Take home users:

Take home user has 12000 records for the information of user behavior in the time of period. It is having the details of user name , organization id , site visited , etc.

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 12000 entries, 0 to 11999

Data columns (total 10 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 object\_id 12000 non-null int64

1 creation\_time 12000 non-null datetime64[ns]

2 name 12000 non-null object

3 email 12000 non-null object

4 creation\_source 12000 non-null object

5 last\_session\_creation\_time 8823 non-null float64

6 opted\_in\_to\_mailing\_list 12000 non-null int64

7 enabled\_for\_marketing\_drip 12000 non-null int64

8 org\_id 12000 non-null int64

9 invited\_by\_user\_id 6417 non-null float64

dtypes: datetime64[ns](1), float64(2), int64(4), object(3)

takehome\_user\_eng:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 207917 entries, 0 to 207916

Data columns (total 3 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 time\_stamp 207917 non-null datetime64[ns]

1 user\_id 207917 non-null int64

2 visited 207917 non-null int64

dtypes: datetime64[ns](1), int64(2)

takehome\_users.describe().T

basic statistics to the take home user of average and max , min user inference as below:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **mean** | **std** | **min** | **0.25** | **0.50** | **0.75** | **max** |
| **object\_id** | 12,000.00 | 6,000.50 | 3,464.25 | 1.00 | 3,000.75 | 6,000.50 | 9,000.25 | 12,000.00 |
| **last\_session\_creation\_time** | 8,823.00 | 1,379,279,000.00 | 19,531,160.00 | 1,338,452,000.00 | 1,363,195,000.00 | 1,382,888,000.00 | 1,398,443,000.00 | 1,402,067,000.00 |
| **opted\_in\_to\_mailing\_list** | 12,000.00 | 0.25 | 0.43 | - | - | - | - | 1.00 |
| **enabled\_for\_marketing\_drip** | 12,000.00 | 0.15 | 0.36 | - | - | - | - | 1.00 |
| **org\_id** | 12,000.00 | 141.88 | 124.06 | - | 29.00 | 108.00 | 238.25 | 416.00 |
| **invited\_by\_user\_id** | 6,417.00 | 5,962.96 | 3,383.76 | 3.00 | 3,058.00 | 5,954.00 | 8,817.00 | 11,999.00 |

takehome\_user\_eng.describe().T

Take home user engagement details of average and max & min user login details with inference:

count mean std min 25% 50% 75% max

user\_id 207917.0 5913.314197 3394.941674 1.0 3087.0 5682.0 8944.0 12000.0

visited 207917.0 1.000000 0.000000 1.0 1.0 1.0 1.0 1.0

df.describe().T

The user who is visited 7 days of time period with average , max & min of basic details.

|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **user\_id** | 207917.0 | 5913.314197 | 3394.941674 | 1.0 | 3087.0 | 5682.0 | 8944.0 | 12000.0 |
| **visited** | 207917.0 | 1.000000 | 0.000000 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **visits\_7\_days** | 207917.0 | 4.381393 | 2.042666 | 1.0 | 3.0 | 4.0 | 6.0 | 7.0 |

df[df.visits\_7\_days >= 3.0]

| **time\_stamp** | **user\_id** | **visited** | **date** | **visits\_7\_days** |
| --- | --- | --- | --- | --- |
| **9** | 2014-02-09 03:45:04 | 2 | 1 | 2014-02-09 | 3.0 |
| **10** | 2014-02-13 03:45:04 | 2 | 1 | 2014-02-13 | 3.0 |
| **27** | 2013-02-19 22:08:03 | 10 | 1 | 2013-02-19 | 3.0 |
| **30** | 2013-03-02 22:08:03 | 10 | 1 | 2013-03-02 | 3.0 |
| **31** | 2013-03-05 22:08:03 | 10 | 1 | 2013-03-05 | 3.0 |

If any of the values in the visits\_7\_days column is >=3, then we can label that user as adopted.

user\_id visits\_7\_days

0 1 1.0

1 2 3.0

2 3 1.0

3 4 1.0

4 5 1.0

Result : The user visited 7days period with 3 time in the over all user.

Adopted user counts: The user who is adopted and not adopted users count:

0 7221

1 1602

Name: adopted\_user, dtype: int64

Adopted user:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 12000 entries, 0 to 11999

Data columns (total 11 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 object\_id 12000 non-null int64

1 creation\_time 12000 non-null datetime64[ns]

2 name 12000 non-null object

3 email 12000 non-null object

4 creation\_source 12000 non-null object

5 last\_session\_creation\_time 8823 non-null datetime64[ns]

6 opted\_in\_to\_mailing\_list 12000 non-null int64

7 enabled\_for\_marketing\_drip 12000 non-null int64

8 org\_id 12000 non-null int64

9 invited\_by\_user\_id 6417 non-null float64

10 adopted\_user 8823 non-null float64

dtypes: datetime64[ns](2), float64(2), int64(4), object(3)

df\_users.describe().T

count mean std min 25% 50% 75% max

object\_id 12000.0 6000.500000 3464.245950 1.0 3000.75 6000.5 9000.25 12000.0

opted\_in\_to\_mailing\_list 12000.0 0.249500 0.432742 0.0 0.00 0.0 0.00 1.0

enabled\_for\_marketing\_drip 12000.0 0.149333 0.356432 0.0 0.00 0.0 0.00 1.0

org\_id 12000.0 141.884583 124.056723 0.0 29.00 108.0 238.25 416.0

invited\_by\_user\_id 6417.0 5962.957145 3383.761968 3.0 3058.00 5954.0 8817.00 11999.0

adopted\_user 12000.0 0.133500 0.340128 0.0 0.00 0.0 0.00 1.0

The site wise user statistics:

gmail.com 3562

yahoo.com 2447

jourrapide.com 1259

cuvox.de 1202

gustr.com 1179

...

lrabg.com 1

mrytw.com 1

ugtav.com 1

hqhll.com 1

iuxiw.com 1

Name: email\_domain, Length: 1184, dtype: int64

count mean std min 25% 50% 75% max

opted\_in\_to\_mailing\_list 12000.0 0.249500 0.432742 0.0 0.0 0.0 0.00 1.0

enabled\_for\_marketing\_drip 12000.0 0.149333 0.356432 0.0 0.0 0.0 0.00 1.0

org\_id 12000.0 141.884583 124.056723 0.0 29.0 108.0 238.25 416.0

invited\_by\_user\_id 12000.0 3188.691333 3869.027693 0.0 0.0 875.0 6317.00 11999.0

adopted\_user 12000.0 0.133500 0.340128 0.0 0.0 0.0 0.00 1.0

days\_since\_creation 12000.0 324.568000 216.646173 6.0 129.0 304.0 506.00 736.0

creation\_source\_GUEST\_INVITE 12000.0 0.180250 0.384412 0.0 0.0 0.0 0.00 1.0

creation\_source\_ORG\_INVITE 12000.0 0.354500 0.478381 0.0 0.0 0.0 1.00 1.0

creation\_source\_PERSONAL\_PROJECTS 12000.0 0.175917 0.380765 0.0 0.0 0.0 0.00 1.0

creation\_source\_SIGNUP 12000.0 0.173917 0.379054 0.0 0.0 0.0 0.00 1.0

creation\_source\_SIGNUP\_GOOGLE\_AUTH 12000.0 0.115417 0.319537 0.0 0.0 0.0 0.00 1.0

Correlation between the various parameter range to infer the associate relationship between different variables.

opted\_in\_to\_mailing\_list 0.008097

enabled\_for\_marketing\_drip 0.004653

org\_id 0.064523

invited\_by\_user\_id 0.021596

adopted\_user 1.000000

days\_since\_creation 0.088020

creation\_source\_GUEST\_INVITE 0.045408

creation\_source\_ORG\_INVITE -0.007636

creation\_source\_PERSONAL\_PROJECTS -0.075817

creation\_source\_SIGNUP 0.009299

creation\_source\_SIGNUP\_GOOGLE\_AUTH 0.036119

Business visualization of adopted user with many variables for comparison to conclude the adopted user of the given 7 days period.















