

```
In [1]: # Importando biblioteca para tratamento dos dados:  
import pandas as pd
```

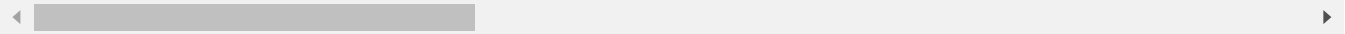
```
In [2]: # Importando o arquivo csv:  
url = ('C:/Users/User/Documents/projeto_fmu/full_data.csv')  
df = pd.read_csv(url, low_memory=False)
```

```
In [3]: # Imprimindo dataframe:  
df
```

Out[3]:

	Data de cadastro	Canal de atendimento	Denúncia emergencial	Denunciante	Cenário da violação	Município	Frequ
0	7/4/20 11:50	TELEFÔNICO	SITUAÇÃO FLAGRANTE - ATÉ 24H DA OCORRÊNCIA - S...	A PRÓPRIA VÍTIMA	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SARAPUÍ	Ú OCORRÉ
1	7/10/20 18:37	TELEFÔNICO	NÃO	ANÔNIMO	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SÃO PAULO	DIARIAM
2	7/9/20 9:26	TELEFÔNICO	NÃO	A PRÓPRIA VÍTIMA	CASA DA VÍTIMA	GUARULHOS	SEMANALM
3	7/13/20 14:15	TELEFÔNICO	NÃO	ANÔNIMO	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	DENUNCIANTE NÃO SOUBE INFORMAR	DIARIAM
4	7/10/20 22:05	TELEFÔNICO	NÃO	A PRÓPRIA VÍTIMA	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	HORTOLÂNDIA	SEMANALM
...	
17857	3/26/23 22:38	PORTAL	NÃO	TERCEIRO	OUTROS	355030 SÃO PAULO	Ú OCORRÉ
17858	3/29/23 12:49	PORTAL	NÃO	TERCEIRO	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	351870 GUARUJÁ	OCASIONALM
17859	3/31/23 19:34	TELEFÔNICO	NÃO	TERCEIRO	CASA DA VÍTIMA	350070 AGUDOS	DIARIAM
17860	3/31/23 19:34	TELEFÔNICO	NÃO	TERCEIRO	CASA DA VÍTIMA	350070 AGUDOS	DIARIAM
17861	3/31/23 19:34	TELEFÔNICO	NÃO	TERCEIRO	CASA DA VÍTIMA	350070 AGUDOS	DIARIAM

17862 rows × 34 columns



```
In [4]: # Verificando as informações dos dados que iremos utilizar:
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17862 entries, 0 to 17861
Data columns (total 34 columns):
 #   Column                                     Non-Null Count  Dtype
---  -
 0   Data de cadastro                         17862 non-null  object
 1   Canal de atendimento                     17862 non-null  object
 2   Denúncia emergencial                     17862 non-null  object
 3   Denunciante                             17862 non-null  object
 4   Cenário da violação                     17836 non-null  object
 5   Município                               17859 non-null  object
 6   Frequência                              17827 non-null  object
 7   Início das violações                     17827 non-null  object
 8   sl_quantidade_vitimas                    13308 non-null  float64
 9   Grupo vulnerável                         17862 non-null  object
10   Motivação                               17862 non-null  object
11   Relação vítima-suspeito                  17861 non-null  object
12   Sexo da vítima                           17862 non-null  object
13   Orientação sexual da vítima              17775 non-null  object
14   Faixa etária da vítima                   17499 non-null  object
15   UF da vítima                             17856 non-null  object
16   Profissão da vítima                     12463 non-null  object
17   Grau de instrução da vítima               13379 non-null  object
18   Religião da vítima                       8527 non-null   object
19   Raça\Cor da vítima                       17821 non-null  object
20   Etnia da vítima                          8515 non-null   object
21   Faixa de renda da vítima                 12563 non-null  object
22   Natureza Jurídica do Suspeito            17862 non-null  object
23   Sexo do suspeito                         17862 non-null  object
24   Faixa etária do suspeito                 17451 non-null  object
25   Suspeito preso                           17390 non-null  object
26   UF do suspeito                           17849 non-null  object
27   Profissão do suspeito                    15360 non-null  object
28   Grau de instrução do suspeito             8494 non-null   object
29   Religião do suspeito                     8495 non-null   object
30   Raça\Cor do suspeito                    14256 non-null  object
31   Suspeito Etnia                           8703 non-null   object
32   Faixa de renda do suspeito              8494 non-null   object
33   Violacao                                 16792 non-null  object
dtypes: float64(1), object(33)
memory usage: 4.6+ MB
```

```
In [5]: # Conversão de alguns dados para o tipo ideal que será utilizado para análise:
date_col = pd.DatetimeIndex(df['Data de cadastro'])

df['Ano'] = date_col.year
df['Mês'] = date_col.month
df['Dia'] = date_col.day

df['Data de cadastro'] = pd.to_datetime(df['Data de cadastro'])

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17862 entries, 0 to 17861
Data columns (total 37 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Data de cadastro                         17862 non-null  datetime64[ns]
1   Canal de atendimento                     17862 non-null  object
2   Denúncia emergencial                     17862 non-null  object
3   Denunciante                             17862 non-null  object
4   Cenário da violação                     17836 non-null  object
5   Município                               17859 non-null  object
6   Frequência                              17827 non-null  object
7   Início das violações                    17827 non-null  object
8   sl_quantidade_vitimas                    13308 non-null  float64
9   Grupo vulnerável                        17862 non-null  object
10  Motivação                               17862 non-null  object
11  Relação vítima-suspeito                  17861 non-null  object
12  Sexo da vítima                          17862 non-null  object
13  Orientação sexual da vítima              17775 non-null  object
14  Faixa etária da vítima                   17499 non-null  object
15  UF da vítima                            17856 non-null  object
16  Profissão da vítima                     12463 non-null  object
17  Grau de instrução da vítima              13379 non-null  object
18  Religião da vítima                      8527 non-null  object
19  Raça\Cor da vítima                      17821 non-null  object
20  Etnia da vítima                         8515 non-null  object
21  Faixa de renda da vítima                 12563 non-null  object
22  Natureza Jurídica do Suspeito            17862 non-null  object
23  Sexo do suspeito                        17862 non-null  object
24  Faixa etária do suspeito                 17451 non-null  object
25  Suspeito preso                          17390 non-null  object
26  UF do suspeito                          17849 non-null  object
27  Profissão do suspeito                    15360 non-null  object
28  Grau de instrução do suspeito             8494 non-null  object
29  Religião do suspeito                     8495 non-null  object
30  Raça\Cor do suspeito                    14256 non-null  object
31  Suspeito Etnia                          8703 non-null  object
32  Faixa de renda do suspeito               8494 non-null  object
33  Violacao                                16792 non-null  object
34  Ano                                       17862 non-null  int64
35  Mês                                       17862 non-null  int64
36  Dia                                       17862 non-null  int64
dtypes: datetime64[ns](1), float64(1), int64(3), object(32)
memory usage: 5.0+ MB
```

```
In [6]: # Renomeando algumas colunas para facilitar a análise:
df = df.rename(columns={'sl_quantidade_vitimas':'Vítimas', 'Data de cadastro':'Data de cadastro'})
```

```
In [7]: df.head(2)
```

Out[7]:

	Data	Canal de atendimento	Denúncia emergencial	Denunciante	Cenário da violação	Município	Frequência	Início da violação
0	2020-07-04 11:50:00	TELEFÔNICO	SITUAÇÃO FLAGRANTE - ATÉ 24H DA OCORRÊNCIA - S...	A PRÓPRIA VÍTIMA	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SARAPUÍ	ÚNICA OCORRÊNCIA	NÃO SE APLIC
1	2020-07-10 18:37:00	TELEFÔNICO	NÃO	ANÔNIMO	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SÃO PAULO	DIARIAMENTE	HÁ MAIS DE SE MESI

2 rows × 37 columns

```
In [8]: # Identificando quantos valores nulos temos na base de dados:
df.isnull().sum()
```

```
Out[8]: Data                                0
Canal de atendimento                       0
Denúncia emergencial                      0
Denunciante                              0
Cenário da violação                       26
Município                                3
Frequência                               35
Início das violações                      35
Vítimas                                  4554
Grupo vulnerável                          0
Motivação                                0
Relação vítima-suspeito                   1
Sexo da vítima                            0
Orientação sexual da vítima               87
Faixa etária da vítima                    363
UF da vítima                              6
Profissão da vítima                      5399
Grau de instrução da vítima               4483
Religião da vítima                       9335
Raça\Cor da vítima                        41
Etnia da vítima                          9347
Faixa de renda da vítima                  5299
Natureza Jurídica do Suspeito             0
Sexo do suspeito                          0
Faixa etária do suspeito                  411
Suspeito preso                           472
UF do suspeito                            13
Profissão do suspeito                     2502
Grau de instrução do suspeito              9368
Religião do suspeito                      9367
Raça\Cor do suspeito                      3606
Suspeito Etnia                           9159
Faixa de renda do suspeito                9368
Violacao                                 1070
Ano                                         0
Mês                                         0
Dia                                         0
dtype: int64
```

```
In [9]: # Verificando em percentual a quantidade de dados nulos de cada coluna:
df.isnull().sum()/df.shape[0]*100
```

```
Out[9]: Data                                0.000000
Canal de atendimento                      0.000000
Denúncia emergencial                     0.000000
Denunciante                             0.000000
Cenário da violação                      0.145560
Município                               0.016795
Frequência                              0.195947
Início das violações                     0.195947
Vítimas                                 25.495465
Grupo vulnerável                         0.000000
Motivação                               0.000000
Relação vítima-suspeito                  0.005598
Sexo da vítima                          0.000000
Orientação sexual da vítima              0.487068
Faixa etária da vítima                   2.032247
UF da vítima                            0.033591
Profissão da vítima                     30.226178
Grau de instrução da vítima              25.097973
Religião da vítima                      52.261785
Raça\Cor da vítima                      0.229538
Etnia da vítima                         52.328967
Faixa de renda da vítima                 29.666331
Natureza Jurídica do Suspeito            0.000000
Sexo do suspeito                        0.000000
Faixa etária do suspeito                 2.300974
Suspeito preso                          2.642481
UF do suspeito                          0.072780
Profissão do suspeito                   14.007390
Grau de instrução do suspeito             52.446535
Religião do suspeito                    52.440936
Raça\Cor do suspeito                   20.188109
Suspeito Etnia                          51.276453
Faixa de renda do suspeito               52.446535
Violacao                                5.990371
Ano                                      0.000000
Mês                                      0.000000
Dia                                      0.000000
dtype: float64
```

```
In [10]: df.head(1)
```

```
Out[10]:
```

	Data	Canal de atendimento	Denúncia emergencial	Denunciante	Cenário da violação	Município	Frequência	Início da violação
0	2020-07-04 11:50:00	TELEFÔNICO	SITUAÇÃO FLAGRANTE - ATÉ 24H DA OCORRÊNCIA - S...	A PRÓPRIA VÍTIMA	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SARAPUÍ	ÚNICA OCORRÊNCIA	NÃO S APLIC

1 rows × 37 columns

```
In [11]: # Criando uma cópia da base de dados e mantendo a versão original caso precise buscar
df2 = df.copy()
df2.head(2)
```

Out[11]:

	Data	Canal de atendimento	Denúncia emergencial	Denunciante	Cenário da violação	Município	Frequência	Início da violação
0	2020-07-04 11:50:00	TELEFÔNICO	SITUAÇÃO FLAGRANTE - ATÉ 24H DA OCORRÊNCIA - S...	A PRÓPRIA VÍTIMA	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SARAPUÍ	ÚNICA OCORRÊNCIA	NÃO SE APLIC
1	2020-07-10 18:37:00	TELEFÔNICO	NÃO	ANÔNIMO	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SÃO PAULO	DIARIAMENTE	HÁ MAIS DE SE MESI

2 rows × 37 columns

```

In [12]: # Removendo colunas com % maior que 20 para não prejudicar na análise de dados:
df2.drop(["Profissão da vítima",
          "Grau de instrução da vítima",
          "Religião da vítima",
          "Etnia da vítima",
          "Faixa de renda da vítima",
          "Grau de instrução do suspeito",
          "Religião do suspeito",
          "Suspeito Etnia",
          "Faixa de renda do suspeito"], axis=1, inplace=True)

In [13]: # Removendo colunas que não serão utilizadas na análise:
df2.drop(["Grupo vulnerável", "Motivação", "Sexo da vítima", "UF da vítima", "Naturalidade da vítima",
          "Sexo do suspeito", "UF do suspeito", "Violacao", "Profissão do suspeito"], axis=1, inplace=True)

In [14]: # Verificando a conversão da coluna data de cadastro em data:
df2.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17862 entries, 0 to 17861
Data columns (total 19 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Data                                     17862 non-null  datetime64[ns]
1   Canal de atendimento                    17862 non-null  object
2   Denúncia emergencial                   17862 non-null  object
3   Denunciante                            17862 non-null  object
4   Cenário da violação                    17836 non-null  object
5   Município                              17859 non-null  object
6   Frequência                             17827 non-null  object
7   Início das violações                   17827 non-null  object
8   Vítimas                                13308 non-null  float64
9   Relação vítima-suspeito                17861 non-null  object
10  Orientação sexual da vítima             17775 non-null  object
11  Faixa etária da vítima                  17499 non-null  object
12  Raça\Cor da vítima                     17821 non-null  object
13  Faixa etária do suspeito                17451 non-null  object
14  Suspeito preso                         17390 non-null  object
15  Raça\Cor do suspeito                   14256 non-null  object
16  Ano                                     17862 non-null  int64
17  Mês                                    17862 non-null  int64
18  Dia                                    17862 non-null  int64
dtypes: datetime64[ns](1), float64(1), int64(3), object(14)
memory usage: 2.6+ MB

```

```

In [15]: # Conferindo o novo formato da base de dados:
df2.head(5)

```


Out[15]:

	Data	Canal de atendimento	Denúncia emergencial	Denunciante	Cenário da violação	Município	Frequência
0	2020-07-04 11:50:00	TELEFÔNICO	SITUAÇÃO FLAGRANTE - ATÉ 24H DA OCORRÊNCIA - S...	A PRÓPRIA VÍTIMA	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SARAPUÍ	ÚNICA OCORRÊNCIA
1	2020-07-10 18:37:00	TELEFÔNICO	NÃO	ANÔNIMO	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	SÃO PAULO	DIARIAMENTE
2	2020-07-09 09:26:00	TELEFÔNICO	NÃO	A PRÓPRIA VÍTIMA	CASA DA VÍTIMA	GUARULHOS	SEMANALMENTE
3	2020-07-13 14:15:00	TELEFÔNICO	NÃO	ANÔNIMO	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	DENUNCIANTE NÃO SOUBE INFORMAR	DIARIAMENTE
4	2020-07-10 22:05:00	TELEFÔNICO	NÃO	A PRÓPRIA VÍTIMA	CASA ONDE RESIDE A VÍTIMA E O SUSPEITO	HORTOLÂNDIA	SEMANALMENTE

```
In [16]: # Salvando em csv pois iremos utilizar a base no Datastudio:
df2.to_csv('new_data_format.csv', index=False)
```