

## **Escopo**

**O que é séries temporais?**

**Por que séries temporais?**

**Por que é interessante estudar a parte de finanças em séries temporais?**

**O que é uma série temporal ?**

- Uma série temporal é uma sequência de observações ou dados coletados ao longo do tempo em intervalos regulares ou irregulares. Esses dados são tipicamente ordenados em uma escala temporal, onde cada observação está associada a um ponto específico no tempo. A série temporal é comumente utilizada em várias áreas, como economia, finanças, meteorologia, ciências sociais, engenharia, medicina, entre outras.

## **Características de uma série temporal**

- Intervalo Regular ou Irregular: As séries temporais podem ter intervalos de tempo regulares, como dados coletados a cada hora, dia ou mês, ou intervalos irregulares, onde as observações não ocorrem em intervalos uniformes.
- Tendência: Refere-se a uma direção geral ou padrão de crescimento ou declínio nos dados ao longo do tempo. Pode ser ascendente (crescente), descendente (decrecente) ou horizontal (constante).

**Por que as finanças se enquadram em séries temporais ?**

- Quando analisamos o preço de uma ação , obrigatoriamente fazemos isso em um intervalo de tempo , pois queremos acompanhar o comportamento dessa série temporal.
- Isso pertence à definição de séries temporais , logo ela vai ter características de uma série temporal tal como tendência , intervalos regulares ou irregulares e , portanto , podemos aplicar métodos estatísticos de séries temporais tal como fazer modelagem e pequenas previsões sobre esse tema . O que é muito interessante pois problemas relacionados a mercado financeiro e análise de dados são muito importantes para enriquecer .

## **DataSet**

**Qual o tema do dataset proposto?**

**Quais são seus atributos?**

**Qual o tema do dataset proposto?**

- Basicamente faz uma comparação de ações negociadas no dia entre grandes companhias (Microsoft, Alphabet, Tesla, Amazon, Apple, Meta, Nvidia)
- Divididas em tipos de ações alcançadas no dia (Fechamento ajustado, Fechamento, Alta, Baixa, Abertura)

**Quais são seus atributos?**

- Date (Data - Índice): Data a qual se refere o devido valor
- Adj Close (Fechamento Ajustado): Preço de fechamento de uma ação que foi ajustado para incluir quaisquer distribuições de dividendos e mudanças devido a eventos corporativos.
- Close (Fechamento): Último preço o qual a ação foi negociado no durante um dia específico.
- High (Alta): O preço mais alto que uma ação foi negociada no dia específico.
- Low (Baixa): O preço mais baixo que uma ação foi negociada no dia específico.
- Open (Abertura): O preço de início da ação no dia em específico.
- Volume: Total de ações negociadas no dia.

## **Metodologia**

### **Análise exploratória**

1. **Compreensão dos Dados**: Antes de aplicar técnicas de análise de série temporal, é crucial entender os dados que você está lidando. A análise exploratória permite que você conheça as características das variáveis, suas distribuições, tendências e possíveis padrões.
2. **Identificação de Problemas**: Durante a exploração, você pode detectar problemas nos dados, como valores ausentes, outliers ou inconsistências. Esses problemas podem afetar os resultados da análise de série temporal, e é importante resolvê-los antes de prosseguir.
3. **Seleção de Variáveis**: Se o seu conjunto de dados tiver várias colunas, você precisa decidir quais variáveis são relevantes para sua análise de série temporal. A exploração ajuda a identificar as colunas de interesse e a entender como elas se relacionam com o tempo.
4. **Tendências e Padrões**: A análise exploratória pode revelar tendências temporais e padrões sazonais, que são informações valiosas para modelar a série temporal. Identificar tendências crescentes, decrescentes ou flutuações sazonais pode influenciar as escolhas dos métodos de análise e modelagem.
5. **Estacionariedade**: A estacionariedade é uma propriedade importante em análise de série temporal, pois muitos modelos assumem que as séries são estacionárias. Através da exploração, você pode verificar se a série é estacionária ou se precisa de transformações para torná-la estacionária.
6. **Definição de Parâmetros**: Muitos modelos de série temporal têm parâmetros que precisam ser definidos. A análise exploratória pode fornecer insights sobre possíveis valores de parâmetros, ajudando na configuração adequada dos modelos.
7. **Visualização**: A exploração gráfica, como gráficos de linha, histogramas, e decomposição de série temporal, ajuda a visualizar o comportamento dos dados ao longo do tempo. Essas visualizações auxiliam na identificação de padrões visuais.
8. **Validação de Suposições**: A exploração permite validar suposições que você pode fazer sobre os dados antes de aplicar modelos. Isso ajuda a garantir que os métodos de análise escolhidos sejam apropriados para o conjunto de dados em questão.

## Pré-Processamento

### Análise de dados estatísticos

- **Contagem (count):** O número de valores não nulos em cada coluna. Isso mostra quantos valores estão presentes e ajuda a identificar valores ausentes.
- **Média (mean):** A média aritmética dos valores presentes em cada coluna.
- **Desvio Padrão (std):** O desvio padrão, que mede a dispersão dos valores em relação à média. Valores mais altos indicam maior variabilidade.
- **Valor Mínimo (min):** O menor valor presente em cada coluna.
- **Percentil 25 (25%):** O valor abaixo do qual 25% dos dados estão localizados. Também conhecido como primeiro quartil.
- **Mediana (50%):** O valor do meio quando os dados estão organizados em ordem crescente. Também conhecida como segundo quartil ou valor médio.
- **Percentil 75 (75%):** O valor abaixo do qual 75% dos dados estão localizados. Também conhecido como terceiro quartil.
- **Valor Máximo (max):** O maior valor presente em cada coluna.

## Verificar a presença de Outliers

Para avaliar a presença de outliers, pode-se calcular o intervalo interquartil (IQR), que é a diferença entre o terceiro quartil (Q3) e o primeiro quartil (Q1):  $IQR = Q3 - Q1$ . Qualquer valor abaixo de  $Q1 - (1.5 * IQR)$  ou acima de  $Q3 + (1.5 * IQR)$  é considerado um outlier.

### Calculo do IQR:

$$IQR = 360.5 - 180 = 180.5$$

Agora, calcula-se os limites inferior e superior para identificar possíveis outliers:

- **Limite Inferior:**  $Q1 - 1.5 * IQR = 180 - 1.5 * 180.5 \approx -90.75$  (Como os valores só fazem sentido até o 0 marca-se até esse ponto.)
- **Limite Superior:**  $Q3 + 1.5 * IQR = 360.5 + 1.5 * 180.5 \approx 631.25$

**Verificar valores faltantes**

**Verificar formatos de datas**

## **Modelos de Decomposição**

### **Quando usar o modelo aditivo ou o modelo multiplicativo?**

- **Modelo multiplicativo:** quando as variações em torno da tendência estão relacionadas ao nível da série temporal. Isso significa que as mudanças percentuais (ou proporcionais) na série são constantes em cada período do ciclo
- $Y[t] = \text{Tendência}[t] * \text{Sazonalidade}[t] * \text{Ciclo}[t] * \text{Erro}[t]$
- o resíduo (ou erro) representa a razão entre o valor observado e o valor previsto pela tendência e pelos componentes sazonais e cíclicos
- **Modelo aditivo:** quando as variações em torno da tendência são constantes em termos absolutos, independentemente do nível da série temporal. Isso significa que as mudanças absolutas na série são constantes em cada período do ciclo
- $Y[t] = \text{Tendência}[t] + \text{Sazonalidade}[t] + \text{Ciclo}[t] + \text{Erro}[t]$
- o resíduo (ou erro) representa a diferença entre o valor observado e o valor previsto pela tendência e pelos componentes sazonais e cíclicos

### **Testes de estacionaridade**

- Testes de estacionaridade são procedimentos estatísticos usados para verificar se uma série temporal é estacionária. Uma série temporal é considerada estacionária quando suas propriedades estatísticas não mudam com o tempo. Isso significa que a média, a variância e a covariância (ou correlação) entre observações em diferentes momentos no tempo permanecem constantes ao longo da série. Se uma série não for estacionária, pode ser mais desafiador aplicar métodos estatísticos tradicionais de análise ou previsão, pois as tendências, sazonalidades e outros padrões podem dificultar a interpretação e a previsão precisa, sendo necessário aplicar técnicas para torná-la estacionária. Para verificar a estacionaridade de uma série temporal, geralmente são realizados testes

estatísticos, como o teste Dickey-Fuller aumentado (ADF) ou o teste de Phillips-Perron.

### **Modelos de Predição ML( AUTOARIMA, theta forecaster, prophet)**

- Existem diversos modelos de previsão em Machine Learning que podem ser usados para prever valores futuros em séries temporais. Alguns dos modelos populares incluem o AutoARIMA, o Theta Forecaster e o Prophet.

#### **AutoARIMA:**

O AutoARIMA é uma extensão do modelo ARIMA (Média Móvel Integrada de AutoRegressão) que automatiza a seleção dos hiperparâmetros do modelo ARIMA, tornando-o mais fácil de usar. Ele realiza uma busca automática para identificar os valores ótimos de  $p$ ,  $d$  e  $q$  (ordens do modelo ARIMA) que melhor se ajustam aos dados da série temporal.

#### **Theta Forecaster:**

O Theta Forecaster é um modelo de previsão simples e eficaz que pode ser aplicado a séries temporais univariadas. Ele baseia-se em uma abordagem chamada "Método Theta," que é uma técnica de suavização exponencial. O modelo Theta é capaz de lidar com tendências e sazonalidades nas séries temporais, tornando-o adequado para previsões de curto e médio prazo.

#### **Prophet:**

Prophet é um modelo de previsão desenvolvido pelo Facebook que é especialmente adequado para séries temporais que possuem sazonalidade anual, variações sazonais e feriados irregulares. Ele lida bem com dados ausentes e outliers e permite incorporar informações específicas do domínio, como feriados. Prophet é projetado para ser fácil de usar, com uma sintaxe simples, tornando-o uma escolha popular para previsões de séries temporais de negócios.

### **Avaliação de Modelos**

- A avaliação de modelos é um processo fundamental no desenvolvimento de modelos de aprendizado de máquina e tem como objetivo medir o desempenho de um modelo em tarefas específicas. Durante essa etapa, os modelos são treinados em um conjunto de dados de treinamento e, em

seguida, testados em um conjunto de dados de teste ou validação para avaliar sua capacidade de generalização. Métricas de avaliação, como Mean Squared Error - MSE, Mean Absolute Error - MAE e Mean Absolute Percentage Error - MAPE, são calculadas para determinar quão bem o modelo realiza a tarefa em questão. A avaliação de modelos auxilia na seleção do melhor modelo, na otimização de hiperparâmetros e na verificação de seu desempenho em relação a outros modelos.

### **Análise de modelos**

- A análise de modelos é um processo complementar à avaliação e tem como objetivo aprofundar a compreensão do funcionamento interno dos modelos. Nessa etapa, examinamos como o modelo toma decisões, identificamos quais características são mais relevantes, interpretamos coeficientes em modelos lineares, analisamos resíduos, detectamos overfitting e avaliamos a robustez do modelo em cenários diversos. A análise de modelos visa fornecer insights sobre a interpretabilidade, interpretações errôneas, tendências e características críticas para a tomada de decisões. Ela é fundamental para garantir que os modelos sejam transparentes, confiáveis e atendam aos requisitos específicos do problema em questão.

### **Análise de indicadores financeiros**

#### **Caudas pesadas**

- Caudas pesadas se referem à ocorrência de eventos extremos em dados financeiros, com maior probabilidade do que o esperado em uma distribuição normal. Esses eventos extremos podem ter um impacto significativo nos resultados financeiros e no gerenciamento de risco. É importante reconhecer as caudas pesadas ao analisar indicadores financeiros e implementar estratégias de gerenciamento de risco adequadas.

#### **Efeitos de alavancagem**

- Os efeitos de alavancagem financeira se referem à influência da utilização de dívida ou endividamento por uma empresa na rentabilidade e na estrutura de capital da mesma. O uso de dívida é uma forma de alavancagem financeira que pode amplificar os resultados financeiros de uma empresa, tanto positivamente quanto negativamente.

### **Efeito calendário (prophet)**

- O "efeito calendário" no contexto do Prophet, um modelo de previsão de séries temporais desenvolvido pelo Facebook, refere-se à capacidade do modelo de lidar com sazonalidades e tendências sazonais nos dados. O Prophet é projetado para lidar com dados temporais que podem mostrar padrões previsíveis ao longo do tempo, como variações diárias, semanais ou anuais. Ele automaticamente incorpora esses efeitos de calendário nas previsões, facilitando a modelagem e a previsão de dados temporais, como vendas sazonais, tráfego na web ou outros tipos de séries temporais que exibem comportamento cíclico. O Prophet é frequentemente usado em aplicações de previsão de negócios e análise de dados para ajudar a entender e antecipar tendências sazonais nos dados.

### **Efeito de Badnews**

- O "efeito bad news" é uma reação negativa dos investidores e a queda nos preços dos ativos financeiros em resposta a notícias ou informações negativas sobre uma empresa, setor ou mercado financeiro. Essas notícias ruins podem levar a vendas de ativos e impactar negativamente a confiança dos investidores. Isso destaca a importância da análise de risco e eventos econômicos para investidores.

