Introdução à análise de dados com Python

A análise de dados com Python é uma habilidade essencial para aqueles que

desejam obter informações significativas de conjuntos de dados e tomar decisões

informadas. A Analítica de Dados com Python pode ser usada por analistas de

dados na exploração para escrever scripts e funções customizadas, manipular

os dados, otimizar fluxos de trabalho e gerar visualizações de dados. Além

disso, o Python está se tornando cada vez mais popular para desenvolver

código para aplicativos personalizados de aprendizado de máquina e análise

preditiva, sendo assim uma linguagem de programação poderosa e amplamente

utilizada.

“Para muitas pessoas, a linguagem de programação Python tem um forte apelo. Desde o seu surgimento em 1991, Python se tornou uma das linguagens de programação interpretadas mais populares, junto com Perl, Ruby e outras. Python e Ruby se tornaram especialmente populares por volta de 2005 para a construção de sites, com seus diversos frameworks web como Rails (Ruby) e Django (Python). Linguagens como essas com frequência são chamadas de linguagens de scripting, pois podem ser usadas para escrever rapidamente pequenos programas ou scripts para automatizar outras tarefas. Não gosto do termo “linguagem de scripting”, pois carrega uma conotação de que essas linguagens não poderiam ser usadas para a construção de softwares sérios. Entre as linguagens interpretadas, por diversos motivos históricos e culturais, Python desenvolveu uma comunidade grande e ativa de processamento científico e análise de dados. Nos últimos dez anos, Python passou de uma linguagem de computação científica inovadora, ou para ser usada “por sua própria conta e risco”, para uma das linguagens mais importantes em ciência de dados, aprendizado de máquina (machine learning) e desenvolvimento de softwares em geral, no ambiente acadêmico e no mercado”(Python Para Análise de Dados, Wes McKinney, 30 de dezembro de 2011)

## **Por que usar Python para análise de dados?**

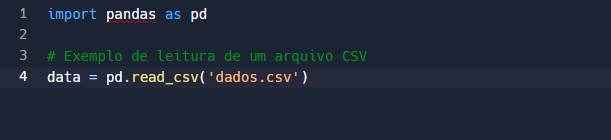
Python é uma escolha popular para análise de dados por várias razões:

* Linguagem de programação acessível: Python é conhecido por sua sintaxe simples e legível, o que o torna acessível até mesmo para iniciantes em programação.
* Ampla comunidade e suporte: Há uma comunidade ativa de desenvolvedores que oferece bibliotecas, tutoriais e recursos online para análise de dados em Python.
* Bibliotecas de análise de dados: Python oferece várias bibliotecas poderosas, como Pandas, NumPy, Matplotlib, Seaborn e Scikit-Learn, que simplificam tarefas comuns de análise de dados.

## **Passos básicos de análise de dados com Python:**

### **1. Coleta de dados**

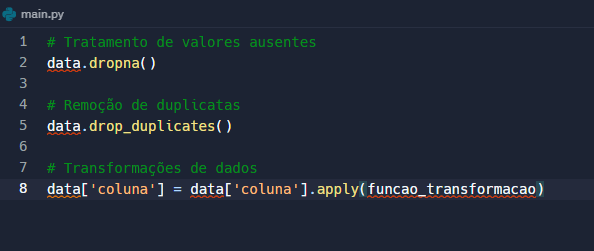
O primeiro passo na análise de dados é coletar os dados relevantes. Os dados podem ser provenientes de diversas fontes, como bancos de dados, arquivos CSV, APIs da web, planilhas, entre outros. Você pode usar bibliotecas Python, como Pandas, para importar e manipular esses dados.



FONTE: GRUPO

### **2. Limpeza e preparação de dados**

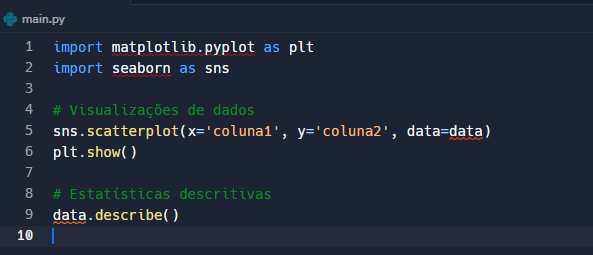
Os dados nem sempre estão em formato perfeito. É comum haver valores ausentes, duplicados ou inconsistentes que precisam ser tratados. Você também pode precisar fazer transformações nos dados para torná-los adequados para análise.

Copiar código

FONTE: GRUPO

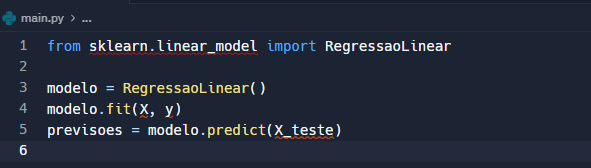
### **3. Exploração de dados**

Neste estágio, você explora os dados para obter insights iniciais. Isso inclui a criação de visualizações, estatísticas descritivas e a identificação de tendências e padrões nos dados.

Copiar códigoFONTE:GRUPO

### **4. Análise de dados**

A análise de dados envolve métodos estatísticos e algoritmos para responder a perguntas específicas ou fazer aplicação com base nos dados. Bibliotecas como Scikit-Learn são úteis para tarefas de aprendizado de máquina e modelagem preditiva.

Copiar código

FONTE:GRUPO

### **5. Comunicação de resultados**

Comunicar os resultados da análise é crucial. Isso pode ser feito por meio de relatórios, visualizações de dados interativos ou apresentações. Bibliotecas como Matplotlib e Seaborn podem ajudar a criar gráficos informativos.

Essa é apenas uma introdução à análise de dados com Python. À medida que você avança, você pode aprofundar seus conhecimentos e explorar tópicos avançados, como análise de séries temporais, aprendizado de máquina, processamento de linguagem natural (PNL) e muito mais. A prática é fundamental, então comece a trabalhar em projetos de análise de dados para aprimorar suas habilidades.

Introdução a biblioteca Pandas.

“O pandas oferece estruturas de dados de alto nível e funções, projetadas para fazer com que trabalhar com dados estruturados ou tabulares seja rápido, fácil e expressivo. Desde o seu surgimento em 2010, o pandas tem ajudado a viabilizar o Python como um ambiente eficaz e produtivo para análise de dados. Os principais objetos do pandas usados neste livro são o Data Frame – uma estrutura de dados tabular, orientada a colunas, com rótulos (labels) tanto para linhas quanto para colunas – e as Séries – um objeto array unidimensional, com rótulo.”(Python Para Análise de Dados, Wes McKinney, 30 de dezembro de 2011)

uma das bibliotecas mais importantes para entender como usar dados em Python principais tarefas, bem como utilizar bibliotecas para manter o processamento de gráficos digitais e usados. forneça uma breve introdução e execute uma biblioteca hoje.

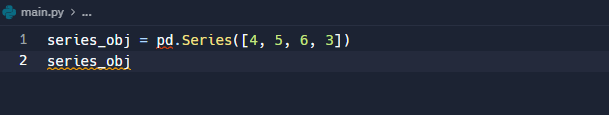
Uma estrutura de dados estrutura chamada Séries ou Data Frame pode ser criada e modificada usando a Biblioteca Panda .Uma série ou Data Frame pode ser criada e modificada usando a Biblioteca Panda .

séries é uma matriz de objetos unidimensional (como a matriz Numpy ) que contém um índice (os resultados das observações) . considerado como uma coleção de dados retangulares ou tabulares em que cada coluna contém um tipo específico de dados, denotado por um índice de colunas e um índice para cada observação (linha).

importar a biblioteca , aplicamos a seguinte convenção :

Séries

Uma matriz unidimensional vinculada a uma tabela de dados chamada "índice" ( ou "índice"). Podemos criar uma série usando o seguinte código .código.

FONTE:GRUPO

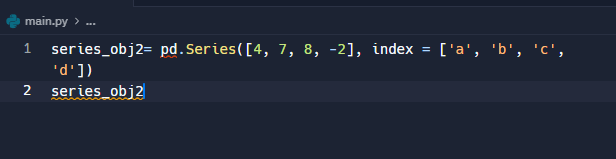
0 4

1 5

2 6

3 3

dtype: int64

Veja que existe um número representando a ordem das observações ao lado de um número valor, representando a ordem das observações ao lado de cada valor. possível criar os valores de um índice .FONTE:GRUPO

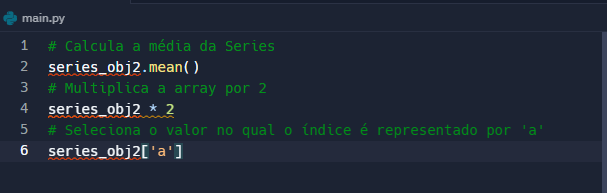
a 4

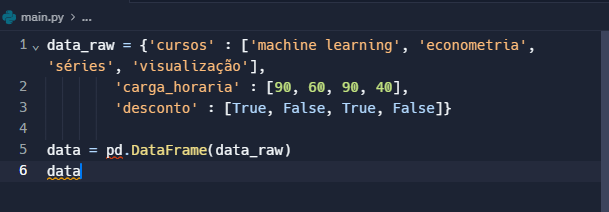
b 7

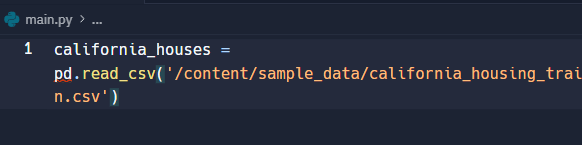
c 8

d -2

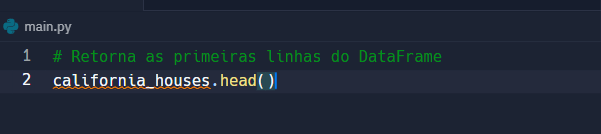
dtype: int64

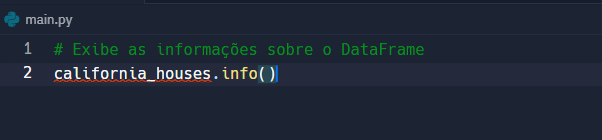
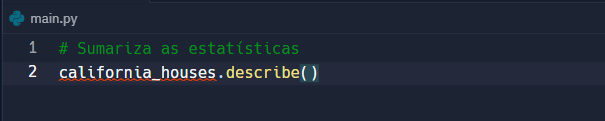
Objeto de série pode ser feito de forma semelhante a um array no Numpy, seja através de funções, multiplicandos escalonados , ou mesmo filtragens. A ordem do índice não muda. ser feito através da rotação da observação para permitir a seleção.FONTE:GRUPO

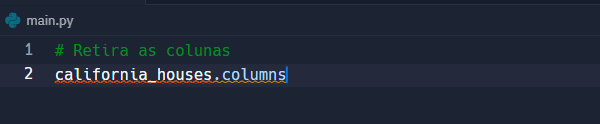
Como dito, o Data Frame simboliza uma tabela de dados retangular tabela de dados com uma coleção de colunas que são pré-preenchidas com uma série de observações , sejam elas numéricas, textuais, booleanas, etc.uma coleção de colunas pré -preenchidas com uma série de observações , sejam elas numéricas , textuais, booleanas , etc. Além disso, tem dois índices, um para observações e outro duas colunas, descreve-o como uma coleção de listas, dicionários ou matrizes unidimensionais. A partir de uma coleção de dicionários , construiremos um DataFrame.FONTE:GRUPO

Além disso , é possível importar dados de outros locais começando pelos pandas. Outros locais começam com pandas. O método mais popular método é importar.csv a importar arquivos.csv .archivos. Vemos um exemplo abaixo, importando um conjunto de dados de treinamento para uma coleção de dados sobre preços, um exemplo casas na Califórnia do Google Cola Baixo, importando um conjunto de dados de treinamento para coleta de dados sobre preços de casas na Califórnia do Google Colab.FONTE:GRUPO

Recebemos essa coleção de dados, o interessante a fazer é investigá-la e entender seu conteúdo. use uma variedade de métodos para fazer isso.

FONTE:GRUPO

FONTE:GRUPOFONTE:GRUPO

Também é possível tirar os números dos índices e os nomes das colunasFONTE:GRUPO

Depois, demos uma breve introdução aos pandas , mas também há uma série de possibilidades adicionais de manipulação e cálculo . Isso inclui filtragem de DataFrame mais sofisticada , agregação e fusão com condições, bem como métodos para configurar um Data Frame para modelagem e visualização.

Introdução a Scikit Learning

Sua construção é baseada na interação com outras bibliotecas Python, numéricas e científicas, como NumPy e SciPy e Matplotlib, e é considerada uma das soluções mais versáteis e populares do mercado.

Para ajuste, seleção e avaliação de modelo , bem como pré-processamento de dados , Scikit-learn é considerada a biblioteca mais prática e confiável para treinamento de aprendizado de máquina (ML) em Python .

biblioteca de alto nível , permite a definição de modelos de dados preditivos usando apenas algumas linhas de código . Nesse sentido, o scikit-learn é bem documentado , comparativamente simples de aprender e útil se você estiver procurando uma introdução ao aprendizado de máquina .

O Scikit-Learn foi desenvolvido como Google Summer of \_projeto Projeto de código e foi idealizado e foi idealizado pelo cientista de dados David Carnapau.pelo cientista de dados David Carnapau. O o nome da biblioteca é derivado é derivado da frase " SciKit" (um SciPy Toolkit ), de extensão de terceiros do SciPy que é vendida separadamente.a frase "SciKit" (um SciPy Toolkit), uma extensão de terceiros do que é vendido separadamente.

Quais são as principais aplicações do Scikit Learn?

O Scikit-learn fornece uma ampla gama de algoritmos integrados que algoritmos podem ser usados ​​para utilizar totalmente projetos de ciência de dados .Usado para utilizar totalmente projetos de ciência de dados . Listamos abaixo seus principais pontos abaixo, juntamente com a finalidade de cada um;

### 1. Algoritmos de Classificação (Classification)

2. Algoritmos de Regressão

3. Algoritmos de Agrupamento (Clustering)

4. Redução de dimensionalidade (Dimensionality reduction)

5. Seleção de Modelo (Model selection)

6. Pré-processamento (Preprocessing)

As ferramentas de pré-processamento do scikit-learn são cruciais para extrair e normalizar recursos durante a análise de dados crucial para extrair e normalizar recursos durante a análise de dados .

Essas ferramentas podem ser usadas no contexto da ciência de dados para modificar dados de entrada ,da ciência de dados como modificar dados de entrada , como texto, e aplicar seus recursos à análise .texto e aplicar seus recursos à análise.

importantes da biblioteca recursos Scikit-aprender

Você aprenderá sobre algumas das funções básicas que o pacote oferece para desenvolver modelos confiáveis ​​e eficazes : algumas das funções básicas que. O pacote fornece para desenvolver modelos de ML confiáveis ​​e eficazes.

O que o Scikit-Learn faz?

Scikit-learn é um grátis,biblioteca biblioteca de código aberto para aprendizado de máquina Python (aprendizado de máquina ) .para aprendizado de máquina Python ( aprendizado de aprendizado de máquina ). variedade de ferramentas eficazes para mineração de dados , análise e modelagem estatística , além de suporte para aprendizagem supervisionada e não supervisionada .

Uma biblioteca de aprendizado de máquina.

Vamos listar algumas:

1. NumPy

2.Pandas

3.SciPy

4.Stats Models

5.Matplotlib

6.Seaborn

7.Plotly

8.Bokeh

9.Pydot

10.Dash

O que é visualização de dados ,visualização exatamente?

visualização de dados o processo de representação de dados com elementos visuais , como elementos diagramas,como diagramas, gráficos ou mapas.gráficos ou mapas. Ela converte dados complicados, de grande volume ou numéricos em uma representação visual mais fácil de processar . visualizar As ferramentas ferramentas de dados melhoram e automatizam o processo de comunicação visual para obter precisão e detalhes finos .melhorar e automatizar o processo de comunicação visual para obter precisão e detalhes finos . representações podem ser usadas para extrair insights acionáveis ​​de dados brutos .

Por que a visualização de dados é tão crucial?

A maioria das empresas modernas empresas processo processar enormes quantidades de dados de muitas fontes de dados , como as seguintes :enormes quantidades de dados de muitas fontes de dados , como as seguintes :

e externo sites internos sites

dispositivos com inteligência

sistemas internos de coleta de dados sistemas de coleta

meios de comunicação Redes sociais

No entanto, os dados brutos podem ser difíceis de compreender e usar.Pode ser difícil de entender e usar. Portanto, os cientistas prepararam e apresentaram os dados no contexto apropriado .Preparou e apresentou os dados no contexto apropriado . Eles fornecer um formato visual para que formato tomadores de decisão para que os tomadores de decisão possam ver como as relações entre os dados se relacionam com padrões ou tendências ocultas .pode ver como as relações entre os dados se relacionam com padrões ou tendências ocultas . a visualização se desenvolve de dados narrativas que apoiam a inteligência corporativa , a tomada de decisões e o planejamento estratégico com base emo planejamento dados.baseado em dados.

Referencias

https://www.usandopy.com/pt/artigo/python-pandas-introducao-a-biblioteca-pandas (788 termos)

https://awari.com.br/aprenda-tudo-sobre-ux-design-com-figma-dicas-e-truques-para-se-tornar-um-especialista-2

https://www.voitto.com.br/blog/artigo/principais-bibliotecas-python

https://www.hashtagtreinamentos.com/bibliotecas-mais-importantes-do-python

https://harve.com.br/blog/programacao-python-blog/numpy-python-o-que-e-vantagens-e-tutorial-inicial

https://analisemacro.com.br/data-science/python/introducao-ao-pandas/

https://reviews.tn/pt/top-meilleurs-sites-genealogie-gratuits

http://computacaointeligente.com.br/outros/intro-sklearn-part-1

https://didatica.tech/a-biblioteca-scikit-learn-pyhton-para-machine-learning

https://dramoniquevalois.com.br