

*Chegou seu...*

GUIA DE BOLSO

# ANÁLISE DE DADOS

*Alguns conceitos e ferramentas*

ALEX SOUZA

## Sumário

ETL .....	4
OLTP e OLAP: O que são e quais as diferenças? .....	4
Em resumo... ..	4
Mas, qual é a diferença entre OLTP e OLAP? .....	4
Referências.....	5
Modelagem Multidimensional (Tabelas: Fato e Dimensão) .....	5
Tabela de Fatos .....	6
Tabela de dimensões.....	6
Observações .....	6
Referências:.....	6
AWS Redshift.....	7
Big Query .....	7
Primeiros passos.....	7
Tópicos sugeridos.....	8
Snowflake (Software) .....	8
Snowflake é projetado em três componentes principais .....	8
Pentaho .....	9
Análise Exploratória de Dados .....	9
Principal objetivo da EDA .....	9
Exemplo.....	9
O que é um Dashboard?.....	10
Tipo de Dashboards.....	10
Dashboard X Relatório .....	10
Referencias.....	10
Power BI .....	10
Exemplos .....	11
Referências.....	11
Tableau .....	11
Google Data Studio .....	11
Looker.....	12
Melhorar a tomada de decisões .....	12
Crie experiências ricas em dados que vão além da BI .....	12
Acelere seus negócios com fluxos de trabalho baseados em dados .....	12
AWS Quicksight .....	12
Principais benefícios.....	13

QlikSense e Qlikview .....	13
Como usar o Qlik Sense? .....	13
O que é o sistema QlikView? .....	14
Por que usar QlikView? .....	14
Diferença entre o Qlik Sense e o QlikView .....	14
Adobe Analytics.....	14
Databricks (SQL Analytics).....	14
Vejam também.....	15
Redash .....	16
Knime .....	16
Dundas BI .....	17
Metabase.....	17
R.....	18
Google Sheets.....	18
Sheets para Análise .....	18
Storytelling .....	19
Como mentir com dados - Usando visualização de dados? .....	19
Exemplo 1 .....	20
Exemplo 2 .....	20

## ETL

### OLTP e OLAP: O que são e quais as diferenças?

**OLTP e OLAP** São duas siglas muito frequentes no universo do business intelligence (BI) e que apesar de apresentarem nomes parecidos, **seus conceitos e aplicações são bem diferentes.**

**OLTP** ou *Online Transaction Processing* (Processamento de Transações Online, em português) **são sistemas que se encarregam de registrar todas as transações contidas em uma determinada operação organizacional.** Por exemplo: sistema de transações bancárias que registra todas as operações efetuadas em um banco, caixas de multibanco, reservas de viagens ou hotel on-line, Cartões de Crédito.

**OLAP** ou *Online Analytical Processing* (Processamento Analítico Online) **é a capacidade para manipular e analisar um grande volume de dados sob múltiplas perspectivas. As aplicações são usadas pelos gestores em qualquer nível da organização para lhes permitir análises comparativas que facilitem a sua tomada de decisões diárias.** A arquitetura OLAP possui ferramentas que são classificadas em cinco tipos que são:

[ROLAP](#) | [MOLAP](#) | [HOLAP](#) | [DOLAP](#) | WOLAP (além de XOLAP).

#### Em resumo...

**OLTP** serve para designar os sistemas operacionais com dados transacionais. **(Ambiente de banco onde você insere, altera e ler dados o tempo todo.)**

**OLAP** é uma habilidade para executar dentro do Data Warehouse e realizar análises dos grandes volumes de dados. **(Ambiente específico para trabalhar a parte de análise, tratamento e processamento dos dados para extrair valor deles)**

#### Mas, qual é a diferença entre OLTP e OLAP?

Ao que tange à performance, o **OLTP tem uma velocidade superior no processamento e uso dos dados operacionais**, sem muita eficiência na análise para deliberações administrativas. Já os dados de **OLAP são menos detalhistas, mais sumarizados e repletos de índices.**

Em questão de armazenamento o do **OLTP acontece em bancos de dados mais convencionais.** Já o **OLAP requer uma estrutura mais avançada, como o Data Warehouse**, o qual é capaz de otimizar a performance com extensos volumes de informações.

	OLAP	OLTP
Foco	Foco no nível estratégico da organização. Visa a análise empresarial e tomada de decisão.	Foco no nível operacional da organização. Visa a execução operacional do negócio.
Performance	Otimização para a leitura e geração de análises e relatórios gerenciais.	Alta velocidade na manipulação de dados operacionais, porém ineficiente para geração de análises gerenciais.
Estrutura dos dados	Os dados estão estruturados na modelagem dimensional. Os dados normalmente possuem alto nível de sumarização.	Os dados são normalmente estruturados em um modelo relacional normalizado, otimizado para a utilização transacional. Os dados possuem alto nível de detalhes.
Armazenamento	O armazenamento é feito em estruturas de <i>Data Warehouse</i> com otimização no desempenho em grandes volumes de dados.	O armazenamento é feito em sistemas convencionais de banco de dados através dos sistemas de informações da organização.
Abrangência	É utilizado pelos gestores e analistas para a tomada de decisão.	É utilizado por técnicos e analistas e engloba vários usuários da organização.
Frequência de atualização	A atualização das informações é feita no processo de carga dos dados. Frequência baixa, podendo ser diária, semanal, mensal ou anual (ou critério específico).	A atualização dos dados é feita no momento da transação. Frequência muito alta de atualizações.
Volatilidade	Dados históricos e não voláteis. Os dados não sofrem alterações, salvo necessidades específicas (por motivos de erros ou inconsistências de informações).	Dados voláteis, passíveis de modificação e exclusão.
Tipos de permissões nos dados	É permitido apenas a inserção e leitura. Sendo que para o usuário está apenas disponível a leitura.	Podem ser feito leitura, inserção, modificação e exclusão dos dados.

Com as diferenças mostradas, percebemos que **não se trata de um conceito ser melhor que o outro, mas sim de conceitos complementares e com objetivos distintos dentro da organização**. Cabe à empresa se posicionar e utilizar ambos da melhor forma possível para conciliar desempenho operacional e o resultado estratégico da organização.

## Referências

- <https://medium.com/@aasouzaconsult/aprofundando-em-data-warehouse-65ed2bca9a33>
- <https://www.youtube.com/watch?v=O7tmupHe9Ts>
- <https://www.dataex.com.br/oltp-e-olap/>
- <https://pt.wikipedia.org/wiki/OLTP>
- <https://pt.wikipedia.org/wiki/OLAP>
- <https://canaltech.com.br/business-intelligence/o-que-significa-oltp-e-olap-na-pratica/>

## Modelagem Multidimensional (Tabelas: Fato e Dimensão)

Um Data Warehouse geralmente é um componente central de uma infra-estrutura BI em uma organização, funciona como um repositório de armazenamento e análise de informações numéricas.

Após o entendimento sobre o OLTP e OLAP, cria-se um ambiente DW utilizando a modelagem multidimensional que tem dois tipos:

### Tabela de Fatos

Normalmente são tabelas com milhões de linhas que contam o que aconteceu, um tipo de histórico dentro de uma empresa (Registros Históricos) ...todas as compras feitas em uma loja por exemplo.

+os fatos são as métricas para análises.

+São quantificadores

+Tabela com todo o histórico

### Tabela de dimensões

Tabela que geralmente apresenta características de alguma coisa, não é regra, mas já ajuda a identificar (características dos produtos que são vendidos ..marca, categoria, custo...

+ torna a tabela Fatos menor

+A dimensão é o que dá personalidade e qualidade aos fatos ocorridos

+é a dimensão que nos permite visualizar as informações por diversos aspectos

+Podemos pensar em atributos quando pensamos em dimensão

+ As dimensões são qualificadoras

+tabela de características

### Observações

\*A relação que o power bi faz é relacionar uma TF com uma TD para que o usuário consiga puxar as informações, elas se complementam

\*exemplo, numa tabela fato eu vou ter poucas informações...já na tabela dimensão eu vou ter as características de algo que está na TF

\*não posso relacionar duas tabelas fatos

### Referências:

- <https://medium.com/@aasouzaconsult/aprofundando-em-data-warehouse-65ed2bca9a33>

## AWS Redshift

Amazon Redshift – Lançado em 2012, é um serviço de data warehouse em nuvem de armazenamento de Big Data e é gerenciado pela Amazon Web Services (AWS), isso significa que o cliente não precisa se preocupar com manutenção, provisionar o cluster, instalar atualização, e qual quer outro serviço estrutural. Foi desenvolvido para atender uma variedade de fonte de dados baseado em SQL. É uma ferramenta orientada para análise como reporting, data warehouse e BI.

Seu gerenciamento é feito por clusters, que são formados por um conjunto de “nós”, mas que apresenta um “nó” líder e um ou mais “nós” de computação. O tipo e quantidade de “nós” que o cliente precisa vai variar com o tamanho dos dados, do número de consultas que o usuário irá realizar.

É baixo custo, quando comparado a outras ferramentas do mercado que possuem a mesma proposta. O cliente pode iniciar com apenas alguns gigabytes de memória e escalar até petabytes.

Há backup contínuo, ou seja, ao longo do dia ele faz backups incrementais acompanhando a atualização do cliente.

Ele também funciona em formato colunar, o que impacta na sua velocidade e forma de processamento dos dados.

Saibam mais em: <https://aws.amazon.com/pt/redshift/>

## Big Query

O BigQuery é um serviço de armazenamento de dados de baixo custo e totalmente gerenciado do Google para análises com escala em petabytes. Essa ferramenta é NoOps, ou seja, ela não requer uma infraestrutura responsável pelo gerenciamento nem um administrador de banco de dados. Com o BigQuery, você se concentra na análise de dados para ter acesso a insights relevantes usando um SQL familiar e aprendizado de máquina integrado. Tudo isso a um custo-benefício inigualável.

Localize o BigQuery no menu do lado esquerdo de Google Cloud Platform Console, em **Big Data**.

### Primeiros passos

- O que é o BigQuery? Veja uma introdução ao tema.
- Guia de início rápido: teste a IU da Web do BigQuery, aprenda a consultar um conjunto de dados público e saiba como carregar dados em uma tabela.
- Todos os guias de instruções: instruções detalhadas para todas as tarefas do BigQuery.

## Tópicos sugeridos

- Consulta de dados
- Controle de custos
- Atalhos do teclado
- Como carregar dados JSON aninhados e repetidos
- Como carregar dados no BigQuery
- Criar um aplicativo simples com a API

## Snowflake (Software)

Criado em 2012 o Snowflake é um SaaS (software as a service) totalmente gerenciado que fornece uma plataforma única para data warehousing, data lakes, engenharia de dados, ciência de dados, desenvolvimento de aplicativos de dados e compartilhamento e consumo seguros de dados em tempo real/compartilhados. O Snowflake apresenta recursos prontos para uso, como separação de armazenamento e computação, computação escalável em tempo real, compartilhamento de dados, clonagem de dados e suporte a ferramentas de terceiros para lidar com as necessidades exigentes de empresas em crescimento.

### Snowflake é projetado em três componentes principais

Serviços em nuvem, processamento de consultas e o armazenamento de banco de dados.

- Serviços em nuvem. O Snowflake usa ANSI SQL para serviços em nuvem, permitindo que os usuários otimizem seus dados e gerenciem sua infraestrutura. Os serviços incluem autenticação, gerenciamento de infraestrutura, análise e otimização de consultas, gerenciamento de metadados e controle de acesso.
- Processamento de consultas. A camada de computação do Snowflake é composta por data warehouses de nuvem virtual que permitem analisar dados por meio de solicitações. Cada armazém virtual Snowflake é um cluster independente e não compete por recursos de computação nem afeta o desempenho um do outro — o que significa que a simultaneidade da carga de trabalho nunca é um problema.
- Armazenamento de banco de dados. Um banco de dados Snowflake é onde os conjuntos de dados estruturados e semiestruturados carregados de uma organização são mantidos para processamento e análise. O Snowflake gerencia automaticamente todas as partes do processo de armazenamento de dados, incluindo organização, estrutura, metadados, tamanho do arquivo, compactação e estatísticas.

Existem vários benefícios no Snowflake, a Escalabilidade instantânea e quase ilimitada, a automação facilitada, uma única cópia de dados, compartilhada com segurança, em qualquer lugar e Integrações de dados de terceiros.

Vejam mais em: <https://www.snowflake.com/en/>



## Pentaho

O **Pentaho**® é um software desenvolvido em Java utilizado na área de **Business Intelligence (BI)** que oferece, principalmente, soluções nas áreas de integração de dados (ETL – Extract Transform Load), relatórios, análises online (OLAP – Online Analytical Processing) e mineração de dados.

Ele se tornou o preferido na organização do Big Data por ser open source (código aberto) e por se integrar facilmente a qualquer infraestrutura de TI. Ainda, os relatórios, vistas de análises e dashboards gerados pelo **Pentaho**® são customizáveis e intuitivos.

Além de trazer de organizar e apresentar os dados de acordo com as necessidades do cliente, o **Pentaho**® é focado nos processos de trabalho da organização e traz soluções para a empresa. Ele implementa esses processos na resolução de problemas detectados, além de apresentar uma visão geral do funcionamento da empresa. Mais...

Documentação:

[https://help.hitachivantara.com/Documentation/Pentaho/8.3/Products/Pentaho\\_Data\\_Integration](https://help.hitachivantara.com/Documentation/Pentaho/8.3/Products/Pentaho_Data_Integration)

Download:

[https://events.pentaho.com/CE-Download\\_Data-Integration-ALL-OS.html](https://events.pentaho.com/CE-Download_Data-Integration-ALL-OS.html)

## Análise Exploratória de Dados

Ou, em resumo, EDA (do inglês: *Exploratory Data Analysis*) é uma abordagem para analisar dados para resumir as principais características dos dados, obter uma melhor compreensão do conjunto de dados, descobrir relacionamentos entre diferentes variáveis e extrair variáveis importantes. EDA, é um passo importante em qualquer projeto de Análise de Dados ou Ciência de Dados.

### Principal objetivo da EDA

O **principal objetivo** da EDA é detectar quaisquer erros ou discrepâncias, bem como entender diferentes padrões nos dados. Ele permite que os analistas entendam melhor os dados antes de fazer qualquer suposição e descubram relacionamentos entre variáveis. Os resultados da EDA ajudam as empresas a conhecer seus clientes, expandir seus negócios e tomar melhores decisões.

### Exemplo

Nosso conjunto de dados de exemplo, se trata de Vendas de uma Concessionária de veículos usados.

<https://medium.com/@aasouzaconsult/python-para-an%C3%A1lise-de-dados-24028d7013b3>

## O que é um Dashboard?

Agora iremos abordar sobre dashboards, para que serve, como é utilizado e qual a sua importância para uma empresa.

**Dashboard** é um painel de interface gráfica que contribui visualmente para a observação de informações, indicadores, ações tomadas pela empresa e métricas, de forma organizada.

**O intuito do uso de um Dashboard** é a capacidade de transformar grandes volumes de dados em informações para criar decisões estratégicas de forma clara e com grande assertividade e eficácia para uma empresa.

### Tipo de Dashboards

**Operacional:** Esse tipo de dashboard tem o intuito para equipes dentro da empresa participarem das operações e auxilia no fluxo de trabalho.

**Tático:** É responsável pela por apresentar indicadores para melhorar regras de negócios.

**Estratégico:** Permite a visualização de dados importantes para os negócios gerais e fazer comparações com anos anteriores por exemplo, assim facilitando a tomada de decisões para passos futuros dentro da empresa.

### Dashboard X Relatório

**Relatórios:** apresenta as informações de forma estática e necessita da criação de outros relatórios para atualização de novos dados.

**Dashboards:** expõe as informações em tempo real, não necessitando da criação de vários dashboards já que ele se atualiza automaticamente conforme os dados são dispostos.

### Referencias

- <https://www.alura.com.br/artigos/dashboards-vs-relatorios>
- <https://neilpatel.com/br/blog/dashboard-o-que-e/>

## Power BI

Antes de falar da ferramenta Power BI é necessário falar sobre o que é o Business Intelligence. O termo Business Intelligence foi criado na década de 80, e de forma resumida, consiste na tomada de decisões empresariais baseadas em dados. E é nesse momento que entram as ferramentas para transformar os dados em inteligência de negócios.

O Power BI foi lançado pela Microsoft em 2015 e se tornou líder do mercado, segundo a Gartner, devido ao seu ambiente amigável e de fácil utilização. O software serve para facilitar a visualização de dados através de relatórios e dashboards dinâmicos.

Com ele é possível conectar a mais de 100 fontes de dados diferentes, incluindo sites e arquivos. E com isso o software soluciona mais uma necessidade do mercado, pois com ele você pode relacionar os dados de diversas tabelas que foram extraídas de fontes diferentes.

Após a disponibilização do relatório ou dashboard na internet, os dados poderão ser atualizados automaticamente em até 8 vezes no mesmo dia. Isso facilita e contribui para que os gestores possam ter o controle das informações para tomar a decisão certa, de acordo com cada situação. No Serviço do Power BI o proprietário do projeto escolhe quais colaboradores da sua empresa terão acesso aos dashboards publicados.

É, também, o local em que as atualizações automáticas são configuradas. Otimizar o Business Intelligence da empresa com o Power BI tem muitos benefícios, pois além de ser uma ferramenta que utiliza funções parecidas com o excel, ela ainda é integrada com diversas funcionalidades da Microsoft.

### Exemplos

- <https://community.powerbi.com/t5/Data-Stories-Gallery/Censo-Escolar-2020/td-p/2807378>
- <https://community.powerbi.com/t5/Data-Stories-Gallery/COVID-19/td-p/2806786>

### Referências

- [https://blogdozouza.wordpress.com/2015/11/18/gartner\\_bi/](https://blogdozouza.wordpress.com/2015/11/18/gartner_bi/)
- <https://blogdozouza.wordpress.com/bi/microsoft/power-bi/>
- <https://powerbi.microsoft.com/pt-br/what-is-power-bi>

## Tableau

O Tableau é uma empresa americana de software de visualização de dados interativo e análise de dados para geração de relatórios que foca em inteligência de negócios.

Objetivo: Ajudar as pessoas a ver e entender os dados, pois colocam o usuário em primeiro lugar com sua interface intuitiva, seja analista de dados, cientista, executivo.

Empresa: Foi comprada em 2019 sendo a maior aquisição da Salesforce na época.

Ferramentas: Os produtos tableau consultam banco de dados relacionais, cubos de processamento analítico on-line, bancos de dados em nuvem e planilhas para gerar a visualização de dados para gráfico, e pode conectar com algumas fontes de dados como Microsoft Excel, arquivos de Texto, entre outros.

Mais informações em: <https://blogdozouza.wordpress.com/bi/tableau/>

## Google Data Studio

Google Data Studio é uma ferramenta de business analytics bastante útil no marketing digital. É possível criar dashboards personalizados e atraentes para analisar dados e apresentar resultados de estratégias, é uma ferramenta poderosa para a criação de relatórios e dashboards, que você pode usar para fazer análises, apresentar os resultados do marketing e tomar melhores decisões.

Algumas características:

- Gratuito.
- Lançado em 2016 - Faz parte do pacote de ferramentas do Google Marketing Platform. - Integração com diversas outras plataformas e ferramentas do Google, como Google Analytics e Google Ads.

Mais informações em: <https://rockcontent.com/br/blog/google-data-studio/>

## Looker

Uma plataforma moderna de business intelligence, análise incorporada e aplicativo de dados.

### Melhorar a tomada de decisões

Analise, visualize e tome decisões com base em insights que as pessoas podem confiar como atualizados. Tome decisões mais fundamentadas com uma arquitetura moderna no banco de dados e uma poderosa camada semântica que oferece métricas regidas em tempo real.

### Crie experiências ricas em dados que vão além da BI

Aproveite uma plataforma com foco na API para fornecer dados regidos por toda a organização, com insights integrados, fluxos de trabalho baseados em dados e aplicativos personalizados.

### Acelere seus negócios com fluxos de trabalho baseados em dados

Otimize processos de negócios e aumente a eficiência com fluxos de trabalho baseados em dados que podem simplificar e automatizar tarefas repetitivas anteriormente.

Link para a plataforma: <https://cloud.google.com/resources/looker-free-trial?hl=pt-br>

## AWS Quicksight

O Amazon Quicksight é um serviço para visualização de dados destinado à inteligência comercial (BI) na nuvem (AWS) que não necessita a preocupação com servidores ou máquinas virtuais (infraestrutura) para ser gerenciado (serverless). Conecta dados na nuvem e permite a combinação de dados alimentados por diversas fontes, como da própria AWS, big data, planilhas, banco de dados e até mesmo de aplicativos como Salesforce por exemplo. Como

serviço baseado completamente na nuvem, oferece segurança de nível empresarial, disponibilidade global e redundância. Fornece insights baseados em aprendizado de máquina (ML), assim como detecção e previsão de anomalias integradas para a melhor visualização e interpretação dos dados para o cliente. O QuickSight é alimentado pelo SPICE, um mecanismo de cálculo super-rápido na memória que oferece desempenho e escala, independentemente de quantos usuários estejam ativos.

## Principais benefícios

- **Alta escalabilidade:** O Quicksight permite escalar de 10 usuário para 10000 sem a preocupação com implantação ou gerenciamento de infraestrutura;
- **Pay as you go:** Não necessita de pagamento fixo mensal, adiantado ou fixado em quantidade de usuários que irão utilizar. Pague somente conforme sua necessidade de uso;
- **Serverless:** Não necessita de software para deploy, manutenção, upgrades ou migrações. Serviço completamente gerenciado pela nuvem;
- **Integração:** O Quicksight integra com suas fontes de dados e outros serviços da AWS que disponibiliza tudo que você precisa para obter uma solução de inteligência comercial de ponta a ponta (E2E BI solution);
- **Análises para diferenciar aplicações:** Incorpore facilmente visualizações e painéis interativos, criação de painéis sofisticados ou recursos de consulta de linguagem natural nas aplicações para diferenciar a experiência do usuário e desbloquear novas oportunidades de monetização;
- **QuickSight Q :** O Q permite que os usuários finais mergulhem profundamente nos dados por meio de perguntas simples, sem treinamento de BI. Faça perguntas e o mecanismo alimentado por ML entrega visualizações relevantes instantaneamente;

## QlikSense e Qlikview

O Qlik Sense é uma ferramenta de visualização de informações que explora profundamente todos os dados e revela suas conexões de modo simples e instantâneo, entregando conhecimento acerca das oportunidades de negócio.

Por meio de seu dashboard intuitivo, a ferramenta permite um olhar profundo acerca dos dados para entender a real situação do negócio e tudo o que acontece no mercado.

O Qlik Sense pode ser alimentado por várias fontes de dados, mantendo uma versão gratuita, mais simples, que já permite explorar vários pontos do software e obter suas vantagens.

## Como usar o Qlik Sense?

Ao acionar os dados no **Qlik Sense**, ele apresenta as opções de visualização disponíveis, tais como gráficos em barra, histogramas e mapas, entre outros. Você precisa apenas clicar e arrastar no tipo de visualização que deseja para acessá-la.

## Vantagens

- Visualização de dados intuitiva
- Possibilidade de compartilhamento
- Geração de relatórios simplificada
- Exploração associativa e busca inteligente
- Mobilidade

## O que é o sistema QlikView?

**QlikView** é uma solução de BI que permite criar análises guiadas orientada ao usuário, que auxilia na tomada de decisões a partir de fontes diversas de conhecimento, dados, pessoas e ambiente.

## Por que usar QlikView?

O **QlikView** proporciona rapidez flexibilidade, facilidade de uso e recursos poderosos, com isso ele permite transformar dados vindos de fontes diferentes em conhecimento de negócio, propiciando tomadas de decisões mais precisas e criando novas oportunidades para a empresa.

## Diferença entre o Qlik Sense e o QlikView

O **Qlik Sense** oferece soluções analíticas modernas. O **QlikView** baseia-se em análises guiadas. Os aplicativos são criados por desenvolvedores de documentos do **QlikView**, que planejam minuciosamente o modelo de dados, o layout, os gráficos e as fórmulas. Em seguida, eles fornecem esses aplicativos aos analistas.

## Adobe Analytics

Essa tecnologia se resume em ETL, tem intuito de criar painéis e relatórios para criar visualizações de dados, para posteriormente tomar uma decisão seja ela corporativa ou pessoal. O Adobe Analytics pode se conectar com qualquer dispositivo web para obter dados, ele possui integrações como por exemplo o adobe experience cloud.

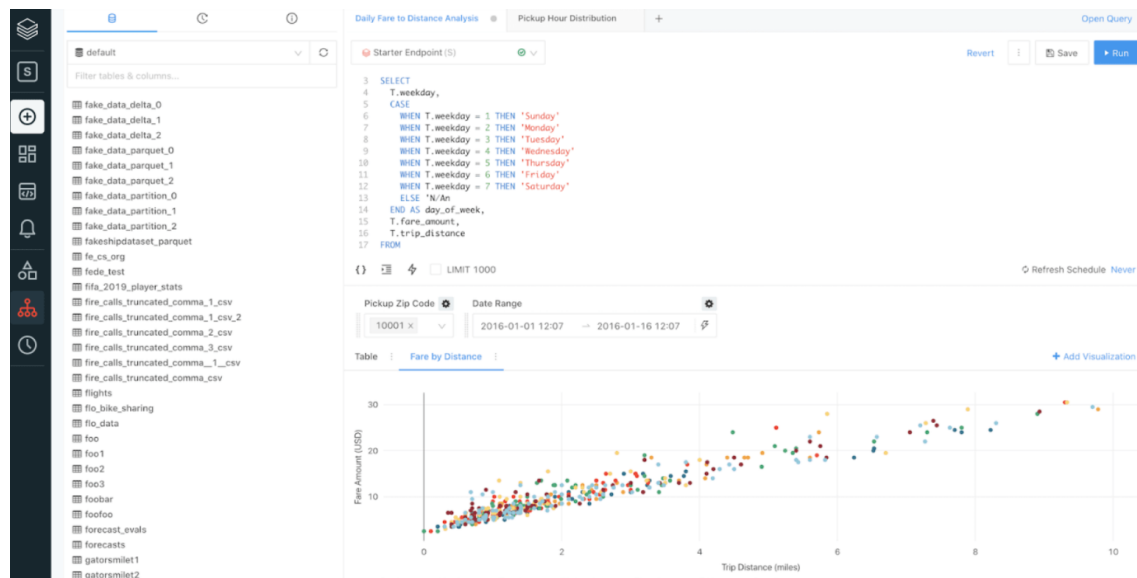
Após extrair os dados da web, é a hora de analisar e quem faz esse papel com excelência são os analistas, eles criam as visões no Adobe analytics e com isso é enviado para o setor estratégico da empresa para a tomada de decisão.

Mais informações: <https://business.adobe.com/br/products/analytics/adobe-analytics.html>

## Databricks (SQL Analytics)

O Databricks SQL permite que os analistas de dados descubram e encontrem rapidamente conjuntos de dados, escrevam consultas em uma sintaxe SQL familiar e explorem facilmente os esquemas de tabela **do Delta Lake** para análise ad hoc. O código SQL usado regularmente pode ser salvo como snippets para reutilização rápida e os resultados da consulta podem ser armazenados em cache para manter os tempos de execução curtos.

### Saber mais...

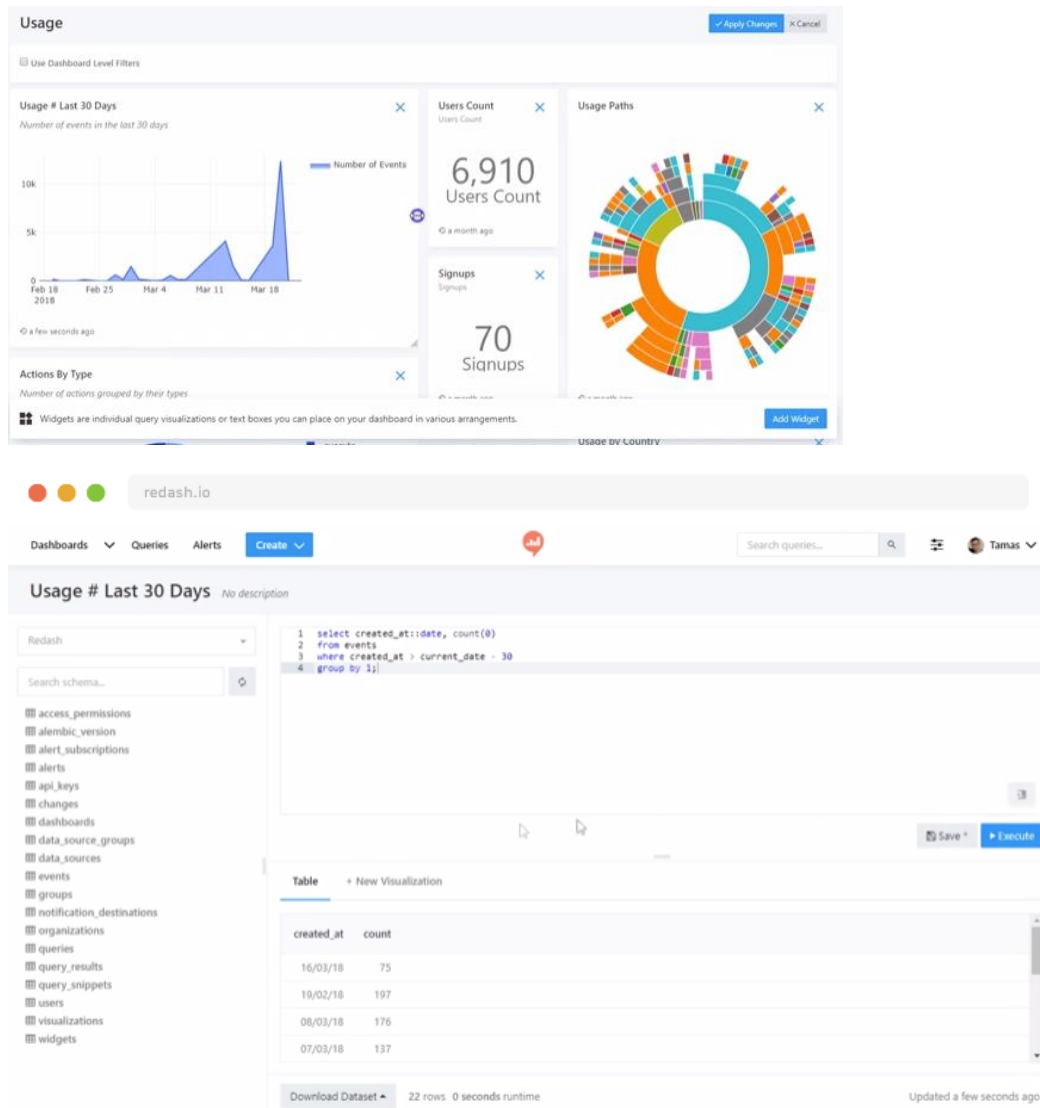


### Vejam também

- <https://www.databricks.com/product/databricks-sql>
- <https://medium.com/blog-do-zouza/lendo-tabelas-do-databricks-community-pelo-power-bi-2a44e801ec76>

## Redash

O Redash é uma solução open-source, parte da Databricks, feita para permitir que qualquer um, independentemente do nível de sofisticação técnica, se beneficie do poder dos dados criando gráficos e dashboards em SQL. Uma vez que você tenha conectado o Redash ao seu Data Warehouse ou Data Lake, a Kondado irá disponibilizar e atualizar os seus dados para consulta no Redash e, a partir de lá, você poderá criar visualizações, gráficos, dashboards e KPIs utilizando o Redash.



Mais informações: <https://redash.io/>

## Knime

Na KNIME, construímos software para criar e produzir ciência de dados usando um ambiente fácil e intuitivo, permitindo que todas as partes interessadas no processo de ciência de dados se concentrem no que fazem de melhor.



KNIME Analytics Platform é o software de código aberto para criar ciência de dados. Intuitivo, aberto e continuamente integrando novos desenvolvimentos, o KNIME torna a compreensão de dados e o design de fluxos de trabalho de ciência de dados e componentes reutilizáveis acessíveis a todos

Mais informações: <https://www.knime.com/>

## Dundas BI

Ofereça uma experiência de análise incomparável em todos os setores, usando a primeira plataforma de BI de ponta a ponta totalmente programável de seu tipo.

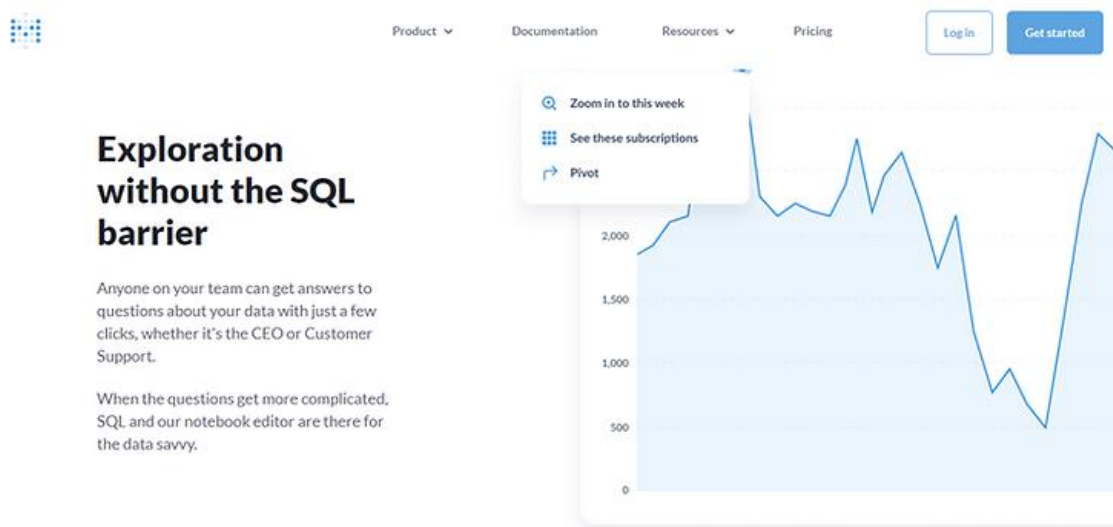
Personalizada para se adequar ao seu caso de uso específico, nossa plataforma flexível combina visualizações de dados inteligentes e interativas, APIs abertas e recursos de BI completos e avançados para análise de dados sem esforço – tudo em uma plataforma integrada.

Mais informações: <https://www.dundas.com/dundas-bi/platform>

## Metabase

O Metabase é uma ferramenta open source especialmente interessante para a parte administrativa de uma empresa que precisa entender mais sobre os dados gerados em sua companhia, sem precisar ser um expert no assunto. A ferramenta de business intelligence é capaz de transformar os dados e suas estruturas em algo palpável através de filtros orgânicos e uma estética confortável para a visualização.

Nela os dados são transformados em dashboards atualizadas, e em relatórios e gráficos simples e rápidos, facilmente entendidos. Ela se adequa a todos que queiram usá-la, ou seja, não importa se você é o técnico mais experiente, o dono da empresa ou o responsável pelas relações financeiras, a plataforma torna possível o compartilhamento dessas informações com didática impecável.



Mais informações: <https://www.metabase.com/>

## R

R é uma linguagem de programação multi-paradigma orientada a objetos, programação funcional, dinâmica, fracamente tipada, voltada à manipulação, análise e visualização de dados. Foi criado originalmente por Ross Ihaka e por Robert Gentleman no departamento de Estatística da Universidade de Auckland, Nova Zelândia.

Permite: extração, limpeza, carregamento e transformação de dados, além de manipulação de bancos de dados e visualização de dados.

Python também pode ser utilizado para [visualização](#) e [análise de dados](#).

Links úteis:

- <https://www.r-project.org/>
- <https://www.rstudio.com/products/rstudio/download/>
- <https://blogdozouza.wordpress.com/ia/linguagens/codigos-em-r-python-matlab/>

## Google Sheets

É um editor de planilhas do Google que possui uma interface amigável e funcionalidades muito semelhantes ao Microsoft Excel, porém, é uma ferramenta gratuita e integrada ao Google Drive. O serviço também inclui Google Docs, Google Slides, Google Drawings, Google Forms, Google Sites e Google Keep

**Acesso** - Para acessá-lo existem dois caminhos, diretamente no site do Google Sheets ou clicando em ou por meio do Google Drive.

**Vantagens** - *Armazenamento em nuvem* - A tecnologia da nuvem salva os documentos de forma segura sem usar a memória do celular ou *laptop*. Os documentos são acessados e compartilhados pelo celular, pelo navegador ou computador.

*Colaboração em tempo real* - Com o Google Planilhas é possível compartilhar sua planilha, com ou sem restrições de modificação. Além disso, o programa permite a colaboração em tempo real com outros usuários e a criação de comentários para os demais colegas.

*Manipulação de dados* - Permite aplicar filtros, consolidar informações, buscar dados duplicados, organizar de forma crescente ou decrescente.

**Desvantagem** - *Limite de células* – A quantidade limite de células do Google Sheets é de 5 milhões. Pensando no contexto de grandes empresas, seriam necessárias diversas planilhas para compilar e analisar dados.

**Preço** - O Google Sheets é uma ferramenta gratuita, acessada a partir da criação de um login no Google.

Sheets para Análise: <https://support.google.com/docs/answer/9330962?hl=pt-BR>

## Storytelling

Exploração de dados com propósito, o objetivo do storytelling é criar uma estratégia para mostrar o negócio e como esse negócio pode ser impulsionado. O storytelling é um conceito que reúne várias técnicas que são usadas para colocar em prática a arte de contar histórias envolventes, impactantes e inesquecíveis que “vendem” uma ideia, propósito, conceito ou projeto e que despertam sentimentos de identificação, esse conceito pode ser aplicado aos mais diversos formatos de conteúdo ajudando não só a promover seu negócio como também a associar a sua marca a valores, experiências e ensinamentos que transcendem a venda, sendo uma forma ou metodologia para você persuadir e convencer.

OBJETIVO - Primeiro, imagine os resultados que você deseja trazer para sua campanha ou ação e comece a usar os dados que estão relacionados a essa linha de raciocínio. Quanto mais concisa for a história e melhor definido o roteiro, maiores são suas chances de impactar positivamente seu público com informações realmente valiosas.

- Para quem será contado?
- O que será contado?
- Qual a finalidade desta história?
- Como você irá contar?

No storytelling é preciso aprender a impactar, gerar empatia e levar a informação da maneira mais sedutora e interessante possível, quanto mais segurança na transmissão de uma mensagem, mais capacidade de conexão emocional com seu público você terá. Assim fica mais fácil fazer com que eles embarquem na sua ideia, trabalhem com você e que seu projeto ganhe a devida atenção.

É preciso descobrir o que toca a sua audiência, as pessoas com quem está dialogando e conduzir sua história mostrando como a sua ideia pode alterar a jornada de um cliente, da sua equipe ou da própria empresa.

Em resumo, essa técnica serve para chamar a atenção do cliente, atraí-lo para a sua marca, criar uma conexão e estabelecer um relacionamento mais concreto e duradouro.

Vejam mais em: <https://medium.com/blog-do-zouza/storytelling-em-an%C3%A1lise-de-dados-f708cca115bb>

## Como mentir com dados - Usando visualização de dados?

Hoje no mundo dos dados possuímos várias formas de etapas para no final interessar uma tomada de decisão seja corporativa ou seja pessoal, uma dessas formas é armazenar os dados, extrair, tratar os dados e criar as chamadas “Visualizações” com base nos dados que se possui,

mas entre os dados e as visualizações pode-se manipular a forma com que esses dados serão mostrados, de certa forma essa manipulação se dá pelo contexto em que a visualização está inserida, digamos que temos esse exemplo abaixo:

**Exemplo 1** (Pessoa qualquer perguntando a um entrevistador de uma empresa):

Pergunta: “Olá qual é a média salarial da empresa?”

Resposta: “Olá, a média salarial é de R\$ 10000,00”

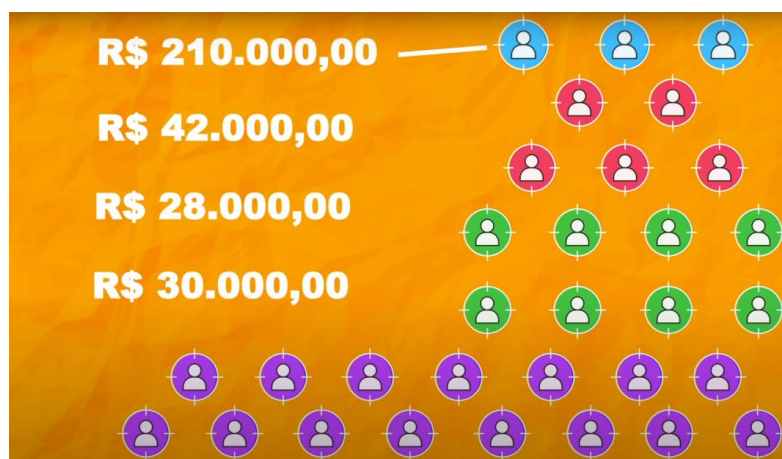
Ao analisar o contexto pode-se tirar várias conclusões e primeiramente vamos aos fatos dos dados de como ele são sem manipulação:

A) Ao todo a empresa possui 31 funcionários.

B) É dividido em: 3 sócios (R\$ 70000,00), 5 gerentes (R\$ 8400,00), 8 supervisores (Cada supervisor ganha R\$ 3500,00) e 15 analistas (Cada analista ganha R\$ 2000,00)

C) Ao fazer um cálculo básico de média pegamos o total de cada Custo dos tipos de funcionários e dividimos por 31 (Total de funcionários da empresa)

Chegamos ao resultado de que a média salarial da empresa é de R\$ 10000,00 (Dez mil reais)



(Imagem mostra o total do custo por cada tipo de funcionário e após a soma do total, dividimos por 31 (total de funcionários, chegamos ao resultado de R\$ 10000,00)

**Exemplo 2** (Pessoa qualquer perguntando a um funcionário de uma empresa):

Pergunta: “Olá qual é a média salarial da empresa que você trabalha?”

Resposta: “Olá, a média salarial é de R\$ 3500,00”

Nesse outro contexto foi feito uma mediana, ou seja, separando 15 grupos de funcionários e 15 grupos de outros funcionários e o valor do meio foi os Supervisores, que ganham R\$ 3500,00 (imagem abaixo ilustrando o cálculo da mediana)



Conclusão: Os dois exemplos estão corretos, porém a forma que estão inseridos em contexto fazem com que temos perspectivas e visões diferentes.

Legenda das imagens:

- Cor roxa = Analistas
- Cor verde = Supervisores
- Cor Rosa = Gerentes
- Cor Azul = Alta gestão

Mais sobre mentir com dados: <https://viz.wtf/>

Conheçam mais sobre a:

