



Módulo 2

Introdução às ferramentas de TI essenciais para análise de dados



Bem-vindo ao Módulo 2!

Neste módulo, vamos mergulhar no universo das ferramentas de TI essenciais para a coleta, organização e análise de dados. Durante as próximas horas, você aprenderá a **utilizar softwares e técnicas** que vão transformar dados brutos em insights valiosos para o seu negócio.

Isso porque a análise de dados depende de ferramentas que nos ajudam a coletar, organizar, interpretar e apresentá-los de maneira eficiente.

Vamos explorar as principais ferramentas que você poderá utilizar ao longo da sua jornada no mundo dos dados.



Antes de falarmos sobre a análise de dados em si, é importante conhecer as ferramentas que possibilitam esse processo.

Alguns exemplos são: Excel, SQL, Python e R. Cada uma dessas ferramentas e linguagens tem suas vantagens e aplicações específicas. Vamos explorar um pouco de cada uma.

Excel (Microsoft): Você provavelmente já ouviu falar (ou até usou) o Excel. Mas, você sabia que o Excel é mais do que apenas um software para criar planilhas? Ele oferece uma série de funcionalidades poderosas para análise de dados, como tabelas dinâmicas, fórmulas avançadas e gráficos interativos. Além disso, ele é extremamente acessível e está presente em praticamente todas as empresas.

Trazendo para um caso prático, imagine que você é gerente de vendas de uma empresa de produtos de beleza. Cada dia você recebe uma planilha com o número de vendas por produto, região e vendedor. Aí entra o Excel: com ele, você consegue organizar todos esses dados, calcular a média de vendas por vendedor e criar gráficos que mostram os principais indicadores.

Além disso, uma de suas principais funcionalidades são as Tabelas Dinâmicas. Você pode rapidamente descobrir qual produto teve o melhor desempenho em uma determinada região ou período. É uma ferramenta prática e acessível, usada por empresas de todos os tamanhos para análises rápidas e visuais.



SQL (Structured Query Language): A Linguagem de consulta estruturada (SQL) é uma linguagem padrão para criação e manipulação de bancos de dados. Muito usada para criar, acessar e gerenciar bancos de dados.

No ambiente empresarial, os dados muitas vezes são armazenados em grandes bancos de dados, como MySQL, PostgreSQL, SQL Server, entre outros. Com SQL, você será capaz de fazer consultas complexas como:

- **SELECT** para buscar dados específicos;
- **JOIN** para combinar informações de várias tabelas;
- **GROUP BY** para agrupar dados por categorias;
- **WHERE** para filtrar informações com base em critérios.

Essas funcionalidades permitem a você extrair apenas os dados relevantes, facilitando a criação de relatórios precisos.

O **MySQL**, por exemplo, é um programa que utiliza essa linguagem para armazenar, organizar e buscar informações de maneira eficiente.

Em outras palavras, o MySQL usa os comandos SQL para executar tarefas como adicionar novos dados, consultar informações existentes ou alterar registros no banco de dados.

Python: Python é uma das linguagens de programação mais populares no mundo da ciência de dados e análise de dados.

Para que você compreenda melhor, uma linguagem de programação é um conjunto de instruções e regras que permite aos desenvolvedores criar programas de software, sites e aplicativos. Ela funciona como uma ponte entre o ser humano e o computador, permitindo que as pessoas comuniquem comandos de forma que o computador possa entender e executar.

Essas linguagens têm uma sintaxe e semântica específicas, ou seja, possuem regras sobre como as instruções devem ser escritas (sintaxe) e o que elas significam (semântica). Cada linguagem de programação tem suas próprias características e é projetada para resolver tipos específicos de problemas.

Algumas linguagens são mais voltadas para o desenvolvimento de sites (como JavaScript), outras para sistemas de controle (como C ou Assembly), e outras ainda para ciência de dados (como Python).



Quando usar Python?

Se você precisa de uma linguagem mais geral, que além de análise de dados pode ser usada para outras tarefas como automação, web, etc.

Se você está lidando com machine learning e grandes volumes de dados.

Se você quer uma linguagem com uma curva de aprendizado mais suave para iniciantes, o Python é uma excelente opção.

Dessa forma, o Python é uma das linguagens mais populares para ciência de dados porque é muito versátil e tem uma vasta quantidade de bibliotecas (ferramentas prontas) que ajudam a trabalhar com dados de maneira eficiente.





R: Assim como o Python, o **R é uma linguagem de programação**, mas com um foco muito forte em análise estatística. Ele é amplamente utilizado por estatísticos e cientistas de dados para manipulação de dados e visualização avançada.

R foi originalmente criado para estatística e análise de dados. Por isso, ele é extremamente poderoso quando você precisa fazer cálculos estatísticos avançados ou trabalhar com modelos estatísticos mais complexos. Se você está em uma área que exige muita análise matemática, o **R oferece uma série de funções prontas para isso.**



Quando usar R?

Se você está em uma área que exige muita estatística e visualizações de alta qualidade.

Se você trabalha em ambientes acadêmicos ou de pesquisa.

Se o foco da sua análise é a modelagem estatística complexa, o R pode ser a melhor escolha.





Python & R

Tanto Python quanto R são amplamente usados em diversos setores, especialmente em ciência de dados, análise de negócios e áreas relacionadas à tecnologia e pesquisa. Abaixo estão alguns dos cargos que podem fazer uso dessas linguagens, com uma breve explicação de como cada linguagem é aplicada em cada posição:



Cargos que podem usar Python:

1.Cientista de Dados (Data Scientist):

Cientistas de dados usam Python para analisar dados, construir modelos de machine learning, e criar automatizações.

2.Engenheiro de Dados (Data Engineer):

Engenheiros de dados usam Python para construir e manter infraestruturas de dados, como bancos de dados e sistemas de processamento de dados em larga escala. Python também é usado para integrar diferentes sistemas e automatizar fluxos de dados.



3. Analista de Dados (Data Analyst):

Um analista de dados usa Python para processar, limpar e visualizar dados. O Python facilita a manipulação de grandes conjuntos de dados e a criação de relatórios e dashboards.

4. Desenvolvedor de Software (Software Developer) ou Engenheiro de Software:

Python é amplamente usado no desenvolvimento de software, especialmente em back-end de aplicações web, automação de tarefas e scripts. Cargos como desenvolvedor web ou engenheiro de automação podem usar Python para criar soluções de software, APIs e até aplicações completas.



5. Especialista em Machine Learning (Machine Learning Engineer)

Python é essencial para engenheiros de aprendizado de máquina, que desenvolvem, treinam e implementam modelos de inteligência artificial. Ainda vamos explorar esse tema em nosso curso.

6. Engenheiro de Automação (Automation Engineer):

Profissionais que automatizam tarefas repetitivas, como testes de software, gerenciamento de servidores e fluxos de trabalho, utilizam Python para escrever scripts de automação.

7. Administrador de Sistemas (SysAdmin/DevOps):

Python é usado para automatizar tarefas administrativas, gerenciar servidores, escrever scripts para manutenção de sistemas e integração contínua de software.

Cargos que podem usar R:



Cientista de Dados (Data Scientist):

Muitos cientistas de dados que trabalham com modelagem estatística complexa e visualização de dados detalhada usam R. Ele é popular em áreas acadêmicas e de pesquisa, especialmente por seus pacotes especializados em estatística avançada.



Estatístico:

Estatísticos frequentemente usam R para realizar análises de dados complexas, como regressão, teste de hipóteses, e modelagem estatística. A linguagem R é projetada para atender diretamente às necessidades da estatística.





Analista de Pesquisa (Research Analyst):

Profissionais que trabalham em áreas como ciências sociais, economia, biologia e outras áreas acadêmicas usam R para análises e experimentos estatísticos, especialmente quando precisam de ferramentas poderosas para manipulação de dados e gráficos.



Analista de Dados (Data Analyst):

Analistas de dados, especialmente em áreas que requerem análises detalhadas ou que seguem abordagens baseadas em métodos estatísticos complexos, usam R para organizar, limpar e interpretar dados.

Muitos profissionais em ciência de dados e análise trabalham com as duas, escolhendo qual usar com base nas exigências do projeto.

