Merge TXT files and Convert to CSV

Automatically merge txt files based on column position and convert to CSV

In [1]:

```python
import os
import pandas as pd
```

In [2]:

```python
# Test to explore file
#txt_file = 'DesmatamentoMunicipios2000.txt'
#encoding_type = 'ISO-8859-1'

#dataset=pd.read_csv(txt_file,encoding=encoding_type)
#dataset.head()
```

In [3]:

```python
# Set the path to the files that you want to process (or use the current path)

path = os.getcwd()
#path = "C:\\Users\\Ricardo\\Desktop\\dados"
files = os.listdir(path)

# Set the encoding type
encoding_type = 'ISO-8859-1'
```

In [4]:

```python
path
```

Out[4]:

```
'C:\\Users\\Ricardo\\Desktop\\dados'
```

In [5]:

```
files_txt = [f for f in files if f[-3:] == 'txt']

files_txt
```

Out[5]:

```
['DesmatamentoMunicipios2000.txt',
 'DesmatamentoMunicipios2001.txt',
 'DesmatamentoMunicipios2002.txt',
 'DesmatamentoMunicipios2003.txt',
 'DesmatamentoMunicipios2004.txt',
 'DesmatamentoMunicipios2005.txt',
 'DesmatamentoMunicipios2006.txt',
 'DesmatamentoMunicipios2007.txt',
 'DesmatamentoMunicipios2008.txt',
 'DesmatamentoMunicipios2009.txt',
 'DesmatamentoMunicipios2010.txt',
 'DesmatamentoMunicipios2011.txt',
 'DesmatamentoMunicipios2012.txt',
 'DesmatamentoMunicipios2013.txt',
 'DesmatamentoMunicipios2014.txt',
 'DesmatamentoMunicipios2015.txt',
 'DesmatamentoMunicipios2016.txt',
 'DesmatamentoMunicipios2017.txt',
 'DesmatamentoMunicipios2018.txt']
```

In [6]:

```
all_data = []
for f in files_txt:
    info = pd.read_csv(f,encoding=encoding_type,header=None,skiprows=1)
    info['File'] = str(f[-8:-4])
    all_data.append(info)
```

In [7]:

```
df = pd.concat(all_data, ignore_index=True, sort=False)
df.columns =['Nr','Lat','Long','Latgms','Longms','Municipio','CodIbge','Estado',\
             'AreaKm2','DesmatadoAno','IncrementoAnoAnt','FlorestaAno','NuvemAno',\
             'NaoObservadoAno','NaoFlorestaAno','HidrografiaAno','Soma','Ano']
```

In [8]:

```
df.columns
```

Out[8]:

```
Index(['Nr', 'Lat', 'Long', 'Latgms', 'Longms', 'Municipio', 'CodIbge',
       'Estado', 'AreaKm2', 'DesmatadoAno', 'IncrementoAnoAnt', 'Florest
aAno',
       'NuvemAno', 'NaoObservadoAno', 'NaoFlorestaAno', 'HidrografiaAn
o',
       'Soma', 'Ano'],
      dtype='object')
```

```
In [9]:
```

```
df.shape
```

```
Out[9]:
```

```
(14440, 18)
```

```
In [10]:
```

```
df.head()
```

```
Out[10]:
```

	Nr	Lat	Long	Latgms	Longms	Municipio	CodIbge	Estado	AreaKm2	Des
0	1	-3.12356	-47.40354	S 3 7 24.82	O 47 24 12.74	Paragominas	1505502	PA	19465	
1	2	-7.46114	-52.46225	S 7 27 40.10	O 52 27 44.10	São Félix do Xingu	1507300	PA	84253	
2	3	-5.52503	-49.94418	S 5 31 30.11	O 49 56 39.05	Marabá	1504208	PA	15161	
3	4	-11.22873	-57.64737	S 11 13 43.43	O 57 38 50.53	Juara	5105101	MT	22666	
4	5	-4.29175	-45.93682	S 4 17 30.30	O 45 56 12.55	Santa Luzia	2110005	MA	6245	

```
In [11]:
```

```
df.to_csv("dados-consolidados.csv", index = False, sep=';', encoding='utf-8')
```

```
In [ ]:
```