

MEC

Pós-Graduação
Reconhecida
Oficialmente

Git e GitHub

para Análise de Dados



X XPERIUN

MBA

Fala meus queridos e queridas!

Sejam bem-vindos ao nosso próximo tópico, um que pode parecer, à primeira vista, pertencer exclusivamente ao mundo dos desenvolvedores de software, mas que, garanto, é uma das competências mais valiosas e diferenciadoras para um profissional de dados hoje: o controle de versão com Git e GitHub.

Muitos de vocês vêm de áreas de negócios e estão mergulhando fundo no universo dos dados. Outros já possuem um background mais técnico. O nosso objetivo aqui é nivelar o conhecimento e mostrar por que estas ferramentas são cruciais para todos, independentemente da sua origem.

O Cenário que Queremos Evitar

Imagine o seguinte: você está trabalhando em um projeto complexo de análise. Sua pasta de trabalho começa a ficar assim:

analise_vendas_v1.ipynb

analise_vendas_v2_final.ipynb

analise_vendas_v2_final_AGORA_VAI.ipynb

analise_vendas_v2_com_correcao_do_chefe.ipynb

Essa confusão é um grande risco para a integridade, a reprodutibilidade e a eficiência de qualquer projeto de dados. É aqui que entram o Git e o GitHub.

O que são Git e GitHub?

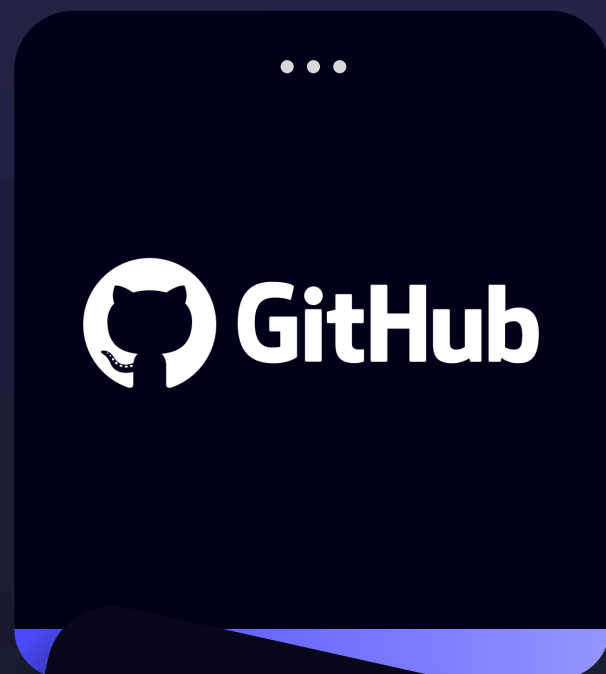
(A Perspectiva para Dados)

Pense neles como a solução definitiva para o caos dos arquivos.

Controle de Versão: A Máquina do Tempo para Seus Projetos de Dados

O conceito central é o

controle de versão, que consiste na prática de gerenciar as alterações feitas nos arquivos de um projeto. Ele funciona como um sistema que registra cada mudança, quem a fez e quando. Se um erro for cometido, você pode simplesmente "voltar no tempo" para uma versão anterior que funcionava perfeitamente.





O Cérebro Local do Seu Projeto

Git é a ferramenta. É um software gratuito e de código aberto que você instala em seu computador para fazer esse controle de versões. Pense em todos os

scripts que você irá criar: os arquivos .py com código Python para extrair, manipular e tratar dados; os arquivos .sql com as queries para consultar o banco de dados; ou até mesmo scripts em outras linguagens. O Git é perfeito para gerenciar esses arquivos de texto, pois ele consegue registrar cada linha de código que foi alterada, adicionada ou removida.



O Hub de Colaboração e Portfólio do Analista de Dados

GitHub é a plataforma na nuvem onde você e sua equipe hospedam o projeto. É o repositório central para sincronizar o trabalho de todos. Para nós, analistas de dados, o GitHub serve para:

Colaboração Estruturada: Vários analistas podem trabalhar em diferentes hipóteses ou modelos (em "branches" ou ramificações separadas) sem interferir no trabalho um do outro.

Reprodutibilidade e Credibilidade: Garante que sua análise possa ser verificada e reproduzida por outros, um pilar fundamental no mundo dos dados.

Construção de Portfólio: Este ponto é crucial para vocês. Hoje, um perfil no GitHub é o portfólio de um profissional de dados. É onde vocês podem exibir seus projetos de análise, suas visualizações e seus modelos, provando suas habilidades de forma prática para recrutadores.



E para Arquivos que Não São Código (Power BI, Dashboards)?

Uma dúvida comum é: "Isso serve apenas para código?". A resposta é não, e aqui o benefício é um pouco diferente, mas igualmente importante.

Peguemos o exemplo de um arquivo .pbix do Power BI ou um dashboard do Tableau. O Git/GitHub não consegue mostrar a

diferença exata dentro desses arquivos da mesma forma que faz com um código (pois são arquivos "binários", como uma caixa-preta).

No entanto, ele ainda oferece um valor imenso. Ele atuará como um repositório central e seguro, um cofre digital. Você saberá que a versão mais recente do dashboard da sua equipe está ali. O sistema registrará quem subiu a última versão e quando. Se o arquivo for corrompido ou perdido no computador de alguém, você terá um backup seguro na nuvem.

Em resumo:

Para código (Python, SQL), o Git/GitHub é uma máquina do tempo detalhada.

Para arquivos como .pbix ou planilhas, ele é um cofre seguro com um registro de atividades.

Quando Você Vai Usar Isso?

Sempre que iniciar um projeto de análise que envolva scripts (Python, SQL, R, etc.).

Quando for trabalhar em equipe em um projeto de dados.

Quando quiser armazenar e controlar as versões de dashboards e relatórios importantes (Power BI, Tableau).

Para construir seu portfólio profissional e compartilhar seus projetos com o mundo.

Nesta introdução, nosso foco foi no "o quê" e no "porquê". Na nossa aula do dia 19 de agosto de 2025, vamos colocar a mão na massa e aprender o "como". Veremos os comandos essenciais e faremos na prática nosso primeiro projeto de análise de dados versionado.

Até lá, sugiro que criem uma conta gratuita no <https://github.com/> para já irem se familiarizando com a plataforma e baixar o Git no computador (<https://git-scm.com/downloads>). Para ajudar, segue o vídeo para auxiliar tanto na instalação do GIT e criação de conta e token do GITHUB:

- <https://youtu.be/e6OfXJwy5SA?si=uBIJT49IEQ8266aL>

Te vejo na aula.

 **Abraços!**