



**Data**Hackers |



STATE OF DATA  
**BRAZIL 2025**

# AULA 02 – EDA

## ANÁLISE EXPLORATÓRIA

# ONTEM VOCÊ APRENDEU

## *AGENDA*

- ✓ O que é pandas e porque ele é o mais usado na área de dados!
- ✓ Enumerar as características de séries e dataframes pandas
- ✓ Importar dados em formato csv no pandas!
- ✓ Organizar o Dataset do State of Data pra começar sua análise VENCEDORA!



# AO FINAL DESSA AULA VOCÊ VAI SABER

## *AGENDA*

- ✓ Descrever a importância de realizar análise exploratória de dados
- ✓ Filtrar e agrupar dataframes pandas para buscar insights e extrair informações
- ✓ Criar gráficos para compreender melhor as informações!
- ✓ Como realizar diversos experimentos em busca da melhor análise!



**ANTES DE MAIS NADA**



# COMO PARTICIPAR DO CHALLENGE?



# CHALLENGE STATE OF DATA

## *PASSOS PARA PARTICIPAR*

1. Criar sua conta no Kaggle ([www.kaggle.com](http://www.kaggle.com))
2. Se inscrever no forms de inscrição!!
3. Ler as regras no regulamento "Challenge State of Data Brazil 2025" na aba Discussion do Kaggle do Data Hackers
4. Adicionar seu notebook no Kaggle do Data Hackers
5. Deixar o notebook público até a data de encerramento (04/08)



# CHALLENGE STATE OF DATA

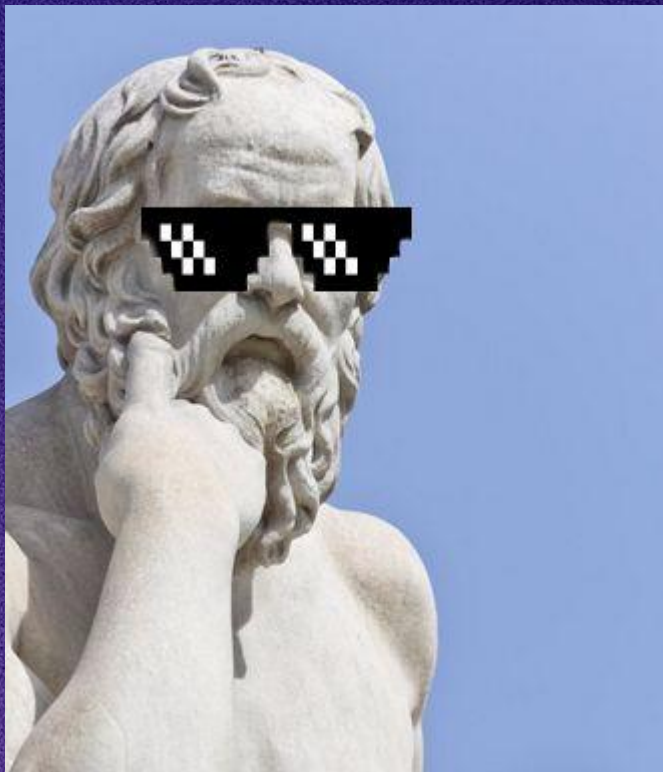
## *REGRAS IMPORTANTES!!*

- Os participantes precisam possuir endereço válido no Brasil
- As análises devem ser compreensíveis em português/Brasil
- Só aceitaremos um cadastro e uma solução por participante
- Os códigos precisam estar em Python ou R
- É permitido usar dados externos, desde que a origem seja informada, pois só serão permitidos dados públicos



**TODA ANÁLISE COMEÇA POR  
UMA HIPÓTESE!**





Leon se achando o Sócrates



# ESTRATÉGIA PARA ANÁLISE

## *CONCEITOS FUNDAMENTAIS*

- ✓ Vocês viram que é fácil se perder num mundaréu de colunas informações!
- ✓ É importante então ter uma ESTRATÉGIA para direcionar a análise
- ✓ Uma forma de conseguir encontrar uma boa estratégia é entendendo melhor os dados a partir da Análise Exploratória



# ESTRATÉGIA PARA ANÁLISE

## *CONCEITOS FUNDAMENTAIS*

- ✓ Para nossa aula vamos focalizar na REMUNERAÇÃO dos profissionais de dados (até porque esses boletos não vão se pagar sozinhos)
- ✓ A partir da Análise Exploratória da remuneração, levando em conta diversos aspectos, podemos formular nossa HIPÓTESE



# ESTRATÉGIA PARA ANÁLISE

## *CONCEITOS FUNDAMENTAIS*

- ✓ Uma hipótese pode ser confirmada ou não!
- ✓ Estatisticamente, precisamos de um rigor
- ✓ Mas no Challenge a ideia é buscar por insights, ou seja, esse rigor não é determinante!



# COMO É A REMUNERAÇÃO DOS PROFISSIONAIS DE DADOS?





# REMUNERAÇÃO DOS PROFISSIONAIS

## *CONCEITOS FUNDAMENTAIS*

- ✓ Vamos DEBULHAR as remunerações, buscar estratificações, comparações, extrair o máximo de conhecimento
- ✓ Para tal vamos criar muitos gráficos em diversos pacotes (bom que vocês já terão um notebook cheio de visualizações diferentes)



# ANÁLISE EXPLORATÓRIA DE DADOS

## *CONCEITOS FUNDAMENTAIS*

- ✓ Aqui sim, padawan, a gente tem que MERGULHAR nos dados
- ✓ A análise exploratória é sempre o primeiro passo uma vez que temos os dados acessíveis de modo a entender em detalhes as informações que os dados podem nos trazer
- ✓ Agrupamentos, contagens, gráficos, vale tudo pra tentar extrair bons insights



# ANÁLISE EXPLORATÓRIA DE DADOS

## *CONCEITOS FUNDAMENTAIS*

- ✓ Hoje a aula vai ser mais TOP DOWN, ou seja, vou mostrar os gráficos sem explicar muito os pacotes
- ✓ A notícia boa é que o notebook é de vocês! Podem usar, abusar e perguntar o quanto quiserem lá no Slack
- ✓ O mais importante é saber extrair insights! Com o notebook prontinho fica mole vocês trabalharem na análise de vocês



**DE NOVO ESSE GORDIN ME  
ENROLANDO! CADÊ ESSA  
ANÁLISE, MACHO!\***

**\* Homenagem à galera de Fortaleza, de onde eu não deveria ter saído, quero férias de novo, aqui tá muito frio**



**BORA PRO MÃO NA MASSA!**