

Cargas para tbl_cidade

In [1]:



```
import pandas as pd  
import numpy as np
```

In [2]:



```
tbl_estados = pd.read_csv('tbl_estados.csv', sep=',')
```

In [3]:

```
tbl_estados
```

Out[3]:

| | id_estado | estado | uf |
|----|-----------|---------------------|----|
| 0 | 1 | Acre | AC |
| 1 | 2 | Alagoas | AL |
| 2 | 3 | Amapá | AP |
| 3 | 4 | Amazonas | AM |
| 4 | 5 | Bahia | BA |
| 5 | 6 | Ceará | CE |
| 6 | 7 | Distrito Federal | DF |
| 7 | 8 | Espírito Santo | ES |
| 8 | 9 | Goiás | GO |
| 9 | 10 | Maranhão | MA |
| 10 | 11 | Mato Grosso | MT |
| 11 | 12 | Mato Grosso do Sul | MS |
| 12 | 13 | Minas Gerais | MG |
| 13 | 14 | Pará | PA |
| 14 | 15 | Paraíba | PB |
| 15 | 16 | Paraná | PR |
| 16 | 17 | Pernambuco | PE |
| 17 | 18 | Piauí | PI |
| 18 | 19 | Rio de Janeiro | RJ |
| 19 | 20 | Rio Grande do Norte | RN |
| 20 | 21 | Rio Grande do Sul | RS |
| 21 | 22 | Rondônia | RO |
| 22 | 23 | Roraima | RR |
| 23 | 24 | Santa Catarina | SC |
| 24 | 25 | São Paulo | SP |
| 25 | 26 | Sergipe | SE |
| 26 | 27 | Tocantins | TO |

In [4]:

```
tbl_ceps = pd.read_table('ceps.txt')
```

In [5]:



tbl_ceps.head(20)

Out[5]:

| | Cep | cidades/uf | Bairro | Logradouro | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 |
|----|---------|--------------|--------|---------------------------------|--|------------|------------|------------|
| 0 | 1001000 | São Paulo/SP | Sé | Praça da Sé - lado ímpar | NaN | NaN | NaN | NaN |
| 1 | 1001001 | São Paulo/SP | Sé | Praça da Sé - lado par | NaN | NaN | NaN | NaN |
| 2 | 1001010 | São Paulo/SP | Sé | Rua Filipe de Oliveira | NaN | NaN | NaN | NaN |
| 3 | 1001900 | São Paulo/SP | Sé | Praça da Sé, 108 | UNESP - Universidade Estadual Júlio de Mesqui... | NaN | NaN | NaN |
| 4 | 1001901 | São Paulo/SP | Sé | Praça da Sé, 371 | Edifício Santa Lídia | NaN | NaN | NaN |
| 5 | 1001902 | São Paulo/SP | Sé | Praça da Sé, 385 | OAB - Ordem dos Advogados do Brasil. | NaN | NaN | NaN |
| 6 | 1002000 | São Paulo/SP | Sé | Rua Direita - lado par | NaN | NaN | NaN | NaN |
| 7 | 1002001 | São Paulo/SP | Sé | Rua Direita - lado ímpar | NaN | NaN | NaN | NaN |
| 8 | 1002010 | São Paulo/SP | Sé | Praça do Patriarca | NaN | NaN | NaN | NaN |
| 9 | 1002020 | São Paulo/SP | Centro | Viaduto do Chá | NaN | NaN | NaN | NaN |
| 10 | 1002900 | São Paulo/SP | Centro | Viaduto do Chá, 15 | Prefeitura do Município de São Paulo | NaN | NaN | NaN |
| 11 | 1002901 | São Paulo/SP | Sé | Rua Direita, 32 | Edifício Piratininga | NaN | NaN | NaN |
| 12 | 1002902 | São Paulo/SP | Sé | Rua Direita, 191 | Edifício Zogbi | NaN | NaN | NaN |
| 13 | 1002903 | São Paulo/SP | Sé | Rua Direita, 250 | Edifício Barão de Iguape | NaN | NaN | NaN |
| 14 | 1003000 | São Paulo/SP | Sé | Rua José Bonifácio - lado par | NaN | NaN | NaN | NaN |
| 15 | 1003001 | São Paulo/SP | Sé | Rua José Bonifácio - lado ímpar | NaN | NaN | NaN | NaN |
| 16 | 1003010 | São Paulo/SP | Sé | Praça Ouvidor Pacheco e Silva | NaN | NaN | NaN | NaN |

| | Cep | cidades/uf | Bairro | Logradouro | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 |
|----|---------|--------------|--------|-------------------------|---------------------|------------|------------|------------|
| 17 | 1003900 | São Paulo/SP | Sé | Rua José Bonifácio, 24 | Edifício Triângulo | NaN | NaN | NaN |
| 18 | 1003901 | São Paulo/SP | Sé | Rua José Bonifácio, 93 | Edifício Santa Cruz | NaN | NaN | NaN |
| 19 | 1003902 | São Paulo/SP | Sé | Rua José Bonifácio, 209 | Edifício Brasília | NaN | NaN | NaN |

In [6]:

```
nova_tabela = tbl_ceps['cidades/uf']
```

In [7]:

```
nova_tabela.head(10)
```

Out[7]:

```
0    São Paulo/SP
1    São Paulo/SP
2    São Paulo/SP
3    São Paulo/SP
4    São Paulo/SP
5    São Paulo/SP
6    São Paulo/SP
7    São Paulo/SP
8    São Paulo/SP
9    São Paulo/SP
Name: cidades/uf, dtype: object
```

In [8]:



```
nova_tabela = nova_tabela.str.split("/")  
nova_tabela.head(50)
```

Out[8]:

```
0    [São Paulo, SP]  
1    [São Paulo, SP]  
2    [São Paulo, SP]  
3    [São Paulo, SP]  
4    [São Paulo, SP]  
5    [São Paulo, SP]  
6    [São Paulo, SP]  
7    [São Paulo, SP]  
8    [São Paulo, SP]  
9    [São Paulo, SP]  
10   [São Paulo, SP]  
11   [São Paulo, SP]  
12   [São Paulo, SP]  
13   [São Paulo, SP]  
14   [São Paulo, SP]  
15   [São Paulo, SP]  
16   [São Paulo, SP]  
17   [São Paulo, SP]  
18   [São Paulo, SP]  
19   [São Paulo, SP]  
20   [São Paulo, SP]  
21   [São Paulo, SP]  
22   [São Paulo, SP]  
23   [São Paulo, SP]  
24   [São Paulo, SP]  
25   [São Paulo, SP]  
26   [São Paulo, SP]  
27   [São Paulo, SP]  
28   [São Paulo, SP]  
29   [São Paulo, SP]  
30   [São Paulo, SP]  
31   [São Paulo, SP]  
32   [São Paulo, SP]  
33   [São Paulo, SP]  
34   [São Paulo, SP]  
35   [São Paulo, SP]  
36   [São Paulo, SP]  
37   [São Paulo, SP]  
38   [São Paulo, SP]  
39   [São Paulo, SP]  
40   [São Paulo, SP]  
41   [São Paulo, SP]  
42   [São Paulo, SP]  
43   [São Paulo, SP]  
44   [São Paulo, SP]  
45   [São Paulo, SP]  
46   [São Paulo, SP]  
47   [São Paulo, SP]  
48   [São Paulo, SP]  
49   [São Paulo, SP]
```

Name: cidades/uf, dtype: object

In [9]:

```
dic = {}
cidades = []
uf = []
for i in range(len(nova_tabela)):
    cidades.append(nova_tabela[i][0])
    uf.append(nova_tabela[i][1])

dic['Cidade'] = cidades
dic['Uf'] = uf

print(dic)
```

IOPub data rate exceeded.

The notebook server will temporarily stop sending output to the client in order to avoid crashing it.

To change this limit, set the config variable

`--NotebookApp.iopub_data_rate_limit`.

Current values:

NotebookApp.iopub_data_rate_limit=1000000.0 (bytes/sec)

NotebookApp.rate_limit_window=3.0 (secs)

In [10]:

```
df = pd.DataFrame(dic)
```

In [11]:

```
df.head(100)
```

Out[11]:

| | Cidade | Uf |
|-----|-----------|-----|
| 0 | São Paulo | SP |
| 1 | São Paulo | SP |
| 2 | São Paulo | SP |
| 3 | São Paulo | SP |
| 4 | São Paulo | SP |
| ... | ... | ... |
| 95 | São Paulo | SP |
| 96 | São Paulo | SP |
| 97 | São Paulo | SP |
| 98 | São Paulo | SP |
| 99 | São Paulo | SP |

100 rows × 2 columns

In [12]:

```
uf = df['Uf']
```

In [13]:

```
uf
```

Out[13]:

```
0          SP
1          SP
2          SP
3          SP
4          SP
```

```
...
732758    RS - Distrito
732759    RS - Distrito
732760         RS
732761         RS
732762         RS
```

```
Name: Uf, Length: 732763, dtype: object
```

In [14]:

```
df['Uf'].replace(to_replace = "RS - Distrito", value = "RS", inplace=True)
df['Uf'].replace(to_replace = "SP - Distrito", value = "SP", inplace=True)
df['Uf'].replace(to_replace = "AC - Distrito", value = "AC", inplace=True)
df['Uf'].replace(to_replace = "AL - Distrito", value = "AL", inplace=True)
df['Uf'].replace(to_replace = "AL - Povoado", value = "AL", inplace=True)
df['Uf'].replace(to_replace = "AM - Distrito", value = "AM", inplace=True)
df['Uf'].replace(to_replace = "AM - Povoado", value = "AM", inplace=True)
df['Uf'].replace(to_replace = "AP - Distrito", value = "AP", inplace=True)
df['Uf'].replace(to_replace = "AP - Povoado", value = "AP", inplace=True)
df['Uf'].replace(to_replace = "BA - Distrito", value = "BA", inplace=True)
df['Uf'].replace(to_replace = "BA - Povoado", value = "BA", inplace=True)
df['Uf'].replace(to_replace = "CE - Distrito", value = "CE", inplace=True)
df['Uf'].replace(to_replace = "CE - Povoado", value = "CE", inplace=True)
df['Uf'].replace(to_replace = "ES - Distrito", value = "ES", inplace=True)
df['Uf'].replace(to_replace = "ES - Povoado", value = "ES", inplace=True)
df['Uf'].replace(to_replace = "GO - Distrito", value = "GO", inplace=True)
df['Uf'].replace(to_replace = "GO - Povoado", value = "GO", inplace=True)
df['Uf'].replace(to_replace = "MA - Distrito", value = "MA", inplace=True)
df['Uf'].replace(to_replace = "MA - Povoado", value = "MA", inplace=True)
df['Uf'].replace(to_replace = "MG - Distrito", value = "MG", inplace=True)
df['Uf'].replace(to_replace = "MG - Povoado", value = "MG", inplace=True)
df['Uf'].replace(to_replace = "MS - Distrito", value = "MS", inplace=True)
df['Uf'].replace(to_replace = "MT - Distrito", value = "MT", inplace=True)
df['Uf'].replace(to_replace = "MT - Povoado", value = "MT", inplace=True)
df['Uf'].replace(to_replace = "PA - Distrito", value = "PA", inplace=True)
df['Uf'].replace(to_replace = "PA - Povoado", value = "PA", inplace=True)
df['Uf'].replace(to_replace = "PB - Distrito", value = "PB", inplace=True)
df['Uf'].replace(to_replace = "PB - Povoado", value = "PB", inplace=True)
df['Uf'].replace(to_replace = "PE - Distrito", value = "PE", inplace=True)
df['Uf'].replace(to_replace = "PE - Povoado", value = "PE", inplace=True)
df['Uf'].replace(to_replace = "PI - Povoado", value = "PI", inplace=True)
df['Uf'].replace(to_replace = "PR - Distrito", value = "PR", inplace=True)
df['Uf'].replace(to_replace = "PR - Povoado", value = "PR", inplace=True)
df['Uf'].replace(to_replace = "RJ - Distrito", value = "RJ", inplace=True)
df['Uf'].replace(to_replace = "RJ - Povoado", value = "RJ", inplace=True)
df['Uf'].replace(to_replace = "RN - Distrito", value = "RN", inplace=True)
df['Uf'].replace(to_replace = "RN - Povoado", value = "RN", inplace=True)
df['Uf'].replace(to_replace = "RO - Distrito", value = "RO", inplace=True)
df['Uf'].replace(to_replace = "RO - Povoado", value = "RO", inplace=True)
df['Uf'].replace(to_replace = "RR - Distrito", value = "RR", inplace=True)
df['Uf'].replace(to_replace = "RS - Povoado", value = "RS", inplace=True)
df['Uf'].replace(to_replace = "SC - Distrito", value = "SC", inplace=True)
df['Uf'].replace(to_replace = "SE - Distrito", value = "SE", inplace=True)
df['Uf'].replace(to_replace = "SP - Povoado", value = "SP", inplace=True)
df['Uf'].replace(to_replace = "TO - Distrito", value = "TO", inplace=True)
df['Uf'].replace(to_replace = "TO - Povoado", value = "TO", inplace=True)
```


In [15]:



```
df
```

Out[15]:

| | Cidade | Uf |
|--------|--------------------------|-----|
| 0 | São Paulo | SP |
| 1 | São Paulo | SP |
| 2 | São Paulo | SP |
| 3 | São Paulo | SP |
| 4 | São Paulo | SP |
| ... | ... | ... |
| 732758 | São João Bosco (Ciríaco) | RS |
| 732759 | Cruzaltinha (Ciríaco) | RS |
| 732760 | David Canabarro | RS |
| 732761 | David Canabarro | RS |
| 732762 | David Canabarro | RS |

732763 rows × 2 columns

In [16]:



```
df.notnull()
```

Out[16]:

| | Cidade | Uf |
|--------|--------|------|
| 0 | True | True |
| 1 | True | True |
| 2 | True | True |
| 3 | True | True |
| 4 | True | True |
| ... | ... | ... |
| 732758 | True | True |
| 732759 | True | True |
| 732760 | True | True |
| 732761 | True | True |
| 732762 | True | True |

732763 rows × 2 columns

In [17]:



```
tbl_estados_query = pd.DataFrame(tbl_estados['uf'])
```

In [18]:



```
tbl_estados_query
```

Out[18]:

| | uf |
|----|----|
| 0 | AC |
| 1 | AL |
| 2 | AP |
| 3 | AM |
| 4 | BA |
| 5 | CE |
| 6 | DF |
| 7 | ES |
| 8 | GO |
| 9 | MA |
| 10 | MT |
| 11 | MS |
| 12 | MG |
| 13 | PA |
| 14 | PB |
| 15 | PR |
| 16 | PE |
| 17 | PI |
| 18 | RJ |
| 19 | RN |
| 20 | RS |
| 21 | RO |
| 22 | RR |
| 23 | SC |
| 24 | SP |
| 25 | SE |
| 26 | TO |

In [19]:



```
tbl_estados_query['cod'] = range(1,28)
```

Type *Markdown* and LaTeX: α^2

In [20]:



```
tbl_estados_query['uf']
```

Out[20]:

```
0    AC
1    AL
2    AP
3    AM
4    BA
5    CE
6    DF
7    ES
8    GO
9    MA
10   MT
11   MS
12   MG
13   PA
14   PB
15   PR
16   PE
17   PI
18   RJ
19   RN
20   RS
21   RO
22   RR
23   SC
24   SP
25   SE
26   TO
```

Name: uf, dtype: object

In [21]:

```
df['Uf'].replace(to_replace = 'AC' ,value = 1 ,inplace=True)
df['Uf'].replace(to_replace = 'AL' ,value = 2 ,inplace=True)
df['Uf'].replace(to_replace = 'AP' ,value = 3 ,inplace=True)
df['Uf'].replace(to_replace = 'AM' ,value = 4 ,inplace=True)
df['Uf'].replace(to_replace = 'BA' ,value = 5 ,inplace=True)
df['Uf'].replace(to_replace = 'CE' ,value = 6 ,inplace=True)
df['Uf'].replace(to_replace = 'DF' ,value = 7 ,inplace=True)
df['Uf'].replace(to_replace = 'ES' ,value = 8 ,inplace=True)
df['Uf'].replace(to_replace = 'GO' ,value = 9 ,inplace=True)
df['Uf'].replace(to_replace = 'MA' ,value = 10 ,inplace=True)
df['Uf'].replace(to_replace = 'MT' ,value = 11 ,inplace=True)
df['Uf'].replace(to_replace = 'MS' ,value = 12 ,inplace=True)
df['Uf'].replace(to_replace = 'MG' ,value = 13 ,inplace=True)
df['Uf'].replace(to_replace = 'PA' ,value = 14 ,inplace=True)
df['Uf'].replace(to_replace = 'PB' ,value = 15 ,inplace=True)
df['Uf'].replace(to_replace = 'PR' ,value = 16 ,inplace=True)
df['Uf'].replace(to_replace = 'PE' ,value = 17 ,inplace=True)
df['Uf'].replace(to_replace = 'PI' ,value = 18 ,inplace=True)
df['Uf'].replace(to_replace = 'RJ' ,value = 19 ,inplace=True)
df['Uf'].replace(to_replace = 'RN' ,value = 20 ,inplace=True)
df['Uf'].replace(to_replace = 'RS' ,value = 21 ,inplace=True)
df['Uf'].replace(to_replace = 'RO' ,value = 22 ,inplace=True)
df['Uf'].replace(to_replace = 'RR' ,value = 23 ,inplace=True)
df['Uf'].replace(to_replace = 'SC' ,value = 24 ,inplace=True)
df['Uf'].replace(to_replace = 'SP' ,value = 25 ,inplace=True)
df['Uf'].replace(to_replace = 'SE' ,value = 26 ,inplace=True)
df['Uf'].replace(to_replace = 'TO' ,value = 27 ,inplace=True)
```

In [22]:

```
df
```

Out[22]:

| | Cidade | Uf |
|--------|--------------------------|-----|
| 0 | São Paulo | 25 |
| 1 | São Paulo | 25 |
| 2 | São Paulo | 25 |
| 3 | São Paulo | 25 |
| 4 | São Paulo | 25 |
| ... | ... | ... |
| 732758 | São João Bosco (Ciríaco) | 21 |
| 732759 | Cruzaltinha (Ciríaco) | 21 |
| 732760 | David Canabarro | 21 |
| 732761 | David Canabarro | 21 |
| 732762 | David Canabarro | 21 |

732763 rows × 2 columns

In [23]:

```
df.drop_duplicates(inplace=True)
df.index = range(df.shape[0])
df
```

Out[23]:

| | Cidade | Uf |
|-------|--------------------------|-----|
| 0 | São Paulo | 25 |
| 1 | Osasco | 25 |
| 2 | Carapicuíba | 25 |
| 3 | Barueri | 25 |
| 4 | Santana de Parnaíba | 25 |
| ... | ... | ... |
| 10438 | São Caetano (Água Santa) | 21 |
| 10439 | Ciríaco | 21 |
| 10440 | São João Bosco (Ciríaco) | 21 |
| 10441 | Cruzaltinha (Ciríaco) | 21 |
| 10442 | David Canabarro | 21 |

10443 rows × 2 columns

In [24]:

```
df
```

Out[24]:

| | Cidade | Uf |
|-------|--------------------------|-----|
| 0 | São Paulo | 25 |
| 1 | Osasco | 25 |
| 2 | Carapicuíba | 25 |
| 3 | Barueri | 25 |
| 4 | Santana de Parnaíba | 25 |
| ... | ... | ... |
| 10438 | São Caetano (Água Santa) | 21 |
| 10439 | Ciríaco | 21 |
| 10440 | São João Bosco (Ciríaco) | 21 |
| 10441 | Cruzaltinha (Ciríaco) | 21 |
| 10442 | David Canabarro | 21 |

10443 rows × 2 columns

In [25]:



```
df.to_csv('cidades_newload.csv', sep=',', index=False)
```

In [26]:



```
df.shape
```

Out[26]:

```
(10443, 2)
```

In []:

