

DE Training

March 25, 2024

1 Python Recap

1.1 Pandas

- Pandas data cleaning
- Renaming columns
- Rearranging column order
- Checking data types
- Removing Text from Column
- Dealing with missing data
- Changing data type

import pandas as pd

```
[1]: pip install pandas
```

```
Requirement already satisfied: pandas in /opt/conda/lib/python3.11/site-packages (2.2.1)
```

```
Requirement already satisfied: numpy<2,>=1.23.2 in /opt/conda/lib/python3.11/site-packages (from pandas) (1.26.4)
```

```
Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.11/site-packages (from pandas) (2.8.2)
```

```
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.11/site-packages (from pandas) (2023.3.post1)
```

```
Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.11/site-packages (from pandas) (2024.1)
```

```
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.11/site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
[2]: import pandas as pd
```

```
[3]: df = pd.read_csv('https://raw.githubusercontent.com/Theophilusakugre/↵Tertiary-GTL/master/fires.csv', index_col=0)
```

```
[4]: df
```

```
[4]:
```

| | ano | mes | estado | numero | encontro |
|---|------|---------|--------|---------|----------|
| 0 | 1998 | Janeiro | Acre | 0 Fires | 1/1/1998 |

| | | | | | |
|------|------|----------|-----------|---------|----------|
| 1 | 1999 | Janeiro | Acre | 0 Fires | 1/1/1999 |
| 2 | 2000 | Janeiro | Acre | 0 Fires | 1/1/2000 |
| 3 | 2001 | Janeiro | Acre | 0 Fires | 1/1/2001 |
| 4 | 2002 | Janeiro | Acre | 0 Fires | 1/1/2002 |
| ... | ... | ... | ... | ... | ... |
| 6449 | 2012 | Dezembro | Tocantins | 128 | 1/1/2012 |
| 6450 | 2013 | Dezembro | Tocantins | 85 | 1/1/2013 |
| 6451 | 2014 | Dezembro | Tocantins | 223 | 1/1/2014 |
| 6452 | 2015 | Dezembro | Tocantins | 373 | 1/1/2015 |
| 6453 | 2016 | Dezembro | Tocantins | 119 | 1/1/2016 |

[6454 rows x 5 columns]

```
[5]: df = pd.read_csv('fires.csv', index_col=0)
```

```
[6]: df
```

```
[6]:
```

| | ano | mes | estado | numero | encontro |
|------|------|----------|-----------|---------|----------|
| 0 | 1998 | Janeiro | Acre | 0 Fires | 1/1/1998 |
| 1 | 1999 | Janeiro | Acre | 0 Fires | 1/1/1999 |
| 2 | 2000 | Janeiro | Acre | 0 Fires | 1/1/2000 |
| 3 | 2001 | Janeiro | Acre | 0 Fires | 1/1/2001 |
| 4 | 2002 | Janeiro | Acre | 0 Fires | 1/1/2002 |
| ... | ... | ... | ... | ... | ... |
| 6449 | 2012 | Dezembro | Tocantins | 128 | 1/1/2012 |
| 6450 | 2013 | Dezembro | Tocantins | 85 | 1/1/2013 |
| 6451 | 2014 | Dezembro | Tocantins | 223 | 1/1/2014 |
| 6452 | 2015 | Dezembro | Tocantins | 373 | 1/1/2015 |
| 6453 | 2016 | Dezembro | Tocantins | 119 | 1/1/2016 |

[6454 rows x 5 columns]

```
[7]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 6454 entries, 0 to 6453
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   ano         6454 non-null   int64
1   mes         6454 non-null   object
2   estado      6454 non-null   object
3   numero      6322 non-null   object
4   encontro    6454 non-null   object
dtypes: int64(1), object(4)
memory usage: 302.5+ KB
```

```
[8]: df.columns
```

```
[8]: Index(['ano', 'mes', 'estado', 'numero', 'encontro'], dtype='object')
```

```
[9]: col= {'ano': 'Year', 'mes': 'Month', 'estado': 'State', 'numero': 'No_of_Fires', 'encontro': 'Date'}
```

```
[10]: col
```

```
[10]: {'ano': 'Year',  
      'mes': 'Month',  
      'estado': 'State',  
      'numero': 'No_of_Fires',  
      'encontro': 'Date'}
```

```
[11]: df=df.rename(columns=col)
```

```
[12]: df
```

```
[12]:
```

| | Year | Month | State | No_of_Fires | Date |
|------|------|----------|-----------|-------------|----------|
| 0 | 1998 | Janeiro | Acre | 0 Fires | 1/1/1998 |
| 1 | 1999 | Janeiro | Acre | 0 Fires | 1/1/1999 |
| 2 | 2000 | Janeiro | Acre | 0 Fires | 1/1/2000 |
| 3 | 2001 | Janeiro | Acre | 0 Fires | 1/1/2001 |
| 4 | 2002 | Janeiro | Acre | 0 Fires | 1/1/2002 |
| ... | ... | ... | ... | ... | ... |
| 6449 | 2012 | Dezembro | Tocantins | 128 | 1/1/2012 |
| 6450 | 2013 | Dezembro | Tocantins | 85 | 1/1/2013 |
| 6451 | 2014 | Dezembro | Tocantins | 223 | 1/1/2014 |
| 6452 | 2015 | Dezembro | Tocantins | 373 | 1/1/2015 |
| 6453 | 2016 | Dezembro | Tocantins | 119 | 1/1/2016 |

```
[6454 rows x 5 columns]
```

```
[13]: df.columns
```

```
[13]: Index(['Year', 'Month', 'State', 'No_of_Fires', 'Date'], dtype='object')
```

```
[14]: df.columns[[4,1,0,2,3]]
```

```
[14]: Index(['Date', 'Month', 'Year', 'State', 'No_of_Fires'], dtype='object')
```

```
[15]: df[['Date', 'State', 'No_of_Fires']]
```

```
[15]:
```

| | Date | State | No_of_Fires |
|---|----------|-------|-------------|
| 0 | 1/1/1998 | Acre | 0 Fires |
| 1 | 1/1/1999 | Acre | 0 Fires |

| | | | |
|------|----------|-----------|---------|
| 2 | 1/1/2000 | Acre | 0 Fires |
| 3 | 1/1/2001 | Acre | 0 Fires |
| 4 | 1/1/2002 | Acre | 0 Fires |
| ... | ... | ... | ... |
| 6449 | 1/1/2012 | Tocantins | 128 |
| 6450 | 1/1/2013 | Tocantins | 85 |
| 6451 | 1/1/2014 | Tocantins | 223 |
| 6452 | 1/1/2015 | Tocantins | 373 |
| 6453 | 1/1/2016 | Tocantins | 119 |

[6454 rows x 3 columns]

```
[16]: df.dtypes
```

```
[16]: Year          int64
      Month         object
      State         object
      No_of_Fires   object
      Date          object
      dtype: object
```

```
[17]: pd.to_datetime(df['Date'])
```

```
[17]: 0      1998-01-01
      1      1999-01-01
      2      2000-01-01
      3      2001-01-01
      4      2002-01-01
      ...
      6449    2012-01-01
      6450    2013-01-01
      6451    2014-01-01
      6452    2015-01-01
      6453    2016-01-01
      Name: Date, Length: 6454, dtype: datetime64[ns]
```

```
[18]: df['Date']=pd.to_datetime(df['Date'])
```

```
[19]: df.dtypes
```

```
[19]: Year          int64
      Month         object
      State         object
      No_of_Fires   object
      Date          datetime64[ns]
      dtype: object
```

```
[20]: df.Date.dt.month
```

```
[20]: 0      1
      1      1
      2      1
      3      1
      4      1
      ..
      6449    1
      6450    1
      6451    1
      6452    1
      6453    1
      Name: Date, Length: 6454, dtype: int32
```

```
[21]: df.Date.dt.year
```

```
[21]: 0      1998
      1      1999
      2      2000
      3      2001
      4      2002
      ...
      6449    2012
      6450    2013
      6451    2014
      6452    2015
      6453    2016
      Name: Date, Length: 6454, dtype: int32
```

```
[22]: df['Day_of_Week']=df.Date.dt.dayofweek
```

```
[23]: df
```

```
[23]:   Year  Month  State No_of_Fires  Date  Day_of_Week
0   1998  Janeiro  Acre      0 Fires 1998-01-01         3
1   1999  Janeiro  Acre      0 Fires 1999-01-01         4
2   2000  Janeiro  Acre      0 Fires 2000-01-01         5
3   2001  Janeiro  Acre      0 Fires 2001-01-01         0
4   2002  Janeiro  Acre      0 Fires 2002-01-01         1
...   ...   ...   ...   ...   ...   ...
6449 2012  Dezembro  Tocantins      128 2012-01-01         6
6450 2013  Dezembro  Tocantins       85 2013-01-01         1
6451 2014  Dezembro  Tocantins      223 2014-01-01         2
6452 2015  Dezembro  Tocantins      373 2015-01-01         3
6453 2016  Dezembro  Tocantins      119 2016-01-01         4
```

[6454 rows x 6 columns]

```
[24]: df.Day_of_Week.unique()
```

```
[24]: array([3, 4, 5, 0, 1, 2, 6], dtype=int32)
```

```
[25]: day= {3:'thur', 4:'fri', 5:'sat', 0:'mon', 1:'tue', 2:'wed', 6:'sun'}
```

```
[26]: df.Day_of_Week.map(day)
```

```
[26]: 0      thur
1      fri
2      sat
3      mon
4      tue
...
6449   sun
6450   tue
6451   wed
6452   thur
6453   fri
Name: Day_of_Week, Length: 6454, dtype: object
```

```
[27]: df
```

```
[27]:
```

| | Year | Month | State | No_of_Fires | Date | Day_of_Week |
|------|------|----------|-----------|-------------|------------|-------------|
| 0 | 1998 | Janeiro | Acre | 0 Fires | 1998-01-01 | 3 |
| 1 | 1999 | Janeiro | Acre | 0 Fires | 1999-01-01 | 4 |
| 2 | 2000 | Janeiro | Acre | 0 Fires | 2000-01-01 | 5 |
| 3 | 2001 | Janeiro | Acre | 0 Fires | 2001-01-01 | 0 |
| 4 | 2002 | Janeiro | Acre | 0 Fires | 2002-01-01 | 1 |
| ... | ... | ... | ... | ... | ... | ... |
| 6449 | 2012 | Dezembro | Tocantins | 128 | 2012-01-01 | 6 |
| 6450 | 2013 | Dezembro | Tocantins | 85 | 2013-01-01 | 1 |
| 6451 | 2014 | Dezembro | Tocantins | 223 | 2014-01-01 | 2 |
| 6452 | 2015 | Dezembro | Tocantins | 373 | 2015-01-01 | 3 |
| 6453 | 2016 | Dezembro | Tocantins | 119 | 2016-01-01 | 4 |

[6454 rows x 6 columns]

```
[28]: df['Day']=df.Day_of_Week.map(day)
```

```
[29]: df
```

```
[29]:
```

| | Year | Month | State | No_of_Fires | Date | Day_of_Week | Day |
|---|------|---------|-------|-------------|------------|-------------|------|
| 0 | 1998 | Janeiro | Acre | 0 Fires | 1998-01-01 | 3 | thur |
| 1 | 1999 | Janeiro | Acre | 0 Fires | 1999-01-01 | 4 | fri |

| | | | | | | | |
|------|------|----------|-----------|---------|------------|-----|------|
| 2 | 2000 | Janeiro | Acre | 0 Fires | 2000-01-01 | 5 | sat |
| 3 | 2001 | Janeiro | Acre | 0 Fires | 2001-01-01 | 0 | mon |
| 4 | 2002 | Janeiro | Acre | 0 Fires | 2002-01-01 | 1 | tue |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 6449 | 2012 | Dezembro | Tocantins | 128 | 2012-01-01 | 6 | sun |
| 6450 | 2013 | Dezembro | Tocantins | 85 | 2013-01-01 | 1 | tue |
| 6451 | 2014 | Dezembro | Tocantins | 223 | 2014-01-01 | 2 | wed |
| 6452 | 2015 | Dezembro | Tocantins | 373 | 2015-01-01 | 3 | thur |
| 6453 | 2016 | Dezembro | Tocantins | 119 | 2016-01-01 | 4 | fri |

[6454 rows x 7 columns]

1.1.1 working on the 'No_of_Fires' column

```
[30]: df['No_of_Fires']
```

```
[30]: 0      0 Fires
      1      0 Fires
      2      0 Fires
      3      0 Fires
      4      0 Fires
      ...
      6449      128
      6450      85
      6451      223
      6452      373
      6453      119
      Name: No_of_Fires, Length: 6454, dtype: object
```

```
[31]: df['No_of_Fires']=df['No_of_Fires'].str.strip(' Fires Date').astype(float)
```

```
[32]: df.columns
```

```
[32]: Index(['Year', 'Month', 'State', 'No_of_Fires', 'Date', 'Day_of_Week', 'Day'],
      dtype='object')
```

```
[ ]:
```

```
[33]: df
```

```
[33]:   Year  Month  State  No_of_Fires  Date  Day_of_Week  Day
      0  1998  Janeiro  Acre        0.0  1998-01-01      3  thur
      1  1999  Janeiro  Acre        0.0  1999-01-01      4  fri
      2  2000  Janeiro  Acre        0.0  2000-01-01      5  sat
      3  2001  Janeiro  Acre        0.0  2001-01-01      0  mon
      4  2002  Janeiro  Acre        0.0  2002-01-01      1  tue
      ...  ...  ...  ...  ...  ...  ...  ...
```

| | | | | | | | |
|------|------|----------|-----------|-------|------------|---|------|
| 6449 | 2012 | Dezembro | Tocantins | 128.0 | 2012-01-01 | 6 | sun |
| 6450 | 2013 | Dezembro | Tocantins | 85.0 | 2013-01-01 | 1 | tue |
| 6451 | 2014 | Dezembro | Tocantins | 223.0 | 2014-01-01 | 2 | wed |
| 6452 | 2015 | Dezembro | Tocantins | 373.0 | 2015-01-01 | 3 | thur |
| 6453 | 2016 | Dezembro | Tocantins | 119.0 | 2016-01-01 | 4 | fri |

[6454 rows x 7 columns]

1.1.2 null values

```
[34]: df.isna().sum()
```

```
[34]: Year          0
      Month        0
      State        0
      No_of_Fires  132
      Date         0
      Day_of_Week  0
      Day          0
      dtype: int64
```

```
[36]: df.dropna()
```

```
[36]:
```

| | Year | Month | State | No_of_Fires | Date | Day_of_Week | Day |
|------|------|----------|-----------|-------------|------------|-------------|------|
| 0 | 1998 | Janeiro | Acre | 0.0 | 1998-01-01 | 3 | thur |
| 1 | 1999 | Janeiro | Acre | 0.0 | 1999-01-01 | 4 | fri |
| 2 | 2000 | Janeiro | Acre | 0.0 | 2000-01-01 | 5 | sat |
| 3 | 2001 | Janeiro | Acre | 0.0 | 2001-01-01 | 0 | mon |
| 4 | 2002 | Janeiro | Acre | 0.0 | 2002-01-01 | 1 | tue |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 6449 | 2012 | Dezembro | Tocantins | 128.0 | 2012-01-01 | 6 | sun |
| 6450 | 2013 | Dezembro | Tocantins | 85.0 | 2013-01-01 | 1 | tue |
| 6451 | 2014 | Dezembro | Tocantins | 223.0 | 2014-01-01 | 2 | wed |
| 6452 | 2015 | Dezembro | Tocantins | 373.0 | 2015-01-01 | 3 | thur |
| 6453 | 2016 | Dezembro | Tocantins | 119.0 | 2016-01-01 | 4 | fri |

[6322 rows x 7 columns]

```
[41]: df.describe()
```

```
[41]:
```

| | Year | No_of_Fires | Date | Day_of_Week |
|-------|-------------|-------------|-------------------------------|-------------|
| count | 6454.000000 | 6322.000000 | 6454 | 6454.000000 |
| mean | 2007.461729 | 110.667972 | 2007-06-18 18:28:53.622559488 | 3.188720 |
| min | 1998.000000 | 0.000000 | 1998-01-01 00:00:00 | 0.000000 |
| 25% | 2002.000000 | 3.000000 | 2002-01-01 00:00:00 | 1.000000 |
| 50% | 2007.000000 | 26.000000 | 2007-01-01 00:00:00 | 3.000000 |
| 75% | 2012.000000 | 117.000000 | 2012-01-01 00:00:00 | 5.000000 |

| | | | | |
|-----|-------------|------------|---------------------|----------|
| max | 2017.000000 | 998.000000 | 2017-01-01 00:00:00 | 6.000000 |
| std | 5.746654 | 192.248217 | NaN | 1.908821 |

```
[43]: round(df['No_of_Fires'].mean(),2)
```

```
[43]: 110.67
```

```
[44]: df['No_of_Fires'].fillna(round(df['No_of_Fires'].mean(),2))
```

```
[44]: 0          0.0
      1          0.0
      2          0.0
      3          0.0
      4          0.0
      ...
      6449      128.0
      6450       85.0
      6451      223.0
      6452      373.0
      6453      119.0
      Name: No_of_Fires, Length: 6454, dtype: float64
```

```
[47]: df.iloc[[109,110,111]]
```

```
[47]:      Year  Month State  No_of_Fires      Date  Day_of_Week  Day
      109  2007  Junho  Acre           4.0  2007-01-01          0  mon
      110  2008  Junho  Acre          NaN  2008-01-01          1  tue
      111  2009  Junho  Acre           0.0  2009-01-01          3  thur
```

```
[51]: df['No_of_Fires'].fillna(method='ffill').iloc[[109,110,111]]
```

/tmp/ipykernel_9941/2932004214.py:1: FutureWarning: Series.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.

```
df['No_of_Fires'].fillna(method='ffill').iloc[[109,110,111]]
```

```
[51]: 109      4.0
      110      4.0
      111      0.0
      Name: No_of_Fires, dtype: float64
```

```
[52]: df['No_of_Fires'].fillna(method='bfill').iloc[[109,110,111]]
```

/tmp/ipykernel_9941/755239207.py:1: FutureWarning: Series.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead.

```
df['No_of_Fires'].fillna(method='bfill').iloc[[109,110,111]]
```

```
[52]: 109    4.0
      110    0.0
      111    0.0
      Name: No_of_Fires, dtype: float64
```

```
[54]: df['No_of_Fires'].ffill().iloc[[109,110,111]]
```

```
[54]: 109    4.0
      110    4.0
      111    0.0
      Name: No_of_Fires, dtype: float64
```

```
[57]: 'Accra, Ghana'.split(',')[0]
```

```
[57]: 'Accra'
```

```
[59]: df.columns
```

```
[59]: Index(['Year', 'Month', 'State', 'No_of_Fires', 'Date', 'Day_of_Week', 'Day'],
      dtype='object')
```

```
[62]: df=df[['Date', 'Year', 'Month', 'Day', 'State', 'No_of_Fires']]
      df
```

```
[62]:
```

| | Date | Year | Month | Day | State | No_of_Fires |
|------|------------|------|----------|------|-----------|-------------|
| 0 | 1998-01-01 | 1998 | Janeiro | thur | Acre | 0.0 |
| 1 | 1999-01-01 | 1999 | Janeiro | fri | Acre | 0.0 |
| 2 | 2000-01-01 | 2000 | Janeiro | sat | Acre | 0.0 |
| 3 | 2001-01-01 | 2001 | Janeiro | mon | Acre | 0.0 |
| 4 | 2002-01-01 | 2002 | Janeiro | tue | Acre | 0.0 |
| ... | ... | ... | ... | ... | ... | ... |
| 6449 | 2012-01-01 | 2012 | Dezembro | sun | Tocantins | 128.0 |
| 6450 | 2013-01-01 | 2013 | Dezembro | tue | Tocantins | 85.0 |
| 6451 | 2014-01-01 | 2014 | Dezembro | wed | Tocantins | 223.0 |
| 6452 | 2015-01-01 | 2015 | Dezembro | thur | Tocantins | 373.0 |
| 6453 | 2016-01-01 | 2016 | Dezembro | fri | Tocantins | 119.0 |

```
[6454 rows x 6 columns]
```

```
[65]: df.drop('Day', axis=1)
```

```
[65]:
```

| | Date | Year | Month | State | No_of_Fires |
|---|------------|------|---------|-------|-------------|
| 0 | 1998-01-01 | 1998 | Janeiro | Acre | 0.0 |
| 1 | 1999-01-01 | 1999 | Janeiro | Acre | 0.0 |
| 2 | 2000-01-01 | 2000 | Janeiro | Acre | 0.0 |
| 3 | 2001-01-01 | 2001 | Janeiro | Acre | 0.0 |
| 4 | 2002-01-01 | 2002 | Janeiro | Acre | 0.0 |

```

...      ...      ...      ...      ...
6449 2012-01-01  2012  Dezembro  Tocantins      128.0
6450 2013-01-01  2013  Dezembro  Tocantins       85.0
6451 2014-01-01  2014  Dezembro  Tocantins      223.0
6452 2015-01-01  2015  Dezembro  Tocantins      373.0
6453 2016-01-01  2016  Dezembro  Tocantins      119.0

```

[6454 rows x 5 columns]

```
[68]: df.drop(['Year', 'Month', 'Day'], axis=1)
```

```

[68]:
      Date      State  No_of_Fires
0  1998-01-01     Acre          0.0
1  1999-01-01     Acre          0.0
2  2000-01-01     Acre          0.0
3  2001-01-01     Acre          0.0
4  2002-01-01     Acre          0.0
...
6449 2012-01-01  Tocantins      128.0
6450 2013-01-01  Tocantins       85.0
6451 2014-01-01  Tocantins      223.0
6452 2015-01-01  Tocantins      373.0
6453 2016-01-01  Tocantins      119.0

```

[6454 rows x 3 columns]

```
[70]: df.drop(2) # drop row
```

```

[70]:
      Date  Year  Month  Day  State  No_of_Fires
0  1998-01-01  1998  Janeiro  thur     Acre          0.0
1  1999-01-01  1999  Janeiro  fri     Acre          0.0
3  2001-01-01  2001  Janeiro  mon     Acre          0.0
4  2002-01-01  2002  Janeiro  tue     Acre          0.0
5  2003-01-01  2003  Janeiro  wed     Acre         10.0
...
6449 2012-01-01  2012  Dezembro  sun  Tocantins      128.0
6450 2013-01-01  2013  Dezembro  tue  Tocantins       85.0
6451 2014-01-01  2014  Dezembro  wed  Tocantins      223.0
6452 2015-01-01  2015  Dezembro  thur  Tocantins      373.0
6453 2016-01-01  2016  Dezembro  fri  Tocantins      119.0

```

[6453 rows x 6 columns]

```
[71]: df.shape
```

```
[71]: (6454, 6)
```

```
[72]: df
```

```
[72]:
```

| | Date | Year | Month | Day | State | No_of_Fires |
|------|------------|------|----------|------|-----------|-------------|
| 0 | 1998-01-01 | 1998 | Janeiro | thur | Acre | 0.0 |
| 1 | 1999-01-01 | 1999 | Janeiro | fri | Acre | 0.0 |
| 2 | 2000-01-01 | 2000 | Janeiro | sat | Acre | 0.0 |
| 3 | 2001-01-01 | 2001 | Janeiro | mon | Acre | 0.0 |
| 4 | 2002-01-01 | 2002 | Janeiro | tue | Acre | 0.0 |
| ... | ... | ... | ... | ... | ... | ... |
| 6449 | 2012-01-01 | 2012 | Dezembro | sun | Tocantins | 128.0 |
| 6450 | 2013-01-01 | 2013 | Dezembro | tue | Tocantins | 85.0 |
| 6451 | 2014-01-01 | 2014 | Dezembro | wed | Tocantins | 223.0 |
| 6452 | 2015-01-01 | 2015 | Dezembro | thur | Tocantins | 373.0 |
| 6453 | 2016-01-01 | 2016 | Dezembro | fri | Tocantins | 119.0 |

```
[6454 rows x 6 columns]
```

2 groupby

```
[76]: round(df.groupby('State')['No_of_Fires'].sum(),2)
```

```
[76]: State
```

| | |
|------------------|----------|
| Acre | 17971.03 |
| Amapa | 20108.58 |
| Amazonas | 29890.13 |
| Bahia | 43411.95 |
| Ceara | 30395.04 |
| Distrito Federal | 3501.00 |
| Espirito Santo | 37002.28 |
| Goiias | 23362.85 |
| Maranhao | 24839.17 |
| Mato Grosso | 95566.98 |
| Minas Gerais | 37002.28 |
| Paraiba | 51723.92 |
| Pernambuco | 22959.00 |
| Piau | 37542.65 |
| Rio | 44799.86 |
| Roraima | 23749.07 |
| Santa Catarina | 23362.85 |
| Sao Paulo | 49888.20 |
| Sergipe | 3167.00 |
| Tocantins | 33441.51 |
| alagoas | 4594.00 |
| par  | 22124.14 |
| rondonia | 19239.43 |

```
Name: No_of_Fires, dtype: float64
```

```
[77]: round(df.groupby('State')['No_of_Fires'].mean(),2)
```

```
[77]: State
Acre                76.80
Amapa               87.05
Amazonas           127.74
Bahia              186.32
Ceara              128.79
Distrito Federal   14.96
Espirito Santo     160.18
Goiás              99.42
Maranhao           105.70
Mato Grosso        202.47
Minas Gerais       160.18
Paraiba            110.29
Pernambuco         100.26
Piau               161.13
Rio                63.28
Roraima            101.49
Santa Catarina     99.42
Sao Paulo          211.39
Sergipe            13.59
Tocantins          143.53
alagoas            19.22
par               94.55
rondonia           82.57
Name: No_of_Fires, dtype: float64
```

```
[80]: round(df.groupby(['State', 'Year', 'Day'])['No_of_Fires'].mean(),2)
```

```
[80]: State  Year  Day
Acre      1998  thur    60.83
          1999  fri    27.75
          2000  sat    36.17
          2001  mon    69.00
          2002  tue   128.66
          ...
rondonia  2013  tue   157.31
          2014  wed    36.39
          2015  thur   77.71
          2016  fri   167.23
          2017  sun    63.85
Name: No_of_Fires, Length: 460, dtype: float64
```

```
[81]: df['State'].unique()
```

```
[81]: array(['Acre', 'alagoas', 'Amapa', 'Amazonas', 'Bahia', 'Ceara',  
          'Distrito Federal', 'Espirito Santo', 'Goias', 'Maranhao',  
          'Mato Grosso', 'Minas Gerais', 'par  ', 'Paraiba', 'Pernambuco',  
          'Piau', 'Rio', 'rondonia', 'Roraima', 'Santa Catarina',  
          'Sao Paulo', 'Sergipe', 'Tocantins'], dtype=object)
```

```
[ ]:
```