Programação Orientada a Objetos para Dados

Programação Orientada a Objetos e o Universo de dados

Prof. Ms. Leonardo Rocha

- Unidade de Ensino: 1
- Competência da Unidade: Compreensão sobre o paradigma de programação de orientação a objetos.
- Resumo: Programação orientada a objetos aplicada à ciência de dados.
- Palavras-chave: POO, ciência de dados, herança.
- Título da Teleaula: Programação orientada objetos e o universo de dados.
- Teleaula nº: 1

Contextualização

Conhecendo a Programação Orientada a Objetos
Ferramentas, Bibliotecas e habilidades técnicas
Estrutura de repetição e condição em Python
Biblioteca Pandas

Habilidades e competências

Habilidade x Competência

A **Habilidade** é o *saber fazer*

A **Competência** é a coordenação desse *saber fazer* que produz algum tipo de benefício ou resultado.

Exemplo de habilidade:

Trabalho em equipe

Exemplo de competência:

Coordenar equipe de desenvolvimento garantindo as entregas em tempo hábil.

Habilidade x Competência

A **Habilidade** é o *saber fazer*

A **Competência** é a coordenação desse *saber fazer* que produz algum tipo de benefício ou resultado.

Exemplo de habilidade:

Trabalho em equipe

Exemplo de competência:

Coordenar equipe de desenvolvimento garantindo as entregas em tempo hábil.

Ciência de Dados

A ciência de dados exige habilidades e competências pois lida com dados, sua análise, sistematização e conversão em informação.

Analista de dados - Programação, fundamentos em estatística, manipulação, tratamento e visualização de dados.

Ciência de Dados

Engenheiro de dados - responsável para lidar com volume grande de dados. Deve esxtrair diferentes tipos de dados, tratá-los e armazená-los em BDs.

Programação, manipulação de dados e engenharia de software.

Ciência de Dados

Cientista de dados - desenvolve modelos preditivos com auxílio de algoritmos e estatística para identificar tendências através da identificação de padrões.

Programação, estatística, Machine Learning.

Habilidades de ciência de dados

HABILIDADES	ANALISTA DE DADOS	ENGENHEIRO DE DADOS	CIENTISTA DE DADOS	ENGENHEIRO DE MACHINE LEARNING
PROGRAMAÇÃO	✓	✓	✓	✓
ESTATÍSTICA	✓	✗	✓	✗
MANIPULAÇÃO DE DADOS	✓	✓	✓	✓
VISUALIZAÇÃO DE DADOS	✓	✗	✓	✗
MACHINE LEARNING	✗	✗	✓	✓
ENGENHARIA DE SOFTWARE	✗	✓	✗	✓

Ferramentas tecnológicas

Perfil em rede social profissional é fonte de busca. Linkedin é uma rede social necessária.

Além disso, ter um repositório (portfólio) com os principais projetos é uma forma de mostrar o trabalho já desenvolvido, a experiência adquirida e a qualidade técnica do que se faz. Github, Gitlab, Bitbucket e outros são exemplos.

Programação para ciência de dados

Onde entra a programação?

A programação é fundamento base, assim como manipulação de dados, para atuação na área de ciência de dados. Note que a programação é base, inclusive, para a manipulação de dados. Portanto, programar é crucial para atuar nessa área.

Paradigmas de programação

Paradigma de programação, a grosso modo, é a forma de escrita (estrutura) de um determinado código, desde que ele a aceite. Existem linguagens que aceitam mais de um paradigma de programação. Um deles, a programação orientada a objetos apresenta como uma das principais características: a reutilização de código. Isso é possível, por exemplo, pela construção de classes.

Outro exemplo é o paradigma estruturado tem como principal característica seguir sequência.

POO

A programação orientada a objetos tem como pilar a:

1 - Abstração
2 - Encapsulamento
3 - Herança
4 - Polimorfismo

Abstração

Abstração significa definir e focar no que é necessário no desenvolvimento

Sistema de notas de alunos

Dados do aluno

Características físicas do aluno?

Encapsulamento

Detalhes interno do funcionamento dos métodos de uma classe permaneça oculto para os objetos, ou seja, é possível esconder propriedades e métodos da classe, impedindo mudanças nos atributos.

Herança

Herança é o termo que representa a relação entre classes diferentes no que tange à reutilização de código. Esse conceito representa a possibilidade de criar nova classes que podem herdar atributos e métodos daquela que passa a ser conhecida como classe pai. Esse movimento de herdar permite realizar alterações que atribuem novas características.

Polimorfismo

Diz respeito à capacidade da linguagem de programação de processar objetos de formas diferentes dependendo do seu tipo de dado ou classe. A vantagem é ter métodos com o mesmo nome mas que são implementados de formas diferentes. Exemplo: Classe conexão com método abrirConexao() será o mesmo para as classes filhas ConexãoOracle e ConexãoPostgree mas a implementação será diferente.

Python

Rank	Language	Score
1	Python	100.0
2	Java	95.4
3	C	94.7
4	C++	92.4
5	JavaScript	88.1
6	C#	82.4
7	R	81.7
8	Go	77.7
9	HTML	75.4
10	Swift	70.4

Fonte: https://spectrum.ieee.org/top-programming-languages-2021

Como pensar a POO para ciência de Dados?

Solução

Como veremos adiante, a Programação Orientada a Objetos é crucial para a utilização na ciência de dados. Isso porque é super importante utilizar de características da Orientação a Objetos para consulta e análise de dados. Por exemplo, o elemento e escrita de código para sua reutilização não só contribui para otimização do tempo mas também para aproveitar algum código que já tenha sido escrito e outras situações que exigem tal código.

Dúvidas

Bibliotecas

Bibliotecas

Uma biblioteca de software é uma coleção de subprogramas, ou seja, funções ou métodos que são compartilhados e que já estão prontas para o uso. Elas provêm serviços a programas independentes, permitindo o compartilhamento e alteração de código e dados de forma modular.

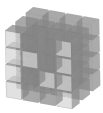
Pandas

Muito popular e uma das mais utilizadas, oferece ao programador uma forma de trabalhar com análise e estruturas de dados de forma siplificada.

pandas

NumPy

Biblioteca que suporta processamento de multi-dimensionais arranjos e matrizes e conta com um grande conjunto de funções matemáticas e métodos para este fim.



NumPy

Matplotlib

Excelente para ser utilizada no trabalho com visualização de diagramas e gráficos.

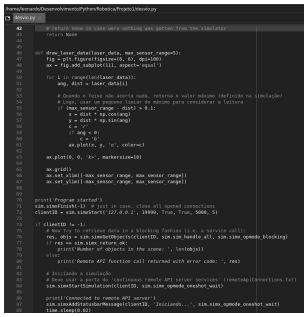
Características do Python

Características do Python

É uma linguagem com sintaxe simples (forma como o código é escrito).

A Indentação é levada a sério no Python pois é ela que determina a execução do código. Sem indentação, o código apresentará comportamentos diferentes daqueles projetados pelo programador.

Indentação



Fonte: autor

Importância da indentação na programação com Python na prática

par ou impar.py

```
#EXEMPLO DE CÓDIGO
#Teste de Indentação
num = int(input("Digite um número: "))
if(num % 2 == 0):
    print("Número par.")
else:
    print("Número impar.")
```

Sugestão de leituras

Estrutura de dados Python e biblioteca Pandas

Estrutura de dados nativas

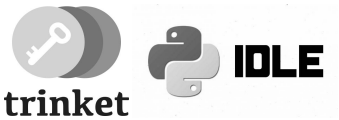
Listas: Podem guardar qualquer tipo de dados ao mesmo tempo pois é considerada estrutura heterogênea e podem ser mutáveis. Escrita com [ ]

Tuplas: As tuplas diferem das listas por serem imutáveis. Elas não podem ser alteradas. Escrita com ( ) ou sem

Dicionários: Coleção de objetos armazenado em pares de chaves e valor separados por ":" e os pares de chave/valor são separados por ",".

fonte: https://bit.ly/3A7koEz

Conhecendo Trinket.io e Idle



Prática com Idle

Listas

Tupla

Dicionário

Estruturas no Trinket

Condição

Repetição

Try e Except

Instalação do pandas

Para instalar o pandas, existe uma documentação que mostra os passos a serem seguidos. Na documentação é possível ver como realizar a instalação no Windows e no Linux.

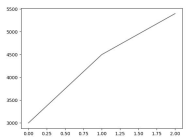
Fonte: https://pandas.pydata.org/docs/getting_started/install.html

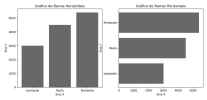
Criando dataFrame com dicionário

```
>>> import pandas as pd
>>> ds = pd.read_excel('~/Documentos/Profissional/Kroton/Disciplinas/ProgramacaoOrientadaAobjetosParaDados/exemplos/pandas/dados1.xlsx')
>>> ds
    cod_pedido  valor_venda  quantidade_de_itens
0            1       2500.0                    5
1            2        130.0                    8
2            3        960.0                    4
3            4        560.0                    2
4            5       3450.0                    2
5            6        400.0                   15
6            7        900.0                   98
7            8       1500.0                   98
8            9        990.0                   18
9           10        600.0                    5
10          11        100.0                   34
>>> ds['valor_venda'].sum()
12098.9
>>> valor_vendas={'vendedor':['Leonardo', 'Paula', 'Fernanda'], 'valor':[3000,4500,5400]}
>>> df = pd.DataFrame(valor_vendas)
>>> df
   vendedor  valor
0  Leonardo   3000
1     Paula   4500
2  Fernanda   5400
>>>
```

Plotando gráficos

A plotagem de gráficos é feita pela biblioteca Matplotlib.





Documentação oficial Pandas

A documentação oficial está repleta de informações e exemplos para prática e compreensão da biblioteca Pandas e sua adoção na análise de dados. Vamos conhecer a documentação:

https://pandas.pydata.org/docs/getting_started/

Como criar código em Python para escrita sequência de Fibonacci?

Solução

```
#############################################################
#Exercício da SP1 da primeira aula                          #
#Construir código para escrita da sequência Fibonacci      #
#Autor: Leonardo Rocha                                      #
#############################################################
print("ALGORITMO PARA ESCRITA DE SEQUÊNCIA FIBONACCI")
qtde=int(input("Informe quantos números Fibonacci você deseja: "))
valor1 = 1
valor2 = 1
sequencia = [1,1]
for count in range(2,qtde):
    valor = sequencia[count -1] + sequencia[count - 2]
    sequencia.append(valor)

print(sequencia)
#ou imprimindo com quebra de linha
#[print(numero) for numero in sequencia]
```

Dúvidas

Recapitulando

Habilidades e Competências

Programação orientada a objetos aplicada à ciência de dados

Bibliotecas em Python

estrutura de código em Python