

# Data Engineering

TechGuide - Alura, FIAP e PM3

---

## Data Engineering

### Nível 1

#### ☐ 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.

#### ☐ 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

## ☐ **Java - Fundamentos:**

- Java é uma linguagem de programação amplamente usada para codificar aplicações Web. Java é uma linguagem multiplataforma, orientada a objetos e centrada em rede que pode ser usada como uma plataforma em si. É uma linguagem de programação rápida, segura e confiável para codificar tudo, desde aplicações móveis e software empresarial até aplicações de big data e tecnologias do servidor.
- 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

## ☐ **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.

#### ☐ **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

#### ☐ **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.

#### ☐ **Pipelines e ETL:**

- Pipelines e ETL (Extract, Transform, Load) são conceitos essenciais no campo da Engenharia de Dados, desempenhando um papel fundamental na coleta, transformação e integração de dados. Compreender os fundamentos dessas áreas é essencial para aqueles que desejam trabalhar com a gestão e o processamento eficiente de grandes volumes de informações.

#### ☐ **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.

#### ☐ **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

## Nível 2

### ☐ **Processamento distribuído de dados:**

- Em processamento distribuído de dados usamos técnicas e ferramentas para realizar tarefas computacionais em um conjunto de máquinas conectadas, conhecido como cluster. Essas máquinas trabalham em conjunto para analisar e processar grandes volumes de dados de maneira mais eficiente do que seria possível em uma única máquina. O profissional pode usar várias ferramentas e frameworks para realizar esse trabalho, incluindo Apache Hadoop, Apache Spark e Databricks.

### ☐ **Bancos de dados relacionais:**

- Quando trabalhamos com bancos de dados relacionais somos responsáveis por projetar, implementar e gerenciar sistemas de bancos de dados que organizam dados em tabelas inter-relacionadas. Elas utilizam a linguagem SQL para manipular e consultar dados, além de técnicas de modelagem de dados para criar estruturas de banco de dados eficientes e eficazes. Os bancos de dados relacionais são usados em uma variedade de aplicações, incluindo sistemas de gestão de clientes, aplicações de e-commerce e sistemas de gerenciamento de recursos empresariais.

### ☐ **Bancos de dados não relacionais:**

- Os bancos de dados não relacionais, também conhecidos como bancos de dados NoSQL, são uma alternativa aos bancos de dados relacionais

tradicionais. Eles são particularmente adequados para lidar com grandes volumes de dados estruturados e não estruturados e são capazes de escalar horizontalmente em vários servidores para suportar cargas de trabalho pesadas. Entre os bancos de dados NoSQL mais populares estão o MongoDB e o Datomic. Neste segmento, você aprenderá os fundamentos dos bancos de dados NoSQL, suas características e usos típicos.

#### ☐ **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.

#### ☐ **Contêineres:**

- Os contêineres são pacotes de software que contêm todos os elementos necessários para serem executados em qualquer ambiente. Gerenciamento de contêineres é uma área crucial na computação em nuvem e DevOps, que envolve o uso de tecnologias para automatizar o processo de criação, implantação, escalonamento e monitoramento de contêineres. Contêineres são unidades de software padronizadas que permitem aos desenvolvedores empacotar todas as dependências de um aplicativo (código, bibliotecas, configurações, etc.) em um único pacote. Isso permite que o aplicativo seja executado de forma consistente em qualquer ambiente de infraestrutura.
- A tecnologia de contêineres, como exemplificada pelo Docker, fornece um ambiente consistente e portátil para desenvolvimento, teste e implantação de aplicativos, o que é vital para o trabalho eficiente de engenharia de dados. Além disso, o Kubernetes, um sistema de orquestração de contêineres, permite o gerenciamento, a automação e a escalabilidade de aplicações baseadas em contêineres em ambientes de produção. Dominar

esses conceitos e tecnologias possibilita a engenheiros de dados construir e manter pipelines de dados eficientes e confiáveis.

- O Kubernetes (também conhecido como k8s ou kube) é uma plataforma de orquestração de containers open source que automatiza grande parte dos processos manuais necessários para implantar, gerenciar e escalar aplicações em containers.
- Isolar seu software para funcionar independentemente
- Implantar software em clusters
- Modularizar seu sistema em pacotes menores
- Conhecer a plataforma Docker
- Conhecer Kubernetes

#### **Data Warehouses:**

- Um Data Warehouse é um sistema de armazenamento de dados projetado para análise de negócios, suportando processamento de transações e consultas complexas. Ele permite a consolidação de dados de várias fontes, fornecendo um ambiente unificado para análise de dados. Com recursos de normalização e limpeza de dados, garante a qualidade e a consistência dos dados armazenados. Soluções de Data Warehouse populares incluem Google BigQuery, Amazon Redshift e Snowflake, cada uma com suas próprias características e benefícios únicos.
- Data Warehouses desempenham um papel crucial na engenharia de dados. Eles fornecem uma infraestrutura que facilita a gestão, o processamento e a análise de grandes volumes de dados. Sua capacidade de armazenar e processar grandes quantidades de dados de várias fontes permite análises avançadas e suporte à tomada de decisões, o que é fundamental no atual cenário empresarial orientado por dados. Os engenheiros de dados projetam, constroem e mantêm data warehouses para garantir que os dados estejam acessíveis, seguros e prontos para análise. Além disso, soluções modernas de data warehouse, como o BigQuery, o Amazon Redshift e o Snowflake, oferecem capacidades de escalabilidade e flexibilidade para lidar com o crescente volume, velocidade e variedade de dados, tornando-se ainda mais valiosos para a engenharia de dados.



## **Data Lakes:**

- Um Data Lake é uma arquitetura de armazenamento de dados que permite o armazenamento de grandes volumes de dados brutos em seu formato original, provenientes de diversas fontes e em vários formatos, como estruturado, semiestruturado e não estruturado. Diferentemente do Data Warehouse, que exige que os dados sejam limpos e estruturados antes de serem armazenados, o Data Lake permite o armazenamento de dados "como estão", proporcionando flexibilidade para transformar e processar os dados conforme necessário para fins de análise.
- Data Lakes são fundamentais para a engenharia de dados devido à sua capacidade de lidar com a crescente variedade, volume e velocidade de dados gerados no mundo digital atual. Eles fornecem um meio econômico e escalável para armazenar e analisar grandes volumes de dados de diversas fontes, o que é fundamental para análises avançadas, como aprendizado de máquina e análise preditiva. Os engenheiros de dados são responsáveis pela construção e manutenção de Data Lakes, garantindo a ingestão de dados de maneira eficiente, o gerenciamento seguro dos dados e a implementação de processos de governança de dados. Além disso, os Data Lakes suportam a democratização dos dados, permitindo que diferentes partes interessadas, desde cientistas de dados até analistas de negócios, acessem os dados necessários para gerar insights valiosos.

## **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

## ☐ **Redes de Computadores - Fundamentos:**

- Rede de computadores é uma malha que interliga milhares de sistemas computacionais para a transmissão de dados. Também conhecidos como nós, esses dispositivos interconectados enviam, recebem e trocam tráfego de dados, voz e vídeo, graças ao hardware e software que compõe o ambiente.
- Configurações de redes são essenciais para que seja possível acessar uma aplicação, principalmente se ela estiver na nuvem.
- Entenda melhor os componentes de rede e a suas atribuições
- Diferenciar os serviços disponibilizados a nível de rede
- Saber as diferenças entre as camadas de rede
- Saber o que é servidor web, proxy reverso e load balancer

## **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, contém 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.

#### ☐ **Data Mesh:**

- O Data Mesh promove uma abordagem descentralizada na arquitetura de dados, capacitando cada equipe ou domínio a ser responsável por seus próprios dados. Isso resulta em maior autonomia, colaboração e eficiência na gestão dos dados.
- A importância do Data Mesh na engenharia de dados está relacionada à descentralização da responsabilidade pelos dados. Essa abordagem promove a colaboração, agilidade e melhoria da qualidade e confiabilidade dos dados na organização.

#### ☐ **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.

#### ☐ **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.

#### ☐ **Infraestrutura como código (IaC):**

- Infraestrutura como código (IaC) é uma abordagem em que a infraestrutura de TI, incluindo servidores, redes e recursos relacionados, é gerenciada e provisionada por meio de código. Em vez de configurar manualmente a infraestrutura, o IaC utiliza arquivos de configuração ou scripts que descrevem a infraestrutura desejada de forma automatizada e reproduzível.
- A infraestrutura como código é de extrema importância na engenharia de dados, pois permite uma gestão eficiente e escalável dos recursos de infraestrutura necessários para processamento e armazenamento de dados. Com o IaC, é possível definir, provisionar e configurar ambientes de maneira consistente, rápida e controlada. Isso facilita a implantação e manutenção de pipelines de dados, ambientes de desenvolvimento e testes, clusters de processamento distribuído e outros componentes necessários para a engenharia de dados. Além disso, o IaC permite a automação e o versionamento da infraestrutura, proporcionando maior agilidade, rastreabilidade e garantia de qualidade no gerenciamento dos recursos de TI.

#### ☐ **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.

#### **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 automatica e com segurança.
- Realizar testes de conexão e testes de carga para evitar que a aplicação apresente problemas ao ser atualizada.

# Habilidade Auxiliar: Arquitetura de software

## ☐ 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.

## ☐ HTTP - Fundamentos:

- HTTP significa Hyper Text Transfer Protocol. A comunicação entre computadores cliente e servidores web é feita enviando solicitações HTTP e recebendo respostas HTTP.
- Entender a diferença dos verbos HTTP
- Testar os requests e ver os status codes no navegador
- Saber fazer uma requisição HTTP na linha de comando com WGET
- Baixar uma imagem com WGET
- Fazer um post

## ☐ Testes de software:

- Testes de software referem-se ao processo de avaliar e validar um software para garantir que ele esteja funcionando corretamente e atenda aos requisitos estabelecidos. Esses testes envolvem a execução de diferentes

cenários e casos de uso, identificando bugs, erros e garantindo a qualidade do software.

- Na engenharia de dados, os testes de software são essenciais para garantir a precisão e confiabilidade dos pipelines de dados e dos sistemas de armazenamento, processamento e distribuição de dados. Ao realizar testes rigorosos, os engenheiros de dados podem identificar e corrigir erros ou inconsistências nos dados, garantindo a integridade e a consistência dos fluxos de dados. Além disso, os testes de software ajudam a validar as transformações e os processos de manipulação de dados, certificando-se de que os dados sejam processados corretamente e estejam em conformidade com as regras de negócio. Isso resulta em um ambiente de engenharia de dados mais confiável, seguro e de alta qualidade, garantindo que os dados sejam consistentes e precisos para análise, tomada de decisões e outras atividades relacionadas aos dados.

#### ☐ **Design Patterns:**

- Na engenharia de software, um "padrão de projeto" (Design Pattern em inglês) é uma solução geral e reutilizável para um problema que ocorre normalmente dentro de um determinado contexto de projeto de software.
- Conhecer e aplicar os principais Design Patterns

#### ☐ **Estruturas de Dados:**

- No contexto dos computadores, uma estrutura de dados é uma forma específica de armazenar e organizar os dados na memória do computador para que esses dados possam ser facilmente recuperados e utilizados de forma eficiente quando necessário posteriormente.
- Conhecer as principais estruturas de dados
- Implementar as principais estruturas de dados

#### ☐ **Analytics by Design:**

- Analytics By Design é uma metodologia para melhorar a prontidão analítica dos projetos. Um projeto pronto para análise permitirá que cientistas de dados, analistas e até mesmo equipes de negócios comecem a analisar e obtenham insights com mais rapidez.

# Habilidade Auxiliar: Dados e Machine Learning

## ☐ 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

## ☐ Extração e Tratamento de Dados:

- A extração de dados é o processo de coleta ou recuperação de tipos diferentes de dados de uma variedade de fontes, muitos dos quais podem estar mal organizados ou completamente desestruturados.
- Obter os dados que serão analisados
- Tratar os dados obtidos, transformando-os, alterando sua estrutura e valores a fim de deixar a base de dados mais coerente e garantir que os dados que serão trabalhados estejam nas melhores condições para serem analisados
- Utilizar as bibliotecas Pandas e Scikit-learn para tratar os dados

## ☐ Machine Learning - Fundamentos:

- O Aprendizado de Máquina ou Machine Learning é um subcampo da Engenharia e da Ciência da Computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial.
- Aprendizado supervisionado
- Utilizar algoritmos de classificação
- Utilizar algoritmos de regressão



- Utilizar o Scikit-learn para criar modelos de machine learning

#### ☐ **MLOps:**

- O MLOps permite que os modelos de Machine Learning sejam implantados de maneira rápida e confiável, o que é especialmente importante em empresas que lidam com grandes volumes de dados.
- Ajuda a garantir a qualidade e confiabilidade dos modelos de Machine Learning em produção, além de facilitar a manutenção e atualização desses modelos.
- Permite que as pessoas engenheiras de dados trabalhem em conjunto com cientistas de dados e desenvolvedores de software para implementar soluções de Machine Learning em larga escala.
- Ajuda a garantir a governança e a conformidade dos modelos de Machine Learning com as políticas e regulamentações da empresa.

#### ☐ **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

Alura, PM3 e FIAP  
O Techguide.sh é um projeto open source