

Biblioteca pandas e geopandas

Disciplina: AUTOMATIZAÇÃO DE ANÁLISES URBANAS EM SIG - PYTHON PARA QGIS

Aluna: Tamires Lenhart

Nível: Doutorado

Semestre: 2024/2

Data: 18/11/2024

Pandas

Pandas é um biblioteca para **manipulação e análise de dados**, escrita em Python. Muito utilizada para iniciar análises e sistematizações de dados no Python, ele permite **leitura, manipulação, agregação e plot de dados** em poucos passos (códigos).

Fonte: Ribeiro, Lucas. Introdução a Biblioteca pandas. Medium. *online*. 2020.
Disponível em:
<https://medium.com/tech-grupozap/introdu%C3%A7%C3%A3o-a-biblioteca-pandas-89fa8ed4fa38>. Acesso em: Nov. 2024.

Dataframe

Dataframe se trata de uma **planilha ou banco de dados**, composto por elementos de **colunas, linhas e índice**.

Quando um arquivo do **Excel** (csv, dbf, xls) é **importado** e lido pelo **pandas** ele é **convertido** em um **Dataframe**.

Índice	Colunas	
Linhas		

Fonte: Ribeiro, Lucas. Introdução a Biblioteca pandas. Medium. *online*. 2020.
Disponível em:
<https://medium.com/tech-grupozap/introdu%C3%A7%C3%A3o-a-biblioteca-pandas-89fa8ed4fa38>. Acesso em: Nov. 2024.

A popularidade da biblioteca pandas

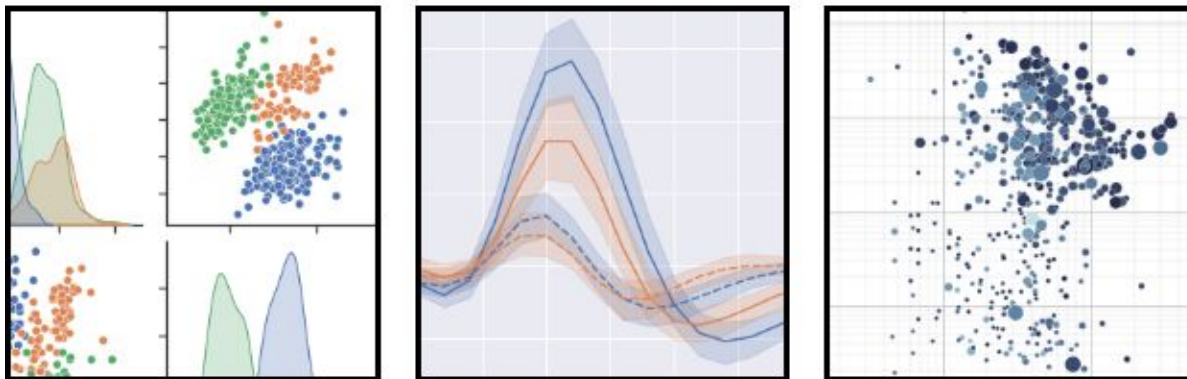
- Um objeto rápido e eficiente para manipulação de dados com indexação integrada em *DataFrame*;
- Ferramentas para leitura e gravação de dados entre estruturas de dados preexistentes e suporte para diferentes formatos;
- Alinhamento inteligente e tratamento integrado de dados ausentes: obtenha alinhamento automático baseado em rótulos, em cálculos e manipule facilmente dados desordenados de forma sistemática;
- Remodelação flexível de conjuntos de dados;
- As colunas e linhas podem ser inseridas e excluídas das estruturas de dados;
- Mesclagem e união de conjuntos de dados de alto desempenho;
- Funcionalidade de série temporal: geração de intervalo de datas e conversão de frequência, estatísticas de janela móvel, mudança e atraso de data. Crie até mesmo compensações de horário específicas de domínio e junte séries temporais sem perda de dados.

Bibliotecas complementares do pandas

Bibliotecas que podem ser utilizadas para refinar os processos de análise:

seaborn

Estatística básica e
visualização de
dados



Fonte:
Ribeiro, Lucas. Introdução a Biblioteca pandas. Medium. *online*. 2020. Disponível em:
<https://medium.com/tech-grupozap/introdu%C3%A7%C3%A3o-a-biblioteca-pandas-89fa8ed4fa38>.
Acesso em: Nov. 2024.
Waskom, Michael. Seaborn: statistical data visualization. PyData. *online*. 2014-2024. Disponível
em: <https://seaborn.pydata.org/>. Acesso em: Nov. 2024.

Bibliotecas complementares do pandas

Bibliotecas que podem ser utilizadas para refinar os processos de análise:



Funções matemáticas
de diversas áreas do
conhecimento

```
To try the examples in the browser:
1. Type code in the input cell and press
   Shift + Enter to execute
2. Or copy paste the code, and click on
   the "Run" button in the toolbar
....

# The standard way to import NumPy:
import numpy as np

# Create a 2-D array, set every second element in
# some rows and find max per row:
```

Fonte:
Ribeiro, Lucas. Introdução a Biblioteca pandas. Medium. *online*. 2020. Disponível em:
<https://medium.com/tech-grupozap/introdu%C3%A7%C3%A3o-a-biblioteca-pandas-89fa8ed4fa38>.

Acesso em: Nov. 2024.

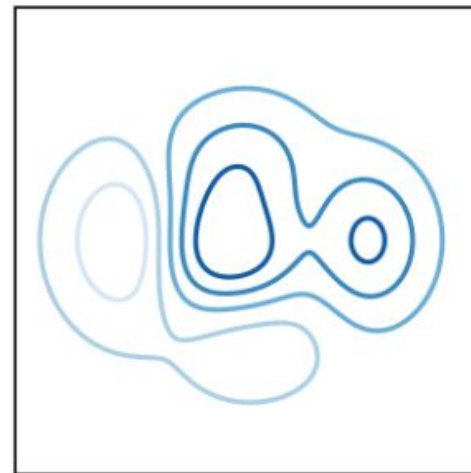
Numpy, Team. NUMPY:The fundamental package for scientific computing with Python. *online*.
2024. Disponível em: <https://numpy.org/>. Acesso em: Nov. 2024.

Bibliotecas complementares do pandas

Bibliotecas que podem ser utilizadas para refinar os processos de análise:



Visualização de dados:
estatísticos, interativos
e animados.



`contour(X, Y, Z)`

Fonte:
Ribeiro, Lucas. Introdução a Biblioteca pandas. Medium. *online*. 2020. Disponível em:
<https://medium.com/tech-grupozap/introdu%C3%A7%C3%A3o-a-biblioteca-pandas-89fa8ed4fa38>.

Acesso em: Nov. 2024.

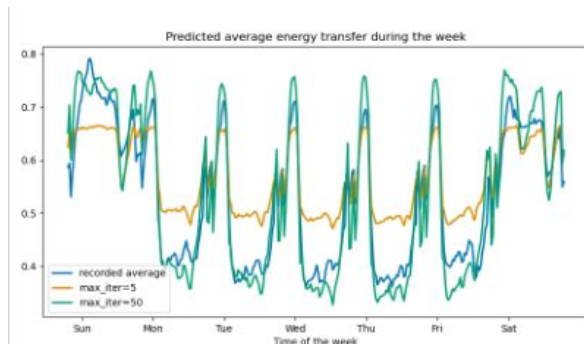
Matplotlib, Team. Matplotlib: Visualization with Python. *online*. 2012-2024. Disponível em:
<https://matplotlib.org/> . Acesso em: Nov. 2024.

Bibliotecas complementares do pandas

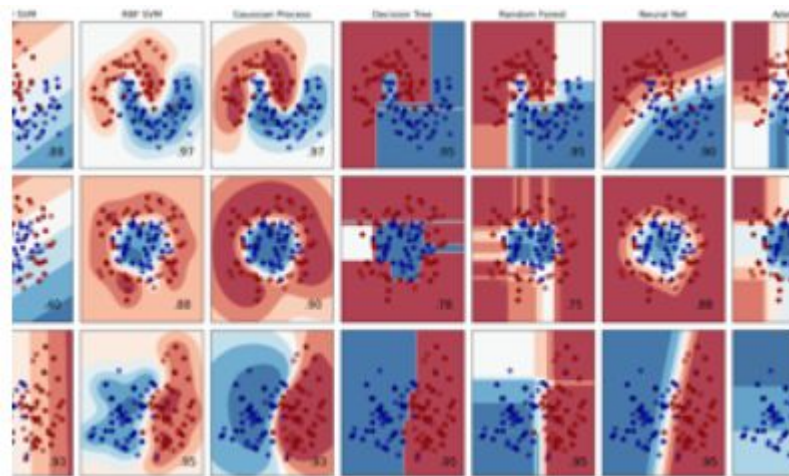
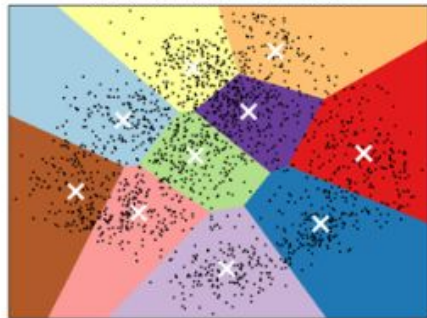
Bibliotecas que podem ser utilizadas para refinar os processos de análise:



Classificação,
cluster e
regressão.



K-means clustering on the digits dataset (PCA-reduced data)
Centroids are marked with white cross



Fonte:

Ribeiro, Lucas. Introdução a Biblioteca pandas. Medium. *online*. 2020.

Disponível em:

<https://medium.com/tech-grupozap/introdu%C3%A7%C3%A3o-a-biblioteca-pandas-89fa8ed4fa38>. Acesso em: Nov. 2024.

Scikit-learn, Team. scikit-learn: Machine Learning in Python. *online*. 2012-2024.

Disponível em: <https://scikit-learn.org/stable/>. Acesso em: Nov. 2024.

Instalar biblioteca pandas

#Jupyter Lab ou Jupyter Notebook:

#Instalando Pandas

```
pip install pandas
```

#Importando a biblioteca

```
importar pandas como pd
```

#Lendo arquivos de dados (csv)

```
df = pd.read_csv('caminho_arquivo.csv', sep='separador')
```

#Lendo arquivos de dados (xls)

```
df = pd.read_excel('caminho_arquivo.xlsx', sheet_name='guia')
```

Fonte:

Ribeiro, Lucas. Introdução a Biblioteca pandas. Medium. *online*. 2020.

Disponível em:

<https://medium.com/tech-grupozap/introdu%C3%A7%C3%A3o-a-biblioteca-pandas-89fa8ed4fa38>. Acesso em: Nov. 2024.

Geopandas

Geopandas é um projeto de **código aberto** para **facilitar o** trabalho com **dados geoespaciais em python**. **Geopandas** estende os tipos de dados usados pelos **pandas** para **permitir operações espaciais em tipos geométricos**.

Geopandas depende ainda da **extensão pyogrio** para **acesso aos arquivos** e da **biblioteca matplotlib** para **plotagem**. As **operações geométricas** são **realizadas pelo shapely**.

Fonte: Developers, Geopandas. Geopandas:
Data Structures. Pydata. *online*. 2013.

Disponível em:

https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov.

2024.

Geopandas

Geopandas é um projeto de código aberto para permitir operações geoespaciais em python. Geopandas depende ainda da extensão `pyogrio` para acesso aos arquivos e da biblioteca `matplotlib` para plotagem. As operações geométricas são realizadas pelo `shapely`.

*Pyogrio fornece acesso de leitura, gravação rápido e orientado em massa as fontes de dados vetoriais **GDAL/OGR**, como ESRI Shapefile, GeoPackage, GeoJSON e vários outros. As fontes de dados vetoriais são pontos, linhas ou polígonos associados às tabelas de atributos.*

Fonte: Developers, Geopandas. Geopandas: Data Structures. Pydata. *online*. 2013.

Disponível em:

https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov.

2024.

Geopandas

Geopandas
geoespaciais
pandas

Shapely envolve geometrias e operações **GEOS** para fornecer uma interface rica em recursos para **geometrias singulares (escalares)** e **ufuncs NumPy de alto desempenho** para operações usando **matrizes de geometrias**.

Geopandas

da biblioteca matplotlib para plotagem. As operações geométricas são realizadas pelo **shapely**.

o trabalho com **dados** de dados usados pelos **geométricos**.

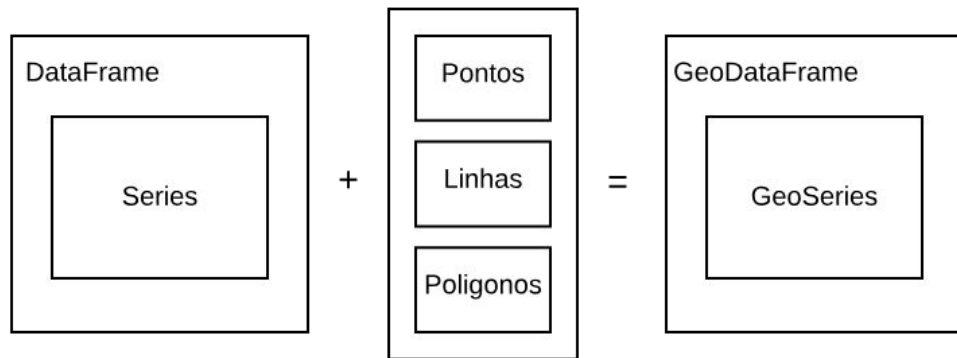
acesso aos arquivos e

Fonte: Developers, Geopandas. Geopandas: Data Structures. Pydata. *online*. 2013. Disponível em: https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov. 2024.

Geopandas

GeoPandas implementa duas estruturas de dados principais, uma **GeoSeries** e uma **GeoDataFrame**. Estas são subclasses de **pandas.Series** e **pandas.DataFrame**, respectivamente.

Combina os recursos de **pandas** e **shapely**, fornecendo **operações geoespaciais** em pandas e uma interface de **alto nível para múltiplas geometrias**.



Fonte: Developers, Geopandas. Geopandas: Data Structures. Pydata. *online*. 2013.

Disponível em:

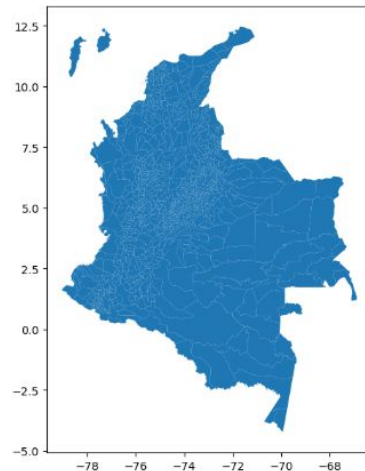
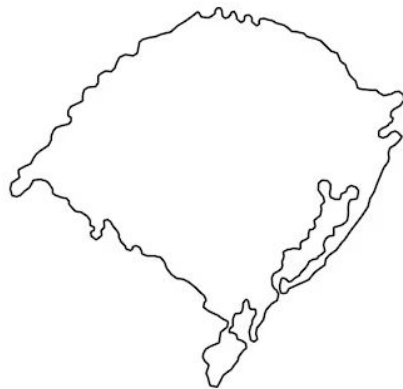
https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov.

2024.

Geopandas - GeoSeries

Uma **GeoSeries** é essencialmente um vetor, onde cada entrada de forma é um conjunto de formas correspondentes a uma observação.

A entrada pode consistir em um polígono único ou múltiplas formas que são incluídas como uma observação:



Fonte: Developers, Geopandas. Geopandas: Data Structures. Pydata. *online*. 2013.
Disponível em:
https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov. 2024.

Geopandas - *GeoSeries_Atributos*

- **área:** área da forma (unidades de projeção);
- **limites:** tupla de coordenadas máximas e mínimas em cada eixo para cada forma;
- **total_bounds:** tupla de coordenadas máximas e mínimas em cada eixo para `GeoSeries` inteiro;
- **geom_type:** tipologia da geometria;
- **is_valid:** testa se as coordenadas formam uma forma geométrica razoável de acordo com o padrão Simple Feature Access.

Fonte: Developers, Geopandas. Geopandas:
Data Structures. Pydata. *online*. 2013.
Disponível em:

https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov.
2024.

Geopandas - *GeoSeries_Atributos*

- **área:** área da forma (unidades de projeção);
- **limites:** tupla de coordenadas máximas e mínimas em cada eixo para cada forma;
- **total_bounds:** tupla de coordenadas máximas e mínimas para o `GeoSeries` inteiro;
- **geom_type:** tipologia geométrica da forma;
- **is_valid:** testa se as coordenadas são válidas de acordo com o padrão da WKT.

CRS - Coordinate reference systems.
O sistema de referência de coordenadas (CRS) é importante porque as **formas geométricas** em um objeto `GeoSeries` ou `GeoDataFrame` são simplesmente uma coleção de coordenadas em um espaço arbitrário.

Fonte: Developers, Geopandas. Geopandas Data Structures. Pydata. *online*. 2013. Disponível em:

https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov. 2024.

Fonte: https://geopandas.org/en/stable/docs/user_guide/projections.html

Geopandas

- **área:** área
- **limites:** tupla
- **total_boun**
- **geom_type:** tipologia
- **is_valid:** testa se as coord

Esta parte do **Simple Features Access (SFA)**, também chamada de **ISO 19125**, descreve a **arquitetura comum para geometria de feature simples**. A classe base **Geometry** tem subclasses para **Point, Curve, Surface e GeometryCollection**. Cada objeto geométrico é associado a um **Coordinate Reference System (CRS)**, que descreve o **espaço de coordenadas no qual o objeto geométrico é definido**.

Fonte: <https://www.ogc.org/pt/publications/standard/sfa/>

Fonte: Developers, Geopandas. Geopandas: Data Structures. Pydata. *online*. 2013. Disponível em: https://geopandas.org/en/stable/docs/user_guide/data_structures.html. Acesso em: Nov. 2024.

Geopandas - *GeoSeries* métodos básicos

- **distância()**: retorna séries com distância mínima de cada entrada para outra;
- **centróide**: retorna *GeoSeries* de centróides;
- ***representative_point()***: *retorna GeoSeries de pontos que estão representados dentro de cada geometria. NÃO retorna centróides.*
- **to_crs()**: altera o sistema de referência de coordenadas.
- **plot()**: plota *GeoSeries*.

Fonte: Developers, Geopandas. Geopandas:
Data Structures. Pydata. *online*. 2013.

Disponível em:

https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov.
2024.

Geopandas - *GeoSeries_testes de relação*

- ***geom_equals_exact()***: tem formato igual a outro (até uma tolerância de casa decimal especificada);
- ***contém()***: é a forma contida em outro;
- ***intersects()***: a forma cruza outra;
-

Fonte: Developers, Geopandas. Geopandas:
Data Structures. Pydata. *online*. 2013.

Disponível em:

https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov. 2024.

Geopandas - *GeoDataFrame*

Um `GeoDataFrame` é uma **estrutura de dados tabular** que contém um `GeoSeries`.

A propriedade mais importante de um `GeoDataFrame` é que ele **sempre possui** uma **coluna `GeoSeries`** que **possui** um ***status especial*** - a “coluna de geometria ativa”.

Fonte: Developers, Geopandas. Geopandas:
Data Structures. Pydata. *online*. 2013.

Disponível em:

https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov.

2024.

Geopandas - *GeoDataFrame*

Um `GeoDataFrame` é uma **estrutura de dados tabular** que contém um `GeoSeries`.

A propriedade mais importante é a **coluna `GeoSeries` de geometria ativa**”.

A coluna de **geometria ativa** – *não importa o nome da `GeoSeries` correspondente* – pode ser **acessada** através do **atributo de geometria** (`gdf.geometry`), e o **nome da coluna de geometria** pode ser encontrado digitando `gdf.geometry.name` OU `gdf.active_geometry_name`.

Fonte: Developers, Geopandas. Geopandas Data Structures. Pydata. *online*. 2013. Disponível em: https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov. 2024.

Instalar biblioteca GeoPandas

#Jupyter Lab ou Jupyter Notebook:

#Instalando GeoPandas

```
pip install geopandas
```

#Importando a biblioteca

```
import geopandas
```

#Opções do GeoPandas:

```
geopandas.options
```

Fonte: Developers, Geopandas. Geopandas: Data Structures. Pydata. *online*. 2013. Disponível em: https://geopandas.org/en/stable/docs/user_guide/data_structures.html . Acesso em: Nov. 2024.