

The data is available as two attached CSV files: `takehome_user_engagement.csv` `takehome_users.csv` The data has the following two tables: 1] A user table ( "takehome\_users" ) with data on 12,000 users who signed up for the product in the last two years. This table includes: • name: the user's name • object\_id: the user's id • email: email address • creation\_source: how their account was created. This takes on one of 5 values: ○ PERSONAL\_PROJECTS: invited to join another user's personal workspace ○ GUEST\_INVITE: invited to an organization as a guest (limited permissions) ○ ORG\_INVITE: invited to an organization (as a full member) ○ SIGNUP: signed up via the website ○ SIGNUP\_GOOGLE\_AUTH: signed up using Google Authentication (using a Google email account for their login id)

• creation\_time: when they created their account • last\_session\_creation\_time: unix timestamp of last login • opted\_in\_to\_mailing\_list: whether they have opted into receiving marketing emails • enabled\_for\_marketing\_drip: whether they are on the regular marketing email drip • org\_id: the organization (group of users) they belong to • invited\_by\_user\_id: which user invited them to join (if applicable). 2] A usage summary table ( "takehome\_user\_engagement" ) that has a row for each day that a user logged into the product.

Defining an "adopted user" as a user who has logged into the product on three separate days in at least one seven-day period , identify which factors predict future user adoption . We suggest spending 12 hours on this, but you're welcome to spend more or less. Please send us a brief writeup of your findings (the more concise, the better no more than one page), along with any summary tables, graphs, code, or queries that can help us understand your approach. Please note any factors you considered or investigation you did, even if they did not pan out. Feel free to identify any further research or data you think would be valuable.

```
In [1]: # importing all the required libraries and modules
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import requests
import json
from pandas.io.json import json_normalize
from datetime import datetime, timedelta
import scipy.stats
import matplotlib.dates as mdates
import plotly.graph_objects as go
plt.style.use('bmh')
```

```
In [44]: #cwd = os.getcwd() # Get the current working directory (cwd)
#files = os.listdir(cwd) # Get all the files in that directory
#print("Files in %r: %s" % (cwd, files))
```

```
In [2]: # Checking the encoding of the csv files
with open('C:/Users/Karthi/Downloads/takehome_users.csv') as f:
    print(f)
    #for text in f:
        #print(text)
with open('C:/Users/Karthi/Downloads/takehome_user_engagement.csv') as f:
    print(f)
```

```
<_io.TextIOWrapper name='C:/Users/Karthi/Downloads/takehome_users.csv' mode='r' encoding='cp1252'>
<_io.TextIOWrapper name='C:/Users/Karthi/Downloads/takehome_user_engagement.csv' mode='r' encoding='cp1252'>
```

```
In [3]: # Importing the csv file into a dataframe using the above encoding

# Load as Pandas dataframe
users = pd.read_csv('C:/Users/Karthi/Downloads/takehome_users.csv', parse_dates = ['creation_time'],
                    encoding = "cp1252")
user_eng = pd.read_csv('C:/Users/Karthi/Downloads/takehome_user_engagement.csv', parse_dates =
                      encoding = "cp1252")
```

```
In [4]: users.info()
```

```
<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)
memory usage: 937.6+ KB
```

```
In [6]: users.describe()
```

```
Out[6]:
```

	object_id	last_session_creation_time	opted_in_to_mailing_list	enabled_for_marketing_drip	org_id
<b>count</b>	12000.00000	8.823000e+03	12000.000000	12000.000000	12000.000000
<b>mean</b>	6000.50000	1.379279e+09	0.249500	0.149333	141.884583
<b>std</b>	3464.24595	1.953116e+07	0.432742	0.356432	124.056723
<b>min</b>	1.00000	1.338452e+09	0.000000	0.000000	0.000000
<b>25%</b>	3000.75000	1.363195e+09	0.000000	0.000000	29.000000
<b>50%</b>	6000.50000	1.382888e+09	0.000000	0.000000	108.000000
<b>75%</b>	9000.25000	1.398443e+09	0.000000	0.000000	238.250000
<b>max</b>	12000.00000	1.402067e+09	1.000000	1.000000	416.000000

```
In [7]: user_eng.info()
```

```
<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)
memory usage: 4.8 MB
```

```
In [8]: user_eng.describe()
```

Out[8]:

	user_id	visited
count	207917.000000	207917.0
mean	5913.314197	1.0
std	3394.941674	0.0
min	1.000000	1.0
25%	3087.000000	1.0
50%	5682.000000	1.0
75%	8944.000000	1.0
max	12000.000000	1.0

In [9]: `user_eng.user_id.nunique()`

Out[9]: 8823

In [10]: `users.head()`

	object_id	creation_time	name	email	creation_source	last_session_creation_time	optec
0	1	2014-04-22 03:53:30	Clausen August	AugustCClausen@yahoo.com	GUEST_INVITE	1.398139e+09	
1	2	2013-11-15 03:45:04	Poole Matthew	MatthewPoole@gustr.com	ORG_INVITE	1.396238e+09	
2	3	2013-03-19 23:14:52	Bottrill Mitchell	MitchellBottrill@gustr.com	ORG_INVITE	1.363735e+09	
3	4	2013-05-21 08:09:28	Clausen Nicklas	NicklasSClausen@yahoo.com	GUEST_INVITE	1.369210e+09	
4	5	2013-01-17 10:14:20	Raw Grace	GraceRaw@yahoo.com	GUEST_INVITE	1.358850e+09	

In [11]: `user_eng.head()`

	time_stamp	user_id	visited
0	2014-04-22 03:53:30	1	1
1	2013-11-15 03:45:04	2	1
2	2013-11-29 03:45:04	2	1
3	2013-12-09 03:45:04	2	1
4	2013-12-25 03:45:04	2	1

Converting the last\_session\_creation\_time column in the users column to datetime.

In [12]: `users['last_session_creation_time'] = pd.to_datetime(users['last_session_creation_time'],unit='s')`

In [13]: `users.last_session_creation_time.dtypes`

Out[13]: `dtype('<M8[ns]')`

In [14]: `users.head(3)`

Out[14]:	object_id	creation_time	name	email	creation_source	last_session_creation_time	optec
<b>0</b>	1	2014-04-22 03:53:30	Clausen August	AugustCClausen@yahoo.com	GUEST_INVITE	2014-04-22 03:53:30	
<b>1</b>	2	2013-11-15 03:45:04	Poole Matthew	MatthewPoole@gustr.com	ORG_INVITE	2014-03-31 03:45:04	
<b>2</b>	3	2013-03-19 23:14:52	Bottrill Mitchell	MitchellBottrill@gustr.com	ORG_INVITE	2013-03-19 23:14:52	

```
In [15]: users['last_session_creation_time'].min(), users['last_session_creation_time'].max()
```

```
Out[15]: (Timestamp('2012-05-31 08:20:06'), Timestamp('2014-06-06 14:58:50'))
```

```
In [16]: users['creation_time'].min(), users['creation_time'].max()
```

```
Out[16]: (Timestamp('2012-05-31 00:43:27'), Timestamp('2014-05-30 23:59:19'))
```

```
In [17]: df = user_eng.copy()
```

```
In [18]: df['date'] = pd.to_datetime(df.time_stamp.dt.date)
```

```
In [19]: def rolling_count(df_group, frequency):
         return df_group.rolling(frequency, on='date')['user_id'].count()
```

```
In [20]: df['visits_7_days'] = df.groupby('user_id', as_index=False, group_keys=False).apply(rolling_count)
```

```
In [21]: df.describe().T
```

Out[21]:		count	mean	std	min	25%	50%	75%	max
	<b>user_id</b>	207917.0	5913.314197	3394.941674	1.0	3087.0	5682.0	8944.0	12000.0
	<b>visited</b>	207917.0	1.000000	0.000000	1.0	1.0	1.0	1.0	1.0
	<b>visits_7_days</b>	207917.0	4.381393	2.042666	1.0	3.0	4.0	6.0	7.0

```
In [22]: df[df.visits_7_days >= 3.0]
```

Out[22]:

	time_stamp	user_id	visited	date	visits_7_days
<b>9</b>	2014-02-09 03:45:04	2	1	2014-02-09	3.0
<b>10</b>	2014-02-13 03:45:04	2	1	2014-02-13	3.0
<b>27</b>	2013-02-19 22:08:03	10	1	2013-02-19	3.0
<b>30</b>	2013-03-02 22:08:03	10	1	2013-03-02	3.0
<b>31</b>	2013-03-05 22:08:03	10	1	2013-03-05	3.0
...	...	...	...	...	...
<b>207898</b>	2014-05-23 11:04:47	11988	1	2014-05-23	6.0
<b>207899</b>	2014-05-24 11:04:47	11988	1	2014-05-24	6.0
<b>207900</b>	2014-05-26 11:04:47	11988	1	2014-05-26	5.0
<b>207901</b>	2014-05-27 11:04:47	11988	1	2014-05-27	5.0
<b>207902</b>	2014-06-01 11:04:47	11988	1	2014-06-01	3.0

160522 rows × 5 columns

In [23]: `user_adopted = df.groupby('user_id')['visits_7_days'].max().reset_index()`

In [24]: `user_adopted['adopted_user'] = user_adopted['visits_7_days'].apply(lambda x: 1 if x>=3 else 0)`

In [25]: `user_adopted.head()`

Out[25]:

	user_id	visits_7_days	adopted_user
<b>0</b>	1	1.0	0
<b>1</b>	2	3.0	1
<b>2</b>	3	1.0	0
<b>3</b>	4	1.0	0
<b>4</b>	5	1.0	0

In [26]: `user_adopted.adopted_user.value_counts()`

Out[26]:

0	7221
1	1602

Name: adopted\_user, dtype: int64

In [27]: `user_adopted.drop('visits_7_days', axis = 1, inplace = True)`  
`user_adopted.rename(columns={"user_id": "object_id"}, inplace=True)`

In [28]: `user_adopted.set_index("object_id", inplace = True)`

In [29]: *#Joining the above user\_adopted dataframe with the original users dataframe.*  
`df_users = users.join(user_adopted, on = 'object_id', how='left')`

In [30]: `df_users.head()`

Out[30]:

	object_id	creation_time	name	email	creation_source	last_session_creation_time	optec
0	1	2014-04-22 03:53:30	Clausen August	AugustCClausen@yahoo.com	GUEST_INVITE	2014-04-22 03:53:30	
1	2	2013-11-15 03:45:04	Poole Matthew	MatthewPoole@gustr.com	ORG_INVITE	2014-03-31 03:45:04	
2	3	2013-03-19 23:14:52	Bottrill Mitchell	MitchellBottrill@gustr.com	ORG_INVITE	2013-03-19 23:14:52	
3	4	2013-05-21 08:09:28	Clausen Nicklas	NicklasSClausen@yahoo.com	GUEST_INVITE	2013-05-22 08:09:28	
4	5	2013-01-17 10:14:20	Raw Grace	GraceRaw@yahoo.com	GUEST_INVITE	2013-01-22 10:14:20	

In [31]:

```
df_users.info()

<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)
memory usage: 1.0+ MB
```

In [32]:

```
df_users['last_session_creation_time'].fillna(0, inplace = True)
df_users['adopted_user'].fillna(0, inplace = True)
```

In [33]:

```
df_users.describe().T
```

Out[33]:

	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

In [34]:

```
df_users['email_domain'] = df_users.email.apply(lambda x: x.split('@')[1])
```

In [35]:

```
df_users['email_domain'].value_counts()
```

```
Out[35]: 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
```

```
In [36]: # Also checking the creation_source for NON NULL invited_by_user column
         df_users[~df_users.invited_by_user_id.isnull()].creation_source.unique()
```

```
Out[36]: array(['GUEST_INVITE', 'ORG_INVITE'], dtype=object)
```

```
In [37]: df_users.drop(['object_id', 'name', 'email', 'email_domain'], axis = 1, inplace = True)
```

```
In [38]: df_users.invited_by_user_id.fillna(0, inplace=True)
```

```
In [39]: df_users['days_since_creation'] = (user_eng.time_stamp.max() - df_users.creation_time).dt.days
```

```
In [40]: df_users.drop(['creation_time', 'last_session_creation_time'], axis = 1, inplace = True)
```

```
In [41]: #Let's OneHotEncode the creation_source column.
         df_users = pd.get_dummies(df_users, columns=['creation_source'])
```

```
In [42]: df_users.describe().T
```

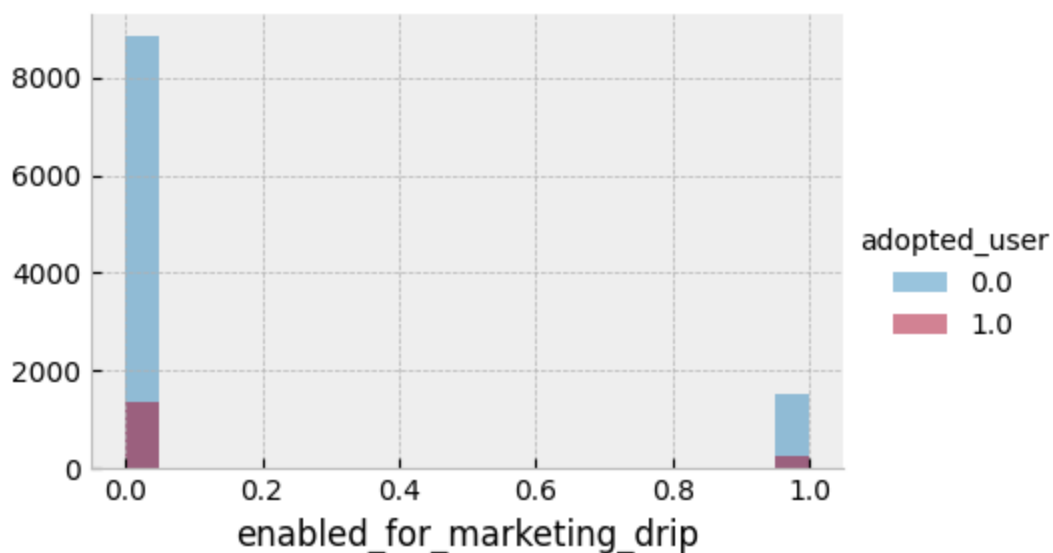
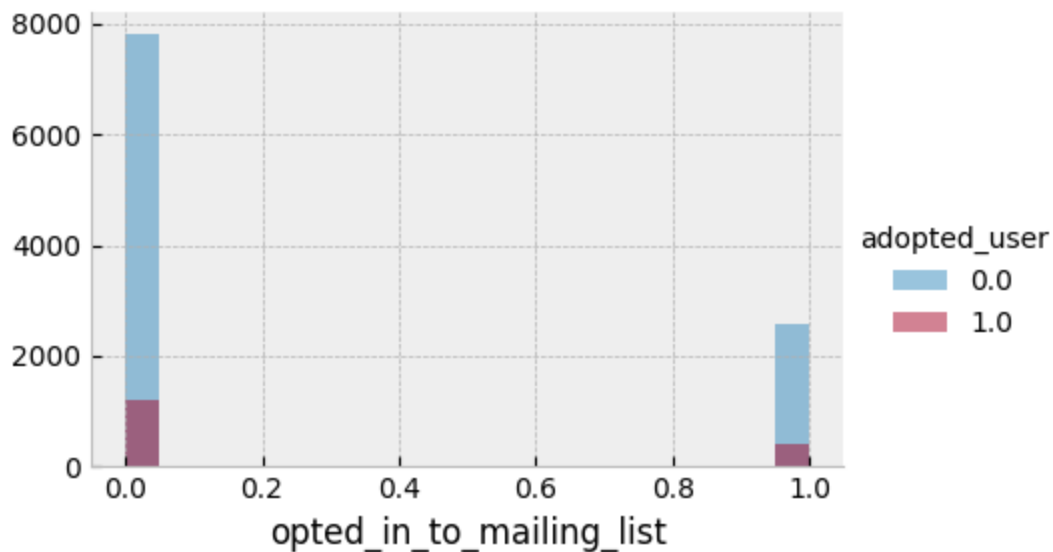
```
Out[42]:
```

	count	mean	std	min	25%	50%	75%	max
<b>opted_in_to_mailing_list</b>	12000.0	0.249500	0.432742	0.0	0.0	0.0	0.00	1.0
<b>enabled_for_marketing_drip</b>	12000.0	0.149333	0.356432	0.0	0.0	0.0	0.00	1.0
<b>org_id</b>	12000.0	141.884583	124.056723	0.0	29.0	108.0	238.25	416.0
<b>invited_by_user_id</b>	12000.0	3188.691333	3869.027693	0.0	0.0	875.0	6317.00	11999.0
<b>adopted_user</b>	12000.0	0.133500	0.340128	0.0	0.0	0.0	0.00	1.0
<b>days_since_creation</b>	12000.0	324.568000	216.646173	6.0	129.0	304.0	506.00	736.0
<b>creation_source_GUEST_INVITE</b>	12000.0	0.180250	0.384412	0.0	0.0	0.0	0.00	1.0
<b>creation_source_ORG_INVITE</b>	12000.0	0.354500	0.478381	0.0	0.0	0.0	1.00	1.0
<b>creation_source_PERSONAL_PROJECTS</b>	12000.0	0.175917	0.380765	0.0	0.0	0.0	0.00	1.0
<b>creation_source_SIGNUP</b>	12000.0	0.173917	0.379054	0.0	0.0	0.0	0.00	1.0
<b>creation_source_SIGNUP_GOOGLE_AUTH</b>	12000.0	0.115417	0.319537	0.0	0.0	0.0	0.00	1.0

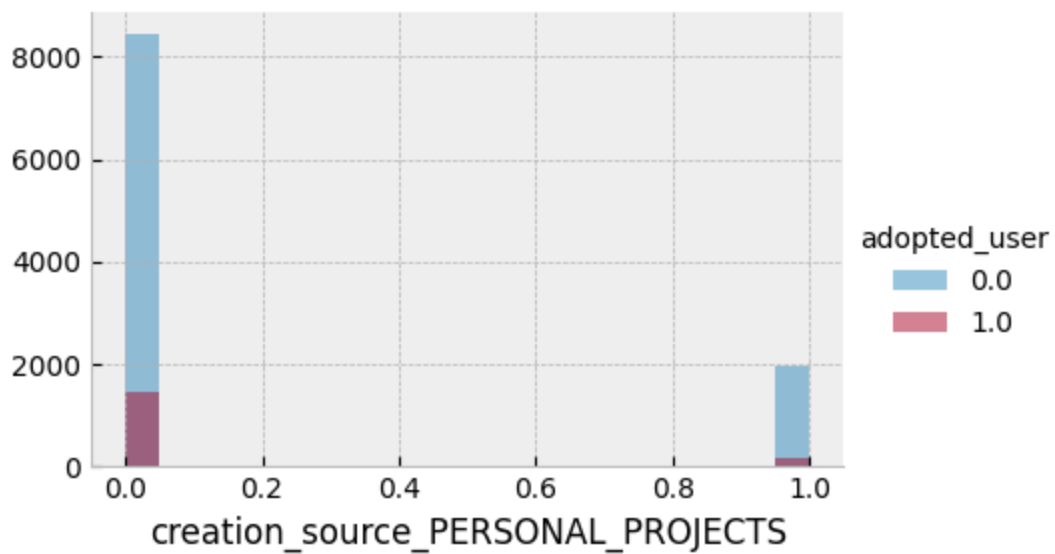
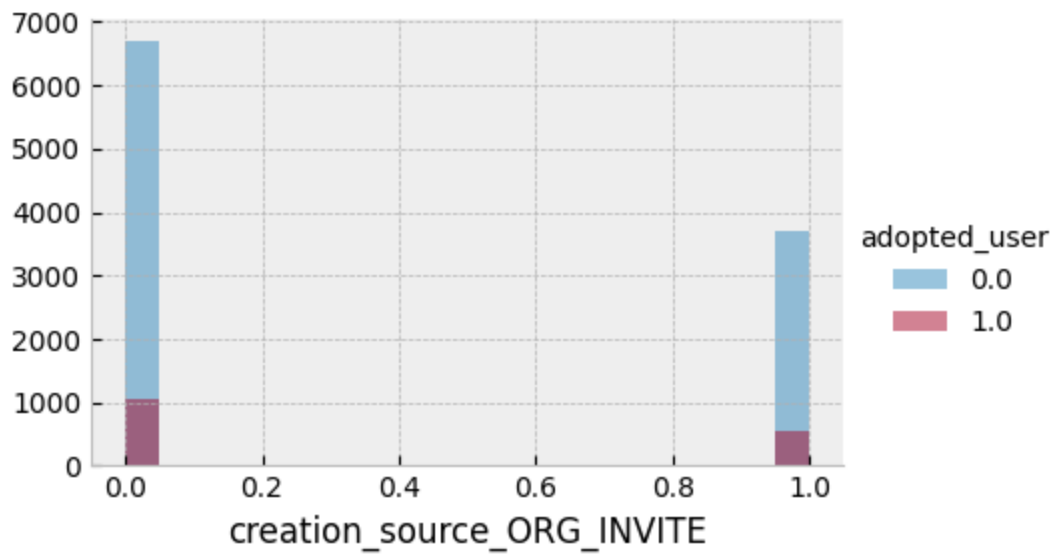
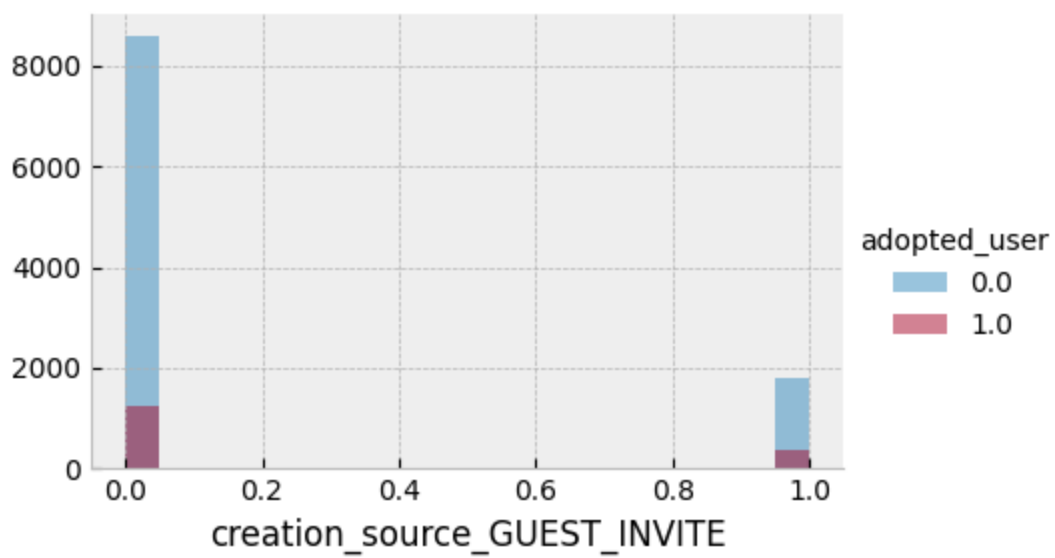
```
In [43]: df_users.corr()['adopted_user']
```

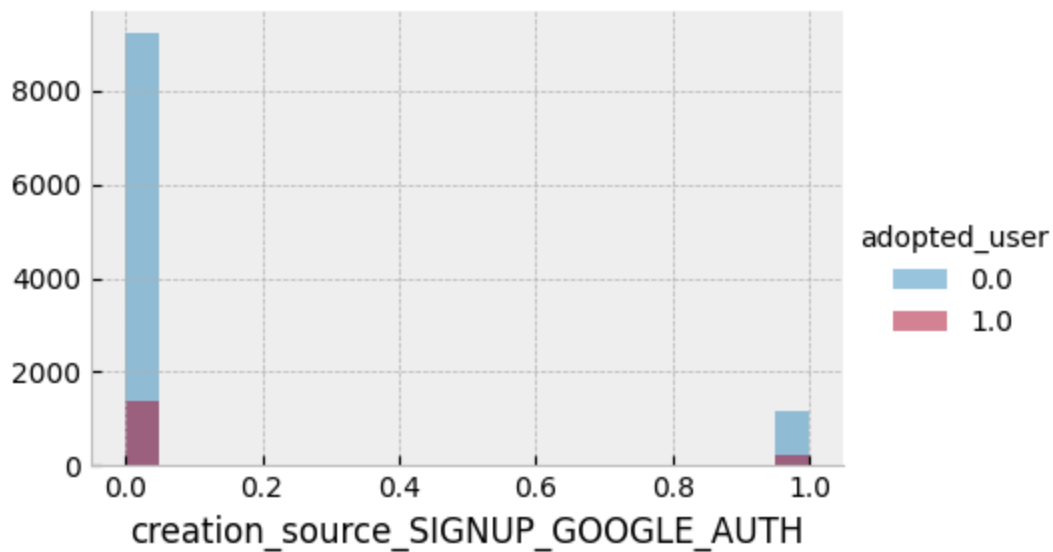
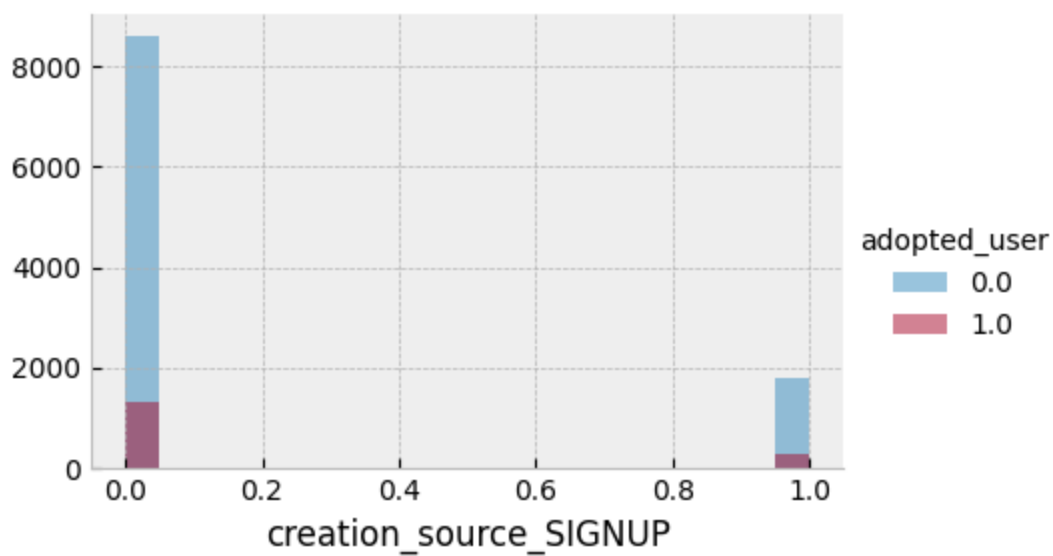
```
Out[43]: opted_in_to_mailing_list      0.008097
enabled_for_marketing_drip          0.004653
org_id                             0.064523
invited_by_user_id                  0.021596
adopted_user                        1.000000
days_since_creation                 0.088020
creation_source_GUEST_INVITE        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
Name: adopted_user, dtype: float64
```

```
In [44]: #sns.pairplot(df_users, hue = 'adopted_user')
for col in ['opted_in_to_mailing_list', 'enabled_for_marketing_drip', 'creation_source_GUEST_INVITE',
            'creation_source_ORG_INVITE', 'creation_source_PERSONAL_PROJECTS', 'creation_source_SIGNUP',
            'creation_source_SIGNUP_GOOGLE_AUTH']:
    g = sns.FacetGrid(df_users, hue = "adopted_user", height=3, aspect=1.5,)
    g.map(plt.hist, col, alpha=.5, bins = 20)
    g.add_legend()
```

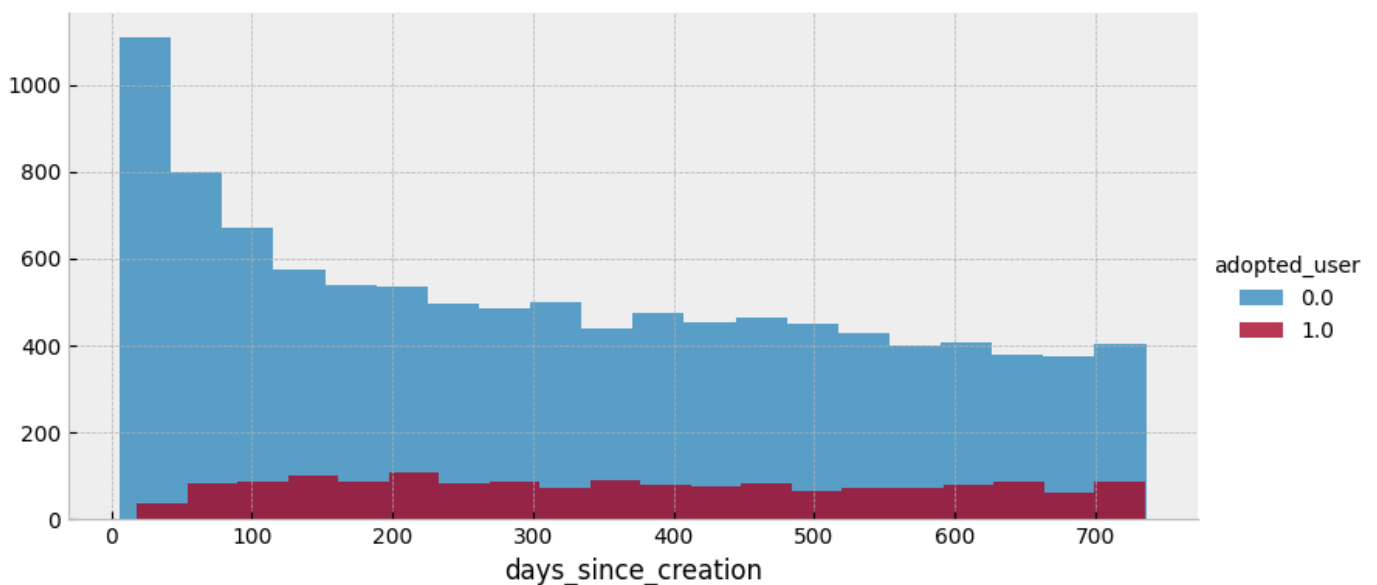








```
In [45]: g = sns.FacetGrid(df_users, hue = "adopted_user", height=4, aspect=2,)
_ = g.map(plt.hist, 'days_since_creation', alpha=0.8, bins = 20)
_ = g.add_legend()
#sns.distplot(df_users['days_since_creation'], kde = False, bins = 20, hue = 'adopted_user')
```



```
In [46]: # Importing necessary packages
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import f1_score, precision_score, recall_score, accuracy_score, confusion_m
```

```

from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV
import time
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score, roc_curve, a
precision_score, recall_score, f1_score

```

```

In [51]: def cv_optimize(model, parameters, Xtrain, ytrain, n_folds = 5):
        """
        Cross validation. Function to hypertune the model "model" with the input paramete distributio
        "parameters" on the training data.
        The output will be the best estimator whose average score on all folds will be best.
        """
        clf = GridSearchCV(model, param_grid = parameters, cv = n_folds, scoring = 'accuracy')
        t0 = time.time()
        clf.fit(Xtrain, ytrain)
        time_fit = time.time() - t0
        print('\n\n\n=====',type(model).__name__,'=====')
        print("It takes %.3f seconds for tuning " % (time_fit))
        print("BEST PARAMS", clf.best_params_)
        best = clf.best_estimator_
        return best

def do_classify(model, parameters, df, targetname, scale = True, cols_to_transform = 'numeric',
               featurenames = 'all', train_size = 0.8):

    # Creating the X and y variables for our model
    if featurenames == 'all':
        X = df.drop([targetname], axis = 1)
    else:
        X = df[featurenames]

    y = df[targetname]

    Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, train_size = train_size)

    model = cv_optimize(model, parameters, Xtrain, ytrain)
    t0 = time.time()
    model = model.fit(Xtrain, ytrain)
    time_fit = time.time() - t0
    print("It takes %.3f seconds for fitting" % (time_fit))
    training_accuracy = model.score(Xtrain, ytrain)
    test_accuracy = model.score(Xtest, ytest)
    precision = precision_score(ytest, model.predict(Xtest))
    recall = recall_score(ytest, model.predict(Xtest))
    AUC = roc_auc_score(ytest, model.predict_proba(Xtest)[:,:1])

    print("Accuracy on training data: {:.2f}".format(training_accuracy))
    print("Accuracy on test data:      {:.2f}".format(test_accuracy))
    print("Precision on test data:     {:.2f}".format(precision))
    print("Recall on test data:       {:.2f}".format(recall))
    print("AUC on test data:          {:.2f}".format(AUC))
    print("=====Confusion Matrix=====")
    print(confusion_matrix(ytest, model.predict(Xtest)))
    print("=====Classification report=====")
    print(classification_report(ytest, model.predict(Xtest)))
    print("="*100)
    print("="*100)
    print("="*100)
    return model, Xtrain, ytrain, Xtest, ytest

```

```

In [52]: # Random Forest model

```





```
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.  
3. To keep the past behaviour, explicitly set `max features='sqrt'` or remove this parameter as
```









[illegible]

```
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
```

```
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max features='sqrt'` or remove this parameter as
```







[illegible]

```
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
    warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
```

```
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p  
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max features='sqrt'` or remove this parameter as
```



it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.

```
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.  
warn(  
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.py:427: FutureWarning: `max features='auto'` has been deprecated in 1.1 and will be removed in 1.
```

[illegible]

```
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
y:427: FutureWarning: `max_features='auto'` has been deprecated in 1.1 and will be removed in 1.
3. To keep the past behaviour, explicitly set `max_features='sqrt'` or remove this parameter as
it is also the default value for RandomForestClassifiers and ExtraTreesClassifiers.
warn(
C:\Users\Karthi\PycharmProjects\pythonProject\venv\lib\site-packages\sklearn\ensemble\_forest.p
```

```
===== RandomForestClassifier =====
```

BEST PARAMS {'criterion': 'gini', 'max\_depth': 12, 'max\_features': 'auto', 'n\_estimators': 75}

	precision	recall	f1-score	support
0.0	0.88	0.75	0.81	2090
1.0	0.17	0.34	0.23	310
accuracy			0.70	2400
macro avg	0.53	0.55	0.52	2400
weighted avg	0.79	0.70	0.74	2400

=====

===

=====

===

=====

===

```
In [54]: feat_imp = pd.DataFrame({'importance': model_rf.feature_importances_})
          feat_imp['feature'] = Xtrain.columns
          feat_imp.sort_values(by='importance', ascending=False, inplace=True)

          feat_imp.sort_values(by='importance', inplace=True)
          feat_imp = feat_imp.set_index('feature', drop=True)
          feat_imp.plot.barh(title = 'Random Forest feature importance', figsize = (12,7))
```

Random Forest feature importance

