

Notebook para concatenar as 27 tabelas de devedores da PFN, uma para cada UF. Total de mais de 20 milhões de linhas.

```
In [1]: import pandas as pd
```

```
In [2]: acre = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_AC_20210
```

```
In [3]: acre
```

```
Out[3]:
```

	CPF_CNPJ	TIPO_PESSOA	TIPO_DEVEDOR	NOME_DEVEDOR	UF_UNIDADE_RESPONSAVEL
0	XXX631.543XX	Pessoa física	PRINCIPAL	VICENTE ABREU NETO	/
1	22.138.364/0001-75	Pessoa jurídica	PRINCIPAL	FARIAS E FARIAS CONSULTORIO MEDICO LTDA	/
2	XXX754.642XX	Pessoa física	PRINCIPAL	CIRENE SOUTO NUNES	/
3	XXX236.462XX	Pessoa física	PRINCIPAL	RUI EMANUEL RODRIGUES ARRUDA	/
4	34.710.061/0001-64	Pessoa jurídica	PRINCIPAL	MADEIREIRA JANEL IND E COM LTDA	/
...
50993	XXX691.992XX	Pessoa física	CORRESPONSAVEL	MARIA DALVA GOMES DE OLIVEIRA	/
50994	00.804.444/0001-80	Pessoa jurídica	PRINCIPAL	MERCADINHO CELEIRO LTDA	/
50995	01.630.624/0001-55	Pessoa jurídica	PRINCIPAL	SUSY PASCOAL NOGUEIRA	/
50996	XXX633.352XX	Pessoa física	CORRESPONSAVEL	MANOEL BRAZ SOUZA CORREIA	/
50997	XXX873.982XX	Pessoa física	CORRESPONSAVEL	JOSE BRANCO DA COSTA	/

50998 rows × 6 columns

```
In [4]: alagoas = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_AL_20
```

```
In [5]: amazonas = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_AM_2
```

```
In [6]: amapa = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_AP_2021
```

```
In [7]: bahia = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_BA_2021
```

```
In [8]: ceara = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_CE_2021
```

```
In [9]: distritofederal = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SI
```

```
In [10]: espiritosanto = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA
In [11]: goias = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_GO_2021
In [12]: maranhao = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_MA_2
In [13]: minasgerais = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_M
In [14]: matogrossodosul = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SI
In [15]: matogrosso = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_MT
In [16]: matogrossodosul = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SI
In [17]: para = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_PA_20210
In [18]: paraiba = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_PB_20
In [19]: pernambuco = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_PE
In [20]: piaui = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_PI_2021
In [21]: parana = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_PR_202
In [22]: riodejaneiro = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_
In [23]: riograndedonorte = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_S
In [24]: rondonia = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_RO_2
In [25]: roraima = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_RR_20
In [26]: riograndedosul = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SID
In [27]: santacatarina = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA
In [28]: sergipe = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_SE_20
In [29]: saopaulo = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_SP_2
In [30]: tocantins = pd.read_csv(r'/home/73594253368/Documentos/PFN/arquivo_lai_SIDA_TO_
```

1) Contatenamos todos os 27, exportando para o concatenado.csv, de 20.090.612 registros

```
In [31]: concatenado = pd.concat([acre,alagoas,amazonas,amapa,bahia,ceara,distritofedera
```

```
In [32]: concatenado.to_csv('/home/73594253368/Documentos/PEN/concatenado.csv')
```

```
In [41]: concatenado.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 20090612 entries, 0 to 127506
Data columns (total 13 columns):
#   Column                                Dtype
---  ---
0   CPF_CNPJ                             object
1   TIPO_PESSOA                           object
2   TIPO_DEVEDOR                           object
3   NOME_DEVEDOR                           object
4   UF_UNIDADE_RESPONSAVEL                 object
5   UNIDADE_RESPONSAVEL                     object
6   NUMERO_INSCRICAO                       int64
7   TIPO_SITUACAO_INSCRICAO                 object
8   SITUACAO_INSCRICAO                     object
9   RECEITA_PRINCIPAL                       object
10  DATA_INSCRICAO                         object
11  INDICADOR_AJUIZADO                       object
12  VALOR_CONSOLIDADO                       float64
dtypes: float64(1), int64(1), object(11)
memory usage: 2.1+ GB
```

```
In [35]: concatenado.head()
```

Out[35]:

	CPF_CNPJ	TIPO_PESSOA	TIPO_DEVEDOR	NOME_DEVEDOR	UF_UNIDADE_RESPONSAVEL	UNI
0	XXX631.543XX	Pessoa física	PRINCIPAL	VICENTE ABREU NETO		AC
1	22.138.364/0001-75	Pessoa jurídica	PRINCIPAL	FARIAS E FARIAS CONSULTORIO MEDICO LTDA		AC
2	XXX754.642XX	Pessoa física	PRINCIPAL	CIRENE SOUTO NUNES		AC
3	XXX236.462XX	Pessoa física	PRINCIPAL	RUI EMANUEL RODRIGUES ARRUDA		AC
4	34.710.061/0001-64	Pessoa jurídica	PRINCIPAL	MADEIREIRA JANEL IND E COM LTDA		AC

```
In [50]: shape = concatenado.shape
print('\nDataFrame Shape :', shape)
print('\nNúmero de linhas :', shape[0])
print('\nNúmero de colunas :', shape[1])
```

DataFrame Shape : (20090612, 13)

Número de linhas : 20090612

Número de colunas : 13

2) Ordenamos em ordem alfabética das UF, exportando para o concatenadoordenado.csv

```
In [36]: concatenadoordenado = pd.concat([acre,alagoas,amazonas,amapa,bahia,ceara,distri
```

In [39]: `concatenadoordenado.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 20090612 entries, 0 to 127506
Data columns (total 13 columns):
#   Column                                Dtype
---  -
0   CPF_CNPJ                             object
1   TIPO_PESSOA                           object
2   TIPO_DEVEDOR                           object
3   NOME_DEVEDOR                           object
4   UF_UNIDADE_RESPONSAVEL                 object
5   UNIDADE_RESPONSAVEL                     object
6   NUMERO_INSCRICAO                       int64
7   TIPO_SITUACAO_INSCRICAO                 object
8   SITUACAO_INSCRICAO                     object
9   RECEITA_PRINCIPAL                       object
10  DATA_INSCRICAO                         object
11  INDICADOR_AJUIZADO                       object
12  VALOR_CONSOLIDADO                       float64
dtypes: float64(1), int64(1), object(11)
memory usage: 2.1+ GB
```

In [42]: `concatenadoordenado.to_csv('/home/73594253368/Documentos/PEN/concatenadoordenad`