

# **ASSIGNMENT**

**Portfólio Individual**

**CLASSIFICAÇÃO E RECOMENDAÇÃO**

# #20



# INSTRUÇÕES ASSIGNMENT 20

---

## INSTRUÇÕES

Este Assignment #20 é um mix de conteúdos de Classificação e Recomendação.

As questões respondidas deste assignment valem nota normalmente como nos anteriores. O desenvolvimento da resolução que será postado no seu github pessoal e servirá como um portfólio individual de Classificação e Recomendação.

Logo, você deverá:

- Desenvolver o modelo e colocar no [Github](#)
- Responder as [questões](#) referentes a solução deste Assignment
- Link do Github: na última questão você deverá inserir o link do seu github contendo o desenvolvimento deste assignment.

## OBJETIVO DO ASSIGNMENT COMO PORTFÓLIO INDIVIDUAL

1. Portfólio individual serve para demonstrar o seu conhecimento adquirido no curso através de projetos intermediários publicados em seu github pessoal. *\* Podem ser apresentados a empresas em entrevistas de emprego.*

2. No final do curso, vocês terão no total de 3 projetos intermediários como portfólio. Cada um deles valerá entregas de Assignments distribuídos nos slots finais 6, 7 e 8.

3. Os conteúdos podem se misturar com módulos anteriores. O importante é mostrar sua ideia, ter um portfólio estruturado.



# INSTRUÇÕES PARA O ASSIGNMENT 20

---

## CONTEXTO

O CEO da Amazon contratou um time de cientistas de dados para trabalhar em diferentes vertentes da empresa, e você ficou encarregado de fazer um sistema de recomendação para os clientes Amazon. Para isso, te deram acesso à uma base de dados não muito estruturada em Json: um arquivo de metadata com informações dos produtos e outro com as avaliações.

## O PROJETO:

O CEO deseja que seu algoritmo de recomendação seja exclusivamente em cima de avaliações verificadas (campo ``verified` = True` no arquivo de avaliações). Contudo, há uma base sem classificação e que o CEO faz questão de que seja adicionada no sistema de recomendação (valores com missing value na coluna ``verified``). Para isso, você precisará classificar se estas avaliações são verificadas ou não, e no caso positivo, adicioná-los no sistema de recomendação.

Adicionalmente, o CEO também deseja saber de possíveis associações e/ou correlações nesta base de dados que você vai trabalhar.



# INSTRUÇÕES PARA O ASSIGNMENT 20

---

## Projeto

Desta forma, eis o que você deve fazer neste projeto:

- Carregar, limpar e fazer análises exploratórias no banco de dados fornecido;
- Classificar a base sem informação para incluir as observações das avaliações verificadas no sistema de recomendação;
- Desenvolver e entregar um sistema de recomendação, com exemplos de aplicação.

OBS: O desenvolvimento e a decisão do modelo é totalmente sua, portanto se achar que deve utilizar um valor/procedimento diferente, sinta-se livre para fazer os testes e validar suas hipóteses para achar o resultado coerente.

LINK PARA DOWNLOAD DOS ARQUIVOS:

<https://bit.ly/3xzk1Tk>



# INSTRUÇÕES PARA O ASSIGNMENT 20

---

## LIMPEZA DOS DADOS

- Aqui você deve limpar e explorar os dados para decidir variáveis que possam ser úteis para classificação de avaliações verificadas. Você também deve levantar associações e/ou correlação para apresentar ao CEO.

## EDA

Realize uma EDA da forma como preferir, explore os dados, levante ideias, avalie correlações.

## 1º ETAPA: CLASSIFICAÇÃO

Aqui você deve utilizar as variáveis estudadas na primeira parte para classificar as avaliações sem label. Deve-se avaliar os modelos criados, tunar ao menos um modelo selecionado e utilizá-lo na classificação.

Após classificação, não se esqueça de remover as avaliações não verificadas.

## 2º ETAPA: RECOMENDAÇÃO

Aqui você deve desenvolver um sistema de recomendação e mostrá-lo na prática para o CEO. Tente prever possíveis erros para evitá-los de antemão. Por exemplo, caso o produto requisitado não esteja na matriz/base, aponte os 10 mais vendidos.



# INSTRUÇÕES PARA O ASSIGNMENT 20

---

## BOAS PRÁTICAS DE UM PORTFÓLIO

Lembrando que ao longo da sua análise, introdução, EDA, transformações e modelagem, é importante ressaltar a ideia por trás da sua ação para que o leitor entenda o raciocínio usado. Essa prática de documentação enriquece ainda mais seu projeto e garante um entendimento melhor do público interessado.

Links como referência:

### 1. Guia para aplicação dos métodos

<https://www.kdnuggets.com/2019/10/clustering-metrics-better-elbow-method.html>

### 2. Elbow Method vs Silhouette Score

<https://vitalflux.com/elbow-method-silhouette-score-which-better/>

dnc



Data Science & Machine Learning

**#HARDWORK**