

Fecomércio Sesc

Data Science – Princípios e Técnicas

Setembro

2024



Onde me encontrar:

https://www.linkedin.com/in/marco-mialaret-junior/

e

https://github.com/MatmJr





Mas o Pysus dá muito problema



Para contornar esse problema vamos extrair os dados da API e converter em um CSV.



Instalando o pysus:

Abra um notebook no colab. Na primeira célula, execute:

!pip install pysus



A base de dados SIH:

from pysus.ftp.databases.sih import SIH import pandas as pd

sih = SIH().load()
sih.metadata['description']



sih.groups

```
{'RD': 'AIH Reduzida',
  'RJ': 'AIH Rejeitada',
  'ER': 'AIH Rejeitada com erro',
  'SP': 'Serviços Profissionais',
  'CH': 'Cadastro Hospitalar',
  'CM': ''}
```



files = sih.get_files("RD", uf="PE", year=2024, month=[1,2,3]) data = sih.download(files)



import pandas as pd

```
df_list = []
for i in range(len(data)):
    df = data[i].to_dataframe()
    df_list.append(df)
```

df_final = pd.concat(df_list, ignore_index=True)



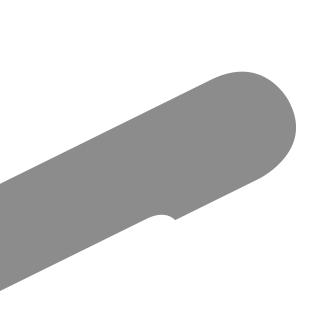


df_final

	UF_ZI	ANO_CMPT	MES_CMPT	ESPEC	CGC_HOSP	N_AIH	IDENT	CEP	MUNIC_RES	NASC	•••	DIAGSEC9	TPDISEC1	TPDISEC2	TPDISEC3	TPDISEC4	TPDISEC5	TPDISEC6	TPDISEC7	TPDISEC8	TPDISEC9
0	260000	2024	01	01	10667814000138	2624100247330	1	55845000	260270	19830208			0	0	0	0	0	0	0	0	0
1	260000	2024	01	01	10667814000138	2624100247340	1	54420241	260790	19880930			0	0	0	0	0	0	0	0	0
2	260000	2024	01	01	10667814000138	2624100247351	1	64040085	221100	19610615			0	0	0	0	0	0	0	0	0
3	260000	2024	01	02	10572048001523	2624105067716	1	55400000	260420	20001217			2	0	0	0	0	0	0	0	0
4	260000	2024	01	02	10572048001523	2624105067727	1	55170000	260260	19901031			2	0	0	0	0	0	0	0	0
	1	202	707	1		100	922	22	123	2.07	222	1222	0.11	023	5201		2.2	1.0	575	7800	2020
153929	261640	2024	03	03	11866365000110	2624109213418	1	55608835	261640	19550215			0	0	0	0	0	0	0	0	0
153930	261640	2024	03	03	11866365000110	2624109213430	1	55604070	261640	19441021			0	0	0	0	0	0	0	0	0
153931	261640	2024	03	01	11866365000110	2624109212274	1	55602010	261640	19900730			0	0	0	0	0	0	0	0	0
153932	261640	2024	03	01	11866365000110	2624109212285	1	55660000	260190	19831208	2050		0	0	0	0	0	0	0	0	0
153933	261640	2024	03	01	11866365000110	2624109212307	1	55610350	261640	19611226			0	0	0	0	0	0	0	0	0

153934 rows × 113 columns









Cada grupo deverá seguir esses passos vistos na aula de hoje.



Dúvidas?







Marco Mialaret, MSc

Telefone:

81 98160 7018

E-mail:

marcomialaret@gmail.com

