In [1]:

```
import dill
#dill.dump_session('notebook_env.db')
#dill.load_session('notebook_env.db')

#https://stackoverflow.com/questions/34342155/how-to-pickle-or-store-jupyter-ipython-no
tebook-session-for-later
#https://www.reddit.com/r/IPython/comments/6reiqp/how_can_i_save_and_load_the_state_of_
the_kernel/dl6f2yn/
```

In [2]:

```
import pandas as pd
import numpy as np

import warnings
import math
import datetime
```

In [3]:

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix

from sklearn import tree
from sklearn.neighbors import KNeighborsClassifier
from sklearn import svm
from sklearn.neural_network import MLPClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.dummy import DummyClassifier
import matplotlib as mt
import pickle
warnings.filterwarnings('ignore')
```

In [4]:

```
from imblearn.over_sampling import SMOTE
#from imblearn.under_sampling import NeighbourhoodCleaningRule
from imblearn.under_sampling import RandomUnderSampler
```

In [5]:

```
%matplotlib notebook
```

```
In [6]:
```

```
i clf tree = 1
i_clf_knn = 2
i_clf_svm = 3
i_clf_mlp = 4
i_clf_naive = 5
i_clf_dummy = 6
nomes_algs = ['Tree', 'KNN', 'SVM', 'MLP', 'Naive', 'Dummy']
def ObterAlgoritmoClf(tp_algoritmo):
    if tp algoritmo == i clf tree:
        return tree.DecisionTreeClassifier(criterion='gini', max_depth=5)
    elif tp_algoritmo == i_clf_knn:
        return KNeighborsClassifier(n_neighbors=3)
    elif tp_algoritmo == i_clf_svm:
        return svm.SVC(C=1.0, kernel='sigmoid')
    elif tp_algoritmo == i_clf_mlp:
        return MLPClassifier(hidden_layer_sizes=100, activation='relu')
    elif tp_algoritmo == i_clf_naive:
        return GaussianNB()
    elif tp_algoritmo == i_clf_dummy:
        return DummyClassifier(strategy='prior')
    else:
        return None
def ObterMatrizConfusao(clf, features_teste, target_teste):
    cls_predict = clf.predict(features_teste)
    mat_conf = confusion_matrix(target_teste, cls_predict)
    score = clf.score(features_teste, cls_predict)
    return mat_conf
def PrepararLista(lista):
    listaCopy = lista.copy()
    modelo = { }
    for col in list(lista.columns.values):
        print(col)
        if (lista.dtypes[col] == 'object'):
            col le = LabelEncoder()
            col_labels = col_le.fit_transform(lista[col])
            col mapping = {index: label for index, label in enumerate(col le.classes )}
            modelo[col] = col_mapping
            listaCopy[col] = col_labels
        else:
            modelo[col] = 'tipado'
    return modelo, listaCopy
def DM_Calcular(mod_matriz_confusao, ret):
    matriz_confusao = mod_matriz_confusao
    VP = matriz confusao[0,0]
    FP = matriz_confusao[0,1]
    FN = matriz confusao[1,0]
    VN = matriz_confusao[1,1]
    sensibilidade = VP/(VP+FN)*100
    especificidade = VN/(VN+FP)*100
    if np.isnan(sensibilidade):
```

```
sensibilidade = 0
    if np.isnan(especificidade):
        especificidade = 0
                  = ((VP+VN)/(VP+VN+FP+FN))*100
                   = (VP*VN - FP*FN) / Math.sqrt((VP + FP)*(VP + FN)*(VN + FP)*(VN + F)
    #fi
N))
    VPP
                  = (VP/(VP+FP))*100
    VPN
                  = (VN/(VN+FN))*100
                 = (sensibilidade+especificidade)/2
    eficiencia
                  = VP+FP
    totalP
    totalN
                  = VN+FN
                 = VP+FN
    totalVPFN
                  = FP+VN
    totalFPVN
    totalG
                  = VP+FP+FN+VN
    if ret == 'sens':
        return round(sensibilidade, 2)
    elif ret == 'esp':
       return round(especificidade, 2)
    elif ret == 'acur':
        return round(acuracia, 2)
    elif ret == 'VPP':
       return round(VPP, 2)
    elif ret == 'VPN':
        return round(VPN, 2)
    elif ret == 'efic':
       return round(eficiencia, 2)
    elif ret == 'totP':
        return totalP
    elif ret == 'totN':
        return totalN
    elif ret == 'totG':
        return totalG
    else:
        return -1
```

In [7]:

```
colsExibir = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'VPP', 'VPN', 'TExe
c' ]
metr_alg = ['sens', 'esp', 'acur', 'VPP', 'VPN', 'efic', 'totP', 'totN', 'totG' ]
cols = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'VPP', 'VPN', 'mat_conf',
'AlgBin', 'TExec' ]
colsExibirMin = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'TExec' ]
```

```
def ExibirMedidas(X_train, X_test, y_train, y_test):
    algs = []
    for idx in range(1, 7):
        dtIni = datetime.datetime.now()
        print(dtIni)
        d = dict(id_alg=idx, algoritmo=nomes_algs[idx - 1])
        print(d)
        clf = ObterAlgoritmoClf(idx)
        # analisar o resultado depois retirar filtro
        #if idx != i_clf_tree:
            continue
        clf.fit(X_train, y_train)
        d['AlgBin'] = clf
        mat_conf = ObterMatrizConfusao(clf, X_test, y_test)
        for mtr in metr_alg:
            d[mtr] = DM_Calcular(mat_conf, mtr)
        d['mat_conf'] = mat_conf
        dtFim = datetime.datetime.now()
        dtDiff = dtFim - dtIni
        d['TExec'] = round(dtDiff.total_seconds() / 60, 2)
        print(clf)
        algs.append(d)
    dfAlg = pd.DataFrame(algs)
    dfAlg = dfAlg.fillna(0)
    print(datetime.datetime.now())
    return dfAlg[cols]
def ReExibirMedidas(dfAlg2, X_test, y_test):
    lstdfAlg = dfAlg2.T.to_dict()
    algs = []
    print('ReExibirMedidas')
    for idx in range(1, 7):
        row = lstdfAlg[idx-1]
        print(row['algoritmo'])
        clf = row['AlgBin']
        dtIni = datetime.datetime.now()
        print(dtIni)
        mat_conf = ObterMatrizConfusao(clf, X_test, y_test)
        for mtr in metr_alg:
            row[mtr] = DM_Calcular(mat_conf, mtr)
        #d['Modelo'] = pickle.dumps(clf)
        print(clf)
        dtFim = datetime.datetime.now()
        dtDiff = dtFim - dtIni
```

```
row['TExec'] = round(dtDiff.total_seconds() / 60, 2)

algs.append(row)

dfAlg = pd.DataFrame(algs)
 dfAlg = dfAlg.fillna(0)

return dfAlg[cols]
```

In [9]:

```
def GerarResampling(tipo, X, y):
    if tipo == 'over':
        sm = SMOTE(random_state=42)
        return sm.fit_resample(X, y)
    elif tipo == 'under':
        #cnn = NeighbourhoodCleaningRule()
        cnn = RandomUnderSampler(random_state=42)
        return cnn.fit_resample(X, y)
```

In [10]:

```
def ExibirDesbanciamento(y):
    dfDesb = pd.DataFrame({'col': y})
    ig_verd = dfDesb['col'][ dfDesb['col'] == 1 ].size
    ig_fals = dfDesb['col'][ dfDesb['col'] == 0 ].size
    lstGrf = []
    dicv = dict()
    dicv['Atr'] = 'Certificado'
    dicv['Qtd'] = ig_verd
    lstGrf.append(dicv)
    dicf = dict()
    dicf['Atr'] = 'Não Certificado'
    dicf['Qtd'] = ig_fals
    lstGrf.append(dicf)
    if ig_verd > ig_fals:
        print(round(ig_fals / ig_verd, 2))
    else:
        print(round(ig_verd / ig_fals, 2))
    return pd.DataFrame(lstGrf)
```

In [11]:

```
path_arq_dir = r'D:\Dados\bstoll\Documents\SuperOneNotes\jupyter-notebook\mooc-dataset'
path_arq = path_arq_dir + '\\' + r'HMXPC13_DI_v2_5-14-14.csv'

df = pd.read_csv(path_arq)
```

In [12]:

df.head(5)

Out[12]:

	course_id	userid_DI	registered	viewed	explored	certified	fina
0	HarvardX/CB22x/2013_Spring	MHxPC130442623	1	0	0	0	
1	HarvardX/CS50x/2012	MHxPC130442623	1	1	0	0	
2	HarvardX/CB22x/2013_Spring	MHxPC130275857	1	0	0	0	
3	HarvardX/CS50x/2012	MHxPC130275857	1	0	0	0	
4	HarvardX/ER22x/2013_Spring	MHxPC130275857	1	0	0	0	
4							

In [13]:

df.describe()

Out[13]:

	registered	viewed	explored	certified	YoB	neve
count	641138.0	641138.000000	641138.000000	641138.000000	544533.000000	441987.000
mean	1.0	0.624299	0.061899	0.027587	1985.253279	431.008
std	0.0	0.484304	0.240973	0.163786	8.891814	1516.116
min	1.0	0.000000	0.000000	0.000000	1931.000000	1.000
25%	1.0	0.000000	0.000000	0.000000	1982.000000	3.000
50%	1.0	1.000000	0.000000	0.000000	1988.000000	24.000
75%	1.0	1.000000	0.000000	0.000000	1991.000000	158.000
max	1.0	1.000000	1.000000	1.000000	2013.000000	197757.000
4						+

```
In [14]:
```

```
df.dtypes
```

Out[14]:

```
course_id
                       object
userid_DI
                       object
registered
                        int64
viewed
                        int64
explored
                        int64
certified
                        int64
final_cc_cname_DI
                       object
LoE_DI
                       object
YoB
                      float64
gender
                       object
grade
                       object
start_time_DI
                       object
last_event_DI
                       object
                      float64
nevents
ndays_act
                      float64
nplay_video
                      float64
nchapters
                      float64
                        int64
nforum_posts
roles
                      float64
                      float64
incomplete_flag
dtype: object
```

In [15]:

```
classe = 'certified'
print(classe)
```

certified

In [16]:

```
#df['gender'] = df['gender'].fillna('')
#df['roles'] = df['roles'].fillna(0)
#df['YoB'] = df['YoB'].fillna(0)
```

In [17]:

```
In [18]:
```

```
del df['nevents']
del df['ndays_act']
del df['nplay_video']
del df['nchapters']
del df['nforum_posts']
df = df.rename(columns=
          {'Q_nevents':
                              'nevents',
           'Q_ndays_act':
                             'ndays_act',
           'Q nplay_video': 'nplay_video',
                             'nchapters',
           'Q_nchapters':
           'Q_nforum_posts': 'nforum_posts',
          })
#start_time_DI
#last_event_DI
#del df['Unnamed']
```

In [19]:

```
df['start_year'] = pd.DatetimeIndex(df['start_time_DI']).year
df['start_month'] = pd.DatetimeIndex(df['start_time_DI']).month

df['last_e_year'] = pd.DatetimeIndex(df['last_event_DI']).year
df['last_e_month'] = pd.DatetimeIndex(df['last_event_DI']).month
```

In [20]:

```
del df['start_time_DI']
del df['last_event_DI']
del df['userid_DI']
#del df['Unnamed: 0']
```

In [21]:

```
# diminuir a eficiencia dos algoritmos
del df['nplay_video']
del df['nforum_posts']
del df['nchapters']
del df['ndays_act']
del df['incomplete_flag']
del df['grade']
del df['explored']
#del df['age']
```

In [22]:

```
for col in df.columns:
    if col == classe:
        next

    df[col] = df[col].fillna(0)
    print('==========')
    print(col)
    print(df[col].unique())
```

```
______
course id
['HarvardX/CB22x/2013_Spring' 'HarvardX/CS50x/2012'
 'HarvardX/ER22x/2013_Spring' 'HarvardX/PH207x/2012_Fall'
'HarvardX/PH278x/2013_Spring' 'MITx/6.002x/2012_Fall'
'MITx/6.002x/2013_Spring' 'MITx/14.73x/2013_Spring'
'MITx/2.01x/2013_Spring' 'MITx/3.091x/2012_Fall'
'MITx/3.091x/2013_Spring' 'MITx/6.00x/2012_Fall' 'MITx/6.00x/2013_Spring'
'MITx/7.00x/2013_Spring' 'MITx/8.02x/2013_Spring'
'MITx/8.MReV/2013_Summer']
______
registered
[1]
______
viewed
[0 1]
______
certified
[0 1]
______
final_cc_cname_DI
['United States' 'France' 'Unknown/Other' 'Mexico' 'Australia' 'India'
 'Canada' 'Russian Federation' 'Other South Asia'
'Other North & Central Amer., Caribbean' 'Other Europe' 'Other Oceania'
'Japan' 'Other Africa' 'Colombia' 'Germany'
'Other Middle East/Central Asia' 'Poland' 'Indonesia' 'Other East Asia'
'Bangladesh' 'China' 'United Kingdom' 'Ukraine' 'Spain' 'Greece'
'Pakistan' 'Brazil' 'Nigeria' 'Egypt' 'Other South America' 'Portugal'
'Philippines' 'Morocco']
______
[0 'Secondary' "Bachelor's" "Master's" 'Doctorate' 'Less than Secondary']
______
YoB
  0. 2012. 1987. 1968. 1989. 1978. 1993. 1988. 1981. 1980. 1991. 1977.
1992. 1990. 1986. 1984. 1982. 1983. 1979. 1994. 1967. 1969. 1985. 1971.
1973. 1974. 1995. 1972. 1976. 1965. 1963. 1964. 1975. 1955. 1944. 1966.
1957. 1997. 2000. 1960. 1970. 1996. 1959. 1961. 1953. 1952. 1956. 1962.
1958. 1999. 1945. 2011. 1954. 1947. 1948. 1998. 1950. 1949. 1951. 1940.
1936. 1941. 1942. 2010. 2008. 2002. 1937. 2001. 1946. 1939. 1938. 2009.
1943. 1935. 2007. 2003. 1931. 1934. 2013.]
______
gender
roles
______
start year
[2012 2013]
______
start month
[12 10 2 9 1 6 7 3 8 4 5 11]
______
last e year
[2013. 0. 2012.]
______
last e month
[11. 0. 5. 3. 6. 8. 1. 12. 4. 2. 7. 9. 10.]
```

In [23]:

\		6				
,		cou	nt cou	ınt	c	ount
course_id	certified					
HarvardX/CB22x/2013_Spring	0	296	18 296	18	2	9618
11d1 val dit/ CD22x/ 2013_3p1 111g	1			884	-	384
HarvardX/CS50x/2012	0	1683			16	8334
11a1 Val ax/ C550x/ 2012	1	1003		.87		1287
HarvardX/ER22x/2013_Spring	0	550				5060
riai vai ux/ Litzzx/ 2013_3pi Tiig	1	23		346		2346
HanvandV/DH207v/2012 Fall					7	
HarvardX/PH207x/2012_Fall	0	397			3	9750
II IV/DU270 /2012 C :	1	18		342	-	1842
HarvardX/PH278x/2013_Spring		388			3	8891
NTT /44 TO /0040 6 :	1			11		711
MITx/14.73x/2013_Spring	0	257			2	5785
	1	20		85		2085
MITx/2.01x/2013_Spring	0	54		18		5418
	1			247	_	247
MITx/3.091x/2012_Fall	0	135			1	.3583
	1			32		632
MITx/3.091x/2013_Spring	0	60		001		6001
	1			.38		138
MITx/6.002x/2012_Fall	0	390	61 396	61	3	9061
	1	17	50 17	'50		1750
MITx/6.002x/2013_Spring	0	216	42 216	542	2	1642
	1	5	93 5	593		593
MITx/6.00x/2012_Fall	0	642	54 642	254	6	4254
	1	24	77 24	177		2477
MITx/6.00x/2013_Spring	0	564	62 564	162	5	6462
	1	12	53 12	253		1253
MITx/7.00x/2013_Spring	0	201	86 201	.86	2	0186
	1			323		823
MITx/8.02x/2013_Spring	0	302			3	0226
, , ,	1			322		822
MITx/8.MReV/2013_Summer	0	91		.80		9180
, o	1			297		297
	_	_	_			
		LoE_DI	YoB	gender	roles	\
		count	count	count	count	`
course_id	certified	Courre	counc	counc	counc	
HarvardX/CB22x/2013 Spring	0	29618	29618	29618	29618	
nai vai ax/ CD22x/ 2013_3pi 11ig	1	384	384	384	384	
HarvardX/CS50x/2012	0	168334	168334	168334	168334	
11a1 Val UX/ C330X/ 2012	1	1287	1287	1287	1287	
Hanyandy/ED22y/2012 Chning	_	55060	55060	55060		
HarvardX/ER22x/2013_Spring	0 1	2346	2346	2346	55060 2346	
HanvandV/DH207v/2012 Fall	0			39750		
HarvardX/PH207x/2012_Fall		39750	39750		39750	
H	1	1842	1842	1842	1842	
HarvardX/PH278x/2013_Spring		38891	38891	38891	38891	
NTT /44 TO /0040 C :	1	711	711	711	711	
MITx/14.73x/2013_Spring	0	25785	25785	25785	25785	
NTT (0.04 (0050 5 :	1	2085	2085	2085	2085	
MITx/2.01x/2013_Spring	0	5418	5418	5418	5418	
/2 22 /22	1	247	247	247	247	
MITx/3.091x/2012_Fall	0	13583	13583	13583	13583	
	1	632	632	632	632	
MITx/3.091x/2013_Spring	0	6001	6001	6001	6001	
	1	138	138	138	138	
MITx/6.002x/2012_Fall	0	39061	39061	39061	39061	
	1	1750	1750	1750	1750	
MITx/6.002x/2013_Spring	0	21642	21642	21642	21642	

1	593	593	593	593
0	64254	64254	64254	64254
1	2477	2477	2477	2477
0	56462	56462	56462	56462
1	1253	1253	1253	1253
0	20186	20186	20186	20186
1	823	823	823	823
0	30226	30226	30226	30226
1	822	822	822	822
0	9180	9180	9180	9180
1	297	297	297	297
	0 1 0 1 0 1 0	0 64254 1 2477 0 56462 1 1253 0 20186 1 823 0 30226 1 822 0 9180	0 64254 64254 1 2477 2477 0 56462 56462 1 1253 1253 0 20186 20186 1 823 823 0 30226 30226 1 822 822 0 9180 9180	0 64254 64254 64254 1 2477 2477 2477 0 56462 56462 56462 1 1253 1253 1253 0 20186 20186 20186 1 823 823 823 0 30226 30226 30226 1 822 822 822 0 9180 9180 9180

start_year start_month last_e_year

,		scar c_ycar	3 car c_morren	rasc_c_ycar
\				
		count	count	count
course_id	certified	20512	20512	22512
HarvardX/CB22x/2013_Spring	0	29618	29618	29618
	1	384	384	384
HarvardX/CS50x/2012	0	168334	168334	168334
	1	1287	1287	1287
HarvardX/ER22x/2013_Spring	0	55060	55060	55060
	1	2346	2346	2346
HarvardX/PH207x/2012_Fall	0	39750	39750	39750
	1	1842	1842	1842
HarvardX/PH278x/2013_Spring	0	38891	38891	38891
	1	711	711	711
MITx/14.73x/2013_Spring	0	25785	25785	25785
	1	2085	2085	2085
MITx/2.01x/2013_Spring	0	5418	5418	5418
	1	247	247	247
MITx/3.091x/2012_Fall	0	13583	13583	13583
_	1	632	632	632
MITx/3.091x/2013_Spring	0	6001	6001	6001
	1	138	138	138
MITx/6.002x/2012_Fall	0	39061	39061	39061
	1	1750	1750	1750
MITx/6.002x/2013_Spring	0	21642	21642	21642
	1	593	593	593
MITx/6.00x/2012 Fall	0	64254	64254	64254
· · · · · -	1	2477	2477	2477
MITx/6.00x/2013_Spring	0	56462	56462	56462
,	1	1253	1253	1253
MITx/7.00x/2013 Spring	0	20186	20186	20186
	1	823	823	823
MITx/8.02x/2013_Spring	0	30226	30226	30226
,	1	822	822	822
MITx/8.MReV/2013_Summer	0	9180	9180	9180
	1	297	297	297
	-	231	231	231

last_e_month count course_id certified HarvardX/CB22x/2013_Spring HarvardX/CS50x/2012 HarvardX/ER22x/2013_Spring HarvardX/PH207x/2012_Fall HarvardX/PH278x/2013_Spring 0

MITx/14.73x/2013_S	nnina	1 0			711 785			
MITIX/14./3X/2013_3	buring	1			765 285			
MITx/2.01x/2013_Sp	ring	0			418			
		1			247			
MITx/3.091x/2012_F	all	0			583			
MITx/3.091x/2013_S	nring	1 0			532 001			
, 5 (6) 1 / / 2 (15) _ 5	P8	1			138			
MITx/6.002x/2012_F	all	0		390	061			
MTT /6 000 /2012 6		1			750			
MITx/6.002x/2013_S	pring	0 1			542 593			
MITx/6.00x/2012_Fa	11	0			254			
, <u>-</u>		1			477			
MITx/6.00x/2013_Sp	ring	0			462			
MITx/7.00x/2013_Sp	ning	1 0			253 1 <i>06</i>			
MITX/7.00X/2013_SP	r.rug	1			186 823			
MITx/8.02x/2013_Sp	ring	0			226			
	_	1			322			
MITx/8.MReV/2013_S	ummer	0			180			
		1 =====			297 ======			
	cour	se_id	viewed	final_cc	cname DI	LoE_DI	YoE	3
\		_			_	_		
		count	count		count	count	count	=
registered certific		23451	623451		623451	623451	623451	l
1		17687	17687		17687		17687	
	_			tart_year		_		\
nogistaned contifi	col		roles s count	tart_year count		nth last _. unt	_e_year count	\
registered certific	co ed	unt	count	count	co	unt	count	\
registered certific 1 0 1	co ed 6234	unt 451 6			co 623	unt		\
1 0	coi ed 6234 170	unt 451 6 687	count 523451 17687	count 623451	co 623	unt 451	count 623451	\
1 0	coi ed 6234 170	unt 451 6 687 _e_mon	count 523451 17687 nth	count 623451	co 623	unt 451	count 623451	\
1 0 1	coo ed 6234 170 last	unt 451 6 687	count 523451 17687 nth	count 623451	co 623	unt 451	count 623451	\
1 0	coo ed 6234 170 last	unt 451 6 687 _e_mon	count 523451 17687 ith int	count 623451	co 623	unt 451	count 623451	\
1 0 1 registered certific	coo ed 6234 170 last	unt 451 6 687 _e_mon cou	count 523451 17687 oth unt	count 623451	co 623	unt 451	count 623451	\
1 0 1 registered certification 1 0 1 ================================	coo ed 6234 170 last ed	unt 451 6 687 _e_mon cou 6234 176	count 523451 17687 1th Int 151 587	623451 17687	623 17	unt 451 687	count 623451 17687	
1 0 1 registered certification 1 0 1 ================================	coded 6234 170 last_ed ====================================	unt 451 6 687 _e_mon cou 6234 176 ===== d regi	count 523451 17687 1th 1nt 151 887 ======	623451 17687	623 17 ======= cname_DI	unt 451 687 ====== LoE_DI	count 623451 17687	\
1 0 1 registered certification 1 0 1 ================================	coo ed 6234 170 last ed	unt 451 6 687 _e_mon cou 6234 176 ===== d regi	count 523451 17687 1th Int 151 587	623451 17687	623 17 ======= cname_DI	unt 451 687	count 623451 17687	
registered certification of the second certified certified of the second certified of the second certified certified of the second certified certified of the second certified c	coded 6234 170 last ed ======= ourse_ic coun	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t	count 523451 17687 1th 151 587 ====== stered count	623451 17687	623 17 ======= cname_DI count 240876	unt 451 687 LoE_DI count 240876	count 623451 17687 YoB count 240876	
registered certification of the second certified of th	coded 6234 170 last ed ====== course_ic coun 240870 382575	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6	240876 323451 17687 151 151 152 153 154 154 154 154 154 154 154 154 154 154	623451 17687	co 623 17 	unt 451 687 ====== LoE_DI count 240876 382575	count 623451 17687 YoB count 240876 382575	
registered certification of the second certified certified of the second certified of the second certified certified of the second certified certified of the second certified c	coded 6234 170 last ed ======= ourse_ic coun	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6	count 223451 17687 1th 151 1587 ====== stered count 240876	623451 17687	623 17 ======= cname_DI count 240876	unt 451 687 LoE_DI count 240876	count 623451 17687 YoB count 240876	
registered certification of the second certified certified of the second certified of the second certified certified of the second certified certified of the second certified	coled 623- 170 last ed ======: ourse_i coun 240870 382570 1768	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6 5	23451 17687 1th int 451 687 ====== .stered count 240876 382575 17687	count 623451 17687 ===================================	co 623 17 17 240876 382575 17687	unt 451 687 ====== LoE_DI count 240876 382575 17687	count 623451 17687 YoB count 240876 382575 17687	
registered certification of the second certified certified of the second certified of the second certified certified of the second certified certified of the second certified	coled 623- 170 last ed ======: ourse_i coun 240870 382570 1768	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6 5	23451 17687 1th 1nt 151 1687 240876 382575 17687	623451 17687	co 623 17 17 240876 382575 17687	unt 451 687 ====== LoE_DI count 240876 382575 17687	count 623451 17687 ===== YoB count 240876 382575 17687 ear \	
registered certification of the second certified of th	coned 6234 170 last_ed	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6 5 7 role coun	240876 382575 17687	count 623451 17687 ======= final_cc_o		======================================	count 623451 17687 YoB count 240876 382575 17687 ear \ unt	
registered certification of the second certified of th	coned 6234 176 last ed	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6 5 7 role coun	count 523451 17687 1th 151 587 ====== count 240876 382575 17687 es start		co 623 17 ======= cname_DI count 240876 382575 17687 rt_month count	unt 451 687 LoE_DI count 240876 382575 17687 last_e_ye	count 623451 17687 17687 240876 382575 17687 ear \ unt	
registered certification of the second certified of th	coled 6234 170 last ed ======= ourse_i coun 240870 382571 1768 gender count 240876 382575	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6 5 7 role coun 24087 38257	23451 17687 17687 14h 151 1687 1587 240876 382575 17687 128 start		co 623 17 623 17 count 240876 382575 17687 rt_month count 240876 382575	unt 451 687 ======= LoE_DI	count 623451 17687 17687 240876 382575 17687 ear \ unt 876 575	
registered certification of the second certified of th	coned 6234 176 last ed	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6 5 7 role coun	23451 17687 17687 14h 151 1687 1587 240876 382575 17687 128 start		co 623 17 ======= cname_DI count 240876 382575 17687 rt_month count	unt 451 687 ======= LoE_DI	count 623451 17687 17687 240876 382575 17687 ear \ unt	
registered certification of the second certified of th	colled 6234 176 1ast_ed =======: coun* 240876 382575 1768* gender count 240876 382575 17687 ast_e_means	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 65 7 role coun 24087 38257 1768 onth	23451 17687 17687 14h 151 1687 1587 240876 382575 17687 128 start		co 623 17 623 17 count 240876 382575 17687 rt_month count 240876 382575	unt 451 687 ======= LoE_DI	count 623451 17687 17687 240876 382575 17687 ear \ unt 876 575	
registered certification of the second certified of th	colled 6234 176 1ast_ed =======: coun* 240876 382575 1768* gender count 240876 382575 17687 ast_e_means	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 6 5 7 role coun 24087 38257 1768	23451 17687 17687 14h 151 1687 1587 240876 382575 17687 128 start		co 623 17 623 17 count 240876 382575 17687 rt_month count 240876 382575	unt 451 687 ======= LoE_DI	count 623451 17687 17687 240876 382575 17687 ear \ unt 876 575	
registered certification of the second certified of th	coled 6234 170 last ed ======= ourse_i coun 240870 382575 17687 gender count 240876 382575 17687 ast_e_mc	unt 451 6 687 _e_mon cou 6234 176 ===== d regi t 65 7 role coun 24087 38257 1768 onth	23451 17687 17687 14h 151 1687 1587 240876 382575 17687 128 start		co 623 17 623 17 count 240876 382575 17687 rt_month count 240876 382575	unt 451 687 ======= LoE_DI	count 623451 17687 17687 240876 382575 17687 ear \ unt 876 575	

1	0 1		382575 17687							
\	====		course_ic	regist	ered	viewed	final_	_ccc	name_DI	== LoE_DI
	ified	certified	count	: с	ount	count	:		count	count
0	TITCU	0		. 62 ' 1	_	623451 17687			623451 17687	623451 17687
cert	ified	certified		gender count	rol cou		rt_year count	star	rt_month count	\
0 1		0	623451 17687	623451 17687	6234 176		623451 17687		623451 17687	
cont	ifiod	certified	last_e_ye cou			nth unt				
0 1	TITEU	0 1	6234 176		623 17	451 687				
====	=====	=======	=======	======	=====	======			======== · ,	==
wed	\								registere	
unt	_							unt	coun	t co
	l_cc_ ralia	cname_DI			ce 0	rtified		5223	622	3 6
					1			196	19	6
196 Bang 148	lades	h			0		3	3148	314	8 3
					1			34	34	4
34 Braz 403	il				0		17	403	1740	3 17
453					1			453	45	3
Cana	da				0		12	405	1240	5 12
405					1			333	33.	3
333 Chin	а				0		5	108	510	8 5
108					1			62	6	2
	mbia				0		4	601	460	1 4
601					1			202	20	2
202 Egyp	t				0		9	166	916	6 9
166					1			120	12	0
120 Fran	CE				0		Δ	1496	449	6 4
496					1		7	204	20	
204 Germ 612	any				0		7	204 '612	761	
~ 										

4.50	1	462	462	
462 Greece	0	4845	4845	4
845	1	317	317	
317 India	0	85491	85491	85
491	1	3205	3205	3
205 Indonesia	0	3314	3314	3
314	1	96	96	
96 Japan	0	2230	2230	2
230	1	40	40	
40 Mexico	0	5470	5470	5
470	1	168	168	
168 Morocco	0	3938	3938	3
938	1	28	28	3
28	Τ	20	20	
•••		• • •	• • •	
Other Middle East/Central Asia 005	0	17005	17005	17
	1	320	320	
320 Other North & Central Amer., Caribbean	0	4265	4265	4
265	1	169	169	
169 Other Oceania	0	341	341	
341	1	5	5	
5 Other South America	0	9624	9624	9
624	1	292	292	
292 Other South Asia	0	12584	12584	12
584	1	408	408	
408 Pakistan	0	10667	10667	10
667	1	157	157	
157 Philippines	0	5293	5293	5
293	1	81	81	
81 Poland	0	4813	4813	4
813	1	413	413	
413 Portugal	0	2081	2081	2
U		-		

081	1	112		112	
112					
Russian Federation 797	0	9797		9797	9
	1	635		635	
635 Spain	0	9166		9166	9
166				J100	
837	1	837		837	
Ukraine	0	3897		3897	3
897	1	203		203	
203	1	203		203	
United Kingdom 261	0	21261	2	1261	21
201	1	870		870	
870 United States	0	170050	17	0050	179
858	Ø	179858	17	9858	1/9
202	1	4382		4382	4
382 Unknown/Other 968	0	81968	8	1968	81
	1	61		61	
61					
		LoE_DI	YoB	gend	er
\					
1					
	certified	count	count	cou	nt
` final_cc_cname_DI Australia	certified 0	count 6223	count 6223	cou 62	
final_cc_cname_DI				62	
final_cc_cname_DI	0	6223	6223	62	23 96
<pre>final_cc_cname_DI Australia Bangladesh</pre>	0 1 0 1	6223 196 3148 34	6223 196 3148 34	62 1 31	23 96 48 34
final_cc_cname_DI Australia	0 1 0 1 0	6223 196 3148 34 17403	6223 196 3148 34 17403	62 1 31 174	23 96 48 34 03
<pre>final_cc_cname_DI Australia Bangladesh Brazil</pre>	0 1 0 1 0	6223 196 3148 34 17403 453	6223 196 3148 34 17403 453	62 1 31 174 4	23 96 48 34 03 53
<pre>final_cc_cname_DI Australia Bangladesh</pre>	0 1 0 1 0 1	6223 196 3148 34 17403 453 12405	6223 196 3148 34 17403 453 12405	62 1 31 174 4 124	23 96 48 34 03 53
final_cc_cname_DI Australia Bangladesh Brazil Canada	0 1 0 1 0 1 0	6223 196 3148 34 17403 453 12405 333	6223 196 3148 34 17403 453 12405 333	62 1 31 174 4 124 3	23 96 48 34 03 53 05
<pre>final_cc_cname_DI Australia Bangladesh Brazil</pre>	0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108	6223 196 3148 34 17403 453 12405 333 5108	62 1 31 174 4 124 3 51	23 96 48 34 03 53 05 33
<pre>final_cc_cname_DI Australia Bangladesh Brazil Canada China</pre>	0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62	6223 196 3148 34 17403 453 12405 333 5108 62	62 1 31 174 4 124 3 51	23 96 48 34 03 53 05 33 08
final_cc_cname_DI Australia Bangladesh Brazil Canada	0 1 0 1 0 1 0 1 0	6223 196 3148 34 17403 453 12405 333 5108 62 4601	6223 196 3148 34 17403 453 12405 333 5108 62 4601	62 1 31 174 4 124 3 51	23 96 48 34 03 53 05 33 08 62
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia	0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62	6223 196 3148 34 17403 453 12405 333 5108 62	62 1 31 174 4 124 3 51	23 96 48 34 03 53 05 33 08 62 01
<pre>final_cc_cname_DI Australia Bangladesh Brazil Canada China</pre>	0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202	62 1 31 174 4 124 3 51 46 2	23 96 48 34 03 53 05 33 08 62 01
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia	0 1 0 1 0 1 0 1 0 1 0	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166	62 1 31 174 4 124 3 51 46 2	23 96 48 34 03 53 05 33 08 62 01 02 66 20
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt	0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120	62 1 31 174 4 124 3 51 46 2 91 1	23 96 48 34 03 53 05 33 08 62 01 02 66 20
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt	0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76	23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany	0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4	23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France	0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48	23 96 48 34 03 53 05 33 06 62 01 02 66 04 12 62 45
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3	23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854	23 96 48 34 03 53 05 33 862 01 02 66 20 96 45 17 91
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece India	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 462 4845 317 85491 3205	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32	23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45 17 91
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32 33	23 96 48 34 03 53 05 33 86 20 96 45 17 91 95 14
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece India Indonesia	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 462 4845 317 85491 3205	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32 33	23 96 48 34 03 53 05 33 06 62 01 02 66 62 04 12 64 51 7 91 95 14 96
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece India	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314 96	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314 96	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32 33	23 96 48 34 03 53 05 33 06 62 01 02 66 62 96 45 17 91 95 14 96
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece India Indonesia	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314 96 2230	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314 96 2230	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32 33	23 96 48 34 03 53 05 33 862 01 02 66 20 96 45 17 91 95 14 96 30 40
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece India Indonesia Japan	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314 96 2230 40	6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314 96 2230 40	62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32 33	23 96 48 34 03 53 05 33 862 01 02 66 20 96 45 17 91 95 14 96 30 40

	1	28	28	28
 Other Middle East/Central Asia	0	 17005	 17005	 17005
Other Middle Last/Central Asia	1	320	320	320
Other North & Central Amer., Caribbean		4265	4265	4265
other north a central raner, car robean	1	169	169	169
Other Oceania	0	341	341	341
	1	5	5	5
Other South America	0	9624	9624	9624
	1	292	292	292
Other South Asia	0	12584	12584	12584
	1	408	408	408
Pakistan	0	10667	10667	10667
DI 171	1	157	157	157
Philippines	0	5293	5293	5293
Daland	1	81	81	81
Poland	0	4813 413	4813 413	4813 413
Portugal	1 0	2081	2081	2081
roi tugai	1	112	112	112
Russian Federation	0	9797	9797	9797
Nassian Featración	1	635	635	635
Spain	0	9166	9166	9166
•	1	837	837	837
Ukraine	0	3897	3897	3897
	1	203	203	203
United Kingdom	0	21261	21261	21261
	1	870	870	870
United States	0	179858	179858	179858
	1	4382	4382	4382
Unknown/Other	0	81968	81968	81968
OHRHOWH/ OTHER				
Olikilowity Octici	1	61	61	61
Olikilowity o'clici		61	61	61
Olikilowity o'clici		61 roles	61 start_ye	61 ar \
	1	61	61	61 ar \
final_cc_cname_DI	1 certified	roles	61 start_ye cou	61 ar \ nt
	1 certified 0	roles count	61 start_ye cou 62	61 ar \ nt
<pre>final_cc_cname_DI Australia</pre>	1 certified 0 1	roles count 6223 196	61 start_ye cou 62 1	61 ar \ nt 23 96
final_cc_cname_DI	1 certified 0 1	61 roles count 6223 196 3148	61 start_ye cou 62 1 31	61 ar \ nt 23 96 48
<pre>final_cc_cname_DI Australia Bangladesh</pre>	certified 0 1 0 1	61 roles count 6223 196 3148 34	61 start_ye cou 62 1 31	61 ar \ nt 23 96 48 34
<pre>final_cc_cname_DI Australia</pre>	1 certified 0 1	61 roles count 6223 196 3148	61 start_ye cou 62 1 31	61 ar \ nt 23 96 48 34
<pre>final_cc_cname_DI Australia Bangladesh</pre>	certified 0 1 0 1	61 roles count 6223 196 3148 34 17403	61 start_ye cou 62 1 31	61 ar \ nt 23 96 48 34 03 53
<pre>final_cc_cname_DI Australia Bangladesh Brazil</pre>	certified 0 1 0 1	61 roles count 6223 196 3148 34 17403 453	61 start_ye cou 62 1 31 174 4 124	61 ar \ nt 23 96 48 34 03 53
<pre>final_cc_cname_DI Australia Bangladesh Brazil</pre>	<pre>certified 0 1 0 1 0 1 0 0</pre>	61 roles count 6223 196 3148 34 17403 453 12405	61 start_ye cou 62 1 31 174 4 124	61 ar \ nt 23 96 48 34 03 53 05 33
<pre>final_cc_cname_DI Australia Bangladesh Brazil Canada China</pre>	1 certified 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62	61 start_ye cou 62 1 31 174 4 124 3 51	61 ar \ nt 23 96 48 34 03 53 05 33 08 62
final_cc_cname_DI Australia Bangladesh Brazil Canada	1 certified 0 1 0 1 0 1 0 1 0	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601	61 start_ye cou 62 1 31 174 4 124 3 51	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia	1 certified 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202	61 start_ye cou 62 1 31 174 4 124 3 51 46 2	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02
<pre>final_cc_cname_DI Australia Bangladesh Brazil Canada China</pre>	1 certified 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt	1 certified 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1	61 ar \ nt 23 96 48 34 03 53 05 33 06 62 01 02 66 20
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia	1 certified 0 1 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 48	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45 17
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 48	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45 17 91
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45 17 91 05
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece India	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32 33	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45 17 91 05
final_cc_cname_DI Australia Bangladesh Brazil Canada China Colombia Egypt France Germany Greece India	1 certified 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	61 roles count 6223 196 3148 34 17403 453 12405 333 5108 62 4601 202 9166 120 4496 204 7612 462 4845 317 85491 3205 3314	61 start_ye cou 62 1 31 174 4 124 3 51 46 2 91 1 44 2 76 4 48 3 854 32 33	61 ar \ nt 23 96 48 34 03 53 05 33 08 62 01 02 66 20 96 04 12 62 45 17 91 05 14

	1	40	40
Mexico	0	5470	5470
	1	168	168
Morocco	0	3938	3938
	1	28	28
•••		• • •	
Other Middle East/Central Asia	0	17005	17005
	1	320	320
Other North & Central Amer., Caribbean	0	4265	4265
	1	169	169
Other Oceania	0	341	341
	1	5	5
Other South America	0	9624	9624
	1	292	292
Other South Asia	0	12584	12584
	1	408	408
Pakistan	0	10667	10667
	1	157	157
Philippines	0	5293	5293
	1	81	81
Poland	0	4813	4813
	1	413	413
Portugal	0	2081	2081
	1	112	112
Russian Federation	0	9797	9797
	1	635	635
Spain	0	9166	9166
	1	837	837
Ukraine	0	3897	3897
	1	203	203
United Kingdom	0	21261	21261
	1	870	870
United States	0	179858	179858
	1	4382	4382
Unknown/Other	0	81968	81968
	1	61	61

start_month last_e_year

		count	count
<pre>final_cc_cname_DI</pre>	certified		
Australia	0	6223	6223
	1	196	196
Bangladesh	0	3148	3148
	1	34	34
Brazil	0	17403	17403
	1	453	453
Canada	0	12405	12405
	1	333	333
China	0	5108	5108
	1	62	62
Colombia	0	4601	4601
	1	202	202
Egypt	0	9166	9166
	1	120	120
France	0	4496	4496
	1	204	204
Germany	0	7612	7612
	1	462	462
Greece	0	4845	4845
	1	317	317

\

India	0	85491	85491
	1	3205	3205
Indonesia	0	3314	3314
	1	96	96
Japan	0	2230	2230
	1	40	40
Mexico	0	5470	5470
	1	168	168
Morocco	0	3938	3938
	1	28	28
Other Middle Fact/Control Acia	0	17005	17005
Other Middle East/Central Asia	0 1	320	320
Other North & Central Amer., Caribbean	_	4265	4265
other North & Central Amer., Caribbean	1	169	169
Other Oceania	0	341	341
other occurre	1	5	5
Other South America	0	9624	9624
	1	292	292
Other South Asia	0	12584	12584
	1	408	408
Pakistan	0	10667	10667
	1	157	157
Philippines	0	5293	5293
	1	81	81
Poland	0	4813	4813
	1	413	413
Portugal	0	2081	2081
	1	112	112
Russian Federation	0	9797	9797
	1	635	635
Spain	0	9166	9166
	1	837	837
Ukraine	0	3897	3897
Hottod Kinadom	1	203	203
United Kingdom	0 1	21261 870	21261 870
United States	0	179858	179858
officed States	1	4382	4382
Unknown/Other	0	81968	81968
OTIKITOWITY O'CITCI	1	61	61
	-	01	01
		last_e_month	
		count	
<pre>final_cc_cname_DI</pre>	certified		
Australia	0	6223	
	1	196	
Bangladesh	0	3148	
	1	34	
Brazil	0	17403	

1 Brazil Canada China ${\tt Colombia}$ Egypt France

Germany	0	7612
	1	462
Greece	0	4845
	1	317
India	0	85491
	1	3205
Indonesia	0	3314
	1	96
Japan	0	2230
	1	40
Mexico	0	5470
	1	168
Morocco	0	3938
	1	28
Other Middle East/Central Asia	0	17005
ocher filadic lase, central fista	1	320
Other North & Central Amer., Caribbean	=	4265
ocher nor en a central raner, caribbean	1	169
Other Oceania	0	341
ocher occania	1	5
Other South America	0	9624
John Journ America	1	292
Other South Asia	0	12584
	1	408
Pakistan	0	10667
	1	157
Philippines	0	5293
	1	81
Poland	0	4813
	1	413
Portugal	0	2081
	1	112
Russian Federation	0	9797
	1	635
Spain	0	9166
	1	837
Ukraine	0	3897
	1	203
United Kingdom	0	21261
	1	870
United States	0	179858
	1	4382
Unknown/Other	0	81968
	1	61

[68 rows x 11 columns]

		course_id	registered	viewed	<pre>final_cc_cname_</pre>
DI \		count	count	count	cou
nt					
LoE_DI	certified				
0	0	102092	102092	102092	1020
92					
	1	3916	3916	3916	39
16					
Bachelor's	0	214898	214898	214898	2148
98	_				
=-	1	4870	4870	4870	48
70					

Doctorate	0	12965	5 1	.2965 1	12965	129
65	1	422	2	422	422	4
22 Less than Secondary	Q	13690	a 1	.3690 1	13690	136
90						
02	1	402	2	402	402	4
Master's	0	113971	l 11	.3971 11	13971	1139
71	1	4218	3	4218	4218	42
18 Secondary	0	165835	5 16	55835 16	55835	1658
35	1	3859	2	3859	3859	38
59	1	3033	7	3033	2029	36
		YoB	gender	roles	start_year	start mon
th \						_
nt		count	count	count	count	cou
LoE_DI	certified	100000	100000	100000	100000	4000
0 92	0	102092	102092	102092	102092	1020
16	1	3916	3916	3916	3916	39
Bachelor's	0	214898	214898	214898	214898	2148
98	1	4870	4870	4870	4870	48
70						
Doctorate 65	0	12965	12965	12965	12965	129
22	1	422	422	422	422	4
22 Less than Secondary	0	13690	13690	13690	13690	136
90	1	402	402	402	402	4
02						
Master's 71	0	113971	113971	113971	113971	1139
	1	4218	4218	4218	4218	42
18 Secondary	0	165835	165835	165835	165835	1658
35	1	3859	3859	3859	3859	38
59	-	3633	3633	3633	3633	38
		last_e_ye	ear last	: e month	า	
		col		count		
LoE_DI 0	certified 0	1026	992	102092	2	
	1		916	3916		
Bachelor's	0 1	2148	398 370	214898 4876		
Doctorate	0	129	965	12965	5	
Loce than Cocondain	1		122	422		
Less than Secondary	1	136	102	13690 402		
Master's	0	1139	971	113971	L	
Socondany	1		218	4218		
Secondary	0	1658	333	165835	•	

1 3859 3859

		т			3033 ===========		_
					final_cc_cname_DI		
er \		count	count	count	count	count	cou
nt							
YoB 0.0	certified 0	92737	92737	92737	92737	92737	927
37	1	3868	3868	3868	3868	3868	38
68 1931.0	0	7	7	7	7	7	
7 1934.0	0	5	5	5	5	5	
5 1935.0	0	36	36	36	36	36	
36 1936.0	0	42	42	42	42	42	
42	1	1	1	1	1	1	
1 1937.0	0	60	60	60	60	60	
60	1	4	4	4	4	4	
4 1938.0	0	72	72	72	72	72	
72	1	2	2	2	2	2	
2 1939.0	0	80	80	80	80	80	
80	1	6	6	6	6	6	
6 1940.0	0	91	91	91	91	91	
91	1	1	1	1	1	1	
1 1941.0	0	94	94	94	94	94	
94	1	2	2	2	2	2	
1942.0	0	193	193	193	193	193	1
93	1	3	3	3	3	3	
1943.0 18	0	218	218	218	218	218	2
5	1	5	5	5	5	5	
1944.0 18	0	218	218	218	218	218	2
3	1	3	3	3	3	3	
1945.0 11	0	211	211	211	211	211	2
5	1	5	5	5	5	5	
1946.0 20	0	320	320	320	320	320	3
11	1	11	11	11	11	11	

1947.0	0	451	451	451	451	451	4
51	1	9	9	9	9	9	
9 1948.0	0	355	355	355	355	355	3
55 •••			•••				
1992.0	0	36789	36789	36789	36789	36789	367
89	1	1057	1057	1057	1057	1057	10
57 1993.0	0	32985	32985	32985	32985	32985	329
85	1	1014	1014	1014	1014	1014	10
14 1994.0	0	23411	23411	23411	23411	23411	234
11 10	1	510	510	510	510	510	5
1995.0	0	13546	13546	13546	13546	13546	135
46	1	260	260	260	260	260	2
60 1996.0	0	7518	7518	7518	7518	7518	75
18 36	1	236	236	236	236	236	2
1997.0 79	0	3979	3979	3979	3979	3979	39
03	1	103	103	103	103	103	1
1998.0 95	0	1895	1895	1895	1895	1895	18
	1	40	40	40	40	40	
1999.0 34	0	934	934	934	934	934	9
18	1	18	18	18	18	18	
2000.0 34	0	334	334	334	334	334	3
6	1	6	6	6	6	6	
2001.0 40	0	140	140	140	140	140	1
2	1	2	2	2	2	2	
2002.0 44	0	44	44	44	44	44	
1	1	1	1	1	1	1	
2003.0 10	0	10	10	10	10	10	
2007.0 6	0	6	6	6	6	6	
2008.0 10	0	10	10	10	10	10	
2009.0	0	8	8	8	8	8	
2010.0	0	17	17	17	17	17	

34	34	34	34	34	
472	472	472	472	472	4
61	61	61	61	61	
	472	472 472	472 472 472	472 472 472 472	472 472 472 472

		roles	start vear	start month	last e vear	last_e_month
		count	count	count	count	count
YoB	certified	counc	counc	counc	counc	counc
0.0	0	92737	92737	92737	92737	92737
	1	3868	3868	3868	3868	3868
1931.0	0	7	7	7	7	7
1934.0	0	5	5	5	5	5
1935.0	0	36	36	36	36	36
1936.0	0	42	42	42	42	42
	1	1	1	1	1	1
1937.0	0	60	60	60	60	60
	1	4	4	4	4	4
1938.0	0	72	72	72	72	72
	1	2	2	2	2	2
1939.0	0	80	80	80	80	80
	1	6	6	6	6	6
1940.0	0	91	91	91	91	91
	1	1	1	1	1	1
1941.0	0	94	94	94	94	94
	1	2	2	2	2	2
1942.0		193	193	193	193	193
	1	3	3	3	3	3
1943.0		218	218	218	218	218
	1	5	5	5	5	5
1944.0		218	218	218	218	218
1015.0	1	3	3	3	3	3
1945.0		211	211	211	211	211
1046 0	1	5	5	5	5	5
1946.0		320	320	320	320	320
1047.0	1	11	11	11	11	11
1947.0	1	451 9	451 9	451 9	451 9	451 9
1948.0						
1940.0	Ø	355	355	355	355	355
1992.0	0	36789	36789	36789	36789	36789
1992.0	1	1057	1057	1057	1057	1057
1993.0		32985	32985	32985	32985	32985
1000.0	1	1014	1014	1014	1014	1014
1994.0		23411	23411	23411	23411	23411
	1	510	510	510	510	510
1995.0		13546	13546	13546	13546	13546
	1	260	260	260	260	260
1996.0	0	7518	7518	7518	7518	7518
	1	236	236	236	236	236
1997.0	0	3979	3979	3979	3979	3979
	1	103	103	103	103	103
1998.0	0	1895	1895	1895	1895	1895
	1	40	40	40	40	40
1999.0	0	934	934	934	934	934
	1	18	18	18	18	18
2000.0	0	334	334	334	334	334
	1	6	6	6	6	6
2001.0	0	140	140	140	140	140

2002.0	1	2 44	2 44			2 14	2 44	2 44
2002.0	1	1	1		•	1	1	1
2003.0	0	10	10)	:	10	10	10
2007.0		6	6			6	6	6
2008.0		10	10		•	10	10	10
2009.0 2010.0		8 17	8 17			8 17	8 17	8 17
2010.0		34	34			34	34	34
2012.0		472	472			72	472	472
2013.0	0	61	61	•	(51	61	61
	ows x 11 c							
=====		course ic				======= d final_cc		
		count	_	unt	coun		count	
_	certified							
0	0	83199		199	83199		83199	83199
f	1	3607		607	360		3607	3607
Т	0 1	138563 4232		563 232	138563 4233		138563 4232	138563 4232
m	0	401672		.672	40167		401672	401672
	1	9848		848	9848		9848	9848
0	0	17	7	17	17	7	17	17
		YoB				start_mont		
gondon	certified	count	count		count	coun	t c	ount
gender.	0	83199	83199		83199	8319	9 8	3199
O	1	3607	3607		3607	360		3607
f	0	138563	138563	1	.38563	13856		8563
	1	4232	4232		4232	423	2	4232
m	0	401672	401672	4	01672	40167	2 40	1672
	1	9848	9848		9848	984		9848
0	0	17	17		17	1	7	17
		last_e_mc	onth					
			ount					
_	certified		1100					
0	0 1		3199 3607					
f	0	_	3563					
•	1		1232					
m	0	401	L672					
	1	9	848					
0	0		17					
=====	=======	counce id	nogistor	:====	·=====:	 final cc	======= cnamo DT	====== \ae DT \
		course_id count	cou		count	TINAT_CC_	count	LoE_DI \
roles	certified	counc			counc		courre	counc
0.0	0	623451	6234	51	623451		623451	623451
:	1	17687	176	87	17687		17687	17687
		_				tart_month		
roles	certified	count	count	C	ount	count	CO	unt
	0	623451 6	523451	62	3451	623451	623	451
	1		17687		.7687	17687		687

last_e_month
 count

25097 25097

43225 43225

25097 25097

43225 43225

1	893	893	893	893	893
	YoB gender	roles s	tart_year	last_e_year last	_e_m

onth		YoB	gender	roles	start_year	last_e_year	last_e_m
Official		count	count	count	count	count	С
ount start month	contified						
1 5388	0	75388	75388	75388	75388	75388	7
2678	1	2678	2678	2678	2678	2678	
2 4828	0	74828	74828	74828	74828	74828	7
2995	1	2995	2995	2995	2995	2995	
3 8007	0	58007	58007	58007	58007	58007	5
1528	1	1528	1528	1528	1528	1528	
4 0279	0	30279	30279	30279	30279	30279	3
457	1	457	457	457	457	457	
5 2080	0	32080	32080	32080	32080	32080	3
651	1	651	651	651	651	651	
6 0598	0	20598	20598	20598	20598	20598	2
416	1	416	416	416	416	416	
7 5454	0	35454	35454	35454	35454	35454	3
969	1	969	969	969	969	969	
8 3369	0	93369	93369	93369	93369	93369	9
2148	1	2148	2148	2148	2148	2148	
9 4145	0	64145	64145	64145	64145	64145	6
2412	1	2412	2412	2412	2412	2412	
10 0981	0	70981	70981	70981	70981	70981	7
2276	1	2276	2276	2276	2276	2276	
11 5097	0	25097	25097	25097	25097	25097	2
264	1	264	264	264	264	264	
12 3225	0	43225	43225	43225	43225	43225	4
893	1	893	893	893	893	893	
========	=======						
,		course_	_id regi	stered	viewed fir	nal_cc_cname_	_DI LoE_

DI \ count count count count cou

last_e_year certified

0.0	0	178942	178942	178942		178942	1789
42 12	1	12	12	12		12	
2012.0 66	0	117866	117866	117866		117866	1178
47	1	347	347	347		347	3
2013.0 43	0	326643	326643	326643		326643	3266
28	1	17328	17328	17328		17328	173
20							
			ender rol count cou		_year count	start_month count	\
last_e_year	certified						
0.0	0	178942 1°	78942 1789	42 1	78942	178942	
	1	12	12	12	12	12	
2012.0	0	117866 1	17866 1178	66 1	17866	117866	
	1	347	347 3	47	347	347	
2013.0	0	326643 3	26643 3266	43 3	26643	326643	
	1	17328	17328 173	28	17328	17328	
		last_e_mon					
last_e_year	certified	Cour	110				
0.0	0	1789	4 2				
0.0	1		12				
2012.0	0	1178					
2012.0	1		47				
2013.0	0	3266					
2013.0	J	2200					
	1	173					
	1	173	28		=====		=
	1		28 =======			 L_cc_cname_DI	= LoE
_DI \	_		28 ======= registered	viewed	final		
_DI \		course_id	28 ======= registered	viewed	final	L_cc_cname_DI	LoE
_DI \ unt last_e_mont	h certified	course_id	28 ======= registered count	viewed count	final	l_cc_cname_DI count	LoE co
_DI \ unt last_e_mont 0.0		course_id	28 ======= registered count	viewed	final	L_cc_cname_DI	LoE
_DI \ unt last_e_mont	h certified	course_id count count 1	28 ======= registered count 178942	viewed count 178942	final	L_cc_cname_DI count 178942	LoE co
_DI \ unt last_e_mont 0.0 942	h certified	course_id	28 ======= registered count 178942	viewed count 178942	final	l_cc_cname_DI count	LoE co
_DI \ unt last_e_mont 0.0 942	h certified 0	course_id count 1 178942	28 ======= registered count 178942	viewed count 178942	fina	L_cc_cname_DI count 178942 12	LoE co 178
_DI \ unt last_e_mont 0.0 942 12 1.0	h certified	course_id count count 1	28 ======= registered count 178942	viewed count 178942	fina	L_cc_cname_DI count 178942	LoE co
_DI \ unt last_e_mont 0.0 942	h certified 0 1	course_id count 178942 12	28 ======== registered count 178942 12	viewed count 178942 12 24499	final	l_cc_cname_DI count 178942 12 24499	LoE co 178
_DI \ unt last_e_mont 0.0 942 12 1.0 499	h certified 0	course_id count 1 178942	28 ======== registered count 178942 12	viewed count 178942 12 24499	final	L_cc_cname_DI count 178942 12	LoE co 178
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639	h certified 0 1	course_id count 178942 12	28 ========= registered count 178942 12 24499	viewed count 178942 12 24499 2639	final	l_cc_cname_DI count 178942 12 24499	LoE co 178
_DI \ unt last_e_mont 0.0 942 12 1.0 499	h certified 0 1 0	course_id count 178942 12 24499 2639	28 ========= registered count 178942 12 24499	viewed count 178942 12 24499 2639	final	L_cc_cname_DI count 178942 12 24499 2639	LoE co 178 24
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0	h certified 0 1 0	course_id count 178942 12 24499 2639	28 ====================================	viewed count 178942 12 24499 2639 44229	fina	L_cc_cname_DI count 178942 12 24499 2639	LoE co 178 24
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0	h certified 0 1 0 1	course_id count 178942 12 24499 2639 44229	28 ====================================	viewed count 178942 12 24499 2639 44229	fina	L_cc_cname_DI count 178942 12 24499 2639 44229	LoE co 178 24 2 44
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229	h certified 0 1 0 1	course_id count 178942 12 24499 2639 44229	28 ====================================	viewed count 178942 12 24499 2639 44229 1391	final	L_cc_cname_DI count 178942 12 24499 2639 44229	LoE co 178 24 2 44
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229	h certified 0 1 0 1 0	course_id count 178942 12 24499 2639 44229 1391	28 ====================================	viewed count 178942 12 24499 2639 44229 1391	final	L_cc_cname_DI	LoE co 178 24 2 44 1
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533	h certified 0 1 0 1 0	course_id count 178942 12 24499 2639 44229 1391	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533	fina	L_cc_cname_DI	LoE co 178 24 2 44 1
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533	h certified 0 1 0 1 0 1 0 1	course_id count 178942 12 24499 2639 44229 1391 71533 589	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533 589	final	L_cc_cname_DI	LoE
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533 589 4.0	h certified 0 1 0 1 0 1 0	course_id count 178942 12 24499 2639 44229 1391 71533	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533 589	final	L_cc_cname_DI	LoE co 178 24 2 44 1
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533	h certified 0 1 0 1 0 1 0 1 0	course_id count 178942 12 24499 2639 44229 1391 71533 589 41145	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533 589 41145	fina	L_cc_cname_DI count 178942 12 24499 2639 44229 1391 71533 589 41145	LoE
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533 589 4.0 145	h certified 0 1 0 1 0 1 0 1	course_id count 178942 12 24499 2639 44229 1391 71533 589	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533 589 41145	fina	L_cc_cname_DI	LoE
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533 589 4.0 145 710	h certified 0 1 0 1 0 1 0 1 0 1	course_id count 178942 12 24499 2639 44229 1391 71533 589 41145 710	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533 589 41145 710	final	L_cc_cname_DI count 178942 12 24499 2639 44229 1391 71533 589 41145 710	LoE
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533 589 4.0 145 710 5.0	h certified 0 1 0 1 0 1 0 1 0	course_id count 178942 12 24499 2639 44229 1391 71533 589 41145	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533 589 41145 710	final	L_cc_cname_DI count 178942 12 24499 2639 44229 1391 71533 589 41145	LoE
_DI \ unt last_e_mont 0.0 942 12 1.0 499 639 2.0 229 391 3.0 533 589 4.0 145 710	h certified 0 1 0 1 0 1 0 1 0 1	course_id count 178942 12 24499 2639 44229 1391 71533 589 41145 710	28 ====================================	viewed count 178942 12 24499 2639 44229 1391 71533 589 41145 710 42267	final	L_cc_cname_DI count 178942 12 24499 2639 44229 1391 71533 589 41145 710	LoE

353								
6.0 339	0	3333	9	33339	33339)	33339	33
	1	272	4	2724	2724		2724	2
724 7.0	0	3125	4	31254	31254	-	31254	31
254	1	262	9	2629	2629)	2629	2
629 8.0	0	4298	6	42986	42986	i	42986	42
986	1	352	8	3528	3528	;	3528	3
528 9.0	0	2773		27737	27737		27737	27
737	1	76		765	765		765	_,
765	1	70	J	703	703		703	
10.0 510	0	4651	0	46510	46510)	46510	46
11.0	0	2422	3	24223	24223		24223	24
223 12.0	0	1478	7	14787	14787	,	14787	14
787	1	34	7	347	347	•	347	
347								
			_				start_month	\
last_e_month	contified	count	count	coun	T	count	count	
0.0	0	178942	178942	17894	.2	178942	178942	
0.0	1	12	12		.2	12	12	
1.0	0	24499	24499			24499	24499	
	1	2639	2639			2639	2639	
2.0	0	44229	44229			44229	44229	
2.0	1	1391	1391			1391	1391	
3.0	0	71533	71533			71533	71533	
4.0	1 0	589 41145	589 41145			589 41145	589 41145	
4.0	1	710	710			710	710	
5.0	0	42267	42267			42267	42267	
	1	2353	2353			2353	2353	
6.0	0	33339	33339			33339	33339	
	1	2724	2724	272	4	2724	2724	
7.0	0	31254	31254			31254	31254	
	1	2629	2629			2629	2629	
8.0	0	42986	42986			42986	42986	
9.0	1 0	3528 27737	3528 27737			3528 27737	3528 27737	
5.0	1	765	765			765	765	
10.0	0	46510	46510			46510	46510	
11.0	0	24223	24223			24223	24223	
12.0	0	14787	14787	1478	7	14787	14787	
	1	347	347	34	.7	347	347	
		_						
		last_e_y						
last_e_month	certified	CO	unt					
0.0	0	178	942					
	1	1,0	12					
1.0	0	24	499					
	1		639					

1 0 1 0 0 0	3528 27737 765 46510 24223 14787 347	
0 1 0 0	27737 765 46510 24223	
0 1 0	27737 765 46510	
0 1	27737 765	
0	27737	
1	3528	
0	42986	
1	2629	
0	31254	
1	2724	
0	33339	
1	2353	
0	42267	
1	710	
0	41145	
1	589	
0	71533	
1	1391	
0	44229	
	1 0 1 0 1 0 1 0 1	1 1391 0 71533 1 589 0 41145 1 710 0 42267 1 2353 0 33339 1 2724 0 31254

In [24]:

4

df.describe()

Out[24]:

							_
	registered	viewed	certified	YoB	roles	start_year	
count	641138.0	641138.000000	641138.000000	641138.000000	641138.0	641138.000000	
mean	1.0	0.624299	0.027587	1686.120498	0.0	2012.530800	
std	0.0	0.484304	0.163786	710.240700	0.0	0.499051	
min	1.0	0.000000	0.000000	0.000000	0.0	2012.000000	
25%	1.0	0.000000	0.000000	1975.000000	0.0	2012.000000	
50%	1.0	1.000000	0.000000	1986.000000	0.0	2013.000000	
75%	1.0	1.000000	0.000000	1991.000000	0.0	2013.000000	
max	1.0	1.000000	1.000000	2013.000000	0.0	2013.000000	~
4						>	

In [25]:

```
data_dum = pd.get_dummies(df, drop_first=False)
#data_dum.head(10).to_csv('novoModelo.csv')
data_dum.shape
```

Out[25]:

(641138, 69)

In [26]:

modelo, dfTrat = PrepararLista(data_dum)

```
registered
viewed
certified
YoB
roles
start year
start month
last_e_year
last e month
course_id_HarvardX/CB22x/2013_Spring
course_id_HarvardX/CS50x/2012
course id HarvardX/ER22x/2013 Spring
course id HarvardX/PH207x/2012 Fall
course_id_HarvardX/PH278x/2013_Spring
course_id_MITx/14.73x/2013_Spring
course_id_MITx/2.01x/2013_Spring
course_id_MITx/3.091x/2012_Fall
course_id_MITx/3.091x/2013_Spring
course id MITx/6.002x/2012 Fall
course id MITx/6.002x/2013 Spring
course_id_MITx/6.00x/2012_Fall
course_id_MITx/6.00x/2013_Spring
course_id_MITx/7.00x/2013_Spring
course_id_MITx/8.02x/2013_Spring
course id MITx/8.MReV/2013 Summer
final_cc_cname_DI_Australia
final_cc_cname_DI_Bangladesh
final_cc_cname_DI_Brazil
final_cc_cname_DI_Canada
final_cc_cname_DI_China
final cc cname DI Colombia
final_cc_cname_DI_Egypt
final cc cname DI France
final_cc_cname_DI_Germany
final_cc_cname_DI_Greece
final_cc_cname_DI_India
final_cc_cname_DI_Indonesia
final_cc_cname_DI_Japan
final cc cname DI Mexico
final cc cname DI Morocco
final_cc_cname_DI_Nigeria
final cc cname DI Other Africa
final cc cname DI Other East Asia
final cc cname DI Other Europe
final cc cname DI Other Middle East/Central Asia
final cc cname DI Other North & Central Amer., Caribbean
final_cc_cname_DI_Other Oceania
final cc cname DI Other South America
final cc cname DI Other South Asia
final cc cname DI Pakistan
final cc cname DI Philippines
final_cc_cname_DI_Poland
final cc cname DI Portugal
final cc cname DI Russian Federation
final cc cname DI Spain
final cc cname DI Ukraine
final cc cname DI United Kingdom
final_cc_cname_DI_United States
final_cc_cname_DI_Unknown/Other
LoE DI 0
LoE DI Bachelor's
```

```
LoE_DI_Doctorate
LoE_DI_Less than Secondary
LoE_DI_Master's
LoE_DI_Secondary
gender_0
gender_f
gender_m
gender_o
```

In [27]:

dfTrat.describe()

Out[27]:

							_
	registered	viewed	certified	YoB	roles	start_year	
count	641138.0	641138.000000	641138.000000	641138.000000	641138.0	641138.000000	
mean	1.0	0.624299	0.027587	1686.120498	0.0	2012.530800	
std	0.0	0.484304	0.163786	710.240700	0.0	0.499051	
min	1.0	0.000000	0.000000	0.000000	0.0	2012.000000	
25%	1.0	0.000000	0.000000	1975.000000	0.0	2012.000000	
50%	1.0	1.000000	0.000000	1986.000000	0.0	2013.000000	
75%	1.0	1.000000	0.000000	1991.000000	0.0	2013.000000	
max	1.0	1.000000	1.000000	2013.000000	0.0	2013.000000	
8 rows	× 69 columi	ns					~
4						•	

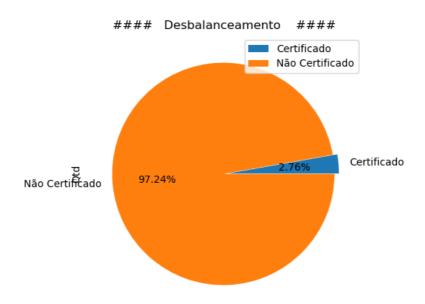
In [28]:

```
val_classes = dfTrat[classe].values
del dfTrat[classe]
```

In [29]:

```
dfDesbOrig = ExibirDesbanciamento(val_classes)
dfDesbOrig.plot.pie(x='Atr', y='Qtd', title='### Desbalanceamento ####', labels=d
fDesbOrig['Atr'].values, explode=(0.03, 0.01),autopct = '%1.2f%%', figsize=(8, 5))
dfDesbOrig
```

0.03



Out[29]:

	Atr	Qtd
0	Certificado	17687
1	Não Certificado	623451

In [30]:

```
X_train, X_test, y_train, y_test = train_test_split(dfTrat, val_classes, test_size=0.3,
random_state=42)
X_test_orig = X_test
y_test_orig = y_test
```

```
In [31]:
```

```
registered [ 0.0 ]
viewed [ 0.093 ]
YoB [ 0.0 ]
roles [ 0.0 ]
start_year [ 0.163 ]
start_month [ 0.228 ]
last_e_year [ 0.108 ]
last_e_month [ 0.217 ]
course_id_HarvardX/CB22x/2013_Spring [ 0.0 ]
course_id_HarvardX/CS50x/2012 [ 0.124 ]
course_id_HarvardX/ER22x/2013_Spring [ 0.012 ]
course_id_HarvardX/PH207x/2012_Fall [ 0.0 ]
course_id_HarvardX/PH278x/2013_Spring [ 0.0 ]
course_id_MITx/14.73x/2013_Spring [ 0.04 ]
course_id_MITx/2.01x/2013_Spring [ 0.0 ]
course_id_MITx/3.091x/2012_Fall [ 0.0 ]
course_id_MITx/3.091x/2013_Spring [ 0.0 ]
course_id_MITx/6.002x/2012_Fall [ 0.012 ]
course_id_MITx/6.002x/2013_Spring [ 0.0 ]
course_id_MITx/6.00x/2012_Fall [ 0.0 ]
course_id_MITx/6.00x/2013_Spring [ 0.0 ]
course_id_MITx/7.00x/2013_Spring [ 0.0 ]
course_id_MITx/8.02x/2013_Spring [ 0.0 ]
course_id_MITx/8.MReV/2013_Summer [ 0.0 ]
final_cc_cname_DI_Australia [ 0.0 ]
final_cc_cname_DI_Bangladesh [ 0.0 ]
final_cc_cname_DI_Brazil [ 0.0 ]
final_cc_cname_DI_Canada [ 0.0 ]
final_cc_cname_DI_China [ 0.0 ]
final_cc_cname_DI_Colombia [ 0.0 ]
final_cc_cname_DI_Egypt [ 0.0 ]
final_cc_cname_DI_France [ 0.0 ]
final_cc_cname_DI_Germany [ 0.0 ]
final_cc_cname_DI_Greece [ 0.0 ]
final_cc_cname_DI_India [ 0.003 ]
final_cc_cname_DI_Indonesia [ 0.0 ]
final_cc_cname_DI_Japan [ 0.0 ]
final_cc_cname_DI_Mexico [ 0.0 ]
final_cc_cname_DI_Morocco [ 0.0 ]
final_cc_cname_DI_Nigeria [ 0.0 ]
final_cc_cname_DI_Other Africa [ 0.0 ]
final cc cname DI Other East Asia [ 0.0 ]
final_cc_cname_DI_Other Europe [ 0.0 ]
final_cc_cname_DI_Other Middle East/Central Asia [ 0.0 ]
final_cc_cname_DI_Other North & Central Amer., Caribbean [ 0.0 ]
final_cc_cname_DI_Other Oceania [ 0.0 ]
final_cc_cname_DI_Other South America [ 0.0 ]
final_cc_cname_DI_Other South Asia [ 0.0 ]
final cc cname DI Pakistan [ 0.0 ]
final_cc_cname_DI_Philippines [ 0.0 ]
final_cc_cname_DI_Poland [ 0.0 ]
final_cc_cname_DI_Portugal [ 0.0 ]
final_cc_cname_DI_Russian Federation [ 0.0 ]
final_cc_cname_DI_Spain [ 0.0 ]
final_cc_cname_DI_Ukraine [ 0.0 ]
final_cc_cname_DI_United Kingdom [ 0.0 ]
final_cc_cname_DI_United States [ 0.0 ]
final_cc_cname_DI_Unknown/Other [ 0.0 ]
LoE_DI_0 [ 0.0 ]
LoE DI Bachelor's [ 0.0 ]
LoE_DI_Doctorate [ 0.0 ]
```

```
LoE_DI_Less than Secondary [ 0.0 ]
LoE_DI_Master's [ 0.0 ]
LoE_DI_Secondary [ 0.0 ]
gender_0 [ 0.0 ]
gender_f [ 0.0 ]
gender_m [ 0.0 ]
gender_o [ 0.0 ]
```

In [32]:

```
dfAlg = ExibirMedidas(X train, X test, y train, y test)
dfAlg[colsExibir]
2019-04-17 15:39:20.079977
{'id_alg': 1, 'algoritmo': 'Tree'}
DecisionTreeClassifier(class weight=None, criterion='gini', max depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
2019-04-17 15:39:22.346298
{'id_alg': 2, 'algoritmo': 'KNN'}
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
2019-04-17 15:46:58.642884
{'id_alg': 3, 'algoritmo': 'SVM'}
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
2019-04-17 16:14:45.229892
{'id_alg': 4, 'algoritmo': 'MLP'}
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
2019-04-17 16:24:43.393790
{'id_alg': 5, 'algoritmo': 'Naive'}
GaussianNB(priors=None, var smoothing=1e-09)
2019-04-17 16:24:45.629701
{'id alg': 6, 'algoritmo': 'Dummy'}
DummyClassifier(constant=None, random_state=None, strategy='prior')
2019-04-17 16:24:45.851322
Out[32]:
```

	id_aig	aigoritmo	acur	sens	esp	епс	VPP	VPN	IEXec
0	1	Tree	97.20	97.20	0.00	48.60	100.00	0.00	0.04
1	2	KNN	96.81	97.97	40.37	69.17	98.76	29.03	7.60
2	3	SVM	97.20	97.20	0.00	48.60	100.00	0.00	27.78
3	4	MLP	97.20	97.20	0.00	48.60	100.00	0.00	9.97
4	5	Naive	65.73	99.40	6.66	53.03	65.14	86.30	0.04
5	6	Dummy	97.20	97.20	0.00	48.60	100.00	0.00	0.00

In [33]:

```
X_original = dfTrat.values
y_original = val_classes
```

In [34]:

```
#over_sampling
#https://imbalanced-learn.readthedocs.io/en/stable/generated/imblearn.over_sampling.SMO
TE.html

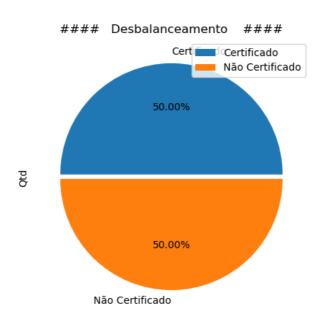
print('over_sampling')
print(datetime.datetime.now())
X_over, y_over = GerarResampling('over', X_original, y_original)
X_train, X_test, y_train, y_test = train_test_split(X_over, y_over, test_size=0.3, rand om_state=42)
print(datetime.datetime.now())
```

over_sampling
2019-04-17 17:07:07.183641
2019-04-17 17:07:17.443654

In [35]:

```
dfDesbOver = ExibirDesbanciamento(y_over)
dfDesbOver.plot.pie(x='Atr', y='Qtd', title='### Desbalanceamento ####', labels=d
fDesbOver['Atr'].values, explode=(0.03, 0.01),autopct = '%1.2f%%', figsize=(8, 5))
dfDesbOver
```

1.0



Out[35]:

	Atr	Qtd
0	Certificado	623451
1	Não Certificado	623451

In [36]:

```
dfAlgOver = ExibirMedidas(X over, X test, y over, y test)
dfAlgOver[colsExibir]
2019-04-17 17:07:52.324262
{'id_alg': 1, 'algoritmo': 'Tree'}
DecisionTreeClassifier(class weight=None, criterion='gini', max depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
2019-04-17 17:08:03.068270
{'id_alg': 2, 'algoritmo': 'KNN'}
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
2019-04-17 17:55:55.605951
{'id_alg': 3, 'algoritmo': 'SVM'}
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
2019-04-20 01:44:16.690193
{'id alg': 4, 'algoritmo': 'MLP'}
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
2019-04-20 02:30:59.781103
{'id_alg': 5, 'algoritmo': 'Naive'}
GaussianNB(priors=None, var smoothing=1e-09)
2019-04-20 02:31:04.307316
{'id alg': 6, 'algoritmo': 'Dummy'}
DummyClassifier(constant=None, random_state=None, strategy='prior')
2019-04-20 02:31:04.779725
Out[36]:
```

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	89.51	89.47	89.54	89.51	89.57	89.44	0.18
1	2	KNN	96.78	96.56	97.00	96.78	97.02	96.54	47.88
2	3	SVM	49.96	0.00	49.96	24.98	0.00	100.00	3348.35
3	4	MLP	90.63	94.52	87.35	90.94	86.27	94.99	46.72
4	5	Naive	76.60	88.63	70.26	79.45	61.06	92.15	0.08
5	6	Dummy	50.04	50.04	0.00	25.02	100.00	0.00	0.01

In [37]:

```
#from imblearn.under_sampling import RandomUnderSampler
#randCnn = RandomUnderSampler(random_state=42)
```

In [38]:

```
#under_sampling
#https://imbalanced-learn.readthedocs.io/en/stable/generated/imblearn.under_sampling.Ne
ighbourhoodCleaningRule.html

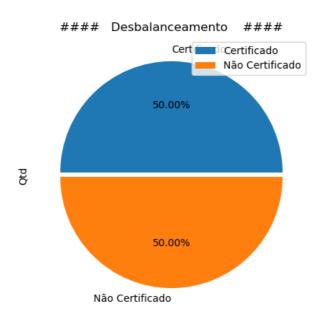
print('under_sampling')
print(datetime.datetime.now())
X_under, y_under = GerarResampling('under', X_original, y_original) #original
#X_under, y_under = randCnn.fit_resample(X_original, y_original)
X_train, X_test, y_train, y_test = train_test_split(X_under, y_under, test_size=0.3, ra
ndom_state=42)
print(datetime.datetime.now())
```

under_sampling 2019-04-20 10:13:37.189287 2019-04-20 10:13:38.198081

In [39]:

```
dfDesbUnder = ExibirDesbanciamento(y_under)
dfDesbUnder.plot.pie(x='Atr', y='Qtd', title='#### Desbalanceamento ####', labels=
dfDesbUnder['Atr'].values, explode=(0.03, 0.01),autopct = '%1.2f%%', figsize=(8, 5))
dfDesbUnder
```

1.0



Out[39]:

	Atr	Qtd
0	Certificado	17687
1	Não Certificado	17687

In [40]:

```
dfAlgUnder = ExibirMedidas(X under, X test, y under, y test)
dfAlgUnder[colsExibir]
2019-04-20 10:13:46.305790
{'id_alg': 1, 'algoritmo': 'Tree'}
DecisionTreeClassifier(class weight=None, criterion='gini', max depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
2019-04-20 10:13:46.464221
{'id_alg': 2, 'algoritmo': 'KNN'}
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
2019-04-20 10:13:53.853508
{'id_alg': 3, 'algoritmo': 'SVM'}
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
2019-04-20 10:16:27.248606
{'id alg': 4, 'algoritmo': 'MLP'}
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
2019-04-20 10:16:31.434311
{'id_alg': 5, 'algoritmo': 'Naive'}
GaussianNB(priors=None, var smoothing=1e-09)
2019-04-20 10:16:31.554682
{'id alg': 6, 'algoritmo': 'Dummy'}
DummyClassifier(constant=None, random_state=None, strategy='prior')
2019-04-20 10:16:31.568692
Out[40]:
                                                 VPN TExec
   id alg algoritmo
                                     ofic
                                           VPP
                  acur
                       sens
                               esn
```

	iu_aig	aigoritiilo	acui	36113	СЭР	CIIC	V 11	V1 14	ILXCC
0	1	Tree	88.53	88.65	88.42	88.53	88.37	88.70	0.00
1	2	KNN	93.69	94.17	93.22	93.69	93.14	94.24	0.12
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	2.56
3	4	MLP	65.84	99.53	59.45	79.49	31.79	99.85	0.07
4	5	Naive	75.94	87.88	69.74	78.81	60.15	91.71	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	0.00

In [41]:

```
#metr_alg = ['sens', 'esp', 'acur', 'VPP', 'VPN', 'efic', 'totP', 'totN', 'totG']
#cols = ['id_alg', 'algoritmo', 'acur', 'sens', 'esp', 'efic', 'VPP', 'VPN', 'mat_con
f', 'AlgBin', 'TExec']
dfAlgUnder2 = ReExibirMedidas(dfAlgUnder, X_test, y_test)
dfAlgUnder2[colsExibir]
ReExibirMedidas
Tree
2019-04-20 10:35:44.445198
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=5,
            max features=None, max leaf nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=Non
e,
            splitter='best')
KNN
2019-04-20 10:35:44.465225
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
SVM
2019-04-20 10:35:51.003182
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
MLP
2019-04-20 10:36:48.670322
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n iter no change=10, nesterovs momentum=True, power t=0.5,
       random state=None, shuffle=True, solver='adam', tol=0.0001,
       validation fraction=0.1, verbose=False, warm start=False)
Naive
2019-04-20 10:36:48.714450
GaussianNB(priors=None, var smoothing=1e-09)
2019-04-20 10:36:48.777620
DummyClassifier(constant=None, random_state=None, strategy='prior')
Out[41]:
```

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	88.53	88.65	88.42	88.53	88.37	88.70	0.00
1	2	KNN	93.69	94.17	93.22	93.69	93.14	94.24	0.11
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	0.96
3	4	MLP	65.84	99.53	59.45	79.49	31.79	99.85	0.00
4	5	Naive	75.94	87.88	69.74	78.81	60.15	91.71	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	0.00

In [42]:

```
dfAlgOver2[colsExibir]
ReExibirMedidas
Tree
2019-04-20 11:50:24.623913
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=5,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min weight fraction leaf=0.0, presort=False, random state=Non
e,
            splitter='best')
KNN
2019-04-20 11:50:24.644495
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
SVM
2019-04-20 11:50:47.275784
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='auto_deprecated',
  kernel='sigmoid', max_iter=-1, probability=False, random_state=None,
  shrinking=True, tol=0.001, verbose=False)
2019-04-20 12:25:42.203951
MLPClassifier(activation='relu', alpha=0.0001, batch_size='auto', beta_1=
0.9,
       beta_2=0.999, early_stopping=False, epsilon=1e-08,
       hidden_layer_sizes=100, learning_rate='constant',
       learning_rate_init=0.001, max_iter=200, momentum=0.9,
       n_iter_no_change=10, nesterovs_momentum=True, power_t=0.5,
       random_state=None, shuffle=True, solver='adam', tol=0.0001,
       validation_fraction=0.1, verbose=False, warm_start=False)
Naive
```

dfAlgOver2 = ReExibirMedidas(dfAlgOver, X test, y test)

Out[42]:

Dummy

2019-04-20 12:25:42.366897

2019-04-20 12:25:42.430069

GaussianNB(priors=None, var smoothing=1e-09)

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	TExec
0	1	Tree	88.02	87.68	88.37	88.03	88.46	87.59	0.00
1	2	KNN	93.19	90.46	96.31	93.38	96.55	89.83	0.38
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	34.92
3	4	MLP	89.88	93.14	87.08	90.11	86.08	93.67	0.00
4	5	Naive	74.72	83.47	69.61	76.54	61.61	87.82	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	0.00

DummyClassifier(constant=None, random_state=None, strategy='prior')

In [43]:

dfAlg

Out[43]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	mat_conf	
0	1	Tree	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	DecisionTreeClassific
1	2	KNN	96.81	97.97	40.37	69.17	98.76	29.03	[[184644, 2310], [3824, 1564]]	KNeighborsClas
2	3	SVM	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	SVC(C:
3	4	MLP	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	MLPClas
4	5	Naive	65.73	99.40	6.66	53.03	65.14	86.30	[[121780, 65174], [738, 4650]]	Gau ,
5	6	Dummy	97.20	97.20	0.00	48.60	100.00	0.00	[[186954, 0], [5388, 0]]	DummyCla
4										>

In [44]:

dfAlgUnder

Out[44]:

										A
	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	mat_conf	
0	1	Tree	88.53	88.65	88.42	88.53	88.37	88.70	[[4686, 617], [600, 4710]]	DecisionTreeCla:
1	2	KNN	93.69	94.17	93.22	93.69	93.14	94.24	[[4939, 364], [306, 5004]]	KNeighbors
2	3	SVM	50.03	0.00	50.03	25.02	0.00	100.00	[[0, 5303], [0, 5310]]	SV
3	4	MLP	65.84	99.53	59.45	79.49	31.79	99.85	[[1686, 3617], [8, 5302]]	MLF
4	5	Naive	75.94	87.88	69.74	78.81	60.15	91.71	[[3190, 2113], [440, 4870]]	
5	6	Dummy	49.97	49.97	0.00	24.98	100.00	0.00	[[5303, 0], [5310, 0]]	Dumm' •
4										+

In [45]:

dfAlgOver

Out[45]:

	id_alg	algoritmo	acur	sens	esp	efic	VPP	VPN	mat_conf	
0	1	Tree	89.51	89.47	89.54	89.51	89.57	89.44	[[167669, 19521], [19726, 167155]]	DecisionTreeClassif
1	2	KNN	96.78	96.56	97.00	96.78	97.02	96.54	[[181607, 5583], [6463, 180418]]	KNeighborsCla
2	3	SVM	49.96	0.00	49.96	24.98	0.00	100.00	[[0, 187190], [0, 186881]]	SVC(C
3	4	MLP	90.63	94.52	87.35	90.94	86.27	94.99	[[161493, 25697], [9363, 177518]]	MLPCk
4	5	Naive	76.60	88.63	70.26	79.45	61.06	92.15	[[114303, 72887], [14661, 172220]]	Ga
5	6	Dummy	50.04	50.04	0.00	25.02	100.00	0.00	[[187190, 0], [186881, 0]]	DummyCla
4										•

In [46]:

print('Original')
dfAlg[colsExibirMin]

Original

Out[46]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	97.20	97.20	0.00	48.60	0.04
1	2	KNN	96.81	97.97	40.37	69.17	7.60
2	3	SVM	97.20	97.20	0.00	48.60	27.78
3	4	MLP	97.20	97.20	0.00	48.60	9.97
4	5	Naive	65.73	99.40	6.66	53.03	0.04
5	6	Dummy	97.20	97.20	0.00	48.60	0.00

In [47]:

```
print('Oversample')
dfAlgOver[colsExibirMin]
```

Oversample

Out[47]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	89.51	89.47	89.54	89.51	0.18
1	2	KNN	96.78	96.56	97.00	96.78	47.88
2	3	SVM	49.96	0.00	49.96	24.98	3348.35
3	4	MLP	90.63	94.52	87.35	90.94	46.72
4	5	Naive	76.60	88.63	70.26	79.45	0.08
5	6	Dummy	50.04	50.04	0.00	25.02	0.01

In [48]:

```
print('Oversample - Teste com dados originais')
dfAlgOver2[colsExibirMin]
```

Oversample - Teste com dados originais

Out[48]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	88.02	87.68	88.37	88.03	0.00
1	2	KNN	93.19	90.46	96.31	93.38	0.38
2	3	SVM	50.03	0.00	50.03	25.02	34.92
3	4	MLP	89.88	93.14	87.08	90.11	0.00
4	5	Naive	74.72	83.47	69.61	76.54	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	0.00

```
In [49]:
```

```
print('Undersample')
dfAlgUnder[colsExibirMin]
```

Undersample

Out[49]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	88.53	88.65	88.42	88.53	0.00
1	2	KNN	93.69	94.17	93.22	93.69	0.12
2	3	SVM	50.03	0.00	50.03	25.02	2.56
3	4	MLP	65.84	99.53	59.45	79.49	0.07
4	5	Naive	75.94	87.88	69.74	78.81	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	0.00

In [50]:

```
print('Undersample - Teste com dados originais')
dfAlgUnder2[colsExibirMin]
```

Undersample - Teste com dados originais

Out[50]:

	id_alg	algoritmo	acur	sens	esp	efic	TExec
0	1	Tree	88.53	88.65	88.42	88.53	0.00
1	2	KNN	93.69	94.17	93.22	93.69	0.11
2	3	SVM	50.03	0.00	50.03	25.02	0.96
3	4	MLP	65.84	99.53	59.45	79.49	0.00
4	5	Naive	75.94	87.88	69.74	78.81	0.00
5	6	Dummy	49.97	49.97	0.00	24.98	0.00

In [51]:

```
dfAlgUnder2.T.to_dict()[0]['algoritmo']
```

Out[51]:

'Tree'

In [52]:

```
dill.dump_session('notebook_env.db')
```

In []: