

Input

```
1 from google.colab import files
2
3 uploaded = files.upload()
4
5 for fn in uploaded.keys():
6     print('User uploaded file "{name}" with length {length} bytes'.format(
7         name=fn, length=len(uploaded[fn])))
```



Choose Files

No file chosen

Upload widget is only available when the cell has been executed in the current

browser session. Please rerun this cell to enable.

Saving final_dataset.csv to final_dataset.csv

User uploaded file "final_dataset.csv" with length 2119572 bytes

```
1 import json
2 import csv
3
4 import numpy as np
5 import pandas as pd
6
7 f_data_200 = pd.read_csv('final_dataset.csv')
8 f_data_200.describe()
```

	Unnamed: 0	ft_total_changes	ft_addns	ft_delns
count	3400.000000	3400.000000	3400.000000	3400.000000
mean	6470.702353	584.115000	385.935882	198.179118

Analysis

25%	2458.750000	2.000000	1.000000	1.000000
-----	-------------	----------	----------	----------

```
1 f_data_200.shape
```

(3400, 19)

```
1 f_data_200.head()
```

	Unnamed: 0	ft_uname	ft_rname	ft_commit_date	author_association	merged_status	pull_title
0	0	Oxflotus	contain-facebook	2018-11-26T22:17:57Z	CONTRIBUTOR	True	fixed Mu Account READM
1	1	AlexLakatos	Reps	2017-03-02T21:40:01Z	CONTRIBUTOR	True	adding t first issue the re newslett
2	2	AlexLakatos	Reps	2017-03-31T22:36:41Z	CONTRIBUTOR	True	added re newslett for mar
3	3	AlexLakatos	Reps	2017-05-05T10:03:54Z	CONTRIBUTOR	True	adding t april iss of t newslett
4	4	AlinSelagea	build-slaveapi	2017-10-25T08:42:22Z	CONTRIBUTOR	True	Revert b 139940

```
1 f_data = f_data_200
```

```
1 f_data.columns
```

Index(['Unnamed: 0', 'ft_uname', 'ft_rname', 'ft_commit_date', 'author_association', 'merged_status', 'pull_title', 'pull_msg', 'pullauthordate', 'pullid', 'ft_total_changes', 'ft_addns', 'ft_delns', 'commitauthordate', 'commitid', 'pullcommitsha', 'pullcommitmsg', 'repodesc', 'repolang'], dtype='object')
--

```
1 f_data.nunique
```

```
<bound method DataFrame.nunique of      Unnamed: 0  ...  repolang
0      0  ...  JavaScript
1      1  ...      HTML
2      2  ...      HTML
3      3  ...      HTML
4      4  ...      Python
...    ...  ...      ...
3395   12634  ...      Python
3396   12635  ...      Python
3397   12681  ...      Python
3398   12682  ...      CSS
3399   12683  ...      Python

[3400 rows x 19 columns]>
```

```
1 f_data.isna().sum()
2 # we can ignore these null messages, repository descriptions and languages
```

```
Unnamed: 0      0
ft_uname      0
ft_rname      0
ft_commit_date 0
author_association 0
merged_status  0
pull_title     0
pull_msg      1007
pullauthordate 0
pullid        0
ft_total_changes 0
ft_addns      0
ft_delns      0
commitauthordate 0
commitid      0
pullcommitsha  0
pullcommitmsg  0
repodesc      109
repolang      392
dtype: int64
```

```
1 # each user contributions to number of different repositories and sorted in decreasing order
2
3 user_diffrepos = f_data.groupby('ft_uname')['ft_rname'].nunique()
4 user_diffrepos.sort_values(ascending=False,inplace=True)
5 print(user_diffrepos)
6
```

```
ft_uname
Mozilla-GitHub-Standards    10
vladikoff                   7
mortonfox                   6
vbudhram                    6
spasovski                   6
..
lucasr                      1
lsblakk                     1
lowks                       1
lougeniaC64                 1
0xfloatus                   1
Name: ft_rname, Length: 750, dtype: int64
```

```

1 # each repo with different user's contributions to it and sorted in decreasing order
2 repo_diffusers = f_data.groupby('ft_rname')['ft_uname'].nunique()
3 repo_diffusers.sort_values(ascending=False,inplace=True)
4 print(repo_diffusers)

```

```

ft_rname
django-browserid      23
django-csp            23
infosec.mozilla.org  22
leadership-training   21
moz-corsica           20
..
fennec-distribution-sample  1
build-relengapi-clobberer    1
etp-search-volume-study      1
eslint-plugin-amo            1
minion-vm                    1
Name: ft_uname, Length: 305, dtype: int64

```

```

1 # user_diffrepos_total
2 f_data.groupby(['ft_uname','ft_rname']).size()
3 #print(user_diffrepos_total_changes_count)

```

```

ft_uname    ft_rname
0xflotus    contain-facebook      1
AlexLakatos  Reps                  3
AlinSelagea  build-slaveapi        1
AlinT        bouncer-tests         1
AlphaRoy14   mozilla-campus-clubs  1
..
zacc         input-tests            1
zalun        Marketplace.Python    10
zbraniecki   b2gperf               1
zee-moz      leadership-training    1
zyegfryed    bleach                 1
Length: 1078, dtype: int64

```

```

1 # # user_diffrepos_total_changes_count
2 user_diffrepos_total_changes_count = f_data.groupby(['ft_uname','ft_rname'])['ft_total_changes']
3 print(user_diffrepos_total_changes_count)

```

```

ft_uname    ft_rname
0xflotus    contain-facebook      2
AlexLakatos  Reps                  3210
AlinSelagea  build-slaveapi        25
AlinT        bouncer-tests         10
AlphaRoy14   mozilla-campus-clubs      8
...
zacc         input-tests            4
zalun        Marketplace.Python    859
zbraniecki   b2gperf               13
zee-moz      leadership-training    2
zyegfryed    bleach                 4
Name: ft_total_changes, Length: 1078, dtype: int64

```

```

1 # # user_diffrepos_total_ft_addns_count
2 user_diffrepos_total_addns_count = f_data.groupby(['ft_uname','ft_rname'])['ft_addns'].sum()
3 print(user_diffrepos_total_addns_count)

```

```

ft_uname    ft_rname

```

0xfloatus	contain-facebook	1
AlexLakatos	Reps	3210
AlinSelagea	build-slaveapi	15
AlinT	bouncer-tests	7
AlphaRoy14	mozilla-campus-clubs	4
	...	
zacc	input-tests	2
zalun	Marketplace.Python	791
zbraniecki	b2gperf	9
zee-moz	leadership-training	1
zyegfryed	bleach	2

Name: ft_addns, Length: 1078, dtype: int64

```

1 # # user_diffrepos_ft_delns_count
2 user_diffrepos_ft_delns_count = f_data.groupby(['ft_uname','ft_rname'])['ft_delns'].sum()
3 print(user_diffrepos_ft_delns_count)

```

ft_uname	ft_rname	
0xfloatus	contain-facebook	1
AlexLakatos	Reps	0
AlinSelagea	build-slaveapi	10
AlinT	bouncer-tests	3
AlphaRoy14	mozilla-campus-clubs	4
	..	
zacc	input-tests	2
zalun	Marketplace.Python	68
zbraniecki	b2gperf	4
zee-moz	leadership-training	1
zyegfryed	bleach	2

Name: ft_delns, Length: 1078, dtype: int64

```

1 # # user_diffrepos_repolang_count
2 user_diffrepos_repolang_count = f_data.groupby(['ft_uname','ft_rname'])['repolang'].nunique()
3 print(user_diffrepos_repolang_count)
4 user_diffrepos_repolang_count.sort_values(ascending=False,inplace=True)
5 print(user_diffrepos_repolang_count)

```

ft_uname	ft_rname	
0xfloatus	contain-facebook	1
AlexLakatos	Reps	1
AlinSelagea	build-slaveapi	1
AlinT	bouncer-tests	0
AlphaRoy14	mozilla-campus-clubs	1
	..	
zacc	input-tests	0
zalun	Marketplace.Python	1
zbraniecki	b2gperf	1
zee-moz	leadership-training	1
zyegfryed	bleach	1

Name: repolang, Length: 1078, dtype: int64

ft_uname	ft_rname	
zyegfryed	bleach	1
emceeaich	moz-corsica	1
espadrine	doctorjs	1
escapewindow	configman	1
erogol	TTS	1
	..	
nchapman	identity-pubkeys	0
tobbi	Addon-Tests	0
	input-tests	0
TDA	fxa-auth-mailer	0
chutten	data-review	0

Name: repolang, Length: 1078, dtype: int64

```

1 #repo_languages_count
2 repo_languages_count = f_data.groupby(['repolang']).size()
3 print(repo_languages_count)
4 repo_languages_count.sort_values(ascending=False,inplace=True)
5 print(repo_languages_count)

```

```

repolang
ApacheConf          1
Brightscript        1
C                   10
C++                 40
CSS                 245
Clojure             8
Dockerfile          2
FreeMarker          55
GCC Machine Description  2
Go                  20
HTML                81
Java                69
JavaScript          1152
Jupyter Notebook    7
Less                18
Lua                 1
Makefile            10
Objective-C          4
PHP                 149
Python              923
Rust                112
Shell               97
Swift               1
dtype: int64
repolang
JavaScript          1152
Python              923
CSS                 245
PHP                 149
Rust                112
Shell               97
HTML                81
Java                69
FreeMarker          55
C++                 40
Go                  20
Less                18
C                   10
Makefile            10
Clojure             8
Jupyter Notebook    7
Objective-C          4
GCC Machine Description  2
Dockerfile          2
Brightscript        1
Swift               1
Lua                 1
ApacheConf          1
dtype: int64

```

```

1 #user_conributions_count
2 conributions_languages_count = f_data.groupby(['ft_urname']).size()
3 print(conributions_languages_count)
4 conributions_languages_count.sort_values(ascending=False,inplace=True)
5 print(conributions_languages_count)

```

```

ft_username
0xflotus      1
AlexLakatos   3
AlinSelagea    1
AlinT          1
AlphaRoy14     1
..
zacc           1
zalun         10
zbraniecki     1
zee-moz        1
zyegfryed      1
Length: 750, dtype: int64
ft_username
oremj          10
akatsoulas     10
rohit-rk       10
andrewhayward  10
robhudson      10
..
kanru          1
keianhzo       1
kinetiknz      1
kingthorin     1
0xflotus       1
Length: 750, dtype: int64

```

```

1 list_users = dict(contributions_languages_count)
2 print(list_users)
3 count = 0
4 for key, pair in list_users.items():
5     if pair == 10:
6         count = count + 1
7 print(count)

```

```

{'oremj': 10, 'akatsoulas': 10, 'rohit-rk': 10, 'andrewhayward': 10, 'robhudson': 10, 'lmorchar': 10, 'zacc': 1, 'zbraniecki': 1, 'zyegfryed': 1, 'kingthorin': 1, 'kanru': 1, 'keianhzo': 1, 'kinetiknz': 1, 'zalun': 10, 'AlexLakatos': 3, 'AlinSelagea': 1, 'AlinT': 1, 'AlphaRoy14': 1, '0xflotus': 1}
206

```

```

1 #user_contributions_count
2 contributions_languages_count = f_data.groupby(['ft_username']).size().agg(['mean', 'min', 'median', 'count'])
3 print(contributions_languages_count)
4 contributions_languages_count.sort_values(ascending=False, inplace=True)
5 print(contributions_languages_count)

```

```

mean          4.533333
min            1.000000
median        3.000000
count        750.000000
dtype: float64
count        750.000000
mean          4.533333
median        3.000000
min            1.000000
dtype: float64

```

```

1 #contributions_languages_count = 10 print
2 #contributions_languages_count.shape

```

```

1 # # user total changes 10 news

```

```

1 # # user_total_changes_10rows
2 user_total_changes_10rows = f_data.groupby(['ft_uname'])#['ft_total_changes'].sum()
3 print(user_total_changes_10rows)

```

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7f4dbc6f2890>

```

1 user_total_changes_10rows.groups

```

```
{'0xfloetus': [0], 'AlexLakatos': [1, 2, 3], 'AlinSelagea': [4], 'AlinT': [5], 'AlphaRoy14': [6]}
```

```

1 #user_total_changes_10rows = user_total_changes_10rows.sort_values(ascending=False,inplace=True)
2 user_total_changes_10rows.groups
3 #user_total_changes_10rows.get_group('AlexLakatos')

```

```
{'0xfloetus': [0], 'AlexLakatos': [1, 2, 3], 'AlinSelagea': [4], 'AlinT': [5], 'AlphaRoy14': [6]}
```

```

1 each_count =0
2 for each in user_total_changes_10rows.groups:
3     #user_total_changes_10rows.get_group(each)
4     #print(each)
5     name = str(each)
6     each_count= each_count+1
7     print(name)
8     print(user_total_changes_10rows.get_group(name))
9
10 print(each_count)

```

Streaming output truncated to the last 5000 lines.

```

bobsilverberg
  Unnamed: 0      ft_uname  ...      repodesc repolang
725      2824 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
726      2827 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
727      2830 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
728      2833 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
729      2836 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
730      2839 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
731      2842 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
732      2845 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
733      2848 bobsilverberg  ...  Tests for Mozilla Input website.      NaN
734      2850 bobsilverberg  ...  Tests for Mozilla Input website.      NaN

```

[10 rows x 19 columns]

```

brampitoyo
  Unnamed: 0  ... repolang
735      2973  ...      NaN
736      2974  ...      NaN
737      2975  ...      NaN

```

[3 rows x 19 columns]

```

bramwelt
  Unnamed: 0  ft_uname  ...      repodesc repolang
738      2976 bramwelt  ...  the grand unified configuration system  Python

```

[1 rows x 19 columns]

```

brandonsavage
  Unnamed: 0  ... repolang
739      2977  ...      PHP
740      2978  ...      PHP
741      2979  ...      PHP

```



```

742      2980  ...    Python
743      2981  ...    Python
744      2982  ...    Python
745      2983  ...    Python
746      2984  ...    Python
747      2985  ...      PHP
748      2987  ...      PHP

```

[10 rows x 19 columns]

brendandahl

```

      Unnamed: 0  ... repolang
749      3016  ...      C++
750      3017  ...      C++
751      3018  ...      C++
752      3019  ...      C++
753      3020  ...      C++
754      3021  ...      C++
755      3022  ...      C++
756      3023  ...      C++
757      3024  ...      C++
758      3025  ...      C++

```

[10 rows x 19 columns]

brettgaylor

```

      Unnamed: 0      ft_uname  ... repodesc      repolang
759      3033  brettgaylor  ...      NaN  JavaScript

```

...

```

1 '''
2 each_count =0
3 for each in user_total_changes_10rows.groups:
4     #user_total_changes_10rows.get_group(each)
5     #print(each)
6     name = str(each)
7     each_count= each_count+1
8     print(name)
9     print(user_total_changes_10rows.get_group(name))
10
11 print(each_count)
12 '''
13 user_name_list =[]
14 user_total_changes =[]
15 user_total_addns =[]
16 user_total_delns =[]
17 user_repo_lang = []
18
19 all_users_total_changes = []
20 all_users_total_addns = []
21 all_users_total_delns = []
22
23 all_users_repo_lang =[]
24 each_count=0
25 for each in user_total_changes_10rows.groups:
26     #user_total_changes_10rows.get_group(each)
27     #print(each)
28     name = str(each)
29     each_count= each_count+1
30     #print(name)
31     #print(user_total_changes_10rows.get_group(name))
32     user_name_list.append(name)
33     user_total_changes.append(user_total_changes_10rows.get_group(name)['ft_total_changes'])
34     user_total_addns.append(user_total_changes_10rows.get_group(name)['ft_addns'])

```

```

34 user_total_addns.append(user_total_changes_10rows.get_group(name)['ft_addns'])
35 user_total_delns.append(user_total_changes_10rows.get_group(name)['ft_delns'])
36 user_repo_lang.append(user_total_changes_10rows.get_group(name)['repo_lang'])
37
38 all_users_total_changes.append(user_total_changes)
39 all_users_total_addns.append(user_total_addns)
40 all_users_total_delns.append(user_total_delns)
41 all_users_repo_lang.append(user_repo_lang)
42
43 print(each_count)
44
45 #print(user_total_changes_10rows.get_group(name)['ft_total_changes'])

```

750

▼ Plots:

All plots from pandas: https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html
<https://towardsdatascience.com/a-guide-to-pandas-and-matplotlib-for-data-exploration-56fad95f951c>
<https://towardsdatascience.com/5-minute-guide-to-plotting-with-pandas-e8c0f40a1df4>

```

1 #all plots for final submission goes here..
2

```

```

1 #1st individual plot for total changes in ten contributions
2 #user= list_users[0]
3 total_changes =[]
4 total_list_of_changes =[]
5
6 for key, value in list_users.items():
7     #if f_data['ft_rname'] == key:
8     #    total_changes.append(f_data['ft_total_changes'])
9     total_list_of_changes.append(total_changes)
10 print(total_changes)
11 print(total_list_of_changes)

```

```

[]
[[], [], [], [], [], [], [], [], [], [], [], [], [], [], [], [], [], [], [], [], []

```

```

1 print(all_users_total_changes[1][0])

```

```

0      2
Name: ft_total_changes, dtype: int64

```

```

1 !pip install mpld3

```

Collecting mpld3

Downloading <https://files.pythonhosted.org/packages/7d/b4/f380b6d58658106870161d703972b74fc2/>
| 890kB 9.0MB/s

Requirement already satisfied: jinja2 in /usr/local/lib/python3.7/dist-packages (from mpld3) (2.11.3)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (from mpld3) (3.3.0)
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from jinja2->mpld3) (2.0.1)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib->mpld3) (2.8.2)

```
Requirement already satisfied: pyparsing!=2.0.4,!2.1.2,!2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib)
Requirement already satisfied: numpy>=1.11 in /usr/local/lib/python3.7/dist-packages (from matplotlib)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from matplotlib)
Building wheels for collected packages: mpld3
  Building wheel for mpld3 (setup.py) ... done
  Created wheel for mpld3: filename=mpld3-0.5.2-cp37-none-any.whl size=200617 sha256=9ceb2a59f6
  Stored in directory: /root/.cache/pip/wheels/21/73/06/ea4b85609301850b1289a282852d92e22fcbf7
Successfully built mpld3
Installing collected packages: mpld3
Successfully installed mpld3-0.5.2
```

```
1 !pip install pandas-bokeh
2
3 import pandas as pd
4 import pandas_bokeh
5
6 pandas_bokeh.output_notebook()
7
8
```

```
Collecting pandas-bokeh
  Downloading https://files.pythonhosted.org/packages/c6/08/a1fb196d9a58893bbf67b13a7db3973f39
Requirement already satisfied: pandas>=0.22.0 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: bokeh>=2.0 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: numpy>=1.15.4 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: Jinja2>=2.7 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: typing-extensions>=3.7.4 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Installing collected packages: pandas-bokeh
Successfully installed pandas-bokeh-0.5.5
```

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 import mpld3
4
5 user_name_list = []
6 user_total_changes = []
7 user_total_addns = []
8 user_total_delns = []
9 user_repo_lang = []
10
11 all_users_total_changes = []
12 all_users_total_addns = []
13 all_users_total_delns = []
14
15 all_users_repo_lang = []
16 each_count=0
17 fig = plt.figure(figsize = (20, 10))
18 df_list = []
```

```

19 for each in user_total_changes_10rows.groups:
20     #user_total_changes_10rows.get_group(each)
21     #print(each)
22     name = str(each)
23     #each_count= each_count+1
24     #print(name)
25     #print(user_total_changes_10rows.get_group(name))
26     user_total_changes = []
27     user_total_addns = []
28     user_total_delns = []
29     user_repo_lang = []
30     indexes = []
31     user_name_list.append(name)
32     for i in user_total_changes_10rows.get_group(name)['ft_total_changes']:
33         user_total_changes.append(i)
34     num = len(user_total_changes)
35     #print(len(user_total_changes))
36     for i in user_total_changes_10rows.get_group(name)['ft_addns']:
37         user_total_addns.append(i)
38     for i in user_total_changes_10rows.get_group(name)['ft_delns']:
39         user_total_delns.append(i)
40     for i in user_total_changes_10rows.get_group(name)['repolang']:
41         user_repo_lang.append(i)
42     for i in user_total_changes_10rows.get_group(name)['ft_rname']:
43         indexes.append(i)
44     '''
45     user_total_addns = np.array(user_total_addns)
46     user_total_changes = np.array(user_total_changes)
47     user_total_delns = np.array(user_total_delns)
48     '''
49     #print(user_total_addns.shape)
50     #print(user_total_changes.shape)
51     idx = []
52     for j in range(0,num):
53         idx.append(j)
54     df = pd.DataFrame({
55         "total_changes": user_total_changes,
56         "addition_changes": user_total_addns,
57         "deletion_changes": user_total_delns,
58         #'repo_changes':user_repo_lang,
59     }, index=indexes,columns = ['total_changes','addition_changes','deletion_changes'])
60     #plt.figure()
61     df['total_changes']= df['total_changes'].astype(int)
62     df['addition_changes']= df['addition_changes'].astype(int)
63     df['deletion_changes']= df['deletion_changes'].astype(int)
64
65     df=df.astype(int)
66     plt.title(name)
67     #df_list.append(df)# for individual comment this line
68     if len(indexes) == 10:
69         lines = df.plot_bokeh(kind = 'line',title = name)
70     '''for fr in df_list:
71         #fr.plot()
72         plt.plot(fr['total_changes'],fr['addition_changes'],fr['deletion_changes'])'''
73 plt.xlim(0,10)
74 plt.ylim(0,100)
75 plt.show()
76
77 html_str = mpld3.fig_to_html(fig)

```

```
78 Html_file= open("contributions.html","w")
79 Html_file.write(html_str)
80 Html_file.close()
81 '''
82 all_users_total_changes.append(user_total_changes)
83 all_users_total_addns.append(user_total_addns)
84 all_users_total_delns.append(user_total_delns)
85 all_users_repo_lang.append(user_repo_lang)'''
```

```

1 '''
2 name = user_name_list[4]
3
4 t_list = all_users_total_changes[]
5 print(t_list)
6 a_list = all_users_total_addns[4]
7 print(len(a_list))
8 d_list = all_users_total_delns[4]
9 print(len(d_list))
10 r_list = all_users_repo_lang[4]
11 print(len(r_list))'''
12
13 '''
14 df = pd.DataFrame({
15     'total_changes':t_list,
16     'addition_changes':a_list,
17     'deletion_changes':d_list,
18     'repo_changes':r_list,
19     }, index=[0,1,2,3,4,5,6,7,8,9])
20
21 lines = df.plot.line()
22 '''

```

```

'\ndf = pd.DataFrame({\n    'total_changes':t_list,\n    'addition_changes':a_list,\n    'deletion_changes':d_list,\n    'repo_changes':r_list,\n    }, index=[0,1,2,3,4,5,6,7,8,9])\n\nlines = df.plot.l:

```

```

1 #repo_languages_count
2 repo_languages_count = f_data.groupby(['repolang']).size()
3 print(repo_languages_count)
4 #repo_languages_count.sort_values(ascending=False,inplace=True)
5 #print(repo_languages_count)
6 lang_names =[]
7 for i in f_data.groupby(['repolang']).groups:
8     name = str(i)
9     lang_names.append(name)
10 lang_count =[]
11 for i in repo_languages_count:
12     lang_count.append(i)
13     #lang_names.append(repo_languages_count.get_group(i))
14
15 #print()
16 print(lang_names)
17 print(lang_count)
18

```

repolang	
ApacheConf	1
Brightscript	1
C	10
C++	40
CSS	245
Clojure	8
Dockerfile	2
FreeMarker	55
GCC Machine Description	2
Go	20

```

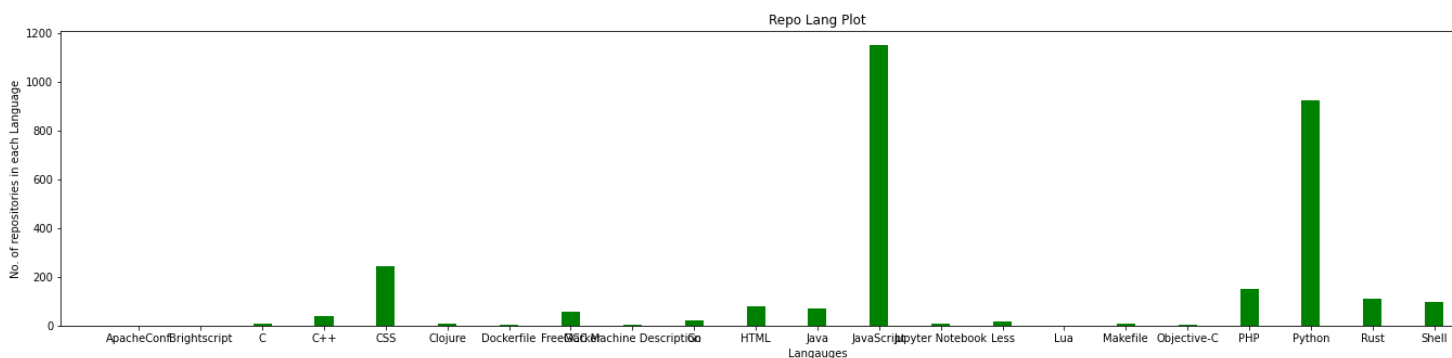
HTML 81
Java 69
JavaScript 1152
Jupyter Notebook 7
Less 18
Lua 1
Makefile 10
Objective-C 4
PHP 149
Python 923
Rust 112
Shell 97
Swift 1
dtype: int64
['ApacheConf', 'Brightscript', 'C', 'C++', 'CSS', 'Clojure', 'Dockerfile', 'FreeMarker', 'GCC M
[1, 1, 10, 40, 245, 8, 2, 55, 2, 20, 81, 69, 1152, 7, 18, 1, 10, 4, 149, 923, 112, 97, 1]

```

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 lang_df = pd.DataFrame({
5     "lang_count": lang_count,
6 },columns = lang_names)
7
8 plt.figure(figsize = (25, 5))
9 '''
10 lang_df=lang_df.astype(int)
11 lang_df.plot.bar()
12 plt.axhline(0, color="k")
13 '''
14 plt.bar(lang_names, lang_count, color ='green',width = 0.3)
15
16 plt.xlabel("Langauges")
17 plt.ylabel("No. of repositories in each Language")
18 plt.title("Repo Lang Plot")
19 plt.show()

```



```

1 # user_diffrepos_total
2 users_five_repositories = f_data.groupby(['ft_uname','ft_rname']).size()
3 #print(user_diffrepos_total_changes_count)
4 users_five_repositories.sort_values(ascending=False,inplace=True)
5 print(users_five_repositories)
6

```

```
ft_uname      ft_rname
```

cadecairos	MakeAPI	10
kparlante	kpiggybank	10
bobsilverberg	input-tests	10
tomato42	cipherscan	10
klahnakoski	measure-noise	10
		..
keianhzo	androidspeech	1
kanru	layerscope	1
k88hudson	make-valet	1
jvillalobos	addons-contribution-notifier	1
0xfloatus	contain-facebook	1
Length: 1078, dtype: int64		

```

1 f_count =0
2 for i in users_five_repositories:
3     if i>=5:
4         #print(i," ")#users_five_repositories.group(i))
5         f_count = f_count +1
6 print("count is", f_count)

```

count is 256

```

1 # user_diffrepos_total
2 users_five_repositories_groups = f_data.groupby(['ft_urname','ft_rname'])#.sort_values(ascending=F
3 #print(user_diffrepos_total_changes_count)
4 #users_five_repositories_groups#.sort_values(ascending=False,inplace=True)
5 #print(users_five_repositories_groups.groups)
6 for each in users_five_repositories_groups.groups:
7     #if(each.size>=5):
8     #print(each)
9     if(users_five_repositories[each]>=5):#users_five_repositories_groups.get_group(each).size >=5
10         print(each)
11
12

```

```

('AutomatedTester', 'input-tests')
('CJKu', 'layerscope')
('Callek', 'build-relengapi-slaveloan')
('Dexterp37', 'firefox-hardware-report')
('EricEdens', 'fennec-search')
('FernandaBernardo', 'games.mozilla.org')
('FrancescoSTL', 'bakery')
('Gozala', 'firefox-narrate-experiment')
('KWierso', 'deprecated-treestatus')
('LegNeato', 'mediawiki-bugzilla')
('Marlena', 'Addon-Tests')
('MartinBest', 'moz-corsica')
('MichaelKohler', 'activate.mozilla.community')
('MihaiTabara', 'build-funsize')
('MikkCZ', 'One-Mozilla-blog')
('Mossop', 'firefox-browser-architecture')
('Pomax', 'friendlycode')
('ReonSaji', 'OpenDesign')
('Rupal-IIITD', 'inclusion')
('ScottDowne', 'make-valet')
('Standard8', 'hello-tests')
('SuYoungHong', 'dscontrib')
('TDA', 'fxa-geodb')
('TrevorFSmith', 'action-input')
('Uberi', 'cerberus')
('WilsonKathleen', 'ca-policy')
('Zearin', 'html5-lint')

```



```
(
('acabunoc', 'leadership-training')
('acmiyaguchi', 'mozilla-schema-generator')
('adamkalis', 'buddyup')
('agjwong', 'jupyter-notebook-gist')
('ajvb', 'frost')
('akatsoulas', 'feedthefox')
('akshitac8', 'inclusion')
('alanmoo', 'mofo-bootstrap')
('alevgibson', 'meao')
('alicoding', 'goggles.mozilla.org')
('almossawi', 'marketplace-stats')
('andreastt', 'geckodriver')
('andrewhayward', 'content-2012.mozillafestival.org')
('andrewkrug', 'guardduty-multi-account-manager')
('april', 'minion-vm')
('armenzg', 'firefox-code-coverage-frontend')
('arunk-s', 'libaudit-go')
('asalga', 'mozilla-ignite-learning-lab-demos')
('bacharakis', 'mozilla-campus-clubs')
('bear', 'briar-patch')
('bebef1987', 'input-tests')
('benadida', 'browserid-crypto')
('bensterthal', 'fxoss')
('bgrins', 'firefox-browser-architecture')
('bhearsum', 'balrog-ui')
('biancadanforth', 'lightbeam-we')
('blairmacintyre', 'aframe-xr')
('bnjbvr', 'arewefastyet')
('bobsilverberg', 'input-tests')
('brandonsavage', 'dragnet')
('brandonsavage', 'mediawiki-bugzilla')
('brendandahl', 'libadb.js')
('brunoleonard', 'libadskit-erl')
```

```
1 # user_diffrepos_total
2 users_five_repositories = f_data.groupby(['ft_urname']).nunique()
3 #print(user_diffrepos_total_changes_count)
4 users_five_repositories.sort_values(by = 'ft_rname',ascending=False,inplace=True)
5 #for i in users_five_repositories##:
6 #     if i >=5:
7 print(users_five_repositories['ft_rname'])
8
9 final_each_user_to_five_repos_series = users_five_repositories['ft_rname']
10
```

```
ft_urname
Mozilla-GitHub-Standards    10
vladikoff                   7
mortonfox                   6
spasovski                   6
clouserw                    6
..
eaon                         1
edsu                         1
edwardabraham               1
ehsan                       1
zyegfried                   1
Name: ft_rname, Length: 750, dtype: int64
```

```
1 # user_diffrepos_total
2 users_five_repositories_groups = f_data.groupby(['ft_urname']).sort_values(ascending=False,inplace=True)
3 #print(user_diffrepos_total_changes_count)
4 #users_five_repositories_groups.sort_values(ascending=False,inplace=True)
5 #print(users_five_repositories_groups_groups)
```

```

5 #print(users_five_repositories_groups.groups)
6 final_each_user_to_five_repos_names = []
7 for each in users_five_repositories_groups.groups:
8     #if(each.size>=5):
9     #print(each)
10    if(final_each_user_to_five_repos_series[each]>=5):#users_five_repositories_groups.get_group(e
11        print(each)
12        final_each_user_to_five_repos_names.append(each)
13
14

```

Mozilla-GitHub-Standards

TDA

adamlofting

clouserw

dannycoates

dependabot[bot]

flodolo

glogiotatidis

mortonfox

mstriemer

mythmon

oremj

ozten

pbkracker

philbooth

spasovski

vbudhram

vladikoff

willkg

```

1 #got each user with contributions to five different repos in first 10 contributions
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5
6 for each in final_each_user_to_five_repos_names:
7     #print(users_five_repositories_groups.get_group(str(each)))
8     group = users_five_repositories_groups.get_group(str(each))
9     #print(group.drop_duplicates('ft_rname'))
10    group_unique = group.drop_duplicates('ft_rname')
11    print(group_unique)
12    group_total = []
13    group_addns = []
14    group_delns = []
15    idx =[]
16    #for i in range(0,5):
17    graph_count =0
18    for k in group_unique['ft_total_changes']:
19        if graph_count==5:
20            break
21        group_total.append(k)
22        graph_count = graph_count +1
23
24    graph_count =0
25    for k in group_unique['ft_addns']:
26        if graph_count==5:
27            break
28        group_addns.append(k)
29        graph_count = graph_count +1
30

```

```
31 graph_count =0
32 for k in group_unique['ft_delns']:
33     if graph_count==5:
34         break
35     group_delns.append(k)
36     graph_count = graph_count +1
37
38 graph_count =0
39 for k in group_unique['ft_rname']:
40     if graph_count==5:
41         break
42     idx.append(k)
43     graph_count = graph_count +1
44
45 df = pd.DataFrame({
46     "total_changes": group_total,
47     "addition_changes": group_addns,
48     "deletion_changes": group_delns,
49     "repo_names": idx,
50     #'repo_changes':user_repo_lang,
51 })#columns =['total_changes','addition_changes','deletion_changes'])
52 plt.figure()
53 df['total_changes']= df['total_changes'].astype(int)
54 df['addition_changes']= df['addition_changes'].astype(int)
55 df['deletion_changes']= df['deletion_changes'].astype(int)
56
57 #df=df.astype(int)
58 #plt.title(each)
59 #df_list.append(df)# for individual comment this line
60 #lines = df.plot()
61 df.plot(x = "repo_names", y=["total_changes","addition_changes","deletion_changes"],kind="bar")
62
```

```

      Unnamed: 0  ...      repolang
206           534  ...          HTML
207           535  ...    JavaScript
208           536  ...          Python
209           537  ...    JavaScript
210           538  ...          Python
211           539  ...          Python
212           540  ...          PHP
213           541  ...          Python
214           542  ...          CSS
215           543  ...          Python

```

[10 rows x 19 columns]

```

      Unnamed: 0  ...      repolang
318           998  ...          NaN
319           999  ...    JavaScript
320          1000  ...    JavaScript
321          1001  ...    JavaScript
322          1002  ...    JavaScript

```

[5 rows x 19 columns]

```

      Unnamed: 0  ...      repolang
406          1221  ...    JavaScript
410          1225  ...    ApacheConf
411          1226  ...    JavaScript
412          1227  ...    JavaScript
413          1228  ...          CSS

```

[5 rows x 19 columns]

```

      Unnamed: 0  ...      repolang

```

```

1 #now do it for 5 repos having first contributions
2
3 # user_diffrepos_total
4 repositories_five_users = f_data.groupby(['ft_rname']).nunique()
5 #print(user_diffrepos_total_changes_count)
6 repositories_five_users.sort_values(by = 'ft_uname',ascending=False,inplace=True)
7 #for i in users_five_repositories##:
8 #     if i >=5:
9 print(repositories_five_users['ft_uname'])
10
11 final_each_repo_to_five_users_series = repositories_five_users['ft_uname']
12

```

```

ft_rname
django-csp          23
django-browserid    23
infosec.mozilla.org 22
leadership-training 21
moz-corsica         20
..
aggredash           1
all-aboard           1
amocontentguidelines 1
androidspeech        1
django-recaptcha     1
Name: ft_uname, Length: 305, dtype: int64

```

```

1322          5042  ...          CSS

```

```

1 # user_diffrepos_total
2 repositories_five_users_groups = f_data.groupby(['ft_rname'])#.sort_values(ascending=False,inplace=True)
3 #print(user_diffrepos_total_changes_count)
4 #users_five_repositories_groups#.sort_values(ascending=False,inplace=True)
5 #print(users_five_repositories_groups.groups)
6 final_each_repo_to_five_users_names = []

```

```
6 final_each_repo_to_five_users_names = []
7 for each in repositories_five_users_groups.groups:
8     #if(each.size>=5):
9     #print(each)
10    if(final_each_repo_to_five_users_series[each]>=5):#users_five_repositories_groups.get_group(e
11        print(each)
12        final_each_repo_to_five_users_names.append(each)
13
14
```

Addon-Tests
Jisort
MakeAPI
Marketplace.PHP
One-Mozilla-blog
OpenDesign
PyBrowserID
activate.mozilla.community
addons
addons-contribution-notifier
advocacy.mozilla.org
arewefastyet
ateam-bootcamp
authenticator-rs
awsbox
basket-client
bleach
blurts-server
bouncer-tests
browserid-crypto
buddyup
build-relengdocs
build.webmaker.org
cerberus
cipherscan
commbadge
configman
connect-cachify
contain-facebook
contribute.json
corsica
darjeeling
deprecated-treestatus
django-badger
django-browserid
django-csp
django-product-details
django-session-csrf
dscontrib
dxr
eslint-plugin-no-unsanitized
fennec-search
firefox-browser-architecture
funfactory
fx36start
fxa-auth-db-mysql
fxa-auth-db-server
fxa-auth-mailer
fxa-local-dev
fxoss
geckodriver
high-fidelity
html5-lint
http-observatory
identity-pubkeys

```
inclusion
infernyx
infosec-risk-management-bugzilla
infosec.mozilla.org
```

```
1 !pip install pandas-bokeh
2
3 import pandas as pd
4 import pandas_bokeh
5
6 pandas_bokeh.output_notebook()
7
8
```

```
Requirement already satisfied: pandas-bokeh in /usr/local/lib/python3.7/dist-packages (0.5.5)
Requirement already satisfied: bokeh>=2.0 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: pandas>=0.22.0 in /usr/local/lib/python3.7/dist-packages (from pandas-bokeh)
Requirement already satisfied: Jinja2>=2.7 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: numpy>=1.11.3 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: typing-extensions>=3.7.4 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from bokeh)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from bokeh)
```

```
1 !pip install cufflinks
2
3 import pandas as pd
4 import cufflinks as cf
5 from IPython.display import display,HTML
6 #making all charts public and setting a global theme
7 cf.set_config_file(sharing='public',theme='ggplot',offline=False)
```

```
Requirement already satisfied: cufflinks in /usr/local/lib/python3.7/dist-packages (0.17.3)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: colorlover>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: plotly>=4.1.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: ipywidgets>=7.0.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: numpy>=1.9.2 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: setuptools>=34.4.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: ipython>=5.3.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: pandas>=0.19.2 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: retrying>=1.3.3 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: widgetsnbextension~>=3.5.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: nbformat>=4.2.0 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: jupyterlab-widgets>=1.0.0; python_version >= "3.6" in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: ipykernel>=4.5.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: traitlets>=4.3.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: decorator in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: pickleshare in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: pexpect; sys_platform != "win32" in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: simplegeneric>=0.8 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.4 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: pygments in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
```

```
Requirement already satisfied: notebook>=4.4.1 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: ipython-genutils in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: jupyter-core in /usr/local/lib/python3.7/dist-packages (from nb
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.7/dist-packages (from :
Requirement already satisfied: tornado>=4.0 in /usr/local/lib/python3.7/dist-packages (from ipy
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: wcwidth in /usr/local/lib/python3.7/dist-packages (from prompt-t
Requirement already satisfied: nbconvert in /usr/local/lib/python3.7/dist-packages (from notebo
Requirement already satisfied: Send2Trash in /usr/local/lib/python3.7/dist-packages (from notel
Requirement already satisfied: jinja2 in /usr/local/lib/python3.7/dist-packages (from notebook:
Requirement already satisfied: terminado>=0.8.1 in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.7/dist-packages (from jupyter
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: testpath in /usr/local/lib/python3.7/dist-packages (from nbconve
Requirement already satisfied: bleach in /usr/local/lib/python3.7/dist-packages (from nbconvert
Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.7/dist-packages (fi
Requirement already satisfied: defusedxml in /usr/local/lib/python3.7/dist-packages (from nbcor
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: webencodings in /usr/local/lib/python3.7/dist-packages (from blea
Requirement already satisfied: packaging in /usr/local/lib/python3.7/dist-packages (from bleach
Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from
```

```
1 #pandas_bokeh.output_file('eachrepo_5users.html')
2 pandas_bokeh.output_notebook()
```

```
1 #got each user with contributions to five different repos in first 10 contributions
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5
6 #fig = plt.figure(figsize =((8,6)))
7
8
9 for each in final_each_repo_to_five_users_names:
10     #print(users_five_repositories_groups.get_group(str(each)))
11     group = repositories_five_users_groups.get_group(str(each))
12     #print(group.drop_duplicates('ft_rname'))
13     group_unique = group.drop_duplicates('ft_uname')
14     print(group_unique)
15     group_total = []
16     group_addns = []
17     group_delns = []
18     idx =[]
19     #for i in range(0,5):
20     graph_count =0
21     for k in group_unique['ft_total_changes']:
22         if graph_count==5:
23             break
24         group_total.append(k)
25         graph_count = graph_count +1
26
27     graph_count =0
28     for k in group_unique['ft_addns']:
29         if graph_count==5:
30             break
31         group_addns.append(k)
32         graph_count = graph_count +1
```

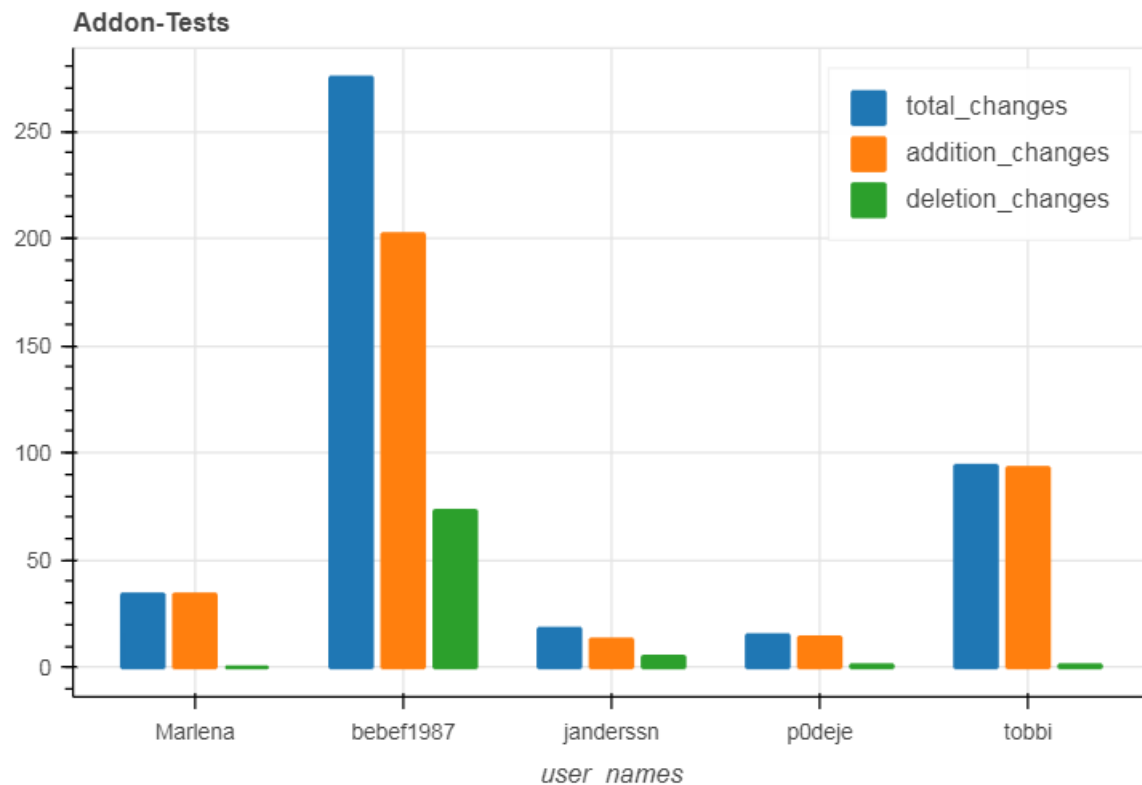
```

33
34     graph_count =0
35     for k in group_unique['ft_delns']:
36         if graph_count==5:
37             break
38         group_delns.append(k)
39         graph_count = graph_count +1
40
41     graph_count =0
42     for k in group_unique['ft_uname']:
43         if graph_count==5:
44             break
45         idx.append(k)
46         graph_count = graph_count +1
47
48     df = pd.DataFrame({
49         "total_changes": group_total,
50         "addition_changes": group_addns,
51         "deletion_changes": group_delns,
52         "user_names": idx,
53         #'repo_changes':user_repo_lang,
54     })#columns =['total_changes','addition_changes','deletion_changes'])
55     plt.figure(figsize =((8,6)))
56     df['total_changes']= df['total_changes'].astype(int)
57     df['addition_changes']= df['addition_changes'].astype(int)
58     df['deletion_changes']= df['deletion_changes'].astype(int)
59
60     #df=df.astype(int)
61     #plt.title(each)
62     #df_list.append(df)# for individual comment this line
63     #lines = df.plot()
64     #df.plot(x = "user_names", y=["total_changes","addition_changes","deletion_changes"],kind="bar")
65     #i=100
66     #df = fig.add_subplot(i)
67     #i=i+1
68     df.plot_bokeh(kind='bar',x = "user_names",title = each)
69     #df.iplot(kind = 'bar')
70     plt.show()
71 #fig.savefig('output_of_5.pdf')

```


	Unnamed: 0	ft_uname	...	reposedesc	repolang
150	393	Marlena	...	Tests for Mozilla's Add-ons website	NaN
642	2654	bebef1987	...	Tests for Mozilla's Add-ons website	NaN
1682	6254	janderssn	...	Tests for Mozilla's Add-ons website	NaN
2540	9918	p0deje	...	Tests for Mozilla's Add-ons website	NaN
3116	11842	tobbi	...	Tests for Mozilla's Add-ons website	NaN

[5 rows x 19 columns]



<Figure size 576x432 with 0 Axes>

	Unnamed: 0	...	repolang
383	1179	...	Java
771	3206	...	Java
1155	4678	...	Java
1625	6193	...	Java

1

plotly Interactive plots

```

1 #done every thing above so commenting these
2 '''
3 !pip install pandas-bokeh
4
5 import pandas as pd
6 import pandas_bokeh
7 '''
8 # pandas_bokeh.output_file(filename): for exporting plots as HTML.
9 # pandas_bokeh.output_notebook(): for embedding plots in Jupyter Notebooks.
10 #dataframe.plot_bokeh()
11 # nifty_data_resample.plot_bokeh(kind='bar',figsize=(10,6));
12 # nifty_data_resample.plot_bokeh(kind='barh',stacked=True);
13
14 ''' pip install cufflinks
15 import pandas as pd
16 import cufflinks as cf
17 from IPython.display import display,HTML
18 #making all charts public and setting a global theme

```

```
18 #making all charts public and setting a global theme
19 cf.set_config_file(sharing='public',theme='white',offline=True)
20
21 nifty_data_resample.iplot(kind='bar');
22 nifty_data_resample.iplot(kind='barh',barmode = 'stack');
23
24 '''
```

```
' pip install cufflinks\nimport pandas as pd\nimport cufflinks as cf\nfrom IPython.display import display,HTML\n#making all charts public and setting a global theme\ncf.set_config_file(sharing='public',theme='white',offline=True)\n\nnifty_data_resample.iplot(kind='bar');\nnifty_data_resample.iplot(kind='barh',barmode = 'stack')\n\n'
```

```
1 !pip install cufflinks
2
3 import pandas as pd
4 import cufflinks as cf
5 from IPython.display import display,HTML
6 #making all charts public and setting a global theme
7 cf.set_config_file(sharing='public',theme='white',offline=True)
```

```
Requirement already satisfied: cufflinks in /usr/local/lib/python3.7/dist-packages (0.17.3)
Requirement already satisfied: colorlover>=0.2.1 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: pandas>=0.19.2 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
Requirement already satisfied: numpy>=1.9.2 in /usr/local/lib/python3.7/dist-packages (from cufflinks)
```

```

1
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from bokeh.plotting import figure, output_file, show
5
6
7 lang_df = pd.DataFrame(lang_count, lang_names)
8
9 p = figure(plot_width=400, plot_height=300)
10
11 p=plt.figure(figsize = (40, 5))
12 '''
13 lang_df=lang_df.astype(int)
14 lang_df.plot.bar()
15 plt.axhline(0, color="k")
16 '''
17 '''
18 plt.bar(lang_names, lang_count, color = 'green', width = 0.3)
19
20 plt.xlabel("Languages")
21 plt.ylabel("No. of repositories in each Language")
22 plt.title("Repo Lang Plot")
23 plt.show()
24 '''
25
26 #p.vbar(lang_count, lang_names)
27 #p.xaxis.major_label_orientation = "vertical"
28 lang_df.plot_bokeh(kind='bar')
29 #show(p)
30
31

```



```
1 lang_df.plot_bokeh(kind='pie')
```

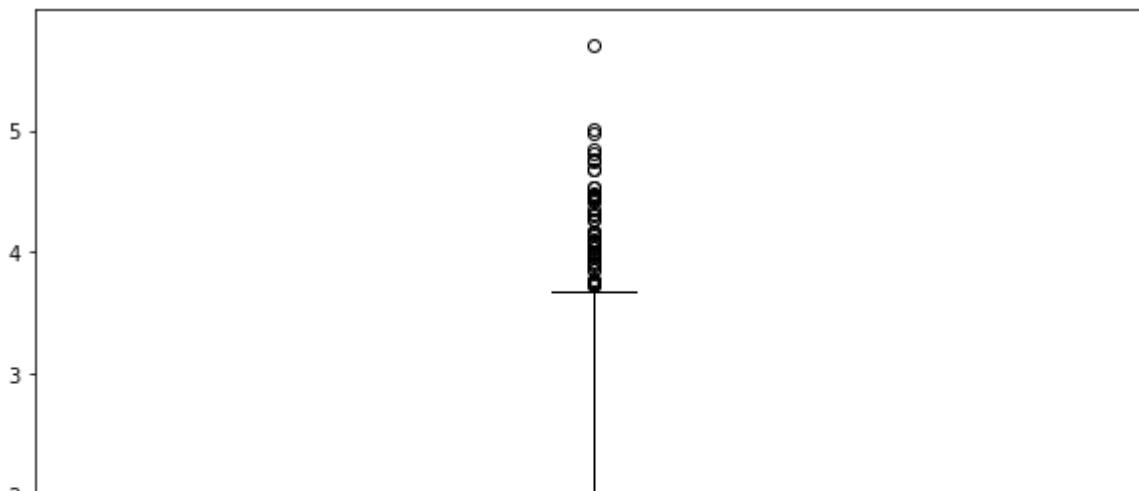
```
__x__values_original
```

```
1 #pandas_bokeh.output_notebook()
2 #lang_df.plot_bokeh(kind = 'bar', figsize=(10,6), axis ='x')
```

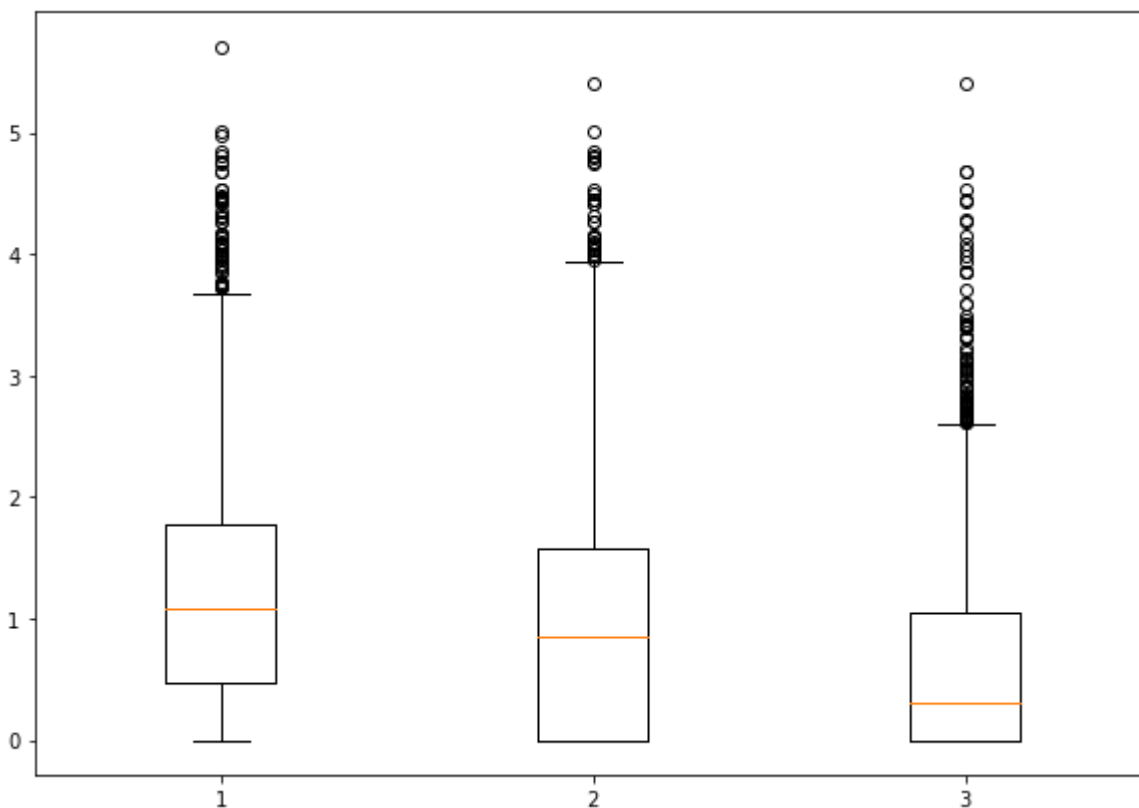
```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 log_total = np.log10(f_data['ft_total_changes'])
5 log_addns = np.log10(f_data['ft_addns'])
6 log_delns = np.log10(f_data['ft_delns'])
7
8
```

```
/usr/local/lib/python3.7/dist-packages/pandas/core/series.py:726: RuntimeWarning:
divide by zero encountered in log10
```

```
1 fig = plt.figure(figsize =(10, 7))
2 plt.boxplot(log_total)
3 plt.show()
```



```
1 data = [log_total, log_addns, log_delns]
2 fig = plt.figure(figsize =(10, 7))
3 bp = plt.boxplot(data)
4 plt.show()
```

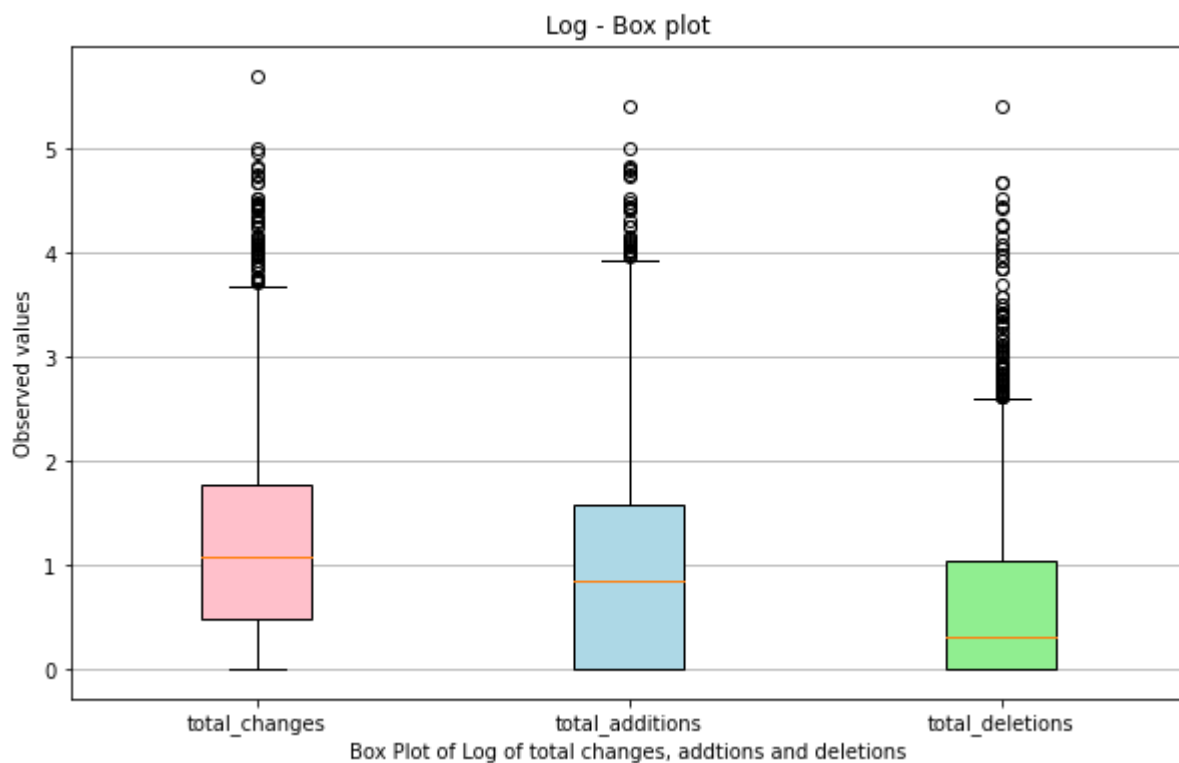


```
1 import matplotlib.pyplot as plt
2 import numpy as np
3
4 labels = ['total_changes', 'total_additions', 'total_deletions']
5 fig, ax = plt.subplots(nrows=1, ncols=1, figsize=(10, 6))
6
7 # rectangular box plot
8 bplot1 = ax.boxplot(data,
9                     vert=True, # vertical box alignment
10                    patch_artist=True, # fill with color
11                    labels=labels) # will be used to label x-ticks
12 ax.set_title('Log - Box plot')
13
14 # fill with colors
15 colors = ['pink', 'lightblue', 'lightgreen']
16 for patch, color in zip(bplot1['boxes'], colors):
17     patch.set_facecolor(color)
```

```

18
19 # adding horizontal grid lines
20 ax.yaxis.grid(True)
21 ax.set_xlabel('Box Plot of Log of total changes, addtions and deletions')
22 ax.set_ylabel('Observed values')
23
24 plt.show()
25

```



```

1 dataf = pd.DataFrame(log_total, log_addns, log_delns)
2 log_total.plot_bokeh(kind='point')

```



