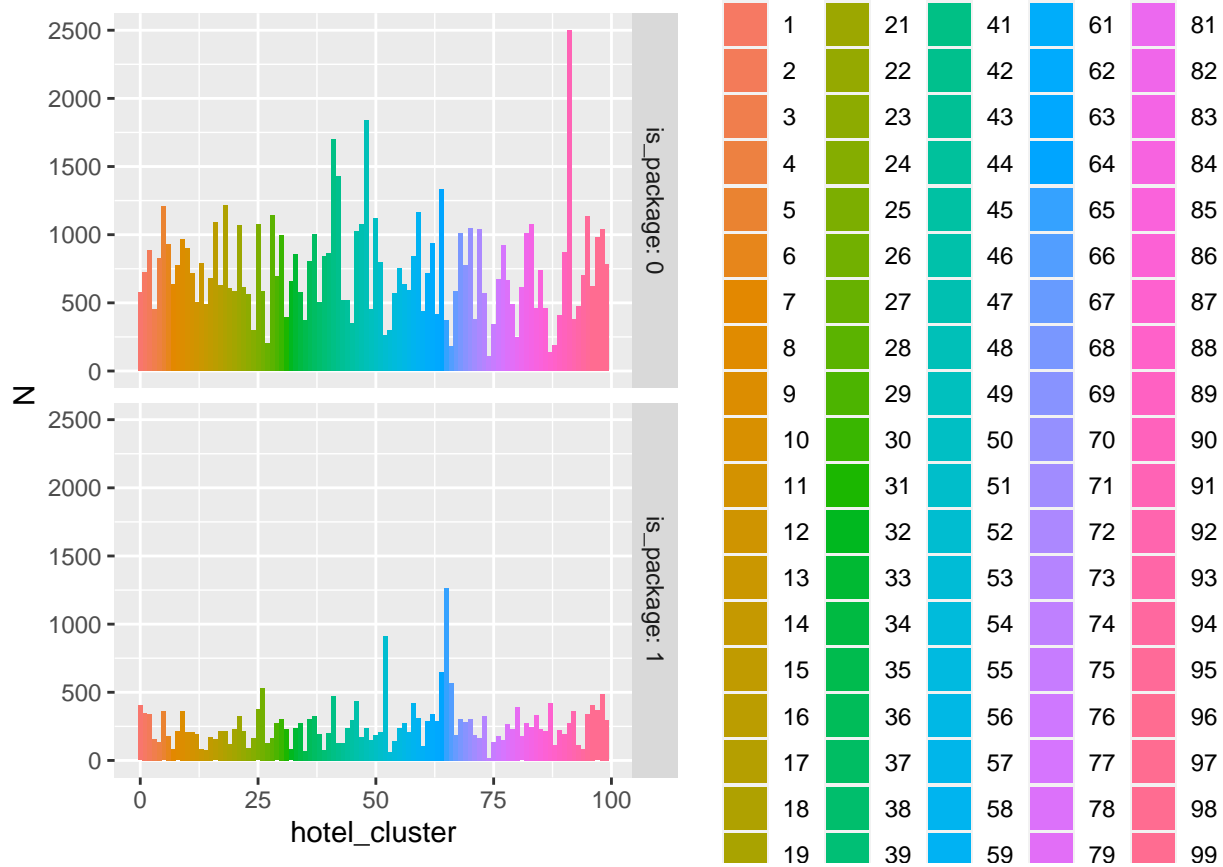


We can see in the above chart majority customers are not mobile user. Lets look at it a bit more by drawing relationship with hotel cluster.

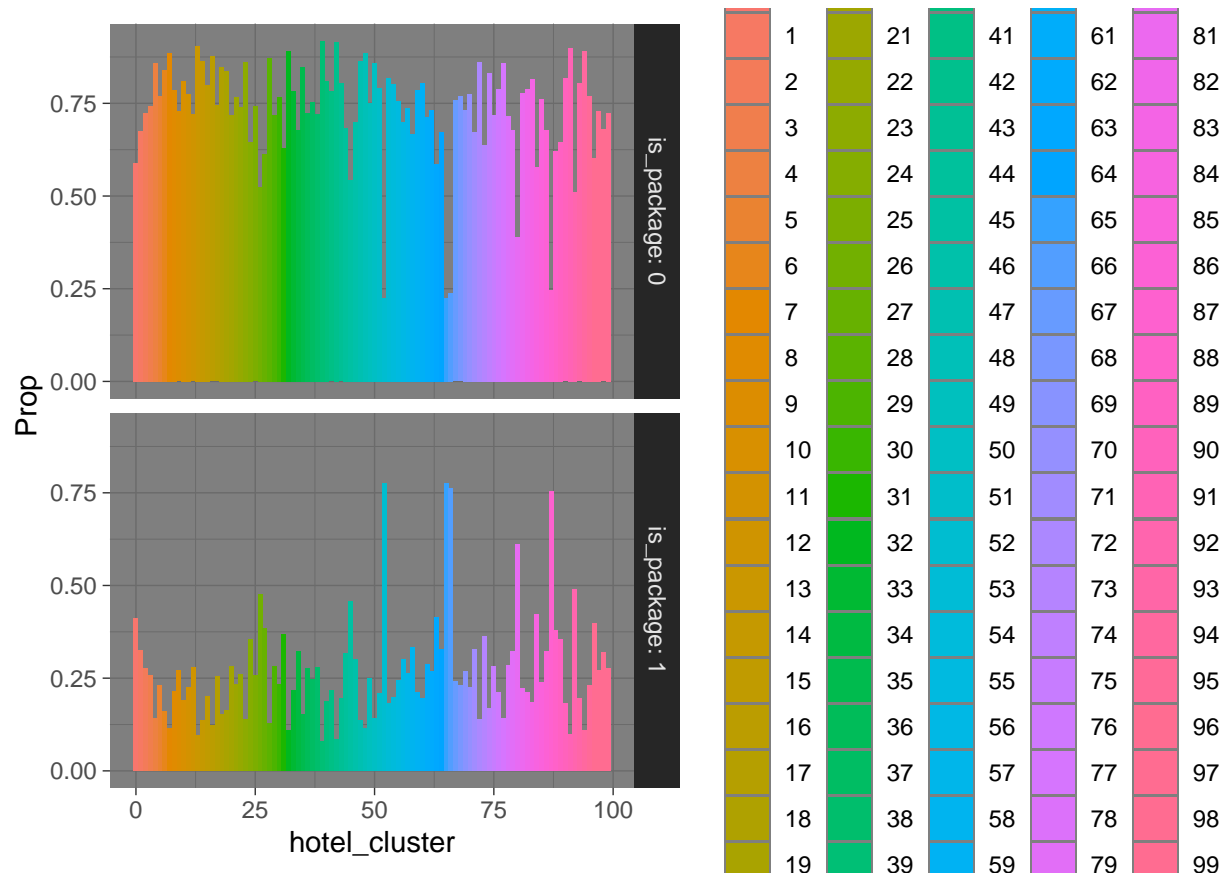


Here calculated proportion over each hotel cluster and we can see all the hotel_clusters have more or less the same proportion among is_mobile 0 and 1, with just a few having a larger proportion than average.

is_package & Hotel cluster

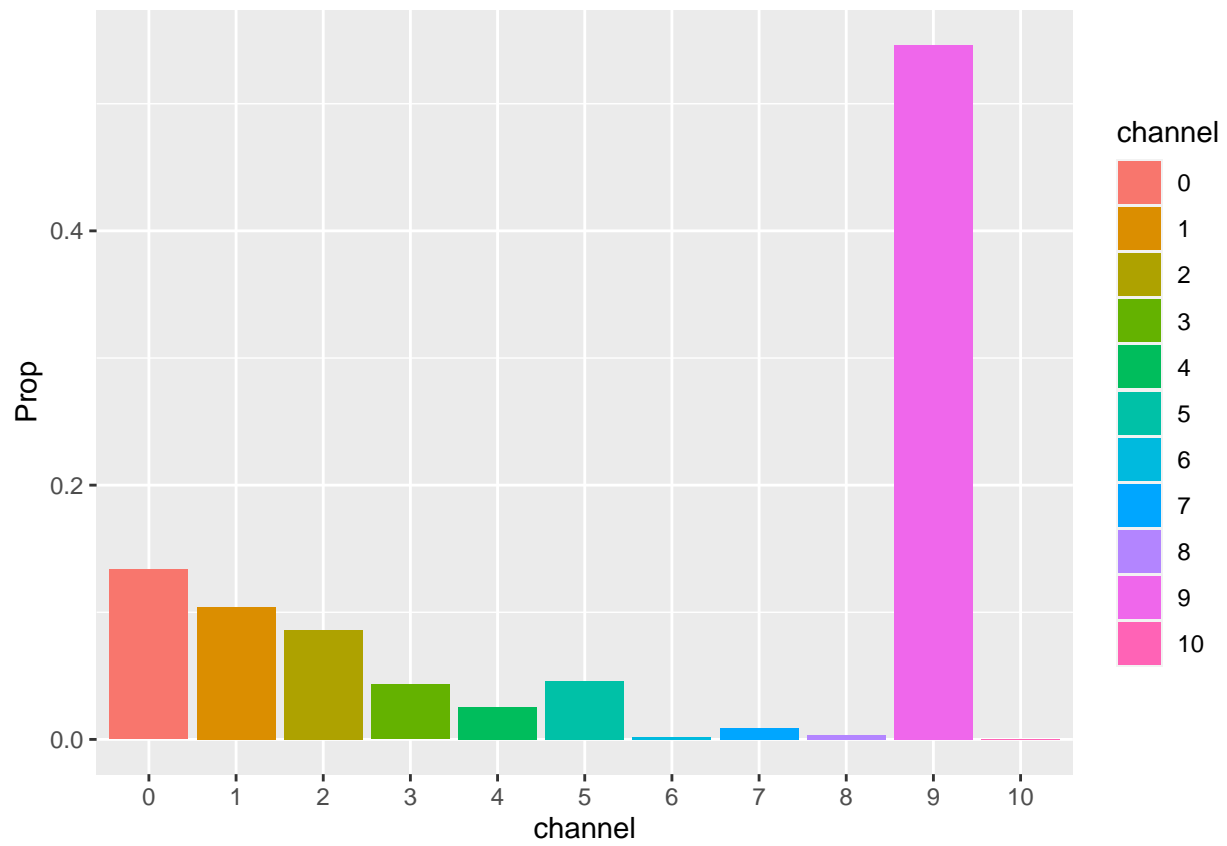


Here we can see that some hotel cluster have more visibility based on package. Let's create a proportion to see the same.

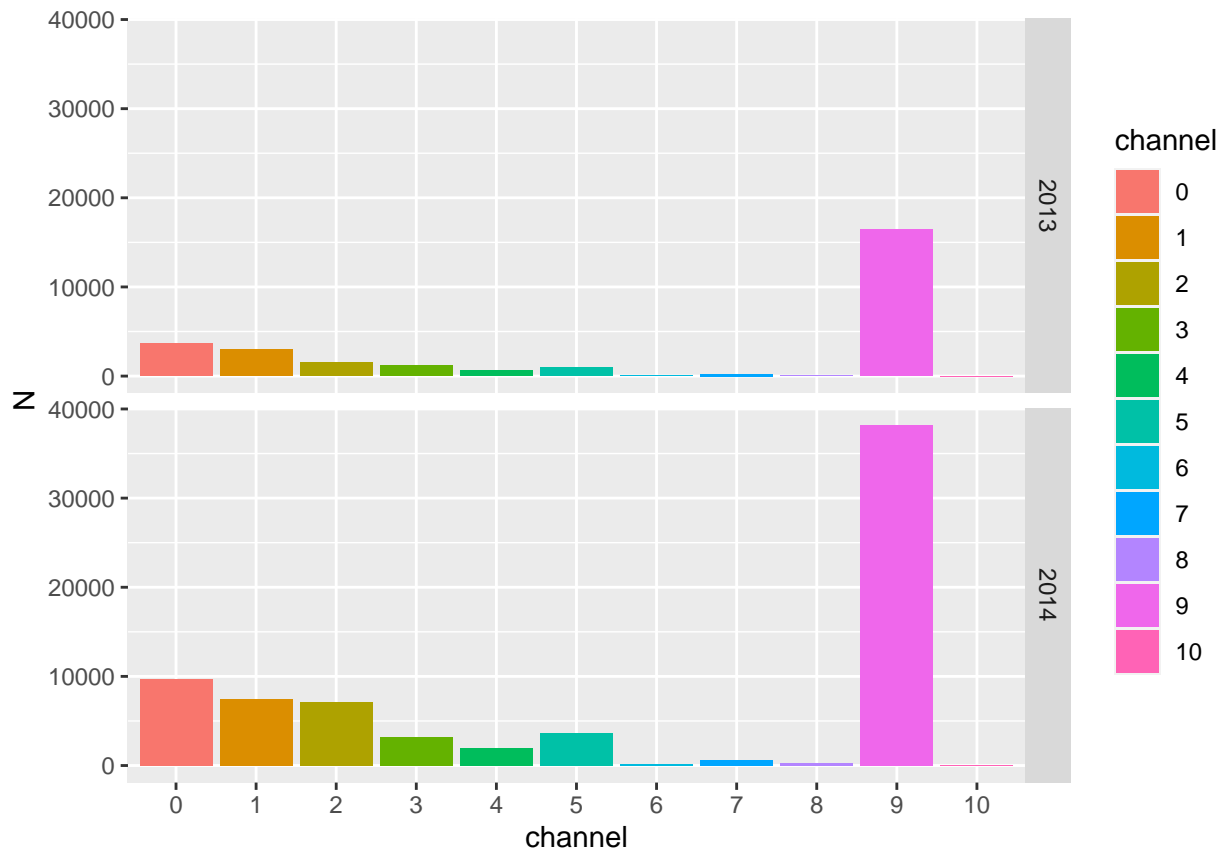


Above chart shows clearly that some hotel cluster have more visibly on package.

marketing channel



We can see in the above chart, channel 9 dominates rest of the channels proportions.



Coverience Check

Convert numeric Factor features into numerics and dropping date_time, srch_ci and srch_co, because we have already extracted required details. Also dropping week_day feature because it will not work for coverience calculation. Also I am concerting year, month and day from factor to integer. Doing all the work in separate dataset I have copied.

Updated Dataset:

```
## Classes 'data.table' and 'data.frame': 100000 obs. of 31 variables:
## $ site_name : int 2 2 2 2 2 2 2 2 2 2 ...
## $ posa_continent : int 4 4 4 4 4 4 4 4 4 4 ...
## $ user_location_country : int 66 66 66 66 66 66 66 66 66 66 ...
## $ user_location_region : int 348 348 348 442 442 442 189 189 189 189 ...
## $ user_location_city : int 48862 48862 48862 35390 35390 35390 10067 10067 10067 10067 ...
## $ orig_destination_distance: num 2234 2234 2234 913 914 ...
## $ user_id : int 12 12 12 93 93 93 501 501 501 501 ...
## $ is_mobile : int 1 1 1 1 1 1 1 1 1 1 ...
## $ is_package : int 2 2 1 1 1 1 1 2 1 1 ...
## $ channel : int 10 10 10 4 4 4 3 3 3 3 ...
## $ srch_adults_cnt : int 2 2 2 2 2 2 2 2 2 2 ...
## $ srch_children_cnt : int 0 0 0 0 0 0 0 0 0 0 ...
## $ srch_rm_cnt : int 1 1 1 1 1 1 1 1 1 1 ...
## $ srch_destination_id : int 8250 8250 8250 14984 14984 14984 8267 8267 8267 8267 ...
## $ srch_destination_type_id : int 1 1 1 1 1 1 1 1 1 1 ...
## $ is_booking : int 0 1 0 0 0 0 0 0 0 0 ...
## $ cnt : int 3 1 1 1 1 1 2 1 1 1 ...
## $ hotel_continent : int 2 2 2 2 2 2 2 2 2 2 ...
```

```
## $ hotel_country      : int  50 50 50 50 50 50 50 50 50 50 ...
## $ hotel_market       : int  628 628 628 1457 1457 1457 675 675 675 675 ...
## $ hotel_cluster      : int   1 1 1 80 21 92 41 41 69 70 ...
## $ date_year          : int   2 2 2 2 2 2 2 2 2 2 ...
## $ date_month         : int   8 8 8 8 8 8 7 7 7 7 ...
## $ date_day           : int  11 11 11 9 9 9 16 16 16 16 ...
## $ date_hour          : int   8 9 9 19 19 19 10 10 10 10 ...
## $ ci_year            : int   2 2 2 2 2 2 2 2 2 2 ...
## $ ci_month           : int   8 8 8 11 11 11 8 8 8 8 ...
## $ ci_day             : int  27 29 29 23 23 23 1 1 1 1 ...
## $ co_year            : int   2 2 2 2 2 2 2 2 2 2 ...
## $ co_month           : int   8 9 9 11 11 11 8 8 8 8 ...
## $ co_day             : int  31 2 2 28 28 28 2 2 2 2 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

Coverience Matrix Plot

Coverience plots are taking a lot of time so I manually looked at the coverience of hotel_cluster coverience agains rest of the variables.

```
##               site_name posa_continent user_location_country
## site_name      1.461901e+02 -6.966554e+00  6.486031e+01
## posa_continent -6.966554e+00  5.640030e-01  5.342680e+00
## user_location_country 6.486031e+01  5.342680e+00  2.950725e+03
## user_location_region  7.339816e+02 -3.694848e+01  2.022779e+02
## user_location_city   2.572333e+02 -1.456183e+02  9.618515e+04
## orig_destination_distance NA NA NA
## user_id           1.300355e+04 -2.418313e+03  4.361892e+04
## is_mobile         -4.612370e-03 -1.922455e-03  6.693122e-02
## is_package        6.772333e-01 -5.040129e-02 -4.258442e-01
## channel           -3.875579e+00  3.004526e-01  1.857710e+01
## srch_adults_cnt    9.633379e-02 -8.177147e-03  1.312492e+00
## srch_children_cnt  -3.697189e-02  2.986724e-03  1.070098e+00
## srch_rm_cnt        1.819160e-01 -1.671586e-02  9.928278e-02
## srch_destination_id 1.348451e+03 -7.959340e+01 -2.588348e+03
## srch_destination_type_id -1.598853e+00  1.129552e-01  1.590796e+00
## is_booking         -5.187210e-02  3.009835e-03  1.633316e-02
## cnt               4.162003e-01 -2.015580e-02  5.143502e-01
## hotel_continent    5.227358e+00 -4.305611e-01 -3.558474e+00
## hotel_country      1.643432e+02 -7.462120e+00  4.001602e+02
## hotel_market       -3.753207e+02  1.842298e+01  5.244106e+02
## hotel_cluster      -6.704607e+00  4.792746e-02 -3.958248e+01
## date_year          -1.542535e-02 -1.643510e-03 -5.753327e-01
## date_month         1.033099e+00 -6.955373e-02 -5.621443e+00
## date_day           3.135529e-01  2.037354e-02 -3.688631e+00
## date_hour          5.051837e+00 -4.236930e-01 -3.264796e+00
## ci_year           NA NA NA
## ci_month          NA NA NA
## ci_day            NA NA NA
## co_year           NA NA NA
## co_month          NA NA NA
## co_day            NA NA NA
##               user_location_region user_location_city
## site_name      7.339816e+02  2.572333e+02
## posa_continent -3.694848e+01 -1.456183e+02
```

## user_location_country	2.022779e+02	9.618515e+04
## user_location_region	4.384800e+04	4.222200e+05
## user_location_city	4.222200e+05	2.830107e+08
## orig_destination_distance	NA	NA
## user_id	6.438935e+05	-2.163802e+07
## is_mobile	3.621664e-01	-1.818167e+02
## is_package	7.275444e+00	1.542370e+02
## channel	-9.898636e+00	1.073624e+03
## srch_adults_cnt	2.463784e+00	-2.218238e+02
## srch_children_cnt	-1.194265e+00	9.936472e+01
## srch_rm_cnt	1.015824e+00	-6.313024e+01
## srch_destination_id	7.913691e+03	7.730898e+05
## srch_destination_type_id	-6.972949e+00	-1.524273e+02
## is_booking	7.345394e-01	1.607119e+00
## cnt	5.676263e-01	-1.162953e+02
## hotel_continent	3.748051e+01	2.773614e+02
## hotel_country	-1.875332e+02	-1.046419e+04
## hotel_market	4.111953e+03	-2.025890e+04
## hotel_cluster	4.872197e+01	-3.204731e+03
## date_year	-3.367795e+00	8.939766e+01
## date_month	-3.108704e+01	-6.699506e+02
## date_day	-1.782053e+01	-2.431550e+03
## date_hour	4.754750e+01	1.073084e+03
## ci_year	NA	NA
## ci_month	NA	NA
## ci_day	NA	NA
## co_year	NA	NA
## co_month	NA	NA
## co_day	NA	NA
##	orig_destination_distance	user_id is_mobile
## site_name	NA	1.300355e+04 -4.612370e-03
## posa_continent	NA	-2.418313e+03 -1.922455e-03
## user_location_country	NA	4.361892e+04 6.693122e-02
## user_location_region	NA	6.438935e+05 3.621664e-01
## user_location_city	NA	-2.163802e+07 -1.818167e+02
## orig_destination_distance	NA	NA NA
## user_id	NA	1.213828e+10 7.798152e+01
## is_mobile	NA	7.798152e+01 1.189789e-01
## is_package	NA	1.479559e+03 1.114744e-02
## channel	NA	-2.646198e+03 -3.809772e-02
## srch_adults_cnt	NA	-5.133352e+02 4.672612e-03
## srch_children_cnt	NA	-1.477800e+02 4.368209e-03
## srch_rm_cnt	NA	-1.224234e+03 -3.956021e-03
## srch_destination_id	NA	-2.434078e+07 -9.879536e+01
## srch_destination_type_id	NA	-6.020098e+03 -1.979243e-02
## is_booking	NA	2.747179e+01 -3.775119e-03
## cnt	NA	1.033958e+03 2.833435e-03
## hotel_continent	NA	-4.693285e+03 3.533974e-03
## hotel_country	NA	7.064692e+04 -3.282240e-01
## hotel_market	NA	-4.857764e+05 -7.748607e-01
## hotel_cluster	NA	4.454176e+03 -2.798753e-02
## date_year	NA	6.523168e+02 3.826433e-03
## date_month	NA	7.131062e+03 4.742587e-02
## date_day	NA	-1.220159e+04 5.621241e-02

## date_hour		NA	1.540414e+04	1.528073e-01
## ci_year		NA		NA
## ci_month		NA		NA
## ci_day		NA		NA
## co_year		NA		NA
## co_month		NA		NA
## co_day		NA		NA
##	is_package	channel	srch_adults_cnt	
## site_name	6.772333e-01	-3.875579e+00	9.633379e-02	
## posa_continent	-5.040129e-02	3.004526e-01	-8.177147e-03	
## user_location_country	-4.258442e-01	1.857710e+01	1.312492e+00	
## user_location_region	7.275444e+00	-9.898636e+00	2.463784e+00	
## user_location_city	1.542370e+02	1.073624e+03	-2.218238e+02	
## orig_destination_distance	NA	NA	NA	
## user_id	1.479559e+03	-2.646198e+03	-5.133352e+02	
## is_mobile	1.114744e-02	-3.809772e-02	4.672612e-03	
## is_package	1.926273e-01	-2.049806e-02	-9.229912e-03	
## channel	-2.049806e-02	1.422162e+01	-1.236481e-01	
## srch_adults_cnt	-9.229912e-03	-1.236481e-01	7.919281e-01	
## srch_children_cnt	-1.695459e-02	-1.320732e-02	7.583124e-02	
## srch_rm_cnt	-5.615025e-03	-6.328643e-04	2.047402e-01	
## srch_destination_id	-6.717753e+02	-1.955735e+02	2.033268e+01	
## srch_destination_type_id	-2.061700e-01	2.211929e-01	-3.095612e-02	
## is_booking	-8.860958e-03	2.526359e-02	-1.299707e-02	
## cnt	6.258587e-02	-6.896389e-02	2.071833e-02	
## hotel_continent	9.803810e-02	-1.582597e-01	-2.689440e-02	
## hotel_country	-2.816711e-01	-4.921966e+00	5.726732e-02	
## hotel_market	8.535669e-01	4.352011e+01	7.835885e+00	
## hotel_cluster	6.601585e-01	-6.502968e-01	1.886259e-01	
## date_year	1.601881e-03	-7.711950e-02	7.617759e-03	
## date_month	-2.780825e-02	-5.314486e-01	-1.439335e-02	
## date_day	-2.924842e-02	-3.583328e-01	-1.138314e-01	
## date_hour	6.926670e-02	-6.069902e-01	7.586029e-02	
## ci_year	NA	NA	NA	
## ci_month	NA	NA	NA	
## ci_day	NA	NA	NA	
## co_year	NA	NA	NA	
## co_month	NA	NA	NA	
## co_day	NA	NA	NA	
##	srch_children_cnt	srch_rm_cnt	srch_destination_id	
## site_name	-3.697189e-02	1.819160e-01	1.348451e+03	
## posa_continent	2.986724e-03	-1.671586e-02	-7.959340e+01	
## user_location_country	1.070098e+00	9.928278e-02	-2.588348e+03	
## user_location_region	-1.194265e+00	1.015824e+00	7.913691e+03	
## user_location_city	9.936472e+01	-6.313024e+01	7.730898e+05	
## orig_destination_distance	NA	NA	NA	
## user_id	-1.477800e+02	-1.224234e+03	-2.434078e+07	
## is_mobile	4.368209e-03	-3.956021e-03	-9.879536e+01	
## is_package	-1.695459e-02	-5.615025e-03	-6.717753e+02	
## channel	-1.320732e-02	-6.328643e-04	-1.955735e+02	
## srch_adults_cnt	7.583124e-02	2.047402e-01	2.033268e+01	
## srch_children_cnt	5.217762e-01	2.821475e-02	-1.320139e+02	
## srch_rm_cnt	2.821475e-02	1.953007e-01	9.867496e+01	
## srch_destination_id	-1.320139e+02	9.867496e+01	1.219497e+08	

## srch_destination_type_id	-1.220011e-02	1.829632e-02	1.035325e+04
## is_booking	-3.860771e-03	1.319723e-03	4.048650e+01
## cnt	2.325854e-02	-6.146851e-04	-2.961600e+02
## hotel_continent	-6.254851e-02	1.621705e-02	5.766938e+02
## hotel_country	-7.362145e-01	6.577123e-01	3.373037e+04
## hotel_market	-9.630824e-01	2.741849e+00	4.997534e+05
## hotel_cluster	1.143598e-01	-7.421789e-02	-2.320459e+03
## date_year	-1.431880e-02	-3.034098e-03	7.079507e+01
## date_month	-1.548784e-01	2.774381e-03	2.561803e+02
## date_day	-7.064358e-02	-5.793568e-02	2.203858e+02
## date_hour	1.441988e-01	-7.585207e-03	-1.207563e+03
## ci_year	NA	NA	NA
## ci_month	NA	NA	NA
## ci_day	NA	NA	NA
## co_year	NA	NA	NA
## co_month	NA	NA	NA
## co_day	NA	NA	NA
##	srch_destination_type_id	is_booking	cnt
## site_name	-1.598853e+00	-0.051872097	4.162003e-01
## posa_continent	1.129552e-01	0.003009835	-2.015580e-02
## user_location_country	1.590796e+00	0.016333155	5.143502e-01
## user_location_region	-6.972949e+00	0.734539434	5.676263e-01
## user_location_city	-1.524273e+02	1.607119313	-1.162953e+02
## orig_destination_distance	NA	NA	NA
## user_id	-6.020098e+03	27.471791865	1.033958e+03
## is_mobile	-1.979243e-02	-0.003775119	2.833435e-03
## is_package	-2.061700e-01	-0.008860958	6.258587e-02
## channel	2.211929e-01	0.025263595	-6.896389e-02
## srch_adults_cnt	-3.095612e-02	-0.012997071	2.071833e-02
## srch_children_cnt	-1.220011e-02	-0.003860771	2.325854e-02
## srch_rm_cnt	1.829632e-02	0.001319723	-6.146851e-04
## srch_destination_id	1.035325e+04	40.486504947	-2.961600e+02
## srch_destination_type_id	4.667569e+00	0.023702579	-6.249292e-02
## is_booking	2.370258e-02	0.075861469	-3.837196e-02
## cnt	-6.249292e-02	-0.038371963	1.433200e+00
## hotel_continent	-1.802431e-01	-0.017082079	7.197775e-02
## hotel_country	-4.470596e+00	-0.222417898	5.449068e-01
## hotel_market	3.737828e+01	1.830018553	-5.437469e+00
## hotel_cluster	-1.880390e+00	-0.202376376	3.168856e-02
## date_year	-3.679630e-02	-0.004938917	-3.390161e-02
## date_month	-2.467221e-01	-0.011957467	-1.950356e-01
## date_day	-1.473991e-01	-0.004250739	3.441891e-02
## date_hour	-3.512442e-02	-0.020589437	9.571720e-02
## ci_year	NA	NA	NA
## ci_month	NA	NA	NA
## ci_day	NA	NA	NA
## co_year	NA	NA	NA
## co_month	NA	NA	NA
## co_day	NA	NA	NA
##	hotel_continent	hotel_country	hotel_market
## site_name	5.227358e+00	1.643432e+02	-3.753207e+02
## posa_continent	-4.305611e-01	-7.462120e+00	1.842298e+01
## user_location_country	-3.558474e+00	4.001602e+02	5.244106e+02
## user_location_region	3.748051e+01	-1.875332e+02	4.111953e+03

## user_location_city	2.773614e+02	-1.046419e+04	-2.025890e+04
## orig_destination_distance	NA	NA	NA
## user_id	-4.693285e+03	7.064692e+04	-4.857764e+05
## is_mobile	3.533974e-03	-3.282240e-01	-7.748607e-01
## is_package	9.803810e-02	-2.816711e-01	8.535669e-01
## channel	-1.582597e-01	-4.921966e+00	4.352011e+01
## srch_adults_cnt	-2.689440e-02	5.726732e-02	7.835885e+00
## srch_children_cnt	-6.254851e-02	-7.362145e-01	-9.630824e-01
## srch_rm_cnt	1.621705e-02	6.577123e-01	2.741849e+00
## srch_destination_id	5.766938e+02	3.373037e+04	4.997534e+05
## srch_destination_type_id	-1.802431e-01	-4.470596e+00	3.737828e+01
## is_booking	-1.708208e-02	-2.224179e-01	1.830019e+00
## cnt	7.197775e-02	5.449068e-01	-5.437469e+00
## hotel_continent	2.575324e+00	2.858335e+01	-6.339839e+01
## hotel_country	2.858335e+01	3.100247e+03	8.144643e+02
## hotel_market	-6.339839e+01	8.144643e+02	2.576828e+05
## hotel_cluster	-2.984685e-01	-3.412569e+01	5.306335e+02
## date_year	1.310210e-02	1.081024e-01	-3.253917e-01
## date_month	-9.279581e-02	3.702829e+00	-3.165622e+01
## date_day	-2.653886e-03	-7.728172e-01	-2.349387e+01
## date_hour	1.583299e-01	1.101404e-01	1.023353e+01
## ci_year	NA	NA	NA
## ci_month	NA	NA	NA
## ci_day	NA	NA	NA
## co_year	NA	NA	NA
## co_month	NA	NA	NA
## co_day	NA	NA	NA
##	hotel_cluster	date_year	date_month
## site_name	-6.704607e+00	-0.015425352	1.033099e+00
## posa_continent	4.792746e-02	-0.001643510	-6.955373e-02
## user_location_country	-3.958248e+01	-0.575332707	-5.621443e+00
## user_location_region	4.872197e+01	-3.367795119	-3.108704e+01
## user_location_city	-3.204731e+03	89.397660323	-6.699506e+02
## orig_destination_distance	NA	NA	NA
## user_id	4.454176e+03	652.316817741	7.131062e+03
## is_mobile	-2.798753e-02	0.003826433	4.742587e-02
## is_package	6.601585e-01	0.001601881	-2.780825e-02
## channel	-6.502968e-01	-0.077119505	-5.314486e-01
## srch_adults_cnt	1.886259e-01	0.007617759	-1.439335e-02
## srch_children_cnt	1.143598e-01	-0.014318798	-1.548784e-01
## srch_rm_cnt	-7.421789e-02	-0.003034098	2.774381e-03
## srch_destination_id	-2.320459e+03	70.795065625	2.561803e+02
## srch_destination_type_id	-1.880390e+00	-0.036796302	-2.467221e-01
## is_booking	-2.023764e-01	-0.004938917	-1.195747e-02
## cnt	3.168856e-02	-0.033901606	-1.950356e-01
## hotel_continent	-2.984685e-01	0.013102097	-9.279581e-02
## hotel_country	-3.412569e+01	0.108102440	3.702829e+00
## hotel_market	5.306335e+02	-0.325391746	-3.165622e+01
## hotel_cluster	8.381390e+02	0.009014932	1.150195e-01
## date_year	9.014932e-03	0.200344338	1.708733e-01
## date_month	1.150195e-01	0.170873296	1.082293e+01
## date_day	-4.589674e-01	-0.023290396	-6.289824e-01
## date_hour	-4.730215e-01	0.002121236	-2.748312e-01
## ci_year	NA	NA	NA

## ci_month	NA	NA	NA		
## ci_day	NA	NA	NA		
## co_year	NA	NA	NA		
## co_month	NA	NA	NA		
## co_day	NA	NA	NA		
##	date_day	date_hour	ci_year	ci_month	ci_day
## site_name	3.135529e-01	5.051837e+00	NA	NA	NA
## posa_continent	2.037354e-02	-4.236930e-01	NA	NA	NA
## user_location_country	-3.688631e+00	-3.264796e+00	NA	NA	NA
## user_location_region	-1.782053e+01	4.754750e+01	NA	NA	NA
## user_location_city	-2.431550e+03	1.073084e+03	NA	NA	NA
## orig_destination_distance	NA	NA	NA	NA	NA
## user_id	-1.220159e+04	1.540414e+04	NA	NA	NA
## is_mobile	5.621241e-02	1.528073e-01	NA	NA	NA
## is_package	-2.924842e-02	6.926670e-02	NA	NA	NA
## channel	-3.583328e-01	-6.069902e-01	NA	NA	NA
## srch_adults_cnt	-1.138314e-01	7.586029e-02	NA	NA	NA
## srch_children_cnt	-7.064358e-02	1.441988e-01	NA	NA	NA
## srch_rm_cnt	-5.793568e-02	-7.585207e-03	NA	NA	NA
## srch_destination_id	2.203858e+02	-1.207563e+03	NA	NA	NA
## srch_destination_type_id	-1.473991e-01	-3.512442e-02	NA	NA	NA
## is_booking	-4.250739e-03	-2.058944e-02	NA	NA	NA
## cnt	3.441891e-02	9.571720e-02	NA	NA	NA
## hotel_continent	-2.653886e-03	1.583299e-01	NA	NA	NA
## hotel_country	-7.728172e-01	1.101404e-01	NA	NA	NA
## hotel_market	-2.349387e+01	1.023353e+01	NA	NA	NA
## hotel_cluster	-4.589674e-01	-4.730215e-01	NA	NA	NA
## date_year	-2.329040e-02	2.121236e-03	NA	NA	NA
## date_month	-6.289824e-01	-2.748312e-01	NA	NA	NA
## date_day	7.751487e+01	-2.164003e-01	NA	NA	NA
## date_hour	-2.164003e-01	3.063032e+01	NA	NA	NA
## ci_year	NA	NA	NA	NA	NA
## ci_month	NA	NA	NA	NA	NA
## ci_day	NA	NA	NA	NA	NA
## co_year	NA	NA	NA	NA	NA
## co_month	NA	NA	NA	NA	NA
## co_day	NA	NA	NA	NA	NA
##	co_year	co_month	co_day		
## site_name	NA	NA	NA		
## posa_continent	NA	NA	NA		
## user_location_country	NA	NA	NA		
## user_location_region	NA	NA	NA		
## user_location_city	NA	NA	NA		
## orig_destination_distance	NA	NA	NA		
## user_id	NA	NA	NA		
## is_mobile	NA	NA	NA		
## is_package	NA	NA	NA		
## channel	NA	NA	NA		
## srch_adults_cnt	NA	NA	NA		
## srch_children_cnt	NA	NA	NA		
## srch_rm_cnt	NA	NA	NA		
## srch_destination_id	NA	NA	NA		
## srch_destination_type_id	NA	NA	NA		
## is_booking	NA	NA	NA		

## cnt	NA	NA	NA
## hotel_continent	NA	NA	NA
## hotel_country	NA	NA	NA
## hotel_market	NA	NA	NA
## hotel_cluster	NA	NA	NA
## date_year	NA	NA	NA
## date_month	NA	NA	NA
## date_day	NA	NA	NA
## date_hour	NA	NA	NA
## ci_year	NA	NA	NA
## ci_month	NA	NA	NA
## ci_day	NA	NA	NA
## co_year	NA	NA	NA
## co_month	NA	NA	NA
## co_day	NA	NA	NA

Analysis Result

We can see over the year usage of all the different channels increased a lot. we can see channel specially which got an increase of 100%. Also based on coverience we can see target variable is not corrected to any other dependent field. Based on this we can say linear model will not work for this dataset.