



escola
britânica de
artes criativas
& tecnologia

Profissão: Analista de dados



GLOSSÁRIO



Projeto: Análise Exploratória de Dados em Python



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

- **Conheça a Kaggle**
- **Analise o problema de negócio**
- **Explore dados**
- **Visualize dados**



Conheça a Kaggle



Conheça a Kaggle

Glossário

● Linguagem R

É uma linguagem de programação multi-paradigma orientada a objetos, programação funcional, dinâmica, fracamente tipada, voltada à manipulação, análise e visualização de dados.

● Notebook

Os notebooks do Colab permitem combinar código executável e rich text em um só documento, além de imagens, HTML, LaTeX.



Analise o problema de negócio



Analise o problema de negócio

Glossário

Latitude

É uma coordenada geográfica que indica a distância de um ponto em relação à linha do equador, medida em graus.

Longitude

É uma coordenada geográfica que indica a distância de um ponto em relação ao meridiano de Greenwich, medida em graus.

Pré-processamento

É uma etapa do processo de análise de dados, em que os dados brutos são processados e transformados em um formato mais adequado para a análise.



Explore dados



Explore dados

Glossário

• Estrutura nativa

É a estrutura de dados básica fornecida pela linguagem de programação, sem a necessidade de bibliotecas ou módulos adicionais. Por exemplo, em Python, as estruturas nativas incluem listas, tuplas, dicionários, conjuntos, entre outros.

• Esquema

É a estrutura organizacional dos dados, que define como os dados são armazenados, organizados e relacionados entre si.

• Subconjunto dos dados

É um conjunto de dados que é extraído de um conjunto maior de dados com a finalidade de analisar ou manipular apenas uma parte específica dos dados.



Visualize dados



Visualize dados

Glossário

● Coordenadas geográficas

As coordenadas geográficas são as coordenadas de latitude e longitude que definem a posição geográfica de um ponto na Terra. O Geopandas utiliza essas coordenadas para representar as geometrias espaciais em seus GeoDataFrames. As coordenadas geográficas são especificadas na ordem (longitude, latitude).

● Geopandas

Biblioteca do Python que permite a análise de dados geoespaciais. Combina as funcionalidades de Pandas (biblioteca para manipulação de dados em Python) com ferramentas de visualização e análise de dados geográficos. Para importar o Geopandas, basta usar o comando: **import geopandas as gpd**. O Geopandas trabalha com diferentes tipos de arquivos de dados geoespaciais, incluindo arquivos shapefile (.shp), arquivos GeoJSON (.json) e arquivos ESRI File Geodatabase (.gdb).



Visualize dados

Glossário

GeoDataFrame

É a estrutura de dados principal do Geopandas, que é semelhante a um DataFrame do Pandas, mas contém uma coluna adicional chamada "geometry" que armazena informações sobre as geometrias espaciais associadas a cada registro do DataFrame.

Índice do GeoDataFrame

Pode ser definido explicitamente por meio do método "set_index()", ou pode ser o índice padrão do DataFrame. O índice é importante para acessar os registros de maneira eficiente.



Bons estudos!

