

ca675_assignment_1

October 25, 2021

CA675 Cloud Assignment

Name - Sarthak Bhagwat Taru StudentID - 21261303 Email - sarthak.taru2@mail.dcu.ie

Task 1 - Acquire the top 2,00,000 posts by ViewCount

Query executed on stackexchange data platform to retrieve the top 2,00,000 posts by viewcount
<https://data.stackexchange.com/stackoverflow/query/new>

SQL Query- `SELECT * FROM (SELECT ps.Id, ps.PostTypeId, ps.ViewCount, ps.Title, ps.AcceptedAnswerId, ps.ParentId, ps.CreationDate, ps.DeletionDate, ps.Score, us.DisplayName, us.EmailHash, ps.LastEditorDisplayName, ps.Tags, ps.ContentLicense, ps.OwnerUserId, ps.FavoriteCount, ps.AnswerCount, ps.CommentCount, ROW_NUMBER() OVER (ORDER BY ViewCount DESC) as row_num FROM Posts ps inner join Users as us on us.Id=ps.OwnerUserId)tab Where row_num BETWEEN 150001 AND 200000`

Above sql returns the 50000 records per execution hence fired it for 4 times to extract 2,00,000 post. Then the data is extracted into 4 csv files.

Task 2 & 3 - Use Pig/Hive/MapReduce - Extract, Transform and Load the data as applicable to get: This assignment is completed on Google Cloud Platform by creating the cluster with Dataproc service. Cluster was configured with Jupyter notebook execute python code. I have used the Hive because of the prior hands-on experience in sql query language. Created the ca675 Database and Table creation is as follows - sql query-CREATE TABLE IF NOT EXISTS ca675.Stackexchange (Id int, PostTypeId int, ViewCount int, Title varchar(255), AcceptedAnswerId int, ParentId int, CreationDate timestamp, DeletionDate timestamp, Score int, OwnerDisplayName varchar(255), EmailHash varchar(255), LastEditorDisplayName varchar(255), Tags varchar(255), ContentLicense varchar(255), OwnerUserId int, FavoriteCount int, AnswerCount int, CommentCount int, row_num int) ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde';

Data loading into hive - Uploaded the csv file in dataproc cloud storage and loaded the data through below sql query LOAD DATA INPATH 'gs://dataproc-staging-us-central1-1078341724532-t5a0gorr/combined.csv' INTO TABLE ca675.Stackexchange;

Below code is written in the Jupyter notebook accessed by public address created by cluster.

```
[1]: from pyhive import hive
     from tabulate import tabulate
     import pandas as pd
```

```
[4]: host_name = "localhost"
     port = 10000
```

```

user = "sarthak_taru2"
password = "1988377416667072609"
database="ca675"
def hiveconnection(host_name, port, user,password, database):
    conn = hive.Connection(host=host_name,
    ↪port=port,username=user,password=password,database=database,auth='CUSTOM')
    return conn
conn = hiveconnection(host_name, port, user,password, database)
cur = conn.cursor()

```

Task 2.1 - The top 10 posts by score

```

[6]: cur.execute('''select Id,OwnerUserId,OwnerDisplayName,score,viewcount
FROM stackexchanges
order by score desc LIMIT 10''')
result = cur.fetchall()
df = pd.DataFrame (result, columns =
    ↪['Id','OwnerUserId','OwnerDisplayName','Score','Viewcount'])
display(df)

```

	Id	OwnerUserId	OwnerDisplayName	Score	Viewcount
0	11227809	87234	GManNickG	25933	1649855
1	927358	89904	Hamza Yerlikaya	23348	10062790
2	2003505	95592	Matthew Rankin	18514	9285139
3	292357	6068	pupeno	12834	3041604
4	231767	18300	Alex. S.	11551	2681330
5	477816	12870	Oli	10921	3269028
6	348170	14069	paxos1977	10079	3985243
7	5767325	364969	Walker	9931	8937271
8	6591213	338204	Forrest	9792	3729583
9	1642028	87234	GManNickG	9560	877861

Task 2.2 - The top 10 users by post score

```

[10]: cur.execute('''
select OwnerUserId,max(OwnerDisplayName) , sum(score) as score
from stackexchanges
group by OwnerUserId
order by score desc limit 10
''')
result = cur.fetchall()
df1 = pd.DataFrame (result, columns =
    ↪['OwnerUserId','OwnerDisplayName','score'])
display(df1)

```

	OwnerUserId	OwnerDisplayName	score
0	87234	GManNickG	37672

1	4883	readonly	28817
2	9951	e-satis	26878
3	6068	pupeno	25944
4	89904	Hamza Yerlikaya	24024
5	51816	Joan Venge	23763
6	49153	Ali	20203
7	179736	TIMEX	19603
8	95592	Matthew Rankin	19479
9	63051	flybywire	19362

Task 3 - The number of distinct users, who used the word “cloud” in one of their posts Here I have considered the Title and Tags to find out the total count of owner user id who used the word cloud in one of their posts.

```
[11]: cur.execute('''
SELECT
COUNT(DISTINCT OwnerUserId) as owner_user_count
FROM stackexchanges
WHERE Title LIKE '%cloud%' or tags LIKE '%cloud%'
''')
result = cur.fetchall()
df2 = pd.DataFrame (result, columns = ['owner_user_count'])
display(df2)
```

	owner_user_count
0	373

Task 4 - calculate the per-user TF-IDF of the top 10 terms for each of the top 10 users So first of all, I have extracted the top 10 owner user id with their titles and created the list of top 10 users.

```
[18]: df3 = pd.read_sql("""
SELECT OwnerUserId,OwnerDisplayName,Title
from stackexchanges
WHERE OwnerUserId
IN
(
select OwnerUserId from(select OwnerUserId,max(OwnerDisplayName) ,
→sum(score) as score
from stackexchanges
group by OwnerUserId
order by score desc limit 10)stack
)
order by OwnerUserId""", conn)

result = cur.fetchall()
top_10_users = list(df3["owneruserid"].unique())
display(top_10_users)
```

[4883, 6068, 9951, 49153, 51816, 63051, 87234, 89904, 95592, 179736]

I have taken the reference from below site to implement tf-idf https://scikitlearn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html for each of the owner and their top 10 words plot the table accordingly.

```
[26]: from sklearn.feature_extraction.text import TfidfVectorizer

def calculate_tfidf(title_each_user):
    vectorizer = TfidfVectorizer(stop_words='english', lowercase=True)
    response = vectorizer.fit_transform(title_each_user['title'])
    df_tfidf_words = pd.DataFrame(response.toarray(), columns=vectorizer.
    ↪get_feature_names()) #calculating tf-idf values
    df_final_result_per_user=df_tfidf_words.sum(axis=0, numeric_only= True)
    ↪#suming up the tf-idf to get the top 10 words
    top_words = df_final_result_per_user.nlargest(n=10)
    top_10_words = list(top_words.index)
    return(df_tfidf_words[top_10_words])

for each_item in top_10_users: # Iterating through top 10 users
    owneruserid = str(each_item)
    selectTitle=df3.loc[df3['owneruserid'] == each_item]
    selectTitle.insert(0, 'Owneruserid', each_item)
    tfidf=calculate_tfidf(selectTitle)
    tfidf.insert(0, 'owneruserid', owneruserid)
    display(tfidf)
```

	owneruserid	python	git	ruby	table	process	list \
0	4883	0.288276	0.000000	0.000000	0.000000	0.000000	0.000000
1	4883	0.000000	0.280138	0.000000	0.000000	0.000000	0.000000
2	4883	0.362262	0.000000	0.000000	0.000000	0.000000	0.000000
3	4883	0.429802	0.000000	0.000000	0.000000	0.000000	0.000000
4	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5	4883	0.000000	0.473839	0.000000	0.000000	0.000000	0.000000
6	4883	0.000000	0.000000	0.000000	0.521156	0.000000	0.000000
7	4883	0.000000	0.000000	0.000000	0.000000	0.350231	0.000000
8	4883	0.000000	0.000000	0.000000	0.332494	0.000000	0.000000
9	4883	0.000000	0.000000	0.439712	0.000000	0.000000	0.000000
10	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.416967
12	4883	0.249965	0.000000	0.000000	0.000000	0.000000	0.000000
13	4883	0.000000	0.000000	0.427544	0.000000	0.000000	0.000000
14	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
15	4883	0.000000	0.000000	0.521156	0.000000	0.000000	0.568902
16	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
17	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

18	4883	0.234332	0.000000	0.000000	0.000000	0.000000	0.000000
19	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	4883	0.288276	0.000000	0.000000	0.000000	0.000000	0.000000
22	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	4883	0.000000	0.000000	0.000000	0.330545	0.330545	0.000000
24	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
26	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
27	4883	0.000000	0.301298	0.000000	0.000000	0.000000	0.000000
28	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
29	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	4883	0.000000	0.000000	0.000000	0.000000	0.350231	0.000000
31	4883	0.000000	0.402233	0.000000	0.000000	0.000000	0.000000
32	4883	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	rename	rails	difference	write
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	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	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.568902	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000
8	0.362955	0.362955	0.000000	0.362955
9	0.000000	0.000000	0.000000	0.479997
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.311276	0.000000
19	0.000000	0.000000	0.534435	0.000000
20	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000	0.000000
29	0.000000	0.534435	0.000000	0.000000
30	0.000000	0.000000	0.000000	0.000000

31	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000

	owneruserid	file	sql	android	java	asp	mvc	\
0	6068	0.000000	0.000000	0.380359	0.000000	0.000000	0.000000	
1	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
2	6068	0.000000	0.000000	0.000000	0.000000	0.395486	0.395486	
3	6068	0.000000	0.000000	0.000000	0.000000	0.358995	0.358995	
4	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
5	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
6	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
7	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
8	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
9	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
10	6068	0.000000	0.000000	0.000000	0.477085	0.000000	0.000000	
11	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
12	6068	0.000000	0.000000	0.366222	0.000000	0.000000	0.000000	
13	6068	0.000000	0.655359	0.000000	0.000000	0.000000	0.000000	
14	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
15	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
16	6068	0.405213	0.000000	0.000000	0.000000	0.000000	0.000000	
17	6068	0.000000	0.000000	0.000000	0.000000	0.319651	0.319651	
18	6068	0.000000	0.000000	0.000000	0.000000	0.395486	0.395486	
19	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
20	6068	0.000000	0.000000	0.424618	0.000000	0.000000	0.000000	
21	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
22	6068	0.592457	0.000000	0.000000	0.000000	0.000000	0.000000	
23	6068	0.000000	0.351970	0.000000	0.000000	0.000000	0.000000	
24	6068	0.000000	0.000000	0.000000	0.366222	0.000000	0.000000	
25	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
26	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
27	6068	0.000000	0.636049	0.000000	0.000000	0.000000	0.000000	
28	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
29	6068	0.380359	0.000000	0.000000	0.000000	0.000000	0.000000	
30	6068	0.000000	0.000000	0.000000	0.371569	0.000000	0.000000	
31	6068	0.405213	0.000000	0.000000	0.000000	0.000000	0.000000	
32	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
33	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
34	6068	0.000000	0.000000	0.416691	0.000000	0.000000	0.000000	
35	6068	0.000000	0.000000	0.000000	0.318516	0.000000	0.000000	
36	6068	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

	net	git	data	difference
0	0.000000	0.000000	0.408388	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.395486	0.000000	0.000000	0.000000
3	0.358995	0.000000	0.000000	0.000000

4	0.000000	0.000000	0.000000	0.000000
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	0.441577	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.000000	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.709610	0.000000	0.354805
16	0.000000	0.000000	0.000000	0.000000
17	0.319651	0.000000	0.000000	0.000000
18	0.395486	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.503558
20	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.318057	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000
25	0.000000	0.000000	0.000000	0.000000
26	0.000000	0.000000	0.000000	0.000000
27	0.000000	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.408388	0.000000
30	0.000000	0.000000	0.000000	0.398950
31	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000
33	0.000000	0.356070	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.000000	0.000000
36	0.000000	0.000000	0.000000	0.000000

	owneruserid	git	python	javascript	string	file	callable	\
0	9951	0.000000	0.000000	0.373615	0.000000	0.000000	0.0	
1	9951	0.000000	0.627334	0.000000	0.000000	0.000000	0.0	
2	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
3	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
4	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
5	9951	0.000000	0.000000	0.000000	0.000000	0.000000	1.0	
6	9951	0.000000	0.448934	0.000000	0.000000	0.000000	0.0	
7	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
8	9951	0.000000	0.000000	0.000000	0.000000	0.531024	0.0	
9	9951	0.000000	0.382546	0.000000	0.000000	0.000000	0.0	
10	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	
11	9951	0.000000	0.000000	0.000000	0.663266	0.000000	0.0	

12	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
13	9951	0.214611	0.000000	0.000000	0.000000	0.586510	0.0
14	9951	0.308479	0.000000	0.000000	0.000000	0.000000	0.0
15	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
16	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
17	9951	0.000000	0.000000	0.000000	0.469000	0.000000	0.0
18	9951	0.260351	0.000000	0.000000	0.000000	0.000000	0.0
19	9951	0.000000	0.000000	0.373615	0.000000	0.000000	0.0
20	9951	0.000000	0.000000	0.494955	0.000000	0.000000	0.0
21	9951	0.000000	0.000000	0.000000	0.000000	0.000000	0.0
22	9951	0.290302	0.000000	0.000000	0.000000	0.000000	0.0
23	9951	0.362225	0.000000	0.000000	0.000000	0.000000	0.0
24	9951	0.544156	0.000000	0.000000	0.000000	0.000000	0.0

	android	dictionary	java	way
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	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	0.531024	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.000000
6	0.493906	0.000000	0.000000	0.493906
7	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.420868	0.000000
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.455516	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.375491	0.000000	0.375491
17	0.000000	0.000000	0.469000	0.000000
18	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.000000
23	0.000000	0.000000	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.000000

	owneruserid	javascript	php	using	jquery	java	array \
0	49153	0.232620	0.000000	0.000000	0.000000	0.000000	0.301414
1	49153	0.000000	0.000000	0.359405	0.434229	0.000000	0.000000
2	49153	0.000000	0.321643	0.000000	0.000000	0.000000	0.000000
3	49153	0.000000	0.000000	0.000000	0.000000	0.31798	0.000000
4	49153	0.291616	0.000000	0.291616	0.352326	0.000000	0.000000

..
74	49153	0.320213	0.000000	0.000000	0.000000	0.000000	0.000000
75	49153	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
76	49153	0.000000	0.000000	0.299801	0.362216	0.000000	0.000000
77	49153	0.325344	0.000000	0.000000	0.000000	0.000000	0.421560
78	49153	0.000000	0.299440	0.000000	0.000000	0.000000	0.000000

	string	file	class	select
0	0.000000	0.0	0.0	0.000000
1	0.000000	0.0	0.0	0.563434
2	0.000000	0.0	0.0	0.000000
3	0.000000	0.0	0.0	0.000000
4	0.000000	0.0	0.0	0.000000
..
74	0.451053	0.0	0.0	0.000000
75	0.000000	0.0	0.0	0.000000
76	0.000000	0.0	0.0	0.469993
77	0.000000	0.0	0.0	0.000000
78	0.000000	0.0	0.0	0.000000

[79 rows x 11 columns]

	owneruserid	python	string	list	wpf	values	index \
0	51816	0.000000	0.000000	0.000000	0.000000	0.0	0.0
1	51816	0.000000	0.000000	0.000000	0.000000	0.0	0.0
2	51816	0.167437	0.000000	0.000000	0.000000	0.0	0.0
3	51816	0.235714	0.000000	0.000000	0.000000	0.0	0.0
4	51816	0.000000	0.000000	0.000000	0.000000	0.0	0.0
..
61	51816	0.000000	0.000000	0.273919	0.283864	0.0	0.0
62	51816	0.000000	0.327286	0.000000	0.000000	0.0	0.0
63	51816	0.238198	0.000000	0.000000	0.000000	0.0	0.0
64	51816	0.283509	0.000000	0.000000	0.000000	0.0	0.0
65	51816	0.000000	0.000000	0.000000	0.000000	0.0	0.0

	function	class	net	value
0	0.000000	0.0	0.0	0.0
1	0.000000	0.0	0.0	0.0
2	0.66757	0.0	0.0	0.0
3	0.000000	0.0	0.0	0.0
4	0.000000	0.0	0.0	0.0
..
61	0.000000	0.0	0.0	0.0
62	0.000000	0.0	0.0	0.0
63	0.000000	0.0	0.0	0.0
64	0.000000	0.0	0.0	0.0
65	0.000000	0.0	0.0	0.0

[66 rows x 11 columns]

	owneruserid	vs	bash	java	list	instance	python	\
0	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
2	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
3	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
4	63051	0.000000	0.000000	0.000000	0.445721	0.000000	0.000000	
5	63051	0.000000	0.251584	0.000000	0.000000	0.000000	0.000000	
6	63051	0.000000	0.422632	0.000000	0.000000	0.000000	0.000000	
7	63051	0.000000	0.000000	0.526922	0.000000	0.000000	0.000000	
8	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
9	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.434555	
10	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
11	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
12	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
13	63051	0.405255	0.000000	0.000000	0.481714	0.000000	0.000000	
14	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
15	63051	0.000000	0.000000	0.000000	0.434555	0.000000	0.000000	
16	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
17	63051	0.230692	0.000000	0.548433	0.000000	0.000000	0.000000	
18	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
19	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
20	63051	0.000000	0.000000	0.000000	0.000000	0.801672	0.000000	
21	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
22	63051	0.000000	0.334818	0.000000	0.000000	0.000000	0.000000	
23	63051	0.286790	0.000000	0.000000	0.000000	0.000000	0.000000	
24	63051	0.000000	0.322928	0.000000	0.000000	0.000000	0.000000	
25	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
26	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
27	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
28	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
29	63051	0.000000	0.322928	0.000000	0.000000	0.000000	0.000000	
30	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
31	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
32	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.457795	
33	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
34	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
35	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
36	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
37	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
38	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
39	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
40	63051	0.000000	0.000000	0.355847	0.000000	0.000000	0.000000	
41	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
42	63051	0.445161	0.000000	0.000000	0.000000	0.000000	0.000000	

43	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
44	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
45	63051	0.000000	0.000000	0.000000	0.000000	0.474157	0.000000
46	63051	0.000000	0.000000	0.000000	0.000000	0.000000	0.355847
47	63051	0.503658	0.000000	0.000000	0.000000	0.000000	0.000000
48	63051	0.279854	0.000000	0.000000	0.000000	0.000000	0.000000

	linux	redirect	kill	output
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	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	0.000000
5	0.000000	0.284276	0.000000	0.284276
6	0.000000	0.477551	0.000000	0.000000
7	0.000000	0.000000	0.569917	0.000000
8	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000
11	0.000000	0.000000	0.000000	0.000000
12	0.332098	0.000000	0.000000	0.000000
13	0.000000	0.000000	0.000000	0.000000
14	0.000000	0.000000	0.000000	0.000000
15	0.000000	0.000000	0.000000	0.000000
16	0.000000	0.000000	0.000000	0.000000
17	0.000000	0.000000	0.000000	0.000000
18	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000
22	0.000000	0.000000	0.000000	0.378327
23	0.000000	0.340899	0.000000	0.000000
24	0.000000	0.000000	0.000000	0.364891
25	0.445721	0.000000	0.482090	0.000000
26	0.000000	0.000000	0.000000	0.000000
27	0.327402	0.000000	0.000000	0.000000
28	0.000000	0.000000	0.000000	0.000000
29	0.000000	0.000000	0.000000	0.000000
30	0.000000	0.000000	0.000000	0.000000
31	0.000000	0.000000	0.000000	0.000000
32	0.000000	0.000000	0.000000	0.000000
33	0.000000	0.000000	0.000000	0.000000
34	0.000000	0.000000	0.000000	0.000000
35	0.000000	0.000000	0.000000	0.000000
36	0.000000	0.000000	0.000000	0.000000
37	0.000000	0.000000	0.000000	0.000000
38	0.000000	0.000000	0.000000	0.000000
39	0.000000	0.000000	0.000000	0.000000

40	0.000000	0.000000	0.000000	0.000000
41	0.000000	0.000000	0.000000	0.000000
42	0.000000	0.000000	0.000000	0.000000
43	0.000000	0.000000	0.000000	0.000000
44	0.000000	0.000000	0.000000	0.000000
45	0.000000	0.000000	0.000000	0.000000
46	0.000000	0.000000	0.000000	0.000000
47	0.000000	0.000000	0.000000	0.000000
48	0.000000	0.000000	0.000000	0.000000

	owneruserid	operator	array	processing	copy	idiom	swap \
0	87234	0.0	0.603023	0.603023	0.000000	0.000000	0.000000
1	87234	0.0	0.000000	0.000000	0.57735	0.57735	0.57735
2	87234	1.0	0.000000	0.000000	0.000000	0.000000	0.000000

	faster	sorted	unsorted
0	0.301511	0.301511	0.301511
1	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000

	owneruserid	java	git	undo	jar	appending \
0	89904	0.000000	0.403194	0.403194	0.000000	0.000000
1	89904	0.313903	0.000000	0.000000	0.000000	0.000000
2	89904	0.000000	0.000000	0.000000	0.391176	0.000000
3	89904	0.000000	0.000000	0.000000	0.000000	0.707107
4	89904	0.423549	0.000000	0.000000	0.000000	0.000000
5	89904	0.356655	0.000000	0.000000	0.000000	0.000000
6	89904	0.000000	0.000000	0.000000	0.000000	0.000000
7	89904	0.000000	0.457985	0.457985	0.000000	0.000000
8	89904	0.313903	0.000000	0.000000	0.000000	0.000000
9	89904	0.000000	0.000000	0.000000	0.440595	0.000000

	objectoutputstream	swing	timer	open	profilers
0	0.000000	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.707107	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.640549	0.640549	0.000000	0.000000
5	0.000000	0.000000	0.000000	0.539382	0.539382
6	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000

	owneruserid	install	pip	installed	python	virtualenv	cache \
0	95592	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

1	95592	0.000000	0.000000	0.000000	0.386479	0.000000	0.000000
2	95592	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	95592	0.454195	0.454195	0.000000	0.000000	0.000000	0.541948
4	95592	0.000000	0.000000	0.435556	0.000000	0.000000	0.000000
5	95592	0.000000	0.000000	0.360493	0.000000	0.360493	0.000000
6	95592	0.347708	0.347708	0.000000	0.347708	0.347708	0.000000
7	95592	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	local	determining	flask	version
0	0.000000	0.000000	0.000000	0.000000
1	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000
3	0.541948	0.000000	0.000000	0.000000
4	0.000000	0.519708	0.519708	0.519708
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

	owneruserid	python	django	javascript	js	node	use	string	\
0	179736	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.0	
1	179736	0.000000	0.000000	0.390101	0.0	0.0	0.000000	0.0	
2	179736	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.0	
3	179736	0.000000	0.000000	0.000000	0.0	0.0	0.322733	0.0	
4	179736	0.338624	0.000000	0.000000	0.0	0.0	0.000000	0.0	
..	
110	179736	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.0	
111	179736	0.000000	0.324994	0.000000	0.0	0.0	0.334224	0.0	
112	179736	0.000000	0.000000	0.000000	0.0	0.0	0.261479	0.0	
113	179736	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.0	
114	179736	0.000000	0.000000	0.000000	0.0	0.0	0.000000	0.0	

	dictionary	remove	query
0	0.000000	0.0	0.0
1	0.000000	0.0	0.0
2	0.000000	0.0	0.0
3	0.000000	0.0	0.0
4	0.000000	0.0	0.0
..
110	0.000000	0.0	0.0
111	0.368768	0.0	0.0
112	0.000000	0.0	0.0
113	0.000000	0.0	0.0
114	0.000000	0.0	0.0

[115 rows x 11 columns]