

A Importância do SQL na Ciência de Dados

Olá! Bem-vindo(a) à nossa newsletter semanal sobre Ciência de Dados!

A linguagem SQL continua sendo uma das mais utilizadas na área de Ciência de Dados e não há como fugir disso. Seja para consultar grandes volumes de dados ou transformar informações para análises mais avançadas, **SQL** é uma ferramenta indispensável.

Seu domínio é essencial para o cientista de dados que deseja se destacar e atuar com eficiência em projetos diversos. O conhecimento de SQL é mais do que uma habilidade técnica – é uma condição básica para o sucesso na área.

Nesta newsletter, trazemos uma tabela que agrupa os principais comandos e funções do SQL, categorizando-os por tipo. Esses comandos são essenciais para a rotina de quem lida com dados e serão úteis em diversas etapas do seu trabalho.

Categoria	Comandos/Funções	Exemplo e Explicação
As Básicas	select, from, where, group by, order by, limit	<code>SELECT * FROM vendas WHERE valor > 100 ORDER BY data DESC LIMIT 10;</code> Esse comando busca as 10 maiores vendas acima de 100 reais, ordenadas pela data de forma decrescente.
Funções de Agregação	min(), max(), count(), avg(), sum()	<code>SELECT AVG(salario) FROM funcionarios;</code> Calcula a média salarial dos funcionários.
Manipulação de Strings	concat(), replace(), upper(), lower(), trim()	<code>SELECT CONCAT(nome, ' ', sobrenome) AS nome_completo FROM clientes;</code> Combina o nome e o sobrenome de clientes para criar um campo de nome completo.
Manipulação de Datas	date_add(), date_format(), datediff(), year()	<code>SELECT YEAR(data_venda) FROM vendas;</code> Extrai o ano da data de venda de cada registro no banco de dados.
Joins	inner join, left join, right join, self join	<code>SELECT a.nome, b.compra FROM clientes a INNER JOIN compras b ON a.id = b.cliente_id;</code> Realiza um inner join entre as tabelas de clientes e compras, unindo informações relacionadas.
Limpeza e Transformação	cast(), coalesce(), case when, union, except	<code>SELECT COALESCE(email, 'sem email') FROM contatos;</code> Usa coalesce para substituir valores nulos no campo email por "sem email".
Funções de Janela	rank(), row_number(), lag(), lead(), sum()	<code>SELECT nome, salario, RANK() OVER(ORDER BY salario DESC) AS rank_salarial FROM funcionarios;</code> Usa a função de janela rank() para atribuir uma classificação de acordo com o salário dos funcionários.
Avançado	CTEs, subqueries, data modeling	<code>WITH total_vendas AS (SELECT SUM(valor) FROM vendas) SELECT * FROM total_vendas;</code> Cria uma <i>Common Table Expression (CTE)</i> para armazenar o total de vendas e reutilizá-lo em consultas futuras.

Por que o SQL é essencial na Ciência de Dados? O SQL é a linguagem padrão para lidar com bancos de dados relacionais e, portanto, uma ferramenta crítica para quem trabalha com dados. Sua simplicidade, combinada com a capacidade de lidar com grandes volumes de informações, a torna ideal para realizar consultas complexas de forma eficiente. Além disso, o SQL oferece uma integração perfeita com outras ferramentas de análise, como Python e R, ampliando ainda mais sua aplicabilidade no contexto da Ciência de Dados.

Cada cientista de dados precisa saber como extrair dados limpos e organizados de fontes variadas, e o SQL facilita esse processo. Usar SQL para realizar agregações, filtrar dados e unir tabelas é parte do

dia a dia da profissão. Além disso, aprender comandos mais avançados, como **funções de janela** e **subqueries**, garante que você seja capaz de lidar com qualquer tipo de consulta, por mais complexa que seja.

Assim, a utilização de **SQL** vai muito além do básico. Quando usada de forma eficiente e estratégica, essa linguagem pode acelerar seu trabalho, melhorar a qualidade dos dados e proporcionar insights que transformam projetos em resultados reais. Se você deseja sucesso na área de dados, o domínio de SQL é indispensável.

Se tiver dúvidas ou sugestões, não hesite em entrar em contato. Fique atento à nossa próxima edição, onde continuaremos a explorar as últimas inovações e ferramentas que estão moldando o futuro da Ciência de Dados.

Saudações,

Prof. Dr. Dilermando Piva Jr.

Coordenador de Ciência de Dados para Negócios - Fatec Votorantim

E-mail: f301.cdn@fatec.sp.gov.br