

Curso de Python do ZERO AO DS

com Meigarom do canal “Seja Um Data
Scientist”

Instagram: @meigarom.datascience

LinkedIn: <https://www.linkedin.com/in/meigarom/>

Telegram: <https://t.me/sejaumdatascientist> (**GRUPO DE ESTUDOS**)

Aula 01 - Começando com o Python

Agenda:

- 1.0. O problema de negócio.
- 2.0. Planejamento de solução.
- 3.0. O que Python?
- 4.0. Como escrever código Python?

- 5.0. Os primeiros comandos Python?
- 6.0. Responder as questões de negócio.
- 7.0. Exercícios

1.0. O problema de negócio.

<https://sejaumdatascientist.com/os-5-projetos-de-data-science-que-fara-o-recrutador-olhar-para-voce/>

- Empresa: House Rocket
- O que a empresa faz:
 - House Rocket plataforma de compras e vendas de imóveis
- Qual o problema ?
 - O CEO da empresa gostaria de maximizar o lucro da empresa encontrando bons negócios.
- A principal estratégia ?
 - Fontes externas para encontrar bons negócios.

- As perguntas do CEO da House Rocket:

1. Quantas casas estão disponíveis para compra?
2. Quantos atributos as casas possuem? (numero de quartos, numero de garagens, m2, vista pro mar)
3. Quais são os atributos?
4. Qual a casa mais cara do portfólio (casa com maior valor)?
5. Qual a casa com o maior número de quartos?

2.0. Planejamento da solução

2.1. Planejamento do PRODUTO FINAL.

1. O que eu vou entregar? (Planilha, Texto, Email, Modelos de ML, ...)

R: Texto com as respostas.

2. Como vai ser a entrega?

R: Perguntas | Respostas

Por exemplo:

1. Quantas casas estão disponíveis para compra?

R: 2300 imóveis disponíveis para compra

2. Quantos atributos as casas possuem? (numero de quartos, numero de garagens, m2, vista pro mar)

R: 10 atributos

3. Quais são os atributos?

4. Qual a casa mais cara do portfólio (casa com maior valor)?

5. Qual a casa com o maior número de quartos?

2.2. Planejamento do PROCESSO.

1. Onde está a informação? (Excel, BD, API, manual)

R: <https://www.kaggle.com/harlfoxem/housesalesprediction>

2. Como coletar essas informações? (SQL, Python, Streamlit, ...)

R: Download (Apertar o botão)

3. Responder as perguntas?

1. Quantas casas estão disponíveis para compra?

R: Contar o numero de linhas do conjunto de dados

2. Quantos atributos as casa possuem? (numero de quartos, garagens, m2, vista pro mar)

R: Contar o numero de colunas do conjunto de dados

3. Quais são os atributos?

R: Mostrar o nome das colunas (automatica)

4. Qual a casa mais cara do portfólio (casa com maior valor)?

R: Ordenar as linhas pela coluna de preco (atributos)

5. Qual a casa com o maior número de quartos?

R: Contar o numero de linhas pela coluna numero de quartos (atributo)

2.3. Planejamento do FERRAMENTAS.

1. Quais ferramentas eu posso usar?

- Excel

- Facil de usar
- Barato
- Muito usado pelos times não técnicos.

- Poder de processamento limitado (1MM)

- Linguagem de programação

- Desenvolvida para criar software
- Processar e Manipular -> Insights

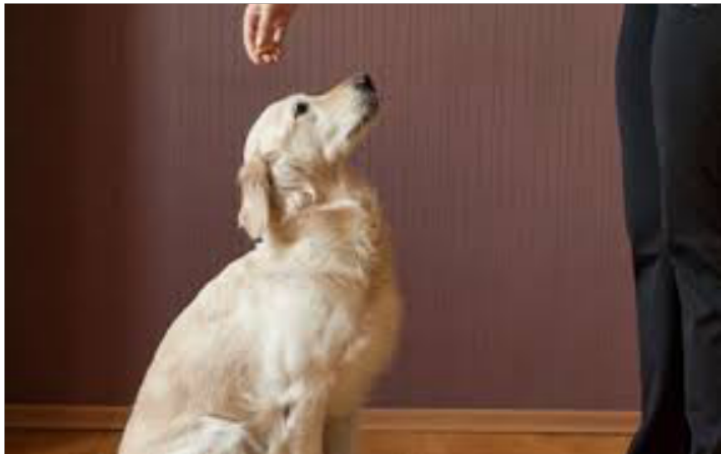
- Escalável
- Modularização

3.0. O que é Python?

1. Linguagem de Programação?

Python é uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum. É uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum. É uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum.

Exemplo 01: Treinando seu Doguinho pidão



Doguinho Pidão (Kika)

Traduzindo para a programação:

Exemplos de Comandos

Python é uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum. É uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum. É uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum.

Python é uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum. É uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum. É uma linguagem de programação de alto nível, interpretada de código, criada por Guido van Rossum.

Para responder as perguntas do CEO, seguinte comandos:

Meigarom -> Contar o numero de linhas do conjunto de dados

Meigarom -> Contar o numero de colunas do conjunto de dados

Meigarom -> Mostrar o nome das colunas (automatica)

Meigarom -> Ordenar as linhas pela coluna de preco (atributos)

Meigarom -> Ordenar as linhas pela coluna de numero de quartos (atributos)

4.0. Como escrever código Python?

Editor de Texto

- Bloco de notas
- Gedit
- TextEdit

Interpretador (Comandos -> Linguagem de máquina (binários) -> processador)

- Python 2.x (2.0, 2.1, 2.2, ..., 2.9)
- Python 3.x (3.0, 3.1, 3.2, ..., 3.8, 3.9)

IDE (Interface Development Environment) - Ambiente de Desenvolvimento

IDEs mais famosas para Python

- PyCharm
- Spyder
- VSCode (Visual Studio Code)
- Jupyter Lab

4.1. Instalando PyCharm

4.1.1. Acesse o link da JetBrains

<https://www.jetbrains.com/pt-br/pycharm/download/#section=windows>

4.1.2. Escolha a sua plataforma

1. Windows
2. Linux
3. Mac

4.1.3. Faça o Download da Versão Free - Versão Community

4.2. Por dentro do Pycharm

1. Clique em “New Project”.
2. **Location:** C:\Users\meigarom.lopes\repos\aula_python_zer
o_ao_ds
3. **New environment using:** Virtualenv
4. Create a main.py welcome script

1. O Editor de Texto
2. O Interpretador do Python
3. O Gerenciador de Pastas
 - Definir a pasta do projeto
 - Definir a pasta onde ficarão os dados
 - Definir a pasta onde ficará os arquivos com os códigos.

5. Os primeiros comandos em Python

Executar o PROCESSO

1. Onde está a informação? (Excel, BD, API, manual)

R: <https://www.kaggle.com/harifoxem/housesalesprediction>

2. Como coletar essas informações? (SQL, Python, Streamlit, ...)

R: Download (Apertar o botão)

Para responder as perguntas do CEO, seguinte comandos:

Meigarom -> Contar o numero de linhas do conjunto de dados
Meigarom -> Contar o numero de colunas do conjunto de dados
Meigarom -> Mostrar o nome das colunas (automatica)
Meigarom -> Ordenar as linhas pela coluna de preco (atributos)
Meigarom -> Ordenar as linhas pela coluna de numero de quartos (atributos)

Como Eu sei qual o comando para contar o numero de linhas?

5.1. A estrutura de um comando

Humano -> Comando Escrito -> computador -> realizar uma ação -> mostrar um resultado

Exemplo: “Como fazer um bolo de chocolate”

Ingredientes:

1. Ovos

2. Chocolate em pó
3. Farinha de trigo
4. Fermento
5. Leite

Modo de Preparo:

1. Adicione os ovos no liquidificador.
2. Adicione o chocolate em pó no liquidificador.
3. Adicione a farinha de trigo no liquidificador.
3. Bata por 5 minutos
4. Adicione o fermento.
5. Misture os Ingredientes.
6. Leve ao Forno

Modo de Preparo -> Manipula -> Ingredientes -> Bolo de Chocolate

Quero abrir um loja de bolo e ensinar para outras pessoas como fazer bolo?
Ingredientes + Modo de Preparo = Receita

Quero abrir um loja de bolo e ensinar para outras pessoas como fazer doces?

Bolo de chocolate

Churros

Doces

Receita de bolo + Receita de doce = Livro de Receitas

Resumindo:

Livro de Receita Organiza

Receita

Receita

Replica

Modo de Preparo

Modo de Preparo Manipula

Ingredientes

Em programação:

Biblioteca Organiza

Funções

Funções Replica

Comandos

Comandos Manipula Variáveis

#pythonehreceitadebolo

5.2. Estruturas

Em programação:

Biblioteca Organiza

Funções

Funções Replica

Comandos

Comandos Manipula

Variáveis

Variáveis:

- Caixa armazenadora
 - Tem nome
 - Guarda um tipo de objeto
 - Guarda objetos diferentes ao mesmo tempo

Exemplo:

- Caixa armazenadora

- nome: caixa_calcas
- armazeno: calças

- Posso guardar sapatos dentro da caixa_calcas? Não
- Posso guardar saia dentro da caixa_calcas? Não

Em programação:

- Caixa armazenadora é um espaço de memória que guarda um objeto
- Nome/ Tipo (não usa dialetos latinos (~, ç,), não usa espaço)

Funções:

- Sequencia de Comandos
 - Tem um nome
 - Pode ou não receber parametros de entrada

- Pode ou não retornar parametros de entrada
- “O que acontece na função, fica na função”

Em programação:

- A função é uma sequencia de comandos
- O nome da função tem que ser o mais próximo de sua responsabilidade
- Função ela pode ou não receber/retornar valores

Biblioteca

- A biblioteca é um agrupamento de funções
- Pode ter um apelido “carinhoso”
- Precisa ser importado

“emprestada”

6. Respondendo as perguntas do CEO

- Perguntas do CEO para você:

1. Quantas casas estão disponíveis para compra?
2. Quantos atributos as casas possuem?
3. Quais são os atributos das casas?
4. Qual a casa mais cara (casa com o maior valor de venda)?
5. Qual a casa com o maior número de quartos?

Conte o **quantidade de linhas** do conjunto de dados

Conte a **quantidade de colunas** do conjunto de dados

Mostre o nome das colunas do conjunto de dados

Ordene o conjunto de dados pelo preço das casas

Vamos fazer no PyCharm!!

7. Exercícios

- Novas Perguntas do CEO para você:

1. Quantas casas estão disponíveis para compra?
2. Quantos atributos as casas possuem?
3. Quais são os atributos das casas?
4. Qual a casa mais cara (casa com o maior valor de venda)?
5. Qual a casa com o maior número de quartos?
6. Qual a soma total de quartos do conjunto de dados?
7. Quantas casas possuem 2 banheiros?
8. Qual o preço médio de todas as casas no conjunto de dados?
9. Qual o preço médio de casas com 2 banheiros?
10. Qual o preço mínimo entre as casas com 3 quartos?
11. Quantas casas possuem mais de 300 metros quadrados na sala de estar?
12. Quantas casas tem mais de 2 andares?
13. Quantas casas tem vista para o mar?
14. Das casas com vista para o mar, quantas tem 3 quartos?
15. Das casas com mais de 300 metros quadrados de sala de estar, quantas tem mais de 2 banheiros?

Agenda:

1. O problema de negócio
(desafio)

2. Planejamento da solução.
3. O que é Python?
4. Como escrever códigos em Python?
5. Os primeiros comandos em Python?
6. Responder as perguntas de negócio
7. Exercícios (Telegram)

1. O problema de negócio
(desafio)

- Empresa: House Rocket
(fictícia)

- O que ela faz:

House Rocket compra imóveis por um preço baixo e revende por um preço alto através de uma plataforma online de compra e venda de imóveis.

- Principal estratégia:

House Rocket compra imóveis por um preço baixo e revende por um preço alto
Lucro = compra - venda

- Perguntas do CEO pra

Você:

1. Quantas casas estão disponíveis para compra?
2. Quantos atributos as casa possuem?
(Tamanho, preço, m2, quartos, ...)?
3. Quais são os atributos para a casa?
4. Qual a casa disponível é a mais cara
(casa de maior valor de venda)?
5. Qual a casa com o maior número de quartos?

2. Planejamento da solução.

2.1. Planejamento do **PRODUTO FINAL** (**output**)

1. O que eu vou entregar? (planilha, email, modelo de ML, ...)

R: Eu vou enviar um text por email.

2. Como será o texto?

R: Perguntas | Respostas

Por exemplo:

1. Quantas casas estão disponíveis para compra?

R: 240 casas disponíveis

2. Quantos atributos as casa possuem?
(Tamanho, preço, m2, quartos, ...)?

R: 30 atributos

2.2. Planejamento do **PROCESSO**.

1. Onde está essa informação? (Site, Excel, bando de dados, API)

R: <https://www.kaggle.com/harlfoxem/housesalesprediction>

2. Como conseguir esses dados? (SQL, Python, Download, pedindo pra alguém)

R: Baixar o site

3. Como responder as perguntas?

1. Quantas casas estão disponíveis

para compra?

R: **Contar a quantidade de linhas** do conjunto de dados

2. Quantos atributos as casa possuem?
(Tamanho, preço, m2, quartos, ...)?

R: **Contar a quantidade de colunas** do conjunto de dados

3. Quais são os atributos para a casa?

R: **Mostrar o nome das colunas** do conjunto de dados

4. Qual a casa disponível é a mais cara
(casa de maior valor de venda)?

R: **Ordenar** o conjunto de dados pelo preço das casas

5. Qual a casa com o maior número de quartos?

R: **Ordenar** o conjunto de dados pelo número.

2.3. Planejamento do **FERRAMENTA**.

1. Qual/Quais ferramentas eu posso usar?

- Excel:
 - Poder de processamento limitado.
 - Não é escalável
 - Gerenciamento complexo
 - Limitado para grandes volumes de dados (Big data VVV (Volume, Velocidade, Variação))
- Linguagem de Programação:
 - Originalmente criada para o desenvolvimento de software
 - Processamento, Análise e Visualização de Dados.
 - Mais complexa.
 - Escalável (poder de processamento depende de muitos computadores)
 - Facil gerenciamento (Git)

3. O que é Python?

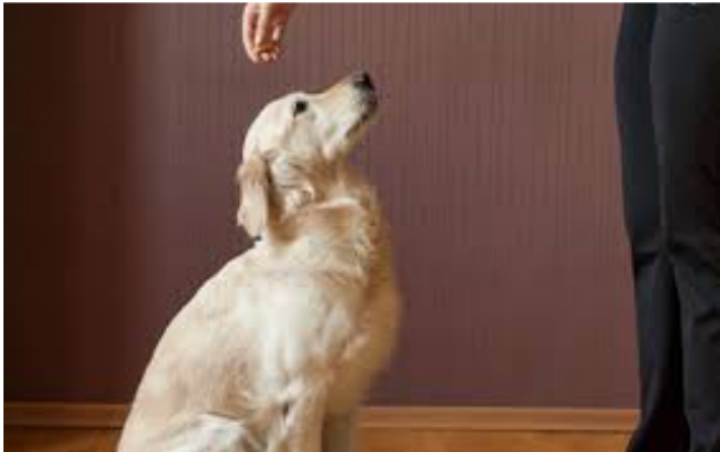
"ferramenta robusta para processamento de dados"

1. O que é uma linguagem de programação?

“A linguagem de programação é um método padronizado, formado por um conjunto de **regras sintáticas** e semânticas, de implementação de um **código fonte** - que pode ser **compilado** e transformado em um **programa de computador**,^[1] ou usado como **script interpretado** - **que informará instruções de processamento ao computador**.^[2] [Nota 1] Permite que um programador especifique precisamente quais os dados que o

computador irá atuar, como estes dados serão armazenados ou transmitidos e, quais ações devem ser tomadas de acordo com as circunstâncias. Linguagens de programação podem ser usadas para expressar **algoritmos** com precisão.” (Wikipedia)

1. Exemplo 01: Treinamento o seu Doguinho pidão



Doguinho Pidão (Kika)

Humano -> **Comandos visuais/verbais/escritos** -> **Doguinho Pidão** -> Realizar um Ação -> Gere um resultado

Meigarom -> “**Senta**” -> **Kika** -> Fica sentada -> Meigarom veria a Kika sentada

Na linguagem de programação:

Humano -> **Comandos Escritos** -> **Computador** -> Realizar um Ação ->

Mostra um resultado

Meigarom -> Seleciona a coluna X -> Compilador/Interpretador -> Gera um instrução de máquina (binário) -> O processador executa -> Mostra um resultado

Exemplos de Comandos:

Meigarom -> Seleciona a coluna X, Y

Meigarom -> Seleciona 2 colunas? [NÃO FUNCIONA]

Meigarom -> data[[X, Y]]

Meigarom -> Ordena as linhas através da coluna Z

Meigarom -> data.sort_values(Z)

Meigarom -> Ordena as linhas através da coluna Z do maior para o menor

Meigarom -> data.sort_values(Z, ascending=False)

Meigarom -> Seleciona todas as linhas, no qual a coluna W é maior que 10

Meigarom -> data[data[W] > 10]

Para responder as perguntas do CEO, temos que usar os seguintes comandos:

Meigarom -> “Conta a quantidade de linhas do conjunto de dados”

Meigarom -> “Mostra o nome das colunas do conjunto de dados”

Meigarom -> “Ordene o conjunto de dados pelo preço das casas”

4. Como escrever códigos em Python?

Editor de Texto:

- Bloco de Notas (Windows)
- Gedit (Linux)
- TextEdit (MacOS)

Interpretador (código -> linguagem de máquina -> executado pela CPU)

- Python 2x - 2.0, 2.1, ..., 2.9
- Python 3x - 3.0, 3.1, ..., 3.9 (beta)

IDE (Interface Development Environment) - Ambiente de Desenvolvimento

- Editor de Texto + Interpretador do Python + Git + SQL Editor + Ambiente Virtual

IDE mais famosas de Python

- PyCharm
- Spyder
- VSCode (Visual Studio Code)

4.1. Instalando o PyCharm

4.1.1. Acesse o site da JetBrains

<https://www.jetbrains.com/pt-br/pycharm/download/#section=windows>

4.1.2. Acesse o site da JetBrains

1. Windows
2. Linux
3. Mac

4.1.3. Faça o Download da Versão Free - Versão Community

4.2. Por dentro do Pycharm

1. Clique em “New Project”.
2. **Location:** C:
\Users\meigarom.lopes\repos\aula_python_zer
o_aos
3. **New environment using:** Virtualenv
4. Create algum texto com o botão direito ->

arquivo em Python

1. O Editor de Texto
2. O Interpretador do Python
3. O Gerenciador de Pastas

- Definir a pasta do projeto
- Definir a pasta onde ficarão os dados
- Definir a pasta onde ficará os arquivos com os códigos.

5. Os primeiros comandos em Python?

A execução do PROCESSO:

1. Onde está essa informação? (Site, Excel, bando de dados, API)

R: <https://www.kaggle.com/harlfoxem/housesalesprediction>

2. Como conseguir esses dados? (SQL, Python, Download, pedindo pra alguém)

R: Baixar o site

Para responder as perguntas do CEO, temos

que usar os seguintes comandos:

Meigarom -> “Conta a quantidade de linhas do conjunto de dados”

Meigarom -> “Mostra o nome das colunas do conjunto de dados”

Meigarom -> “Ordene o conjunto de dados pelo preço das casas”

Qual é o comando para: “Conte a quantidade de linhas do conjunto de dados”

5.1. A estrutura de um comandos

Humano -> Comandos Escritos -> Computador -> Realizar um Ação -> Mostra um resultado

Exemplo: “Como fazer um bolo de chocolate”

Ingredientes:

1. Ovos
2. Chocolate em pó
3. Farinha de Trigo
4. Fermento
5. Leite

Modo de Preparo:

1. Adicione os ovos na batedeira
2. Adicionar o chocolate em pó na batedeira
3. Adicione a farinha de trigo no liquidificador.

4. Bata por 5 minutos
5. Adicione o fermento.
6. Misture os Ingredientes.
7. Leve ao Forno

Modo de Preparo -> Manipula -> Ingredientes -> Bolo de Chocolate

Quero ensinar outras pessoas a fazer o meu bolo de chocolate:

Ingrediente + modo de preparo = Receita

Eu tenho várias receitas de outros bolos e também de doces

Receita de Bolo de Chocolate

Receita de Bolo de Churros

Receita de Doce de Leite

Bolo de Chocolate + Churros + Banana = Livros de Receitas:

Resumindo:

Livro de Receitas ORGANIZA Receita
Receita REPLICA Modo de Preparo
Modo de Preparo MANIPULA os Ingredientes

Em Programação:

Bibliotecas ORGANIZA Funções
Funções REPLICA Comando
Comando MANIPULA Variáveis

O que é uma Variável:

- Caixa armazenadora
 - Tem um nome
 - Guarda apenas um tipo de objeto
 - Não guarda objetos diferentes

Por exemplo:

- Caixa Armazenadora de saias.
 - nome: caixa_saia
 - armazena: saias

Posso guardar sapatos na caixa armazenadora de saias? Não

Posso guardar calças na caixa armazenadora de saias? Não

Posso guardar camisteas na caixa armazenadora de saias? Não

Em programação;

- Caixa armazenadora é um espaço de memória para guardar um objeto de dados

O que é um Comando

