



Módulo 2

Coleta de dados: fontes internas e externas e automação de coleta de dados



Agora que já conhecemos as principais ferramentas, o próximo passo é entender com mais profundidade como podemos coletar os dados de que você precisa para suas análises e decisões estratégicas. A coleta de dados é uma **etapa fundamental para qualquer empresa** que busca se apoiar em informações precisas e relevantes para melhorar suas operações, desenvolver novos produtos ou simplesmente entender melhor seu mercado e seus clientes.

Existem diversas formas de realizar essa coleta, e elas podem ser divididas em fontes internas e externas, além de contarem com a automação para simplificar processos e tornar a análise mais eficiente.

Vamos explorar cada uma dessas opções com mais detalhes.



Fontes Internas de Dados

Imagine que você trabalha em uma startup que está desenvolvendo um novo aplicativo de entrega. As informações sobre os clientes, como o histórico de pedidos, feedbacks fornecidos após cada entrega e o comportamento dentro do app (como as interações e o tempo gasto em cada funcionalidade), são exemplos clássicos de **dados internos**. Esses dados já estão disponíveis dentro da empresa, armazenados em sistemas como **CRMs, bancos de dados, sistemas de atendimento ao cliente e até mesmo logs do aplicativo**.

Esses dados são extremamente valiosos, pois refletem diretamente a experiência do cliente com a sua marca. Eles podem revelar padrões de comportamento que ajudam a identificar oportunidades de melhoria no produto ou serviço, além de fornecer insights sobre como otimizar a experiência do usuário. A coleta e organização adequada desses dados internos permite que eles sejam estruturados de uma maneira que facilite a análise posterior. O desafio aqui está em garantir que esses dados sejam **coletados corretamente** e estejam sempre **atualizados e organizados** dentro das ferramentas e softwares internos da empresa.

Além disso, é crucial garantir a **qualidade dos dados internos**: informações **duplicadas, incompletas ou desatualizadas** podem comprometer toda a análise. Por isso, uma parte importante do processo de coleta é também garantir que os dados sejam **higienizados e padronizados**, eliminando erros e inconsistências.



Fontes Externas de Dados:

Por outro lado, nem todas as informações relevantes estão disponíveis dentro da sua empresa. Muitas vezes, é preciso buscar dados externos para complementar suas análises. **Isso pode incluir dados de mercado, informações sobre concorrentes, tendências do setor ou até mesmo dados demográficos de potenciais clientes.** Para obter essas informações, há várias técnicas que podem ser usadas, como o acesso a APIs públicas (mencionadas no módulo anterior) ou a realização de web scraping em sites públicos.



APIs são interfaces que permitem a comunicação entre sistemas e fornecem dados de forma estruturada. Muitas plataformas, como redes sociais, oferecem APIs que permitem que você colete dados de forma organizada e legal. No entanto, nem sempre existe uma API disponível para os dados que você precisa. Nesses casos, o web scraping se torna uma solução viável. Com o web scraping, você pode capturar dados diretamente de sites, mesmo que esses dados não estejam organizados em uma API ou de maneira fácil de acessar.

Um exemplo comum de web scraping é utilizado por empresas que monitoram a concorrência. Elas podem, por exemplo, **automatizar a captura de preços de produtos em sites concorrentes** para realizar uma análise de competitividade de forma periódica. Também é possível utilizar o web scraping para **coletar notícias** ou **realizar pesquisas de dados de mercado** que não estão prontamente disponíveis em fontes estruturadas.



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, desenvolvimento web, entre outras. 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.

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.





Automação da Coleta de Dados:

A automação é um recurso que facilita (e muito) o processo de coleta de dados, especialmente quando estamos lidando com grandes volumes de informações. Dependendo do **tamanho** da base de dados ou da **frequência** com que os dados são atualizados, a automação pode economizar tempo e garantir uma coleta mais precisa e regular. Ferramentas de **automação**, como scripts em Python, são amplamente utilizadas para esse propósito.



Por exemplo, é possível criar um script que faça a coleta automática de dados de redes sociais sobre a sua marca, capturando menções, comentários e outras interações dos usuários. Da mesma forma, é possível automatizar a captura de preços de concorrentes em sites de e-commerce, coletando essas informações semanalmente e armazenando-as em uma planilha para análise. Dessa forma, a equipe não precisa realizar esse processo **manualmente**, o que não só economiza tempo, mas também reduz a chance de erros humanos.

Outro exemplo prático seria o monitoramento de palavras-chave ou tendências de busca, onde um script pode rodar diariamente, coletando informações de ferramentas como **Google Trends** e armazenando esses dados para análise. A automação permite que você defina uma rotina de captura de dados, seja diária, semanal, mensal ou conforme a necessidade do seu negócio.

Portanto, ao integrar essas diferentes fontes de dados com a automação, você cria um sistema eficiente e confiável de coleta de informações que alimenta suas decisões com dados atualizados e precisos.

