

## Fundamentos: Dados, dados, em todos os lugares.



Figure 1: Dados, dados, em todos os lugares - imagem: <https://img.freepik.com>

### Estudo de caso: novas perspectivas de dados.

Como você aprenderá, é possível encontrar dados em praticamente qualquer lugar. Sempre que você observa e avalia algo no mundo, está coletando e analisando dados. Sua análise ajuda a encontrar maneiras mais fáceis de realizar tarefas, identificar padrões para economizar seu tempo e descobrir novas perspectivas surpreendentes que podem mudar completamente a maneira como você vivencia tudo.

Veja aqui um exemplo real de como um grupo de analistas de dados usou as seis etapas do processo de análise de dados para melhorar o local de trabalho e os processos de negócios. A história deles envolve algo chamado análise de pessoas, também conhecida como análise de recursos humanos ou da força de trabalho. A análise de pessoas é a prática de coletar e analisar dados sobre as pessoas que compõem a força de trabalho de uma empresa com o objetivo de receber insights para melhorar a forma como a empresa opera.

Ser um analista de pessoas envolve o uso de análise de dados para receber insights sobre os funcionários e como eles vivenciam as vidas profissionais. Os insights são usados para definir e criar um local de trabalho mais produtivo e incentivador. Isso pode liberar o potencial dos funcionários, motivá-los a ter o melhor desempenho possível e garantir uma cultura empresarial justa e inclusiva.

As seis etapas do processo de análise de dados que você aprenderá são: **perguntar**, **preparar**, **processar**, **analisar**, **compartilhar** e **agir**. Essas seis etapas se aplicam a qualquer análise de dados. Continue lendo para aprender como

uma equipe de analistas de pessoas usou essas seis etapas para responder a uma pergunta de negócios.

Uma organização tinha uma taxa de rotatividade alta entre as contratações novas. Muitos funcionários deixavam a empresa antes do fim do primeiro ano de trabalho. Os analistas usaram o processo de análise de dados para responder à seguinte pergunta: como a organização pode melhorar a taxa de retenção de funcionários novos?

Veja aqui um resumo do que essa equipe fez, passo a passo.

**PERGUNTAR (ask)** Em primeiro lugar, os analistas precisavam definir como seria o projeto e o que seria um resultado de êxito. Portanto, para determinar isso, eles fizeram perguntas eficazes e colaboraram com líderes e gerentes interessados no resultado da análise do pessoal. Estes foram os tipos de pergunta que eles fizeram:

- O que você acha que os funcionários novos precisam aprender para ter sucesso no primeiro ano de trabalho?
- Você já coletou dados de funcionários novos antes? Em caso afirmativo, podemos ter acesso aos dados históricos?
- Você acredita que os gerentes com taxas de retenção maiores oferecem algo extra ou exclusivo aos funcionários novos?
- Qual você acha que é a principal causa de insatisfação entre os funcionários novos?
- Em que porcentagem você gostaria que a retenção de funcionários aumentasse no próximo ano fiscal?

**PREPARAR (Prepare)** Tudo começou com uma preparação sólida. O grupo criou um cronograma de três meses e decidiu como queria transmitir o progresso às partes interessadas. Também durante essa etapa, os analistas identificaram quais dados precisavam para atingir o resultado bem-sucedido que identificaram na etapa anterior. Nesse caso, os analistas optaram por coletar os dados de uma pesquisa on-line com funcionários novos. Eles desenvolveram perguntas específicas sobre a satisfação dos funcionários com diferentes processos de negócios, como contratação e integração, e a remuneração geral deles. Foram estabelecidas regras para quem teria acesso aos dados coletados. Nesse caso, nenhuma pessoa fora do grupo teria acesso aos dados brutos, mas poderia visualizar os dados resumidos ou agregados. Por exemplo, a remuneração de um indivíduo não estaria disponível, mas as faixas salariais para grupos de indivíduos seriam visíveis. Eles finalizaram com quais informações específicas seriam coletadas e a melhor forma de apresentar os dados visualmente. Os analistas discutiram ideias sobre possíveis problemas relacionados a projetos e dados e como evitá-los.

**PROCESSAR (Process)** O grupo enviou a pesquisa. Grandes analistas sabem respeitar os dados e as pessoas que os fornecem. Como foram os funcionários que forneceram os dados, era importante garantir que todos os funcionários tinham concordado em participar. Os analistas de dados também garantiram que os funcionários entendessem como os dados deles seriam coletados, armazenados, gerenciados e protegidos. Para manter a confidencialidade,

proteger e armazenar os dados de modo eficaz, o acesso ficou restrito a um número limitado de analistas. Coletar e usar dados de maneira ética é uma das responsabilidades de um analista de dados. Em seguida, os dados foram limpos para garantir que estivessem completos e corretos e que eram relevantes. Certos dados foram agregados e resumidos sem revelar respostas individuais. Os dados brutos foram carregados em um armazenamento de dados interno para garantir uma camada extra de segurança.

**ANALISAR (Analyse)** Então, os analistas fizeram o que sabem fazer de melhor: analisar. Pelas pesquisas concluídas, os analistas de dados descobriram que a experiência de um funcionário novo com certos processos era um indicador-chave da satisfação geral com o trabalho. Os analistas descobriram que os funcionários que passavam por um processo de contratação longo e complicado tinham maior probabilidade de deixar a empresa. Os funcionários que passavam por um processo de avaliação e feedback eficiente e transparente tinham maior probabilidade de permanecer na empresa. O grupo sabia que era importante documentar exatamente o que haviam encontrado na análise, não importando os resultados. Fazer de outra forma diminuiria a confiança no processo da pesquisa e reduziria a capacidade de coletar dados verdadeiros dos funcionários no futuro.

**COMPARTILHAR (Share)** Assim como se certificaram de que os dados estivessem cuidadosamente protegidos, os analistas também tiveram cuidado no momento de compartilhar o relatório. Apenas os gerentes que atingiram ou ultrapassaram o número mínimo de subordinados diretos com respostas enviadas à pesquisa estavam qualificados para receber o relatório. Primeiro, o grupo apresentou os resultados aos gerentes qualificados para ter certeza de que eles tinham a visão completa. Em seguida, pediram que entregassem os resultados às equipes. Isso deu aos gerentes a oportunidade de comunicarem os resultados no contexto certo. Assim, eles poderiam ter conversas produtivas com a equipe sobre as próximas etapas para melhorar o envolvimento dos funcionários.

**Agir (Act)** A última etapa do processo para a equipe de analistas foi trabalhar com os líderes da empresa e decidir a melhor forma de implementar as mudanças e agir com base nas descobertas. Os analistas recomendaram padronizar o processo de contratação e avaliação de todas as contratações novas com base nas práticas mais eficientes e transparentes. Um ano depois, a mesma pesquisa foi distribuída aos funcionários. Os analistas anteciparam que uma comparação entre os dois conjuntos de resultados indicaria que o plano de ação funcionou. Acontece que as mudanças melhoraram o índice de retenção de funcionários novos e as ações das lideranças foram bem-sucedidas.

## Dados e instinto

Detetives e analistas de dados têm muito em comum. Os dois grupos dependem de fatos e pistas para tomar decisões. Ambos coletam e examinam as evidências. Os dois conversam com pessoas que conhecem parte da história. E ambos podem até seguir algumas pistas para ver aonde elas levam. Seja você um detetive ou um analista de dados, seu trabalho consiste em seguir etapas para coletar e compreender fatos.

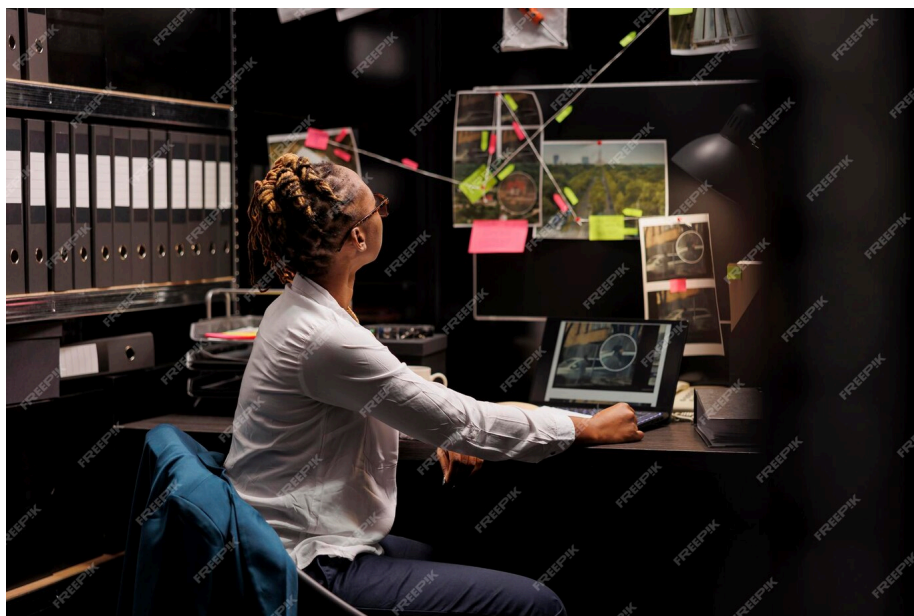


Figure 2: Dados e instinto - imagem: <https://img.freepik>

Os analistas usam a tomada de decisões com base em dados e seguem um processo passo a passo. Você aprendeu que há seis etapas nesse processo:

1. **Fazer perguntas** e definir o problema.
2. **Preparar** os dados, coletando e armazenando as informações.
3. **Processar** os dados, limpando e verificando as informações.
4. **Analisar** os dados para encontrar padrões, relações e tendências.
5. **Compartilhar** dados com seu público.
6. **Agir** com base nos dados e usar os resultados da análise.

No entanto, há outros fatores que influenciam o processo de tomada de decisões. Você pode ter lido histórias de mistério em que o detetive usava o instinto e seguia um palpite que o ajudava a resolver o caso. Instinto é uma compreensão intuitiva de algo com pouca ou nenhuma explicação. Nem sempre isso é algo consciente. Com frequência, captamos os sinais sem nem perceber. Você apenas tem uma “sensação” de que tem razão.

Pense no exemplo de uma licitação de uma incorporadora imobiliária para reconstruir uma parte do distrito central de uma cidade. Ela era conhecida pela preservação de edifícios históricos. Com base na reputação que tinha, os planejadores da agência seguiram o instinto e incluíram a preservação de vários edifícios para conseguir apoio e aprovação para o projeto. No entanto, as doações privadas ficaram aquém e nenhuma parceria chegou a tempo para resolver a situação. Depois de muito atraso e uma cara disputa com a cidade, os edifícios tiveram de ser demolidos.

Quanto mais você entender os dados relacionados a um projeto, mais fácil será descobrir o que é necessário. Esses esforços também ajudarão a identificar erros e lacunas nos dados para conseguir comunicar as descobertas com mais eficácia.

Às vezes, a experiência anterior ajuda você a fazer uma conexão que ninguém mais notaria. Por exemplo, um detetive pode resolver um caso porque se lembra de outro antigo igual ao que está resolvendo no momento. Isso não é apenas instinto.

### **Porque o instinto pode ser um problema.**

Os dados estão no centro da tomada de decisões com base em dados. Portanto, é essencial que os analistas de dados se concentrem nos dados para garantir que tomem decisões informadas. Se você ignorar os dados, preferindo tomar decisões com base na própria experiência, elas poderão ser tendenciosas. Mas, pior ainda, decisões com base no instinto e sem nenhum dado para apoiá-las podem causar erros.

Pense no exemplo de uma licitação de uma incorporadora imobiliária para reconstruir uma parte do distrito central de uma cidade. Ela era conhecida pela preservação de edifícios históricos. Com base na reputação que tinha, os planejadores da agência seguiram o instinto e incluíram a preservação de vários edifícios para conseguir apoio e aprovação para o projeto. No entanto, as doações privadas ficaram aquém e nenhuma parceria chegou a tempo para resolver a situação. Depois de muito atraso e uma cara disputa com a cidade, os edifícios tiveram de ser demolidos.

Quanto mais você entender os dados relacionados a um projeto, mais fácil será descobrir o que é necessário. Esses esforços também ajudarão a identificar erros e lacunas nos dados para conseguir comunicar as descobertas com mais eficácia. Às vezes, a experiência anterior ajuda você a fazer uma conexão que ninguém mais notaria. Por exemplo, um detetive pode resolver um caso porque se lembra de outro antigo igual ao que está resolvendo no momento. Isso não é apenas instinto.

### **Dados + conhecimento dos negócios = mistério resolvido.**

Combinar dados com conhecimento de negócios e talvez com um toque de instinto será uma parte comum do seu processo como analista de dados júnior. A chave é descobrir a combinação exata para cada projeto específico. Muitas vezes, isso dependerá das metas de sua análise. É por isso que os analistas costumam perguntar: “Como posso definir o sucesso desse projeto?”.

Além disso, tente fazer estas perguntas sobre um projeto para ajudar a encontrar o equilíbrio perfeito:

- Que tipo de resultado é necessário?
- Quem será informado?
- Estou respondendo à pergunta que está sendo feita?
- Com que rapidez uma decisão precisa ser tomada?

próprio conhecimento e em sua experiência mais do que o normal. Não há tempo suficiente para analisar completamente todos os dados disponíveis. No entanto, se você tiver um projeto que envolva muito tempo e recursos, a melhor estratégia será se basear mais nos dados. Cabe a você, como analista de dados, fazer a melhor escolha possível. Provavelmente, você combinará dados e conhecimento

de um milhão de maneiras diferentes ao longo da carreira em análise de dados. Quanto mais você praticar, melhor encontrará a combinação perfeita.

### Origens do processo de análise de dados.



Figure 3: Análise de Dados - imagem: <https://img.freepik.com>

Quando você decidiu aderir a este curso, provou ser uma pessoa curiosa. Portanto, vamos explorar sua curiosidade e falar das origens da análise de dados. Não sabemos totalmente quando ou por que a primeira pessoa decidiu registrar dados sobre pessoas e coisas. Mas sabemos que isso foi útil, porque a ideia existe ainda hoje.

Também sabemos que as raízes da análise de dados estão na estatística, que também têm uma história longa. Os arqueólogos consideram a construção das pirâmides, no antigo Egito, como o início da estatística. Os egípcios antigos eram mestres na organização de dados. Eles documentaram os cálculos e as teorias em papiros (materiais semelhantes a papel), que agora são vistos como os primeiros exemplos de planilhas e listas de verificação. Os analistas de dados atuais devem muito a esses escribas brilhantes, que ajudaram a criar um processo mais técnico e eficiente.

É hora de entrar no ciclo de vida da análise de dados, o processo de ir dos dados à decisão. Os dados passam por várias fases conforme são criados, consumidos, testados, processados e reutilizados. Com um modelo de ciclo de vida, todos os principais membros da equipe podem alcançar o sucesso, planejando o trabalho tanto no início quanto no fim do processo de análise de dados. O ciclo de vida da análise de dados é bem conhecido entre os especialistas, mas não existe uma única estrutura definida dessas fases. Pode não haver uma única arquitetura uniformemente seguida por todos os especialistas em análise de dados, mas há

alguns fundamentos compartilhados em todos os processos de análise de dados. Neste texto, fornecemos uma visão geral de vários deles, começando com o processo que constitui a base do Certificado de Análise de Dados do Google.

O processo apresentado é valioso à medida que você avança no curso:

1. **Perguntar:** pergunta/objetivo/desafio de negócios.
2. **Preparar:** geração, coleta, armazenamento e gerenciamento de dados.
3. **Processar:** limpeza/integridade de dados.
4. **Analisar:** exploração, visualização e análise de dados.
5. **Compartilhar:** comunicação e interpretação de resultados.
6. **Agir:** aplicação dos insights com o objetivo de resolver o problema.

Compreender esse processo (e todas as iterações que ajudaram a torná-lo conhecido) será uma grande parte na orientação de sua própria análise e de seu trabalho neste programa. Vamos examinar algumas outras variações do ciclo de vida da análise de dados corporativos.

## CICLO DE VIDA DE DADOS CORPORATIVOS.

### Ciclo de vida da análise de dados da EMC.

O ciclo de vida da análise de dados da EMC Corporation é cíclico, com seis etapas:

1. Descoberta.
2. Pré-processamento de dados.
3. Planejamento de modelo.
4. Criação de modelos.
5. Comunicação dos resultados.
6. Operacionalização.

A EMC Corporation agora é Dell EMC. Este modelo, criado por David Dietrich, reflete a natureza cíclica dos projetos reais. As fases não são marcos estáticos: cada etapa se conecta e leva à próxima e, no fim das contas, se repete. As perguntas-chave ajudam os analistas a testar se realizaram o suficiente para seguir em frente e garantir que as equipes tenham passado tempo suficiente em cada uma das fases e não comecem a modelar antes de os dados ficarem prontos. É um pouco diferente do ciclo de vida da análise de dados em que este programa se baseia, mas tem algumas ideias centrais em comum: a primeira fase está interessada em descobrir e fazer perguntas, os dados precisam ser preparados antes de serem analisados e usados, e então as descobertas precisam ser compartilhadas e colocadas em prática.

### Ciclo de vida iterativo da SAS

Um ciclo de vida iterativo foi criado por uma empresa chamada SAS, uma provedora de soluções de análise de dados líder no mercado. Ele pode ser usado para produzir resultados repetíveis, confiáveis e preditivos:

1. Perguntar.
2. Preparar.
3. Explorar.

4. Modelar.
5. Implementar.
6. Agir.
7. Avaliar.

O modelo SAS enfatiza a natureza cíclica do modelo, visualizando-o como um símbolo de infinito. O ciclo de vida dele tem sete etapas, muitas das quais vimos em outros modelos, como Perguntar, Preparar, Modelar e Agir. Mas esse ciclo de vida também é um pouco diferente. Ele inclui uma etapa após a fase de ação, desenvolvida para ajudar os analistas a avaliar as soluções e, potencialmente, retornar de novo à fase das perguntas.

### **Ciclo de vida da análise de dados com base em projeto**

Um ciclo de vida da análise de dados com base em projeto tem cinco etapas simples:

1. Identificação do problema.
2. Desenvolvimento dos requisitos de dados.
3. Pré-processamento de dados.
4. Execução da análise de dados.
5. Visualização dos dados.

Esse ciclo de vida do projeto de análise de dados foi desenvolvido por Vignesh Prajapati. Não inclui a sexta fase, ou o que chamamos fase da ação (Agir). No entanto, ainda aborda muitas das mesmas etapas dos ciclos de vida que já descrevemos. Ele começa com a identificação do problema e a preparação e o processamento dos dados antes da análise e termina com a visualização dos dados.

### **Ciclo de vida da análise de big data**

Os autores Thomas Erl, Wajid Khattak e Paul Buhler propuseram um ciclo de vida da análise de big data no livro *Big Data Fundamentals: Concepts, Drivers & Techniques*. O ciclo de vida deles sugere fases divididas em nove etapas:

1. Avaliação do caso de negócios.
2. Identificação dos dados.
3. Aquisição e filtragem dos dados.
4. Extração dos dados.
5. Validação e limpeza dos dados.
6. Agregação e representação dos dados.
7. Análise dos dados.
8. Visualização dos dados.
9. Utilização dos resultados da análise.

Esse ciclo de vida parece ter três ou quatro etapas a mais do que os modelos de ciclo de vida anteriores. Mas, na realidade, eles apenas dividiram o que chamamos de fases “Preparar” e “Processar” em etapas menores. Isso enfatiza as tarefas individuais necessárias para coletar, preparar e limpar os dados antes da fase de análise.



### Principal conclusão.

De nossa jornada até as pirâmides e os dados do antigo Egito até agora, a forma como analisamos os dados evoluiu (e continua evoluindo). O processo de análise de dados é como a arquitetura da vida real: há diferentes maneiras de realizar as tarefas, mas as mesmas ideias centrais ainda aparecem em cada modelo do processo.

### Variações do ciclo de vida dos dados.

Há seis estágios no ciclo de vida dos dados:

1. **Planejamento:** decida que tipos de dados são necessários, como serão gerenciados e quem será o responsável por eles.
2. **Captura:** colete ou traga dados de várias fontes diferentes.
3. **Gerenciamento:** cuide e mantenha os dados. Isso inclui determinar como e onde eles são armazenados e as ferramentas usadas para fazer isso.
4. **Análise:** Use os dados para resolver problemas, tomar decisões e auxiliar as metas de negócios.
5. **Arquivamento:** mantenha os dados relevantes armazenados para referência futura e de longo prazo.
6. **Destruição:** remova os dados do armazenamento e exclua todas as cópias compartilhadas deles.

**Aviso:** tenha cuidado para não misturar ou confundir os seis estágios do ciclo de vida dos dados (planejamento, captura, gerenciamento, análise, arquivamento e destruição) com as seis fases do ciclo de vida da análise de dados (pergunta, preparação, processamento, análise, compartilhamento e ação). Eles não devem ser usados ou mencionados de maneira intercambiável.

O ciclo de vida dos dados fornece uma estrutura genérica ou comum de como os dados são gerenciados. Talvez você lembre que as variações do ciclo de vida da análise de dados foram descritas em Origens do processo de análise de dados. O mesmo pode ser feito para o ciclo de vida dos dados. No restante deste texto, você terá uma ideia de como as instituições governamentais, financeiras e educacionais podem ver os ciclos de vida dos dados de maneira um pouco diferente.

### U.S. Fish and Wildlife Service.

O U.S. Fish and Wildlife Service (Serviço de pesca e vida selvagem dos EUA) usa o seguinte ciclo de vida de dados:

1. Planejamento.
2. Aquisição.
3. Manutenção.
4. Acesso.
5. Avaliação.
6. Arquivamento.

### O U.S. Geological Survey (USGS).

O USGS (Pesquisa geológica dos EUA) usa o ciclo de vida dos dados abaixo:

1. Planejamento.
2. Aquisição.
3. Processamento.
4. Análise.
5. Preservação.
6. Publicação/compartilhamento.

Várias atividades transversais ou abrangentes também são realizadas durante cada estágio do ciclo de vida:

- Descrição (metadados e documentação)
- Gerenciamento da qualidade
- Backup e proteção

### **Instituições financeiras.**

As instituições financeiras podem adotar uma abordagem um pouco diferente do ciclo de vida dos dados, conforme descrito em The Data Life Cycle (O ciclo de vida dos dados, em inglês), um artigo da revista Strategic Finance:

1. Captura.
2. Qualificação.
3. Transformação.
4. Utilização.
5. Relatório.
6. Arquivamento.
7. Limpeza.

### **Harvard Business School (HBS).**

1. Geração.
2. Coleta.
3. Processamento.
4. Armazenamento.
5. Gerenciamento.
6. Análise.
7. Visualização.
8. Interpretação.

### **Principal conclusão.**

Compreender a importância do ciclo de vida dos dados vai preparar você para o sucesso como analista de dados. Os estágios individuais do ciclo de vida dos dados variam de empresa para empresa ou por indústria ou setor. Os dados históricos são importantes para o U.S. Fish and Wildlife Service e o USGS. Portanto, o ciclo de vida dos dados deles se concentra no arquivamento e no backup de dados. Os interesses da Universidade de Harvard são pesquisa e ensino. Portanto, o ciclo de vida dos dados inclui visualização e interpretação, mesmo que associados com mais frequência a um ciclo de vida de análise de dados. O ciclo de vida dos dados da HBS também não exige um estágio para limpar ou destruir os dados. Em contraste, o ciclo de vida dos dados de finanças identifica claramente os estágios de arquivamento e limpeza. Resumindo, os

ciclos de vida dos dados variam, mas um princípio de gerenciamento de dados é universal. Controle como os dados são tratados para que sejam precisos, seguros e estejam disponíveis para atender às necessidades da sua organização.

## **Principais ferramentas do analista de dados.**

Conforme você está aprendendo, os programas e soluções mais comuns usados por analistas de dados incluem planilhas, linguagens de consulta e ferramentas de visualização. Neste texto, você aprenderá mais sobre cada um deles. Você verá quando usá-los e por que eles são tão importantes na análise de dados.

### **Planilhas**

Os analistas de dados contam com planilhas para coletar e organizar os dados. Dois aplicativos de planilhas conhecidos que você provavelmente usará muito em seu futuro papel como analista de dados são o Microsoft Excel e o Planilhas Google.

As planilhas estruturam os dados de maneira significativa, possibilitando que você:

- Colete, armazene, organize e classifique as informações.
- Identifique padrões e reúna os dados de uma maneira que funcione para cada projeto de dados específico.
- Crie excelentes visualizações de dados, como gráficos e tabelas.

### **Bancos de dados e linguagens de consulta**

Um banco de dados é uma coleção de dados estruturados armazenados em um sistema de computador. Alguns programas conhecidos de linguagem de consulta estruturada (SQL, na sigla em inglês) incluem MySQL, Microsoft SQL Server e BigQuery.

Linguagens de consulta:

- Permita que os analistas isolem informações específicas de um banco de dados.
- Facilite o aprendizado e a compreensão das solicitações feitas aos bancos de dados.
- Permita que os analistas selecionem, criem, adicionem ou façam o download de dados de um banco de dados para análise.

### **Ferramentas de visualização**

Os analistas de dados usam várias ferramentas de visualização, como gráficos, mapas, tabelas, gráficos e muito mais. Duas ferramentas de visualização conhecidas são o Tableau e o Looker.

Essas ferramentas:

- Transformam números complexos em uma história que as pessoas conseguem entender.
- Ajudam as partes interessadas a chegar a conclusões que levem a decisões informadas e a estratégias de negócios eficazes.

- Têm vários recursos.

O simples recurso de arrastar e soltar do **Tableau** permite que os usuários criem gráficos interativos em painéis e planilhas.

O **Looker** se comunica diretamente com um banco de dados, permitindo que você conecte seus dados diretamente à ferramenta visual que você escolher.

Uma carreira como analista de dados também envolve o uso de linguagens de programação, como R e Python, que são muito usadas para análise estatística, visualização e outras análises de dados.

### Principal conclusão

Você tem muitas ferramentas como analista de dados. Esta é uma apresentação das possibilidades, e você conhecerá muitas dessas ferramentas profundamente ao longo deste curso.

### Como escolher a ferramenta certa para o trabalho.

Como analista de dados, geralmente você terá que decidir qual solução ou programa é o certo para o projeto específico em que está trabalhando. Neste texto, você aprenderá mais sobre como escolher qual ferramenta e quando.

Dependendo da fase do processo de análise de dados em que você se encontrar, será preciso usar ferramentas diferentes. Por exemplo, se você estiver se concentrando na criação de visualizações complexas e atraentes, as ferramentas de visualização que discutimos anteriormente serão a melhor escolha. Mas, se você estiver se concentrando na organização, limpeza e análise de dados, provavelmente escolherá entre planilhas e bancos de dados que usam consultas. Planilhas e bancos de dados oferecem maneiras de armazenar, gerenciar e usar dados. O conteúdo básico para ambas as ferramentas são conjuntos de valores. No entanto, também existem algumas diferenças importantes:

Planilhas	Bancos de dados
Aplicativos de software	Armazenamentos de dados: acessados com o uso de uma linguagem de consulta (por exemplo, SQL)
Estruturar dados em formato de linha e coluna	Estruturar dados usando regras e relacionamentos
Organizar informações em células	Organizar informações em coleções complexas
Conceder acesso a uma quantidade limitada de dados	Conceder acesso a grandes quantidades de dados
Entrada manual de dados	Entrada de dados estrita e consistente
Geralmente um usuário por vez	Vários usuários
Controlado pelo usuário	Controlado por um sistema de gerenciamento de banco de dados

Você não precisa escolher entre uma ou outra, pois cada uma serve a um propósito específico. Geralmente, os analistas de dados trabalham com uma

combinação das duas, pois ambas as ferramentas são muito úteis na análise de dados. Por exemplo, é possível armazenar dados em um banco de dados e, em seguida, exportá-los para uma planilha para realizar uma análise. Ou, se você estiver coletando informações em uma planilha e isso se tornar demais para a plataforma específica, será possível importá-las para um banco de dados.

Referências: - Certificado de Análise de Dados do Google.