

# tf-idf

October 31, 2021

## 1 CA 675 Anushka Walsepatil

## 2 Installing all the required packages

```
[101]: pip install sasl
```

Requirement already satisfied: sasl in /opt/conda/anaconda/lib/python3.7/site-packages (0.3.1)

Requirement already satisfied: six in /opt/conda/anaconda/lib/python3.7/site-packages (from sasl) (1.15.0)

Note: you may need to restart the kernel to use updated packages.

```
[102]: pip install thrift
```

Requirement already satisfied: thrift in /opt/conda/anaconda/lib/python3.7/site-packages (0.15.0)

Requirement already satisfied: six>=1.7.2 in /opt/conda/anaconda/lib/python3.7/site-packages (from thrift) (1.15.0)

Note: you may need to restart the kernel to use updated packages.

```
[103]: pip install thrift-sasl
```

Requirement already satisfied: thrift-sasl in /opt/conda/anaconda/lib/python3.7/site-packages (0.4.3)

Requirement already satisfied: six>=1.13.0 in /opt/conda/anaconda/lib/python3.7/site-packages (from thrift-sasl) (1.15.0)

Requirement already satisfied: thrift>=0.10.0; python\_version >= "3.0" in /opt/conda/anaconda/lib/python3.7/site-packages (from thrift-sasl) (0.15.0)

Requirement already satisfied: pure-sasl>=0.6.2 in /opt/conda/anaconda/lib/python3.7/site-packages (from thrift-sasl) (0.6.2)

Note: you may need to restart the kernel to use updated packages.

```
[104]: pip install PyHive
```

Requirement already satisfied: PyHive in /opt/conda/anaconda/lib/python3.7/site-packages (0.6.4)

Requirement already satisfied: future in /opt/conda/anaconda/lib/python3.7/site-

packages (from PyHive) (0.17.1)  
Requirement already satisfied: python-dateutil in  
/opt/conda/anaconda/lib/python3.7/site-packages (from PyHive) (2.8.0)  
Requirement already satisfied: six>=1.5 in  
/opt/conda/anaconda/lib/python3.7/site-packages (from python-dateutil->PyHive)  
(1.15.0)  
Note: you may need to restart the kernel to use updated packages.

### 3 Hive connection

```
[105]: from pyhive import hive
from sklearn.feature_extraction.text import TfidfVectorizer as tfv
import pandas as pd

host= "127.0.0.1"
port = 10000
username="cluster-1-m"
password="3378257863336058986"
anushka_hive = hive.Connection(host, port, username, password)
```

### 4 Top 10 users to calculate the TF-ID

```
[106]: q1 ="select OwnerUserId, Sum(score) as Score from ca675_data1 group by
↳OwnerUserId order by Score desc limit 10"
data = pd.read_sql(q1, anushka_hive)
display(data)
```

	owneruserid	score
0		81009387
1	87234	37672
2	4883	28817
3	9951	26878
4	6068	25944
5	89904	24024
6	51816	23763
7	49153	20203
8	179736	19603
9	95592	19479

```
[107]: q2="SELECT OWNERUSERID,TITLE,BODY FROM ca675_data1 WHERE OWNERUSERID IN (87234,
↳4883, 9951, 6068, 89904, 51816, 49153, 179736, 95592, 63051)"
data = pd.read_sql(q2, anushka_hive)
data['Body-Title'] = data['title'] + data['body']
```

```
display(data)
```

```
      owneruserid                                title \
0          89904  How do I undo the most recent local commits in...
1          95592  How do I delete a Git branch locally and remot...
2           4883  How can you find out which process is listenin...
3          49153  How to replace all occurrences of a string in ...
4           6068           How do you get a timestamp in JavaScript?
..          ...
420         51816           Clamping floating numbers in Python?
421          9951  Is there any Python-like interactive console f...
422         51816           Checking delegates for null
423         63051  What are the advantages and disadvantages of j...
424        179736  How can I remove all nil elements in a Swift a...
```

```
                                body \
0  I accidentally committed the wrong files to Gi...
1  I want to delete a branch both locally and rem...
2  How can you find out which process is listenin...
3  I have this string in my JavaScript code quotT...
4  Something similar to Unixs timestamp that is a...
..          ...
420  Is there a builtin function for this in Python...
421  I spent a lot of time programming in Java rece...
422  I was reading the Essential C    book and am w...
423  What are the advantages and disadvantages of j...
424  Basic way doesnt work  for index in  lt listc...
```

```
                                Body-Title
0  How do I undo the most recent local commits in...
1  How do I delete a Git branch locally and remot...
2  How can you find out which process is listenin...
3  How to replace all occurrences of a string in ...
4  How do you get a timestamp in JavaScript?Somet...
..          ...
420  Clamping floating numbers in Python?Is there a...
421  Is there any Python-like interactive console f...
422  Checking delegates for nullI was reading the E...
423  What are the advantages and disadvantages of j...
424  How can I remove all nil elements in a Swift a...
```

```
[425 rows x 4 columns]
```

```
[108]: vectorizer = tfv(stop_words='english', lowercase=True)

owners = list(data["owneruserid"].unique())
```

```

for owner in owners:

    owner_awp_df=data[data["owneruserid"] == owner]
    nw_col = vectorizer.fit_transform(owner_awp_df["Body-Title"])
    final_op_aw= pd.DataFrame(nw_col.toarray(),columns=vectorizer.
    ↳get_feature_names())

    final_max_10 = final_op_aw.sum(axis=0)
    top_10= final_max_10.nlargest(10)
    top_10_aw = list(top_10.index)

    final_tfidf_10=final_op_aw[top_10_aw]

    print("-----")
    display(" Top 10 terms of the owner id : "+owner)
    print("-----")

    display(final_tfidf_10)

```

-----

' Top 10 terms of the owner id : 89904'

-----

	file	new	java	timer	setpreferredwidth	\
0	0.000000	0.000000	0.000000	0.000000	0.000000	
1	0.000000	0.000000	0.040077	0.000000	0.606105	
2	0.356757	0.000000	0.000000	0.000000	0.000000	
3	0.000000	0.000000	0.000000	0.000000	0.000000	
4	0.470020	0.000000	0.000000	0.000000	0.000000	
5	0.000000	0.000000	0.313903	0.000000	0.000000	
6	0.000000	0.264469	0.102856	0.622212	0.000000	
7	0.077083	0.440536	0.000000	0.000000	0.000000	
8	0.000000	0.000000	0.000000	0.000000	0.000000	
9	0.000000	0.000000	0.174024	0.000000	0.000000	

  

	tablegetcolumnmodelgetcolumn	bits	byte	iterate	application
0		0.000000	0.000000	0.000000	0.000000
1		0.606105	0.000000	0.000000	0.000000
2		0.000000	0.000000	0.000000	0.000000
3		0.000000	0.000000	0.000000	0.000000
4		0.000000	0.000000	0.000000	0.268619
5		0.000000	0.000000	0.000000	0.000000
6		0.000000	0.000000	0.000000	0.000000
7		0.000000	0.000000	0.000000	0.000000
8		0.000000	0.000000	0.000000	0.254880

9	0.000000	0.526366	0.526366	0.526366	0.000000
---	----------	----------	----------	----------	----------

-----

' Top 10 terms of the owner id : 95592'

	install	python	branch	pip	flask	version	started \
0	0.000000	0.000000	0.697882	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.482939	0.241470	0.000000	0.301837	0.000000	0.000000	0.000000
3	0.262056	0.104822	0.000000	0.104822	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.555725	0.555725	0.000000
5	0.427251	0.000000	0.000000	0.284834	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.382895	0.000000	0.000000	0.000000	0.000000	0.529451

	way	navigation	django
0	0.000000	0.000000	0.000000
1	0.015182	0.48611	0.000000
2	0.062569	0.000000	0.000000
3	0.000000	0.000000	0.060737
4	0.138851	0.000000	0.000000
5	0.098407	0.000000	0.000000
6	0.062210	0.000000	0.417339
7	0.132286	0.000000	0.000000

-----

' Top 10 terms of the owner id : 4883'

	python	use	list	table	way	difference	branch \
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.636117
5	0.000000	0.170564	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.191672	0.000000	0.000000	0.000000	0.491935	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.276282	0.000000	0.000000	0.000000	0.152003	0.000000	0.000000
10	0.203719	0.112081	0.000000	0.000000	0.000000	0.287660	0.000000
11	0.000000	0.000000	0.144907	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

13	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.108905	0.239666	0.428552	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.120015	0.000000	0.308023	0.120015	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.231741	0.000000	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.091125	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.156136	0.000000	0.000000	0.000000	0.000000	0.220471	0.000000
21	0.229290	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000	0.181739	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000	0.123956	0.000000	0.347285
25	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
27	0.075810	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	0.000000	0.198917	0.000000	0.255265	0.099459	0.000000	0.000000
29	0.000000	0.000000	0.000000	0.521156	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.202044	0.000000	0.169489	0.000000	0.000000
31	0.164277	0.000000	0.430964	0.000000	0.180761	0.000000	0.000000
32	0.127064	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	method	process	rename
0	0.000000	0.350231	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000
8	0.209051	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.336243
16	0.000000	0.187004	0.000000
17	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000
21	0.401012	0.000000	0.000000
22	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000

```

26 0.000000 0.000000 0.000000
27 0.099439 0.000000 0.000000
28 0.000000 0.382898 0.000000
29 0.000000 0.000000 0.568902
30 0.000000 0.000000 0.000000
31 0.215482 0.000000 0.000000
32 0.000000 0.000000 0.000000

```

---

' Top 10 terms of the owner id : 49153'

---

	javascript	array	using	php	jquery	file	want	gt	\
0	0.060398	0.000000	0.000000	0.00000	0.000000	0.0	0.00000	0.0	
1	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	0.00000	0.0	
2	0.000000	0.000000	0.235622	0.00000	0.360941	0.0	0.00000	0.0	
3	0.150462	0.393168	0.000000	0.00000	0.000000	0.0	0.00000	0.0	
4	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	0.00000	0.0	
..	...	...	...	...	...	...	...	...	
74	0.000000	0.613169	0.000000	0.12268	0.000000	0.0	0.00000	0.0	
75	0.000000	0.000000	0.034143	0.00000	0.000000	0.0	0.00000	0.0	
76	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	0.00000	0.0	
77	0.000000	0.000000	0.000000	0.00000	0.000000	0.0	0.00000	0.0	
78	0.000000	0.000000	0.076544	0.00000	0.000000	0.0	0.09578	0.0	

	string	class
0	0.193969	0.000000
1	0.000000	0.000000
2	0.000000	0.000000
3	0.000000	0.000000
4	0.111014	0.000000
..	...	...
74	0.000000	0.000000
75	0.137205	0.304663
76	0.000000	0.050179
77	0.000000	0.000000
78	0.205064	0.000000

[79 rows x 10 columns]

---

' Top 10 terms of the owner id : 6068'

---

	file	java	android	git	sql	way	like \
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.728914	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.130150	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.173267	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.140540	0.000000
5	0.000000	0.386413	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.273927	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.217416	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.047141
9	0.000000	0.000000	0.148040	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.334555	0.000000	0.000000	0.124201	0.000000
11	0.000000	0.000000	0.000000	0.000000	0.000000	0.076296	0.158251
12	0.000000	0.000000	0.000000	0.000000	0.105437	0.000000	0.072167
13	0.656317	0.000000	0.000000	0.417075	0.000000	0.085457	0.000000
14	0.138946	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.371197	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.352152	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.094212	0.293117
19	0.000000	0.364640	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.306959	0.000000	0.000000	0.000000	0.000000	0.119905	0.124351
22	0.083346	0.370875	0.000000	0.000000	0.000000	0.000000	0.067528
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.090766	0.000000
24	0.000000	0.000000	0.000000	0.000000	0.000000	0.029544	0.061279
25	0.000000	0.000000	0.000000	0.142581	0.000000	0.087643	0.090893
26	0.000000	0.154458	0.000000	0.000000	0.000000	0.108471	0.000000
27	0.000000	0.000000	0.000000	0.000000	0.000000	0.083163	0.000000
28	0.000000	0.390291	0.000000	0.000000	0.000000	0.000000	0.071063
29	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.106695
30	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000	0.000000	0.082497	0.000000
32	0.116987	0.000000	0.000000	0.000000	0.692403	0.000000	0.094784
33	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	0.166516	0.000000	0.000000	0.000000	0.262812	0.130089	0.000000
36	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.037880

	time	code	tables
0	0.198534	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.389852	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.233437



8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.353714	0.000000
11	0.000000	0.325928	0.000000
12	0.000000	0.000000	0.452826
13	0.000000	0.000000	0.000000
14	0.000000	0.022082	0.000000
15	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.277321
21	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000
23	0.137528	0.000000	0.000000
24	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000
26	0.328708	0.000000	0.000000
27	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000
33	0.000000	0.215485	0.000000
34	0.000000	0.000000	0.000000
35	0.000000	0.061747	0.000000
36	0.000000	0.000000	0.000000

-----  
' Top 10 terms of the owner id : 51816'

	python	like	list	string	want	class	value \
0	0.159158	0.071328	0.000000	0.436149	0.000000	0.000000	0.000000
1	0.075039	0.067258	0.31736	0.000000	0.185491	0.000000	0.000000
2	0.000000	0.055537	0.000000	0.000000	0.076582	0.000000	0.000000
3	0.098312	0.088118	0.000000	0.538819	0.000000	0.000000	0.000000
4	0.229219	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
..	...	...	...	...	...	...	...
61	0.133283	0.000000	0.000000	0.000000	0.000000	0.000000	0.199787
62	0.000000	0.046250	0.000000	0.000000	0.063775	0.000000	0.000000
63	0.000000	0.000000	0.000000	0.000000	0.102406	0.620990	0.000000
64	0.295204	0.132297	0.000000	0.000000	0.000000	0.000000	0.000000
65	0.000000	0.000000	0.000000	0.000000	0.000000	0.068095	0.204286

	index	use	function
0	0.000000	0.000000	0.000000
1	0.125243	0.000000	0.000000
2	0.310247	0.000000	0.000000
3	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000
..	...	...	...
61	0.000000	0.000000	0.000000
62	0.000000	0.072743	0.000000
63	0.000000	0.000000	0.000000
64	0.000000	0.000000	0.237068
65	0.000000	0.000000	0.000000

[66 rows x 10 columns]

-----  
' Top 10 terms of the owner id : 9951'

	python	like	git	head	does	using	way \
0	0.000000	0.066630	0.078782	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.101237	0.000000	0.000000	0.119702	0.000000	0.000000
2	0.525068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.435896	0.000000	0.000000	0.000000	0.000000	0.000000	0.273815
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.083623
5	0.000000	0.000000	0.000000	0.000000	0.417524	0.000000	0.157365
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.119729	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.227740	0.000000	0.000000
8	0.000000	0.000000	0.376139	0.411180	0.075228	0.000000	0.000000
9	0.155490	0.328761	0.000000	0.000000	0.000000	0.000000	0.146510
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.155518	0.000000	0.000000	0.000000	0.000000	0.173264
12	0.000000	0.000000	0.075446	0.000000	0.000000	0.150893	0.000000
13	0.000000	0.188069	0.000000	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000	0.000000	0.236308	0.000000
15	0.000000	0.000000	0.179128	0.734307	0.000000	0.000000	0.000000
16	0.081895	0.138523	0.000000	0.000000	0.163789	0.000000	0.000000
17	0.000000	0.033760	0.000000	0.000000	0.000000	0.159673	0.112839
18	0.000000	0.000000	0.338695	0.000000	0.000000	0.000000	0.000000
19	0.109076	0.092250	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.187925	0.000000	0.062642	0.000000	0.059024
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000	0.000000	0.299606	0.000000
23	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
24	0.079724	0.202277	0.000000	0.000000	0.000000	0.079724	0.000000

	use	know	id
0	0.000000	0.070219	0.000000
1	0.000000	0.000000	0.000000
2	0.262534	0.000000	0.000000
3	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000
5	0.000000	0.223286	0.000000
6	0.119729	0.000000	0.000000
7	0.113870	0.000000	0.000000
8	0.150456	0.000000	0.000000
9	0.000000	0.069295	0.178117
10	0.000000	0.000000	0.000000
11	0.000000	0.081948	0.000000
12	0.000000	0.134492	0.000000
13	0.222372	0.000000	0.000000
14	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000
16	0.000000	0.072993	0.000000
17	0.119755	0.142318	0.045727
18	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.124948
20	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.474805
24	0.000000	0.142117	0.000000

-----  
' Top 10 terms of the owner id : 179736'

	want	python	user	string	dictionary	return	file \
0	0.000000	0.160778	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.186977	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.207866
3	0.000000	0.000000	0.000000	0.172983	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
..	...	...	...	...	...	...	...
110	0.000000	0.000000	0.000000	0.000000	0.000000	0.261357	0.000000
111	0.000000	0.069213	0.000000	0.000000	0.088713	0.000000	0.000000
112	0.042307	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
113	0.000000	0.000000	0.083829	0.000000	0.000000	0.083829	0.000000
114	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

use way just

0	0.000000	0.000000	0.0
1	0.000000	0.000000	0.0
2	0.083676	0.000000	0.0
3	0.000000	0.172983	0.0
4	0.000000	0.000000	0.0
..	...	...	...
110	0.000000	0.121287	0.0
111	0.000000	0.000000	0.0
112	0.000000	0.000000	0.0
113	0.000000	0.000000	0.0
114	0.000000	0.141434	0.0

[115 rows x 10 columns]

-----  
' Top 10 terms of the owner id : 63051'

	file	python	vs	want	use	java	standard \
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.112601	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.116685	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.245377	0.000000	0.000000	0.000000	0.173695	0.000000	0.550574
5	0.171867	0.000000	0.000000	0.160447	0.000000	0.000000	0.482042
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.588666	0.000000
7	0.110772	0.124803	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.580077	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.122889	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.130342	0.000000	0.000000	0.121681	0.000000	0.000000	0.000000
12	0.000000	0.160209	0.167762	0.000000	0.167762	0.000000	0.000000
13	0.078173	0.000000	0.000000	0.291914	0.000000	0.000000	0.000000
14	0.000000	0.159918	0.000000	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.438875	0.000000	0.219437	0.521677	0.000000
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.086165	0.054446	0.000000	0.000000
22	0.000000	0.000000	0.227035	0.000000	0.000000	0.000000	0.000000
23	0.071658	0.000000	0.000000	0.066896	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.172612	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

27	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	0.318679	0.000000	0.000000	0.099168	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.343812	0.000000	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
33	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.000000	0.069348	0.000000	0.000000	0.000000
36	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
37	0.192038	0.000000	0.000000	0.089639	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
39	0.000000	0.000000	0.000000	0.000000	0.000000	0.163161	0.000000
40	0.285320	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
41	0.000000	0.000000	0.246946	0.000000	0.493891	0.000000	0.000000
42	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.217125
43	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
44	0.319440	0.000000	0.000000	0.149107	0.188434	0.000000	0.000000
45	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
46	0.000000	0.321788	0.000000	0.000000	0.000000	0.000000	0.000000
47	0.000000	0.000000	0.229219	0.000000	0.000000	0.000000	0.000000
48	0.000000	0.000000	0.251779	0.000000	0.000000	0.000000	0.000000

	output	instance	kill
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000
4	0.322330	0.000000	0.000000
5	0.225767	0.000000	0.000000
6	0.000000	0.000000	0.636699
7	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000
19	0.000000	0.864911	0.000000
20	0.000000	0.000000	0.000000
21	0.363732	0.000000	0.000000
22	0.000000	0.000000	0.000000
23	0.282391	0.000000	0.000000

24	0.000000	0.000000	0.540266
25	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000
33	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.000000
36	0.000000	0.000000	0.000000
37	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000
39	0.000000	0.000000	0.000000
40	0.000000	0.000000	0.000000
41	0.000000	0.000000	0.000000
42	0.000000	0.000000	0.000000
43	0.000000	0.000000	0.000000
44	0.000000	0.000000	0.000000
45	0.000000	0.317777	0.000000
46	0.000000	0.000000	0.000000
47	0.000000	0.000000	0.000000
48	0.000000	0.000000	0.000000

-----  
 ' Top 10 terms of the owner id : 87234'

	idiom	int	copy	array	arraysize	data	code \
0	0.000000	0.203351	0.000000	0.261452	0.30558	0.30558	0.145251
1	0.000000	0.267377	0.000000	0.000000	0.00000	0.00000	0.133688
2	0.480997	0.000000	0.360748	0.091453	0.00000	0.00000	0.000000

  

	does	mentioned	places
0	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000
2	0.240499	0.240499	0.240499

[ ]: