

Mapa das escolas do RS por taxa de distorção de série

Willer Gomes Junior

```
In [18]: #!/pip3 install folium
```

```
In [19]: # Importar bibliotecas
import pandas as pd
import folium

# Configurando o matplotlib
%matplotlib inline

df = pd.read_csv('fee-2013-mun-taxa-de-distorcao-idade-serie-total-102524.csv', sep=',', encoding = "cp1252", skiprows=1)
df.head()
```

Out[19]:

	Município	ibge	latitude	longitude	/Educação/Ens...de Série/Total 2013 (-)
0	Aceguá	4300034	-31.86076	-54.16706	25,7
1	Água Santa	4300059	-28.16720	-52.03100	14,7
2	Agudo	4300109	-29.64470	-53.25150	26,4
3	Ajuricaba	4300208	-28.23420	-53.77570	24,1
4	Alecrim	4300307	-27.65790	-54.76490	18,9

```
In [20]: # renomeando coluna
df.rename(columns={'/Educação/Ens...de Série/Total 2013 (-)': 'tx_distorcao'}, inplace=True)
df.head()
```

Out[20]:

	Município	ibge	latitude	longitude	tx_distorcao
0	Aceguá	4300034	-31.86076	-54.16706	25,7
1	Água Santa	4300059	-28.16720	-52.03100	14,7
2	Agudo	4300109	-29.64470	-53.25150	26,4
3	Ajuricaba	4300208	-28.23420	-53.77570	24,1
4	Alecrim	4300307	-27.65790	-54.76490	18,9

```
In [21]: # verificar o dataframe
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 495 entries, 0 to 494
Data columns (total 5 columns):
Município      495 non-null object
ibge           495 non-null int64
latitude       495 non-null float64
longitude      495 non-null float64
tx_distorcao   495 non-null object
dtypes: float64(2), int64(1), object(2)
memory usage: 15.5+ KB
```

```
In [22]: # corrigir a tx_distorção, converter para float

# inicialmente trocar virgula por ponto
df['tx_distorcao'] = df['tx_distorcao'].str.replace(',', '.')
```

```
In [23]: df.head()
```

```
Out[23]:
```

	Município	ibge	latitude	longitude	tx_distorcao
0	Aceguá	4300034	-31.86076	-54.16706	25.7
1	Água Santa	4300059	-28.16720	-52.03100	14.7
2	Agudo	4300109	-29.64470	-53.25150	26.4
3	Ajuricaba	4300208	-28.23420	-53.77570	24.1
4	Alecrim	4300307	-27.65790	-54.76490	18.9

```
In [24]: # verificar o dataframe
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 495 entries, 0 to 494
Data columns (total 5 columns):
Município      495 non-null object
ibge           495 non-null int64
latitude       495 non-null float64
longitude      495 non-null float64
tx_distorcao   495 non-null object
dtypes: float64(2), int64(1), object(2)
memory usage: 15.5+ KB
```

```
In [25]: # Transtormando o tipo para float
df['tx_distorcao'] = df['tx_distorcao'].astype(float)
```

```
In [26]: # verificar o dataframe
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 495 entries, 0 to 494
Data columns (total 5 columns):
Município      495 non-null object
ibge           495 non-null int64
latitude       495 non-null float64
longitude      495 non-null float64
tx_distorcao   495 non-null float64
dtypes: float64(3), int64(1), object(1)
memory usage: 17.5+ KB
```

```
In [27]: # 10 municípios com menores tx_distorcao
df.nsmallest(10, 'tx_distorcao')
```

Out[27]:

	Município	ibge	latitude	longitude	tx_distorcao
490	Vista Alegre do Prata	4323606	-28.8052	-51.7946	2.6
375	Santo Antônio do Palma	4317558	-28.4956	-52.0267	4.3
28	Áurea	4301552	-27.6936	-52.0505	5.8
268	Nicolau Vergueiro	4312674	-28.5298	-52.4676	7.0
128	Derrubadas	4306320	-27.2642	-53.8645	7.4
119	Cotiporã	4305959	-28.9891	-51.6971	7.5
68	Campestre da Serra	4303673	-28.7926	-51.0941	7.6
476	Vanini	4322558	-28.4758	-51.8447	7.8
236	Linha Nova	4311643	-29.4679	-51.2003	8.0
392	São José do Inhacorá	4318499	-27.7251	-54.1275	8.0

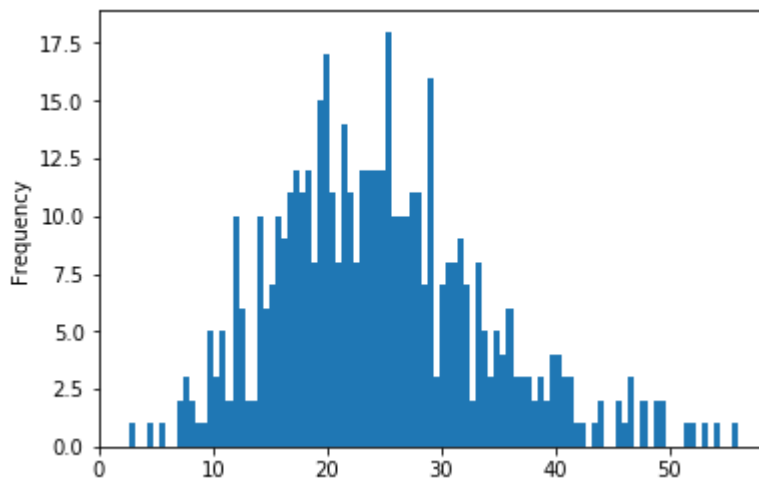
```
In [28]: # 10 municípios com maiores tx_distorcao
df.nlargest(10, 'tx_distorcao')
```

Out[28]:

	Município	ibge	latitude	longitude	tx_distorcao
307	Pedras Altas	4314175	-31.7217	-53.5838	56.0
316	Pinheiro Machado	4314506	-31.5794	-53.3798	53.9
17	Arambaré	4300851	-30.9093	-51.5046	52.8
140	Eldorado do Sul	4306767	-30.0847	-51.6187	51.9
35	Barra do Quaraí	4301875	-30.2029	-57.5497	51.5
393	São José do Norte	4318507	-32.0151	-52.0331	49.4
413	São Valério do Sul	4319737	-27.7906	-53.9368	49.2
193	Hulha Negra	4309654	-31.4067	-53.8667	49.0
106	Chuí	4305439	-33.6866	-53.4594	48.6
233	Lavras do Sul	4311502	-30.8071	-53.8931	47.7

```
In [29]: df['tx_distorcao'].plot.hist(bins=100)
```

```
Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0xaa53da6c>
```



```
In [30]: # Quantos municípios tem tx_distorcao <= 10
df[df['tx_distorcao'] <= 10].count()
```

```
Out[30]: Município      17
         ibge           17
         latitude       17
         longitude      17
         tx_distorcao    17
         dtype: int64
```

```
In [31]: # Quantos municípios tem tx_distorcao >= 10
df[df['tx_distorcao'] >= 45].count()
```

```
Out[31]: Município      17
         ibge           17
         latitude       17
         longitude      17
         tx_distorcao    17
         dtype: int64
```

```
In [32]: # Mapa do Brasil
brasil = folium.Map(location=[-13.6603615, -69.6775883], #coordenadas Google
                    zoom_start=4
                    )
brasil
```

Out[32]:



```
In [33]: # Localizando o RS
RioSul = folium.Map(location=[-29.8627852, -53.1200958], #coordenadas Google maps
                    zoom_start=6
                    )
RioSul # -29.8627852, -53.1200958, 7z
```

Out[33]:



```
In [34]: # Percorrendo o dataframe com base nas melhores escolas[tx_distorcao <= 10]
# desenhar o marcador para cada escola usando o for e iterrows
```

```
for indice, municipio in df[df['tx_distorcao'] <= 10].iterrows():
    folium.Marker(
        location=[municipio['latitude'], municipio['longitude']],
        popup=municipio['Município'],
        icon=folium.map.Icon(color='green')
    ).add_to(RioSul)
```

```
# Percorrendo o dataframe com base nas piores escolas[tx_distorcao >= 45]
# desenhar o marcador para cada escola usando o for e iterrows
```

```
for indice, municipio in df[df['tx_distorcao'] >= 45].iterrows():
    folium.Marker(
        location=[municipio['latitude'], municipio['longitude']],
        popup=municipio['Município'],
        icon=folium.map.Icon(color='red')
    ).add_to(RioSul)
```

RioSul

Out[34]:



In []: