

Analytics Engineering

TechGuide - Alura, FIAP e PM3

Analytics Engineering

Nível 1

☐ **Análise e Visualização de Dados (BI):**

- Esses conteúdos abrangem o campo da análise de dados e visualização de informações, com foco no Power BI, Tableau, Qlik Sense e Google Sheets.
- Os vídeos e cursos oferecem introdução ao Power BI, análise de dados da Microsoft, preparação de dados, modelagem de dados, visualização de dados, análise de dados e gerenciamento de workspaces e conjuntos de dados no Power BI.
- Os cursos da Alura oferecem formação em Power BI, Data Analysis com Google Sheets, Tableau, Looker Studio e Qlik Sense, fornecendo conhecimentos abrangentes nessas ferramentas.
- Os artigos fornecem informações sobre o que é Qlik, Tableau, Power BI e Looker Studio, bem como seus produtos e funcionalidades.
- Esses conteúdos são úteis para profissionais que desejam aprimorar suas habilidades em análise de dados, visualização de informações e utilização de ferramentas como Power BI, Tableau, Qlik Sense e Google Sheets.

☐ **Business Intelligence (BI) - Fundamentos:**

- Business Intelligence é um conjunto de teorias, metodologias, processos e tecnologias que possibilitam a transformação dos dados "crus" em

informações extremamente relevantes para tomada de decisão de uma empresa.

- Conhecer o processo de ETL
- Realizar a modelagem e estruturação de tabelas em um Data Warehouse
- Criar visualizações que façam sentido
- Conhecer o PowerBI

☐ **Python - Fundamentos:**

- Python é uma linguagem de programação de alto nível, de uso geral, amplamente utilizada em aplicações web, desenvolvimento de software, ciência de dados e Machine Learning. Sua filosofia de projeto enfatiza a legibilidade do código com o uso de indentação significativa. Python é dinamicamente tipada e tem um garbage collector.
- Conhecer os tipos primitivos
- Declarar variáveis, considerando os diferentes tipos
- Usar estruturas condicionais ('if', 'else')
- Conhecer os operadores de atribuição e comparação
- Usar estruturas de repetição e laços ('while', 'for')
- Usar funções, passando parâmetros e argumentos
- Manipular métodos
- Manipular arrays e listas
- Obter dados de uma API
- Criar construtores
- Funções anônimas

☐ **SQL - Fundamentos:**

- SQL (Structured Query Language, traduzindo, Linguagem de Consulta Estruturada) é uma linguagem de programação padronizada que é usada para gerenciar bancos de dados relacionais e realizar várias operações sobre os dados neles contidos.

- Conhecer os comandos mais comuns do SQL
- Usar SELECT para consultar uma tabela
- Usar INSERT para inserir dados em uma tabela
- Usar UPDATE para atualizar uma tabela
- Usar DELETE para remover dados de uma tabela
- Usar JOIN para conectar os dados de múltiplas tabelas
- Conhecer as cláusulas (FROM, ORDER BY, etc)

☐ **Engenharia de Dados - Fundamentos:**

- Uma pessoa Engenheira de Dados desempenha um papel crucial ao projetar, implementar e manter as infraestruturas de dados que permitem que uma organização trabalhe de maneira eficiente com suas informações. A pessoa engenheira de dados cria pipelines de dados para integrar, limpar e transformar dados provenientes de várias fontes e formatos. Isso possibilita a geração de insights valiosos para o negócio e melhora a tomada de decisões estratégicas.

☐ **Estatística e Matemática - Fundamentos:**

- Equações, funções e limites
- Logaritmos
- Matrizes, determinantes, vetores e espaço vetorial
- Derivadas e integrais
- Diferença entre média, mediana e moda
- Distribuição de frequência
- Variância e desvio padrão
- Distribuição binomial, poisson e normal
- Nível e intervalo de confiança
- Técnicas de amostragem
- Introdução à Regressão linear
- Séries temporais

☐ **Modelagem de dados:**

- A Modelagem de Dados é uma área essencial para a compreensão e organização das informações em um sistema ou banco de dados.
- Nesse conjunto de conteúdos, você aprenderá as técnicas e conceitos fundamentais para modelar dados de forma eficiente e estruturada, visando obter informações relevantes para a tomada de decisões.

☐ **Cloud - Provedores:**

- Conhecer os principais provedores de serviços em nuvem e suas ofertas.
- Comparar as características e serviços oferecidos por diferentes provedores.
- Entender como selecionar o provedor de serviços em nuvem adequado para suas necessidades.
- Explorar as tendências e inovações no mercado de serviços em nuvem.

☐ **Bancos de Dados - Fundamentos:**

- Os bancos de dados são sistemas de armazenamento de dados que permitem a organização, a manipulação e o acesso eficiente às informações. Existem dois principais tipos de bancos de dados, sendo eles SQL (Structured Query Language) e NoSQL (Not Only SQL).
- O conhecimento dos fundamentos de bancos de dados, tanto SQL quanto NoSQL, é essencial na engenharia de dados. Os bancos de dados são a base para armazenar e manipular os dados utilizados nas análises e nos processos de tomada de decisão. Compreender os diferentes tipos de bancos de dados permite aos engenheiros de dados escolher a solução mais adequada para cada caso, levando em consideração os requisitos de escalabilidade, desempenho, consistência e flexibilidade. Além disso, a habilidade de escrever consultas SQL e de utilizar as características específicas dos bancos de dados NoSQL é essencial para extrair, transformar e carregar dados de forma eficiente e eficaz.

Nível 2

☐ **Análise de Dados Avançada:**

- A análise de dados avançada é uma habilidade essencial para profissionais de ciência de dados e engenharia de analytics, permitindo identificação de padrões, tendências e insights valiosos a partir de grandes volumes de dados. Compreender técnicas avançadas de análise de dados, como machine learning, estatística avançada e análise preditiva, é fundamental para a tomada de decisões baseadas em dados e a criação de modelos de previsão precisos.
- Combinar técnicas de estatística, machine learning e análise de séries temporais para identificar padrões e tendências em dados históricos, prever comportamentos futuros e detectar anomalias em séries temporais.

☐ **Python para Ciência de Dados:**

- Python é uma linguagem de programação interpretada de alto nível e que suporta múltiplos paradigmas de programação, como imperativo, orientado a objetos e funcional. É uma linguagem com tipagem dinâmica e gerenciamento automático de memória.
- Aprender lógica de programação em Python
- Aprenda os fundamentos da linguagem como variáveis, funções, listas, condicionais e imports
- Criar análises de dados
- Utilizar o Matplotlib para gerar gráficos
- Usar e manipular listas para agrupar dados
- Conhecer a biblioteca NumPy
- Conhecer a biblioteca Pandas

☐ **SQL - Avançado:**

- Conhecer uma variedade abrangente de conteúdos relacionados a bancos de dados e linguagem SQL.

- Uso de triggers em SQL, linguagens internas de SGBDs, modelagem de dados e manipulação de consultas avançadas.
- Utilize os principais bancos de dados relacionais (Oracle, Microsoft SQL Server, PostgreSQL e MySQL...)
- Aprenda sobre procedimentos, funções, pacotes, otimização do banco, segurança e administração dos bancos de dados.

☐ **Orquestração de Data Pipelines:**

- A orquestração de data pipelines é um processo crucial na engenharia de dados que envolve o gerenciamento e a automação de fluxos de trabalho de dados. É através da orquestração que os pipelines de dados são configurados, programados e monitorados, garantindo a eficiência e a confiabilidade dos processos de extração, transformação e carga (ETL). Uma das ferramentas mais populares para orquestração de data pipelines é o Apache Airflow. Neste segmento, você aprenderá os fundamentos da orquestração de data pipelines, incluindo conceitos importantes, técnicas e melhores práticas.

☐ **Cloud - Fundamentos:**

- Cloud, ou computação em nuvem é a distribuição de serviços de computação pela Internet usando um modelo de preço pago conforme o uso. Uma nuvem é composta de vários recursos de computação, que abrangem desde os próprios computadores (ou instâncias, na terminologia de nuvem) até redes, armazenamento, bancos de dados e o que estiver em torno deles. Ou seja, tudo o que normalmente é necessário para montar o equivalente a uma sala de servidores, ou mesmo um data center completo, estará pronto para ser utilizado, configurado e executado.
- Conhecer a diferença entre IaaS, PaaS e SaaS
- Conhecer os maiores provedores de cloud
- Especializar-se em algum provedor

☐ **Big Data - Fundamentos:**

- Big Data refere-se a conjuntos de dados extremamente grandes e complexos, que não podem ser facilmente processados ou gerenciados por

métodos tradicionais. Esses conjuntos de dados são caracterizados por seu volume massivo, velocidade de geração e variedade de tipos e formatos. Com a explosão da quantidade de dados gerados diariamente, provenientes de diversas fontes como redes sociais, dispositivos IoT e transações comerciais, o Big Data apresenta desafios e oportunidades para extrair insights valiosos e tomar decisões informadas.

- A análise de Big Data envolve o uso de técnicas e tecnologias avançadas, como armazenamento distribuído, processamento em paralelo e aprendizado de máquina, para explorar e transformar esses dados em informações significativas para empresas, organizações e pesquisadores.

Nível 3

☐ **Deploy de Data Pipelines:**

- O deploy de data pipelines envolve a implantação e execução de fluxos automatizados para coletar, transformar e carregar dados de diferentes fontes para um destino específico. Ele garante a disponibilidade contínua dos dados para análise e outras aplicações, incluindo agendamento, monitoramento de desempenho e resolução de problemas durante o processo. O objetivo do deploy de data pipelines é assegurar a ingestão confiável de dados para suportar análises, relatórios e tomada de decisões informadas.
- Na engenharia de dados, o deploy de data pipelines desempenha um papel crucial ao garantir a disponibilidade, fluxo contínuo, escalabilidade, desempenho e monitoramento dos dados, proporcionando dados confiáveis e atualizados para impulsionar insights e ações.

☐ **Gerenciamento de logs:**

- O gerenciamento de logs é o processo de coleta, armazenamento, monitoramento e análise de registros de eventos gerados por sistemas, aplicativos e dispositivos. Esses registros de eventos, conhecidos como logs, contêm informações importantes sobre o funcionamento e o desempenho do sistema, bem como eventos relevantes para fins de segurança e auditoria.

- O gerenciamento de logs desempenha um papel crucial na engenharia de dados, pois fornece insights valiosos para a monitorização, resolução de problemas, otimização de desempenho e análise de sistemas e aplicativos. Os logs podem ser usados para identificar e solucionar problemas, detectar atividades suspeitas ou não autorizadas, rastrear eventos relevantes e analisar o comportamento do sistema ao longo do tempo. Além disso, o gerenciamento de logs é essencial para fins de conformidade, auditoria e conformidade regulatória, permitindo o acompanhamento e a documentação adequados das atividades relacionadas aos dados. Para isso, são utilizadas ferramentas e técnicas para coletar, centralizar, analisar e visualizar os logs, tornando-os mais eficientes e eficazes para a engenharia de dados.

☐ **Proteção de dados:**

- A proteção de dados é o processo de proteger informações importantes de forma que garanta a confidencialidade, integridade e a disponibilidade destes dados.

☐ **Governança de Dados:**

- Governança de dados é um conjunto de políticas, processos e práticas que estabelecem a responsabilidade, a integridade, a qualidade e o uso adequado dos dados em uma organização. Ela envolve a definição de regras e padrões para a coleta, armazenamento, gerenciamento, compartilhamento e uso dos dados, garantindo que sejam confiáveis, consistentes e seguros ao longo de sua vida útil.
- A governança de dados desempenha um papel fundamental na engenharia de dados, pois garante que os dados sejam gerenciados de forma eficiente, confiável e em conformidade com as regulamentações e políticas internas. Ela estabelece diretrizes para a gestão dos dados, incluindo a definição de metadados, a identificação de proprietários de dados, a documentação de políticas de acesso e privacidade, a implementação de medidas de segurança e a garantia da qualidade dos dados.
- A governança de dados também promove a colaboração entre as equipes, estabelecendo processos de tomada de decisão baseados em dados confiáveis e padronizados. Isso resulta em uma melhor qualidade das análises, uma base sólida para tomada de decisões estratégicas e uma

maior confiança nos dados utilizados para impulsionar as iniciativas de engenharia de dados.

☐ **Arquiteturas de Dados - Data Lake e Data Warehouse:**

- As arquiteturas de dados, como Data Lakes e Data Warehouses, são fundamentais para a engenharia de dados, permitindo a coleta, armazenamento, processamento e análise de grandes volumes de dados de diversas fontes. Data Lakes são projetados para armazenar dados brutos em seu formato original, enquanto Data Warehouses são otimizados para análises de negócios. Compreender as diferenças e os casos de uso de cada arquitetura é essencial para projetar e implementar soluções de dados eficazes.
- Na Engenharia de Analytics e Ciência de Dados, a escolha da arquitetura de dados correta é crucial para garantir a qualidade, a confiabilidade e a acessibilidade dos dados. Data Lakes são ideais para armazenar grandes volumes de dados brutos e semiestruturados, permitindo flexibilidade na análise e processamento dos dados. Já os Data Warehouses são otimizados para consultas analíticas complexas e relatórios de negócios, garantindo a consistência e a qualidade dos dados para tomada de decisões.

☐ **Storytelling com dados:**

- Storytelling é uma forma de contar histórias que engajam e chamam a atenção da pessoa que está ouvindo. Dentro da análise de dados, é algo muito importante para passar as informações ao receptor de modo que o mesmo compreenda não apenas os dados, mas também todo o contexto.

Habilidade Auxiliar: Comunicação, Negócios e Colaboração

☐ **Análise de negócios:**

- A análise de negócios auxilia os analistas de BI a coletar e analisar as necessidades empresariais, alinhando as soluções de BI com os objetivos.
- Permitir uma comunicação eficaz e tradução de requisitos entre usuários de negócios e equipes técnicas.

- Aprimora a modelagem de dados e o design de relatórios para saídas de BI precisas e perspicazes.
- Facilitar a colaboração com os stakeholders, garantindo que os projetos de BI atendam às expectativas.
- Contribuir para uma melhor tomada de decisões, analisando processos e propondo soluções baseadas em dados.

☐ **Habilidades de comunicação:**

- Um bom nível de comunicação facilita o atingimento de objetivos, resolução de problemas, além de aumentar a produtividade, porque cada profissional saberá exatamente o que se espera dele e transmitir com clareza suas ideias.

☐ **Inglês técnico:**

- Um bom nível de inglês técnico facilita o atingimento de objetivos, resolução de problemas, além de aumentar a produtividade, e também permite o consumo de materiais de diferentes fontes, principalmente de documentações oficiais.

Habilidade Auxiliar: Análise e Sistemas

☐ **Excel:**

- O Microsoft Excel é um editor de planilhas produzido pela Microsoft com ferramentas de cálculo e de construção de tabelas.
- Realizar as operações matemáticas básicas com seus operadores (soma, subtração, multiplicação e divisão)
- Conhecer as principais fórmulas, como 'MÉDIA' (AVERAGE), 'ARRED' (ROUND), 'MÁXIMO' (MAX), 'MÍNIMO' (MIN), etc
- Realizar buscas em colunas com a função 'PROCV'
- Criar gráficos

☐ **Criação e manipulação de APIs:**

- Criação e manipulação de APIs envolvem o desenvolvimento e interação com Interfaces de Programação de Aplicativos. APIs são conjuntos de regras que permitem que diferentes sistemas se comuniquem e compartilhem dados. A criação de APIs implica em projetar endpoints que permitem o acesso e interação com recursos específicos, enquanto a manipulação envolve utilizar APIs existentes para realizar tarefas específicas, como autenticação, envio de solicitações e processamento de dados retornados. Essas práticas são cruciais para a integração eficiente de sistemas e aplicativos em diferentes áreas de desenvolvimento.
- A criação e manipulação de APIs são fundamentais na engenharia de dados, permitindo a troca eficiente de dados entre sistemas, a integração com serviços externos e a colaboração entre equipes. As APIs simplificam a coleta, processamento e distribuição de informações, ajudando a construir pipelines de dados robustos e escaláveis.

☐ **Entrega e integração contínuas (CI/CD):**

- CI/CD é a abreviação de Continuous Integration/Continuous Delivery, traduzindo para o português "entrega e integração contínuas". Trata-se de uma prática de desenvolvimento de software que visa tornar a integração de código mais eficiente por meio de builds e testes automatizados.
- Automatizar a integração de código entre varias partes da equipe se tornou cada vez mais importante, ja que assim é possível acelerar o desenvolvimento e diminuir o tempo de entrega de software.
- Executar testes automatizados da aplicação para verificar seu funcionamento.
- Realizar a entrega de atualizações de forma automática e com segurança.
- Realizar testes de conexão e testes de carga para evitar que a aplicação apresente problemas ao ser atualizada.

☐ **DevOps e Cloud - Conceitos:**

- DevOps desempenha um papel importante na integração contínua e na entrega contínua de projetos, ao garantir a colaboração eficaz entre desenvolvedores, operadores de sistemas e outros stakeholders importantes em um projeto. Essa pessoa também é fundamental na

implementação e gerenciamento de soluções de computação em nuvem, facilitando o acesso a recursos e serviços, bem como a escalabilidade e a segurança dos sistemas.

- DevOps é essencial para a engenharia de dados, pois permite a integração contínua e a entrega eficiente de projetos. A colaboração entre equipes de desenvolvimento, operações e engenharia de dados facilita a integração de fontes de dados, processamento em tempo real e entrega de insights. Além disso, práticas DevOps garantem testes automatizados, monitoramento e governança de dados, garantindo qualidade e segurança nas soluções analíticas. Essa abordagem agiliza a gestão de dados em escala, possibilitando tomadas de decisões informadas.

☐ **Linux - Fundamentos:**

- Linux é um termo popularmente empregado para se referir a sistemas operacionais que utilizam o Kernel Linux. As distribuições incluem o Kernel Linux, além de softwares de sistema e bibliotecas.
- Conhecer o sistema de diretórios do Linux
- Compactar e descompactar arquivos
- Editar arquivos no console com o VI
- Gerenciar os processos rodando na máquina
- Conhecer as variáveis de ambiente e o PATH
- Gerenciar pacotes
- Realizar comunicação remota com o SSH e SCP

☐ **Git e GitHub - Fundamentos:**

- Git é um sistema de controle de versão distribuído gratuito e de código aberto projetado para lidar com tudo, desde projetos pequenos a muito grandes com velocidade e eficiência.
- GitHub é um serviço de hospedagem para desenvolvimento de software e controle de versão usando Git.
- Criar um repositório
- Clonar um repositório

- Fazer commit, push e pull de e para o repositório
- Reverter um commit
- Criar branches e pull requests
- Lidar com merge e conflitos

Streaming:

- Streaming é o processamento e transmissão contínua de dados em tempo real, à medida que são gerados. Diferentemente do armazenamento tradicional, onde os dados são coletados e processados posteriormente, o streaming permite a análise e o uso imediato dos dados à medida que são recebidos. Essa abordagem é ideal para lidar com grandes volumes de dados em alta velocidade, permitindo a detecção de padrões em tempo real, tomada de decisões instantâneas e resposta rápida a eventos em andamento.
- O streaming desempenha um papel essencial na engenharia de dados, capacitando as organizações a processar, analisar e tomar decisões com base em dados em tempo real. Isso resulta em maior agilidade, escalabilidade e reatividade, possibilitando uma tomada de decisões mais informada e orientada por insights atualizados.

TechGuide - Alura

Alura, PM3 e FIAP

O Techguide.sh é um projeto open source