In [174]:

```
import pandas as pd
df = pd.read_csv('dataSet.csv')
df.head()
```

Out[174]:

	Unkown	DATABASE_NAME	USER_NAME	SESSION_ID	PLAN_ID	TOTAL_SECONDS
0	6	CIW_STAGE	CIW_ETLUSER	8475199	10510659	0
1	11	CIW_STAGE	CIW_ETLUSER	8434579	10463758	4
2	12	CIW_STAGE	CIW_ETLUSER	8434681	10463851	0
3	13	CIW_STAGE	CIW_ETLUSER	8439700	10471018	0
4	20	CIW_STAGE	CIW_ETLUSER	8471503	10506752	0

In [175]:

```
df = df.drop(['Unkown','USER_NAME','SQL_TEXT_TYPE','DATABASE_NAME','SESSION_ID',
'PLAN_ID','TOTAL_SECONDS','SNIPPETS','THROUGH_PUT_ROWS','THROUGH_PUT_SIZE','newl
ine'],axis=1)
df.head()
```

Out[175]:

	SQL_TEXT_HASH	Clusters
0	1580626441	4
1	1463824509	4
2	1953664530	4
3	845098298	4
4	883804858	4

In [176]:

```
class my_dictionary(dict):

    # __init__ function
    def __init__(self):
        self = dict()

    # Function to add key:value
    def add(self, key, value):
        self[key] = value

    def checkKey(self, key):

        if key in self.keys():
            return True
        else:
            return False

    # Increment value
    def increment(self,key):

        if key in self.keys():

            self[key] += 1
```

In [177]:

```
"""
dic = dict()
key = 'Cluster1'
dict1.setdefault(key,{})
"""
Dict = my_dictionary()

for i in range(5,6):
    print(i)
    values = []
    rep_list = []
    nrep_list = []
    for data in df.index:
        if df.loc[data,'Clusters']==i:

            Ext = df['SQL_TEXT_HASH'][data]
            values.append(Ext)
    #print('all')
    #print(values)

    for j in range(len(values)):
        if values[j] not in values[j + 1:]:
            rep_list.append(values[j])

    for k in range(len(values)):
        if values[k] in values[k + 1:]:
            nrep_list.append(values[k])
    #print('non rep')
    #print(rep_list)
    #print('rep')
    #print(nrep_list)

    for j in range(len(rep_list)):
        key = str(rep_list[j])
        value = 1
        Dict.add(key,value)
    #print(Dict)

    for k in range(len(nrep_list)):
        key = str(nrep_list[k])
        present = Dict.checkKey(key)
        #print(present)
        if present == True:
            #print(key)
            Dict.increment(key) #increment(self,key
    print(Dict)
    #values.update({key : 1})
    #values = l(values)
```

```
5
{'1876377380': 1, '1654330104': 1, '75778317': 2, '1707657497': 2,
'379684407': 1, '754658653': 1, '1420412979': 2}
```

In [ ]: