

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 08 - Projeto de Insights + Comunidade DS

Agenda:

1. As etapas de um projeto de Ciência de Dados.
2. Transformando o Python do Zero ao DS em um projeto de Portfólio.
3. Tarefa para casa.
4. Recapitulando o que for aprendido até agora.
5. Comunidade DS

1. As etapas de um projeto de Ciência de Dados

Questão de Negócio -> Entendimento do Negócio -> Coleta de Dados -> Limpeza de Dados -> Exploração de dados -> Modelagem de dados -> Algoritmos de ML -> Avaliação dos algoritmos -> Modelo em Produção.

1.1. O projeto do tipo Insights.

Objetivo: Gerar insights através da análise e manipulação dos dados para auxiliar na tomada de decisão pelo time de negócio.

Etapas: Questão de Negócio -> Entendimento do Negócio -> Coleta de Dados -> Limpeza de Dados -> Exploração de dados.

1.1.1. Questão de Negócio

1. Quais são os imóveis que a House Rocket deveria comprar e por qual preço ?
2. Uma vez a casa comprada, qual o melhor momento para vendê-las e por qual preço ?

1.1.2. Entendimento do Negócio

1. Produto final (o que eu vou entregar efetivamente ?)

- Relatório com as sugestões de compra de apartamento por um valor recomendado.
- Relatório com as sugestões de venda de um apartamento por um valor recomendado.

2. Ferramentas (quais ferramentas eu vou usar no processo?)

- Python 3.8.0
- PyCharm
- Jupyter Notebook

3. Processo (quais os passos necessário para alcançar meu objetivo?)

1. Quais são os imóveis que a House Rocket deveria comprar e por qual preço ?

- Plano 01:

- Coletar os dados do site do Kaggle.
- Agrupar os imóveis por região (zipcode).
- Dentro de cada região, eu vou encontrar a mediana do preço do imóvel.
- Vou sugerir os imóveis que estão abaixo do preço mediano da região e que estejam e boas condições e vista para água.

- Exemplo:

do Imóvel		Imóvel Cod		Região	Preço
Preço	na Mediana	Condições	Status		
450.000,00	I R\$ 500.000,00	10330	I 3	Vila Olímpia	I R\$
					I Compra
750.000,00	I R\$ 500.000,00	10335	I 3	Vila Olímpia	I R\$
					I Não Compra

150.000,00 I R\$ 500.000,00 10335 I 1 I Vila Olímpia I R\$
I Não Compra

2. Uma vez a casa comprada, qual o melhor momento para vendê-las e por qual preço ?

- Plano 01:

- Como os dados já tratados e organizados.
- Agrupar os imóveis por região (zipcode) e por sazonalidade (Summer, Inter).
- Dentro de cada região e sazonalidade, eu vou calcular a mediana do preço.
- Condições de venda:

1. Se o preço da compra for maior que a mediana da região + sazonalidade.

O preço da venda será igual ao preço da compra + 10%

1. Se o preço da compra for menor que a mediana da região + sazonalidade.

O

preço da venda será igual ao preço da compra + 30%

- Exemplo:

- Imóvel Cod | Região |

Temporada | Preço da Mediana | Preço de Compra | Preço de Venda | Lucro

Verão | R\$ 800.000,00 | 10330 | Vila Olímpia | R\$ 450.000,00 + 30% | ??

Inverno | R\$ 400.000,00 | 10335 | Vila Olímpia | R\$ 500.000,00 + 10% | ??

1.1.3. Coleta e Limpeza dos Dados

**- Coletar os dados do site do Kaggle
(<https://www.kaggle.com/harlfoxem/housesalesprediction>)**

- Limpeza de Dados

- Remover datas erradas.

- Remover outliers devido à erro do sistema.

1.1.4. Exploração de Dados:

1. Descobrir Insights para o time de negócio.

2. Explorar os dados para identificar o impacto dos atributos nos algoritmos de ML.

3. O que são Insights para o negócio?

- Insights são descobertas, através dos dados, que são inesperada pelas pessoas.
- Insight precisa ser acionável, caso contrário, ele é apenas uma curiosidade.

Exemplo:

- Durante o período de natal vendê-se mais casas do que na páscoa (Descoberta)
- Imóveis com porão são maiores do que imóveis sem porão (Não-acionável)
- Imóveis com porão são 40% mais caros do que os outros, na média (Acionável - Compra imóveis sem porão)

4. Como fazemos para criar as hipóteses ?

- Toda hipótese de negócio precisa ter 3 características:
1. Precisa ser uma afirmação.
 2. Precisa fazer uma comparação entre 2 variáveis.
 3. Precisa de um valor base.

H1: Imóveis que possuem vista para água, são 30% mais caros, na média.

H2: Imóveis com data de construção menor que 1955, são 50% mais baratos, na média.

H3: Imóveis sem porão possuem sqrt_lot, são 50% maiores do que com porão.

H4: O crescimento do preço dos imóveis YoY (Year over Year) é de 10%

H5: Imóveis com 3 banheiros tem um crescimento MoM

(Month over Month) de 15%

2. Transformando o curso Python do ZERO ao DS em um projeto de Portfólio.

2.1. O que é obrigatório em um projeto de portfólio.

1. Questão de negócio.

- O que você quer resolver ?
 - Encontrar as melhores oportunidades de compra de imóveis do portfólio da House Rocket.
- Qual a problema, a dor, a necessidade do time de negócio ?
 - O time do negócio não consegue tomar boas decisões de compra sem analisar os dados.
 - O portfólio é muito grande, muito tempo para fazer o trabalho manualmente.

2. Premissas de negócio.

- O que você está assumindo para realizar o projeto ?
 - Retirando casas com valor de venda maior que R\$ 500.000,00 (possível erro do sistema).
 - A média do preço das casas na região de Seattle é de R\$ 300.000,00.
 - Todas as casas estão em ótimas condições.

3. Planejamento da solução:

- Qual o seu plano para resolver o problema?

4. Os 5 principais insights dos dados:

- O que há de relevante nos dados que pode ajudar o time de negócio à tomar decisão.

5. Resultados financeiros para o negócio:

- Quanto a empresa espera lucrar com a sua solução ?

6. Conclusão:

- Seu objetivo inicial foi alcançado? Sim ? Não ?
Porque ?

7. Conclusão:

- Seu objetivo inicial foi alcançado? Sim ? Não ?
Porque ?

4. Tarefa para casa:

1. Criar visualizações para responder cada uma das 10 hipóteses de negócio.

2. Construir uma tabela com recomendações de compra ou não compra.

3. Construir uma tabela com recomendações de venda com acréscimo de 10 ou 30%.

4. Fornecer as hipóteses e as tabelas no Streamlit.

5. Transformar o projeto do curso Python do ZERO ao DS em um projeto de Portfólio.

6. Salvar os código no Github.

7. Escrever o README com os requisitos obrigatórios para um portfólio de projeto.

5. Recapitulando o que aprendemos até agora.

Aula 01: Começando com python do ZERO

Aula 02: Extração e manipulação de dados I - Básico

Aula 03: Transformação de dados I - Básico

Aula 04: Estruturas de controle

Aula 05: Funções e organização de código

Aula 06: Tipos de visualizações de dados

Aula 07: Visualização de dados II

Aula 08: Projeto de portfólio

Módulo 01: Resolução dos exercícios do Python do ZERO ao DS

Módulo 02: Instalação das ferramentas (Linux, Jupyter Lab,

Git, Github)

Módulo 03: O novo problema de negócio.

Módulo 04: Desenvolvimento do pensamento analítico

Módulo 05: Extração de dados em HTML usando o BeautifulSoup

Módulo 06: Extração de dados assíncrono

Módulo 07: Banco de Dados e SQL com Python

Módulo 08: Ambiente Virtual em Python

Módulo 09: Agendamento automático de tarefas

Módulo 10: Primeira rodada de perguntas do CEO e análises

Módulo 11: Extração de dados em HTML usando o Selenium

Módulo 12: Segunda rodada de perguntas do CEO e análises

Bônus: Storytelling

Em torno de 50 Aulas de conteúdo sobre Python

**Formação de Python =
Curso de Python do ZERO ao DS
+
Curso de Python do DS ao DEV**

Aulas estão gravadas e cada módulo será liberado na sexta-feira.

