Directing Customers to Subscription Through App Behavior Analysis

Import essential libraries

In [ ]:

```
import numpy as np # for numeric calculation
import pandas as pd # for data analysis and manupulation
import matplotlib.pyplot as plt # for data visualization
import seaborn as sns # for data visualization
from dateutil import parser # convert time in date time data type
```

Import dataset & explore

In [ ]:

```
fineTech_appData = pd.read_csv("FineTech_appData.csv")
```

In [ ]:

```
fineTech_appData.shape
```

Out[ ]:

```
(50000, 12)
```

In [ ]:

```
fineTech_appData.head(6) # show fisrt 6 rows of fineTech_appData DataFrame  *****code 1
```

Out[ ]:

	user	first_open	dayofweek	hour	age	screen_list	numscreens	minigame
0	235136	2012-12-27 02:14:51.273	3	02:00:00	23	idscreen,joinscreen,Cycle,product_review,ScanP...	15	0
1	333588	2012-12-02 01:16:00.905	6	01:00:00	24	joinscreen,product_review,product_review2,Scan...	13	0
2	254414	2013-03-19 19:19:09.157	1	19:00:00	23	Splash,Cycle,Loan	3	0
3	234192	2013-07-05 16:08:46.354	4	16:00:00	28	product_review,Home,product_review,Loan3,Finan...	40	0
4	51549	2013-02-26 18:50:48.661	1	18:00:00	31	idscreen,joinscreen,Cycle,Credit3Container,Sca...	32	0
5	56480	2013-04-03 09:58:15.752	2	09:00:00	20	idscreen,Cycle,Home,ScanPreview,VerifyPhone,Ve...	14	0

In [ ]:

```
fineTech_appData.tail(6) # show last 6 rows of fineTech_appData DataFrame  *****code 2
```

Out[ ]:

	user	first_open	dayofweek	hour	age	screen_list	numscreens	minig
49994	90813	2013-02-25 19:35:12.691	0	19:00:00	36	idscreen,joinscreen,Cycle,product_review,produ...	25	

	user	first_open	dayofweek	hour	age	screen_list	numscreens	minig
49995	222774	2013-05-09 13:48:17.871	6	13:00:00	33	Splash,Home,ScanPreview,VerifyPhone,VerifySSN,...	13	
49996	169179	2013-04-09 00:05:17.823	1	00:00:00	35	Cycle,Splash,Home,RewardsContainer	4	
49997	302367	2013-02-20 22:41:51.165	2	22:00:00	39	joinscreen,product_review,product_review2,Scan...	25	
49998	324905	2013-04-28 12:33:04.288	6	12:00:00	27	Cycle,Home,product_review,product_review,produ...	26	
49999	27047	2012-12-14 01:22:44.638	4	01:00:00	25	product_review,ScanPreview,VerifyDateOfBirth,V...	26	

In [ ]:

```
for i in [1,2,3,4,5]:
    print(fineTech_appData.loc[i,'screen_list'],'\n')
```

```
joinscreen,product_review,product_review2,ScanPreview,VerifyDateOfBirth,location,VerifyCo
untry,VerifyPhone,VerifyToken,Institutions,Loan2

Splash,Cycle,Loan

product_review,Home,product_review,Loan3,Finances,Credit3,ReferralContainer,Leaderboard,R
ewards,RewardDetail,ScanPreview,location,VerifyDateOfBirth,VerifyPhone,VerifySSN,Credit1,
Credit2

idscreen,joinscreen,Cycle,Credit3Container,ScanPreview,VerifyPhone,VerifySSN,Credit1,Loan
2,Home,Institutions,SelectInstitution,BankVerification,ReferralContainer,product_review,p
roduct_review2,VerifyCountry,VerifyToken,product_review

idscreen,Cycle,Home,ScanPreview,VerifyPhone,VerifySSN,Credit1,Credit3Dashboard,Loan2,Inst
itutions,product_review,product_review,product_review3
```

In [ ]:

```
fineTech_appData.isnull().sum() # take summation of null values
```

Out[ ]:

```
user                       0
first_open                 0
dayofweek                  0
hour                       0
age                        0
screen_list                0
numscreens                 0
minigame                   0
used_premium_feature       0
enrolled                   0
enrolled_date          18926
liked                      0
dtype: int64
```

In [ ]:

```
fineTech_appData.info() # brief inforamtion about Dataset
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 12 columns):
user                    50000 non-null int64
first_open              50000 non-null object
dayofweek               50000 non-null int64
hour                    50000 non-null object
age                     50000 non-null int64
screen_list             50000 non-null object
numscreens              50000 non-null int64
minigame                50000 non-null int64
used_premium_feature    50000 non-null int64
```

```
enrolled                50000 non-null int64
enrolled_date           31074 non-null object
liked                   50000 non-null int64
dtypes: int64(8), object(4)
memory usage: 4.6+ MB
```

In [ ]:

```
fineTech_appData.describe() # give the distribution of numerical variables  *****code 3
```

Out[ ]:

	user	dayofweek	age	numscreens	minigame	used_premium_feature	enrolled	l
count	50000.000000	50000.000000	50000.00000	50000.000000	50000.000000	50000.000000	50000.000000	50000.000
mean	186889.729900	3.029860	31.72436	21.095900	0.107820	0.172020	0.621480	0.165
std	107768.520361	2.031997	10.80331	15.728812	0.310156	0.377402	0.485023	0.371
min	13.000000	0.000000	16.00000	1.000000	0.000000	0.000000	0.000000	0.000
25%	93526.750000	1.000000	24.00000	10.000000	0.000000	0.000000	0.000000	0.000
50%	187193.500000	3.000000	29.00000	18.000000	0.000000	0.000000	1.000000	0.000
75%	279984.250000	5.000000	37.00000	28.000000	0.000000	0.000000	1.000000	0.000
max	373662.000000	6.000000	101.00000	325.000000	1.000000	1.000000	1.000000	1.000

In [ ]:

```
# Get the unique value of each columns and it's length
features = fineTech_appData.columns
for i in features:
    print("""Unique value of {}\n{}\nlen is {} \n.....................\n
          """.format(i, fineTech_appData[i].unique(), len(fineTech_appData[i].unique()))
)
```

```
Unique value of user
[235136 333588 254414 ... 302367 324905  27047]
len is 49874
.....................


Unique value of first_open
['2012-12-27 02:14:51.273' '2012-12-02 01:16:00.905'
 '2013-03-19 19:19:09.157' ... '2013-02-20 22:41:51.165'
 '2013-04-28 12:33:04.288' '2012-12-14 01:22:44.638']
len is 49747
.....................


Unique value of dayofweek
[3 6 1 4 2 0 5]
len is 7
.....................


Unique value of hour
[' 02:00:00' ' 01:00:00' ' 19:00:00' ' 16:00:00' ' 18:00:00' ' 09:00:00'
 ' 03:00:00' ' 14:00:00' ' 04:00:00' ' 11:00:00' ' 06:00:00' ' 21:00:00'
 ' 05:00:00' ' 17:00:00' ' 20:00:00' ' 00:00:00' ' 22:00:00' ' 10:00:00'
 ' 08:00:00' ' 15:00:00' ' 13:00:00' ' 23:00:00' ' 12:00:00' ' 07:00:00']
len is 24
.....................


Unique value of age
[ 23  24  28  31  20  35  26  29  39  32  25  17  21  55  38  27  48  37
  22  36  30  58  40  33  57  19  45  34  46  56  42  43  41  47  18  53
  44  49  60  50  52  62  63  16  54  70  51  69  68  59  76  75  66  61
  72  65  90  64  67  73  77  71  74  89  78  86  80  82  79  87  81  85
 101  88  83 100  84  98]
```

```
 101  88  83 100  84  96]
len is 78
.....................


Unique value of screen_list
['idscreen,joinscreen,Cycle,product_review,ScanPreview,VerifyDateOfBirth,VerifyPhone,Veri
fyToken,ProfileVerifySSN,Loan2,Settings,ForgotPassword,Login'
 'joinscreen,product_review,product_review2,ScanPreview,VerifyDateOfBirth,location,Verify
Country,VerifyPhone,VerifyToken,Institutions,Loan2'
 'Splash,Cycle,Loan' ...
 'joinscreen,product_review,product_review2,ScanPreview,VerifyCountry,VerifyPhone,VerifyT
oken,VerifyDateOfBirth,location,Home'
 'Cycle,Home,product_review,product_review,product_review3,ScanPreview,VerifyDateOfBirth,
location,VerifyCountry,VerifyPhone,VerifyToken,product_review,product_review,VerifySSN,pr
oduct_review,SelectInstitution,BankVerification,product_review,product_review'
 'product_review,ScanPreview,VerifyDateOfBirth,VerifyCountry,ProfileVerifySSN,ProfilePage
,ProfileEducation,ProfileEducationMajor,Saving2Amount,Saving8,ProfileMaritalStatus,Profil
eChildren,Saving2,Saving9,Saving7,Saving6,Saving5,Home,Loan2']
len is 38799
.....................


Unique value of numscreens
[ 15  13   3  40  32  14  41  33  19  25  11   4   9  26   6  20   5   8
  42   1  38  49  35  10  52  50  76  37  16  47  90  24  45  31  39  17
  28  27  57  23  21  12   7  18  48  29 136  34  59  89  22  43  36  56
  30   2  44  92  51  70  58  66  46  55  61  75  71  78  85  62  53  54
  73  68  69  63  64  88 106  80 127  74  72 137  83  77  65 104  60  67
  94  81 110  91  82  96 165  79  86 116  99  98 187  84 111 109 107 162
  97 100  95  87 122 216 115 102 128 234 112 108 114 125 119  93 185 192
 189 153 243 103 101 118 325 141 129 133 126 120 123 134 121 105 113 117
 200 247 179 132 144 130 148]
len is 151
.....................


Unique value of minigame
[0 1]
len is 2
.....................


Unique value of used_premium_feature
[0 1]
len is 2
.....................


Unique value of enrolled
[0 1]
len is 2
.....................


Unique value of enrolled_date
[nan '2013-07-05 16:11:49.513' '2013-02-26 18:56:37.841' ...
 '2013-02-25 19:36:56.082' '2013-05-09 13:47:52.875'
 '2013-04-28 12:35:38.709']
len is 31002
.....................


Unique value of liked
[0 1]
len is 2
.....................
```

In [ ]:

```
fineTech_appData.dtypes
```

Out[ ]:

```
user                     int64
first_open              object
dayofweek                int64
hour                    object
age                      int64
screen_list             object
numscreens               int64
minigame                 int64
used_premium_feature     int64
enrolled                 int64
enrolled_date           object
liked                    int64
dtype: object
```

In [ ]:

```
#  hour data convert string to int
fineTech_appData['hour'] = fineTech_appData.hour.str.slice(1,3).astype(int)
```

In [ ]:

```
# get data type of each columns
fineTech_appData.dtypes
```

Out[ ]:

```
user                     int64
first_open              object
dayofweek                int64
hour                     int32
age                      int64
screen_list             object
numscreens               int64
minigame                 int64
used_premium_feature     int64
enrolled                 int64
enrolled_date           object
liked                    int64
dtype: object
```

In [ ]:

```
fineTech_appData.columns
```

Out[ ]:

```
Index(['user', 'first_open', 'dayofweek', 'hour', 'age', 'screen_list',
       'numscreens', 'minigame', 'used_premium_feature', 'enrolled',
       'enrolled_date', 'liked'],
      dtype='object')
```

In [ ]:

```
# drop object dtype columns
fineTech_appData2 = fineTech_appData.drop(['user', 'first_open', 'screen_list', 'enrolled_date'], axis = 1)
```

In [ ]:

```
fineTech_appData2.head(6) # head of numeric dataFrame *****code 4
```

Out[ ]:

	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked
0	3	2	23	15	0	0	0	0
1	6	1	24	13	0	0	0	0
2	1	19	23	3	0	1	0	1

	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked
3	4	16	28	40	0	0	1	0
4	1	18	31	32	0	0	1	1
5	2	9	20	14	0	0	1	0

Data Visualization

Heatmap Using Correlation matrix

In [ ]:

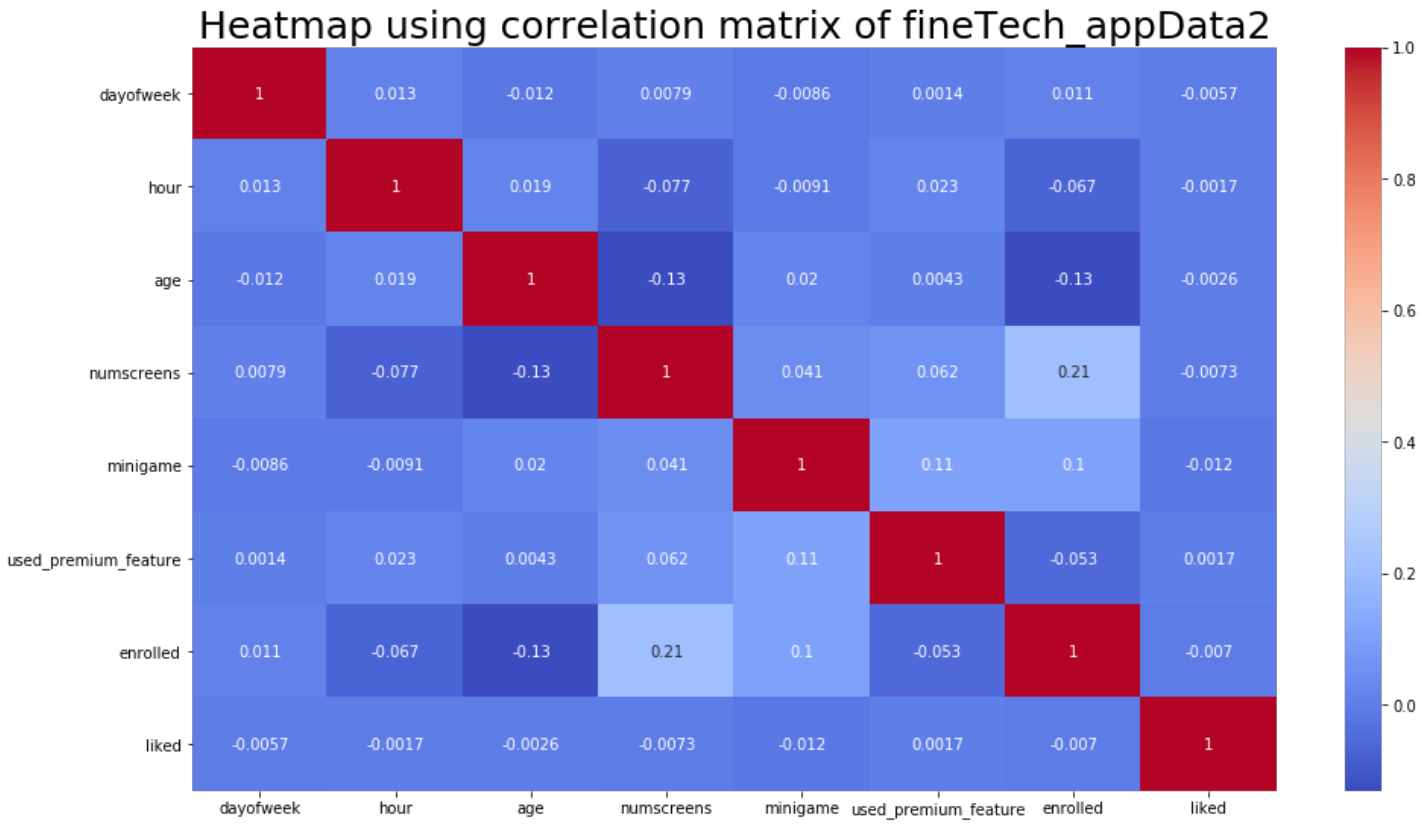
```
# Heatmap
plt.figure(figsize=(16,9)) # heatmap size in ratio 16:9

sns.heatmap(fineTech_appData2.corr(), annot = True, cmap ='coolwarm') # show heatmap

plt.title("Heatmap using correlation matrix of fineTech_appData2", fontsize = 25) # title of heatmap *****code 5
```

Out[ ]:

```
Text(0.5, 1.0, 'Heatmap using correlation matrix of fineTech_appData2')
```



Pairplot of fineTech_appData2

In [ ]:

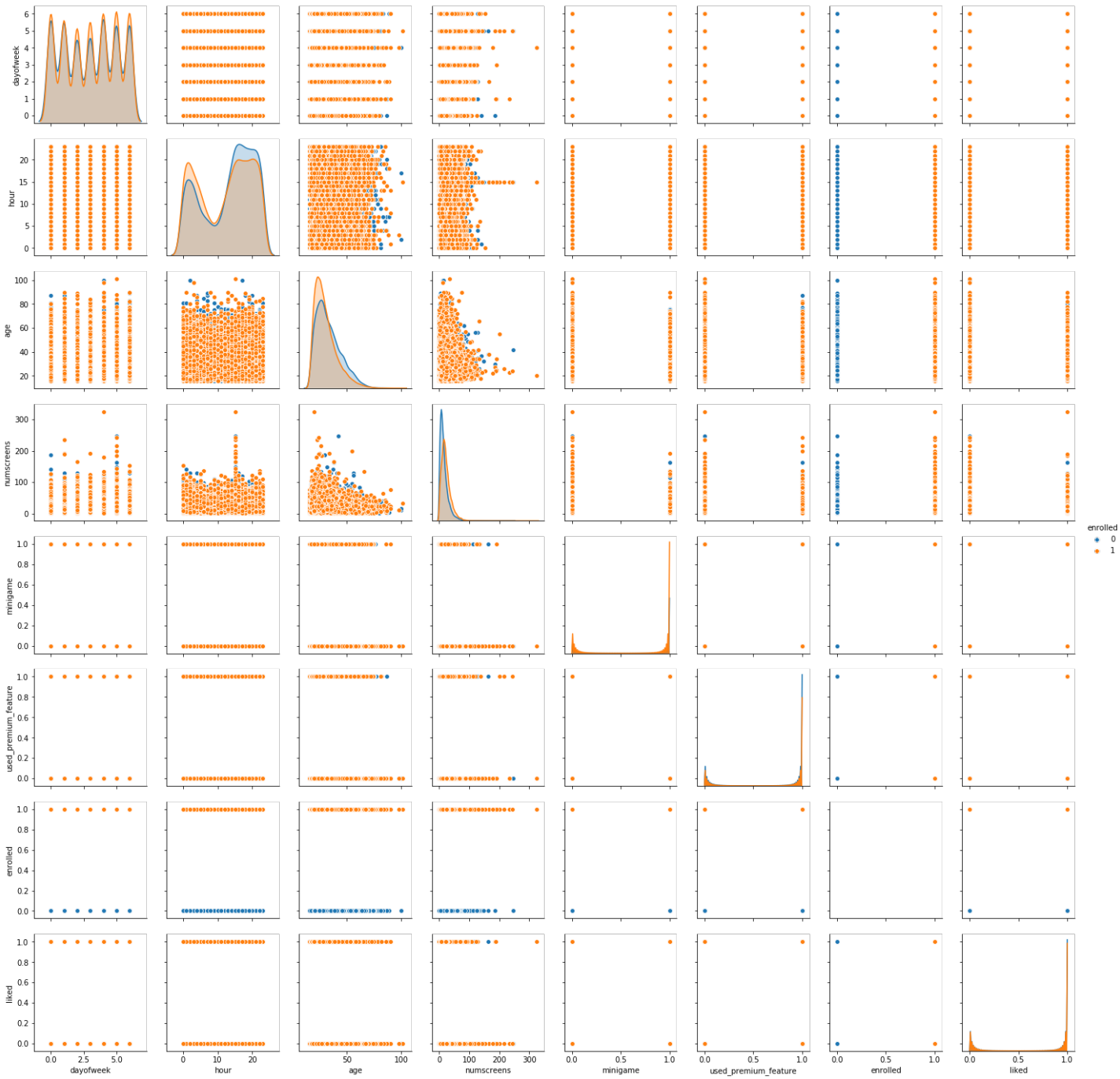
```
# Pailplot of fineTech_appData2 Dataset

#%matplotlib qt5 # for show graph in seperate window
sns.pairplot(fineTech_appData2, hue  = 'enrolled') # *****code 6
```

```
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\nonparametric\kde.py:488: RuntimeW
arning: invalid value encountered in true_divide
  binned = fast_linbin(X, a, b, gridsize) / (delta * nobs)
C:\ProgramData\Anaconda3\lib\site-packages\statsmodels\nonparametric\kdetools.py:34: Runt
imeWarning: invalid value encountered in double_scalars
  FAC1 = 2*(np.pi*bw/RANGE)**2
```

Out[ ]:

```
<seaborn.axisgrid.PairGrid at 0x246e12c6eb8>
```



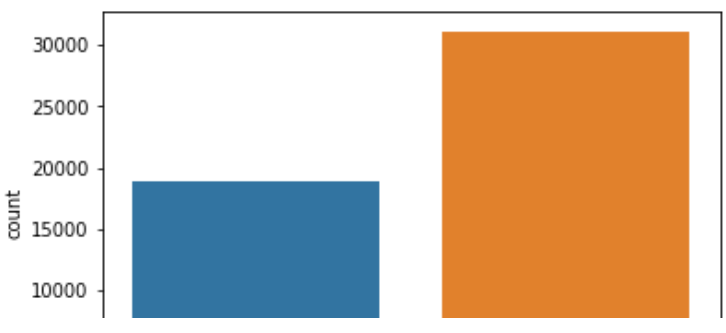
Countplot of enrolled

In [ ]:

```
# Show counterplot of 'enrolled' feature
sns.countplot(fineTech_appData.enrolled) # *****code 7
```

Out[ ]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x246e34a9438>
```



In [ ]:

```
# value enrolled and not enrolled customers
print("Not enrolled user = ", (fineTech_appData.enrolled < 1).sum(), "out of 50000")
print("Enrolled user = ",50000-(fineTech_appData.enrolled < 1).sum(),  "out of 50000")
```

```
Not enrolled user =  18926 out of 50000
Enrolled user =  31074 out of 50000
```

Histogram of each feature of fineTech_appData2

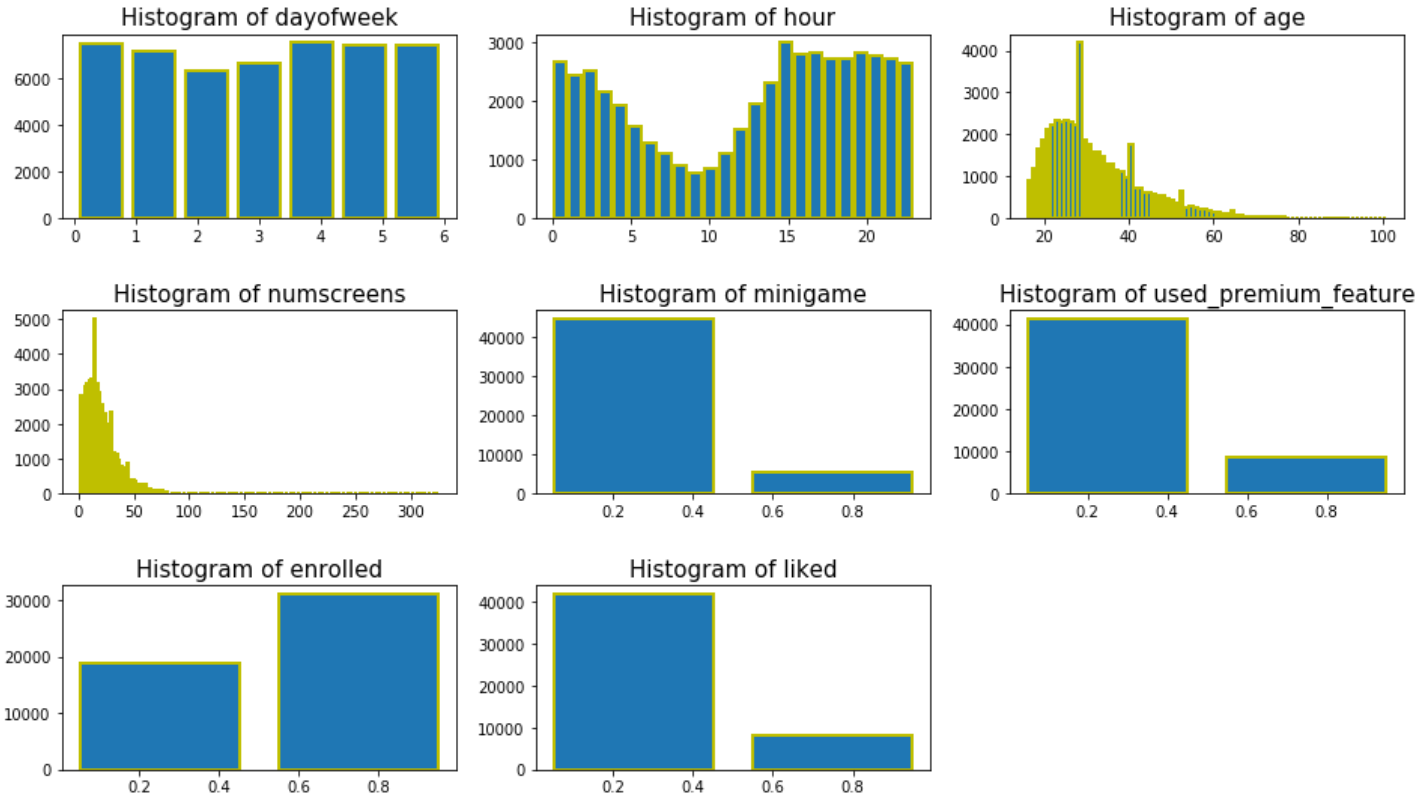
In [ ]:

```
# plot histogram

plt.figure(figsize = (16,9)) # figure size in ratio 16:9
features = fineTech_appData2.columns # list of columns name
for i,j in enumerate(features):
    plt.subplot(3,3,i+1) # create subplot for histogram
    plt.title("Histogram of {}".format(j), fontsize = 15) # title of histogram

    bins = len(fineTech_appData2[j].unique()) # bins for histogram
    plt.hist(fineTech_appData2[j], bins = bins, rwidth = 0.8, edgecolor = "y", linewidth
= 2, ) # plot histogram

plt.subplots_adjust(hspace=0.5) # space between horixontal axes (subplots) *****code 8
```



In [ ]:

```
for i,j in enumerate(features):
    print(i,j)
```

```
0 dayofweek
1 hour
2 age
3 numscreens
4 minigame
```

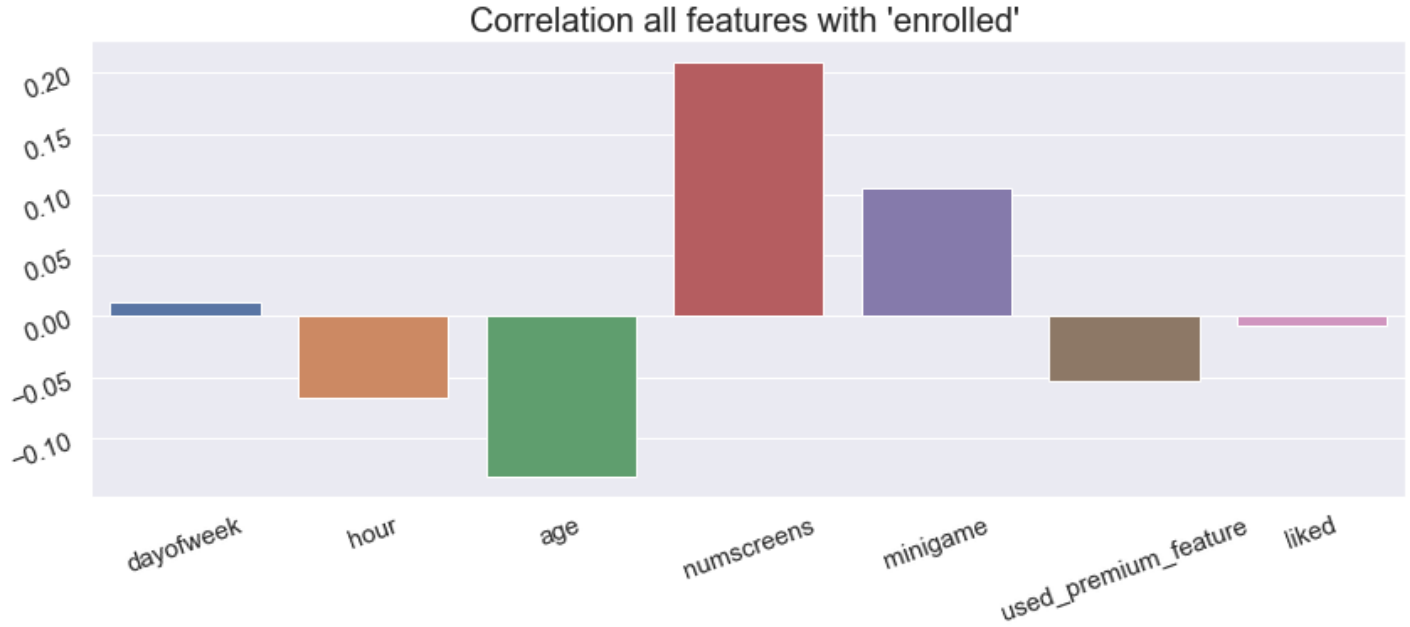
```
5 used_premium_feature
6 enrolled
7 liked
```

Correlation barplot with 'enrolled' feature

In [ ]:

```
# show corelation barplot

sns.set() # set background dark grid
plt.figure(figsize = (14,5))
plt.title("Correlation all features with 'enrolled' ", fontsize = 20)
fineTech_appData3 = fineTech_appData2.drop(['enrolled'], axis = 1) # drop 'enrolled' fea
ture
ax =sns.barplot(fineTech_appData3.columns,fineTech_appData3.corrwith(fineTech_appData2.e
nrolled)) # plot barplot
ax.tick_params(labelsize=15, labelrotation = 20, color ="k") # decorate x & y ticks font
*****code 9
```



In [ ]:

```
# parsing object data into data time format

fineTech_appData['first_open'] =[parser.parse(i) for i in fineTech_appData['first_open']
]
```

In [ ]:

```
fineTech_appData['enrolled_date'] =[parser.parse(i) if isinstance(i, str) else i for i i
n fineTech_appData['enrolled_date']]
```

In [ ]:

```
fineTech_appData.dtypes
```

Out[ ]:

```
user                            int64
first_open              datetime64[ns]
dayofweek                       int64
hour                            int32
age                             int64
screen_list                    object
numscreens                      int64
minigame                        int64
used_premium_feature            int64
enrolled                        int64
enrolled_date           datetime64[ns]
```

```
liked                                int64
dtype: object
```

In [ ]:

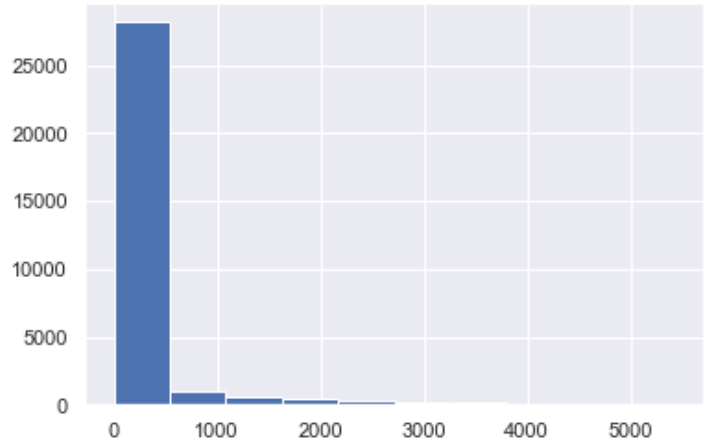
```
fineTech_appData['time_to_enrolled']  = (fineTech_appData.enrolled_date - fineTech_appDat
a.first_open).astype('timedelta64[h]')
```

In [ ]:

```
# plot histogram
plt.hist(fineTech_appData['time_to_enrolled'].dropna()) # *****code 10
```

Out[ ]:

```
(array([2.8195e+04, 1.0320e+03, 5.6700e+02, 4.2500e+02, 2.8800e+02,
        1.7900e+02, 1.6500e+02, 9.7000e+01, 1.0400e+02, 2.2000e+01]),
 array([   0. ,  543.4, 1086.8, 1630.2, 2173.6, 2717. , 3260.4, 3803.8,
        4347.2, 4890.6, 5434. ]),
 <a list of 10 Patch objects>)
```

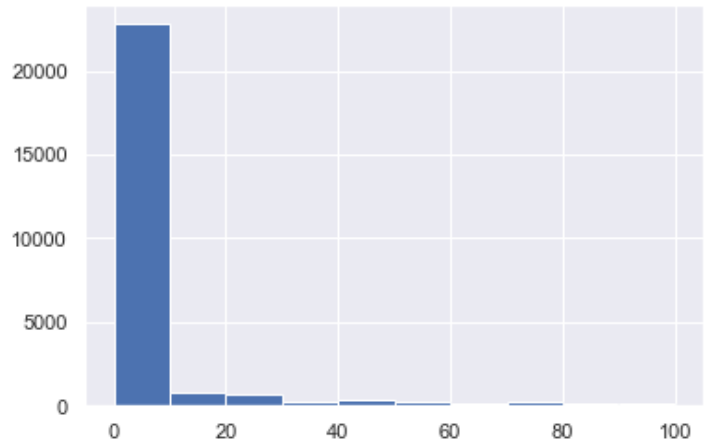


In [ ]:

```
# Plot histogram
plt.hist(fineTech_appData['time_to_enrolled'].dropna(), range = (0,100)) # *****code 11
```

Out[ ]:

```
(array([22793.,   755.,   707.,   288.,   347.,   210.,   187.,   212.,
          135.,   194.]),
 array([  0.,  10.,  20.,  30.,  40.,  50.,  60.,  70.,  80.,  90., 100.]),
 <a list of 10 Patch objects>)
```



In [ ]:

```
# Those customers have enrolled after 48 hours set as 0
fineTech_appData.loc[fineTech_appData.time_to_enrolled > 48, 'enrolled'] = 0
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	first_open	dayofweek	hour	age	screen_list	numscreens	minigar
0	235136	2012-12-27 02:14:51.273	3	2	23	idscreen,joinscreen,Cycle,product_review,ScanP...	15	
1	333588	2012-12-02 01:16:00.905	6	1	24	joinscreen,product_review,product_review2,Scan...	13	
2	254414	2013-03-19 19:19:09.157	1	19	23	Splash,Cycle,Loan	3	
3	234192	2013-07-05 16:08:46.354	4	16	28	product_review,Home,product_review,Loan3,Finan...	40	
4	51549	2013-02-26 18:50:48.661	1	18	31	idscreen,joinscreen,Cycle,Credit3Container,Sca...	32	
5	56480	2013-04-03 09:58:15.752	2	9	20	idscreen,Cycle,Home,ScanPreview,VerifyPhone,Ve...	14	
6	144649	2012-12-25 02:33:18.461	1	2	35	product_review,product_review2,ScanPreview	3	
7	249366	2012-12-11 03:07:49.875	1	3	26	Splash,Cycle,Home,Credit3Container,Credit3Dash...	41	
8	372004	2013-03-20 14:22:01.569	2	14	29	product_review,product_review2,ScanPreview,Ver...	33	
9	338013	2013-04-26 18:22:16.013	4	18	26	Home,Loan2,product_review,product_review,produ...	19	
10	43555	2013-05-14 04:48:27.597	1	4	39	Splash,idscreen,Home,RewardsContainer,Settings...	14	
11	317454	2013-05-28 11:07:07.358	1	11	32	product_review,Home,Loan2,Credit3Container,Ver...	25	
12	205375	2012-12-17 06:28:45.903	0	6	25	idscreen,joinscreen,Cycle,product_review,produ...	11	
13	307608	2013-05-25 19:52:31.798	5	19	23	Alerts,ProfilePage,Home,Credit3Container	4	
14	359855	2013-02-18 04:48:48.912	0	4	17	joinscreen,product_review,product_review2,Scan...	9	
15	284938	2013-02-02 18:41:35.724	5	18	25	idscreen,joinscreen,Cycle,Loan2,product_review...	26	
16	235143	2013-07-07 16:07:35.057	6	16	21	product_review,product_review,product_review,p...	6	
17	141402	2013-02-02 21:12:46.888	5	21	55	joinscreen,Cycle,product_review,Loan2,product_...	20	
18	257945	2013-05-10 05:59:43.405	4	5	32	Splash,product_review,Home,Loan2,product_revie...	15	
19	54931	2013-07-06 17:34:46.439	5	17	25	idscreen,Loan3,product_review,product_review,Home	5	
20	165432	2013-05-24 09:19:49.648	4	9	28	Splash,idscreen,Cycle,Home,Loan2,ProfilePage,B...	8	
21	236951	2013-04-20 04:02:18.337	5	4	38	Cycle,Splash,Home,Loan2,product_review,product...	42	
22	110461	2013-06-08 17:11:46.125	5	17	31	idscreen,Home,ScanPreview,location,VerifyPhone...	9	
23	200187	2013-05-12 02:17:36.514	6	2	27	Splash,idscreen,Cycle,Home,ScanPreview,VerifyP...	42	
24	180427	2013-05-19 20:23:46.939	6	20	48	Home	1	
25	359383	2013-06-23 18:34:40.824	6	18	37	idscreen,product_review,product_review,product...	38	

	user	first_open	dayofweek	hour	age	screen_list	numscreens	minigar
26	9810	2013-04-20 16:36:36.854	0	16	22	Home,product_review,product_review,product_rev...	49	
27	85089	2013-01-23 01:43:05.398	2	1	35	idscreen,joinscreen,Home,Loan2,ProfilePage,Pro...	35	
28	143818	2013-05-14 11:48:41.143	1	11	32	Cycle,Credit3Container,ScanPreview,VerifyDateO...	19	
29	210060	2013-04-27 17:41:24.374	5	17	20	idscreen,Cycle,Home,Loan2,product_review,produ...	10	
...	...	...	...	...	...	...	...	
49970	14381	2013-02-05 15:39:42.553	1	15	33	joinscreen,product_review,product_review2,prod...	12	
49971	347180	2013-03-23 11:18:49.529	5	11	19	joinscreen,Cycle,Credit3Container,ScanPreview,...	23	
49972	169084	2013-01-28 15:53:22.863	0	15	24	Splash,Cycle,Home,Loan2,Loan1,Institutions,Pro...	41	
49973	283019	2013-02-25 18:13:55.305	0	18	49	joinscreen,product_review,product_review2,Scan...	11	
49974	205143	2013-05-05 20:49:08.445	6	20	23	Cycle,product_review,product_review,product_re...	25	
49975	182812	2013-06-18 18:22:22.407	1	18	22	product_review,Home,Loan3,product_review,produ...	18	
49976	361326	2013-03-06 20:47:48.321	2	20	35	idscreen,joinscreen,Cycle,product_review,produ...	69	
49977	180292	2012-12-02 15:00:00.434	6	15	30	SelectInstitution,WelcomeBankVerification,Comm...	14	
49978	288318	2013-04-19 17:16:29.863	4	17	22	idscreen,Cycle,Home,product_review,product_rev...	24	
49979	9215	2013-03-25 14:36:21.560	0	14	26	Home,Loan2,product_review,product_review,produ...	15	
49980	66561	2013-05-21 07:10:44.095	1	7	35	Home,product_review,product_review,product_rev...	13	
49981	15782	2013-05-15 04:01:16.993	2	4	28	Credit2,Credit3Dashboard,Loan2,product_review,...	18	
49982	352052	2013-01-04 02:24:36.509	4	2	58	product_review,ScanPreview	2	
49983	325275	2013-04-15 13:41:11.079	0	13	28	Loan2,product_review,product_review,product_re...	8	
49984	27126	2013-03-18 16:34:39.122	0	16	46	Home,Loan2,Institutions,Credit3Container,Refer...	28	
49985	150486	2013-07-05 23:19:47.630	4	23	24	idscreen,Cycle,Home,product_review,product_rev...	24	
49986	89415	2013-01-21 18:16:17.842	0	18	21	Cycle,Home,Institutions,Loan2,Credit3Dashboard...	13	
49987	255265	2013-06-17 19:27:27.910	0	19	25	product_review,Home,product_review,product_rev...	14	
49988	286847	2013-01-20 02:49:41.102	6	2	48	Loan2,Loan1,History	5	
49989	31525	2013-04-12 17:34:40.117	4	17	27	Home	1	
49990	179308	2013-05-25 17:30:47.675	5	17	20	Splash,idscreen,Cycle,Home,product_review,prod...	8	
49991	85532	2013-02-01 22:33:59.502	4	22	45	Splash,Cycle,Home,Loan2,Loan1,MLWebView,Instit...	30	
49992	96155	2013-02-03 15:41:52.059	6	15	50	idscreen,joinscreen,product_review,product_rev...	28	
49993	343026	2012-11-24 02:02:56.012	5	2	28	joinscreen,product_review,product_review2,Scan...	4	

	user	first_open	dayofweek	hour	age	screen_list	numscreens	minigar
49994	90813	2013-02-25 19:35:12.691	0	19	36	idscreen,joinscreen,Cycle,product_review,produ...	25	
49995	222774	2013-05-09 13:46:17.871	3	13	32	Splash,Home,ScanPreview,VerifyPhone,VerifySSN,...	13	
49996	169179	2013-04-09 00:05:17.823	1	0	35	Cycle,Splash,Home,RewardsContainer	4	
49997	302367	2013-02-20 22:41:51.165	2	22	39	joinscreen,product_review,product_review2,Scan...	25	
49998	324905	2013-04-28 12:33:04.288	6	12	27	Cycle,Home,product_review,product_review,produ...	26	
49999	27047	2012-12-14 01:22:44.638	4	1	25	product_review,ScanPreview,VerifyDateOfBirth,V...	26	

50000 rows × 13 columns

In [ ]:

```
fineTech_appData.drop(columns = ['time_to_enrolled', 'enrolled_date', 'first_open'], inplace=True)
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	dayofweek	hour	age	screen_list	numscreens	minigame	used_prer
0	235136	3	2	23	idscreen,joinscreen,Cycle,product_review,ScanP...	15	0	
1	333588	6	1	24	joinscreen,product_review,product_review2,Scan...	13	0	
2	254414	1	19	23	Splash,Cycle,Loan	3	0	
3	234192	4	16	28	product_review,Home,product_review,Loan3,Finan...	40	0	
4	51549	1	18	31	idscreen,joinscreen,Cycle,Credit3Container,Sca...	32	0	
5	56480	2	9	20	idscreen,Cycle,Home,ScanPreview,VerifyPhone,Ve...	14	0	
6	144649	1	2	35	product_review,product_review2,ScanPreview	3	0	
7	249366	1	3	26	Splash,Cycle,Home,Credit3Container,Credit3Dash...	41	0	
8	372004	2	14	29	product_review,product_review2,ScanPreview,Ver...	33	1	
9	338013	4	18	26	Home,Loan2,product_review,product_review,produ...	19	0	
10	43555	1	4	39	Splash,idscreen,Home,RewardsContainer,Settings...	14	0	
11	317454	1	11	32	product_review,Home,Loan2,Credit3Container,Ver...	25	1	
12	205375	0	6	25	idscreen,joinscreen,Cycle,product_review,produ...	11	0	
13	307608	5	19	23	Alerts,ProfilePage,Home,Credit3Container	4	0	
14	359855	0	4	17	joinscreen,product_review,product_review2,Scan...	9	0	
15	284938	5	18	25	idscreen,joinscreen,Cycle,Loan2,product_review...	26	1	
16	235143	6	16	21	product_review,product_review,product_review,p...	6	0	
17	141402	5	21	55	joinscreen,Cycle,product_review,Loan2,product_...	20	0	
18	257945	4	5	32	Splash,product_review,Home,Loan2,product_revie...	15	0	
19	54931	5	17	25	idscreen,Loan3,product_review,product_review,Home	5	0	
20	165432	4	9	28	Splash,idscreen,Cycle,Home,Loan2,ProfilePage,B...	8	0	
21	236951	5	4	38	Cycle,Splash,Home,Loan2,product_review,product...	42	0	
22	110461	5	17	31	idscreen,Home,ScanPreview,location,VerifyPhone...	9	1	
23	200187	6	2	27	Splash,idscreen,Cycle,Home,ScanPreview,VerifyP...	42	0	

	user	dayofweek	hour	age	screen_list	numscreens	minigame	used_prem
24	180427	6	20	48	Home	1	1	
25	359383	6	18	37	idscreen,product_review,product_review,product...	38	0	
26	9810	0	16	22	Home,product_review,product_review,product_rev...	49	1	
27	85089	2	1	35	idscreen,joinscreen,Home,Loan2,ProfilePage,Pro...	35	0	
28	143818	1	11	32	Cycle,Credit3Container,ScanPreview,VerifyDateO...	19	0	
29	210060	5	17	20	idscreen,Cycle,Home,Loan2,product_review,produ...	10	0	
...	...	...	...	...	...	...	...	
49970	14381	1	15	33	joinscreen,product_review,product_review2,prod...	12	0	
49971	347180	5	11	19	joinscreen,Cycle,Credit3Container,ScanPreview,...	23	0	
49972	169084	0	15	24	Splash,Cycle,Home,Loan2,Loan1,Institutions,Pro...	41	0	
49973	283019	0	18	49	joinscreen,product_review,product_review2,Scan...	11	0	
49974	205143	6	20	23	Cycle,product_review,product_review,product_re...	25	0	
49975	182812	1	18	22	product_review,Home,Loan3,product_review,produ...	18	0	
49976	361326	2	20	35	idscreen,joinscreen,Cycle,product_review,produ...	69	0	
49977	180292	6	15	30	SelectInstitution,WelcomeBankVerification,Comm...	14	0	
49978	288318	4	17	22	idscreen,Cycle,Home,product_review,product_rev...	24	0	
49979	9215	0	14	26	Home,Loan2,product_review,product_review,produ...	15	0	
49980	66561	1	7	35	Home,product_review,product_review,product_rev...	13	0	
49981	15782	2	4	28	Credit2,Credit3Dashboard,Loan2,product_review,...	18	0	
49982	352052	4	2	58	product_review,ScanPreview	2	0	
49983	325275	0	13	28	Loan2,product_review,product_review,product_re...	8	0	
49984	27126	0	16	46	Home,Loan2,Institutions,Credit3Container,Refer...	28	0	
49985	150486	4	23	24	idscreen,Cycle,Home,product_review,product_rev...	24	0	
49986	89415	0	18	21	Cycle,Home,Institutions,Loan2,Credit3Dashboard...	13	0	
49987	255265	0	19	25	product_review,Home,product_review,product_rev...	14	0	
49988	286847	6	2	48	Loan2,Loan1,History	5	0	
49989	31525	4	17	27	Home	1	0	
49990	179308	5	17	20	Splash,idscreen,Cycle,Home,product_review,prod...	8	0	
49991	85532	4	22	45	Splash,Cycle,Home,Loan2,Loan1,MLWebView,Instit...	30	1	
49992	96155	6	15	50	idscreen,joinscreen,product_review,product_rev...	28	0	
49993	343026	5	2	28	joinscreen,product_review,product_review2,Scan...	4	0	
49994	90813	0	19	36	idscreen,joinscreen,Cycle,product_review,produ...	25	0	
49995	222774	3	13	32	Splash,Home,ScanPreview,VerifyPhone,VerifySSN,...	13	0	
49996	169179	1	0	35	Cycle,Splash,Home,RewardsContainer	4	0	
49997	302367	2	22	39	joinscreen,product_review,product_review2,Scan...	25	0	
49998	324905	6	12	27	Cycle,Home,product_review,product_review,produ...	26	0	
49999	27047	4	1	25	product_review,ScanPreview,VerifyDateOfBirth,V...	26	0	

50000 rows × 10 columns

In [ ]:

```
# read csv file and convert it into numpy array
fineTech_app_screen_Data = pd.read_csv("top_screens.csv").top_screens.values
```

In [ ]:

```
fineTech_app_screen_Data
```

Out[ ]:

```
array(['Loan2', 'location', 'Institutions', 'Credit3Container',
       'VerifyPhone', 'BankVerification', 'VerifyDateOfBirth',
       'ProfilePage', 'VerifyCountry', 'Cycle', 'idscreen',
       'Credit3Dashboard', 'Loan3', 'CC1Category', 'Splash', 'Loan',
       'CC1', 'RewardsContainer', 'Credit3', 'Credit1', 'EditProfile',
       'Credit2', 'Finances', 'CC3', 'Saving9', 'Saving1', 'Alerts',
       'Saving8', 'Saving10', 'Leaderboard', 'Saving4', 'VerifyMobile',
       'VerifyHousing', 'RewardDetail', 'VerifyHousingAmount',
       'ProfileMaritalStatus', 'ProfileChildren ', 'ProfileEducation',
       'Saving7', 'ProfileEducationMajor', 'Rewards', 'AccountView',
       'VerifyAnnualIncome', 'VerifyIncomeType', 'Saving2', 'Saving6',
       'Saving2Amount', 'Saving5', 'ProfileJobTitle', 'Login',
       'ProfileEmploymentLength', 'WebView', 'SecurityModal', 'Loan4',
       'ResendToken', 'TransactionList', 'NetworkFailure', 'ListPicker'],
      dtype=object)
```

In [ ]:

```
type(fineTech_app_screen_Data)
```

Out[ ]:

```
numpy.ndarray
```

In [ ]:

```
# Add ',' at the end of each string of  'sreen_list' for further operation.
fineTech_appData['screen_list'] = fineTech_appData.screen_list.astype(str) + ','
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	dayofweek	hour	age	screen_list	numscreens	minigame	used_prem
0	235136	3	2	23	idscreen,joinscreen,Cycle,product_review,ScanP...	15	0	
1	333588	6	1	24	joinscreen,product_review,product_review2,Scan...	13	0	
2	254414	1	19	23	Splash,Cycle,Loan,	3	0	
3	234192	4	16	28	product_review,Home,product_review,Loan3,Finan...	40	0	
4	51549	1	18	31	idscreen,joinscreen,Cycle,Credit3Container,Sca...	32	0	
5	56480	2	9	20	idscreen,Cycle,Home,ScanPreview,VerifyPhone,Ve...	14	0	
6	144649	1	2	35	product_review,product_review2,ScanPreview,	3	0	
7	249366	1	3	26	Splash,Cycle,Home,Credit3Container,Credit3Dash...	41	0	
8	372004	2	14	29	product_review,product_review2,ScanPreview,Ver...	33	1	
9	338013	4	18	26	Home,Loan2,product_review,product_review,produ...	19	0	
10	43555	1	4	39	Splash,idscreen,Home,RewardsContainer,Settings...	14	0	
11	317454	1	11	32	product_review,Home,Loan2,Credit3Container,Ver...	25	1	
12	205375	0	6	25	idscreen,joinscreen,Cycle,product_review,produ...	11	0	
13	307608	5	19	23	Alerts,ProfilePage,Home,Credit3Container,	4	0	
14	359855	0	4	17	joinscreen,product_review,product_review2,Scan...	9	0	
15	284938	5	18	25	idscreen,joinscreen,Cycle,Loan2,product_review...	26	1	
16	235143	6	16	21	product_review,product_review,product_review,p...	6	0	
17	141402	5	21	55	joinscreen,Cycle,product_review,Loan2,product_...	20	0	
18	257945	4	5	32	Splash,product_review,Home,Loan2,product_revie...	15	0	
19	54931	5	17	25	idscreen,Loan3,product_review,product_review,H...	5	0	
20	165432	4	9	28	Splash,idscreen,Cycle,Home,Loan2,ProfilePage,B...	8	0	

	user	dayofweek	hour	age	screen_list	numscreens	minigame	used_prem
21	236951	5	4	38	Cycle,Splash,Home,Loan2,product_review,product...	42	0	
22	110461	5	17	31	idscreen,Home,ScanPreview,location,VerifyPhone...	9	1	
23	200187	6	2	27	Splash,idscreen,Cycle,Home,ScanPreview,VerifyP...	42	0	
24	180427	6	20	48	Home,	1	1	
25	359383	6	18	37	idscreen,product_review,product_review,product...	38	0	
26	9810	0	16	22	Home,product_review,product_review,product_rev...	49	1	
27	85089	2	1	35	idscreen,joinscreen,Home,Loan2,ProfilePage,Pro...	35	0	
28	143818	1	11	32	Cycle,Credit3Container,ScanPreview,VerifyDateO...	19	0	
29	210060	5	17	20	idscreen,Cycle,Home,Loan2,product_review,produ...	10	0	
...	...	...	...	...	...	...	...	
49970	14381	1	15	33	joinscreen,product_review,product_review2,prod...	12	0	
49971	347180	5	11	19	joinscreen,Cycle,Credit3Container,ScanPreview,...	23	0	
49972	169084	0	15	24	Splash,Cycle,Home,Loan2,Loan1,Institutions,Pro...	41	0	
49973	283019	0	18	49	joinscreen,product_review,product_review2,Scan...	11	0	
49974	205143	6	20	23	Cycle,product_review,product_review,product_re...	25	0	
49975	182812	1	18	22	product_review,Home,Loan3,product_review,produ...	18	0	
49976	361326	2	20	35	idscreen,joinscreen,Cycle,product_review,produ...	69	0	
49977	180292	6	15	30	SelectInstitution,WelcomeBankVerification,Comm...	14	0	
49978	288318	4	17	22	idscreen,Cycle,Home,product_review,product_rev...	24	0	
49979	9215	0	14	26	Home,Loan2,product_review,product_review,produ...	15	0	
49980	66561	1	7	35	Home,product_review,product_review,product_rev...	13	0	
49981	15782	2	4	28	Credit2,Credit3Dashboard,Loan2,product_review,...	18	0	
49982	352052	4	2	58	product_review,ScanPreview,	2	0	
49983	325275	0	13	28	Loan2,product_review,product_review,product_re...	8	0	
49984	27126	0	16	46	Home,Loan2,Institutions,Credit3Container,Refer...	28	0	
49985	150486	4	23	24	idscreen,Cycle,Home,product_review,product_rev...	24	0	
49986	89415	0	18	21	Cycle,Home,Institutions,Loan2,Credit3Dashboard...	13	0	
49987	255265	0	19	25	product_review,Home,product_review,product_rev...	14	0	
49988	286847	6	2	48	Loan2,Loan1,History,	5	0	
49989	31525	4	17	27	Home,	1	0	
49990	179308	5	17	20	Splash,idscreen,Cycle,Home,product_review,prod...	8	0	
49991	85532	4	22	45	Splash,Cycle,Home,Loan2,Loan1,MLWebView,Instit...	30	1	
49992	96155	6	15	50	idscreen,joinscreen,product_review,product_rev...	28	0	
49993	343026	5	2	28	joinscreen,product_review,product_review2,Scan...	4	0	
49994	90813	0	19	36	idscreen,joinscreen,Cycle,product_review,produ...	25	0	
49995	222774	3	13	32	Splash,Home,ScanPreview,VerifyPhone,VerifySSN,...	13	0	
49996	169179	1	0	35	Cycle,Splash,Home,RewardsContainer,	4	0	
49997	302367	2	22	39	joinscreen,product_review,product_review2,Scan...	25	0	
49998	324905	6	12	27	Cycle,Home,product_review,product_review,produ...	26	0	
49999	27047	4	1	25	product_review,ScanPreview,VerifyDateOfBirth,V...	26	0	

50000 rows × 10 columns

In [ ]:

```
# string into to number
```

```
for screen_name in fineTech_app_screen_Data:
    fineTech_appData[screen_name] = fineTech_appData.screen_list.str.contains(screen_nam
e).astype(int)
    fineTech_appData['screen_list'] = fineTech_appData.screen_list.str.replace(screen_na
me+",", "")
```

In [ ]:

```
# test
fineTech_appData.screen_list.str.contains('Splash').astype(int)
```

Out[ ]:

```
0         0
1         0
2         0
3         0
4         0
5         0
6         0
7         0
8         0
9         0
10        0
11        0
12        0
13        0
14        0
15        0
16        0
17        0
18        0
19        0
20        0
21        0
22        0
23        0
24        0
25        0
26        0
27        0
28        0
29        0
        ..
49970     0
49971     0
49972     0
49973     0
49974     0
49975     0
49976     0
49977     0
49978     0
49979     0
49980     0
49981     0
49982     0
49983     0
49984     0
49985     0
49986     0
49987     0
49988     0
49989     0
49990     0
49991     0
49992     0
49993     0
49994     0
49995     0
49996     0
49997     0
```

```
49998    0
49999    0
Name: screen_list, Length: 50000, dtype: int32
```

In [ ]:

```
# test
fineTech_appData.screen_list.str.replace('Splash'+",", "")
```

Out[ ]:

```
0        joinscreen,product_review,ScanPreview,VerifyTo...
1        joinscreen,product_review,product_review2,Scan...
2
3        product_review,Home,product_review,ReferralCon...
4        joinscreen,ScanPreview,VerifySSN,Home,SelectIn...
5        Home,ScanPreview,VerifySSN,product_review,prod...
6              product_review,product_review2,ScanPreview,
7        Home,product_review,product_review2,ScanPrevie...
8        product_review,product_review2,ScanPreview,Ver...
9        Home,product_review,product_review,product_rev...
10            Home,Settings,product_review,product_review,
11       product_review,Home,SelectInstitution,product_...
12       joinscreen,product_review,product_review2,Scan...
13                                                  Home,
14       joinscreen,product_review,product_review2,Scan...
15       joinscreen,product_review,product_review2,Veri...
16       product_review,product_review,product_review,p...
17       joinscreen,product_review,product_review2,Veri...
18       product_review,Home,product_review,product_rev...
19                    product_review,product_review,Home,
20                                                  Home,
21       Home,product_review,product_review,product_rev...
22                             Home,ScanPreview,VerifySSN,
23       Home,ScanPreview,VerifySSN,SelectInstitution,R...
24                                                  Home,
25       product_review,product_review,product_review,p...
26       Home,product_review,product_review,product_rev...
27       joinscreen,Home,ProfileChildren,SelectInstitut...
28       ScanPreview,VerifySSN,product_review3,product_...
29       Home,product_review,product_review,product_rev...
                               ...
49970    joinscreen,product_review,product_review2,prod...
49971    joinscreen,ScanPreview,VerifySSN,Home,SelectIn...
49972                                    Home,Loan1,Grouped
49973    joinscreen,product_review,product_review2,Scan...
49974    product_review,product_review,product_review,p...
49975    product_review,Home,product_review,product_rev...
49976    joinscreen,product_review,product_review2,Scan...
49977    SelectInstitution,WelcomeCommunityAndInvites,S...
49978    Home,product_review,product_review,product_rev...
49979    Home,product_review,product_review,product_rev...
49980    Home,product_review,product_review,product_rev...
49981    product_review,product_review,product_review3,...
49982                           product_review,ScanPreview,
49983    product_review,product_review,product_review3,...
49984    Home,ReferralContainer,product_review,LoginFor...
49985    Home,product_review,product_review,product_rev...
49986                              Home,CommunityAndInvites,
49987    product_review,Home,product_review,product_rev...
49988                                       Loan1,History,
49989                                                Home,
49990      Home,product_review,product_review,ScanPreview,
49991                                Home,Loan1,MLVerifySSN,
49992    joinscreen,product_review,product_review2,Veri...
49993    joinscreen,product_review,product_review2,Scan...
49994    joinscreen,product_review,product_review2,Scan...
49995    Home,ScanPreview,VerifySSN,product_review,prod...
49996                                                Home,
49997    joinscreen,product_review,product_review2,Scan...
49998    Home,product_review,product_review,product_rev...
49999    product_review,ScanPreview,ProfileVerifySSN,Pr...
Name: screen_list, Length: 50000, dtype: object
```

In [ ]:

```
# get shape
fineTech_appData.shape
```

Out[ ]:

```
(50000, 68)
```

In [ ]:

```
# head of DataFrame
fineTech_appData.head(6) # *****code 12
```

Out[ ]:

	user	dayofweek	hour	age	screen_list	numscreens	minigame	used_premium_
0	235136	3	2	23	joinscreen,product_review,ScanPreview,VerifyTo...	15	0	
1	333588	6	1	24	joinscreen,product_review,product_review2,Scan...	13	0	
2	254414	1	19	23		3	0	
3	234192	4	16	28	product_review,Home,product_review,ReferralCon...	40	0	
4	51549	1	18	31	joinscreen,ScanPreview,VerifySSN,Home,SelectIn...	32	0	
5	56480	2	9	20	Home,ScanPreview,VerifySSN,product_review,prod...	14	0	

6 rows × 68 columns

In [ ]:

```
# remain screen in 'screen_list'
fineTech_appData.loc[0,'screen_list']
```

Out[ ]:

```
'joinscreen,product_review,ScanPreview,VerifyToken,ProfileVerifySSN,Settings,ForgotPasswo
rd,'
```

In [ ]:

```
fineTech_appData.screen_list.str.count(",").head(6)
```

Out[ ]:

```
0     7
1     5
2     0
3     6
4    10
5     6
Name: screen_list, dtype: int64
```

In [ ]:

```
# count remain screen list and store counted number in 'remain_screen_list'

fineTech_appData['remain_screen_list'] = fineTech_appData.screen_list.str.count(",")
```

In [ ]:

```
# Drop the 'screen_list'
fineTech_appData.drop(columns = ['screen_list'], inplace=True)
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	Login	Profil
0	235136	3	2	23	15	0	0	0	0	1	...	1	
1	333588	6	1	24	13	0	0	0	0	1	...	0	
2	254414	1	19	23	3	0	1	0	1	0	...	0	
3	234192	4	16	28	40	0	0	1	0	0	...	0	
4	51549	1	18	31	32	0	0	1	1	1	...	0	
5	56480	2	9	20	14	0	0	1	0	1	...	0	
6	144649	1	2	35	3	0	0	0	0	0	...	0	
7	249366	1	3	26	41	0	1	0	0	1	...	0	
8	372004	2	14	29	33	1	1	0	0	1	...	0	
9	338013	4	18	26	19	0	0	1	0	1	...	0	
10	43555	1	4	39	14	0	0	1	0	1	...	0	
11	317454	1	11	32	25	1	1	0	0	1	...	0	
12	205375	0	6	25	11	0	0	0	0	0	...	0	
13	307608	5	19	23	4	0	0	0	0	0	...	0	
14	359855	0	4	17	9	0	0	0	0	0	...	0	
15	284938	5	18	25	26	1	0	0	0	1	...	0	
16	235143	6	16	21	6	0	0	1	0	0	...	0	
17	141402	5	21	55	20	0	0	0	0	1	...	0	
18	257945	4	5	32	15	0	0	1	1	1	...	0	
19	54931	5	17	25	5	0	0	1	0	0	...	0	
20	165432	4	9	28	8	0	1	1	0	1	...	0	
21	236951	5	4	38	42	0	1	0	0	1	...	0	
22	110461	5	17	31	9	1	0	1	0	0	...	0	
23	200187	6	2	27	42	0	0	1	1	1	...	0	
24	180427	6	20	48	1	1	0	1	0	0	...	0	
25	359383	6	18	37	38	0	0	1	0	0	...	0	
26	9810	0	16	22	49	1	0	1	0	0	...	0	
27	85089	2	1	35	35	0	0	0	0	1	...	0	
28	143818	1	11	32	19	0	1	1	0	1	...	1	
29	210060	5	17	20	10	0	0	0	0	1	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
49970	14381	1	15	33	12	0	0	1	0	0	...	0	
49971	347180	5	11	19	23	0	0	1	1	1	...	0	
49972	169084	0	15	24	41	0	1	0	0	1	...	0	
49973	283019	0	18	49	11	0	0	1	0	0	...	0	
49974	205143	6	20	23	25	0	0	1	0	1	...	0	
49975	182812	1	18	22	18	0	0	1	0	0	...	0	
49976	361326	2	20	35	69	0	0	1	0	1	...	1	
49977	180292	6	15	30	14	0	1	0	0	1	...	1	
49978	288318	4	17	22	24	0	1	1	0	0	...	0	
49979	9215	0	14	26	15	0	0	1	0	1	...	0	
49980	66561	1	7	35	13	0	1	1	0	0	...	0	
49981	15782	2	4	28	18	0	0	0	1	1	...	0	

49982	352552	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	Logi
49983	325275	0	13	28	8	0	0	0	1	1	...	0
49984	27126	0	16	46	28	0	0	0	0	1	...	1
49985	150486	4	23	24	24	0	0	1	0	0	...	1
49986	89415	0	18	21	13	0	0	0	0	1	...	0
49987	255265	0	19	25	14	0	0	0	0	0	...	0
49988	286847	6	2	48	5	0	0	0	0	1	...	0
49989	31525	4	17	27	1	0	0	0	0	0	...	0
49990	179308	5	17	20	8	0	0	1	1	0	...	0
49991	85532	4	22	45	30	1	1	1	0	1	...	0
49992	96155	6	15	50	28	0	0	1	0	1	...	0
49993	343026	5	2	28	4	0	0	0	1	0	...	0
49994	90813	0	19	36	25	0	0	1	0	1	...	0
49995	222774	3	13	32	13	0	0	1	0	0	...	0
49996	169179	1	0	35	4	0	1	0	0	0	...	0
49997	302367	2	22	39	25	0	0	0	0	0	...	0
49998	324905	6	12	27	26	0	0	1	0	0	...	0
49999	27047	4	1	25	26	0	0	0	1	1	...	0

50000 rows × 68 columns

In [ ]:

```
# total columns
fineTech_appData.columns
```

Out[ ]:

```
Index(['user', 'dayofweek', 'hour', 'age', 'numscreens', 'minigame',
       'used_premium_feature', 'enrolled', 'liked', 'Loan2', 'location',
       'Institutions', 'Credit3Container', 'VerifyPhone', 'BankVerification',
       'VerifyDateOfBirth', 'ProfilePage', 'VerifyCountry', 'Cycle',
       'idscreen', 'Credit3Dashboard', 'Loan3', 'CC1Category', 'Splash',
       'Loan', 'CC1', 'RewardsContainer', 'Credit3', 'Credit1', 'EditProfile',
       'Credit2', 'Finances', 'CC3', 'Saving9', 'Saving1', 'Alerts', 'Saving8',
       'Saving10', 'Leaderboard', 'Saving4', 'VerifyMobile', 'VerifyHousing',
       'RewardDetail', 'VerifyHousingAmount', 'ProfileMaritalStatus',
       'ProfileChildren ', 'ProfileEducation', 'Saving7',
       'ProfileEducationMajor', 'Rewards', 'AccountView', 'VerifyAnnualIncome',
       'VerifyIncomeType', 'Saving2', 'Saving6', 'Saving2Amount', 'Saving5',
       'ProfileJobTitle', 'Login', 'ProfileEmploymentLength', 'WebView',
       'SecurityModal', 'Loan4', 'ResendToken', 'TransactionList',
       'NetworkFailure', 'ListPicker', 'remain_screen_list'],
      dtype='object')
```

In [ ]:

```
# take sum of all saving screen in one place
saving_screens = ['Saving1',
                  'Saving2',
                  'Saving2Amount',
                  'Saving4',
                  'Saving5',
                  'Saving6',
                  'Saving7',
                  'Saving8',
                  'Saving9',
                  'Saving10',
                 ]
fineTech_appData['saving_screens_count'] = fineTech_appData[saving_screens].sum(axis = 1
```

```
)
fineTech_appData.drop(columns = saving_screens, inplace = True)
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	ProfileEmplo
0	235136	3	2	23	15	0	0	0	0	1	...	
1	333588	6	1	24	13	0	0	0	0	1	...	
2	254414	1	19	23	3	0	1	0	1	0	...	
3	234192	4	16	28	40	0	0	1	0	0	...	
4	51549	1	18	31	32	0	0	1	1	1	...	
5	56480	2	9	20	14	0	0	1	0	1	...	
6	144649	1	2	35	3	0	0	0	0	0	...	
7	249366	1	3	26	41	0	1	0	0	1	...	
8	372004	2	14	29	33	1	1	0	0	1	...	
9	338013	4	18	26	19	0	0	1	0	1	...	
10	43555	1	4	39	14	0	0	1	0	1	...	
11	317454	1	11	32	25	1	1	0	0	1	...	
12	205375	0	6	25	11	0	0	0	0	0	...	
13	307608	5	19	23	4	0	0	0	0	0	...	
14	359855	0	4	17	9	0	0	0	0	0	...	
15	284938	5	18	25	26	1	0	0	0	1	...	
16	235143	6	16	21	6	0	0	1	0	0	...	
17	141402	5	21	55	20	0	0	0	0	1	...	
18	257945	4	5	32	15	0	0	1	1	1	...	
19	54931	5	17	25	5	0	0	1	0	0	...	
20	165432	4	9	28	8	0	1	1	0	1	...	
21	236951	5	4	38	42	0	1	0	0	1	...	
22	110461	5	17	31	9	1	0	1	0	0	...	
23	200187	6	2	27	42	0	0	1	1	1	...	
24	180427	6	20	48	1	1	0	1	0	0	...	
25	359383	6	18	37	38	0	0	1	0	0	...	
26	9810	0	16	22	49	1	0	1	0	0	...	
27	85089	2	1	35	35	0	0	0	0	1	...	
28	143818	1	11	32	19	0	1	1	0	1	...	
29	210060	5	17	20	10	0	0	0	0	1	...	
...	...	...	...	...	...	...	...	...	...	...	...	
49970	14381	1	15	33	12	0	0	1	0	0	...	
49971	347180	5	11	19	23	0	0	1	1	1	...	
49972	169084	0	15	24	41	0	1	0	0	1	...	
49973	283019	0	18	49	11	0	0	1	0	0	...	
49974	205143	6	20	23	25	0	0	1	0	1	...	
49975	182812	1	18	22	18	0	0	1	0	0	...	
49976	361326	2	20	35	69	0	0	1	0	1	...	

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	ProfileEmplo
49977	180292	6	15	30	14	0	1	0	0	1	...	
49978	288318	4	17	22	24	0	1	1	0	0	...	
49979	9215	0	14	26	15	0	0	1	0	1	...	
49980	66561	1	7	35	13	0	1	1	0	0	...	
49981	15782	2	4	28	18	0	0	0	1	1	...	
49982	352052	4	2	58	2	0	0	0	0	0	...	
49983	325275	0	13	28	8	0	0	0	1	1	...	
49984	27126	0	16	46	28	0	0	0	0	1	...	
49985	150486	4	23	24	24	0	0	1	0	0	...	
49986	89415	0	18	21	13	0	0	0	0	1	...	
49987	255265	0	19	25	14	0	0	0	0	0	...	
49988	286847	6	2	48	5	0	0	0	0	1	...	
49989	31525	4	17	27	1	0	0	0	0	0	...	
49990	179308	5	17	20	8	0	0	1	1	0	...	
49991	85532	4	22	45	30	1	1	1	0	1	...	
49992	96155	6	15	50	28	0	0	1	0	1	...	
49993	343026	5	2	28	4	0	0	0	1	0	...	
49994	90813	0	19	36	25	0	0	1	0	1	...	
49995	222774	3	13	32	13	0	0	1	0	0	...	
49996	169179	1	0	35	4	0	1	0	0	0	...	
49997	302367	2	22	39	25	0	0	0	0	0	...	
49998	324905	6	12	27	26	0	0	1	0	0	...	
49999	27047	4	1	25	26	0	0	0	1	1	...	

50000 rows × 59 columns

In [ ]:

```
credit_screens = ['Credit1',
                  'Credit2',
                  'Credit3',
                  'Credit3Container',
                  'Credit3Dashboard',
                 ]
fineTech_appData['credit_screens_count'] = fineTech_appData[credit_screens].sum(axis = 1
)
fineTech_appData.drop(columns = credit_screens, axis = 1, inplace = True)
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	WebView	S
0	235136	3	2	23	15	0	0	0	0	1	...	0	
1	333588	6	1	24	13	0	0	0	0	1	...	0	
2	254414	1	19	23	3	0	1	0	1	0	...	0	
3	234192	4	16	28	40	0	0	1	0	0	...	0	
4	51549	1	18	31	32	0	0	1	1	1	...	0	
5	56480	2	9	20	14	0	0	1	0	1	...	0	
6	144649	1	2	35	3	0	0	0	0	0	...	0	

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	WebView	S
7	249366	1	2	26	41	0	1	0	0	1	...	0	
8	372004	2	14	29	33	1	1	0	0	1	...	0	
9	338013	4	18	26	19	0	0	1	0	1	...	0	
10	43555	1	4	39	14	0	0	1	0	1	...	0	
11	317454	1	11	32	25	1	1	0	0	1	...	0	
12	205375	0	6	25	11	0	0	0	0	0	...	0	
13	307608	5	19	23	4	0	0	0	0	0	...	0	
14	359855	0	4	17	9	0	0	0	0	0	...	0	
15	284938	5	18	25	26	1	0	0	0	1	...	0	
16	235143	6	16	21	6	0	0	1	0	0	...	0	
17	141402	5	21	55	20	0	0	0	0	1	...	0	
18	257945	4	5	32	15	0	0	1	1	1	...	0	
19	54931	5	17	25	5	0	0	1	0	0	...	0	
20	165432	4	9	28	8	0	1	1	0	1	...	0	
21	236951	5	4	38	42	0	1	0	0	1	...	1	
22	110461	5	17	31	9	1	0	1	0	0	...	0	
23	200187	6	2	27	42	0	0	1	1	1	...	0	
24	180427	6	20	48	1	1	0	1	0	0	...	0	
25	359383	6	18	37	38	0	0	1	0	0	...	0	
26	9810	0	16	22	49	1	0	1	0	0	...	1	
27	85089	2	1	35	35	0	0	0	0	1	...	0	
28	143818	1	11	32	19	0	1	1	0	1	...	0	
29	210060	5	17	20	10	0	0	0	0	1	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
49970	14381	1	15	33	12	0	0	1	0	0	...	0	
49971	347180	5	11	19	23	0	0	1	1	1	...	0	
49972	169084	0	15	24	41	0	1	0	0	1	...	0	
49973	283019	0	18	49	11	0	0	1	0	0	...	0	
49974	205143	6	20	23	25	0	0	1	0	1	...	0	
49975	182812	1	18	22	18	0	0	1	0	0	...	0	
49976	361326	2	20	35	69	0	0	1	0	1	...	0	
49977	180292	6	15	30	14	0	1	0	0	1	...	0	
49978	288318	4	17	22	24	0	1	1	0	0	...	0	
49979	9215	0	14	26	15	0	0	1	0	1	...	0	
49980	66561	1	7	35	13	0	1	1	0	0	...	0	
49981	15782	2	4	28	18	0	0	0	1	1	...	0	
49982	352052	4	2	58	2	0	0	0	0	0	...	0	
49983	325275	0	13	28	8	0	0	0	1	1	...	0	
49984	27126	0	16	46	28	0	0	0	0	1	...	0	
49985	150486	4	23	24	24	0	0	1	0	0	...	1	
49986	89415	0	18	21	13	0	0	0	0	1	...	0	
49987	255265	0	19	25	14	0	0	0	0	0	...	0	
49988	286847	6	2	48	5	0	0	0	0	1	...	0	
49989	31525	4	17	27	1	0	0	0	0	0	...	0	
49990	179308	5	17	20	8	0	0	1	1	0	...	0	

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	WebView	S
49991	85532	4	22	45	39	1	1	1	0	1	...	1	
49992	96155	6	15	50	28	0	0	1	0	1	...	0	
49993	343026	5	2	28	4	0	0	0	1	0	...	0	
49994	90813	0	19	36	25	0	0	1	0	1	...	0	
49995	222774	3	13	32	13	0	0	1	0	0	...	1	
49996	169179	1	0	35	4	0	1	0	0	0	...	0	
49997	302367	2	22	39	25	0	0	0	0	0	...	0	
49998	324905	6	12	27	26	0	0	1	0	0	...	0	
49999	27047	4	1	25	26	0	0	0	1	1	...	0	

50000 rows × 55 columns

In [ ]:

```
cc_screens = ['CC1',
              'CC1Category',
              'CC3',
             ]
fineTech_appData['cc_screens_count'] = fineTech_appData[cc_screens].sum(axis = 1)
fineTech_appData.drop(columns = cc_screens, inplace = True)
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	SecurityMod
0	235136	3	2	23	15	0	0	0	0	1	...	
1	333588	6	1	24	13	0	0	0	0	1	...	
2	254414	1	19	23	3	0	1	0	1	0	...	
3	234192	4	16	28	40	0	0	1	0	0	...	
4	51549	1	18	31	32	0	0	1	1	1	...	
5	56480	2	9	20	14	0	0	1	0	1	...	
6	144649	1	2	35	3	0	0	0	0	0	...	
7	249366	1	3	26	41	0	1	0	0	1	...	
8	372004	2	14	29	33	1	1	0	0	1	...	
9	338013	4	18	26	19	0	0	1	0	1	...	
10	43555	1	4	39	14	0	0	1	0	1	...	
11	317454	1	11	32	25	1	1	0	0	1	...	
12	205375	0	6	25	11	0	0	0	0	0	...	
13	307608	5	19	23	4	0	0	0	0	0	...	
14	359855	0	4	17	9	0	0	0	0	0	...	
15	284938	5	18	25	26	1	0	0	0	1	...	
16	235143	6	16	21	6	0	0	1	0	0	...	
17	141402	5	21	55	20	0	0	0	0	1	...	
18	257945	4	5	32	15	0	0	1	1	1	...	
19	54931	5	17	25	5	0	0	1	0	0	...	
20	165432	4	9	28	8	0	1	1	0	1	...	
21	236951	5	4	38	42	0	1	0	0	1	...	
22	110461	5	17	31	9	1	0	1	0	0	...	

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	Loan2	...	SecurityMod
23	200167	6	2	27	42	0	0	1	1	1	...	
24	180427	6	20	48	1	1	0	1	0	0	...	
25	359383	6	18	37	38	0	0	1	0	0	...	
26	9810	0	16	22	49	1	0	1	0	0	...	
27	85089	2	1	35	35	0	0	0	0	1	...	
28	143818	1	11	32	19	0	1	1	0	1	...	
29	210060	5	17	20	10	0	0	0	0	1	...	
...	...	...	...	...	...	...	...	...	...	...	...	
49970	14381	1	15	33	12	0	0	1	0	0	...	
49971	347180	5	11	19	23	0	0	1	1	1	...	
49972	169084	0	15	24	41	0	1	0	0	1	...	
49973	283019	0	18	49	11	0	0	1	0	0	...	
49974	205143	6	20	23	25	0	0	1	0	1	...	
49975	182812	1	18	22	18	0	0	1	0	0	...	
49976	361326	2	20	35	69	0	0	1	0	1	...	
49977	180292	6	15	30	14	0	1	0	0	1	...	
49978	288318	4	17	22	24	0	1	1	0	0	...	
49979	9215	0	14	26	15	0	0	1	0	1	...	
49980	66561	1	7	35	13	0	1	1	0	0	...	
49981	15782	2	4	28	18	0	0	0	1	1	...	
49982	352052	4	2	58	2	0	0	0	0	0	...	
49983	325275	0	13	28	8	0	0	0	1	1	...	
49984	27126	0	16	46	28	0	0	0	0	1	...	
49985	150486	4	23	24	24	0	0	1	0	0	...	
49986	89415	0	18	21	13	0	0	0	0	1	...	
49987	255265	0	19	25	14	0	0	0	0	0	...	
49988	286847	6	2	48	5	0	0	0	0	1	...	
49989	31525	4	17	27	1	0	0	0	0	0	...	
49990	179308	5	17	20	8	0	0	1	1	0	...	
49991	85532	4	22	45	30	1	1	1	0	1	...	
49992	96155	6	15	50	28	0	0	1	0	1	...	
49993	343026	5	2	28	4	0	0	0	1	0	...	
49994	90813	0	19	36	25	0	0	1	0	1	...	
49995	222774	3	13	32	13	0	0	1	0	0	...	
49996	169179	1	0	35	4	0	1	0	0	0	...	
49997	302367	2	22	39	25	0	0	0	0	0	...	
49998	324905	6	12	27	26	0	0	1	0	0	...	
49999	27047	4	1	25	26	0	0	0	1	1	...	

50000 rows × 53 columns

In [ ]:

```
loan_screens = ['Loan',
                'Loan2',
                'Loan3',
                'Loan4',
```

```
                ]
fineTech_appData['loan_screens_count'] = fineTech_appData[loan_screens].sum(axis = 1)
fineTech_appData.drop(columns = loan_screens, inplace = True)
```

In [ ]:

```
fineTech_appData
```

Out[ ]:

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	location	...	SecurityM
0	235136	3	2	23	15	0	0	0	0	0	...	
1	333588	6	1	24	13	0	0	0	0	1	...	
2	254414	1	19	23	3	0	1	0	1	0	...	
3	234192	4	16	28	40	0	0	1	0	1	...	
4	51549	1	18	31	32	0	0	1	1	0	...	
5	56480	2	9	20	14	0	0	1	0	0	...	
6	144649	1	2	35	3	0	0	0	0	0	...	
7	249366	1	3	26	41	0	1	0	0	0	...	
8	372004	2	14	29	33	1	1	0	0	1	...	
9	338013	4	18	26	19	0	0	1	0	1	...	
10	43555	1	4	39	14	0	0	1	0	0	...	
11	317454	1	11	32	25	1	1	0	0	0	...	
12	205375	0	6	25	11	0	0	0	0	0	...	
13	307608	5	19	23	4	0	0	0	0	0	...	
14	359855	0	4	17	9	0	0	0	0	0	...	
15	284938	5	18	25	26	1	0	0	0	1	...	
16	235143	6	16	21	6	0	0	1	0	1	...	
17	141402	5	21	55	20	0	0	0	0	1	...	
18	257945	4	5	32	15	0	0	1	1	1	...	
19	54931	5	17	25	5	0	0	1	0	0	...	
20	165432	4	9	28	8	0	1	1	0	0	...	
21	236951	5	4	38	42	0	1	0	0	1	...	
22	110461	5	17	31	9	1	0	1	0	1	...	
23	200187	6	2	27	42	0	0	1	1	0	...	
24	180427	6	20	48	1	1	0	1	0	0	...	
25	359383	6	18	37	38	0	0	1	0	0	...	
26	9810	0	16	22	49	1	0	1	0	1	...	
27	85089	2	1	35	35	0	0	0	0	1	...	
28	143818	1	11	32	19	0	1	1	0	0	...	
29	210060	5	17	20	10	0	0	0	0	1	...	
...	...	...	...	...	...	...	...	...	...	...	...	
49970	14381	1	15	33	12	0	0	1	0	1	...	
49971	347180	5	11	19	23	0	0	1	1	0	...	
49972	169084	0	15	24	41	0	1	0	0	0	...	
49973	283019	0	18	49	11	0	0	1	0	0	...	
49974	205143	6	20	23	25	0	0	1	0	1	...	
49975	182812	1	18	22	18	0	0	1	0	1	...	
49976	361326	2	20	35	69	0	0	1	0	1		

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enrolled	liked	location	...	SecurityMo
49977	180292	6	15	30	14	0	1	0	0	0	...	
49978	288318	4	17	22	24	0	1	1	0	1	...	
49979	9215	0	14	26	15	0	0	1	0	1	...	
49980	66561	1	7	35	13	0	1	1	0	1	...	
49981	15782	2	4	28	18	0	0	0	1	1	...	
49982	352052	4	2	58	2	0	0	0	0	0	...	
49983	325275	0	13	28	8	0	0	0	1	1	...	
49984	27126	0	16	46	28	0	0	0	0	0	...	
49985	150486	4	23	24	24	0	0	1	0	1	...	
49986	89415	0	18	21	13	0	0	0	0	0	...	
49987	255265	0	19	25	14	0	0	0	0	0	...	
49988	286847	6	2	48	5	0	0	0	0	0	...	
49989	31525	4	17	27	1	0	0	0	0	0	...	
49990	179308	5	17	20	8	0	0	1	1	0	...	
49991	85532	4	22	45	30	1	1	1	0	1	...	
49992	96155	6	15	50	28	0	0	1	0	1	...	
49993	343026	5	2	28	4	0	0	0	1	0	...	
49994	90813	0	19	36	25	0	0	1	0	1	...	
49995	222774	3	13	32	13	0	0	1	0	0	...	
49996	169179	1	0	35	4	0	1	0	0	0	...	
49997	302367	2	22	39	25	0	0	0	0	1	...	
49998	324905	6	12	27	26	0	0	1	0	1	...	
49999	27047	4	1	25	26	0	0	0	1	0	...	

50000 rows × 50 columns

In [ ]:

```
fineTech_appData.shape
```

Out[ ]:

```
(50000, 50)
```

In [ ]:

```
fineTech_appData.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50000 entries, 0 to 49999
Data columns (total 50 columns):
user                      50000 non-null int64
dayofweek                 50000 non-null int64
hour                      50000 non-null int32
age                       50000 non-null int64
numscreens                50000 non-null int64
minigame                  50000 non-null int64
used_premium_feature      50000 non-null int64
enrolled                  50000 non-null int64
liked                     50000 non-null int64
location                  50000 non-null int32
Institutions              50000 non-null int32
VerifyPhone               50000 non-null int32
BankVerification          50000 non-null int32
VerifyDateOfBirth         50000 non-null int32
ProfilePage               50000 non-null int32
VerifyCountry             50000 non-null int32
```

```
Cycle                     50000 non-null int32
idscreen                  50000 non-null int32
Splash                    50000 non-null int32
RewardsContainer          50000 non-null int32
EditProfile               50000 non-null int32
Finances                  50000 non-null int32
Alerts                    50000 non-null int32
Leaderboard               50000 non-null int32
VerifyMobile              50000 non-null int32
VerifyHousing             50000 non-null int32
RewardDetail              50000 non-null int32
VerifyHousingAmount       50000 non-null int32
ProfileMaritalStatus      50000 non-null int32
ProfileChildren           50000 non-null int32
ProfileEducation          50000 non-null int32
ProfileEducationMajor     50000 non-null int32
Rewards                   50000 non-null int32
AccountView               50000 non-null int32
VerifyAnnualIncome        50000 non-null int32
VerifyIncomeType          50000 non-null int32
ProfileJobTitle           50000 non-null int32
Login                     50000 non-null int32
ProfileEmploymentLength   50000 non-null int32
WebView                   50000 non-null int32
SecurityModal             50000 non-null int32
ResendToken               50000 non-null int32
TransactionList           50000 non-null int32
NetworkFailure            50000 non-null int32
ListPicker                50000 non-null int32
remain_screen_list        50000 non-null int64
saving_screens_count      50000 non-null int64
credit_screens_count      50000 non-null int64
cc_screens_count          50000 non-null int64
loan_screens_count        50000 non-null int64
dtypes: int32(37), int64(13)
memory usage: 12.0 MB
```

In [ ]:

```
fineTech_appData.describe()
```

Out[ ]:

	user	dayofweek	hour	age	numscreens	minigame	used_premium_feature	enro
count	50000.000000	50000.000000	50000.000000	50000.00000	50000.000000	50000.000000	50000.000000	50000.000
mean	186889.729900	3.029860	12.557220	31.72436	21.095900	0.107820	0.172020	0.497
std	107768.520361	2.031997	7.438072	10.80331	15.728812	0.310156	0.377402	0.499
min	13.000000	0.000000	0.000000	16.00000	1.000000	0.000000	0.000000	0.000
25%	93526.750000	1.000000	5.000000	24.00000	10.000000	0.000000	0.000000	0.000
50%	187193.500000	3.000000	14.000000	29.00000	18.000000	0.000000	0.000000	0.000
75%	279984.250000	5.000000	19.000000	37.00000	28.000000	0.000000	0.000000	1.000
max	373662.000000	6.000000	23.000000	101.00000	325.000000	1.000000	1.000000	1.000

8 rows × 50 columns

Heatmap with correlation matrix of new fineTech_appData

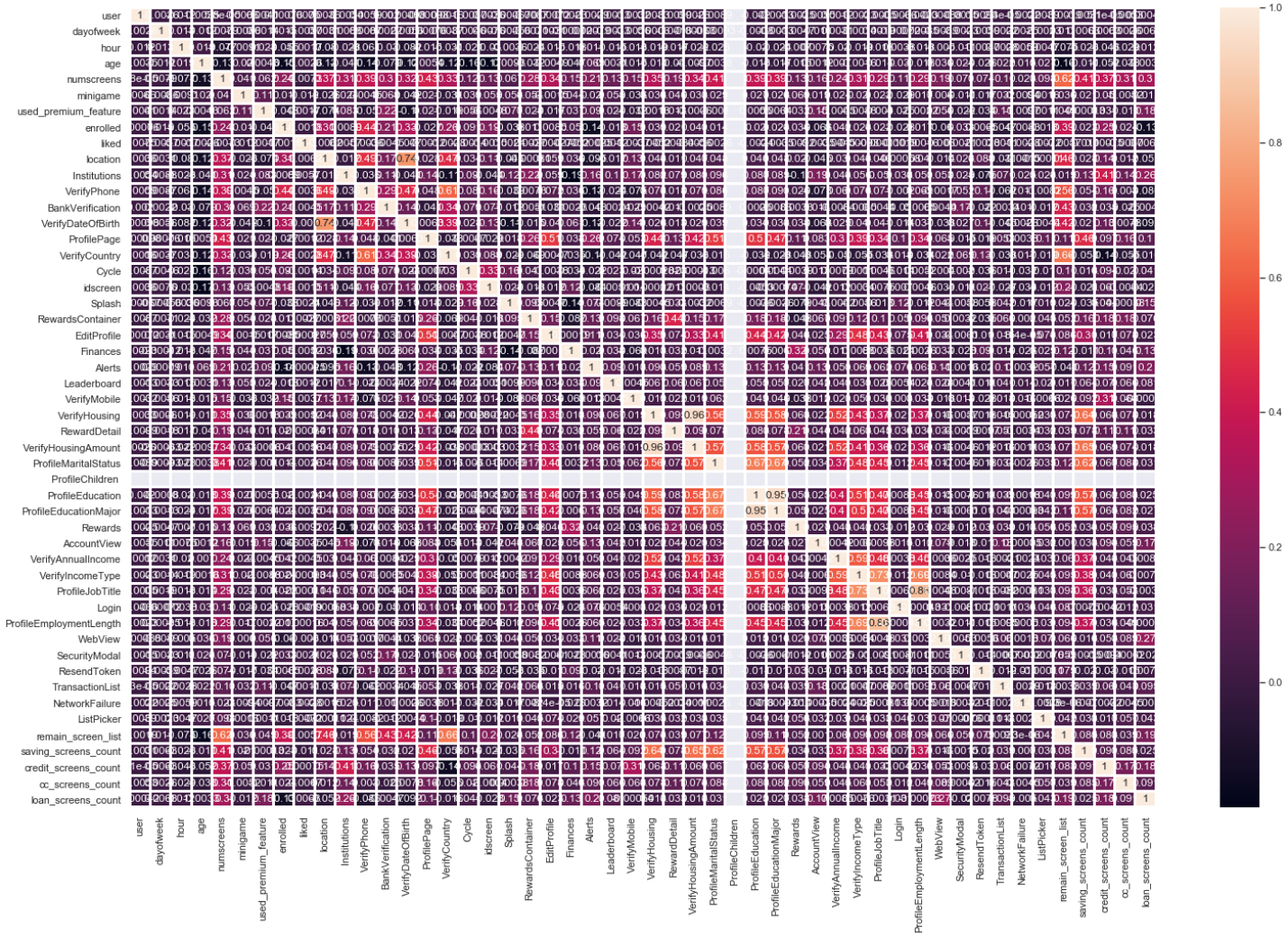
In [ ]:

```
# Heatmap with correlation matrix of new fineTech_appData

plt.figure(figsize = (25,16))
sns.heatmap(fineTech_appData.corr(), annot = True, linewidth =2) #*****code 13
```

Out[ ]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x246e92a37f0>
```



In [ ]:

```
fineTech_appData.columns
```

Out[ ]:

```
Index(['user', 'dayofweek', 'hour', 'age', 'numscreens', 'minigame',
       'used_premium_feature', 'enrolled', 'liked', 'location', 'Institutions',
       'VerifyPhone', 'BankVerification', 'VerifyDateOfBirth', 'ProfilePage',
       'VerifyCountry', 'Cycle', 'idscreen', 'Splash', 'RewardsContainer',
       'EditProfile', 'Finances', 'Alerts', 'Leaderboard', 'VerifyMobile',
       'VerifyHousing', 'RewardDetail', 'VerifyHousingAmount',
       'ProfileMaritalStatus', 'ProfileChildren ', 'ProfileEducation',
       'ProfileEducationMajor', 'Rewards', 'AccountView', 'VerifyAnnualIncome',
       'VerifyIncomeType', 'ProfileJobTitle', 'Login',
       'ProfileEmploymentLength', 'WebView', 'SecurityModal', 'ResendToken',
       'TransactionList', 'NetworkFailure', 'ListPicker', 'remain_screen_list',
       'saving_screens_count', 'credit_screens_count', 'cc_screens_count',
       'loan_screens_count'],
      dtype='object')
```

In [ ]:

```
fineTech_appData['ProfileChildren '].unique()
```

Out[ ]:

```
array([0], dtype=int64)
```

In [ ]:

```
corr_matrix = fineTech_appData.corr()
corr_matrix['ProfileChildren ']
```

Out[ ]:

```
user                        NaN
dayofweek                   NaN
hour                        NaN
age                         NaN
numscreens                  NaN
minigame                    NaN
used_premium_feature        NaN
enrolled                    NaN
liked                       NaN
location                    NaN
Institutions                NaN
VerifyPhone                 NaN
BankVerification            NaN
VerifyDateOfBirth           NaN
ProfilePage                 NaN
VerifyCountry               NaN
Cycle                       NaN
idscreen                    NaN
Splash                      NaN
RewardsContainer            NaN
EditProfile                 NaN
Finances                    NaN
Alerts                      NaN
Leaderboard                 NaN
VerifyMobile                NaN
VerifyHousing               NaN
RewardDetail                NaN
VerifyHousingAmount         NaN
ProfileMaritalStatus        NaN
ProfileChildren             NaN
ProfileEducation            NaN
ProfileEducationMajor       NaN
Rewards                     NaN
AccountView                 NaN
VerifyAnnualIncome          NaN
VerifyIncomeType            NaN
ProfileJobTitle             NaN
Login                       NaN
ProfileEmploymentLength     NaN
WebView                     NaN
SecurityModal               NaN
ResendToken                 NaN
TransactionList             NaN
NetworkFailure              NaN
ListPicker                  NaN
remain_screen_list          NaN
saving_screens_count        NaN
credit_screens_count        NaN
cc_screens_count            NaN
loan_screens_count          NaN
Name: ProfileChildren , dtype: float64
```

In [ ]:

```
fineTech_appData['ProfileChildren ']
```

Out[ ]:

```
0         0
1         0
2         0
3         0
4         0
5         0
6         0
7         0
8         0
9         0
10        0
11        0
12        0
```

```
12        0
13        0
14        0
15        0
16        0
17        0
18        0
19        0
20        0
21        0
22        0
23        0
24        0
25        0
26        0
27        0
28        0
29        0
         ..
49970     0
49971     0
49972     0
49973     0
49974     0
49975     0
49976     0
49977     0
49978     0
49979     0
49980     0
49981     0
49982     0
49983     0
49984     0
49985     0
49986     0
49987     0
49988     0
49989     0
49990     0
49991     0
49992     0
49993     0
49994     0
49995     0
49996     0
49997     0
49998     0
49999     0
Name: ProfileChildren , Length: 50000, dtype: int32
```

Data Preprocessing

Split dataset in Train and Test

In [ ]:

```
clean_fineTech_appData = fineTech_appData
target = fineTech_appData['enrolled']
fineTech_appData.drop(columns = 'enrolled', inplace = True)
```

In [ ]:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(fineTech_appData, target, test_size
= 0.2, random_state = 0)
```

In [ ]:

```
print('Shape of X_train = ', X_train.shape)
print('Shape of X_test = ', X_test.shape)
print('Shape of y_train = ', y_train.shape)
print('Shape of y_test = ', y_test.shape)
```

```
Shape of X_train =  (40000, 49)
Shape of X_test =  (10000, 49)
Shape of y_train =  (40000,)
Shape of y_test =  (10000,)
```

In [ ]:

```
# take User ID in another variable
train_userID = X_train['user']
X_train.drop(columns= 'user', inplace =True)
test_userID = X_test['user']
X_test.drop(columns= 'user', inplace =True)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:3940: SettingWithCopyWarn
ing:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexin
g.html#indexing-view-versus-copy
  errors=errors)
```

In [ ]:

```
print('Shape of X_train = ', X_train.shape)
print('Shape of X_test = ', X_test.shape)
print('Shape of train_userID = ', train_userID.shape)
print('Shape of test_userID = ', test_userID.shape)
```

```
Shape of X_train =  (40000, 48)
Shape of X_test =  (10000, 48)
Shape of train_userID =  (40000,)
Shape of test_userID =  (10000,)
```

Feature Scaling

In [ ]:

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train_sc = sc.fit_transform(X_train)
X_test_sc = sc.transform(X_test)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\preprocessing\data.py:645: DataConvers
ionWarning: Data with input dtype int32, int64 were all converted to float64 by StandardS
caler.
  return self.partial_fit(X, y)
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\base.py:464: DataConversionWarning: Da
ta with input dtype int32, int64 were all converted to float64 by StandardScaler.
  return self.fit(X, **fit_params).transform(X)
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:4: DataConversionWarning
: Data with input dtype int32, int64 were all converted to float64 by StandardScaler.
  after removing the cwd from sys.path.
```

Model Building

In [ ]:

```
# impoer requiede packages
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score
```

Decision Tree

In [ ]:

```
# Decision Tree Classifier
from sklearn.tree import DecisionTreeClassifier
dt_model = DecisionTreeClassifier(criterion= 'entropy', random_state=0)
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)

accuracy_score(y_test, y_pred_dt)
```

Out[ ]:

```
0.6936
```

In [ ]:

```
# train with Standert Scaling dataset
dt_model2 = DecisionTreeClassifier(criterion= 'entropy', random_state=0)
dt_model2.fit(X_train_sc, y_train)
y_pred_dt_sc = dt_model2.predict(X_test_sc)

accuracy_score(y_test, y_pred_dt_sc)
```

Out[ ]:

```
0.6932
```

K-NN

In [ ]:

```
from sklearn.neighbors import KNeighborsClassifier
knn_model = KNeighborsClassifier(n_neighbors=5, metric='minkowski', p=2,)
knn_model.fit(X_train, y_train)
y_pred_knn = knn_model.predict(X_test)

accuracy_score(y_test, y_pred_knn)
```

Out[ ]:

```
0.6994
```

In [ ]:

```
# train with Standert Scaling dataset
knn_model2 = KNeighborsClassifier(n_neighbors=5, metric='minkowski', p=2,)
knn_model2.fit(X_train_sc, y_train)
y_pred_knn_sc = knn_model2.predict(X_test_sc)

accuracy_score(y_test, y_pred_knn_sc)
```

Out[ ]:

```
0.7314
```

Naive Bayes

In [ ]:

```
# Naive Bayes
from sklearn.naive_bayes import GaussianNB
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)
y_pred_nb = nb_model.predict(X_test)

accuracy_score(y_test, y_pred_nb)
```

Out[ ]:

0.7114

In [ ]:

```
# train with Standert Scaling dataset
nb_model2 = GaussianNB()
nb_model2.fit(X_train_sc, y_train)
y_pred_nb_sc = nb_model2.predict(X_test_sc)

accuracy_score(y_test, y_pred_nb_sc)
```

Out[ ]:

0.7114

Random Forest

In [ ]:

```
# Random Forest Classifier
from sklearn.ensemble import RandomForestClassifier
rf_model = RandomForestClassifier(n_estimators=10, criterion='entropy', random_state=0)
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)

accuracy_score(y_test, y_pred_rf)
```

Out[ ]:

0.7621

In [ ]:

```
# train with Standert Scaling dataset
rf_model2 = RandomForestClassifier(n_estimators=10, criterion='entropy', random_state=0)
rf_model2.fit(X_train_sc, y_train)
y_pred_rf_sc = rf_model2.predict(X_test_sc)

accuracy_score(y_test, y_pred_rf_sc)
```

Out[ ]:

0.7616

Logistic Regression

In [ ]:

```
# Logistic Regression
from sklearn.linear_model import LogisticRegression
lr_model = LogisticRegression(random_state = 0, penalty = 'l1')
lr_model.fit(X_train, y_train)
y_pred_lr = lr_model.predict(X_test)

accuracy_score(y_test, y_pred_lr)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWa
rning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence thi
s warning.
  FutureWarning)
```

Out[ ]:

0.7684

In [ ]:

```
# train with Standert Scaling dataset
lr_model2 = LogisticRegression(random_state = 0, penalty = 'l1')
```

```
lr_model2.fit(X_train_sc, y_train)
y_pred_lr_sc = lr_model2.predict(X_test_sc)

accuracy_score(y_test, y_pred_lr_sc)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:433: FutureWa
rning: Default solver will be changed to 'lbfgs' in 0.22. Specify a solver to silence thi
s warning.
  FutureWarning)
```

Out[ ]:

0.7681

Support Vector Machine

In [ ]:

```
# Support Vector Machine
from sklearn.svm import SVC
svc_model = SVC()
svc_model.fit(X_train, y_train)
y_pred_svc = svc_model.predict(X_test)

accuracy_score(y_test, y_pred_svc)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The de
fault value of gamma will change from 'auto' to 'scale' in version 0.22 to account better
for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.
  "avoid this warning.", FutureWarning)
```

Out[ ]:

0.7616

In [ ]:

```
# train with Standert Scaling dataset
svc_model2 = SVC()
svc_model2.fit(X_train_sc, y_train)
y_pred_svc_sc = svc_model2.predict(X_test_sc)

accuracy_score(y_test, y_pred_svc_sc)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\svm\base.py:196: FutureWarning: The de
fault value of gamma will change from 'auto' to 'scale' in version 0.22 to account better
for unscaled features. Set gamma explicitly to 'auto' or 'scale' to avoid this warning.
  "avoid this warning.", FutureWarning)
```

Out[ ]:

0.779

'''from sklearn.svm import SVC grid_para = {'C':[1,10,100], 'gamma':[1, 0.01, 0.001], 'kernel':['rbf']} from sklearn.model_selection import GridSearchCV grid_lr = GridSearchCV(SVC(), param_grid = grid_para, refit = True, verbose = 4, n_jobs = -1) grid_lr.fit(X_train, y_train) grid_pred_lr = grid_lr.predict(X_test)

cm_grid_lr = confusion_matrix(y_test, grid_pred_lr) sns.heatmap(cm_grid_lr, annot = True, fmt = 'g')

accuracy_score(y_test, grid_pred_lr)'''

XGBoost

In [ ]:

```
# XGBoost Classifier
from xgboost import XGBClassifier
xgb_model = XGBClassifier()
```

```
xgb_model.fit(X_train, y_train)
y_pred_xgb = xgb_model.predict(X_test)

accuracy_score(y_test, y_pred_xgb)
```

Out[ ]:

0.7748

In [ ]:

```
# train with Standert Scaling dataset
xgb_model2 = XGBClassifier()
xgb_model2.fit(X_train_sc, y_train)
y_pred_xgb_sc = xgb_model2.predict(X_test_sc)

accuracy_score(y_test, y_pred_xgb_sc)
```

Out[ ]:

0.7748

In [ ]:

```
# XGB classifier with parameter tuning
xgb_model_pt1 = XGBClassifier(
 learning_rate =0.01,
 n_estimators=5000,
 max_depth=4,
 min_child_weight=6,
 gamma=0,
 subsample=0.8,
 colsample_bytree=0.8,
 reg_alpha=0.005,
 objective= 'binary:logistic',
 nthread=4,
 scale_pos_weight=1,
 seed=27)

xgb_model_pt1.fit(X_train, y_train)
y_pred_xgb_pt1 = xgb_model_pt1.predict(X_test)

accuracy_score(y_test, y_pred_xgb_pt1)
```

Out[ ]:

0.7887

In [ ]:

```
# XGB classifier with parameter tuning
# train with Standert Scaling dataset
xgb_model_pt2 = XGBClassifier(
 learning_rate =0.01,
 n_estimators=5000,
 max_depth=4,
 min_child_weight=6,
 gamma=0,
 subsample=0.8,
 colsample_bytree=0.8,
 reg_alpha=0.005,
 objective= 'binary:logistic',
 nthread=4,
 scale_pos_weight=1,
 seed=27)

xgb_model_pt2.fit(X_train_sc, y_train)
y_pred_xgb_sc_pt2 = xgb_model_pt2.predict(X_test_sc)

accuracy_score(y_test, y_pred_xgb_sc_pt2)
```

Out[ ]:

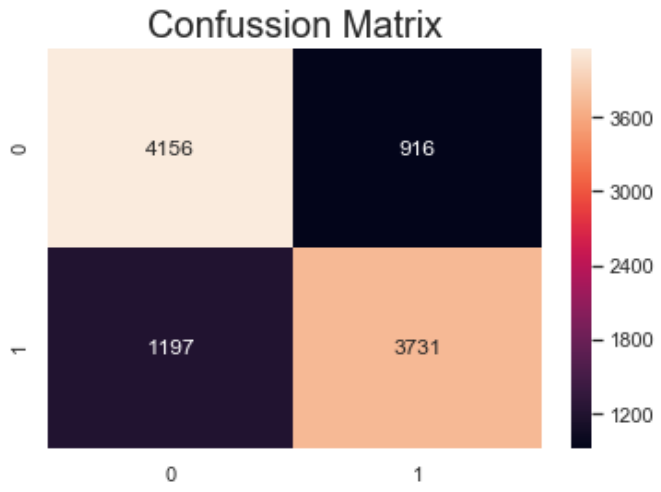
0.7887

In [ ]:

```
# confussion matrix
cm_xgb_pt2 = confusion_matrix(y_test, y_pred_xgb_sc_pt2)
sns.heatmap(cm_xgb_pt2, annot = True, fmt = 'g')
plt.title("Confussion Matrix", fontsize = 20)  # *****code 14
```

Out[ ]:

```
Text(0.5, 1.0, 'Confussion Matrix')
```



Mapping predicted output to the target

In [ ]:

```
final_result = pd.concat([test_userID, y_test], axis = 1)
final_result['predicted result'] = y_pred_xgb_sc_pt2

final_result
```

Out[ ]:

	user	enrolled	predicted result
11841	239786	1	1
19602	279644	1	1
45519	98290	0	0
25747	170150	1	1
42642	237568	1	0
31902	65042	1	0
30346	207226	1	1
12363	363062	0	0
32490	152296	1	1
26128	64484	0	0
14227	38108	1	1
26376	359940	0	0
44173	136089	0	0
12968	14231	1	1
32104	216038	0	0
17844	18918	1	1
43460	316730	1	1
8369	28308	1	0

	user	enrolled	predicted result
15055	228387	1	1
6338	69640	1	1
15301	358264	0	0
46250	348059	0	0
45580	178743	1	1
24647	167556	0	0
46712	294101	0	0
4150	192801	0	0
42460	163983	1	1
29079	298830	0	0
19412	151790	1	1
34839	20200	1	1
...	...	...	...
3380	348989	0	0
37623	248593	1	0
24852	316086	1	1
29372	192540	1	1
49639	256833	0	0
2930	273991	1	1
1210	365937	0	0
22652	295129	0	0
32360	255715	1	0
9171	37332	0	1
49037	164886	1	0
17793	309967	0	0
28887	14907	0	0
567	244737	1	1
662	284862	0	0
46038	60719	1	1
16778	262103	1	0
3075	243679	1	1
34793	280000	1	1
6557	255074	0	0
19150	347521	0	0
40096	335029	1	0
7869	37271	1	1
49546	240006	1	1
45202	279449	0	1
25091	143036	1	1
27853	91158	1	1
47278	248318	0	0
37020	142418	1	1
2217	279355	1	0

10000 rows × 3 columns