

# Sazonalidade dos preços no mercado de ações brasileiro

Dagoberto Reis Medeiros  
Instituto Federal da Bahia - IFBA  
Salvador, Brasil  
E-mail: dagobertomedeiros@gmail.com

Grinaldo Lopes de Oliveira  
Instituto Federal da Bahia - IFBA  
Salvador, Brasil  
E-mail: grinaldo@gmail.com

**Resumo**—O presente trabalho tem por meta explorar o mercado financeiro brasileiro por meio de ferramentas e técnicas de mineração e exploração de dados. Com isto analisar se há algum movimento sazonal na flutuação dos preços de ações negociadas na bolsa de valores brasileira, a B3. Importante registrar que esta análise é realizada interpretando os dados dentro de um espectro temporal, ou seja, entendendo que eles ocorrem em função da passagem do tempo, dos dias, meses e anos. Assim, diferentemente de outros dados que podem ser analisados de forma aleatória, no presente estudo estes precisam estar ordenados obedecendo a sua passagem natural e regular do tempo.

**Keywords**—mercado financeiro, análise técnica, análise fundamentalista, mineração de dados, análise exploratória, séries temporais, sazonalidade, estacionariedade.

## I. INTRODUÇÃO

Nos últimos anos viu-se no Brasil um crescimento significativo do número de investidores, especialmente pessoas físicas. A B3 chegou inclusive a publicar um estudo demonstrando esse crescimento [1]. Neste mesmo estudo destaca-se a figura 1 que aqui é reproduzida. Esta dimensiona o volume e a velocidade que este movimento vem tomando.

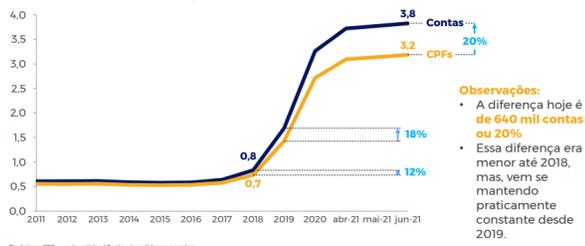
Segundo a própria B3 entre o final de 2020 e início de 2021 houve um incremento de cerca de 500 mil novos investidores no semestre. As pessoas físicas já são responsáveis por 26% de todo o volume negociado na bolsa de valores. Tudo isso confirma que o investimento em renda variável é cada vez mais crescente e popularizado entre os brasileiros.

### Número de Contas e Número de Investidores

O número de investidores PFs na negociação aumentou consideravelmente nos últimos anos, desde 2020 a média de investidores que fazem ao menos um negócio no mês está acima de 1 milhão. Esses investidores também ganharam participação no volume negociado e hoje são responsáveis por mais de 25% do volume na B3

#### Evolução da Posição das Pessoas Físicas na B3

Conceito de Contas x Investidores por CPFs Distintos | em milhares



Disclaimer: CPF = um investidor / Conta = investidor na corretora

Figura 1. Evolução da Posição das Pessoas Físicas na B3

Com isso fez surgir e crescer todo um ecossistema, na bolsa de valores tivemos o surgimento e/ou crescimento de alguns produtos, a exemplo dos mini índices ou mini contratos em

2001, as BDR's (*Brazilian Depositary Receipts*) em 2020 e, fora da bolsa surgiram novas corretoras, destaque-se a XP que popularizou e abriu portas para muitas outras. E, como não pode-se deixar de falar na era do YouTube, surgiu também uma infinidade de canais e cursos prometendo retornos enormes neste mercado imenso. Ou seja, há uma cadeia de produtos e serviços em torno da crescente participação dos brasileiros no mercado financeiros, estes apenas confirmam a importância crescente deste novo nicho de negócios.

É nesse universo que o presente estudo pretende se debruçar. Entre os diversos produtos negociados na bolsa de valores objeto a ser aqui analisado será a ação, ou melhor, na verdade é a variação do preço das ações em função do tempo. A meta é investigar se há alguma influência do tempo sobre o comportamento dos preços, ou seja, se existe sazonalidade nestes.

Para fazer essa análise investigativa e exploratória será necessário fazer uso de algumas tecnologias e conceitos matemáticos e estatísticos para interpretar e automatizar a leitura desses dados. Para isso quase todas as análises são realizadas utilizando como linguagem principal *python* e, desenvolvendo-se prioritariamente no *jupyter notebook*, que trata-se de um aplicativo *open source* que oferece um ambiente no formato de *notebook* (linhas como células para desenvolver os códigos, onde pode-se executar parte do código de forma alternada, sem a necessidade de executar todo o código como numa IDE convencional). E para análise visual, gráfica dos dados foi utilizado principalmente as bibliotecas *Matplotlib* e *Seaborn*, ambas do *python* e, eventualmente também foi utilizado o *Power Bi* para plotar alguns gráficos e realizar algumas análises. Já para análise tabular dos dados a principal ferramenta utilizada foi a estrutura de *Data Frame*, da biblioteca de análise de dados *Pandas*. Trata-se de uma estrutura de tabela com uma série de recursos que dão suporte a tarefas analíticas e eventualmente para análises mais simples e, para armazenar alguns dados foi utilizado o *Excel*, do pacote *office* da *Microsoft* como ferramenta. Além disso, alguns outros recursos são consumidos como as bibliotecas *python Numpy*, de matemática, *Statsmodels*, de estatística, *Time* e *Datetime* para realizar operações com datas e o *Pandas-datareader* para acessar os registros de movimentação na bolsa de valores.

Importante registrar que este trabalho debruça-se sobre o existir ou não sazonalidade na variação de preços dos ativos negociados na bolsa e, tal observação far-se-á com a análise de um volume significativo de dados histórico de cotações destes. Este trabalho não se destina a encontrar fórmulas mirabolantes

para ficar rico com renda variável, muito longe disso. A meta é apenas conhecer melhor o comportamento dos preços, saber se há algo que explique em parte como estes se movimentam.

## II. REFERENCIAL TEÓRICO

Para começar faz-se necessário entender o cenário sobre o qual este trabalho se debruça. Desta forma, dedica-se esta primeira parte do artigo a investigação do mercado brasileiro de ações. Assim, subdividindo-se este em 4 subcapítulos, são eles: Sistema Financeiro Nacional - SFN, o Mercado de Ações, o conceito do que é Ação e por fim um pouco de algumas técnicas utilizadas para entender este mercado, conhecidas por análise técnica e análise fundamentalista.

### A. Sistema Financeiro Nacional

Conceituando, entende-se por Sistema Financeiro:

*... um conjunto de instituições, instrumentos e mercados agrupados de forma harmônica, com a finalidade de canalizar a poupança das unidades superavitárias até o investimento demandado pelas deficitárias.*

*O sistema financeiro pode ser visto como uma rede de mercados e instituições cuja função é transferir os fundos disponíveis dos poupadores, ou seja, aqueles cuja renda é maior do que seus gastos, para investidores, isto é, aqueles cujas oportunidades de gastos são maiores do que sua renda [2].*

A ideia acima exposta parece um tanto generalista e superficial, mas, diz de forma clara o que é um sistema financeiro e quais as relações que são mediadas por ele. No mundo inteiro existem sistemas dessa natureza, nos países desenvolvidos e em países em desenvolvimento. Isso cabe registrar porque atualmente países que querem investimentos em território nacional ou mesmo àqueles que desejam participar do comércio internacional obrigam-se ou sentem-se obrigados a estruturar sistemas como esse. Assim, de país para país este terá algumas nuances próprias, mas, no geral esse o conceito utilizado na maior parte do mundo: um grande balcão para mediação de interesses entre "poupadores" e "gastadores".

E quanto ao sistema financeiro brasileiro é composto por diversas instituições e instrumentos financeiros, conhecido por SFN - Sistema Financeiro Nacional. Tem seu surgimento em 1808, com a chegada da família real portuguesa no Brasil. Em função das relações comerciais estabelecidas com países europeus, exigiu-se então a criação de um intermediário, resultando no Banco do Brasil [3].

Tal história do SFN pode ser dividida em 4 fases segundo Brito, sendo elas:

- 1ª fase: com a vinda da família real para o Brasil, tem-se a abertura dos portos para acordos diretos internacionais, a criação do Banco do Brasil e, a permissão para instalação de bancos estrangeiros em território nacional;
- 2ª fase: entre a Iª e IIª Guerras Mundiais é criada a Inspeção Geral dos Bancos, em 1920, também foi criada a Câmara de Compensação, em 1921 e por fim, houve um fortalecimento das regras de intermediação financeira;

- 3ª fase: após a IIª Guerra Mundial foi criada a Superintendência da Moeda e do Crédito (Sumoc) e também a criação do BNDE, que viria a ser posteriormente o atual BNDES (Banco Nacional de Desenvolvimento Econômico e Social);
- 4ª fase: está vai de 1964 até os dias atuais e começa com a reforma financeira em 1964 e, avança com a criação do Banco Central, com a organização do SFN, Conselho Monetário Nacional, Banco do Brasil e demais instituições financeiras nacionais e estrangeiras. Posteriormente tem-se a regulamentação do mercado de capitais, o estabelecimento de bancos múltiplos em 1988 e em 2002 o início do novo Sistema de Pagamento Brasileiro [4].

Ainda sobre o SFN não pode-se deixar de destacar que sua história evidencia sua constante evolução, que é parte patrocinada pelo avanço da presença do Brasil nos mercados internacionais, mas, também, por constantes investimentos nacionais de seus entes (bancos e outros), inclusive com uso intenso de tecnologia [4]. Destaque-se o Pix, meio de pagamento lançado em 2020, que hoje é um dos meios de pagamento mais utilizado em território nacional ficando atrás dos pagamentos em espécie e se destacando até entre as transações com cartão de crédito e de débito [5].

ÓRGÃOS NORMATIVOS	ENTIDADES SUPERVISORAS	OPERADORES	
CONSELHO MONETÁRIO NACIONAL (CMN)	Banco Central do Brasil (BACEN)	Instituições Financeiras Captadoras de Depósitos à vista	Outros intermediários financeiros e administradores de recursos de terceiros
	Comissão de Valores Mobiliários (CVM)	Bolsas de Mercadorias e Futuros Bolsas de Valores	
CONSELHO NACIONAL DE SEGUROS PRIVADOS (CNSP)	Superintendência de Seguros Privados (SUSEP)	Resseguradores	Entidades Abertas de Previdência Complementar
		Sociedades Seguradoras Sociedades de Capitalização	
CONSELHO NACIONAL DE PREVIDÊNCIA COMPLEMENTAR (CNPC)	Superintendência Nacional de Previdência Complementar (PREVIC)	Entidades fechadas de previdência complementar (fundos de pensão)	

Figura 2. SFN - Sistema Financeiro Nacional

Como pode-se ver na figura 2, atualmente o SFN possui uma estrutura robusta, rica e organizada, exigência mínima para qualquer país atualmente ter condições mínimas de negociar no mercado internacional e, também de fazer negócios internamente, no âmbito nacional. Essa estrutura é composta por:

#### 1) Órgãos Normativos

- a) *Conselho Monetário Nacional - CMN* - responsável por formular a política da moeda e do crédito, objetivando a estabilidade da moeda e o desenvolvimento econômico e social do país;

- b) *Conselho Nacional de Seguros Privados - CNSP* - sua principal função é determinar as normas gerais da política governamental para os segmentos de seguros privados e capitalização;
- c) *Conselho Nacional de Previdência Complementar - CNPC* - define as políticas dos fundos de previdência complementar fechados (fundos de pensão).

## 2) Entidades Supervisoras

- a) *Banco Central do Brasil - BCB ou BACEN* - é um autarquia federal e tem como principal função assegurar a estabilidade do poder de compra da moeda nacional e um sistema financeiro sólido e eficiente;
- b) *Comissão de Valores Mobiliários - CVM* - é uma autarquia vinculada ao Ministério da Fazenda, responsável por regulamentar, desenvolver, controlar e fiscalizar o mercado de valores mobiliários do país;
- c) *Superintendência de Seguros Privados - SUSEP* - órgão responsável pelo controle e fiscalização dos mercados de seguros, previdência privada aberta, capitalização e resseguros;
- d) *Superintendência Nacional de Previdência Complementar - PREVIC* - atua como entidade de fiscalização e supervisão das atividades das entidades fechadas de previdência complementar (fundos de pensão).

## 3) Operadores

- a) *Instituições Financeiras Captadoras de Depósitos à vista* - são os bancos comerciais, bancos múltiplos, cooperativas de créditos e outros, como Banco do Brasil;
- b) *Bolsas de Mercadorias e Futuros e Bolsas de Valores* - é a Bolsa de Valores de São Paulo, a B3;
- c) *Resseguradores, Sociedades Seguradoras, Sociedades de Capitalização e Entidades abertas de Previdência Complementar* - neste grupo inclui-se uma série de entidades que operam seguros, resseguros, títulos de capitalização e outras;
- d) *Entidades fechadas de previdência complementar (fundos de pensão)* - neste grupo estão todas as entidades que operam fundos de pensão, como o Postalís, que é o fundo de pensão dos funcionários dos Correios [6].

Então, o SFN brasileiro caracteriza-se como um importante ente dentro do panorama econômico mundial, responsável por comportar diversas entidades e regular a atuação destes dentro do território nacional e na relação dos mesmos com entidades no exterior, a exemplo do envio e recebimentos de recursos de fora do país.

### B. Conceito de Ação

O objeto desta pesquisa é o preço da ação e, partindo dessa perspectiva carece então esclarecer, definir o que é ação. Assim, uma definição dada por Elder: *Ação é um certificado de propriedade de uma empresa* [7].

Ainda não está óbvio o conceito, então mergulhar um pouco na história deverá esclarecer. As ações em sua forma jurídica, surgem na Inglaterra por volta de 1553 para financiar a Muscovy Company. Assim, para captar 6 mil libras, cada mercador subscreveu uma parte deste montante, adquirindo o direito de sua cota-parte. O sistema foi tão bem sucedido que foi estendido a outros projetos e empresas [2].

Então ação trata exatamente de um pedacinho de uma empresa que pode ser adquirido, uma cota-parte. Desta forma, quando compra-se uma ação, está na verdade assumindo parte da empresa e, com isso o ônus e bônus que a empresa pode oferecer. Pinheiro definiu que:

*As ações são títulos de participação negociáveis, que representam parte do capital social de uma sociedade econômica, que confere ao seu possuidor o direito de participação de sua vida social* [2].

Então ação é exatamente isso, uma parte, uma cota-parte de uma empresa, que ao ser adquirida por quem quer que seja terá sobre a empresa alguns direitos, desde voto nas assembleias (se ações ordinárias) e o mais importante terá direito a dividendos, isto é, terá direito a sua cota-parte nos lucros auferidos pela empresa e, se houver prejuízos não há nada a ganhar, como em qualquer empresa. Todavia, quando se trata de comercializar ação ganha-se também de outra forma, com a valorização do ativo, exemplo: digamos que um comprador qualquer tenha adquirido 100 ações da Companhia Siderúrgica Nacional (na bolsa de valores identificada pelo código CSN3) em 01/05/2019 à R\$ 15,82, ao longo do tempo recebeu dividendos e, agora resolveu se desfazer destes papéis e, em 01/03/2021, vendeu suas ações (ou como é comumente reclamado, fechou sua posição) por R\$ 46,31, desta forma, além de ganhar com os dividendos, este comprador ganhou com a valorização do papel que foi muito significativo, um crescimento na ordem de 292%.

Então quando se fala de lucrar no mercado acionário, fala-se sim de ganhar com dividendos, mas, também se quer dizer que é possível auferir lucros com essa especulação, em que uma empresa pode se valorizar e no futuro pode-se auferir lucros com a venda desses ativos valorizados no decorrer do tempo.

### C. Mercado de Capitais

A CVM (Comissão de Valores Mobiliários) em seu livro "O mercado de valores mobiliários brasileiro" define o mercado de capitais como:

*...o segmento do mercado financeiro em que são criadas as condições para que as empresas captem recursos diretamente dos investidores, através da emissão de instrumentos financeiros, com o objetivo principal de financiar suas atividades ou viabilizar projetos de investimentos* [6].

É neste mercado de capitais, ou melhor, é na bolsa de valores (no Brasil a B3) que as ações são diariamente negociadas, compradas e vendidas, ofertadas. É onde o poupador encontra o tomador de recursos. Em outras palavras, este é o balcão onde os negócios, as transações são realizadas.

E porque buscar recursos para o financiamento de uma empresa na bolsa de valores, ao invés de lograr tais recursos

em instituições financeiras e de fomento com bancos comerciais ou mesmo bancos de desenvolvimentos, a exemplo dos conhecidos: BNDES, Banco do Nordeste e outras instituições. Bem, há algumas vantagens significativas em captar recursos em uma bolsa de valores, destaca-se:

- *Valorização da marca da empresa* - a companhia passa a ter uma imagem institucional e, há um esforço pelo zelo dessa imagem, com isso algumas práticas de gestão passam a fazer parte do dia a dia das empresas;
- *Menor custo na eventual captação de recursos na rede bancária* - por ser empresa de capital aberto e, assim mais transparente, se torna melhor avaliada por bancos, conseguindo concomitantemente melhores taxas;
- *Maior facilidade na obtenção de recursos* - a empresa pode conseguir os fundos necessários entre seus acionistas;
- *Profissionalização da administração* - a empresa precisa atender as exigências legais, com isso os ganhos na melhoria da administração e gestão da empresa são significativos;
- *Liquidez das ações* - pequenos e grandes investidores podem negociar mais facilmente seu patrimônio e lucrar com ganhos e dividendos [8].

Vislumbra-se então o cenário do que é a bolsa de valores e, assim reafirma-se a figura de balcão como uma das melhores formas de ver e entender essa entidade do SFN, um espaço próprio para negociar, comprar e vender, com regras que todos conhecem. E é neste espaço que as ações são transacionadas as milhares diariamente. Durante o expediente da bolsa as ações são precificadas, valorizadas ou desvalorizadas, de acordo com a influência de uma infinidade de fatores, destaco a relação entre compradores e vendedores, por exemplo, se há muitos compradores e poucos vendedores a tendência é que o papel (eventualmente chamaremos ação assim) se valorize, ou o contrário, quando muitos vendedores querem se desfazer e a procura é pouca a tendência será de desvalorização. Bem, no próximo tópico será discutido as formas mais comuns de interpretação desta variação dos preços e mais algumas influências.

#### ***D. Análise Fundamentalista e Análise Técnica***

Para tentar entender o comportamento dos preços das ações nas bolsas de valores desenvolveram-se algumas técnicas, a priori serão destacadas apenas duas que são amplamente conhecidas e amadurecidas. São chamadas: Análise Fundamentalista e Análise Técnica. Registre-se que além dessas técnicas existem outras para análise e interpretação dos movimentos dos preços nas bolsa de valores.

Walter, da Fundação Getúlio Vargas, define Análise Fundamentalista como uma técnica que objetiva encontrar títulos que estão sendo vendidos abaixo de seu preço de equilíbrio, em seguida deve-se manter estes até que seu valor se eleve, para enfim o comprador realizar um ganho de capital o que aumentará seu retorno após a retenção do título [9].

A escola dita fundamentalista, tem por meta avaliar as empresas a partir de seu resultado na sua respectiva atividade

econômica desenvolvida, ou seja, avaliar a produtividade da empresa, sua capacidade de oferecer resultados financeiros e, a partir disso apresentar boas perspectivas para o futuro [10]. Note que a avaliação fundamentalista do valor e/ou da ação em si, consiste na verdade em avaliar o quão eficiente é uma empresa. E para tal avaliação usar-se-á de diversas informações e equações também. O importante a destacar aqui é: numa avaliação fundamentalista o preço da ação deve ser o fim e não a premissa da análise, pois, note que a tônica dessa técnica passa por conhecer a empresa, investigar se é uma empresa com administração boa, com perspectivas claras de crescimento, com prestação de contas transparente e, assim constituir-se em um bom investimento. Cabe ressaltar que a análise fundamentalista não despreza o valor da ação em sua análise, ela faz parte, como será demonstrado a seguir, todavia, vê-se que a marca principal desta técnica é na verdade o conhecimento dos fundamentos da empresa, por isso, análise fundamentalista.

Segundo esta escola existem várias informações a serem analisadas pelo investidor antes de escolher em qual papel aplicar seus recursos, um desses aspectos é o macroeconômico em que a empresa está inserida e, nisto deve-se observar: política, economia nacional e global, cadeia de suprimentos, setor econômico de atuação da empresa e outros [11]. Em linhas gerais trata-se de agregar um arcabouço de informações que lhe diga as chances que determinada empresa tem de crescer e competir dentro de seu nicho produtivo, por exemplo: em 2021, ano em que este trabalho é escrito a cadeia produtiva automotiva está mundialmente prejudicada pela carência de semicondutores no mercado, com isso o mundo inteiro diminuiu sua produção de veículos novos, isso afeta diretamente a capacidade dessas montadoras de produzir riqueza para suas empresas [12]. Fica evidente que investir numa empresa que tem sua capacidade produtiva limitada pela carência de suprimentos pode não ser uma boa ideia.

Ressalte-se que não há receita mágica quando se trata de analisar onde investir. O investidor adepto da escola fundamentalista na sua maioria não esta interessado em comprar ações num dia e desfazer-se no dia seguinte ou até no mesmo dia, ele conta com a valorização da empresa e isso se dá no médio e/ou longo tempo. Ainda sobre o cenário discutido no parágrafo anterior. Fazendo-se uso de diversos dados nota-se que uma montadora hipotética "A" teve forte desvalorização das suas ações, ao passo que esta empresa mantém uma administração eficiente, moderna, inovadora e transparente. Sabe-se que o mundo inteiro esta fazendo uma enorme discussão a cerca do aquecimento global, com investimentos públicos pesados em produção e consumo de energias limpas [13] [14]. E veja essa empresa esta disparada na frente de outras montadoras em suas pesquisas e desenvolvimento de automóveis que fazem uso de fontes energéticas limpas. Desta forma, apesar da forte desvalorização, do tenebroso cenário atual, no horizonte esta empresa apresenta para o investidor um conjunto de qualidades que a permitirá se destacar num futuro muito próximo na concorrência do advento desse novo mercado automobilístico que se avizinha. Note não existe certezas, mas, um conjunto de informações que lhe sinalizam que o futuro pode ser fortuito ou infeliz.

Este trabalho não pretende fazer uma discussão aprofundada do que é e como se faz análise fundamentalista. Todavia,

nos primeiros parágrafos foi introduzido um conceito que carece melhor explicação, trata-se do "ponto de equilíbrio". Esse termo refere-se a um indicador muito utilizado na escola fundamentalista (e não só por esta). Este indicador é comumente conhecido por P/VP - Preço sobre Valor Patrimonial, é obtido dividindo o valor de uma ação negociada na bolsa pelo valor patrimonial da empresa [15]. E é esta razão que sinaliza o quão barata ou cara é uma ação, ou seja, o ponto de equilíbrio citado por Walter que indica o momento ou o preço que determinada ação se alinha ao valor patrimonial da empresa, em linhas gerais seria: se uma ação esta barata é porque seu valor é inferior ao valor patrimonial da respectiva empresa [9]. E assim, a partir dessa relação pode-se estimar se o valor de um papel está caro ou barato ajudando o investidor a decidir se vale a pena o investimento. Em linhas gerais, seria comprar quando esta barata e, vender ou realizar como é dito, quanto chegou ao valor ideal, ao dito ponto de equilíbrio, ou no preço justo, expressão também muito utilizada.

Análise Técnica: esta é outra forma de ler e interpretar a flutuação dos preços das ações. E verdade seja dita é uma das formas mais utilizada entre aqueles que esperam fazer operações rápidas, curtas, de poucos dias, até de segundos entre a compra e venda do ativo. Para clarear:

"... a Escola Técnica concentra seu foco exclusivamente no mercado, desconsiderando a situação financeira das empresas, seus balanços e indicadores de performance, ou ainda, perspectivas do setor de atividade da empresa... acredita na repetitividade do comportamento humano e no poder da ciência estatística como forma de determinar, com base no comportamento passado, as perspectivas para o mercado futuro [16]."



Figura 3. Gráfico com indicadores utilizados para Análise Técnica

O gráfico através de uma infinidade de indicadores constitui-se como uma das principais ferramentas desta escola (figura 3). Neste é possível ver a variação dos preços ao longo do tempo, inclusive fazendo o uso de várias escalas temporais, pode-se ver um gráfico diário, semanal, mensal e, até de minutos. As corretoras através de seus *home brokers*<sup>1</sup> entregam gráficos com intervalos de minutos, com atualização em tempo real, permitindo ao investidor acompanhar a negociação minuto a minuto.

A figura 3 apresenta alguns indicadores utilizados em análises técnicas, como: volume financeiro negociado, média móvel dos últimos períodos (no gráfico apresentado está sendo utilizada média móvel exponencial de 9 e 21 períodos, também

<sup>1</sup>Aplicação pela qual é possível acompanhar, comprar e vender ações e contratos na bolsa de valores.

identificada por MME9 e MME21), além do RSI *Relative Strength Index* e, outros.

Pode-se ainda a partir do reconhecimento de padrões gráficos como os exibidos na figura 4 decidir se é hora de abrir ou fechar uma posição. Enfim, a escola da análise técnica basicamente diz que o investidor deve aprender a reconhecer alguns padrões e, o reconhecimento de um conjunto de padrões ou mesmo apenas um isoladamente deve dizer o que deve ser feito, "se acontecer isto faça aquilo, se não, faça isso" [17]. Por exemplo, a figura 4 apresenta dois padrões gráficos, o primeiro mostra uma curva descendente dos preços e, no fundo o aparecimento de um *candle*<sup>2</sup> com o corpo conforme mostrado, sinalizando possível reversão altista, ou seja, a partir de então o mercado deverá sofrer uma reversão e os preços começarão a subir e, na segunda figura indica o contrário, o surgimento da mesma no gráfico sinaliza possível reversão baixista. A leitura destes padrões e outros permite ao adepto da escola técnica decidir quando entrar numa operação, ou mesmo como entrar numa operação, com uma posição de comprado ou vendido.

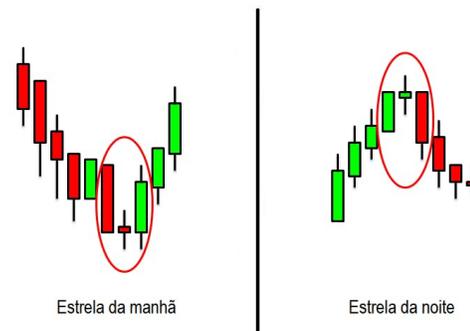


Figura 4. Padrões de candles

É importante registrar que essa é uma técnica que utiliza muitos recursos computacionais, seja para identificar padrões e produzir indicadores, como também faz uso de inteligência artificial para automatizar robôs que reconhecem padrões de movimento na bolsa. Há também outras investidas computacionais de outras naturezas, como desenvolver novos indicadores, alarmes para ajudar a identificar momentos de abrir ou fechar uma posição. Uma serie de possibilidades. Inclusive permitindo por exemplo, uma mescla entre análise técnica e fundamentalista, aproveitando-se o melhor de cada escola para certificar-se das melhores oportunidades de negócios.

### III. KDD E DATA MINING

#### A. KDD - Knowledge Discovery in Databases

Antes de falar propriamente de Mineração de Dados, ou *Data Mining*, expressão em inglês mais comumente conhecida.

<sup>2</sup>Gráfico de candle ou de vela é muito comum na leitura dos mercados financeiros, basicamente a figura é composta de um pavio que registra o valor mínimo e máximo no intervalo de tempo compreendido pela figura, enquanto a cor que aparece comumente como vermelho indica que o valor de fechamento do respectivo período foi menor que a abertura, assim o valor de fechamento estará na parte baixa da vela e na parte alta o valor de abertura e, para a vela verde ocorre o contrário o valor de abertura é menor que o valor de fechamento

Carece de apresentar o processo que coordena os esforços deste trabalho, abreviadamente chamado de KDD - *Knowledge Discovery in Databases*. Método exposto por Fayyad, Piatetsky-Shapiro e Smyth em 1996, em seu artigo: *From Data Mining to Knowledge Discovery in Databases* [18]. Os autores perceberam à época que o volume gerado de dados crescia de forma exponencial e as técnicas usuais até então não eram capazes de interpretar esse montante de informação. Diante disso, é proposto que algumas atividades neste processo precisam ser automatizadas. Assim, pode-se dizer que o KDD constitui-se como um conjunto de tarefas que tem por meta entender e analisar um enorme conjunto de dados e transformá-lo em formas mais simples de apreensão do conhecimento, como um relatório, artigo, ou outras formas e, também, sendo possível utilizar este conhecimento através de técnicas estatísticas para fazer análises preditivas e outras tantas possibilidades.

Desta forma, os mesmos autores estabeleceram um processo para o KDD, conforme descrito na figura 5. Que consiste em 5 etapas até o propriamente dito "conhecimento". São elas:

- **Seleção** - etapa em que separa-se, ou melhor, extrai-se um conjunto seletivo de dados sobre o qual pretende-se conhecer;
- **Processamento** - momento no qual é realizada a limpeza dos dados, excluindo os atributos desnecessários;
- **Transformação** - nesta parte do processo os dados são transformados, formatados para que possam ser analisados na próxima etapa;
- **Mineração de dados ou Data Mining** - é a etapa onde os dados são lidos, onde regras e métricas são aplicadas para a descoberta de padrões, enfim, é onde a informação é produzida;
- **Interpretação** - por fim, nesta etapa a informação produzida é analisada e o conhecimento é então produzido, ou melhor, onde o conhecimento é apreendido.

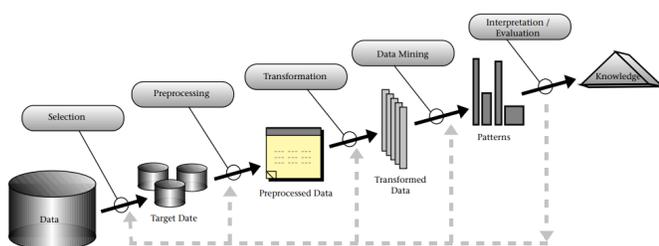


Figura 5. Visão geral das etapas que compõem o processo KDD

Como pode ser observado, KDD é um processo rico e cuidadoso, onde os dados são lapidados afim de produzir saber. Tal técnica eventualmete é tratada como sinônimo de mineração de dados (ver Han, Kamber e Pei) [19]. Todavia, dentro deste trabalho utilizaremos o conceito de mineração de dados como uma etapa distinta dentro de todo o processo que constitui o KDD.

Tentou-se apresentar de forma breve e concisa o que vem a ser o processo do KDD que norteia a produção de todo este trabalho entendendo o KDD como uma dinâmica extenuante e imprescindível para transformação de dados em conhecimento.

Apesar do diagrama proposto na figura 5 listar uma escada de etapas claramente definidas, a realização de cada etapa constitui um emaranhado de atividades que cooperam para a produção do saber neste sentido cabe ainda ressaltar que cada etapa pode ser refeita inúmeras vezes, para garantir o sucesso no final de todas as tarefas.

## B. Data Mining

Uma parte importante na produção deste trabalho refere-se a Mineração de Dados. Segundo Olson, mineração de dados diz respeito a análise de enorme quantidade de dados armazenados em computadores [20]. Já Goldschmidt entende esta como a busca efetiva por conhecimentos úteis no contexto da aplicação do KDD, para ele esta é a parte principal do KDD [21]. E Han vai dizer que diante da imensa quantidade de dados coletados diariamente, analisá-los é uma necessidade importante [19]. Assim, uma meta deste trabalho é analisar, entender, conhecer um pouco dos dados da bolsa de valores brasileira, compreender o comportamento de seus preços e eventualmente, se houver, identificar possíveis padrões sazonais.

Em mineração de dados há uma clara distinção entre usar os dados para encontrar regras e padrões e, fazer o uso destas métricas conhecidas para testar probabilidades [22], como pode ser observado na figura 6.

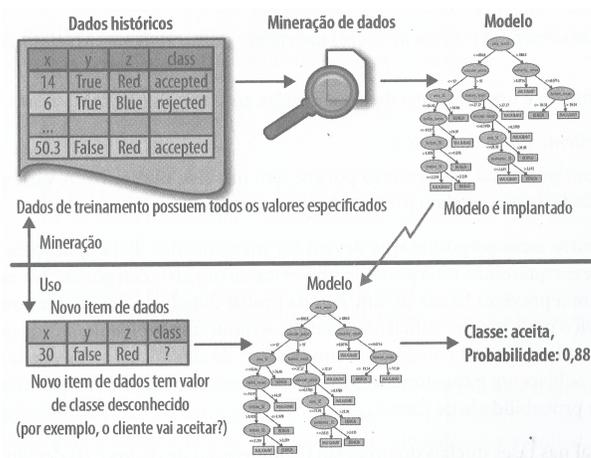


Figura 6. Mineração de dados versus o uso dos resultados da mineração de dados

A compreensão destas duas atividades distintas na mineração de dados é fundamental para entender que minerar constitui uma atividade para conhecer ou reconhecer a existência de possíveis padrões dentro de determinado conjunto de dados, isto é, apreender o conhecimento sobre a dinâmica de funcionamento de determinado domínio, para enfim decidir se precisa ou não realizar análises preditivas. É imprescindível avaliar o modelo construído a partir dos dados históricos, pois, este tem o dever de ser o mais fidedigno possível da interpretação da realidade dos dados. O uso posterior do modelo gerado permitirá fazer estimativas e outras possibilidades.

Desta forma, as tarefas podem ser classificadas em duas categorias: descritivas e preditivas. Sendo a primeira responsável por caracterizar as propriedades de um conjunto de dados,

enquanto a segunda realiza indução nos dados atuais para fazer previsões [19].

Isto posto, para a execução da mineração de dados há um processo próprio, ou melhor, um ciclo chamado CRISP-DM - *Cross-Industry Standard Process of Data Mining*, que quer dizer Processo Padrão de Indústria Cruzada para Exploração de Dados. E trata-se da organização das atividades, conforme demonstrado na figura 7. Onde há um processo exaustivo de idas e vindas, uma etapa pode evoluir e, posteriormente regredir sempre que necessário, afim de que possa-se identificar padrões e outras métricas [23]. Assim, as repetições no processo não são exceções, mas, sim uma regra [22].

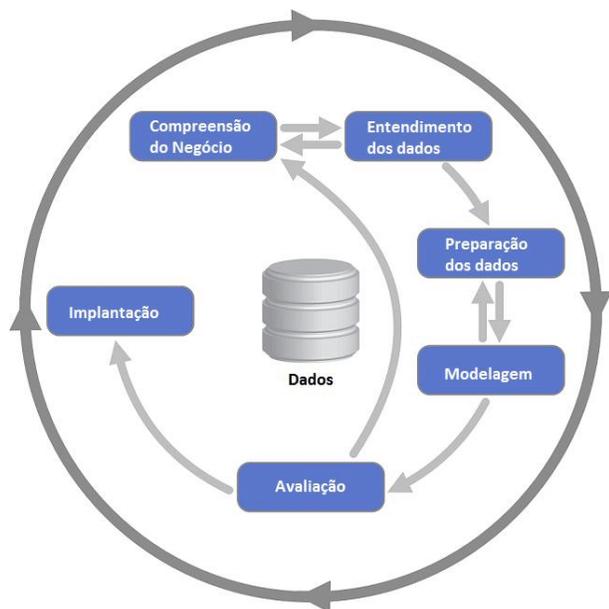


Figura 7. Fases do CRISP-DM

O ciclo do CRISP-DM pode ser assim entendido:

- **Compreensão do negócio** - constitui-se na fase de entendimento do domínio, da situação, do estabelecimento de metas, de desenvolvimento do plano de projeto. Nesta etapa deve-se pensar cuidadosamente sobre o cenário de uso e o problema a ser resolvido [22].
- **Entendimento dos dados** - nesta fase busca-se conhecer, compreender os dados, suas conexões com o negócio, identificar os dados relevantes. Neste momento técnicas de agrupamento e exploração visual também são utilizadas [20].
- **Preparação dos dados** - uma vez identificada a viabilidade dos dados, este é o momento em que os dados são preparados, ou seja, há um trabalho intenso para preparar os dados para sua utilização final, selecionando variáveis, analisando, transformando alguns valores conforme a necessidade e limpando linhas dos dados [23].
- **Modelagem** - é o principal local onde as técnicas de mineração de dados são aplicadas aos dados. É onde a maior parte das técnicas da ciência e tecnologia serão

exercidas [22]. É a fase em que os algoritmos são calibrados e otimizados [23].

- **Avaliação** - o objetivo dessa fase é estimar os resultados da mineração de dados, se os resultados são válidos, avaliar a confiabilidade do modelo. Importante destacar que alguns padrões encontrados podem não sobreviver a um exame minucioso [22].
- **Implantação** - é o momento de fazer conhecer a todos envolvidos os resultados obtidos [24]. É a fase em que os resultados da mineração de dados são colocados em uso no mundo real, a fim de se constatar algum retorno sobre o investimento [22].

### C. Os Dados

Dentro das ciências de dados e suas inúmeras ramificações e tarefas (análise de dados, visualização de dados, aprendizado de máquina, classificação, clusterização e outras) a matéria prima bruta que é a base de tudo isso são os dados. Que podem ser de diversas naturezas e produzidos de múltiplas formas, desde dados de cliques de usuários num site de *e-commerce*, coletas diárias de condições climáticas como velocidade dos ventos, temperatura, ou mesmo dados mecânicos, como quantidade de impressões, estado de funcionamento de uma fornalha industrial, enfim, uma infinidade de produtores e modos de produzir dados existem. E no trabalho em voga os dados são fornecidos pela B3, que é a bolsa de valores brasileira e estes são produzidos a partir da interação entre milhares de vendedores e compradores que, ao realizar essas transações por meio da B3, deixam um registro que é a fonte desta análise. E por isso cabe neste contexto um destaque sobre como coletar, tratar, transformar e usar dados.

Assim, entende-se que os dados são divididos em 2 grupos: de variáveis categóricas ou qualitativas e numéricas ou quantitativas. O primeiro grupo sendo subdividido em dicotômicas, nominais e ordinais. E o segundo grupo subdivide-se em contínuas e discretas [25].

Os dados *qualitativos* e seus subtipos:

- **Dicotômicas ou binárias** - são aquelas que compreendem apenas duas categorias, como sexo, que pode ser apenas masculino ou feminino, ou uma resposta de um questionário com "sim" ou "não";
- **Nominais** - são aquelas que apresentam mais de duas categorias, como país de origem, que pode ser Brasil, Argentina, Paraguai, etc. Ou seja, quando há múltiplas alternativas;
- **Ordinais** - refere-se a dados que possuem uma ordem explícita, como classes sociais A, B, C, D, E [25].

Os dados *quantitativos* e seus subtipos:

- **Discretos** - quando os dados observados podem ser exclusivamente números inteiros, como por exemplo número de filhos, quantidade de idas ao médico;
- **Contínuos** - são dados medidos em escala contínua, ou seja, que podem ter tantas casas decimais quanto for preciso, como por exemplo a temperatura, ou a pressão sanguínea, etc [25].

Tal conhecimento é importante para saber como trabalhar sobre os dados, que técnicas podem ser utilizadas, quais os possíveis resultados obtidos. E a partir disso uma nova tarefa se impõe, que é a limpeza e transformação dos dados. Para Grus, os dados do mundo real são sujos e, para fazer uso deles, antes deve-se trabalhar neles, limpá-los e transformá-los [26].

Pode-se compreender essa atividade de limpeza e transformação num processo maior, muito conhecido sob a sigla *ETL - Extract Transform Load*, que é um processo para extrair dados de sistema, em seguida processá-los e transformá-los [27].

Esta tarefa é importante para que sejam tratados os dados ausentes, ou mesmo com alguma inconsistência ou ruído, que podem levar a interpretação imprecisa dos dados [19]. Por exemplo, no caso deste trabalho estamos utilizando dados das operações realizadas na bolsa de valores, se uma empresa tem o valor de sua ação sendo negociada ao longo da história com valores entre R\$ 12,50 e R\$ 18,75 e, quando observase determinada linha vê-se o valor de R\$ 600,00. Neste caso, o analista precisa fazer uma escolha que pode ser substituir esse valor pela média do período, ou mesmo remover esse registro, mantê-lo só contribuirá para uma leitura equivocada da realidade e poderá induzir o modelo produzido ao erro. O mesmo ocorre com dados ausentes, se determinada empresa tem seus dados negociados diariamente há mais de 10 anos e por ventura durante a extração algum período ou intervalo de registros não foi coletado, novamente o analista precisa fazer uma escolha, fazer nova coleta de dados para recuperar esses dados ausente, ou tentar descobrir se há um motivo para isso e escolher alguma técnica para lidar com os dados ausentes.

Além desse processo de limpeza, inúmeras vezes os dados precisam ser transformados, por exemplo, neste trabalho estão sendo usando dados temporais e, para fazer uso de algumas técnicas de análise de séries temporais é imprescindível que as séries sejam estacionárias (no próximo capítulo este conceito será melhor apresentado), para isso uma transformação dos dados é necessária. Vassiliadis vai dizer que a ação de filtrar ou limpar dados constitui uma operação de transformação [28]. Após esta etapa, os dados são carregados, ou dá-se a carga dos dados como pode ser visto na figura 8 em que os dados são remetidos para o DW e, posteriormente utilizados para análises diversas.

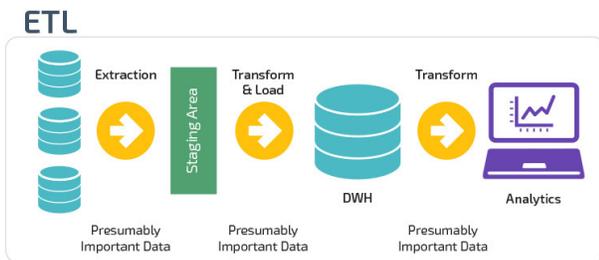


Figura 8. ETL - Extract Transform Load

#### IV. SÉRIES TEMPORAIS

Os dados a serem utilizados neste trabalho tem características "especiais", referem-se a séries temporais, que são:

... uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo

de dados é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência. Enquanto em modelos de regressão por exemplo a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é crucial [29].

É importante conhecer as propriedades do dado a ser analisado e, neste trabalho, os dados têm uma relação direta em função do tempo, ou melhor, em relação a passagem do tempo. Por isso, aprofundar-se-á um pouco os conceitos relacionados a este tipo de informação, que diferentemente de outros tipos de dados possui características próprias e, por isso, técnicas também próprias para tratar e transformar para enfim, para analisá-lo.

Desta forma, serão apresentados os seguintes conceitos: noções gerais, tendência, sazonalidade e autocorrelação.

##### A. Noções Gerais

Existem muitos objetivos para analisar séries temporais, por exemplo fazer previsões de valores futuros, analisar o comportamento da série temporal em gráficos para encontrar periodicidades relevantes [30]. Para esses objetivos e tantos outros existem algumas abordagens que podem ser utilizadas como: técnicas descritivas, identificação de padrões, modelos probabilísticos, análise espectral, métodos não paramétricos (alisamento ou suavização), modelos não lineares, séries multivariadas, volatilidade e outros [29].

Nesse contexto um novo conceito precisa ser inserido, trata-se da estacionariedade em séries temporais. Assim, uma série estacionária é aquela que se desenvolve aleatoriamente no tempo, em torno de uma média constante, refletindo algum equilíbrio estável [30]. Os dois gráficos na figura 9 distinguem visualmente uma série da outra.

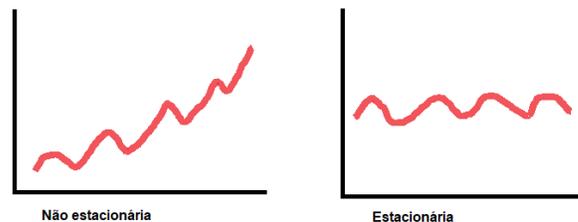


Figura 9. Série estacionária e não estacionária

É importante introduzir este conceito porque no mundo real, na prática a maior parte das séries temporais não se apresentam de forma estacionária [30]. E para analisar esses dados existem técnicas que permitem transformar a série temporal em estacionária como a *diferenciação* que realiza a diferença entre dois pontos sucessivos, de maneira que tenha média e variância constantes e, assim a série poderá ser tratada como estacionária [31].

Antes de aplicar qualquer técnica para tornar a série estacionária a primeira tarefa consiste em testar a mesma. Para isso existem algumas técnicas como teste de Dickey-Fuller, teste do Phillips-Perron e teste KPSS. Neste trabalho usou-se o algoritmo de Dickey-Fuller, que foi desenvolvido motivado pela necessidade de verificar se uma série precisa ser diferenciada para se tornar estacionária. E esta constitui

a primeira etapa para a construção de modelos Autorregressivos Integrados Médias Móveis (ARIMA) [32]. Na figura 10 podemos observar o resultado da aplicação do teste Dickey-Fuller em uma série temporal. Neste caso o teste foi realizado utilizando a linguagem Python e a biblioteca de estatística Statsmodels que implementa uma série de recursos estatísticos.

```

▶ adf_test(y)
↳ Resultado do teste de Dick-Fuller:
Teste          0.815369
Valor P        0.991880
# de lags      13.000000
# de observações 130.000000
Valor crítico(1%) -3.481682
Valor crítico(5%) -2.884042
Valor crítico(10%) -2.578770
dtype: float64

```

Figura 10. Teste Dick-Fuller

Assim, se após o teste de Dickey-Fuller for observado que a série temporal não tem um comportamento estacionário, pode-se inferir a necessidade de diferenciação da série [33]. Entende-se que a série temporal é estacionária se *P-Value* for menor que 0,05, no caso observado da figura 10 nota-se *P-Value* de 0,9918, dessa forma necessitando de alguma intervenção para tornar a série estacionária.

Observe-se que para os dados usados para a construção desse exemplo, após a aplicação da técnica de diferenciação, obteve-se um *P-Value* de 0,05421, ou seja, a série ainda não é estacionária, em função disso foi reaplicada a diferenciação uma segunda vez, obtendo-se assim um *P-Value* de 2,732892e-29, assim obtendo-se a estacionariedade da série na sequencia plotou-se num gráfico que pode ser comparado com a série original, conforme a figura 11.

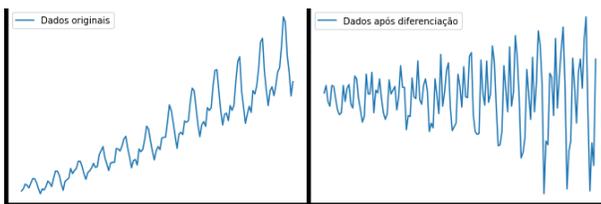


Figura 11. Série temporal não estacionária após aplicação da técnica de diferenciação

### B. Tendência

Uma série temporal é constituída por três componentes: tendência, sazonalidade e ruído (ou erro) [34]. Existem técnicas para decompor essas partes da série, pois esta é uma das principais formas de analisá-las [33]. Uma demonstração da decomposição desses componentes pode ser graficamente observada na figura 12. Note-se que o primeiro quadro refere-se a série original, no segundo vê-se a tendência, no terceiro a sazonalidade e no último o ruído ou erro. Assim como na figura 11 este gráfico utiliza os mesmos dados.

Os gráficos de linhas ou de velas, como observados na figura 3, são algumas alternativas para identificar tendências, que são movimentos de incremento ou decremento do valor da função ao longo do tempo [31]. Além dos gráficos existem modelos que podem ser utilizados para determinar a tendência e até o tipo de tendência, seja ela linear, cíclica ou de cosseno [30]. Aproximando a tendência do objeto deste trabalho, o preço das ações. A tendência surge então como importante componente para analisar, organizar, estudar e a partir dessas informações, fundamentar-se para a tomada de decisão sobre qual o momento oportuno para investir [35]. Importante destacar que numa série temporal o valor da função não necessariamente segue fielmente a tendência, mas, a análise dos dados passados permite identificar qual a direção da tendência e quando ela mudou de direção ou sentido, ou seja, eventualmente os preços das ações podem sim contrariar um movimento de tendência previamente definido, por isso essas séries podem ser interpretadas como estocásticas ou com tendências estocásticas [33].

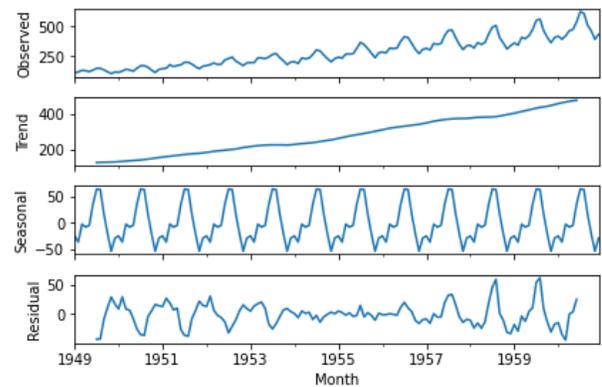


Figura 12. Decomposição de série temporal

Isto posto, pode-se então interpretar a tendência numa série temporal a partir de dois fenômenos ou características, assim elas podem ser do tipo determinístico ou estocástico. Entende-se por tendência determinística aquela que obedece a alguma função no tempo, de forma previsível, podendo evoluir de várias formas. E a tendência estocástica trata-se de um movimento aleatório no tempo, que não há previsibilidade. Para analisar estes tipos diferentes de tendência existem diversos modelos como passeio aleatório (*random walk*) e outros [33].

### C. Sazonalidade

Sazonalidade ou fenômenos sazonais refere-se a eventos que ocorrem com regularidade, ano após ano, como o aumento de vendas de passagens aéreas no verão, crescimento das vendas do comércio varejista no final do ano, etc [36]. Ou seja, são os eventos de crescimento ou decréscimo que ocorre com alguma regularidade, ou melhor, com alguma periodicidade. Isso pode ser percebido no gráfico da figura 12, onde o componente sazonal foi isolado e pode-se notar um movimento periódico regular no comportamento da série temporal analisada.

Este é o principal componente das séries temporais a ser investigado neste trabalho, pois, o que se quer é justamente identificar o quão forte e regular é a sazonalidade de algumas

ações. Óbvio que é sabido que este componente não será forte em todas as ações, o que se quer é investigar, conhecer, para que como foi dito no item anterior, a informação obtida possa servir de suporte para a tomada de decisão. Dito isto, impele-se então saber quais as formas de decompor um série temporal, com suas vantagens e desvantagens para o resgate e avaliação do componente sazonal.

O método utilizado para fazer a decomposição da figura 12 é conhecido por decomposição clássica que é dividida em 2 tipos: decomposição aditiva e multiplicativa.

Na decomposição aditiva a série temporal é a soma de seus componentes:

$$Y = T + S + R$$

Em na decomposição multiplicativa, a série temporal será o produto de seus respectivos componentes:

$$Y = T * S * R$$

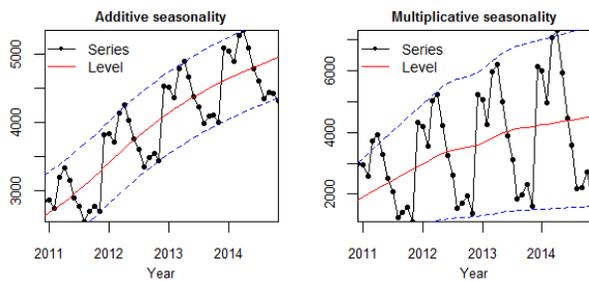


Figura 13. Aditivo versus Multiplicativo

A forma de identificar o tipo de série é a partir da magnitude do componente sazonal se ele muda com o tempo é multiplicativa, se não é aditiva [37], como pode ser observado na figura 13.

Essa abordagem de decomposição da série temporal tem alguns problemas, como: assume que o componente sazonal é constante ao longo do tempo, o que tende a ser insustentável, principalmente em longos períodos. Outro problema é que a linha de tendência é suavizada, não respondendo bem a flutuações bruscas [37]. Contudo, existem outras abordagens para executar esta tarefa e será discutido brevemente o método chamado *STL: Seasonal-Trend Decomposition Procedure Based on Loess*.

O STL tem algumas vantagens: lida bem com qualquer tipo de sazonalidade, o usuário pode controlar a taxa de variação do componente sazonal e é muito robusto para tratar outliers [37]. A figura 14 exibe a diferença entre os métodos. Em ambos os gráficos foi utilizado a mesma série temporal, os mesmos dados, na abordagem clássica foi utilizado o tipo aditivo. Observa-se no STL uma suavização da tendência, ao mesmo tempo no método clássico a sazonalidade é exposta com uma regularidade e constância não percebida no STL e, também, o resíduo se comporta de forma diferente em ambos os métodos, apresentando muito menos outliers no STL.

STL pode ser descrito da seguinte forma:

$$Y_v = T_v + S_v + R_v$$

Onde  $v = 1$  para  $N$  pontos de dados medidos. A decomposição STL é feita por meio de 2 loops. No primeiro loop,

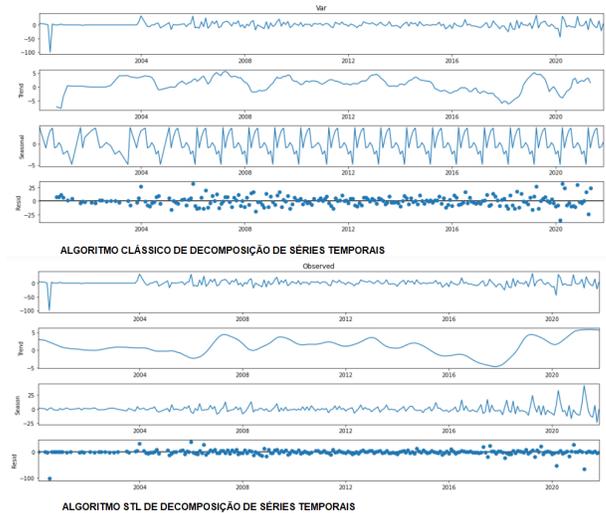


Figura 14. Comparação de algoritmos de decomposição de séries temporais

os pesos de robustez são atribuídos a cada ponto de dados dependendo do tamanho do restante. Isso permite reduzir ou eliminar os efeitos de outliers [38].

Desta forma, tem-se então duas abordagens distintas para lidar com a sazonalidade, com o método de decomposição clássica e com o STL os excessivos testes destes com os dados definirão as melhores técnicas e formas de utilizá-las.

#### D. Autocorrelação

Entende-se por autocorrelação quando diversos pontos amostrais de uma série apresentam algum grau de correlação com lags de 12 meses, ou seja, existe correlação entre  $Y_{t-12}$  e  $Y_{t+12}$  com os valores de  $Y_t$  [33]. Com isto, séries sazonais em geral têm alta autocorrelação que estas apresentam em múltiplos períodos [31].

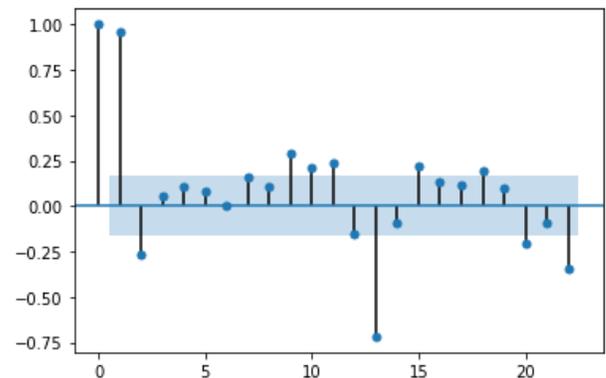


Figura 15. Correlograma

Uma forma de avaliar a autocorrelação de uma série temporal é utilizar um gráfico chamado de correlograma (ver figura 15), que tem em seu eixo 'X' as lags e, por padrão, a primeira lag é sempre igual a 1. Quanto mais próximo o valor de 1, maior é o nível de autocorrelação da série.

## V. ESTUDO DE CASO

Este estudo de caso passou por uma série de etapas para que seu resultado pudesse ser possível, ou melhor, factível. Segundo a dinâmica do KDD expressa anteriormente. Ao longo das próximas linhas será demonstrado as etapas vencidas para a conclusão deste trabalho.

### A. Coleta e seleção dos dados

A primeira etapa constituiu por definir como coletar os dados, para isso foi necessário conhecer quais as empresas listadas na B3, isso foi possível por meio do portal <https://www.ivalor.com.br/empresas/listagem>, que forneceu uma listagem completa de todas as empresas (ver figura 16), com seus respectivos códigos de negociação, setor e classificação, que foi transportado para um arquivo excel, visando sua posterior utilização.

	A	B	C	D	E
	EMPRESA	ACÇÕES	CLASSIFICAÇÃO	SETOR	TICK_API
2	ALLIAR	AALR3	Small Cap	Saúde	AALR3.SA
3	ABC BRASIL	ABCB4	Small Cap	Financeiro	ABCB4.SA
4	AMBEV S/A	ABEV3	Mega Cap	Consumo não Cíclico	ABEV3.SA
5	AERIS	AERIS3	Mid Cap	Bens Industriais	AERIS3.SA
6	AES BRASIL	AESB3	Mid Cap	Utilidade Pública	AESB3.SA
7	AFLUENTE T	AFLT3	Micro Cap	Utilidade Pública	AFLT3.SA
8	BRASILAGRO	AGRO3	Small Cap	Consumo não Cíclico	AGRO3.SA
9	SPTURIS	AHEB3	Micro Cap	Consumo Cíclico	AHEB3.SA
10	SPTURIS	AHEB5	Micro Cap	Consumo Cíclico	AHEB5.SA
11	SPTURIS	AHEB6	Micro Cap	Consumo Cíclico	AHEB6.SA
12	ALLIED	ALLD3	Small Cap	Consumo Cíclico	ALLD3.SA
13	ALPARGATAS	ALPA3	Big Cap	Consumo Cíclico	ALPA3.SA
14	ALPARGATAS	ALPA4	Big Cap	Consumo Cíclico	ALPA4.SA
15	ESTAPAR	ALPK3	Small Cap	Bens Industriais	ALPK3.SA
16	ALIANSCOE	ALSC3	Mid Cap	Financeiro e Outros	ALSC3.SA
17	ALIANSCONAE	ALSO3	Mid Cap	Financeiro	ALSO3.SA
18	ALUPAR	ALUP11	Mid Cap	Utilidade Pública	ALUP11.SA
19	ALUPAR	ALUP3	Mid Cap	Utilidade Pública	ALUP3.SA
20	ALUPAR	ALUP4	Mid Cap	Utilidade Pública	ALUP4.SA
21	LOIAS MARISSA	AMAR3	Small Cap	Consumo Cíclico	AMAR3.SA
22	AMBIPAR	AMB3	Small Cap	Utilidade Pública	AMB3.SA
23	ANIMA	ANIM3	Mid Cap	Consumo Cíclico	ANIM3.SA

Figura 16. Empresas listadas na B3

Após a obtenção da relação das empresas e respectivos códigos, impeliu-se a obter os dados das negociações. Normalmente não são dados fáceis de obter de forma gratuita, em geral nem as plataformas (*home brokers*) e nem a própria B3 oferecem um ponto de acesso para coleta destes dados em massa. A partir disso assumiu-se duas linhas de trabalho desenvolver alguma ferramenta de *web scrapping* ou encontrar algum meio preexistente que permita essa coleta em massa dos dados negociados. Como a linguagem escolhida para realizar este estudo de caso foi Python, após algumas pesquisas chegou-se a uma biblioteca que fornece dados de negociação e outros dados financeiros de diversas empresas e bolsas de valores pelo mundo, através da extração dos dados no portal do *Yahoo Finance* - <https://finance.yahoo.com/>. A biblioteca é a *Pandas-Datareader*.

Como pode ser observado na figura 17, a biblioteca *Pandas-Datareader* oferece um ponto de acesso bem simples, onde basicamente precisa-se: do código da empresa, o período desejado com início e fim (neste caso o fim pode ser omitido, caso deseje-se até os dados mais recentes) e, o intervalo desejado (diário, semanal, mensal, etc). E no caso específico dos ativos brasileiros, os códigos sempre terminam com '.SA', por isso, houve uma etapa a mais antes de realizar as consultas para transformar os códigos para o padrão exigido pela biblioteca,

por exemplo 'BBSE3', é transformado para 'BBSE3.SA' antes de coletar os dados via *Pandas-Datareader*.

```
Coletando os dados na api
In [6]: 1 for tik in tickers:
        2 try:
        3     list_df.append(wb.get_data_yahoo(str(tik), start='2000-1-1', end='2021-11-30', interval="m"))
        4     list_df[len(list_df)-1]['ticker'] = tik
        5     if len(list_df[-1]) < 36:
        6         list_df.pop(-1)
        7     elif list_df[-1]['Volume'][:36:len(list_df[-1])-1].mean() < 20_000_000.00:
        8         list_df.pop(-1)
        9 except:
        10     list_error.append(tik)
        11 pass
```

Figura 17. *Pandas-Datareader*

O *pandas-datareader* devolve os dados na forma de um *Data Frame*, que é uma estrutura de dados utilizada na biblioteca *Pandas*, também no python, que nada mais é que uma tabela com diversas linhas, colunas e índices que permitem fazer uma série de análises e transformações.

Vencida estas etapas, foi preciso selecionar em quais empresas analisar esse efeito sazonal. Para isso não foi feito nenhum corte setorial ou mesmo de classificação das empresas. Optou-se por analisar empresas que tem um fluxo de negociação mais intenso. Essa escolha teve por base recomendações encontradas no livro de Fayh, sobre Fundos Imobiliários, que é outro tipo de ativo, também negociado na B3 e em outras bolsas de valores no mundo. Onde o autor indica escolher fundos para investir de acordo com seu volume negociado diariamente, na ordem de 1 milhão de Reais [39]. Por isso, selecionou-se empresas que tenham em média volume negociado de 20 milhões de Reais por mês. Com isso a lista anterior que contava com 538 empresas, foi reduzida para 128 empresas.

Por fim, o período de análise foi escolhido em um longo espaço de tempo para que houvesse mais dados a serem observados, com isso esta janela temporal vai de janeiro de 2000 até novembro de 2021. Este intervalo gerou 22.750 linhas de registros de negociações, com valores de abertura, fechamento, máxima e mínima.

Importante registrar que o intervalo dos dados negociados é mensal, ou seja, foi obtido os valores de abertura, fechamento, máxima e mínima por mês, por isso, para obter mais dados, ir mais longe no passado se mostrou uma necessidade imperativa.

### B. Limpeza, Tratamento e Transformação dos dados

A biblioteca do *Pandas-Datareader* facilitou bastante o trabalho, ao devolver os registros na forma de *Data Frames*, todavia, ainda foi necessário realizar algum esforço para a limpeza e tratamento dos dados. Para isso utilizou-se a biblioteca do *Pandas*, que é uma das mais utilizadas para análise de dados com python atualmente [40]. Assim, o *Pandas* ajudou a resolver a demanda dos dados ausentes que segundo a bibliografia existente, há diferentes formas de tratar dados inexistentes [41], como eliminar as linhas inteiras de registros, preenche-las com as médias observadas ou preenche-las com zeros, enfim, inúmeras alternativas, que dependem de várias variáveis sobre quais usar, cabendo ao analista a decisão final sobre o que fazer. Como este trabalho pretende investigar a sazonalidade existente ou não nas amostras analisadas, optou-se por remover as linhas com registros ausentes ou zerados, desta forma esperando não comprometer ou minimizar interferências no resultado final da análise. Assim, o que pode ocorrer é alguma empresa apresentar resultados não

satisfatórios, impossibilitando confirmar se tal evento existe na mesma.

Após essa etapa de limpeza, precisou-se realizar uma série de transformações nos dados. Importante registrar que boa parte das transformações brotaram do exercício exaustivo e repetitivo do KDD, especialmente focando nas fases do CRISP-DM e na mineração de dados que foi discutido anteriormente. Desta forma, diversas análises e transformações foram produzidas para interpretar os dados e encontrar padrões. Por fim, algumas transformações destacaram-se. Primeiro entendeu-se os dados das negociações na B3 como 'séries temporais' e, em função disso algumas exigências impeliram-se: um, os dados precisam usar um índice cronológico ordenado, ao criar um data frame no pandas, automaticamente um índice é criado, neste caso, este teve que ser alterado para as datas de movimentação, ou seja, diferentemente de outros data frames com índices incrementais, neste estudo de caso, os data frames possuem índices cronológicos; dois, como se tratam de índices temporais e a análise leva em consideração a passagem do tempo, estes então precisaram ser ordenados cronologicamente, de forma crescente e; por fim, como já foi discutido anteriormente eventualmente os dados de alguma empresa precisaram passar por um processo a mais para transformá-lo em uma série estacionária.

Outra transformação importante foi necessária diretamente sobre os dados observados, pois, a leitura dos dados brutos mostrou-se ineficaz para o objeto deste estudo de caso, que centra-se na existência ou não de sazonalidade nas negociações das empresas na bolsa de valores. Em função disso, acrescentou-se mais um valor que é a variação percentual do fechamento em relação a abertura, ou seja, se a empresa BBSE3, abriu a negociação em 01-05-2013 a R\$ 17,05 e fechou no fim do mesmo mês sendo comercializada a R\$ 18,47, a variação percentual foi positiva, na ordem de 8,33%. A introdução dessa informação abriu uma nova janela de análise, permitindo investigar movimentos sazonais de forma localizada. Com isso, mais uma nova transformação foi possível, a categorização das variações percentuais, agrupando os valores em grupos, de valores negativos, passando por valores nulos e chegando as variações positivas. Isto permitiu unir grupos com variações percentuais próximas, facilitando a leitura e interpretação dos dados. Uma amostra do resultado pode ser visto na figura 18.

Tais transformações foram importantes para a devida apreensão do conhecimento sobre os dados analisados. Basicamente as variações foram agrupadas, isolando variações iguais a 0,00 (zero) em um só grupo, por exemplo, valores que vai acima de 0,00 % até 2,00 %, chamou-se de grupo '2' e a variação na mesma proporção negativa, intitulou-se de '-2' e, assim foi de 2 em 2 %, e por fim, os valores acima de 8 % e abaixo de -8 %, são identificados respectivamente com '10' e '-10'. Com o auxílio do Power Bi<sup>3</sup> pôde-se observar que grande partes das linhas desse data frame apresenta um número significativo de variações percentuais mensais que localizam-se acima dos 8,00% e abaixo do -8,00%. A figura 19 através de um gráfico de pizza, demonstra bem isso.

A medida que se conhece os dados novos saberes são

<sup>3</sup>Ferramenta da Microsoft para análise de negócios, que oferece uma interface interativa com recursos vários para *business intelligence*

```
In [23]: 1 list_df[10].head(20)
```

```
Out[23]:
```

Date	High	Low	Open	Close	Volume	Adj Close	Ticker	Var	Category
2013-05-01	18.900000	16.980000	17.049999	18.469999	235634100.0	10.486530	BBSE3.SA	8.328447	10
2013-06-01	18.440001	15.980000	18.360001	17.600000	124878200.0	9.992579	BBSE3.SA	-4.139435	-6
2013-07-01	18.450001	16.799999	17.270000	18.440001	80988300.0	10.469497	BBSE3.SA	6.774754	8
2013-08-01	20.100000	17.800001	18.440001	19.240000	98911900.0	10.923703	BBSE3.SA	4.338391	6
2013-09-01	22.160000	19.000000	19.040001	21.809999	102208000.0	12.670671	BBSE3.SA	14.548311	10
2013-10-01	24.600000	21.350000	21.750000	24.469999	96452900.0	14.216017	BBSE3.SA	12.505744	10
2013-11-01	25.590000	22.660000	24.469999	25.400000	76612400.0	14.756308	BBSE3.SA	3.800573	4
2013-12-01	25.500000	23.620001	25.209999	24.500000	59691100.0	14.233447	BBSE3.SA	-2.816339	-4
2014-01-01	24.500000	22.150000	24.500000	22.700001	115349700.0	13.187724	BBSE3.SA	-7.346936	-8
2014-02-01	24.600000	21.540001	22.600000	23.299999	100657300.0	13.536298	BBSE3.SA	3.097340	4
2014-03-01	25.120001	22.070000	23.299999	25.110001	127215800.0	14.959227	BBSE3.SA	7.768246	8
2014-04-01	26.840000	24.520000	25.110001	26.150000	109069500.0	15.578005	BBSE3.SA	4.141772	6
2014-05-01	28.990000	26.150000	26.150000	28.500000	96894200.0	16.978811	BBSE3.SA	8.986617	10
2014-06-01	32.880001	28.230000	28.590000	32.439999	76369300.0	19.326057	BBSE3.SA	13.466242	10
2014-07-01	35.299999	31.370001	32.310001	33.099998	78426100.0	19.719250	BBSE3.SA	2.445054	4
2014-08-01	35.809999	31.900000	33.080002	35.689999	90724000.0	21.262236	BBSE3.SA	7.889954	8
2014-09-01	35.980000	29.890000	35.520000	32.200001	95581700.0	19.522953	BBSE3.SA	-9.346846	-10
2014-10-01	33.279999	27.980000	32.040001	33.060001	124141500.0	20.044384	BBSE3.SA	3.183522	4
2014-11-01	35.200001	30.700001	33.180000	33.520000	75836900.0	20.323282	BBSE3.SA	1.024714	2
2014-12-01	32.810001	29.209999	32.509998	32.160000	35389482.0	19.498707	BBSE3.SA	-1.076587	-2

Figura 18. Data Frame

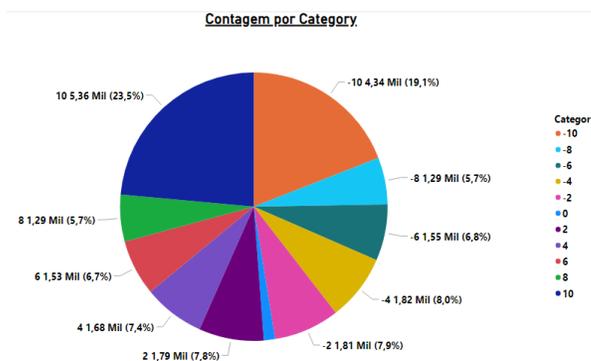


Figura 19. Variação percentual organizada por categoria de valores

agregados a análise. Ao plotar os dados em gráficos de linha, apesar da aparente confusão visual, pois, centenas de linhas estão sendo exibidas no mesmo eixo simultaneamente, informações podem ser extraídas. Cabe registrar que a meta do estudo de caso é automatizar a interpretação desses dados, mas, tal tarefa exige do analista o conhecimento dos dados, para poder fazer uso das melhores técnicas e ferramentas para produzir os saberes a partir da respectiva fonte de dados. Com isso, foi possível identificar visualmente que existem alguns valores outliers, além disso, no gráfico por ano percebe-se claramente que apesar de usar no estudo uma janela de mais de 20 anos para coletar os dados de negociações, algumas empresas só foram ter papéis negociados na bolsa anos depois de 2000, com isso diminuindo o espectro de observação e de confirmação (ver figura 20). Também foi possível visualizar que boa parte das empresas possuem características estacionárias em seus gráficos, com isso nem todas precisaram ser transformadas, todavia, para isso um teste dinâmico e automatizado foi implementado e utilizado para validar essa condição em todas as séries temporais, agrupadas por empresa.

Também foi possível perceber os volumes negociados e quais as empresas mais negociadas na B3, como pode ser observado nos gráficos representados nas figuras 21 e 22. Analisando no longo tempo a Petrobras se destaca com PETR4 e PETR3 com volumes bem significativos, todavia, ao olhar apenas o gráfico relativo apenas ao ano de 2020, o cenário

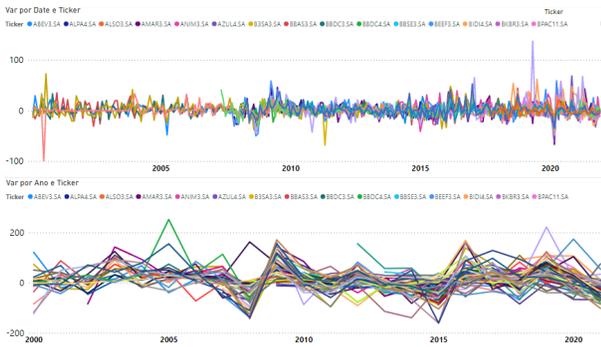


Figura 20. Gráficos de linha da variação percentual, agrupados por mês e por ano

se repete, contudo, com muito menos força e, um quadro um pouco mais equânime se expõe, especialmente quando comparado com boa parte das grandes empresas. Isso ocorre porque a Petrobras é negociada há muito tempo na B3, faz parte de um grupo de ações que estão entre as mais negociadas na B3 [42] [43], naturalmente seu volume no longo tempo tende a ser grande. Além disso, há de se perceber que algumas empresas só começam a ser negociadas na B3 anos depois da Petrobras, mais um motivo para que algumas tenham um volume negociado menor quando observado no longo tempo.

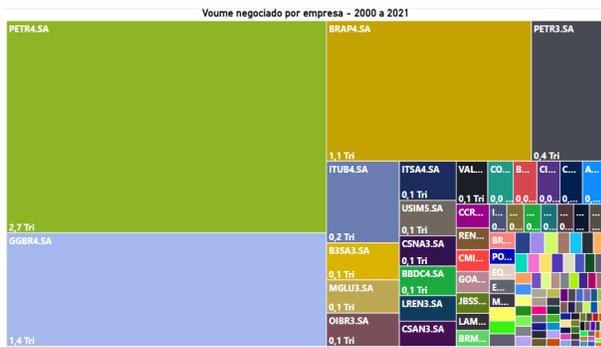


Figura 21. Volume negociado por empresa de 2000 a 2021



Figura 22. Volume negociado por empresa em 2020

Para o objeto deste estudo os dados no longo tempo são bastante significativos, especialmente porque está sendo utilizado um ciclo observacional mensal, de 12 meses, com isso a amostra por empresa e por ano resume-se a 12 linhas de registros, assim para empresas negociadas há pelo menos 20 anos, teremos uma amostra de aproximadamente 240 registros,

já para empresas que tiveram seu *IPO*<sup>4</sup> muito recentemente, diminui muito a base de dados a ser observada e, eventualmente não contribuindo para responder o questionamento inicial deste estudo de caso, se há ou não sazonalidade em determinado ativo.

Este ciclo observacional mensal foi selecionado por ser uma métrica utilizada não apenas no mercado financeiro, mas, em vários aspectos da vida em geral. A passagem dos meses é de alguma forma percebida e a passagem dos anos, apresenta-se como um ciclo que obedece algum grau de repetitividade, ou melhor, de sazonalidade. Por exemplo, o verão é uma estação do ano, que compreende um período de 3 meses, que normalmente coincide com as férias escolares, que coincide com o período em que as agências de turismo tem um crescimento do seu fluxo de atendimento, também coincide com o período em que as instituições educacionais tem uma redução de suas atividades, em outro momento, no fim do verão observa-se o retorno as aulas, o declínio da atividade turística, ao mesmo que ocorre um crescimento do comércio das editoras escolares, livrarias e congêneres. Pode-se ainda lembrar que março e abril é o período da entre-safra da produção açucareira no Brasil, automaticamente diminuindo a oferta nacional de álcool para as empresas que comercializam esse tipo de combustível, ou mesmo no final de ano que é o ponto alto do comercio varejista, com as festas de fim de ano. Ainda tem-se as publicações oficiais como demonstrações de financeiras das empresas, indicadores econômicos como inflação, PIB, fornecidos pelo IBGE e outros, que obedecem espaços temporais que variam de semanal, mensal, trimestral, anual, etc. O objetivo é perceber que o intervalo mensal a priori oferece um ponto de acesso onde a 'natureza' da organização social moderna já percebe um ciclo que carrega em si alguma dose de sazonalidade a depender do dado a ser observado. E partindo desse pressuposto entendeu-se que as empresas negociadas na bolsa de valores pertencem a diversos segmentos que ocasionalmente podem sofrer alguma influência sazonal que pese sobre como seus ativos são negociados, como no caso da Petrobras, que tem em seu leque de produtos o álcool, ou a COGNA, grupo educacional dos maiores do Brasil, ou a CVC, grande empresa do setor de turismo, ou o Magazine Luíza, varejista importante, todos negociados na B3. Por isso, foi escolhido a priori a leitura dos dados obedecendo um ciclo mensal, isso não quer dizer que o mesmo estudo de caso não possa ser realizado obedecendo outros intervalos, como 1 minuto, 5 minutos, 1 hora, 4 horas, diário, semanal, anual, etc.

### C. A Sazonalidade

Para identificar a sazonalidade alguns métodos e a mescla de outros foram utilizados. Cabe registrar que não foi possível identificar este movimento sazonal em todos os ativos analisados e que as diversas técnicas como era de se esperar apresentam resultados com algum grau de diferença e, que a ideia foi mostrar o quão intenso é o movimento sazonal de determinado ativo, sem querer dizer categoricamente que sim ou que não, até porque este seria um tremendo erro, por não respeitar a 'humanidade' daqueles que negociam na bolsa de valores e, que por isso, sofrem diversas influências, que

<sup>4</sup>Expressão utilizada para descrever uma oferta pública inicial de uma empresa na bolsa de valores, ou seja, o dia em que começa a ser negociada.

concomitantemente refletem-se na forma como se negocia. Por exemplo, quando uma empresa esta muito bem avaliada, com índices cada vez melhores é comum ocorrer o efeito chamado de 'manada', onde muitos correm para comprar o mesmo ativo por entenderem que aquele continuará num grande ciclo de valorização e, essa procura incessante tende a valorizar a o preço da ação da respectiva empresa.

Neste sentido foram testados 2 métodos para identificar o efeito sazonal nas negociações analisadas. Para isso, fez-se uso de algumas bibliotecas python para análise de dados, além das já mencionadas, utilizou-se a biblioteca de estatística statsmodels, especialmente os pacotes seasonal-decompose, STL e adfuller, que permitem respectivamente decompor uma série temporal em tendência, sazonalidade e resíduo usando um algoritmo chamado de 'clássico', também um outro algoritmo chamado de STL que realiza as mesmas operações de decomposição de séries temporais, todavia, com uma técnica diferente e, também o pacote adfuller, que utiliza o teste de Dickey-Fuller para determinar se determinada série temporal é ou não estacionária. Também foi utilizado a biblioteca de matemática Numpy que permite através de uma técnica de logaritmos transformar uma determinada série temporal em estacionária.

1) **Algoritmo Clássico:** O primeiro método de decomposição utilizado foi o clássico. O trabalho consistiu em produzir e organizar um grupo de funções que fossem capazes de analisar individualmente as séries de cada ativo, realizando as transformações necessárias. Feito isto, foi necessário desenvolver uma estrutura que pudesse enfileirar estas séries e analisar seus retornos.

A tarefa primeira foi validar se a série temporal a ser utilizada é estacionária, para isso foi desenvolvido o método 'test\_and\_transform\_stationary\_ts' (ver figura 23). Este foi responsável por testar as séries isoladamente, devolvendo-a no formato necessário para avançar na análise. É certo que o método escolhido para transformar a série em estacionária foi o de logaritmo, onde a transformação é aplicada no máximo até 3 vezes, em caso de insucesso a série não poderá "a priori" avançar neste estudo de caso, especificamente. Todavia, a transformação inicial onde obteve-se os valores de variação percentual, permitiu que o comportamento da maior parte das séries fosse "naturalmente"estacionário.

```
In [4]: 1 def test_and_transform_stationary_ts(ts : pd.Series):
2     """
3     Recebe uma série temporal, testa e se preciso transforma e devolve uma serie estacionária.
4     Retorna None em caso de falha
5     """
6     adfuller_ts = adfuller(ts, autolag='AIC')
7     i = 0
8     while(adfuller_ts[1] < 0.05 and i < 4):
9         if adfuller_ts[1] < 0.05:
10            i = 4
11            return ts
12        elif i < 3:
13            adfuller_ts = np.log(adfuller_ts)
14            adfuller_ts = adfuller(adfuller_ts, autolag='AIC')
15            i += 1
16        else:
17            return
```

Figura 23. Teste de estacionariedade e transformação da série temporal

A partir deste ponto foi utilizado através de métodos próprios o pacote seasonal\_decompose, da biblioteca já mencionada do Statsmodels, que possibilita isolar o componente sazonal. Em posse deste, foi implementada uma função para avaliar a intensidade do movimento sazonal dentro da série analisada. Como pôde ser observado na figura 14, a sazonalidade move-se dentro de um intervalo de números negativos a

positivos, passando pelo zero. Assim as curvas do gráfico são mais evidentes, quanto mais distantes ficam do ponto do zero. Por isso, em linhas bem gerais o método avalia a extensão da série e define um ponto do meio para avaliar os dados. Devolvendo para o analista duas *tuplas*<sup>5</sup>, cada uma com duas listas, que apresentam os meses e os valores que se destacaram na avaliação e, com a separação das *tuplas*, conhece-se quais períodos tiveram movimentos de alta ou baixa, seja, apresenta a tendência do respectivo período.

Também foram definidos formas de avaliar os resultados, comparando os retornos das funções com a série temporal em si. Com isso criou-se um viés de confirmação, confrontando os dados obtidos nos algoritmos com os dados de negociação inicialmente conseguidos. Assim, definiu-se 4 escalas de confirmação, descritas na tabela da figura 24 como 'Hit 0', onde valida-se se a tendência esta correta, por exemplo, se os dados processados indicam tendência de valorização naquele mês, quanto maior os casos em que o mês tem variação positiva, mais assertiva é percentualmente a avaliação, o mesmo vale para o contrário, se o resultado indicar tendência e desvalorização e confirma-se isto nos dados, aumenta-se o percentual de acerto. Para as demais avaliações analisou-se além da tendência correta, o percentual de valorização no período, dividido em acima de 2%, acima de 4%, e acima de 6%, considerando-se os valores negativos, se a tendência for de desvalorização.

Ticker	Variable	Trend	Month	Total Months	Hit 0	Hit 2	Hit 4	Hit 6
ABEV3.SA	Var	max	12	22	59,1%	54,5%	45,5%	31,8%
ABEV3.SA	Var	min	2	22	50,0%	36,4%	18,2%	18,2%
ABEV3.SA	Var	min	5	22	59,1%	40,9%	18,2%	9,1%
ALPA4.SA	Var	max	10	14	78,6%	78,6%	71,4%	64,3%
ALPA4.SA	Var	max	4	14	71,4%	57,1%	50,0%	50,0%
ALPA4.SA	Var	min	8	14	50,0%	42,9%	35,7%	28,6%
ALPA4.SA	Var	min	9	14	50,0%	50,0%	50,0%	35,7%
ALPA4.SA	Var	min	3	14	35,7%	21,4%	14,3%	14,3%
ALPA4.SA	Var	min	7	14	42,9%	28,6%	28,6%	21,4%
ALSO3.SA	Var	max	10	11	63,6%	63,6%	36,4%	36,4%
ALSO3.SA	Var	max	7	11	45,5%	45,5%	27,3%	27,3%
ALSO3.SA	Var	min	11	11	45,5%	45,5%	27,3%	27,3%
ALSO3.SA	Var	min	5	11	63,6%	54,5%	36,4%	18,2%
ALSO3.SA	Var	min	6	11	54,5%	45,5%	36,4%	18,2%
AMAR3.SF	Var	max	8	14	78,6%	71,4%	64,3%	42,9%
AMAR3.SF	Var	max	4	14	57,1%	57,1%	57,1%	50,0%
AMAR3.SF	Var	min	11	15	60,0%	60,0%	60,0%	53,3%
ANIM3.SA	Var	max	11	9	88,9%	66,7%	66,7%	44,4%
ANIM3.SA	Var	min	2	8	62,5%	50,0%	50,0%	37,5%
AZUL4.SA	Var	max	11	5	60,0%	40,0%	40,0%	40,0%
AZUL4.SA	Var	min	3	4	50,0%	50,0%	50,0%	50,0%

Figura 24. Decomposição Clássica - amostra dos dados obtidos

Tais testes foram realizados duas vezes, considerando-se as variáveis criadas no início do estudo, que é a variação percentual entre abertura e fechamento por mês dos ativos e, a outra variável deriva desta a medida que categoriza essas variações, então os resultados serão apresentados considerando 'variação percentual' e 'categoria'. Isto posto, no gráfico da figura 25 pode-se fazer algumas observações. Todavia, antes carece registrar que apesar de não ser recomendado utilizar gráficos com mais de 2 eixos, segundo Knaffic, esse tipo de solução não tem uma leitura fácil e quase sempre reclama explicação [44]. Entretanto, neste caso escolheu-se incluir uma linha para que possa-se avaliar os resultados acompanhado de um valor importante para a melhor interpretação, que é o total dos dados observados para o período e empresa selecionados.

<sup>5</sup>Tupla é uma estrutura de dados nativa da linguagem python que se assemelha muito a listas, arrays observados em outras linguagens. Trata-se de um objeto iterável, mas, sua principal característica é a imutabilidade.

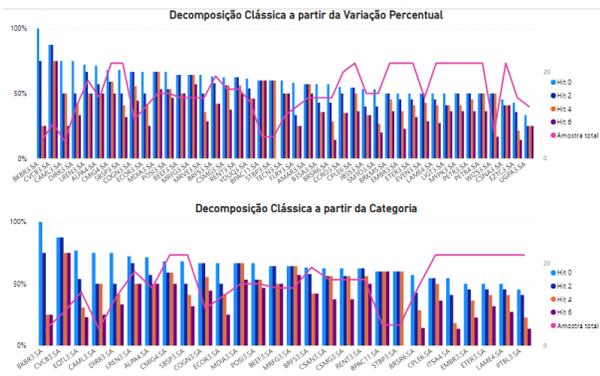


Figura 25. Resultados para o mês de abril, através do método de decomposição clássica

Na figura 25 as barras exprimem o percentual de acerto de 0 a 100%. E a linha rosa exibe a partir do eixo a direita o volume total amostral de cada empresa. Neste gráfico está sendo observado o mês de abril e, como pode ser visto há um grupo de empresas que confirmam a existência de algum movimento sazonal nas negociações da bolsa de valores no Brasil. Todavia, a intensidade desse movimento é confirmado a partir do seu volume amostral, observe-se o caso do código BKBR3, onde tem-se 100% para Hit 0 e 75% para Hit 2, todavia, só foram encontradas referência para 4 meses de abril em seu histórico de negociação na B3, contrastando com PETR3, que tem Hit 0 e Hit 2 de 50%, enquanto a sua amostra tem 22 registros para o mês de abril. Isso expõe um risco na análise, as empresas que possuem um histórico de movimentação menor podem exibir resultados que supervalorizam a sazonalidade, que poderá ser continuamente confirmada ou não, enquanto empresas com mais tempo de negociação, em função da existência de mais dados, demonstram resultados mais sólidos. Observe-se o caso da CMIG4, que tem Hit 0 de 68,18% e Hit 2 de 59,09%, com 22 registros para o mesmo mês de abril, assim, confirmando sim, que existe uma tendência sazonal no mês de abril para o ativo.

Ainda cabe registrar que a mesma análise a partir da variável 'categoria' apresentou resultados pouco diferentes do observado na 'variação percentual', como já era de se esperar, pois, uma variável deriva da outra.

2) **Algoritmo STL:** A obtenção de dados da série sazonal e os diversos testes foram utilizados praticamente os mesmos métodos que com o algoritmo clássico, a exceção se deu com o algoritmo para gerar e isolar o componente sazonal, que neste caso foi utilizado o pacote STL da biblioteca de estatística do python, Statsmodels. Como já foi explicado o funcionamento do algoritmo de decomposição de séries temporal STL, na figura 26 apenas demonstra-se comparativamente como os dados foram isolados em cada técnica.

Os resultados para o STL é que apresentam características bem diferentes do algoritmo clássico. Pois, como o STL ao decompor os componentes da série temporal dá um peso maior para os eventos recentes e, menor para os mais antigos, tal comportamento conflita-se com os métodos aqui utilizados para testar e analisar as séries, pois, busca-se um viés de confirmação no conjunto completo dos dados, desprezando-se por exemplo que um ciclo sazonal por razões diversas pode

```

1 def components_seasonal_trend_resid_STL(ts : pd.Series, period: int):
2     """
3     Retorna uma tupla com os componentes isolados, sazonalidade, tendência e resíduo.
4     Método STL.
5     """
6     ts = ts.fillna(method='ffill')
7     ts = ts.asfreq('MS')
8     result = STL(ts, robust=True, period=period).fit()
9     return (result.seasonal, result.trend, result.resid)

1 def components_seasonal_trend_resid_classic(ts : pd.Series, freq : int):
2     """
3     Retorna uma tupla com os componentes isolados, sazonalidade, tendência e resíduo.
4     Método classico.
5     """
6     ts.sort_index(inplace=True)
7     result = seasonal_decompose(ts, period=freq, model='additive')
8     return (result.seasonal, result.trend, result.resid)

```

Figura 26. Decomposição de série temporal com STL

com o tempo mudar o período em que ocorre. Em função disso, não raras as vezes o retorno das avaliações foi apresentando sem valor, como pode ser visto na tabela da figura 27.

Ticker	Variable	Trend	Month	Total Months	Hit 0	Hit 2	Hit 4	Hit 6
ABEV3.SA	Var			0	0,0%	0,0%	0,0%	0,0%
ABEV3.SA	Var			0	0,0%	0,0%	0,0%	0,0%
ALPA4.SA	Var	max	5	14	64,3%	64,3%	64,3%	64,3%
ALPA4.SA	Var	min	2	14	50,0%	42,9%	42,9%	28,6%
ALSO3.SA	Var	max	11	11	54,5%	54,5%	54,5%	27,3%
ALSO3.SA	Var	max	12	11	63,6%	54,5%	45,5%	18,2%
ALSO3.SA	Var	min	3	11	27,3%	18,2%	18,2%	18,2%
AMAR3.SA	Var			0	0,0%	0,0%	0,0%	0,0%
AMAR3.SA	Var			0	0,0%	0,0%	0,0%	0,0%
ANIM3.SA	Var			0	0,0%	0,0%	0,0%	0,0%
ANIM3.SA	Var			0	0,0%	0,0%	0,0%	0,0%
AZUL4.SA	Var	max	5	5	40,0%	40,0%	40,0%	40,0%
AZUL4.SA	Var	min	10	5	60,0%	60,0%	60,0%	60,0%
B3SA3.SA	Var	max	5	14	50,0%	35,7%	35,7%	28,6%
B3SA3.SA	Var	max	11	15	33,3%	26,7%	13,3%	6,7%
B3SA3.SA	Var	max	12	15	46,7%	40,0%	26,7%	20,0%
B3SA3.SA	Var	max	4	14	57,1%	42,9%	35,7%	35,7%
B3SA3.SA	Var	min	8	14	64,3%	64,3%	57,1%	42,9%
B3SA3.SA	Var	min	2	14	50,0%	35,7%	21,4%	21,4%
B3SA3.SA	Var	min	10	14	42,9%	28,6%	28,6%	28,6%
BBA53.SA	Var			0	0,0%	0,0%	0,0%	0,0%

Figura 27. Decomposição STL - amostra dos dados obtidos

Tal resultado não compromete a qualidade dos mesmos, apenas diminui a quantidade de confirmações ou negativas para sazonalidade em alguns ativos. Na figura 28 tem-se a os mesmos gráficos e dados utilizados na figura 25, assim, o volume de empresas no qual foi possível realizar alguma confirmação foi menor. Em todo caso, a interpretação dos dados continua muito similar, nota-se que as empresas com um volume de registros de negociação menor e, com sazonalidade confirmada, tem índices de confirmação maiores.

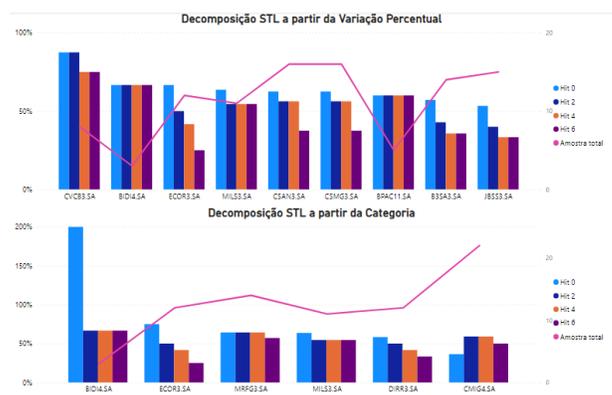


Figura 28. Resultados para o mês de abril, através do método de decomposição STL

## D. Resultados

Neste estágio do estudo de caso já é possível identificar algum vestígio de sazonalidade nos ativos da Bolsa de Valor do Brasil, não em todos os ativos, nem com a mesma regularidade ou intensidade em todos.

Quanto aos algoritmos utilizados para a decomposição das séries temporais, clássico e STL. Ambos funcionaram muito bem, mas, a forma como cada um funciona em relação a forma como os dados foram analisados determinou a quantidade (e não a qualidade) de resultados obtidos.

Ticke	Variable	Trend	Mont	Total Month	Hit 0	Hit 2	Hit 4	Hit 6
POM04.S	Var	max	5	22	68,2%	54,5%	31,8%	27,3%
RAN13.SA	Var	max	6	22	54,5%	36,4%	36,4%	27,3%
RAN13.SA	Var	max	10	22	59,1%	45,5%	45,5%	36,4%
RAN13.SA	Var	min	5	22	54,5%	54,5%	50,0%	40,9%
RAPT4.SA	Var	max	7	22	59,1%	54,5%	54,5%	50,0%
SBSF3.SA	Var	max	4	22	68,2%	50,0%	40,9%	31,8%
SBSF3.SA	Var	max	12	22	72,7%	68,2%	63,6%	50,0%
SBSF3.SA	Var	min	9	22	59,1%	50,0%	31,8%	22,7%
TRPL4.SA	Var	max	12	22	86,4%	72,7%	50,0%	36,4%
USIM5.SA	Var	max	3	22	63,6%	63,6%	50,0%	45,5%
USIM5.SA	Var	max	7	22	68,2%	68,2%	63,6%	45,5%
USIM5.SA	Var	min	5	22	68,2%	68,2%	68,2%	54,5%
VALE3.SA	Var	max	9	22	54,5%	50,0%	45,5%	36,4%
VALE3.SA	Var	max	2	22	63,6%	50,0%	40,9%	36,4%
VALE3.SA	Var	max	12	22	72,7%	50,0%	50,0%	40,9%
VALE3.SA	Var	max	7	22	59,1%	59,1%	59,1%	59,1%
VALE3.SA	Var	min	8	22	59,1%	59,1%	50,0%	40,9%
VALE3.SA	Var	min	3	22	54,5%	45,5%	40,9%	18,2%
VALE3.SA	Var	min	6	22	59,1%	45,5%	31,8%	27,3%
Total								66

Figura 29. Total de registros no longo tempo com algoritmo clássico de decomposição

Ticke	Variable	Trend	Mont	Total Month	Hit 0	Hit 2	Hit 4	Hit 6
BBDC3.SA	Var	max	11	22	63,6%	59,1%	54,5%	40,9%
BRKM5.SA	Var	max	8	22	86,4%	77,3%	54,5%	45,5%
BRKM5.SA	Var	min	3	22	59,1%	59,1%	50,0%	45,5%
CMIG4.SA	Var	max	7	22	63,6%	50,0%	45,5%	27,3%
CPLE6.SA	Var	max	11	22	59,1%	50,0%	50,0%	45,5%
PETR3.SA	Var	max	7	22	63,6%	50,0%	27,3%	22,7%
PETR4.SA	Var	max	7	22	59,1%	40,9%	31,8%	27,3%
PETR4.SA	Var	min	11	22	63,6%	50,0%	45,5%	31,8%
SBSF3.SA	Var	max	2	22	54,5%	50,0%	40,9%	36,4%
SBSF3.SA	Var	min	5	22	59,1%	40,9%	40,9%	27,3%
USIM5.SA	Var	min	5	22	68,2%	68,2%	68,2%	54,5%
VALE3.SA	Var	max	9	22	54,5%	50,0%	45,5%	36,4%
VALE3.SA	Var	min	3	22	54,5%	45,5%	40,9%	18,2%
Total								13

Figura 30. Total de registros no longo tempo com algoritmo STL de decomposição

As figuras 29 e 30 demonstram bem o desempenho de cada técnica em relação a quantidade de resultados obtidos. Para esses resultados filtrou-se apenas os registros que tem 22 meses de amostras para os mesmos meses em todo o conjunto de dados, destacando-se as linhas que tiveram Hit 0 superior a 50%. Note-se que com STL (figura 30) tem-se apenas 13 registros identificados, enquanto para o algoritmo clássico (figura 29) vê-se 66 registros para as mesmas condições.

Desta forma, a partir das técnicas e métodos aqui utilizados o algoritmo clássico obteve uma avaliação melhor a medida que é capaz de oferecer uma gama maior de resultados para o analista, para enfim decidir quando e em qual ativo investir.

## E. Trabalhos futuros

Existe um universo imenso de possibilidades a ser explorado ainda dentro da ótica da sazonalidade dos preços entre os ativos negociados na B3. Inclusive oferecendo outros suportes a decisão. Destaque-se uma tentativa neste mesmo estudo de caso que não pôde ser concluída, que é um algoritmo

próprio que ignora as características temporais da amostra e pressiona por resultados melhores ao utilizar pesos para comparar a qualidade dos resultados e, a normalização dos resultados como meio de facilitar a leitura para o analista sem muito conhecimento do problema.

Ainda é possível melhorar os métodos aqui utilizados para que como o STL faz, garantir um peso maior para os eventos mais recentes, assim, permitindo uma classificação que aumenta a importância dos eventos mais próximos temporalmente do analista ou investidor.

Pode-se ainda utilizar modelos preditivos para séries temporais como ARIMA e SARIMA, oferecendo uma informação estatística sobre um possível futuro. Inclusive durante a produção deste estudo de caso, foram realizadas experimentos com ambos os modelos, sendo que o segundo obteve melhor resultado, mas, ainda carecendo refinamento.

Outra possibilidade é usar técnicas já conhecidas de *Machine Learning* transformando o modelo. Usando *Logistic Regression*, *Ridge Regression*, *Support Vector Regression*, *Decision Tree Regressor*, *KNN Regressor*, *Random Forest Regressor*, *Light GBM* e *XGBoost* [45]. O presente trabalho dedicou-se a analisar a influência univariada do tempo sobre a variação do preço, com o uso de outras técnicas pode-se avaliar a influência de outras variáveis sobre o preço, como os valores de mínima e máxima para o período, as notícias sobre as empresas no período, usando algoritmos para análise e mineração de texto [22], ou mesmo analisando o comportamento desta variação em função do setor de atuação econômica ou classificação na B3 (small caps, big caps, etc). Ou seja, trata-se de incluir no contexto outras variáveis para entender o comportamento das séries e eventualmente prever os movimentos para um curto intervalo de tempo a frente.

Enfim, existem muitas possibilidades para investigar a sazonalidade dos preços na bolsa de valores, este trabalho em algum momento testou alguns modelos e, aqui apresentou um espectro do resultado que mostrou-se importante para o conhecimento das séries em si, além de apontar horizontes para a continuidade e avanços do estudo. Não foi, nem é meta deste estudo esvaziar as possibilidades do tema, apenas um 'pontapé' inicial para avançar e aprofundar em estudos futuros.

## VI. CONCLUSÃO

Este estudo de caso mostrou-se desafiador, do início ao fim e, foi justamente o uso das técnicas do KDD e mineração de dados que permitiram avançar. Começou-se investigando a flutuação diária dos preços, o valor de fechamento de um dia em relação ao de outro dia, buscou-se encontrar um ponto do meio para analisar, optando-se pela média entre mínimas e máximas mensais. Enfim, foram inúmeras tentativas de interpretação dos dados, testes e retestes, inúmeros algoritmos escritos e posteriormente desprezados, até chegar ao modelo aqui construído e relatado. Com este é possível pode-se dizer que o movimento dos preços de ações na bolsa de valores sofre alguma influência do tempo (ainda que pequena), por isso, tem algum componente sazonal. Outrossim, é errado e imprudente afirmar que a sazonalidade é o único componente nessa composição dinâmica dos preços, existem muitas outras variáveis que pesam sobre o resultado final, em todo caso, a sazonalidade participa dessa composição com maior ou menor

intensidade a depender da empresa e do período em que observam-se as negociações.

Mais uma vez, não é meta deste trabalho oferecer uma fórmula para o enriquecimento de quem quer que seja. Todavia, os resultados percebidos apresentam dados que aliados a outras técnicas analíticas do mercado financeiro como análise fundamentalista e análise técnica podem auferir resultados satisfatórios para algumas negociações.

Apesar do presente estudo ter se centrado na observação dos dados a partir de um ciclo mensal das negociações na bolsa de valores. Boa parte dos algoritmos desenvolvidos, foram escritos pensando outros ciclos temporais, com isso, para trabalhos futuros tem-se por meta realizar o mesmo método aqui aplicado para dados com outras frequências, como diária ou mesmo semanal. Tais frequências por sua dinâmica tem a possibilidade de oferecer muito mais informações que os dados obtidos mensalmente, podendo apresentar resultados diferentes, com vieses de confirmação mais fortes ou simplesmente negativos.

Enfim, ainda há muito por se fazer para conhecer melhor e mais intensamente este componente sazonal nos preços das ações na bolsa de valores, inclusive eventualmente validando o método em outros tipos de ativos como índices, que são contratos com vencimento, que com isso podem sofrer algum ajuste nos dias mais próximos que antecedem o vencimento. Além disso, não foi utilizado o mesmo método para avaliar o comportamento dos preços em outras bolsas de valores do mundo ou mesmo da negociação de cripto-moedas, que tem uma característica especial: são negociadas 24 horas por dia, todos os dias, não há interrupção das negociações em função de feriados e, também não há órgãos reguladores como o Banco Central, ou a CVM. O importante é que com esse estudo de caso, crivou-se um marco inicial para outros estudos, com a possibilidade de ampliar o conhecimento sobre a sazonalidade de preços no mercado financeiro, em especial, o brasileiro.

## REFERÊNCIAS

- [1] “Porcentagem de investidores pessoa física cresce na b3.” [Online]. Available: [http://www.b3.com.br/pt\\_br/noticias/porcentagem-de-investidores-pessoa-fisica-cresce-na-b3.htm](http://www.b3.com.br/pt_br/noticias/porcentagem-de-investidores-pessoa-fisica-cresce-na-b3.htm)
- [2] J. L. Pinheiro, *Mercado de capitais*, 9th ed. Atlas, 2019.
- [3] S. W. Silva, J. E. Gonçalves, D. V. R. Souza, W. F. Pereira, and L. R. d. Fonseca, “O sistema financeiro nacional brasileiro: contexto, estrutura e evolução,” *Revista da Universidade Vale do Rio Verde*, 2016.
- [4] O. Brito, *Mercado financeiro*. Saraiva Educação SA, 2019.
- [5] E. Rodrigues. (2021) Pix já disputa com dinheiro e cartão de débito como principal forma de pagamento.
- [6] C. de Valores Mobiliários, *O mercado de valores mobiliários brasileiro*. Comissão de Valores Mobiliários, 2014.
- [7] A. Elder, *Aprenda a operar no mercado de ações*. Elsevier Brasil, 2006.
- [8] R. M. R. Neto and R. Famá, “Uma alternativa de crescimento para o mercado de capitais brasileiro o novo mercado,” *Revista de Administração*, vol. 37, no. 1, pp. 29–38, 2002, <http://rausp.usp.br/wp-content/uploads/files/v37n1p29a38.pdf>.
- [9] R. G. Walter, “Análise fundamentalista e avaliação de títulos: aspectos teóricos,” *Revista de Administração de Empresas*, vol. 14, pp. 15–32, 1974.
- [10] C. A. Debastiani and F. A. Russo, *Avaliando empresas, investindo em ações: a aplicação prática da análise fundamentalista na avaliação de empresas*. Novatec Editora, 2008.
- [11] J. Kobori, *Análise Fundamentalista: Como obter uma performance superior e constante no mercado de ações*. Alta Books Editora, 2019.
- [12] “Crise de semicondutores alimentada pela covid-19 desgasta montadoras,” <https://www.terra.com.br/noticias/tecnologia/crise-de-semicondutores-alimentada-pela-covid-19-desgasta-montadoras,7d88f2f0abd18531f2b66ec5f786720153bigx6o.html>, acessado em 11/09/2021.
- [13] “Merkel pede investimentos nas energias renováveis em África,” <https://www.dw.com/pt-002/merkel-pede-investimentos-nas-energias-renov%C3%A1veis-em-%C3%A1frica/a-59009224>, acessado em 11/09/2021.
- [14] “Minas tem investimentos previstos de R\$ 52 bi para energias renováveis,” [https://www.em.com.br/app/columnistas/marta-vieira/2021/09/03/interna\\_marta\\_vieira,1302119/minastem-investimentosprevistosder52biparaenergiasrenovaveis.shtml](https://www.em.com.br/app/columnistas/marta-vieira/2021/09/03/interna_marta_vieira,1302119/minastem-investimentosprevistosder52biparaenergiasrenovaveis.shtml), acessado em 11/09/2021.
- [15] M. R. da Rosa and R. Bered, “A importância da análise fundamentalista para avaliar o preço das ações de companhias listadas na bolsa de valores (b3),” *Revista Eletrônica de Ciências Contábeis*, vol. 7, no. 2, pp. 124–150, 2018, <http://seer.faccat.br/index.php/contabeis/article/view/749/564>.
- [16] C. A. Debastiani, *Análise Técnica de Ações: identificando oportunidades de compra e venda*. Novatec Editora, 2008.
- [17] F. A. C. D. A. LEMOS, *Análise técnica dos mercados financeiros*. Saraiva Educação SA, 2017.
- [18] U. Fayyad, G. Piattetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” *AI magazine*, vol. 17, no. 3, pp. 37–37, 1996.
- [19] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [20] D. L. Olson and D. Delen, *Advanced data mining techniques*. Springer Science & Business Media, 2008.
- [21] R. Goldschmidt, E. Passos, and E. Bezerra, *Data Mining*. Elsevier Brasil, 2015.
- [22] T. Fawcett and F. Provost, *Data Science para Negócios: O que você precisa saber sobre mineração de dados e pensamento analítico de dados*. Alta Books Editora, 2018.
- [23] D. T. Larose and C. D. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014, vol. 4.
- [24] C. O. Camilo and J. C. d. Silva, “Mineração de dados: Conceitos, tarefas, métodos e ferramentas,” *Universidade Federal de Goiás (UFG)*, vol. 1, no. 1, pp. 1–29, 2009.
- [25] J. L. D. Bastos and R. P. Duquia, “Tipos de dados e formas de apresentação na pesquisa clínico-epidemiológica,” *Scientia medica*, vol. 16, no. 3, pp. 133–138, 2006.
- [26] J. Grus, *Data science do zero*. Alta Books Editora, 2018.
- [27] J. Ferreira, M. Miranda, A. Abelha, and J. Machado, “O processo etl em sistemas data warehouse,” in *INForum*, 2010, pp. 757–765.
- [28] P. Vassiliadis, A. Simitsis, and S. Skiadopoulos, “Conceptual modeling for etl processes,” in *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*, 2002, pp. 14–21.
- [29] R. S. Ehlers, “Análise de séries temporais,” *Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná*, vol. 1, pp. 1–118, 2007.
- [30] M. I. S. Bezerra, “Apostila de análise de séries temporais,” *UNESP: Curso de estatística. São Paulo, SP*, 2006.
- [31] T. O. Favari, “Análise de séries temporais: comparação entre modelos preditivos em estudo de caso,” *UFRGS: Departamento de Engenharia Química. Porto Alegre-RS*, 2018.
- [32] E. D. Torres, J. M. G. da Silva, M. M. Provenza, I. C. de Almeida Lima, and J. L. de Jesus Goulart, “Utilização dos modelos arima para previsão da taxa de churn: Estudo de caso para uma empresa de e-commerce,” *Cadernos do IME-Série Estatística*, vol. 48, p. 36, 2020.
- [33] W. M. Lamounier, “Tendência, ciclos e sazonalidade nos preços spot do café brasileiro na nybot,” *Gestão & Produção*, vol. 14, pp. 13–23, 2007.
- [34] T. R. Pellegrini, “Uma avaliação de métodos de previsão aplicados à grandes quantidades de séries temporais univariadas,” *Universidade Federal de São Carlos*, 2012.

- [35] B. M. COTES and A. F. C. do CARMO, “Análise histórica do índice bovespa: Previsões e tomada de decisão,” *ETIC-ENCONTRO DE INICIAÇÃO CIENTÍFICA-ISSN 21-76-8498*, vol. 15, no. 15, 2019.
- [36] P. A. Morettin and C. M. Tolo, *Análise de séries temporais: modelos lineares univariados*. Editora Blucher, 2018.
- [37] A. Plummer, “Different Types of Time Series Decomposition,” Sep. 2020. [Online]. Available: <https://towardsdatascience.com/different-types-of-time-series-decomposition-396c09f92693>
- [38] D. R. Gardner, “STL Algorithm Explained: STL Part II.” [Online]. Available: <http://www.gardner.fyi/blog/STL-Part-II/>
- [39] M. Fayh, *Método Fayh: Descubra Como Escolher os Melhores Fundos Imobiliários do Mercado e Viva de Renda*. Marcelo Fayh Paulitsch, 2020.
- [40] J. A. V. Ferreira, “Introducao a Analise de Dados - Python e pandas | Kaggle.” [Online]. Available: <https://www.kaggle.com/joaoavf/introducao-a-analise-de-dados-python-e-pandas>
- [41] W. McKinney, *Python para análise de dados: Tratamento de dados com Pandas, NumPy e IPython*. Novatec Editora, 2019.
- [42] “As ações mais negociadas na b3 (brasil bolsa balcão),” Nov. 2017. [Online]. Available: <https://www.iq.com.br/investimentos/artigos/acoes-mais-negociadas-na-b3>
- [43] S. Dana and D. D. N. e. J. Ribeiro, “Análise completa das TOP 5 ações mais negociadas da bolsa.” [Online]. Available: <https://investnews.com.br/cafeina/as-acoes-mais-negociadas-sao-as-que-rentabilizam-melhor/>
- [44] C. N. Knaflic, *Storytelling com Dados: Um guia sobre visualização de dados para profissionais de negócios*. Alta Books, 2019.
- [45] C. E. Souza, “Séries temporais com machine learning — parte 4,” Feb. 2020. [Online]. Available: <https://medium.com/data-hackers/s%C3%A9ries-temporais-com-machine-learning-parte-4-6beb3efb2e99>