

# **TIPOS DE ANÁLISES E PROJETOS**

**Formadora: Alexandra Souza**

# A análise descritiva

## 1.0 A análise descritiva

A Análise Descritiva é um método de análise estatística que ajuda a descrever as características de um conjunto de dados. Ela fornece resumos simples de uma planilha de dados, através de indicadores, gráficos e tabelas.

Além disso, a análise descritiva é útil na exploração inicial dos dados, permitindo compreender a distribuição dos dados, o seu valor central, a dispersão dos dados e a presença de possíveis outliers.

Essa análise é frequentemente usada como um passo inicial na exploração de dados antes de realizar métodos de análise mais complexos.

## 2.0 Características da análise descritiva

A análise descritiva tem várias características:

☐☐☐ **Resumo de dados:** A análise descritiva resume grandes conjuntos de dados brutos em uma forma que é fácil de entender. Isso é feito através de medidas de tendência central (como a média, mediana e moda) e medidas de dispersão (como o desvio padrão e a variância).

☐☐☐ **Visualização de dados:** A análise descritiva também envolve a criação de gráficos e diagramas para visualizar os dados e as relações entre as variáveis. Isso pode incluir histogramas, gráficos de barras, gráficos de pizza, gráficos de dispersão, entre outros.

☐☐☐ **Identificação de tendências e padrões:** A análise descritiva pode ajudar a identificar tendências e padrões nos dados. Por exemplo, pode revelar se um conjunto de dados está centrado em torno de um valor particular,

se é simétrico, se tem uma dispersão larga ou estreita, se há picos nos dados, entre outros.

## ▼ 3.0 Principais ferramentas

Algumas das principais ferramentas utilizadas na Análise Descritiva incluem:

- ☰☰☰ **Planilhas de cálculo:** Softwares como o Excel e Google Sheets são frequentemente utilizados para a análise descritiva, já que permitem a visualização fácil dos dados e a realização de cálculos básicos.
- ☰☰☰ **Software estatístico:** Ferramentas como SPSS e Stata são usadas para realizar análises descritivas mais complexas, incluindo a geração de estatísticas descritivas e a criação de visualizações de dados.
- ☰☰☰ **Software de visualização de dados:** Aplicativos como Tableau e PowerBI permitem criar visualizações de dados interativas e atraentes.
- ☰☰☰ **Linguagem de Programação:** A linguagem de programação Python e R, são as mais utilizadas para realizar a análise descritiva, devido as suas bibliotecas que contém um conjunto de ferramentas de análise.

## ▼ 4.0 Exemplos de análises descritiva

<https://www.kaggle.com/code/sergiyostrovsky/data-analysis-airline-loyalty-campaign-program>

## ▼ 5.0 Resumo

- A análise descritiva é a base para a compreensão do que já aconteceu ou do que está acontecendo atualmente com um conjunto de dados, sendo essencial para análises estatísticas mais complexas.
- As características da análise descritiva incluem resumo de dados, visualização de dados e identificação de tendências e padrões.
- As principais ferramentas utilizadas na Análise Descritiva são planilhas de cálculo, software estatístico, software de visualização de dados e linguagens de programação.

# A análise diagnóstica

## 1.0 A análise diagnóstica

A análise diagnóstica é um tipo de análise que busca entender os motivos ou causas por trás de determinado fenômeno ou resultado.

Diferente da análise descritiva, que se concentra em descrever o que está acontecendo, a análise diagnóstica procura responder à pergunta "por quê?".

Ela busca identificar padrões e relações nos dados que podem explicar mudanças ou tendências observadas. Por exemplo, uma análise diagnóstica pode ser usada para determinar por que as vendas de uma empresa caíram em um determinado trimestre.

Este tipo de análise pode envolver a consideração de fatores externos, como mudanças no mercado ou na economia, bem como fatores internos, como mudanças nas práticas de negócios ou na estratégia de marketing.

## ▼ 2.0 Características da análise diagnóstica

A análise diagnóstica tem várias características distintas:

**Investigativa:** A análise diagnóstica é investigativa por natureza, buscando entender as causas subjacentes de um fenômeno ou problema. Ela vai além da simples descrição dos dados para entender por que algo está acontecendo.

**Baseada em Hipóteses:** Normalmente, a análise diagnóstica começa com uma hipótese sobre o que pode estar causando um determinado resultado. A análise é então usada para testar essa hipótese.

☐☐☐ **Uso de Dados Históricos:** A análise diagnóstica frequentemente envolve a análise de dados históricos para identificar tendências, padrões e relações que podem explicar os resultados atuais.

☐☐☐ **Análise Multivariada:** A análise diagnóstica muitas vezes envolve a análise de múltiplas variáveis ao mesmo tempo para entender as interações complexas que podem estar influenciando um resultado.

☐☐☐ **Interpretação Cuidadosa:** Dada a natureza complexa da análise diagnóstica, é necessária uma interpretação cuidadosa dos resultados. Errar ao interpretar a causa de um fenômeno pode levar a ações ineficazes ou prejudiciais.

## ▼ 3.0 Principais ferramentas

Algumas das principais ferramentas utilizadas em uma análise diagnóstica incluem:

☐☐☐ **Análise de regressão:** Esta técnica estatística é usada para identificar a relação entre variáveis dependentes e independentes. Ela é frequentemente usada em análise diagnóstica para identificar variáveis que podem estar causando um determinado resultado.

☐☐☐ **Análise de correlação:** A análise de correlação é usada para determinar a relação entre duas ou mais variáveis. Quando duas variáveis estão correlacionadas, uma mudança em uma variável é associada a uma mudança na outra.

☐☐☐ **Análise de cluster:** Esta técnica divide os dados em grupos com características semelhantes. Ela pode ser útil na análise diagnóstica para identificar grupos ou segmentos que estão contribuindo para um determinado resultado.

☐☐☐ **Análise de causa e efeito e Diagrama de Ishikawa):** Esta ferramenta é usada para identificar as possíveis causas de um problema. Ela ajuda a visualizar as principais causas e suas sub-causas, fornecendo uma visão abrangente das possíveis causas de um problema.

## ▼ 4.0 Exemplos de análises diagnóstica



## 5.0 Resumo

- A análise diagnóstica busca entender as causas de um fenômeno ou resultado, indo além da descrição dos dados para responder à pergunta "por quê?".
- Ela é caracterizada por ser investigativa, baseada em hipóteses, usar dados históricos, realizar análise multivariada e requerer uma interpretação cuidadosa dos resultados.
- Algumas das principais ferramentas utilizadas em uma análise diagnóstica incluem análise de regressão, análise de correlação, análise de cluster e análise de causa e efeito (Diagrama de Ishikawa).
- A análise diagnóstica pode ser usada em várias situações, como para determinar por que as vendas de uma empresa caíram em um determinado trimestre.

# A análise preditiva

## 1.0 A análise preditiva

A análise preditiva é um tipo de análise estatística que tem como objetivo prever resultados futuros com base em dados históricos e técnicas de modelagem. Esta forma de análise pode ser extremamente útil em muitos campos, desde a previsão do tempo até a previsão de tendências de vendas ou riscos de saúde.

O processo de análise preditiva geralmente envolve a coleta de dados, o desenvolvimento de um modelo estatístico e a previsão de resultados futuros com base nesse modelo. Esses modelos podem ser simples, envolvendo apenas algumas variáveis, ou extremamente complexos, envolvendo muitas variáveis e relações interdependentes.

A análise preditiva é amplamente utilizada em uma variedade de setores. No setor de negócios, por exemplo, pode ser usada para prever tendências de consumo, permitindo que as empresas otimizem suas estratégias de marketing. No setor de saúde, pode ser usada para prever o risco de doenças, permitindo intervenções preventivas.

É importante notar, no entanto, que a análise preditiva não é infalível. Embora possa fornecer previsões precisas em muitos casos, também é suscetível a erros. Portanto, é essencial que os resultados da análise preditiva sejam interpretados com cautela e usados como uma de várias ferramentas na tomada de decisões.

## ▼ 2.0 Características da análise preditiva

A análise preditiva possui várias características distintas que a definem como uma ferramenta valiosa de análise de dados:

☐☐☐ **Baseada em modelos:** A análise preditiva depende fortemente do uso de modelos estatísticos e de machine learning para prever resultados futuros.

☐☐☐ **Usa dados históricos:** A análise preditiva utiliza dados passados para treinar seus modelos e fazer previsões precisas sobre eventos futuros.

☐☐☐ **Multi-variável:** Muitas análises preditivas consideram uma variedade de variáveis e suas interações para fazer previsões precisas.

☐☐☐ **Preditiva, mas não infalível:** Embora a análise preditiva possa ser extremamente precisa, ela não é 100% precisa e as previsões devem sempre ser consideradas com algum grau de incerteza.

☐☐☐ **Orientada para a ação:** As previsões feitas pela análise preditiva são muitas vezes usadas para informar ações e decisões estratégicas.

## ▼ 3.0 Principais ferramentas

Algumas das principais ferramentas utilizadas na análise preditiva incluem:

☐☐☐ **Python:** Esta linguagem de programação é amplamente utilizada na análise de dados devido à sua facilidade de uso e à disponibilidade de bibliotecas como Pandas, NumPy e Sci-kit Learn para manipulação de dados, análise estatística e modelagem preditiva.

☐☐☐ **R:** R é uma linguagem de programação específica para análise estatística e gráficos. É amplamente usada em análise preditiva, especialmente em campos acadêmicos e de pesquisa.

☐☐☐ **SAS:** SAS (Statistical Analysis System) é um software de análise estatística que é amplamente utilizado em análises preditivas, especialmente em setores como o financeiro e o de saúde.

☐☐☐ **SPSS:** Este é outro software de análise estatística comumente usado em análise preditiva. Ele é conhecido por sua interface de usuário fácil de usar e ampla gama de técnicas estatísticas disponíveis.

☐☐☐ **Excel:** Embora não seja tão poderoso quanto algumas das outras ferramentas nesta lista, o Excel ainda é usado em algumas análises preditivas, especialmente aquelas que não requerem técnicas estatísticas avançadas.

## ▼ 4.0 Exemplos de análises preditiva



<https://www.kaggle.com/code/caesarmario/forecasting-future-passengers-w-pycaret>

## ▼ 5.0 Resumo

- A análise preditiva é uma técnica estatística que prevê futuros resultados com base em dados históricos e técnicas de modelagem.
- É usado em uma variedade de setores, como negócios e saúde, para prever tendências de consumo e riscos de saúde, respectivamente.
- A análise preditiva depende fortemente do uso de modelos estatísticos e de aprendizado de máquina, mas não é infalível, então as previsões devem ser consideradas com algum grau de incerteza.
- As principais ferramentas utilizadas na análise preditiva incluem Python, R, SAS, SPSS e Excel.

# A análise prescritiva

## 1.0 A análise prescritiva

A análise prescritiva é o quarto e último passo no ciclo da análise de dados, seguindo a análise descritiva e a análise preditiva.

Enquanto a análise descritiva ajuda a entender o que aconteceu e a análise preditiva tenta prever o que pode acontecer, a análise prescritiva fornece recomendações sobre o que deve ser feito em resposta.

Isso é feito através do uso de técnicas avançadas que levam em consideração uma variedade de cenários e possíveis resultados, para ajudar os tomadores de decisão a entender as implicações de diferentes cursos de ação.

Em suma, a análise prescritiva ajuda as organizações a tomar decisões mais informadas e orientadas por dados, permitindo que otimizem suas operações e alcancem seus objetivos estratégicos.

## ▼ 2.0 Características da análise prescritiva

A análise prescritiva possui diversas características marcantes. Aqui estão algumas delas:

**Recomendações de ação:** A análise prescritiva fornece conselhos sobre possíveis ações que podem ser tomadas com base nos resultados da análise.

**Consideração de diferentes cenários:** A análise prescritiva leva em conta uma variedade de cenários e resultados possíveis.

**Otimização de operações:** A análise prescritiva pode ajudar a otimizar operações, tornando-as mais eficientes e eficazes.

☰☰ **Auxílio à tomada de decisões:** A análise prescritiva fornece informações valiosas que podem ajudar na tomada de decisões estratégicas.

☰☰ **Adaptação a mudanças nas condições:** A análise prescritiva pode se adaptar a mudanças nas condições e incorporar novos dados à medida que se tornam disponíveis.

## ▼ 3.0 Principais ferramentas

Algumas das principais ferramentas utilizadas na Análise Prescritiva incluem:

☰☰ **Software de modelagem:** Ferramentas como o SAS, o SPSS e o R oferecem capacidades de modelagem que podem ser usadas para realizar análises prescritivas. Estas ferramentas permitem aos analistas construir modelos que podem simular diferentes cenários e resultados.

☰☰ **Ferramentas de otimização:** Softwares como o Gurobi e o CPLEX podem ser usados para otimizar processos e operações com base nos resultados da análise.

☰☰ **Ferramentas de simulação:** Softwares como o Simul8 e o AnyLogic podem ser usados para simular diferentes cenários e entender as implicações de diferentes cursos de ação.

☰☰ **Ferramentas de visualização de dados:** Ferramentas como o Tableau e o Power BI podem ser usadas para visualizar os resultados da análise prescritiva, tornando-os mais compreensíveis para os tomadores de decisão.

☰☰ **Plataformas de Machine Learning:** Plataformas como a Azure Machine Learning e a Amazon SageMaker podem ser usadas para construir e treinar modelos de machine learning que podem ser usados na análise prescritiva.

## ▼ 5.0 Resumo

- A análise prescritiva é o quarto e último passo no ciclo da análise de dados, fornecendo recomendações sobre o que deve ser feito em resposta.
- A análise prescritiva fornece conselhos sobre possíveis ações, considera diferentes cenários, ajuda na otimização de operações, auxilia

na tomada de decisões e se adapta a mudanças nas condições.

- Ferramentas principais utilizadas na análise prescritiva incluem software de modelagem (SAS, SPSS, R).

# Problema de negócio

## 1. O problema de negócio

Vamos imaginar uma empresa de e-commerce chamada **"XYP"**. Ela foi fundada há dez anos por um empreendedor visionário, John, que viu uma oportunidade no mercado para criar uma plataforma online que oferecesse uma ampla variedade de produtos, desde eletrônicos e roupas até produtos de beleza e utensílios domésticos.

Nos primeiros anos, a **"XYP"** enfrentou desafios significativos ao estabelecer sua presença online e construir parcerias com fornecedores em diversas categorias de produtos.

Agora, após uma década de operação bem-sucedida, a **"XYP"** está enfrentando uma decisão estratégica crucial sobre seu crescimento futuro. Com um mercado de e-commerce em constante expansão e a concorrência aumentando, John e sua equipe reconhecem a importância de planejar cuidadosamente os próximos passos da empresa para garantir sua posição de liderança no mercado.

Para tomar essa decisão estratégica, a **"XYP"** precisa entender melhor o histórico de seus números de pedidos. A equipe de análise de dados da **"XYP"** está trabalhando arduamente para extrair e analisar os dados de vendas históricos da empresa.

Os líderes da empresa precisam entender como foram os pedidos dos últimos meses da empresa, para conseguirem tomar uma decisão com mais confiança.

Você foi contratado como um analista de dados para mostrar como estão os números dos pedidos da empresa.

## ▼ 2. O tipo de análise de dados

Você foi contratado como um analista de dados para mostrar **como estão os números dos pedidos da empresa**.

O entendimento melhor de qualquer problema de negócio se inicia pela descrição do desafio. Nesse caso, precisamos desenvolver um resumo mostrando como foram o número de pedidos da empresa historicamente.

Será que foi bom ou ruim? Qual é o comportamento histórico? Existe sazonalidade? Quais produtos venderam mais e quais venderam menos?

Nesse caso, pelo contexto do desafio, uma análise descritiva do número de pedidos é um bom caminho.



# O Projeto

## 1.0 Planeamento

---

O Projeto é definido com um passo a passo para determinar o que será entregue a pessoas de negócio, empresa ou cliente, em termos de produto de dados.

### Passos:

#### ▼ Passo 1- Determinar a saída

A saída de um projeto de dados é o produto que será entregue a pessoa de negócio, a área da empresa ou ao cliente.

O produto varia desde uma análise, alguns indicadores da área, um algoritmo de Machine Learning treinado para realizar previsões **ou até mesmo um relatório com os principais Insights.**

Nesse caso, o analista pode usar o **conceito de protótipo**, no qual ele cria uma versão mínimo do produto final usando dados falsos, apenas para com a pessoa de negócio, se o produto de dados realmente resolve o problema por ele proposto.

#### ▼ Exemplo:

Pergunta de negócio: “Eu preciso de indicadores para entender se a base de cliente está aumentado”

#### ▼ Passo 2- Planejar o processo

O processo de um projeto de dados são as tarefas que precisam ser realizadas para que o produto de dados seja construído e entregue.



O processo assume que o analista terá disponível todos os dados que precisa. A partir dessa consideração é preciso desenhar quais são os passos, começando pela inspeção, limpeza e preparação dos dados até o tipo de gráfico, o posicionamento no painel etc.

▼ **Exemplo:**

Quais são os passos para construir o produto de dados?

▼ **Passo 3- Identificar as entradas**

As entradas de um projeto de dados são as fontes disponíveis de informação, como por exemplo, planilhas do excel, google sheets, banco de dados, websites, API ou qualquer outra fonte de dados.

O mapeamento dessas fontes é extremamente importante para entender quais dados estão disponíveis e se temos permissão de acesso.

A finalização dessa etapa, pode gerar modificações de alguns processos e até mesmo mudanças no produto final, devido a disponibilidade dos dados e permissões das fontes de dados.

▼ **Exemplo:**

Quais são as fontes de dados disponíveis para construir o produto de dados final.

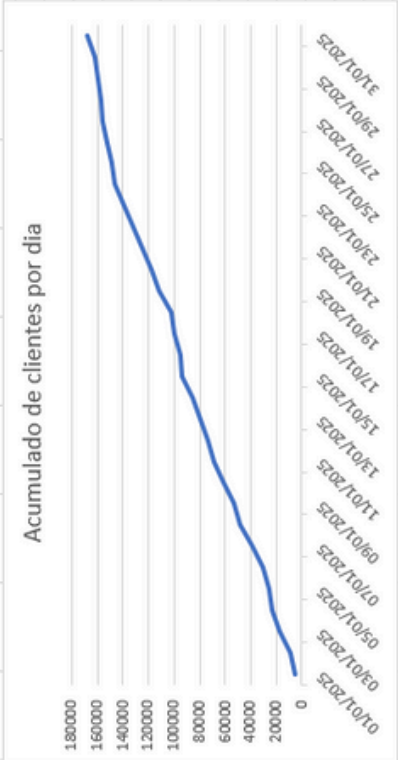
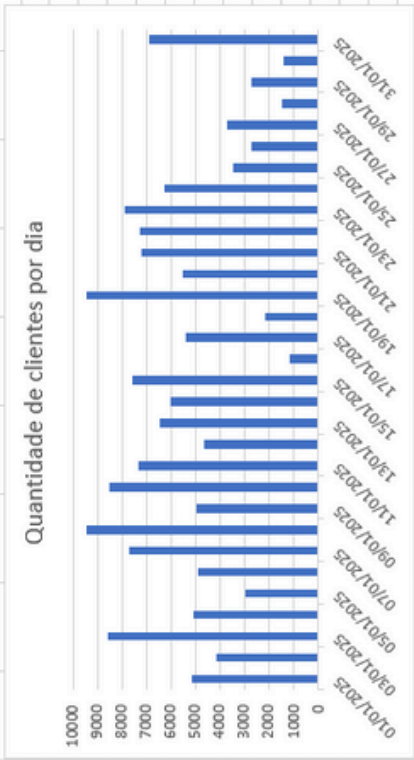
▼ **2.0 Resumo**

---

Essas etapas servirá para determinar o que será entregue a clientes ou empresas em termos de produto de dados.

- **O primeiro passo é determinar a saída (Protótipo)**, ou seja, o produto final que será entregue, que pode variar desde uma análise Descritiva até um algoritmo de Machine Learning.
- **O segundo passo é planejar o processo**, que envolve todas as tarefas necessárias para construir e entregar o produto de dados.
- **O terceiro passo é identificar as entradas**, que são as fontes de dados disponíveis para a construção do produto de dados.
- **O mapeamento dessas fontes** é crucial para entender quais dados estão disponíveis e verificar as permissões de acesso, local de origem, quem é responsável pela entrega de dados, Dicionário de dados e o que mais achar necessário para entendimento e dos dados. Se for um banco de dados relacional, quais tabelas estão envolvidas? eu preciso de todas?

“Eu preciso de indicadores para entender se a base de cliente está aumentando



DATA	Num Cliente	Acumulado
01/01/2025	5141	5141
02/01/2025	4137	9278
03/01/2025	8559	17837
04/01/2025	5070	22907
05/01/2025	2999	25906
06/01/2025	4893	30799
07/01/2025	7704	38503
08/01/2025	9477	47980
09/01/2025	4944	52924
10/01/2025	8519	61443
11/01/2025	7355	68798
12/01/2025	4687	73485
13/01/2025	6474	79959
14/01/2025	6032	85991
15/01/2025	7605	93596
16/01/2025	1148	94744
17/01/2025	5417	100161
18/01/2025	2180	102341
19/01/2025	9472	111813
20/01/2025	5527	117340
21/01/2025	7204	124544
22/01/2025	7264	131808
23/01/2025	7905	139713
24/01/2025	6283	145996
25/01/2025	3476	149472
26/01/2025	2751	152223
27/01/2025	3755	155978
28/01/2025	1510	157488
29/01/2025	2746	160234
30/01/2025	1420	161654
31/01/2025	6902	168556

Passo 2

Tarefas

- Atividades:
- 1 Coletar as colunas Data e Cliente
  - 2 Contagem do Numero de Clientes por dia
  - 3 Desenhar um grafico de barras por dia
  - 4 Desenhar um grafico de linhas acumuladas por dia

Passo 3

Fontes

Será que eu tenho esses dados que eu preciso?

Onde estão esses dados?

Tenho acesso?

Oracle/MYSQL/XLS/CSV

Pedir acesso

Exemplo Macro de Planeamento de projeto

# Planejando a solução

## 1. Planejamento da solução

---

### ▼ 1.1. O problema de negócio

A SuperStore é uma rede de supermercado com várias unidades físicas espalhadas por todo, com o objetivo de fornecer alimentos e comercializar os mais diversos produtos para consumo.

Você foi um dos primeiros contratados para o time de dados da SuperStore e após o onboarding da empresa, já recebeu o seu primeiro desafio:

Analisar os pedidos de compra e trazer insights para o área de negócio. Em A time de gerentes da SuperStore decidiu construir um time de dados para facilitar as decisões da empresa com relação ao abastecimento de produtos das lojas e também das vendas. Hoje, os gerentes tem acessos a algumas planilhas de dados com informações pontuais que ajudam nas decisões, mas eles gostariam de que todos os dados da empresa fossem integrados de modo com que cada área acompanhasse os mesmos indicadores.

O time de dados vai desempenhar um papel importante na criação desse indicadores da empresa, pois terão acesso aos dados transacionais para analisar o comportamento dos clientes, descrevendo os produtos que mais compram, a quantidade comprada, o tamanho da cesta de compra, o valor médio de gasto em cada compra, os produtos devolvidos, o números de pedidos feitos dentro de um determinado tempo e assim por diante.

Essas análises de dados geram insights que abrem espaço para entender como o negócio da empresa funciona do ponto de vista analítico, as razões de aumento ou diminuições do números de pedidos, entre outras pergunta cuja resposta ajudam na tomada de decisão.