

## Etapas do Projeto

### 1. Captura e Pré-processamento dos Dados

- Importação dos Dados: Carregar o arquivo CSV fornecido para explorar e preparar os dados.
- Tratamento de Dados: Limpar e transformar os dados, lidando com valores ausentes e transformando dados de string para formatos numéricos apropriados.
- Exploração Inicial: Analisar estatísticas descritivas básicas e realizar visualizações para entender melhor a distribuição dos dados.

### 2. Análise Exploratória e Feature Engineering

- Análise Temporal: Visualização das séries temporais dos dados históricos de fechamento para identificar padrões, tendências e sazonalidades.
- Criação de Atributos: Criar novos atributos se necessário, como médias móveis, para enriquecer o modelo.

### 3. Escolha do Modelo

- Modelagem: Optar por um modelo de série temporal adequado, como o ARIMA, Prophet ou LSTM, justificando a escolha com base nas características dos dados.
- Justificativa da Técnica: Por exemplo, ARIMA é muitas vezes escolhido para dados estacionários, enquanto o Prophet do Facebook pode lidar bem com sazonalidades diárias ou semanais e feriados.

### 4. Treinamento e Validação do Modelo

- Divisão dos Dados: Separar os dados em conjuntos de treinamento e teste.
- Treinamento do Modelo: Ajustar o modelo utilizando o conjunto de treinamento.
- Validação: Avaliar o modelo com métricas de validação, ajustando hiperparâmetros para melhorar a performance.

### 5. Previsão e Acuracidade

- Forecast: Utilizar o modelo treinado para prever os fechamentos diários para o Q1 de 2025.
- Avaliação de Acuracidade: Comparar os resultados do modelo com um conjunto de dados de teste, visando uma acuracidade superior a 70%.

### 6. Storytelling e Recomendação

- Visualização dos Resultados: Gráficos de previsão vs. real para comunicar os resultados de maneira clara.

- Recomendações: Baseado nas previsões, oferecer recomendações estratégicas ao time de investimentos.

## Tratamento e processamento dos dados

### Ajustes no Pré-processamento

1. Tratar Valores em Milhares e Milhões: Vamos tratar os valores de `Vol` adequadamente para converter strings como "8,90M" e "424.32K" em valores numéricos.
2. Continuar Pré-processamento e Análise: Após corrigirmos esses problemas, podemos continuar com o pré-processamento dos dados.

Vamos refazer o pré-processamento com essas considerações:

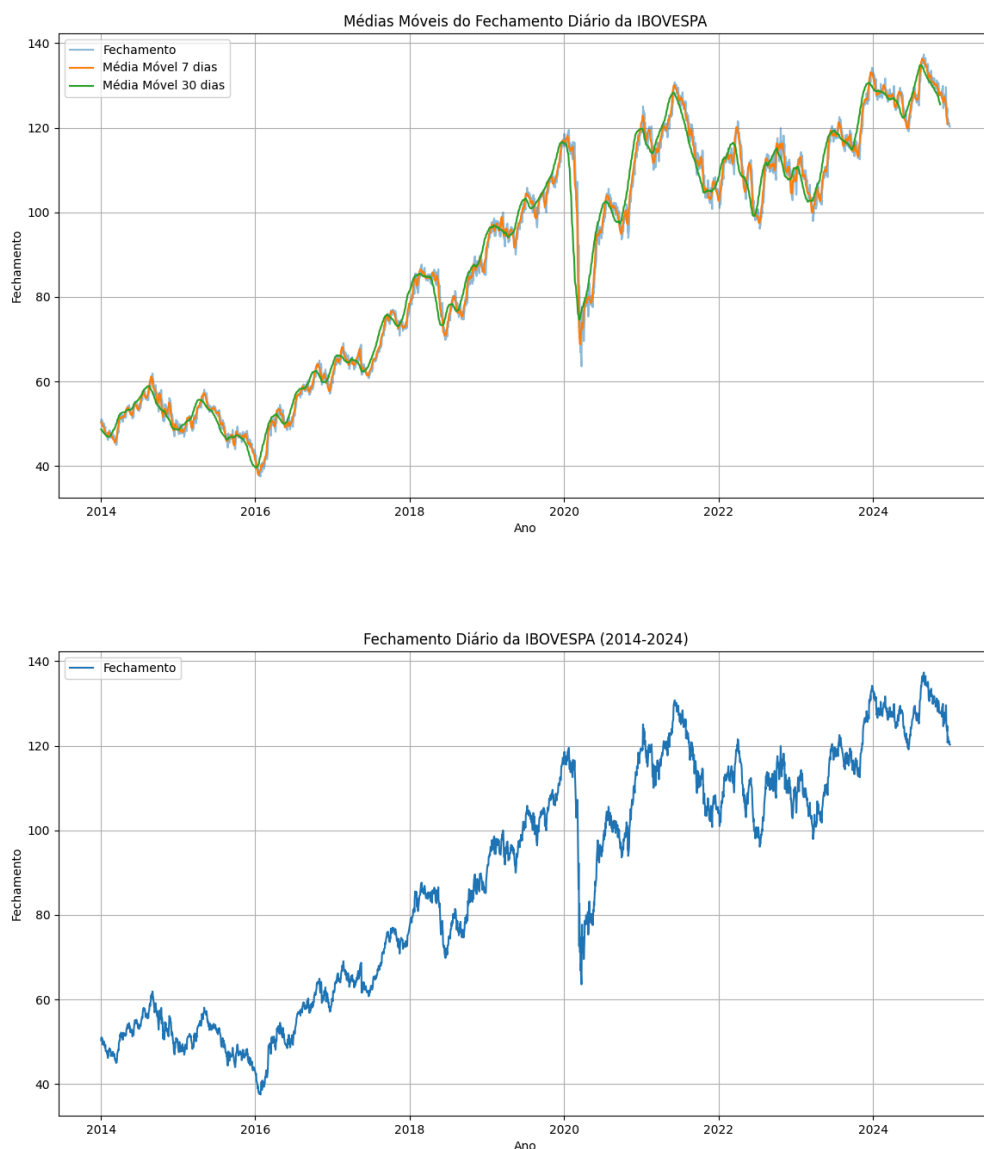
O erro identificado indica que a função de conversão estava sendo aplicada também a valores que já são do tipo float, o que causou uma incompatibilidade. Vou ajustar o código para garantir que a conversão seja aplicada apenas quando o valor for uma string.

Vamos corrigir esse erro e realizar a pré-processamento novamente:

Os dados foram carregados e pré-processados com sucesso. Aqui estão os detalhes importantes do conjunto de dados:

- Total de Registros: 2727 entradas, cobrindo o período de janeiro de 2014 a dezembro de 2024.
- Colunas Convertidas:
  - `Data` foi convertida para o tipo `datetime`.
  - `Vol` agora é um número em notação decimal, lidando adequadamente com valores em milhões e milhares.
  - `Var%` também foi convertida para número decimal.
- Informação ausente: Há um único valor ausente na coluna `Vol`, que poderemos tratar nas etapas posteriores, dependendo de como ele afetará a modelagem.

## Insights da exploração (gráficos)



### Tendência geral:

Há uma tendência geral de alta no índice IBOVESPA na última década, indicando crescimento no mercado. No entanto, essa tendência é acompanhada por flutuações significativas e períodos de aumento e declínio.

### Volatilidade:

O mercado exibe considerável volatilidade, o que é típico para índices de ações. Há picos e quedas visíveis, refletindo reações a eventos econômicos, mudanças de política ou influências do mercado global.

### Flutuações de curto prazo:

A média móvel de 7 dias destaca as flutuações de curto prazo e fornece uma linha mais suave que ainda reflete as variações semanais. Isso pode ajudar a identificar breves movimentos ascendentes ou descendentes.

Tendências de médio prazo:

A média móvel de 30 dias exibe as tendências de médio a longo prazo, filtrando efetivamente parte do ruído associado às flutuações diárias. Ela destaca padrões cíclicos e ajuda a identificar tendências mais amplas no mercado.

Pontos de reversão:

Pontos perceptíveis onde as médias móveis e o preço de fechamento real se cruzam. Esses pontos geralmente indicam reversões potenciais em tendências.

Eventos significativos:

Pode haver certos períodos em que o índice experimenta picos ou quedas incomuns. Isso pode estar relacionado a eventos financeiros globais ou locais, anúncios de políticas ou mudanças no sentimento do investidor.

## Modelagem

O modelo ARIMA foi treinado com sucesso, e agora temos as previsões diárias para o primeiro trimestre de 2025. Aqui estão os detalhes de parte das previsões:

- Previsões para o Início de Janeiro de 2025:
  - 1º de janeiro: Previsão de fechamento em 120.30, com um intervalo de confiança de 117.80 a 122.80.
  - 2 de janeiro: Previsão de fechamento em 120.25, com um intervalo de confiança de 116.86 a 123.64.
  - 3 de janeiro: Previsão de fechamento em 120.27, com um intervalo de confiança de 116.13 a 124.41.

## Avaliação

Podemos usar algumas técnicas para avaliar a performance prévia do modelo em um conjunto de dados conhecido, como o último ano disponível (2024), para obter uma ideia de quão bem o modelo pode prever dados futuros.

### Avaliação Usando Dados de 2024

1. Dividir o Conjunto de Dados:
  - Usaremos uma parte dos dados de 2024 como um conjunto de teste. Isso permite validar o modelo com dados históricos antes de prever o futuro.
2. Métricas de Avaliação:

- Utilizaremos métricas como RMSE (Root Mean Square Error) para avaliar a precisão do modelo.

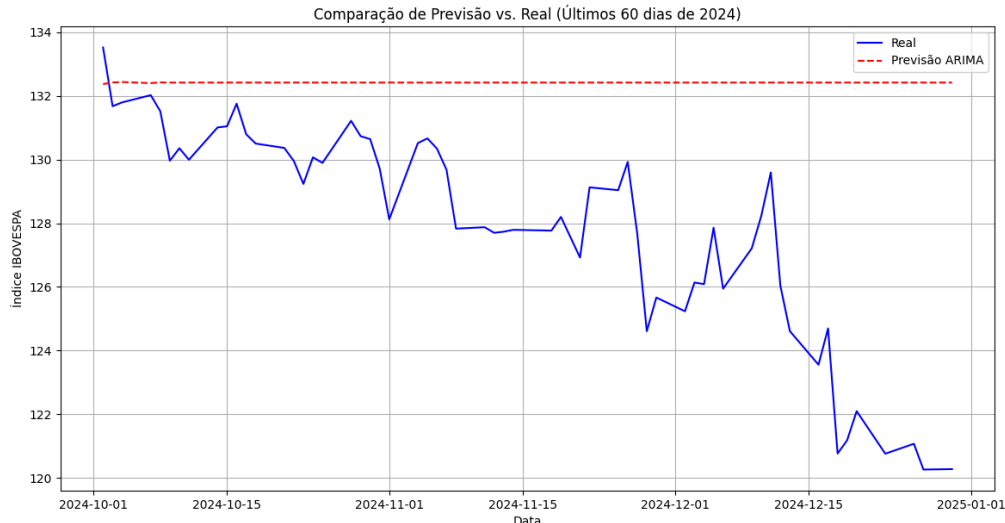
### 3. Comparação das Previsões:

- Comparar as previsões do modelo com os dados reais de 2024 e calcular a acuracidade.

O RMSE (Root Mean Square Error) para o modelo avaliado nos últimos 60 dias de 2024 é aproximadamente 5.57. Este valor nos dá uma ideia da magnitude do erro que podemos esperar ao usar este modelo para previsões.

### Interpretação do RMSE

- RMSE de 5.57: Isso significa que o modelo, em média, erra em cerca de 5.57 unidades do índice IBOVESPA ao prever dados de teste.
- O valor do RMSE deve ser considerado dentro do contexto da escala do índice de fechamento. Se o índice varia em milhares de pontos, um RMSE de aproximadamente 5.57 indica uma previsão bastante precisa, mas devemos compará-lo com os valores reais do fechamento no período de teste para obtenção de uma perspectiva mais informativa.



A acuracidade do modelo ARIMA ao prever os últimos 60 dias de 2024 é de aproximadamente 96.41%. Isso indica que o modelo foi bastante eficaz em capturar as flutuações dos dados reais durante o período de teste.

### O que esta métrica indica:

- Uma acuracidade de 96.41% sugere que o modelo conseguiu prever os fechamentos do IBOVESPA com uma precisão bastante alta, errando apenas em cerca de 3.59% em média.
- Este é um valor muito favorável, especialmente considerando a complexidade dos dados de mercado financeiro.

## Recomendações

### Recomendações

1. Investimento e Gestão de Risco:
  - Considerar as previsões de fechamento diário, junto com o intervalo de confiança, para ajustar estratégias de investimento, especialmente nas decisões de curto prazo.
2. Monitoramento Contínuo:
  - É recomendado realizar uma reavaliação contínua do modelo com dados reais assim que disponíveis para ajustar parâmetros e melhorar a precisão.
  - Incorporar dados mais recentes e fatores externos pode ajudar a reforçar a robustez do modelo.
3. Diversificação Estratégica:
  - Dado a natural volatilidade observada, recomenda-se não depender exclusivamente das previsões mas usá-las como complemento em uma estratégia de investimento diversificada.
4. Capacitação de Time:
  - Oferecer treinamento em análise de dados e interpretação de séries temporais pode melhorar ainda mais as estratégias do time de investimentos.