



Módulo 1

Principais conceitos de análise de dados - estatística básica, mineração de dados e visualização de dados estruturados e não estruturados



Agora que entendemos o que é um dado e sua importância para as empresas e organizações de forma geral, vamos trazer conceitos mais técnicos e como aplicá-los.

Nos dias de hoje, fala-se muito sobre
"tomar decisão baseada em dados".

Mas como obter essas informações?

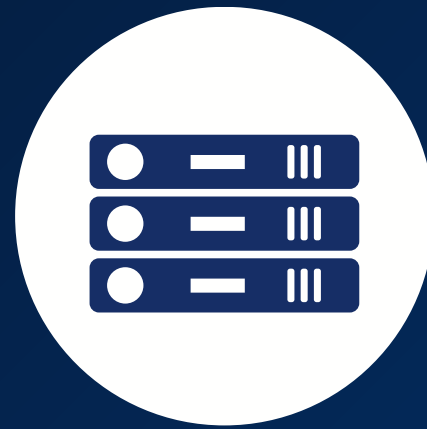
Você já aprendeu no exemplo do aplicativo de Delivery do último módulo que suas informações como nome e localização tornam-se dados para a empresa.

Dessa forma, a empresa precisa agora ter todo um processo para que, de fato, seus funcionários consigam tomar as suas decisões baseadas em dados.

Abaixo, segue um framework do que seria o macroprocesso necessário para isso:



Coleta de Dados



Harmonização
dos Dados



Visualização dos
Dados



Insights

Em primeiro lugar, precisamos "**Coletar os Dados**". Sabe quando você está preenchendo o formulário de um e-commerce ou de um aplicativo no seu celular?

Nesse momento, a empresa está coletando as suas informações. Essa coleta pode ser feita a partir, principalmente, de formulários. Sendo nesse caso, um dado primário ou manual. Google forms e Typeform são ferramentas muito utilizadas para isso.

Além disso, existem outras formas de coletar a informação. Por exemplo, as redes sociais geram dados sobre alcance, audiência, visitas ao perfil.

É muito comum órgãos governamentais disponibilizarem também **APIs públicas para consultar alguma informação ou também extrair dados de sites**, por exemplo. O próprio governo português tem a sua página administrada pelo Portal de dados abertos a Administração Pública.

Tudo isso faz parte da coleta de dados e existem diversas tecnologias que ajudam nesse trabalho.

Mas é fundamental que você entenda alguns principais conceitos:



APIs (Application Programming Interfaces)

É um conjunto de regras e protocolos que permite que diferentes **sistemas de software se comuniquem entre si**. Em termos simples, uma API é uma ponte que permite que um software envie e receba dados de outro software;

Web Scrapping

É uma técnica utilizada para extrair dados de websites automaticamente. Essa prática busca **navegar em páginas da web, identificar e extrair informações específicas**, como textos, imagens, tabelas ou outros elementos de interesse;



Dados Estruturados

Dados estruturados são **informações organizadas em um formato específico e pré-definido**, geralmente em tabelas ou planilhas, onde cada campo ou coluna tem um significado, como números, datas e textos

Dados Não - Estruturados

Dados não estruturados são **informações que não estão organizadas em um formato pré-definido ou padronizado**, como tabelas ou planilhas. Esses dados não seguem um modelo padrão, o que os torna mais complexos para armazenar, processar e analisar. Eles podem incluir uma grande variedade de tipos de dados, como texto livre, imagens, vídeos, áudios, documentos, e-mails, postagens em redes sociais, entre outros. Um ótimo exemplo de um dado não estruturado são as notícias.



Seguindo a nossa explicação, depois que você coletou os dados, você precisa armazená-los em algum lugar.

Para isso você pode então ter um banco de dados local ou na nuvem (ou híbrido). O banco de dados local ou fora da nuvem são os On-Premises.

E o que significa isso?

Basicamente, isso significa que você está armazenando aquela informação em algum dispositivo localizado na empresa ou em algum servidor específico da empresa.

Já no armazenamento na nuvem, é como se os dados estivessem 'na internet' e não em um dispositivo específico.

Por exemplo, o Google tem Google Cloud Platform, onde você pode armazenar uma série de documentos de diferentes tipos de formato. Já a Amazon desenvolveu a AWS (Amazon Web Services), a Microsoft, a Azure e Onedrive, a Apple, o iCloud. Essas, são as principais ferramentas mais conhecidas de armazenamento na nuvem no mercado.



Agora que temos os dados coletados e armazenados, precisamos "Higienizar" essa base. Basicamente, a higienização de dados, também conhecida como limpeza de dados, é o **processo de identificar e corrigir ou remover dados incorretos, incompletos, duplicados ou inconsistentes em um conjunto de dados**. Esse processo é essencial para garantir que as análises e decisões baseadas nos dados sejam precisas e confiáveis.

Então, a limpeza dos dados garante que aquela informação estará correta quando a analisarmos. Existe uma expressão muito interessante utilizada no mercado de tecnologia que explica bem a importância da higienização da informação. "**Garbage in, garbage out**". Na tradução literal para o português significa "lixo dentro, lixo fora", o que entende-se que se você for analisar uma base de dados que não está bem higienizada, você não terá insights bons porque aquela informação está errada, o que pode levar a uma tomada de decisão equivocada.



Seguindo a linha de raciocínio, primeiro você coletou os dados, definiu um local de armazenamento e depois higienizou aquela base. A próxima etapa agora é a "Visualização dos dados".

Essa etapa é fundamental para o sucesso da sua análise. Quando trabalhamos com banco de dados muito grandes, **conhecidos como Big Data**, torna-se impossível você analisar aquela informação sem um gráfico que resume o que está naquele banco de dados.

A visualização dos dados permitirá que você entenda e leia todas as informações. Existem algumas formas de realizar isso, mas a principal é a partir dos Dashboards.

Dessa forma, os **Dashboards** são a **combinação de gráficos, tabelas, indicadores de desempenho (KPIs)** e outras visualizações que irão te proporcionar uma visão clara e imediata do desempenho de uma empresa, projeto, processo ou sistema.



Relatório de Vendas - Delivery



Mas o mais importante não é o gráfico em si, e sim a pergunta que você quer responder com aquela informação.



Existe um conceito que auxilia muito nesse momento que é o **Storytelling**. Esse conceito é muito utilizado por times de marketing e vendas, mas cada vez mais profissionais de dados estão se beneficiando disso.

Quando os dashboards são construídos com um Storytelling e os números, gráficos e tabelas respondem perguntas que de fato podem mudar aquela empresa de patamar ou que ajudou em alguma tomada de decisão, **o time de dados atingiu o seu objetivo e gerou insights significativos para a empresa.**

