

EDGARD JOSEPH KIRIYAMA

POSTECH

DATA ANALYTICS

ANÁLISE DE NEGÓCIOS

AULA 04

SUMÁRIO

O QUE VEM POR AÍ?	3
HANDS ON	4
SAIBA MAIS	5
O QUE VOCÊ VIU NESTA AULA?	13
REFERÊNCIAS.....	14
PALAVRAS-CHAVE	15

EMANDA

O QUE VEM POR AÍ?

Fala, Data Analytics! Tudo certo?

Visualização de dados, o poder de ver e impressionar! Quando se trata de análise de negócios, não há nada como transformar dados complexos em gráficos e visualizações incríveis. Claro, há alguns segredinhos por trás dessa magia. Primeiro, precisamos seguir os princípios de design, como simplicidade e clareza. Falando em gráficos, sabemos que escolher o tipo certo é fundamental. Gráficos de barras, linhas, dispersão e até mesmo aqueles mapas temáticos são como figurinhas especiais, que traduzem informações de maneira poderosa. Ah, não se esqueça de ferramentas incríveis como Tableau, Power BI e Python, que tornam tudo ainda mais interativo e divertido!

HANDS ON

Agora, vamos ver os dados que exploramos! Mas será que conseguimos tirar mais insights interessantes, que atendem a nossa regra de negócios? Vamos utilizar as técnicas de visualização de dados para identificar e entender a melhor forma de representar os nossos dados.

As boas práticas na visualização de dados são o segredo para um verdadeiro espetáculo, como o uso de cores vibrantes, legendas cheias de personalidade, rótulos bem-humorados e títulos que chamam a atenção. No entanto, é importante lembrar que devemos sempre seguir uma linha de raciocínio.

Na análise de negócios, essas visualizações são como o mapa do tesouro, revelando padrões, tendências e conexões valiosas para melhores tomadas de decisão. Claro, sempre enfrentando os desafios e considerações éticas no caminho, garantindo a precisão, evitando vieses e protegendo a privacidade dos dados.

Vamos lá, experts em data?

Acesse o [GitHub da nossa turma](#) para visualizar os arquivos propostos na disciplina.

SAIBA MAIS

VISUALIZAÇÃO DE DADOS

Conforme aprendemos na primeira fase do nosso curso de Data Analytics, visualização de dados é a arte de transformar dados em gráficos, tabelas e representações visuais intuitivas. É como contar histórias com números e padrões que permitam entender melhor o mundo ao nosso redor. Ao apresentar dados de forma visualmente atrativa, podemos identificar tendências, padrões e insights, que poderiam passar despercebidos em uma tabela cheia de números.

Além disso, a visualização de dados nos permite comunicar essas informações complexas de forma mais eficaz. É como ter um idioma universal, que transcende barreiras linguísticas e técnicas. Por meio de gráficos coloridos, infográficos interativos e dashboards dinâmicos, podemos transmitir ideias de maneira acessível e envolvente para diferentes públicos.

PRINCÍPIOS DE DESIGN DE VISUALIZAÇÃO DE DADOS

Os princípios de design de visualização de dados são diretrizes essenciais para criar representações visuais claras e impactantes. Alguns dos mais importantes são: simplificar a complexidade dos dados, garantir a clareza da mensagem, manter a consistência visual e adaptar-se ao público-alvo. Ao aplicar essas diretrizes, é possível criar visualizações eficazes que facilitam a compreensão e a interpretação dos dados, permitindo uma comunicação visual marcante.

De acordo com Knaflitz (2018), o design de visualização de dados geralmente inclui as seguintes características essenciais:

- **Clareza:** a visualização deve ser clara e fácil de entender, transmitindo a mensagem de forma direta e sem ambiguidade. Isso envolve evitar elementos desnecessários ou confusos que possam prejudicar a compreensão dos dados.
- **Simplicidade:** é importante manter a simplicidade na apresentação dos dados, utilizando elementos visuais mínimos para evitar sobrecarregar o

público com informações excessivas. Gráficos e visualizações mais simples geralmente são mais eficazes na transmissão da mensagem.

- **Consistência:** a consistência visual é fundamental para garantir que a visualização seja coerente e fácil de seguir. Isso envolve o uso de cores, fontes, estilos de gráfico e elementos de design consistentes em toda a visualização.
- **Foco no público-alvo:** ao criar uma visualização de dados, é crucial considerar o público-alvo e adaptar a apresentação às suas necessidades e ao seu nível de conhecimento. A visualização deve ser projetada levando em consideração a familiaridade do público com os conceitos apresentados.
- **Hierarquia visual:** refere-se à organização e ênfase dos elementos visuais. É importante destacar os pontos-chave e insights mais importantes, utilizando técnicas como tamanho, cor ou posicionamento para atrair a atenção do público para os aspectos relevantes da visualização.
- **Acessibilidade:** as visualizações devem ser acessíveis a todos os públicos, incluindo pessoas com deficiências visuais ou cognitivas. Isso significa garantir a legibilidade das fontes, o uso adequado de cores para evitar problemas de contraste e fornecer alternativas textuais para elementos visuais importantes.
- **Integridade dos dados:** é essencial garantir que a visualização reflita com precisão os dados a que se refere. Isso envolve verificar a consistência dos números, evitar distorções ou manipulações enganosas e fornecer contexto adequado para uma interpretação correta.

TIPOS DE VISUALIZAÇÃO DE DADOS

Mudanças ao longo do tempo

A técnica de visualização de mudanças ao longo do tempo é uma poderosa ferramenta na análise de dados, que permite identificar padrões, tendências e

variações no decorrer de um período específico. Essa técnica nos ajuda a compreender a evolução de dados ao longo do tempo e a extrair insights valiosos.

De acordo com Strachnyi (2023), uma das formas mais comuns de visualização de mudanças ao longo do tempo é o gráfico de linha. Nele, podemos plotar dados em um eixo temporal, enquanto o outro eixo representa a medida que estamos acompanhando. Por exemplo, podemos usar um gráfico de linha para visualizar a variação das vendas mensais ao longo de um ano.

Outros tipos de gráficos também podem ser utilizados, como o gráfico de barras empilhadas ou o gráfico de área. Essas visualizações permitem comparar a evolução de várias categorias em relação ao tempo, fornecendo uma visão mais abrangente das mudanças no período representado.

Ao criar visualizações de mudanças ao longo do tempo, é importante considerar a escala dos eixos, a legibilidade das informações e o uso de elementos visuais adequados para destacar pontos-chave. Também é importante fornecer contexto e explicar quaisquer eventos ou fatores externos que possam ter influenciado as mudanças observadas (Strachnyi, 2023).

Ao explorar a técnica de visualização de mudanças ao longo do tempo, podemos obter insights significativos sobre o comportamento dos dados ao longo de períodos distintos, auxiliando na tomada de decisões informadas e no planejamento estratégico de ações futuras.



Figura 1: exemplo de visualização de dados ao longo do tempo.
Fonte: Strachnyi (2023)

Determinando a frequência

A técnica de visualização de determinação da frequência é uma abordagem valiosa para entender a distribuição e a ocorrência de dados. Essa técnica permite identificar padrões e proporções em um conjunto de dados, revelando a frequência com que valores específicos aparecem.

Uma das formas mais comuns de visualização para determinar a frequência é o histograma. Esse gráfico divide o eixo horizontal em intervalos (bins) e representa a quantidade de ocorrências de cada intervalo no eixo vertical. O histograma nos ajuda a entender a dispersão dos dados, bem como identificar picos, assimetrias ou distribuições específicas (Strachnyi, 2023).

Outra técnica útil é o gráfico de barras, que também pode representar a frequência de ocorrência de valores distintos. Cada barra representa uma categoria ou valor único, enquanto a altura da barra indica a frequência de ocorrência. Esse tipo de visualização é eficaz quando temos categorias discretas ou quando desejamos comparar a frequência entre diferentes elementos.

Além disso, diagramas de setores (gráficos de pizza) também podem ser usados para mostrar proporções e frequências relativas dentro de um conjunto de dados. Essa técnica é especialmente útil quando queremos visualizar a distribuição percentual de diferentes categorias.

A Knafllic (2018) faz um alerta quanto ao uso de gráficos de pizza, pois quando os tamanhos dos elementos não são semelhantes, é difícil quantificar exatamente a diferença entre eles. Uma solução eficaz é estabelecer que um é maior que o outro, contudo, não é possível avaliar a magnitude dessa diferença apenas pela visualização. Para superar essa limitação, é possível adicionar legendas explicativas para fornecer informações adicionais e contextualizar a comparação entre os elementos. Dessa forma, mesmo que não seja possível visualizar a diferença em termos de tamanho, as legendas fornecem a descrição necessária para compreender a relação entre os elementos representados.

Ao aplicar a técnica de visualização de determinação da frequência, é importante definir o número adequado de intervalos ou categorias para evitar uma representação distorcida. Também é importante escolher cores, rótulos e títulos apropriados, para facilitar a interpretação dos dados e destacar os pontos relevantes (Strachnyi, 2023).

CLIENTES QUE VISITARAM O LAVA-RÁPIDO NOS ÚLTIMOS 30 DIAS



Figura 2: Exemplo de histograma.
Fonte: Strachnyi (2023), adaptado por FIAP (2023)

Determinando relacionamentos

A técnica de visualização de determinação de relacionamentos é uma abordagem eficiente para identificar e compreender as conexões entre diferentes variáveis ou conjuntos de dados. Essa técnica permite explorar as relações entre os elementos e descobrir padrões ou tendências subjacentes.

Um dos métodos comuns de visualização para determinar relacionamentos é o gráfico de dispersão. Esse tipo de gráfico representa pontos de dados em um plano cartesiano, onde cada ponto corresponde a um par de valores das variáveis em estudo. Ao plotar esses pontos, podemos analisar a dispersão dos dados e identificar se há alguma correlação ou relação entre as variáveis. Por exemplo, se houver uma tendência geral de pontos se aproximando de uma linha diagonal ascendente, isso indica uma relação positiva entre as variáveis (Strachnyi, 2023).

Outra técnica útil é a matriz de correlação, uma matriz tabular que mostra as correlações entre diversas variáveis. Os valores na matriz variam de -1 a 1, onde -1 indica uma correlação negativa perfeita, 1 indica uma correlação positiva perfeita e 0 indica nenhuma correlação. Essa visualização ajuda a identificar quais variáveis estão mais fortemente relacionadas umas com as outras.

Além disso, as redes ou grafos são amplamente utilizados para visualizar relacionamentos complexos entre entidades. Nesse tipo de visualização, os nós representam as entidades e as arestas indicam as conexões entre elas. Essa técnica é útil para analisar redes sociais, relações de dependência ou qualquer outro sistema interconectado.

Ao aplicar a técnica de visualização para determinar relacionamentos, é importante escolher a técnica adequada, de acordo com o tipo de dado e o objetivo da análise. Também é fundamental interpretar corretamente os resultados, considerar a causalidade versus correlação e evitar inferências errôneas (Strachnyi, 2023).

Visualização de dados através de mapas

A técnica de visualização de dados utilizando mapas é uma ferramenta para representar informações geográficas e espaciais. Essa técnica permite que os dados sejam mapeados em um contexto geográfico, fornecendo insights visuais sobre a distribuição geográfica, correlações regionais e padrões espaciais.

Os mapas podem ser usados para representar uma variedade de informações, como densidade populacional, distribuição de vendas por região, áreas de concentração de eventos, entre outros. Eles ajudam a identificar padrões geográficos, tendências regionais e variações significativas em um determinado fenômeno (IBGE, 2022).

Existem várias formas de visualização de dados utilizando mapas. Um exemplo comum é o mapa de coroplético, que utiliza diferentes tons de cores ou padrões para representar valores ou categorias diferentes em áreas geográficas específicas. Isso permite que sejam identificadas diferenças e variações nas regiões representadas.

Outra técnica popular é o uso de marcadores ou pontos para indicar a localização de eventos ou informações específicas em um mapa. Isso pode ser útil para mostrar a distribuição de lojas, a localização de acidentes ou a incidência de certos eventos em áreas geográficas.

Além disso, as visualizações de mapas podem ser interativas, permitindo que os usuários explorem e obtenham informações detalhadas em diferentes níveis de zoom ou filtros específicos. Essa interatividade facilita a análise dos dados em um contexto geográfico e fornece uma experiência envolvente ao usuário.

Ao utilizar a técnica de visualização de dados utilizando mapas, é importante garantir que as informações geográficas sejam precisas, que os rótulos e legendas sejam claros e legíveis, e que os aspectos estéticos e de design do mapa sejam considerados para torná-lo visualmente atraente e informativo (IBGE, 2013).

TIPOS DE GRÁFICOS PARA VISUALIZAÇÃO DE DADOS

Existem diversos tipos de gráficos que podem ser utilizados na visualização de dados, cada um adequado para diferentes propósitos e tipos de informação. Vamos explorar alguns dos principais tipos de gráficos e suas características:

- **Gráfico de Barras:** um dos gráficos mais comuns e eficazes para representar dados categorizados. As barras são utilizadas para mostrar a comparação entre diferentes categorias ou grupos. Podem ser horizontais ou verticais, sendo úteis para visualizar quantidades, frequências ou proporções.
- **Gráfico de Linhas:** ideal para exibir a tendência e a evolução de dados ao longo do tempo. É formado por pontos conectados por linhas, representando a mudança de valores em um eixo temporal. É útil para identificar padrões, tendências ascendentes ou descendentes e variações sazonais.
- **Gráfico de Pizza:** também conhecido como gráfico de setores, é utilizado para mostrar a composição de um todo em relação às partes que o compõem. Cada fatia representa uma categoria, sendo a área da fatia proporcional à sua contribuição percentual. É útil para destacar a participação relativa de diferentes elementos.
- **Gráfico de Dispersão:** utilizado para identificar a relação entre duas variáveis numéricas. Cada ponto no gráfico representa uma observação, sendo posicionado em um plano cartesiano de acordo com os valores das variáveis. É útil para identificar correlações ou padrões de dispersão dos dados.
- **Gráfico de Área:** similar ao gráfico de linhas, porém preenchido abaixo da linha para formar uma área. É útil para mostrar a contribuição relativa de diferentes categorias ao longo do tempo e também para visualizar tendências ascendentes ou descendentes.

- **Gráfico de Histograma:** utilizado para representar a distribuição de dados contínuos ou discretos em intervalos. As barras do histograma mostram a frequência ou a densidade das observações em cada intervalo, sendo úteis para identificar a forma e a dispersão dos dados.
- **Gráfico de Radar:** é um gráfico com eixos que se irradiam a partir de um ponto central, utilizado para comparar várias categorias em múltiplos atributos. É útil para visualizar o desempenho relativo em diferentes áreas, como em avaliações de habilidades.

Segue abaixo um resumo de quando e como utilizar os gráficos, de acordo com o site [portal DATAcated](https://portal.datacated.com).

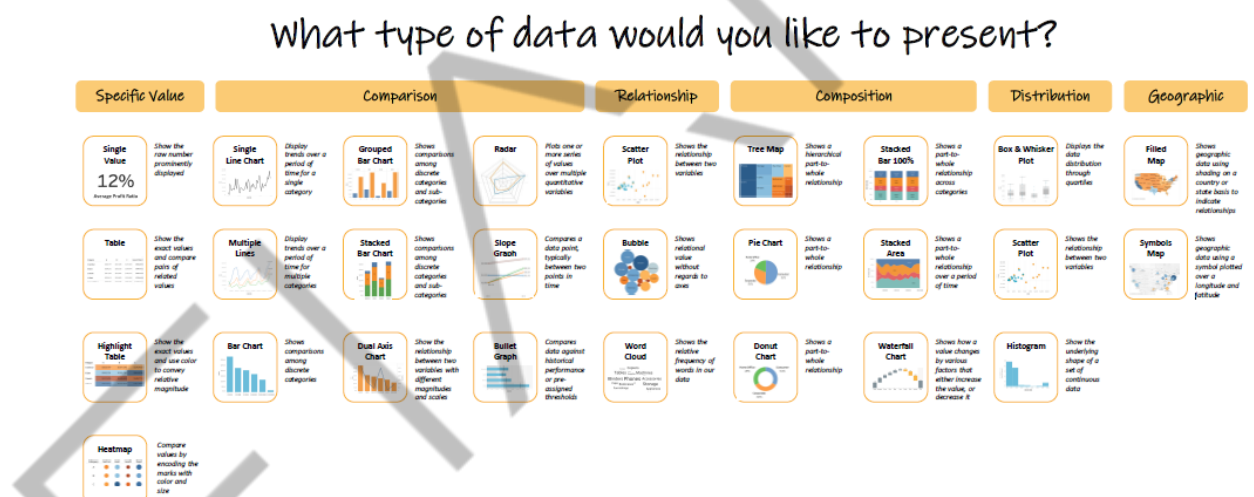


Figura 3: tipos de gráficos para representação dos dados.
Fonte: DATAcated (2023)

O QUE VOCÊ VIU NESTA AULA?

Nessa aula, aprendemos que a visualização de dados é uma ferramenta essencial na análise de negócios, ajudando a compreender e comunicar informações complexas de maneira clara e envolvente. As técnicas mencionadas, como visualização de mudanças ao longo do tempo, determinação da frequência, visualização de relacionamentos e uso de mapas, permitem explorar diferentes aspectos dos dados e obter insights para a tomada de decisões informadas.

No Hands on, trabalhamos bem essas técnicas, lembrando sobre a plotagem de dados para construir algumas visualizações para o nosso entendimento do problema com os preços dos postos de combustível, bem como manipular os dados e gerar valor para eles!

Muito bom, não é verdade? Agora é praticar e entender bem como todo esse conteúdo é aplicado. Bons estudos!

IMPORTANTE: não esqueça de praticar com o desafio da disciplina, para que você possa aprimorar os seus conhecimentos!

Você não está sozinho(a) nesta jornada! Te esperamos no Discord e nas lives com os professores especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

REFERÊNCIAS

DATAcATED. **DATAcated Chart Selector Guide**. 2023. Disponível em: <<https://www.datacated.com/>>. Acesso em: 21 jul. 2023.

IBGE. **Manual Técnico de Uso da Terra**. 3. ed. Rio de Janeiro: Ibge, 2013. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv81615.pdf>>. Acesso em: 21 jul. 2023.

KNAFLIC, C. **Storytelling com dados**: um guia sobre visualização de dados para profissionais de negócios. Rio de Janeiro: Alta Books, 2018.

PROVOST, F; FAWCETT, T. **Data Science para Negócios**: o que você precisa saber sobre mineração de dados e pensamento analítico de dados. Rio de Janeiro: Alta Books, 2016.

STRACHNYI, K. **A Cor dos Dados**: um guia para o uso de cores em storytelling de dados. São Paulo: Novatec, 2023.

PALAVRAS-CHAVE

Análise de Negócios, Análise de Dados, Business Intelligence, Insights, Visualização de dados, Gráficos, Storytelling

EMENDAS

The background is a dark blue field filled with numerous small, light blue dots, resembling a starry sky. Overlaid on this are several large, wavy, translucent lines in shades of blue, yellow, and red, which create a sense of depth and movement. Scattered throughout the composition are various geometric shapes: a circle containing the number '7' in the upper center, a small circle on the left, a cross-like shape on the left, a small circle on the left, and a hexagon in the bottom right corner.

POSTECH