

Curso de Python do ZERO AO DS

com Meigarom do canal “Seja Um
Data Scientist”

Instagram:

@meigarom.datascience (Mais
informações sobre o Curso)

LinkedIn: <https://www.linkedin.com/in/meigarom/> (Contato
Profissional)

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

ESTUDOS)

Aula 02 - Extração e Manipulação de Dados I - Básico

Agenda:

- 1. Novas perguntas de negócio.**
- 2. Planejamento da Solução.**

3. Tipos de Variáveis.

4. Manipulação de Variáveis.

5. Exercícios Práticos.

1. Novas perguntas de negócio.

1.1. Recapitulando o

desafio: ([https://](https://sejaumdatascientist.com/os-5-projetos-de-data-science-que-fara-o-recrutador-olhar-para-voce/)

sejaumdatascientist.com/os-5-projetos-de-data-science-que-fara-o-recrutador-olhar-para-voce/)

- **EMPRESA:** House Rocket

- **MODELO DE NEGÓCIO:** Compra casas com preço baixo e revendo com o preço mais alto.

- **QUAL O DESAFIO:** Encontrar bons negócios dentro do portfólio disponível, ou seja, encontrar casas com preço baixo, em ótima localização e que tenham um ótimo potencial de revenda por um preço mais alto.

1.2. Novas perguntas do

CEO:

1. Qual a data do imóvel mais antigo no portfólio?

2. Quantos imóveis possuem o número máximo de andares?

3. Criar uma classificação para o imóveis, separando-os em baixo e alto padrão, de acordo com preço.

Acima de R\$ 540.000 -> alto padrão

Abaixo de R\$ 540.000 ->

baixo padrão

**4. Gostaria de um relatório ordenado pelo preço e contento as seguintes informações:
(id do imóvel,
data que o imóvel ficou disponível para compra,
o número de quartos,
o tamanho total to terreno,
o preço,
a classificação do imóvel
(alto e baixo padrão)**

5. Gostaria de um Mapa indicando onde as casas

**estão localizadas
geograficamente.**

2. Planejamento de solução:

**2.1. Produto Final (O que
eu vou entregar? Planilha,
email, gráfico, modelo de
ML ...)**

- Email + 2 anexos:**
 - Email:**
 - Texto: Perguntas**

I Respostas

- Anexo:**

- Um relatório em .csv**

- A foto de um mapa em html.**

2.2. Ferramenta (Qual ferramenta eu vou usar?)

- Python 3.8.0**
- PyCharm**

2.3. Processo (Como fazer?)

1. Qual a data do imóvel mais antigo no portfólio?

- Ordenar o conjunto de dados pela menor data

2. Quantos imóveis possuem o número máximo de andares?

- Encontrar os números de andares e determinar o maior

- Contar quanto imóveis eu tenho por andar.

**3. Criar uma
classificação para o
imóveis, separando-os em
baixo e alto padrão,
de acordo com preço.**

**Acima de R\$ 540.000 ->
alto padrão
(high_standard)**

**Abaixo de R\$ 540.000 ->
baixo padrão
(low_standard)**

**- Criar uma nova
coluna no conjunto de**

**dados chamada
“standard”**

**- Para cada linha, eu
vou comparar a coluna
“price”**

**- Se “price” for
maior que 540.000, eu vou
escrever “high_standard”
na coluna “standard”**

**- Se “price” for
menor que 540.000, eu vou
escrever “low_standard”
na coluna “standard”**

**4. Gostaria de um relatório ordenado pelo preço e contento as seguintes informações:
(id do imóvel,
data que o imóvel ficou disponível para compra,
o número de quartos,
o tamanho total to terreno,
o preço,
a classificação do imóvel
(alto e baixo padrão)**

**- Selecionar as
colunas desejadas/**

demandas.

- **Deletar as colunas não desejadas/demandas.**

5. Gostaria de um Mapa indicando onde as casas estão localizadas geograficamente.

- **Procurar uma biblioteca em Python que armazena uma função que desenha mapa.**

- **Aprender a usar a função que desenha**

mapas.

3. Os tipos das variáveis:

- Caixa armazenadora (espaço de memória)
- Precisa ter um NOME e um TIPO

- Boas práticas para o nome:

- expressar a responsabilidade da variável

- seguir estilo “Kamel Case” e “Snake case”

- “Kamel Case”:
HousePrice,

- “Snake Case”:
house_price

- Tipos de variáveis em Python

- Numérica (Inteiro, Float) - Inteiro: Valor sem vírgula, Float: com vírgula.
(4, 3.8)

- Categórica

(characters, strings) “o”

“m” - “meigarom”

**- Dates (date,
timestamp) - Date: Ano-
Mes-Dia, Timestamp: Ano-
Mes-Dia H:M:S**

**- Identificar os tipos das
variáveis:**

- Comando “dtypes”

4. Manipulando as

variáveis:

- Criar (colunas de variáveis e novas linha)

- Deletar (colunas de variáveis e novas linha)

- Selecionar:

- 4 formas de seleciona dados:

- Forma 01:
Direto pelo nome das colunas.

- Forma 02:
Pelos índices das colunas.

- Forma 03:
Pelos índices das linhas e

pelo nome das colunas
- Forma 04:
Pelos índices booleanos.
(True e False)

5. Executando o

PROCESSO

planejado:

1. Qual a data do imóvel
mais antigo no portfólio?
- Ordenar o conjunto
de dados pela menor data

7. Exercícios:

Novas perguntas do CEO para você:

1. Crie uma nova coluna chamada: “house_age”

- Se o valor da coluna “date” for maior que 2014-01-01 => ‘new_house’

- Se o valor da coluna “date” for menor

**que 2014-01-01 =>
'old_house'**

**2. Crie uma nova coluna
chamada:**

“dormitory_type”

**- Se o valor da
coluna “bedrooms” for
igual à 1 => ‘studio’**

**- Se o valor da
coluna “bedrooms” for
igual a 2 => ‘apartament’**

**- Se o valor da
coluna “bedrooms” for
maior que 2 => ‘house’**

3. Crie uma nova coluna chamada:

“condition_type”

- Se o valor da coluna “condition” for menor ou igual à 2 => ‘bad’

- Se o valor da coluna “condition” for igual à 3 ou 4 => ‘regular’

- Se o valor da coluna “condition” for igual à 5 => ‘good’

4. Modifique o TIPO a

Coluna “condition” para STRING

**5. Delete as colunas:
“sqft_living15” e
“sqft_lot15”**

**6. Modifique o TIPO a
Coluna “yr_build” para
DATE**

**7. Modifique o TIPO a
Coluna “yr_renovated”
para DATE**

8. Qual a data mais antiga de construção de um imóvel?

9. Qual a data mais antiga de renovação de um imóvel?

10. Quantos imóveis tem 2 andares?

11. Quantos imóveis estão com a condição igual a “regular” ?

12. Quantos imóveis estão com a condição igual a “bad” e possuem “vista para água” ?

13. Quantos imóveis estão com a condição igual a “good” e são “new_house”?

14. Qual o valor do imóvel mais caro do tipo “studio” ?

15. Quantos imóveis do

tipo “apartment” foram reformados em 2015 ?

16. Qual o maior número de quartos que um imóveis do tipo “house” possui ?

17. Quantos imóveis “new_house” foram reformados no ano de 2014?

18. Selecione as colunas: “id”, “date”, “price”, “floors”,

“zipcode” pelo método:

10.1. Direto pelo nome das colunas.

10.2. Pelos Índices.

10.3. Pelos Índices das linhas e o nome das colunas

10.4. Índices

Booleanos

19. Salve um arquivo .csv com somente as colunas do item 10 ao 17.

20. Modifique a cor dos

**pontos no mapa de “pink”
para “verde-escuro”**

