

Análise da Correlação Entre a Divulgação do EBITDA e Outros Indicadores na Evolução da Cotação de Ações de Empresas Listadas na B3

Silvio Ferraz dos Santos

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Análise da Correlação Entre a
Divulgação do EBITDA e Outros
Indicadores na Evolução da Cotação
de Ações de Empresas Listadas na
B3

Silvio Ferraz dos Santos

USP - São Carlos

2024

Silvio Ferraz dos Santos

Título do trabalho: Análise da Correlação Entre a Divulgação do EBITDA e Outros Indicadores na Evolução da Cotação de Ações de Empresas Listadas na B3

Trabalho de conclusão de curso apresentado ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial

Orientador: Prof. Dr. Zhao Liang

USP - São Carlos

2024

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

S237a Santos, Silvio Ferraz dos
Análise da Correlação Entre a Divulgação do EBITDA e Outros Indicadores na Evolução da Cotação de Ações de Empresas Listadas na B3 / Silvio Ferraz dos Santos; orientador Prof. Dr. Zhao Liang. -- São Carlos, 2024.
56 p.

Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2024.

1. EBITDA. 2. Ações. 3. Big Data. 4. Correlação. 5. Extração de Dados. I. Liang, Prof. Dr. Zhao, orient. II. Título.

DEDICATÓRIA

À minha esposa Renata e às minhas filhas Marina e Gabriela pela compreensão sobre o tempo que não pude dedicar a elas no decorrer desse MBA e pelo carinho e apoio nos momentos de dificuldade e cansaço. Amo vocês!!!

AGRADECIMENTOS

À Profa. Dra. Solange Oliveria Rezende, pelos ensinamentos, incentivo, apoio, parceria e principalmente compreensão em relação às dificuldades que tive em conciliar as muitas atividades com o desenvolvimento deste MBA.

Ao Prof. Dr. Zhao Liang por, com todos os meus atrasos, continuar acreditando que esse trabalho poderia ser concluído com qualidade.

RESUMO

SANTOS, S.F. **Análise da Correlação Entre a Divulgação do EBITDA e Outros Indicadores na Evolução da Cotação de Ações de Empresas Listadas na B3.** 2024. 68 p. Trabalho de conclusão de curso (MBA em Inteligência Artificial e Big Data) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2024.

Esse trabalho analisa se existe uma correlação entre a evolução da cotação de ações de empresas listadas na B3 e a publicação do balanço patrimonial delas, momento em que diversas informações e indicadores importantes, como o EBITDA, o Lucro Líquido, a Receita Líquida e o Patrimônio Líquido são divulgadas, com o objetivo de identificar uma possível influência da divulgação destes indicadores no comportamento do mercado de ações. Em um mercado eficiente, o valor de negociação das ações deve refletir as informações conhecidas sobre a Sociedade Anônima que as emitiu. Dividimos o desenvolvimento deste trabalho em cinco etapas (Identificação do Problema, Pré-Processamento, Extração de Padrões, Pós-Processamento e Utilização do Conhecimento). Os dados usados no trabalho foram os históricos de cotações de ações na B3 no período de um ano e os balanços patrimoniais trimestrais de dezenove empresas, com volume expressivo de negociações diárias, representando diversos setores de atividade, no período de dois trimestres. O banco de dados relacional POSTGRESQL foi usado para armazenar os dados obtidos e notebooks em PHYTON foram criados para o seu processamento. A partir desses dados, foi calculada a diferença entre a média das cinco cotações médias anteriores e posteriores à data de divulgação do balanço de cada empresa, evidenciando a evolução das cotações a partir desse evento. Como os dados possuem escala de valores muito diferentes, foram normalizados com a função MaxAbsScaler, que mantém as características dos dados originais, para permitir a busca por uma correlação entre eles. Verificamos que os dados não possuem uma distribuição normal. O coeficiente de correlação de postos de Spearman, mais eficiente quando a distribuição dos dados não segue uma normal, foi usado para calcular a correlação entre os indicadores selecionados do balanço das empresas e a evolução das cotações após a publicação deles, concluindo que não existe uma correlação significativa entre eles. A importância deste trabalho é demonstrar ao mercado que o EBITDA e outros indicadores encontrados no Balanço Patrimonial das empresas não deve ser usado como referência para decidir sobre a compra ou venda de ações no mercado de capitais.

Palavras-chave: EBITDA; Ações; Big Data; Correlação; Extração de dados

ABSTRACT

SANTOS, S.F. **Analysis of the Correlation Between the Disclosure of EBITDA and Other Indicators in the Evolution of the Price of Shares of Companies Listed on the B3.** 2024. 68 p. Final paper (MBA in Artificial Intelligence and Big Data) – Institute of Mathematical and Computer Sciences, University of São Paulo, São Carlos, 2024.

This paper analyzes whether there is a correlation between the evolution of the stock price of companies listed on B3 and the publication of their balance sheets, a time when several important information and indicators, such as EBITDA, Net Income, Net Revenue and Net Equity are disclosed, with the aim of identifying a possible influence of the disclosure of these indicators on the behavior of the stock market. In an efficient market, the trading value of shares should reflect the known information about the corporation that issued them. We divided the development of this paper into five stages (Problem Identification, Pre-Processing, Pattern Extraction, Post-Processing and Use of Knowledge). The data used in the paper were the historical stock prices on B3 over a period of one year and the quarterly balance sheets of nineteen companies, with significant daily trading volume, representing various sectors of activity, over a period of two quarters. The POSTGRESQL relational database was used to store the data obtained and notebooks in PHYTON were created for their processing. From these data, the difference between the average of the five average prices before and after the date of publication of each company's balance sheet was calculated, highlighting the evolution of the prices after this event. Since the data have very different scales of values, they were normalized with the MaxAbsScaler function, which maintains the characteristics of the original data, to allow the search for a correlation between them. We found that the data do not have a normal distribution. The Spearman rank correlation coefficient, more efficient when the distribution of the data does not follow a normal pattern, was used to calculate the correlation between the selected indicators of the companies' balance sheets and the evolution of the prices after their publication, concluding that there is no significant correlation between them. The importance of this work is to demonstrate to the market that EBITDA and other indicators found in the companies' Balance Sheets should not be used as a reference for deciding on the purchase or sale of shares in the capital market.

Keywords: EBITDA; Shares; Big Data; Correlation; Data extraction

LISTA DE ILUSTRAÇÕES

Figura 1 – Etapas do processo de extração de conhecimento de uma base de dados	25
Figura 2 – Diagrama de dispersão do atributo ebitda	33
Figura 3 – Diagrama de dispersão do atributo lcro_lqdo	34
Figura 4 – Diagrama de dispersão do atributo rcta_lqda	34
Figura 5 – Diagrama de dispersão do atributo prtmo_lqdo	34
Figura 6 – Diagrama de dispersão do atributo diferença	35
Figura 7 – Mapa de Correlação dos atributos usando o coeficiente de correlação linear de PEARSON	36
Figura 8 – Mapa de Correlação dos atributos usando o coeficiente de correlação pontual de SPEARMAN	37

LISTA DE TABELAS

Tabela 1 – Empresas selecionadas para o estudo e o seu principal setor de atuação	27
Tabela 2 – Atributos da tabela cotacoes	28
Tabela 3 – Atributos da tabela balancos	29
Tabela 4 – Tuplas recuperadas da tabela balancos após a inclusão dos dados	30
Tabela 5 – Atributos da tabela dadosanalise	31

SUMÁRIO

1 INTRODUÇÃO.....	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 Ações	21
2.2 O EBITDA	23
2.3 Estudos Correlatos	24
3 MATERIAIS E MÉTODOS	25
3.1 Método	25
3.1.1 Identificação do problema	25
3.1.2 Pré-processamento	25
3.1.2.1 Aquisição dos dados	25
3.1.2.1.1 Histórico de cotações de ações na B3	26
3.1.2.1.2 Balanços Patrimoniais	26
3.1.2.2 Carga, tratamento e limpeza dos dados	27
3.1.2.2.1 Cotações	27
3.1.2.2.1.1 EXCEL	27
3.1.2.2.1.2 POSTGRES	28
3.1.2.2.2 Balanços	29
3.1.2.3 Transformação dos dados	30
3.1.3 Extração de Padrões	31
3.1.4 Pós-Processamento	37
4 CONCLUSÃO	38
REFERÊNCIAS	40
Apêndice A – 0-TCC Pré-Processamento - Carga e Limpeza na Base	41
Apêndice B – 1-TCC Pré-Processamento – Transformação	45
Apêndice C – 2-TCC Extração de Padrões	49
ANEXO A – Layout do Histórico de Cotações de Ações na B3	53

1 INTRODUÇÃO

Segundo Assaf Neto (2008), as sociedades anônimas contam com duas formas de financiamento com recursos próprios. O autofinanciamento e o lançamento de ações no mercado de capitais. Nesse último caso, a empresa precisa estar registrada junto à Comissão de Valores Mobiliários (CVM) como “Companhia Aberta”.

Ações são a menor fração do capital social de uma sociedade anônima e o seu proprietário, o acionista, é sócio e tem participação em seus resultados, positivos ou negativos. São representadas por títulos físicos (cautelas ou certificados) ou do tipo escritural, que é o registro de quem é o proprietário da ação, mantido em uma instituição depositária.

A colocação de novas ações no mercado é realizada no mercado primário (onde o emitente negocia as ações diretamente com o público com a intermediação de uma instituição financeira. Na distribuição primária acontece entrada de recursos no capital da empresa. Os acionistas possuem direitos junto à empresa que emitiu as ações. O principal deles é o dividendo, distribuição de lucro aos acionistas.

Após a emissão das novas ações no mercado primário, elas passam a ser negociadas no mercado secundário diretamente entre investidores em operações de compra e venda que acontecem ou no mercado de balcão ou, principalmente, em Bolsas de Valores, como a B3, em um espaço e horário destinado para essa negociação chamado pregão. No pregão instituições como Corretoras e Distribuidoras de Valores Mobiliários representam os compradores e vendedores, mediante a cobrança de uma corretagem, sendo responsáveis pelo fechamento dos negócios. No mercado secundário os valores são trocados entre compradores e vendedores, não tendo participação da empresa emitente das ações.

A oferta e a procura por cada ação na bolsa de valores define o seu preço, ou seja, os investidores comprando e vendendo influenciam o seu preço aproximando-o de seu valor justo. No entanto, a situação da empresa no momento e a esperada no futuro, indicadores econômicos e políticos e outros fatores também influenciam no preço de mercado das ações, aumentando ou diminuindo o interesse dos investidores em mantê-la como ativo. Desta forma os preços das ações no mercado de valores mobiliários são influenciados por novas informações divulgadas sobre a conjuntura econômica e do mercado e sobre a própria empresa emissora das ações. Isso é uma característica de um mercado dito eficiente. A sensibilidade dos preços de mercado faz com que qualquer projeção sobre o comportamento futuro das ações de uma determinada companhia deve incorporar informações novas e não se basear apenas em informações

divulgadas no passado. Em contrapartida, quando os preços de uma ação não refletem as informações disponíveis no mercado sobre a empresa emissora, acontece uma situação conhecida como “mercado ineficiente”, e isso também é possível no mercado de ações.

Conforme Coelho* (2005 apud Sousa e Sales, 2021, p.2), o Lucro Antes dos Juros, Impostos, Depreciação e Amortização (LAJIDA) ou Earnings Before Interest, Taxes, Depreciation and Amortization (EBITDA), como é mais conhecido e divulgado, é um importante indicador da eficiência da empresa que mede a sua capacidade de gerar caixa e não a geração de caixa. O seu cálculo é feito com os seguintes componentes do balanço patrimonial: ao resultado líquido do período apurado, deve-se acrescentar as despesas financeiras líquidas, os tributos calculados sobre o lucro, as receitas financeiras e as depreciações, amortizações e exaustões. Sua divulgação é opcional, porém se for feita, o seu cálculo é regulamentado pela Instrução nº 527 da Comissão de Valores Mobiliários (CVM) (2012).

Padoveze (2020), descreve os riscos que as empresas incorrem ao atribuir uma importância exagerada ao EBITDA, como usá-lo para calcular a remuneração variável dos gestores, já que distorções são possíveis, sendo a mais grave a empresa apresentar EBITDA positivo e apresentar prejuízo líquido.

O objetivo deste trabalho é analisar se há uma correlação efetiva e significativa entre a divulgação do EBITDA pelas empresas e uma variação posterior positiva ou negativa no valor da cotação das ações em relação ao valor médio praticado no mercado em um período anterior. Serão avaliados outros indicadores do balanço financeiro da empresa com o mesmo intuito, buscando a mesma correlação com a variação da cotação das ações.

A contribuição deste estudo está em disponibilizar informações para a tomada de decisão de gestores do mercado de capitais apresentando o resultado das análises desenvolvidas.

* Coelho, F. **EBITDA**: a busca de uma melhor compreensão do maior vox populi do mercado financeiro. Rio de Janeiro: *Revista Pensar Contábil*, 6(26), 2005

2 REVISÃO BIBLIOGRÁFICA

2.1 Ações

Há três formas de uma empresa se capitalizar, conforme Assaf Neto (2018). O primeiro é receber um aporte direto de capital de seus proprietários. O segundo, buscar o sistema financeiro para buscar empréstimos, que no momento pactuado, serão pagos com o acréscimo de juros. O terceiro é abrir o capital da empresa, emitindo títulos e recebendo aporte de capital diretamente dos investidores, que passam a ser sócios da empresa.

O mercado de capitais é dividido em mercado primário e mercado secundário. O mercado primário é responsável pela disponibilização dos títulos emitidos no momento da abertura de capital da empresa. Para que isso seja possível é necessário o seu registro na Comissão de Valores Mobiliários – CVM, como “Companhia Aberta”, a sua transformação em sociedade anônima – S.A. e a intermediação de uma instituição financeira que será responsável pela oferta pública dos títulos. Neste trabalho nos concentraremos em um tipo específico de títulos que são as ações. A primeira oferta no mercado primário de ações de uma companhia, recebe o nome de IPO – Inicial Public Offering em inglês, ou Oferta Pública Inicial de Ações.

Após a abertura do capital no mercado primário, onde o comprador da ação faz diretamente um aporte de capital à empresa, as ações passam a ser negociadas no mercado secundário, em uma Bolsa de Valores, onde o comprador e o vendedor, com a intermediação de uma Corretora de Valores, realizam a troca de titularidade das ações e a liquidação financeira da transação. No mercado secundário a empresa não participa da transação.

Ação é a menor fração do capital social de uma sociedade anônima. O seu proprietário, o acionista, é sócio da empresa e tem participação em seus resultados, tanto positivos quanto negativos. As ações podem ser representadas por títulos físicos (cautelas ou certificados) ou do tipo escritural, que é o registro de quem é o proprietário da ação, mantido em uma instituição depositária. Atualmente o padrão usual é o tipo escritural, com raras exceções. Também podem ser ordinárias ou preferenciais. A principal diferença entre os dois tipos é que as ações ordinárias dão direito a voto nas assembleias de acionistas proporcional à participação da quantidade de ações possuídas em relação à quantidade total emitida. Já as ações preferenciais dão preferência no recebimento dos dividendos distribuídos quando da apuração do lucro líquido no balanço da empresa.

O retorno do investimento em ações se dá pela distribuição do lucro da empresa aos acionistas através da distribuição de dividendos (pagamento em dinheiro) ou bonificação (pagamento em novas ações) ou na sua comercialização em bolsa de valores.

Segundo Arceno (2019), no ano 2000 um acordo estabeleceu a Bolsa de Valores de São Paulo - BOVESPA como a única bolsa de valores a negociar ações no Brasil. Já em 2008 a Bolsa de Mercadorias e Futuros – BM&F e a BOVESPA sofreram uma fusão que passou a ser denominada BM&F/BOVESPA. Em 2017, a BM&F/BOVESPA e a Central de Custódia e de Liquidação Financeira de Títulos - CETIP se fundiram gerando a B3, uma sociedade de capital aberto, que é uma das maiores bolsas no mercado de capitais do mundo.

As operações de compra e venda de ações acontecem de forma eletrônica e acontecem em um ambiente denominado “Mega Bolsa”. A liquidação das operações é efetuada fisicamente (transferência da titularidade das ações) em d+2 dias úteis e a liquidação financeira acontece em d+3 dias úteis.

Não existe um prazo de resgate fixo para uma ação. A forma de converter o investimento em dinheiro é a sua negociação em Bolsa de Valores e o seu valor de mercado é definido como o valor de equilíbrio resultante da oferta e da procura pelo título durante o pregão. A ação é um investimento de renda variável, sendo que o resultado obtido pode ser um ganho ou perda em relação ao valor investido e depende das condições do mercado quando a negociação acontece, do desempenho da empresa que emitiu a ação e da condição da economia como um todo.

Para Assaf Neto (2018):

Os preços livremente praticados pelos agentes participantes são normalmente bastante sensíveis às novas informações introduzidas pelo ambiente conjuntural, pelo comportamento do mercado e da própria empresa emitente do título, determinando ajustes rápidos em seus valores. Toda nova informação relevante trazida ao mercado tido como eficiente tem o poder de promover alterações nos valores dos ativos negociados, modificando seus livres preços de negociação e resultados de análises. Esse comportamento extremamente sensível dos preços de mercado ensina que projeções que venham a ser efetuadas com relação a determinado cenário futuro devem sempre incorporar novas informações esperadas, e não somente se basear em dados verificados no passado. Um mercado eficiente é entendido como sendo aquele em que os preços refletem as informações disponíveis e apresentam grande sensibilidade a novos dados, ajustando-se rapidamente a outros ambientes.

Isso não significa que em um mercado eficiente os preços sempre estarão ajustados a um valor ideal. Há períodos em que o retorno é anormal pois as informações não estão disponíveis de forma homogênea a todos, ou acontecem indecisões dos investidores e o preço das ações se distanciam do seu valor de mercado justo. Estes retornos anormais não se mantêm como padrão, e como regra, os valores praticados pelo mercado permanecem próximos ao valor justo que reflete as informações disponíveis.

2.2 O EBITDA

Para Sousa e Sales (2021), as empresas têm a necessidade de verificar o quanto são eficientes como organização para poder mensurar o quanto as estratégias adotadas estão corretas. Um dos indicadores escolhidos é o Lucro Antes dos Juros, Impostos, Depreciação e Amortização (LAJIDA) ou Earnings Before Interest, Taxes, Depreciation and Amortization – EBITDA – em inglês, como é mais conhecido. Este indicador é formado a partir de informações operacionais e não financeiras ou econômicas, e tem como finalidade formar métrica sobre a capacidade das empresas gerarem caixa e a sua eficiência operacional.

Segundo Verriest et al (2018), o EBITDA é um índice de desempenho que incorpora propósitos como a avaliação da performance corporativa, eficiência gerencial e medição de solvência. Na década de 1980 se popularizou o uso do EBITDA com aquisições alavancadas, sendo usado para demonstrar a capacidade de uma companhia honrar suas dívidas. Depois disso o EBITDA passou a ser usado na avaliação interna das empresas e como medida de desempenho alternativa ao lucro líquido, lucro operacional, fluxo de caixa operacional e fluxo de caixa livre na comunicação com os acionistas. O EBITDA é calculado e divulgado voluntariamente, pois não faz parte do balanço e não é auditado nem padronizado. Devido a diversos problemas envolvendo cálculos indevidos por parte das empresas e erros em sua divulgação, a CVM editou a Instrução nº 527 (2012), regulamentando o seu cálculo para as empresas registradas nessa instituição.

Padoveze (2020), após constatar que o EBITDA continua a ser usado na avaliação de gestores divisionais e que é usado para o cálculo da remuneração variável dos gestores da empresa, aponta para distorções geradas pelo uso do EBITDA, sendo as principais:

- A margem EBITDA calculada como o percentual do valor do EBITDA em relação ao valor da Receita Operacional Líquida serve apenas precariamente para avaliar internamente na empresa uma evolução em relação aos anos anteriores. Não serve para comparar com a margem EBITDA de outras empresas pois cada uma tem diferente estrutura de ativos operacionais.
- Não demonstra claramente a dimensão do fluxo de caixa, já que desconsidera as despesas financeiras, o Imposto de Renda – IR - e a Contribuição Social sobre o Lucro Líquido – CSLL – que impactarão o caixa.
- Como o EBITDA não considera depreciações, não faz a mensuração da reserva financeira necessária para a reposição dos investimentos.
- Não considera o legado do passado, como passivo ambiental e parcelamento de tributos.

- O interesse dos acionistas minoritários está na distribuição de dividendos que só são conhecidos após a verificação do lucro líquido.
- Como o EBITDA despreza um enorme conjunto de despesas operacionais, financeiras e o IR/CSLL, não pode ser usado de forma correta para avaliar os gestores da empresa.
- Pode gerar indisciplina financeira por parte dos gestores, que tem sua remuneração variável ancorada neste índice e podem ser estimulados a tomar decisões de curto prazo para induzir um resultado favorável.
- Não serve para comparar empresas de qualquer país, mesmo dentro do mesmo segmento de atividades, pois cada país tem ambiente econômico particular.
- Por fim, a mais importante, a empresa pode divulgar um EBITDA positivo e na verdade estar em prejuízo e ainda pode não ter caixa, o que torna o uso deste indicador como gerador de caixa futuro muito perigoso.

2.3 Estudos Correlatos

Sousa e Sales, 2021, em seu “Relação entre o Ebitda e o Retorno das Ações das Empresas Listadas na B3” desenvolveram um estudo com o objetivo de responder se “A divulgação do EBITDA das empresas listadas na B3 gera impactos na cotação de suas ações dentro do curto prazo?”. Trabalhou com as cotações de 90 empresas comparando as cotações de 2 dias antes e 2 dias depois da divulgação do EBITDA pelas empresas para buscar uma correlação entre a evolução das cotações e o EBITDA divulgado.

3 MATERIAIS E MÉTODOS

3.1 Método

De acordo com Rezende (2003), o processo de extração de conhecimento de uma base de dados pode ser dividido em diversas etapas. Três se referem diretamente ao processo da Data Mining, que são Pré-Processamento, Extração de Padrões e Pós-Processamento. Duas outras foram incluídas, sendo uma fase anterior ao processo de Data Mining, Identificação do Problema, e outra posterior que é a Utilização do Conhecimento obtido nas etapas anteriores.

Figura 1 – Etapas do processo de extração de conhecimento de uma base de dados

ra



Fonte: Rezende, S. O. *Sistemas inteligentes: fundamentos e aplicações*. Editora Manole Ltda, 2003.

3.1.1 Identificação do problema

O objetivo deste estudo é buscar uma relação entre a divulgação do EBITDA e de outros indicadores do Balanço Patrimonial de empresas listadas na B3, sobre a evolução da cotação de suas ações no Mercado à Vista, após a data de sua publicação. Para isso selecionamos as cotações das ações no período de 5 dias anteriores e posteriores à data de publicação do Balanço Patrimonial de cada empresa.

3.1.2 Pré-processamento

3.1.2.1 Aquisição dos dados

Neste trabalho, incluímos na etapa de Pré-Processamento a aquisição dos dados necessários para o estudo, um histórico de cotações de ações na B3 cobrindo o período de um ano (setembro de 2023 a agosto de 2024) e os balanços patrimoniais das empresas. Para o armazenamento destes dados escolhemos o Banco de Dados Relacional POSTGRESQL, por possuir grande capacidade de armazenamento e a facilidade do uso de comandos SQL para a criação e manipulação dos dados, além de ser um BD de distribuição gratuita.

3.1.2.1.1 Histórico de cotações de ações na B3

O histórico de cotações de ações usado neste estudo está disponível no sítio de internet da B3 para download no link:

https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/cotacoes-historicas/

Fizemos o download das cotações mês a mês do período de um ano, iniciando em setembro de 2023, até agosto de 2024. Optamos pelo download dos arquivos mensais por conta do volume de dados no arquivo anual, caso optássemos pelo download do mesmo período estudado em um só arquivo.

Os dados estavam armazenados em arquivos com o formato sequencial com tamanho fixo para os atributos. Layout no ANEXO A.

3.1.2.1.2 Balanços Patrimoniais

Para este estudo selecionamos dezenove empresas com volume expressivo de negociações na B3, para calcular a correlação entre os indicadores do Balanço Patrimonial publicado e a evolução da cotação em um período anterior e posterior à publicação dele. Procuramos incluir no estudo empresas de diversos setores de atuação, como energia elétrica; bebidas; companhias aéreas; comércio varejista; produtoras e distribuidoras de combustíveis; indústria aeronáutica; produção de aço; processamento de carnes; papel e celulose; comércio de roupas; mineração e siderurgia, para que os diversos setores da economia estivessem representados. Abaixo (Tabela 1) listamos as empresas selecionadas e os respectivos setores de atuação:

Tabela 1 – Empresas selecionadas para o estudo e o seu principal setor de atuação

EMPRESA	PRINCIPAL SETOR DE ATUAÇÃO
AES BRASIL	ENERGIA ELÉTRICA
AMBEV S/A	BEBIDAS
AZUL	COMPANHIA AÉREA
CASAS BAHIA	COMÉRCIO VAREJISTA
COSAN	DISTR. COMBUSTÍVEIS E PRODUTORA DE ETANOL
EMBRAER	INDÚSTRIA AERONÁUTICA
GERDAU	PRODUTORA DE AÇO
GOL	COMPANHIA AÉREA
JBS	PROCESSAMENTO DE CARNES
KLABIN S/A	PAPEL E CELULOSE
LOJAS RENNER	COMÉRCIO DE ROUPAS
MAGAZINE LUIZA	COMÉRCIO VAREJISTA
MARFRIG	PROCESSAMENTO DE CARNES
NEOENERGIA	ENERGIA ELÉTRICA
PETROBRAS	INDÚSTRIA DO ÓLEO, GÁS NATURAL E ENERGIA
SUZANO S.A.	PAPEL E CELULOSE
ULTRAPAR	PETROQUÍMICO
VALE	MINERAÇÃO E SIDERURGIA

Os dados dos balanços patrimoniais trimestrais das empresas foram obtidos diretamente em seus sítios de internet. Destes balanços patrimoniais trimestrais recuperamos manualmente as informações de Nome da Empresa, Data da Publicação, EBITDA, Margem EBITDA, Lucro Líquido, Receita Líquida e Patrimônio Líquido. Optamos, para efeito deste estudo, pelos balanços patrimoniais do primeiro trimestre de 2024 e segundo trimestre de 2024. Essas informações alimentaram uma planilha eletrônica EXCEL, do pacote Office da Microsoft, e posteriormente foram exportadas para um arquivo .CSV com a finalidade de alimentar um banco de dados.

3.1.2.2 Carga, tratamento e limpeza dos dados

3.1.2.2.1 Cotações

3.1.2.2.1.1 EXCEL

Os arquivos sequenciais foram importados um a um no EXCEL, planilha de cálculo eletrônica do pacote Office da Microsoft, e convertidos no formato da planilha, com tamanho fixo, conforme o layout disponibilizado pela B3 em seu sítio de internet. Nessa conversão

mantivemos os zeros à esquerda nos campos numéricos e, como neste layout não há “,” separando as casas decimais, os valores se apresentam multiplicados por 100. Os atributos que não tinham relação com o Mercado à Vista foram excluídos, com a finalidade de reduzir o volume de informações na base de dados. Fizemos a conversão do atributo “Data do Pregão”, que originalmente estava no formato DDMMAAAA para um formato de data do Excel – DD/MM/AAAA, para posteriormente facilitar a sua carga em um banco de dados. Por fim, os dados foram exportados, também um a um, do EXCEL como arquivo .CSV tendo “;” como separador entre os atributos, para possibilitar a sua importação em um notebook PYTHON e serem inseridos em um banco de dados.

3.1.2.2.1.2 POSTGRES

O banco de dados escolhido foi o POSTGRESQL, por atender aos requisitos do estudo, por permitir o uso de comandos SQL para a recuperação dos dados, pela facilidade da instalação e por ser um banco de dados com licença livre. Após a instalação do POSTGRESQL criamos um Database TCC.

Para a carga da base de dados no POSTGRESQL foi criado o JUPYTER NOTEBOOK “0-TCC Pré-Processamento – Carga e Limpeza na Base.ipynb”, em PYTHON, que se encontra no APÊNDICE A deste trabalho.

Neste notebook instalamos e importamos as bibliotecas e os módulos necessários.

O próximo passo foi criar uma conexão entre o JUPYTER NOTEBOOK e o Banco de Dados POSTGRESQL.

Em seguida, criamos a tabela “cotacoes” (Tabela 2) e efetuamos a carga dos arquivos mensais do histórico de cotações de ações na B3. Optamos por carregar na tabela somente os arquivos do mês 04/2024 até 08/2024, por ser esse o range das datas de publicação dos balanços patrimoniais trimestrais das empresas selecionadas. Foram inseridas um milhão, cento e dezenove mil, novecentos e cinco tuplas neste processo.

Tabela 2 – Atributos da tabela cotacoes

ATRIBUTO	TIPO DE DADO	DESCRIÇÃO
DATA_PREGAO	DATE,	-- Data do Pregão
CDGO_NGCCO_PPL	TEXT,	-- Código de Negociação do Papel
TIPO_MERC	NUMERIC(5),	-- Tipo de Mercado - 010 - À Vista
NOME_RESMDO_EMPRSA	TEXT,	-- Nome Resumido da Empresa Emissora
ESPCFCCO_PPL	TEXT,	-- Especificação do Papel
MOEDA_RFRNCA	TEXT,	-- Moeda Usada na Data do Pregão

PRECO_ABRTRA	DECIMAL(11,2),	-- Preço Abertura
PRECO_MEDIO	DECIMAL(11,2),	-- Preço Médio
PRECO_ULTMO	DECIMAL(11,2),	-- Preço do Último Negócio Fechado no Pregão
FTR_CTCO_PPL	DECIMAL(7)	-- Fator Cotação do Papel 1=Cotação unitária 1000=Cotação lote mil ações

Como os arquivos de histórico de cotações de ações na B3 contém informações de todos os tipos de mercado, excluimos da tabela todas as linhas com tipo de mercado diferente de 10 - Mercado à Vista, alvo deste trabalho.

No próximo passo excluimos as cotações de todas as empresas que não são alvo deste estudo e verificamos se há dados nulos na tabela “cotacoes”, concluindo que não há.

Como os atributos valores nos arquivos sequenciais de cotações tiveram suas duas casas decimais incorporadas à parte inteira do valor, ou seja, estão multiplicados por 100, atualizamos esses atributos dividindo-os por 100, recuperando a informação original dos valores, com duas casas decimais. Em seguida verificamos se existem dados duplicados na tabela “cotacoes” e não foram encontrados.

Ao término da limpeza de dados de cotações, permaneceram 2697 tuplas na tabela “cotações”.

3.1.2.2.2 Balanços

Para completar a carga dos dados, criamos a tabela “balancos” (Tabela 3) e efetuamos a carga dos dados dos balanços patrimoniais das empresas.

Tabela 3 – Atributos da tabela balanços

ATRIBUTO	TIPO DE DADO	DESCRIÇÃO
NOME_EMPRSA	TEXT,	-- Nome da Empresa
DATA_BALANCO	DATE,	-- Data do Balanco
EBITDA	DECIMAL(11,2),	-- EBITDA
MARGEM	NUMERIC(5),	-- Margem EBITDA
LCRO_LQDO	DECIMAL(11,2),	-- Lucro Líquido
RCTA_LQDA	DECIMAL(11,2),	-- Receita Líquida
PTRMO_LQDO	DECIMAL(11,2),	-- Patrimônio Líquido

Verificamos se foram carregados dados nulos na tabela e não foram encontrados.

Os atributos referentes aos indicadores recuperados dos balanços patrimoniais tiveram a sua casa decimal incorporada à sua parte inteira, ou seja, estão multiplicados por 10. Para

recuperar a informação da casa decimal, dividimos esses atributos por 10. Os valores representados nos indicadores estão em milhões de reais.

Após a carga e tratamento a tabela de “balancos” se apresenta conforme a Tabela 4 abaixo:

Tabela 4 – Tuplas recuperadas da tabela balanços após a inclusão dos dados

nome_empresa	data_balanco	ebitda	margem	lcro_lqdo	rccta_lqda	ptrmo_lqdo
NEOENERGIA	2024-04-23	3507.00	31	1127.00	11020.00	31181.00
VALE	2024-04-24	17155.60	41	8418.10	42210.40	194674.80
KLABIN S/A	2024-04-24	1652.00	37	460.00	4429.50	13546.10
GERDAU	2024-05-02	2813.00	17	1245.00	16210.00	52110.90
AES BRASIL	2024-05-03	340.20	41	-102.40	828.60	5458.80
LOJAS RENNER	2024-05-06	377.90	15	139.30	2461.30	10099.80
EMBRAER	2024-05-07	233.70	5	161.40	4448.30	15199.70
CASAS BAHIA	2024-05-08	387.00	6	-261.00	6347.00	3202.00
AMBEV S/A	2024-05-08	6534.80	32	3804.20	20276.30	87376.70
SUZANO S.A.	2024-05-09	4534.20	48	220.00	9458.60	44724.50
ULTRAPAR	2024-05-09	1357.70	4	455.40	30395.90	14375.70
MAGAZ LUIZA	2024-05-10	687.80	7	29.80	9239.30	10899.90
AZUL	2024-05-10	1400.00	30	-1050.20	4700.00	20895.80
GOL	2024-05-14	1426.90	30	-130.20	4713.90	-19348.00
PETROBRAS	2024-05-14	60044.00	51	23700.00	117721.00	409922.00
JBS	2024-05-14	6428.80	7	1807.10	89147.10	48087.00
MARFRIG	2024-05-15	2646.00	9	63.00	31590.90	24654.00
COSAN	2024-05-28	2934.50	30	263.50	9842.00	50920.30
NEOENERGIA	2024-07-23	2970.00	27	815.00	10983.00	31826.00
VALE	2024-07-25	22161.10	40	15367.90	55056.00	213047.80
KLABIN S/A	2024-07-29	2052.00	49	640.90	4205.80	11618.30
GERDAU	2024-07-31	2624.00	16	945.00	16616.00	55341.90
AMBEV S/A	2024-08-01	5811.00	29	2451.90	20044.20	96035.40
AES BRASIL	2024-08-02	366.20	42	-108.70	871.90	5310.30
LOJAS RENNER	2024-08-05	670.50	22	314.90	2804.20	10317.10
CASAS BAHIA	2024-08-07	452.00	7	37.00	6479.00	3242.00
ULTRAPAR	2024-08-08	1335.60	4	491.20	32343.90	14853.80
EMBRAER	2024-08-08	995.50	13	415.70	7847.00	17309.10
SUZANO S.A.	2024-08-08	11276.60	54	-3545.40	20952.70	40614.10
PETROBRAS	2024-08-09	49740.00	41	-2605.00	122258.00	376040.00
MAGAZ LUIZA	2024-08-09	710.70	8	23.60	9010.00	10974.00
AZUL	2024-08-10	1100.00	25	-3809.60	4200.00	22831.20
JBS	2024-08-13	9882.10	10	1715.20	100606.30	50783.10
MARFRIG	2024-08-14	3400.00	10	75.40	34800.00	23207.20
GOL	2024-08-14	410.00	10	-3908.00	3937.00	-23219.00
COSAN	2024-08-14	2370.10	22	-227.10	10694.00	48255.90

3.1.2.3 Transformação dos dados

Para realizar a análise da correlação entre a divulgação do EBITDA e outros indicadores do Balanço Patrimonial e a evolução da cotação das ações de empresas listadas na B3 selecionadas para este estudo, precisamos extrair da nossa base de dados as informações necessárias para efetuar os cálculos.

O JUPYTER NOTEBOOK “1-TCC Pré-Processamento - Transformação.ipynb”, que se encontra no APÊNDICE B deste trabalho foi criado para isso.

Iniciamos o trabalho neste notebook repetindo as etapas de instalação e importação das bibliotecas necessárias e a conexão com o banco de dados TCC no POSTGRESQL.

Em seguida, criamos uma tabela, “dadosanalise” (Tabela 5) que recebe os dados selecionados nas tabelas “cotacoes” e “balancos” necessários para os cálculos. Os atributos desta tabela são:

Tabela 5 – Atributos da tabela dadosanalise

ATRIBUTO	TIPO DE DADO	DESCRIÇÃO
NOME_EMPRSA	TEXT NOT NULL,	-- Nome da Empresa Emissora
DATA_BALANCO	DATE NOT NULL,	-- Data do Balanço
CDGO_NGCCO_PPL	TEXT NOT NULL,	-- Código de Negociação do Papel
EBITDA	FLOAT(13),	-- EBITDA
MARGEM	FLOAT(5),	-- Margem EBITDA
LCRO_LQDO	FLOAT(13),	-- Lucro Líquido
RCTA_LQDA	FLOAT(13),	-- Receita Líquida
PTRMO_LQDO	FLOAT(13),	-- Patrimônio Líquido
MEDIA_ANT	FLOAT(13),	-- Média Anterior
MEDIA_POST	FLOAT(13),	-- Média Posterior
DIFERENCA	FLOAT(13),	-- Diferença Média Anterior e Posterior

Na inserção dos dados na tabela calculamos a média do preço médio das ações de cada empresa nos 5 dias anteriores à data de publicação do balanço que foi armazenada no atributo MEDIA_ANT.

Em seguida calculamos a média do preço médio das ações de cada empresa na data da publicação do balanço, mais 5 dias posteriores e armazenamos no atributo MEDIA_POST.

Exluímos da tabela as linhas que têm MEDIA_ANT ou MEDIA_POST igual a zeros.

Em um próximo passo, calculamos a diferença entre a MEDIA_ANT e a MEDIA_POST e armazenamos no atributo DIFERENCA, completando a carga, tratamento e preparação dos dados para a realização do estudo.

3.1.3 Extração de Padrões

Para a extração de padrões dos dados obtidos no passo anterior, criamos o JUPYTER NOTEBOOK “2 – TCC – Extração de Padrões.ipynb”, APÊNDICE C neste trabalho.

Como nos notebooks anteriores iniciamos o trabalho instalando e importando as bibliotecas necessárias e a fazemos a conexão com o banco de dados TCC no POSTGRESQL.

Para realizarmos a extração de padrões dos dados, o primeiro passo é recuperar os dados da tabela “dadosanalise” e armazenar em um Data Frame do Pandas, pois isso possibilita que

manipulemos os dados usando funções pré-existentes nas bibliotecas importadas para o PYTHON. Os atributos recuperados para a extração de padrões foram “EBITDA”, “LCRO_LQDO”, “RCTA_LQDA”, “PTRMO_LQDO” e “DIFERENCA”. Não foi necessária a conversão dos valores para FLOAT porque esse é o tipo de dado que foi usado para esses atributos na criação na tabela “dadosanalise”.

A escolha do atributo “EBITDA”, está justificada acima, neste trabalho, por ser o indicador escolhido pelo mercado para entender a capacidade da empresa gerar caixa.

O atributo “LCRO_LQDO” foi escolhido porque os dividendos pagos aos acionistas têm como base o lucro líquido da empresa no período.

“RCTA_LQDA” é a receita líquida que a empresa auferiu durante o período compreendido pela demonstração no Balanço Patrimonial e demonstra a capacidade que possui de gerar receitas.

O “PTRMO_LQDO” foi o último indicador do Balanço Patrimonial escolhido para o estudo, por ser a demonstração do valor da empresa e ter uma relação direta com o valor intrínseco das ações.

O próximo passo foi criar uma função para plotar o gráfico histograma de cada um dos atributos e executá-lo para conhecer a distribuição dos dados na escala original.

Como os parâmetros escolhidos para o estudo tem como característica intervalos de valores muito diferentes, faz-se necessário normalizá-los para que um parâmetro não influencie o resultado mais do que outro. Quando os parâmetros escolhidos possuem escalas semelhantes, não é necessário normalizá-los.

A normalização tem como objetivo trazer para uma escala comum os valores numéricos de cada coluna da base de dados, sem descaracterizar as diferenças existentes nos intervalos de valores, permitindo a busca por uma correlação entre eles.

Para a normalização, primeiramente aplicamos aos dados a função StandardScaler da biblioteca SKLEARN, que faz uso da distribuição normal, transformando os valores dos atributos para que a média dos dados padronizados seja igual a zero e o desvio padrão igual a um. Após a normalização, plotamos os novos histogramas para avaliar o resultado.

Analisando os novos histogramas gerados, concluímos que a normalização dos dados usando StandardScaler preservou a sua variância, porém gerou, para atributos como “EBITDA” que só continham valores positivos, valores negativos, não preservando sua característica original, o que não é interessante para a busca de correlação.

Substituímos, então, a StandardScaler pela função MaxAbsScaler da biblioteca SKLEARN, para normalizar os dados. Essa função usa o valor absoluto máximo de cada

parâmetro, para transformá-los em valores entre $[-1, 1]$, sem centralizar ou descolar os dados, preservando a escala de valores com relação aos dados originais. O uso da `MaxAbsScaler` garante que cada característica contribua igualmente para a análise sem perder a estrutura esparsa. Geramos novos histogramas dos dados normalizados com essa função e observamos que as características dos dados originais foram preservadas.

Com os dados normalizados podemos avaliar se existe uma correlação entre os indicadores obtidos no Balanço Patrimonial das empresas (“EBITDA”, “LCRO_LQDO”, “RCTA_LQDA” e “PTRMO_LQDO”) e a evolução da cotação de suas ações após a publicação dele (“DIFERENCA”).

Segundo Mattos (2017), “Para avaliar a existência de correlação entre duas variáveis ou até mesmo verificar se ela pode ser linear ou não, e ainda ter uma ideia de sua intensidade e sentido, é possível recorrer a uma representação gráfica muito simples: os pares de observações das duas variáveis são plotados num diagrama cartesiano chamado “diagrama de dispersão”.”

Geramos o diagrama de dispersão dos atributos `ebitda`, `lcro_lqdo`, `rcta_lqda`, `ptrmo_lqdo` e diferença, com relação à diferença encontrada entre a média das cotações nos cinco dias anteriores e posteriores à publicação do balanço.

Figura 2 – Diagrama de dispersão do atributo `ebitda`

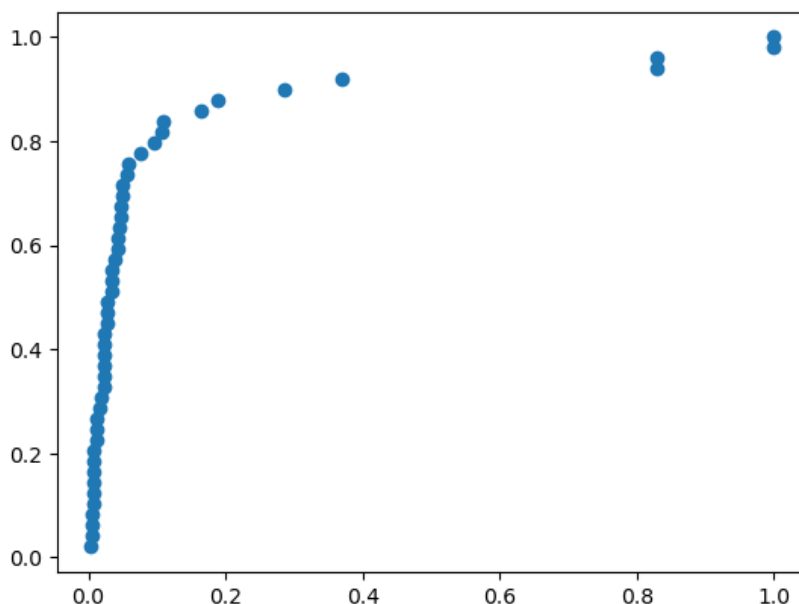


Figura 3 – Diagrama de dispersão do atributo lcro_lqdo

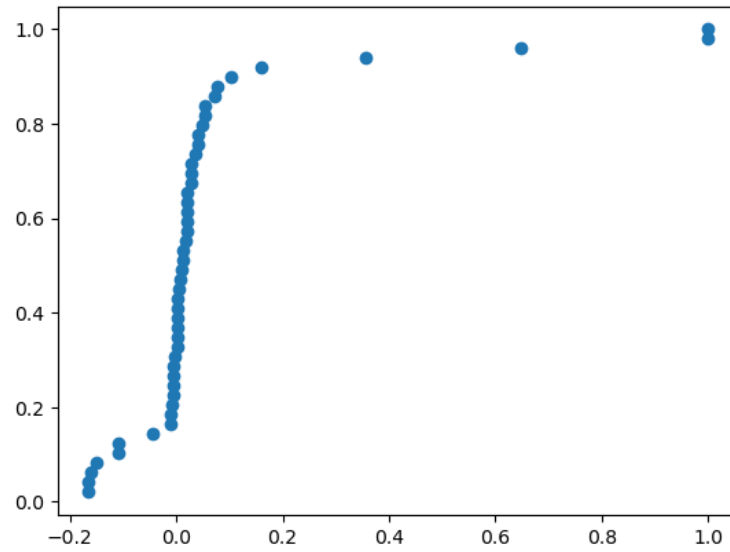


Figura 4 – Diagrama de dispersão do atributo rcta_lqda

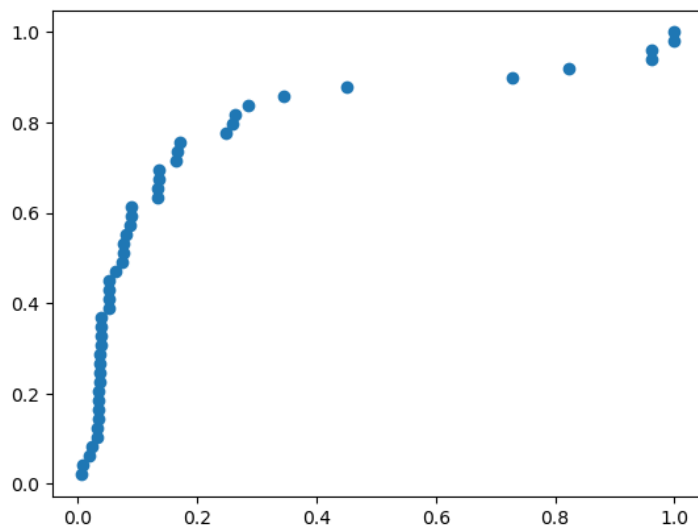


Figura 5 – Diagrama de dispersão do atributo prtmo_lqdo

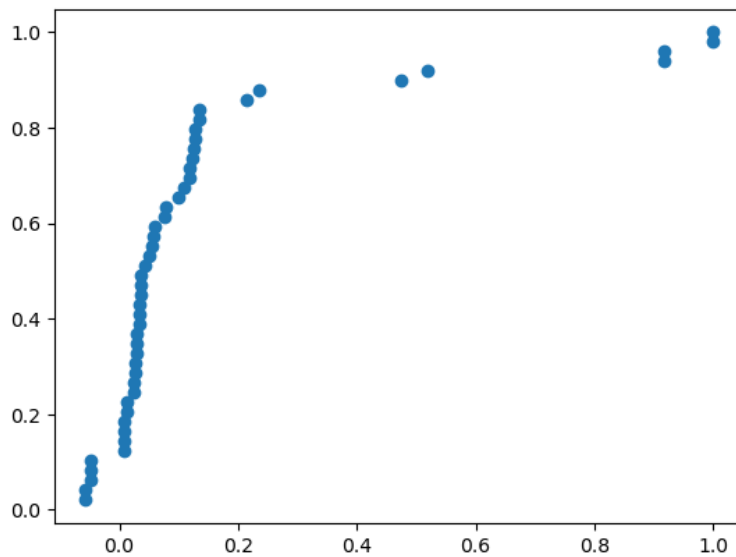
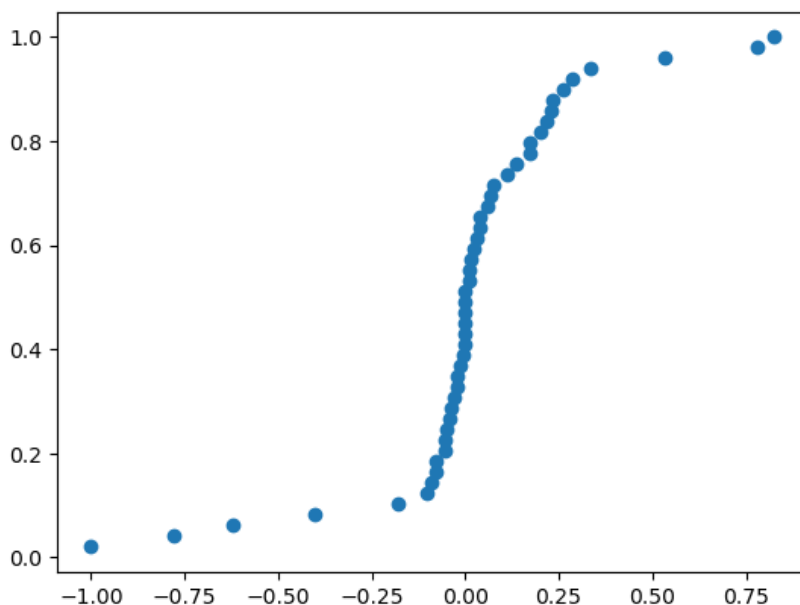


Figura 6 – Diagrama de dispersão do atributo diferença



Os diagramas de dispersão gerados mostram que os dados dos cinco atributos não seguem uma distribuição linear.

Para avaliar se há uma correlação entre os indicadores e a diferença calculada entre a média das cotações dos cinco dias anteriores e posteriores à data da publicação do Balanço Patrimonial das empresas, podemos usar dois coeficientes de correlação: O coeficiente de correlação linear de Pearson e o coeficiente de correlação de postos de Spearman.

Sobre o coeficiente de correlação linear de Pearson, Mattos, Konrath e Azambuja (2017) definem:

Quando a correlação é linear, a mensuração pode ser feita pelo coeficiente de correlação linear de Pearson, representado por r_{xy} , quando for determinado a partir de uma amostra. ...

Esse coeficiente pode variar entre -1 e $+1$. A correlação será tanto mais forte quanto mais próximo o coeficiente estiver desses valores e será tanto mais fraca quanto mais próximo estiver de zero.

Já quanto ao coeficiente de correlação de postos de Spearman, as autoras dizem:

Esse coeficiente não paramétrico destina-se a determinar o grau de associação entre duas variáveis x e y , dispostas em pontos ordenados, ou seja, o objetivo é estudar a correlação entre duas classificações. Aplica-se, igualmente, em variáveis intervalares e de razão como alternativa ao coeficiente de correlação linear de Pearson, quando, neste último, viola-se a suposição de normalidade.

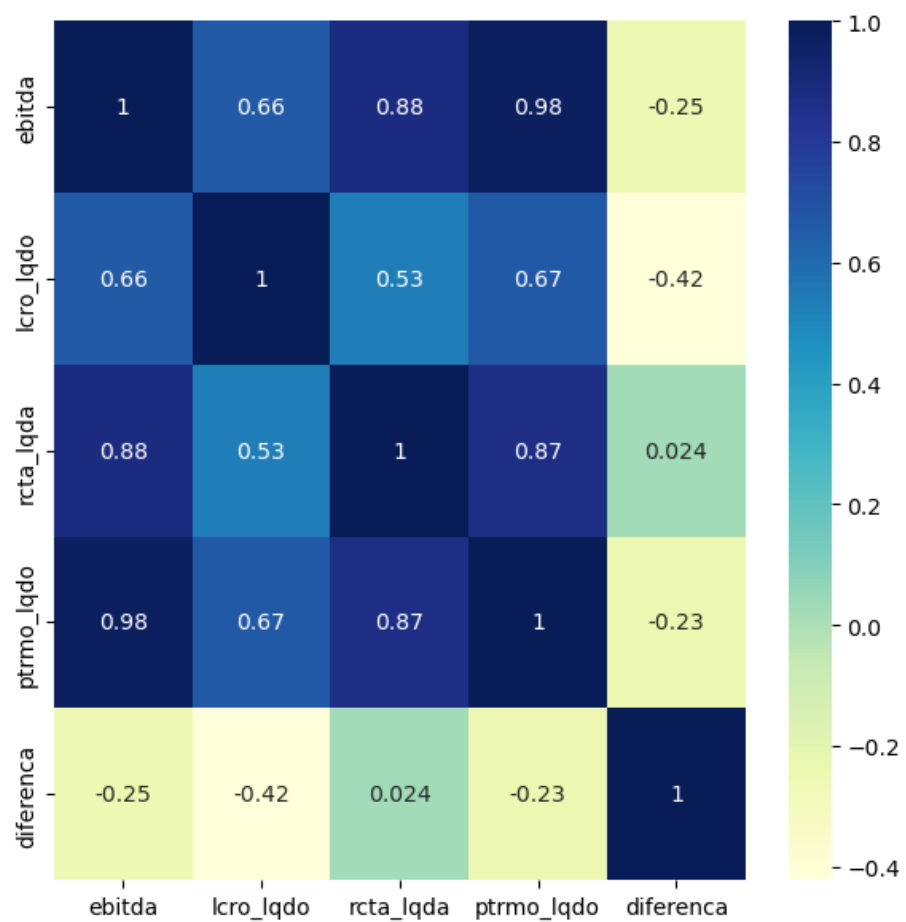
Embora fique evidente que o coeficiente de correlação ideal para os dados deste estudo seja o de Spearman, fizemos a avaliação da correlação do atributo EBITDA em relação à

DIFERENÇA usando os dois coeficientes, Pearson e Spearman. Não encontramos uma correlação forte entre os atributos usando os dois coeficientes (Pearson = -0,24 e Spearman = 0,06)

Na sequência geramos um mapa de correlação de todos os atributos e obtivemos para Pearson:

Figura 7 – Mapa de Correlação dos atributos usando o coeficiente de correlação linear de Pearson

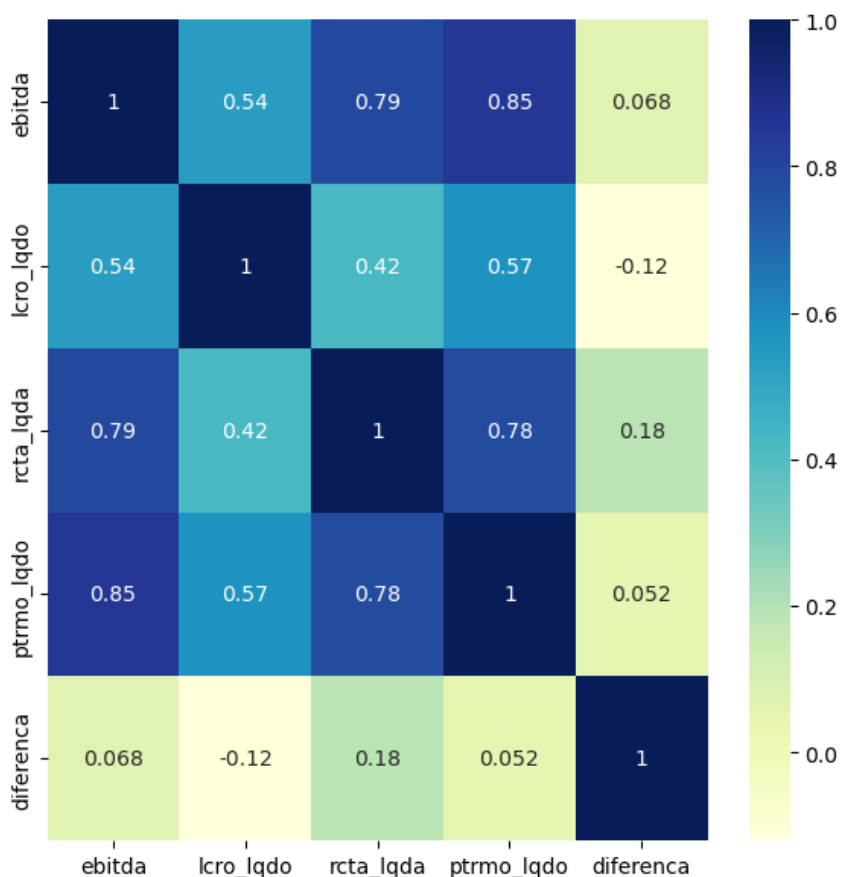
Mapa de Correlação dos atributos usando método de Pearson



Para Spearman:

Figura 8 – Mapa de Correlação dos atributos usando o coeficiente de correlação pontual de Spearman

Mapa de Correlação dos atributos usando método de Spearman



3.1.4 Pós-Processamento

Analizamos a correlação calculada entre cada um dos atributos - ebitda, lcro_lqdo, rcta_lqda, ptrmo_lqdo com relação ao atributo diferenca e concluímos que nenhuma correlação significativa foi encontrada.

No mapa de correlação de Pearson, o maior coeficiente de correlação encontrado foi entre o Lucro Líquido e a variação da média das cotações anteriores e posteriores à data de divulgação do balanço - -0,48. Uma correlação negativa e não significativa.

No mapa de correlação de Spearman, o maior coeficiente de correlação encontrado foi 0,18, entre a Receita Líquida e a variação da média das cotações. Também uma correlação nada significativa e positiva.

4 CONCLUSÃO

Em um mercado eficiente, o valor de negociação das ações emitidas pelas Sociedades Anônimas deve refletir as informações disponibilizadas ao público em geral ou a um grupo restrito. A publicação do balanço patrimonial das empresas traz aos acionistas e investidores informações importantes para o entendimento da capacidade da empresa de gerar lucro, crescer e honrar com seus compromissos de curto e longo prazo, entre elas o EBITDA, indicador que tem como utilidade o entendimento da eficiência operacional e a capacidade da empresa para gerar caixa. Outros indicadores também são importantes, dentre eles o lucro líquido, a receita líquida e o patrimônio líquido. Esse estudo realizou a aquisição, limpeza e transformação dos dados do histórico de cotações de ações obtidos na B3 e do balanço patrimonial de dezenove empresas de diversos setores de atividade, e efetuou o cálculo da correlação entre os indicadores elencados acima e a diferença da média do preço médio das ações das empresas nos cinco dias anteriores e posteriores à data de publicação do balanço patrimonial, concluindo que não existe uma correlação significativa entre esses dados.

Isso pode estar relacionado ao fato de que mercado eficiente é uma característica de mercados maduros, como as Bolsas estadunidenses e europeias, onde, além dos grandes investidores, também o poupador investe em ações. O mercado de capitais no Brasil é altamente especulativo e volátil, não sendo o palco de atuação de pequenos investidores e poupadores.

Como sugestão para futuros estudos, pode-se trabalhar com um período maior de histórico de cotação de ações, uma série maior de balanços patrimoniais e um universo maior de empresas. O volume maior de dados pode apresentar resultados mais precisos, confirmando ou refutando as conclusões deste estudo. Trabalhar com dados de mercados maduros também pode apresentar resultados diferentes.

REFERÊNCIAS

ARCENO, C. **Mercado de ações**: Uma abordagem no âmbito da educação financeira. 2019. 90p. Dissertação (mestrado profissional) - Centro de Ciências Físicas e Matemáticas, Programa de Pós Graduação em Matemática, Universidade Federal de Santa Catarina, Florianópolis, 2019.

ASSAF NETO, A. **Mercado financeiro**. 14. ed. São Paulo: Atlas, 2018.

Comissão de Valores Mobiliários **Instrução nº 527** São Paulo: CVM, 2012 Disponível em: <https://conteudo.cvm.gov.br/export/sites/cvm/legislacao/instrucoes/anexos/500/inst527.pdf>

MATTOS, V.L.D.; KONRATH, C.; AZAMBUJA, A.M.V **Introdução à estatística**: aplicações em ciências exatas. 1. ed. Rio de Janeiro: LTC, 2017.

PADOVEZE, C.L. **Mitos e Lendas em Finanças**: a “maldição” do EBITDA. São Paulo: Almedina, 2020.

REZENDE, S.O. **Sistemas inteligentes**: fundamentos e aplicações. Barueri: Manole, 2003

SOUSA, J.D.S.; SALES, E.N. **Relação entre o Ebitda e o Retorno das Ações das Empresas Listadas na B3**. In: 21ª USP INTERNACIONAL CONFERENCE IN ACCOUNTING “Accounting and Actuarial Sciences improving economic and social development” São Paulo, 2021 Disponível em: <https://congressousp.fipecafi.org/anais/21UspInternational/ArtigosDownload/3154.pdf>

VERRIEST, A.; BOUWENS, J.; DE KOK, T. **The Prevalence and Validity of EBITDA as a Performance Measure**. 2018 Disponível em SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3171131

Apêndice A – 0-TCC Pré-Processamento - Carga e Limpeza na Base

MBI em Inteligência Artificial e Big Data

Trabalho de Conclusão de Curso

1. Preparação da base de dados usada no TCC

Este TCC foi desenvolvida assumindo que a base de dados está armazenada em um servidor do tipo Sistema de Gerenciamento de Bases de Dados (SGBD), o tipo de ferramenta mais apropriado para isso.

Assumimos que:

- O **servidor** usado é o SGBD de domínio público **POSTGRESQL**,
 - que ele já está instalado,
 - com a base de dados TCC definida,
 - e que você (o usuário que estabelece a conexão) tem acesso para executar as operações de consultar, criar, atualizar e apagar qualquer objeto.
- O ambiente de execução do **notebook Jupyter** está instalado, com os seguintes módulos disponíveis:
 - `matplotlib` para traçar gráficos,
 - `pandas` e `numpy` para trabalhar com vetores e matrizes,
 - `ipywidgets` para interadores para experimentar com parâmetros das consultas,
 - `io` para operadores de entrada e saída de dados.
 - `create_engine` para criar uma conexão com o banco de dados.
 - `interact` para manipular uma lista com o resultado das consultas SQL **SELECT**.

```
In [ ]: %pip install matplotlib pandas numpy ipywidgets psycopg2-binary sqlalchemy ipython-sql
```

```
In [ ]: ##### Importar os módulos necessários para o Notebook:
import matplotlib.pyplot as plt
import pandas.io.sql as psql
import numpy as np
import pandas as pd
import ipywidgets as widgets
import psycopg2
from ipywidgets import interact
from sqlalchemy import create_engine
```

1.1. Conectar com uma base de dados

Para iniciar qualquer operação de acesso em um servidor de bases de dados, é necessário que a **base de dados** já esteja carregada com todas a(s) **tabela(s)** que armazenam os dados que devem ser manipulados.\

Então, para começar é necessário estabelecer a conexão com o servidor, e indicar a base de dados alvo.\

```
In [ ]: ##### Conectar com um servidor SQL ##### --> Postgres
%load_ext sql

# Connection format: %sql dialect+driver://username:password@host:port/database
engine = create_engine('postgresql://postgres:pgadmin@localhost/TCC')
%sql postgresql://postgres:Sucesso8.,@localhost/TCC

#### verificar qual é a base de dados que está sendo assessada:
%sql DB << SELECT Current_Database();
print('Conectado à base de dados:\n', DB)
```

Agora estamos conectados à base de dados **TCC**.

Como este **notebook** está sendo executado sequencialmente, por enquanto ela está vazia.

⚠ Para evitar erros por re-execução de células no Notebook, assumimos que os objetos que pretendemos criar até já possam existir. Por isso usamos construções tais como: `DROP DATABASE IF EXISTS` antes de criá-los (o que em geral é desnecessário).

1.3. Carregar e Tratar o Histórico de Cotações das Ações na B3

Os dados estão disponíveis em um arquivo de tipo *comma separated values* (.csv) que contem os atributos de interesse. A carga é feita em dois passos:

1. Criar uma tabela com todos os atributos de: cotacoes

```
In [ ]: %%sql
DROP TABLE IF EXISTS cotacoes;
CREATE TABLE cotacoes (
    DATA_PREGAO DATE,           -- Data do Pregão
    CDGO_NGCO_PPL TEXT,         -- Código de Negociação do Papel
    TIPO_MERC NUMERIC(5),       -- Tipo de Mercado - 010 - À Vista
    NOME_RESMDO_EMPRSA TEXT,    -- Nome Resumido da Empresa Emissora
    ESPCFCCO_PPL TEXT,         -- Especificação do Papel
    MOEDA_RFRNCA TEXT,         -- Moeda Usada na Data do Pregão
    PRECO_ABRTRA DECIMAL(11,2), -- Preço Abertura
    PRECO_MEDIO DECIMAL(11,2),  -- Preço Médio
    PRECO_ULTMO DECIMAL(11,2),  -- Preço do Último Negócio Fechado no Pregão
    FTR_CTCO_PPL DECIMAL(7)     -- Fator Cotação do Papel 1=Cotação unitária 1000=Cotação lote mil ações
);
```

2. Copiamos os dados das cotações de ações na B3 existentes nos arquivos mensais .csv nessa tabela.
 - A. No arquivo, as datas estão no formato Dia/Mes/Ano.

É aconselhável indicar esse formato de data para o gerenciador.

```
In [ ]: %%sql
SET DateStyle to DMY, ISO;
COPY Cotacoes FROM 'C:\Dados\COTAHIST_M042024_CONVERTIDO.csv'
WITH (DELIMITER ';', NULL '', ENCODING 'UTF8', HEADER true, FORMAT CSV);
COMMIT;
```

```
In [ ]: %%sql
SET DateStyle to DMY, ISO;
COPY Cotacoes FROM 'C:\Dados\COTAHIST_M052024_CONVERTIDO.csv'
WITH (DELIMITER ';', NULL '', ENCODING 'UTF8', HEADER true, FORMAT CSV);
COMMIT;
```

```
In [ ]: %%sql
SET DateStyle to DMY, ISO;
COPY Cotacoes FROM 'C:\Dados\COTAHIST_M062024_CONVERTIDO.csv'
WITH (DELIMITER ';', NULL '', ENCODING 'UTF8', HEADER true, FORMAT CSV);
COMMIT;
```

```
In [ ]: %%sql
SET DateStyle to DMY, ISO;
COPY Cotacoes FROM 'C:\Dados\COTAHIST_M072024_CONVERTIDO.csv'
WITH (DELIMITER ';', NULL '', ENCODING 'UTF8', HEADER true, FORMAT CSV);
COMMIT;
```

```
In [ ]: %%sql
SET DateStyle to DMY, ISO;
COPY Cotacoes FROM 'C:\Dados\COTAHIST_M082024_CONVERTIDO.csv'
WITH (DELIMITER ';', NULL '', ENCODING 'UTF8', HEADER true, FORMAT CSV);
COMMIT;
```

1.3.1 Verificando os dados carregados

Vamos conhecer a base de dados.

Vamos conferir os atributos da relação de cotacoes, tal como registrado na base de dados.

Para isso, listamos a sequência de atributos que definem a tabela armazenada.

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
        FROM cotacoes
        ORDER BY cotacoes.data_pregao
        LIMIT :Limit;
    #####
    print(result_set)
```

1.3.2 Deletando os dados que não são objeto deste trabalho

a - Vamos deletar todas as cotações que tem Tipo de Mercado - tipo_merc diferente de 10 - Mercado à Vista

```
In [ ]: %%sql
DELETE FROM cotacoes
WHERE tipo_merc <> 10;
```

b - Vamos deletar todas as cotações de empresas que não fazem parte do estudo

```
In [ ]: %%sql
DELETE FROM cotacoes
WHERE nome_resmdo_empresa <> 'AES BRASIL' AND
      nome_resmdo_empresa <> 'AMBEV S/A' AND
      nome_resmdo_empresa <> 'AZUL' AND
      nome_resmdo_empresa <> 'CASAS BAHIA' AND
      nome_resmdo_empresa <> 'COSAN' AND
      nome_resmdo_empresa <> 'EMBRAER' AND
      nome_resmdo_empresa <> 'GERDAU' AND
      nome_resmdo_empresa <> 'GOL' AND
      nome_resmdo_empresa <> 'JBS' AND
      nome_resmdo_empresa <> 'KLABIN S/A' AND
      nome_resmdo_empresa <> 'LOJAS RENNER' AND
      nome_resmdo_empresa <> 'MAGAZ LUIZA' AND
      nome_resmdo_empresa <> 'MARFRIG' AND
      nome_resmdo_empresa <> 'NEOENERGIA' AND
      nome_resmdo_empresa <> 'PETROBRAS' AND
      nome_resmdo_empresa <> 'SUZANO S.A.' AND
      nome_resmdo_empresa <> 'ULTRAPAR' AND
      nome_resmdo_empresa <> 'VALE';
```

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
    FROM cotacoes
    ORDER BY cotacoes.data_pregao
    LIMIT :Limit;
    #####
    print(result_set)
```

Vamos verificar se há dados nulos na base selecionada

```
In [ ]: np.any(pd.isna(result_set))
```

1.3.3 Efetuar o ajuste das casas decimais dos valores

Vamos ajustar o valor das cotações dividindo por 100 para adequar as casas decimais:

```
In [ ]: %%sql
UPDATE cotacoes SET preco_abrtra = (preco_abrtra / 100),
                        preco_medio = (preco_medio / 100),
                        preco_ultmo = (preco_ultmo / 100);
```

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
    FROM cotacoes
    ORDER BY cotacoes.data_pregao
    LIMIT :Limit;
    #####
    print(result_set)
```

Vamos verificar se há dados duplicados na tabela cotacoes

```
In [ ]: df = pd.DataFrame(result_set)
#
df.duplicated()
```

1.3. Carregar e Tratar os Dados dos Balanços das Empresas Selecionadas

Os dados estão disponíveis em um arquivo de tipo *comma separated values* (.csv) que contem os atributos de interesse. A carga é feita em dois passos:

1. Criar uma tabela com todos os atributos de: `balancos`

```
In [ ]: %%sql
DROP TABLE IF EXISTS balancos;
CREATE TABLE balancos (
    NOME_EMPRSA      TEXT,          -- Nome da Empresa
    DATA_BALANCO    DATE,          -- Data do Balanco
    EBITDA            DECIMAL(11,2), -- EBITDA
    MARGEM            NUMERIC(5),    -- Margem EBITDA
    LCRO_LQDO         DECIMAL(11,2), -- Lucro Líquido
    RCTA_LQDA         DECIMAL(11,2), -- Receita Líquida
    PTRMO_LQDO        DECIMAL(11,2), -- Patrimônio Líquido
    CONSTRAINT balancos_PK PRIMARY KEY (NOME_EMPRSA, DATA_BALANCO)
);
```

2. Copiamos os dados dos balanços trimestrais das empresas (primeiro e segundo trimestre de 2024) existentes no arquivo `.csv` nessa tabela.
 - A. No arquivo, as datas estão no formato `Dia/Mes/Ano`.
É aconselhável indicar esse formato de data para o gerenciador.
 - B. Os valores estão em milhões de reais, porém o arquivo original não tem ponto decimal, incorporando os décimos ao número inteiro (numero x 10).
Pra recuperar as casas decimais é preciso dividir os valores por 10.

```
In [ ]: %%sql
SET DateStyle to DMY, ISO;
COPY balancos FROM 'C:\Dados\BALANCOS.csv'
WITH (DELIMITER ';', NULL '', ENCODING 'UTF8', HEADER true, FORMAT CSV);
COMMIT;
```

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
        FROM balancos
        ORDER BY balancos.data_balanco
    LIMIT :Limit;
    #####
    print(result_set)
```

Vamos verificar se há dados nulos na base de balanços

```
In [ ]: np.any(pd.isna(result_set))
```

```
In [ ]: %%sql
UPDATE balancos SET ebitda = (ebitda / 10),
    margem = (margem / 10),
    lcro_lqdo = (lcro_lqdo / 10),
    rcta_lqda = (rcta_lqda / 10),
    ptrmo_lqdo = (ptrmo_lqdo / 10);
```

```
In [ ]: %sql result_set <<
SELECT *
    FROM balancos
    ORDER BY balancos.data_balanco;
#####
print(result_set)
```

MBI em Inteligência Artificial e Big Data
Preparação da base de dados usada no TCC

FIM

Apêndice B – 1-TCC Pré-Processamento – Transformação

MBI em Inteligência Artificial e Big Data

Trabalho de Conclusão de Curso

1. Prepara os dados para o estudo

Vamos recuperar na base de dados as cotações de cada empresa de 5 dias antes e 5 dias depois da publicação do balanço trimestral. Com essas informações vamos calcular a cotação média dos 5 dias antes e o mesmo para 5 dias depois. Vamos subtrair a média das cotações anteriores da média das cotações posteriores para identificar se houve diferença, qual o valor da diferença e se esse valor é positivo ou negativo.

Para esse trabalho assumimos que:

- O **servidor** usado é o SGBD de domínio público **Postgres**,
 - que ele já está instalado,
 - com a base de dados TCC carregada,
 - e que você (o usuário que estabelece a conexão) tem acesso para executar as operações de consultar, criar, atualizar e apagar qualquer objeto.
- O ambiente de execução do **notebook Jupyter** está instalado, com os seguintes módulos disponíveis:
 - **matplotlib** para traçar gráficos,
 - **pandas** e **numpy** para trabalhar com vetores e matrizes,
 - **ipywidgets** para interadores para experimentar com parâmetros das consultas,
 - **io** para operadores de entrada e saída de dados.
 - **create_engine** para criar uma conexão com o banco de dados.
 - **interact** para manipular uma lista com o resultado das consultas SQL SELECT.

```
In [ ]: %pip install matplotlib pandas numpy ipywidgets psycopg2-binary sqlalchemy ipython-sql
```

```
In [ ]: ##### Importar os módulos necessários para o Notebook:
import matplotlib.pyplot as plt
import pandas.io.sql as psql
import numpy as np
import pandas as pd
import ipywidgets as widgets
import psycopg2
from ipywidgets import interact
from sqlalchemy import create_engine
```

1.1. Conectar com uma base de dados

Para iniciar qualquer operação de acesso em um servidor de bases de dados, é necessário que a **base de dados** já esteja carregada com todas a(s) **tabela(s)** que armazenam os dados que devem ser manipulados.\

Então, para começar é necessário estabelecer a conexão com o servidor, e indicar a base de dados alvo.

Assumimos aqui que o gerenciador **Postgres** já tem os dados carregados na Base de Dados padrão, chamada **TCC**, e que o usuário tem acesso ao esquema padrão **public** dessa base de dados.

```
In [ ]: ##### Conectar com um servidor SQL ##### --> Postgres
%load_ext sql

# Connection format: %sql dialect+driver://username:password@host:port/database
engine = create_engine('postgres://postgres:pgadmin@localhost/TCC')
%sql postgres://postgres:Sucesso8.,@localhost/TCC

#### verificar qual é a base de dados que está sendo acessada:
%sql DB << SELECT Current_Database();
print('Conectado à base de dados:\n', DB)
```

Agora estamos conectados à base de dados **TCC**.\

1.3. Carregar e Tratar a média da cotação das ações dos 5 dias anteriores e a média dos 5 dias posteriores à publicação do balanço trimestral das empresas. Depois calcular a diferença entre a média anterior e a posterior

A carga é feita em dois passos:

1. Criar uma tabela com os atributos necessários para fazer a análise: `dadosanalise`

```
In [ ]: %%sql
DROP TABLE IF EXISTS dadosanalise;
CREATE TABLE dadosanalise (
    NOME_EMPRSA      TEXT NOT NULL,      -- Nome da Empresa Emissora
    DATA_BALANCO    DATE NOT NULL,      -- Data do Pregão
    CDGO_NGCCO_PPL   TEXT NOT NULL,      -- Código de Negociação do Papel
    EBITDA            FLOAT(13),          -- EBITDA
    MARGEM            FLOAT(5),           -- Margem EBITDA
    LCRO_LQDO         FLOAT(13),          -- Lucro Líquido
    RCTA_LQDA         FLOAT(13),          -- Receita Líquida
    PTRMO_LQDO        FLOAT(13),          -- Patrimônio Líquido
    MEDIA_ANT         FLOAT(13),          -- Média Anterior
    MEDIA_POST        FLOAT(13),          -- Média Posterior
    DIFERENCA         FLOAT(13),          -- Diferença Média Anterior e Posterior
    CONSTRAINT dadosanalise_PK PRIMARY KEY (NOME_EMPRSA, DATA_BALANCO, CDGO_NGCCO_PPL)
);
COMMIT;
```

2. Selecionamos os dados das cotações de 5 dias anteriores à data de publicação do balanço das empresas e inserimos na tabela 'dadosanalise'. Para isso criamos uma função `EntraDados()`.

Na função `EntraDados()` usamos um cursor (`cur1`) para fazer a carga dos dados usando o comando SQL INSERT.

```
In [ ]: def EntraDados():
    #####
    %%sql result_ant <<
    SELECT  a.nome_empresa AS NOME,
            a.data_balanco AS DATA,
            b.cdgo_ngcco_ppl AS CODIGO,
            a.EBITDA AS EBIT,
            a.MARGEM AS MRGM,
            a.LCRO_LQDO AS LUCRO,
            a.RCTA_LQDA AS RECEITA,
            a.PTRMO_LQDO AS PTRMO,
            avg(b.preco_medio) AS MEDIAANT
    FROM    balancos a,
            cotacoes b
    WHERE   a.nome_empresa = b.nome_resmdo_empresa and
            b.data_pregao between (a.data_balanco - 6) and (a.data_balanco - 1)
    GROUP BY a.nome_empresa, a.data_balanco, b.cdgo_ngcco_ppl;

    #####
    linhas = len(result_ant)
    num = 0
    df = pd.DataFrame(result_ant)
    conn = psycopg2.connect(host="localhost", port="5432", dbname="TCC", user="postgres", password="Sucesso8.")
    cur1 = conn.cursor()
    while (num < linhas):
        NOME1 = df.nome[num]
        DATA1 = df.data[num]
        CODIGO1 = df.codigo[num]
        EBIT1 = df.ebit[num]
        MRGM1 = df.mrgm[num]
        LUCRO1 = df.lucro[num]
        RECEITA1 = df.receita[num]
        PTRMO1 = df.ptrmo[num]
        MEDIAANT1 = df.mediaant[num]
        num = num + 1
        sql = """INSERT INTO dadosanalise (NOME_EMPRSA, DATA_BALANCO, CDGO_NGCCO_PPL, EBITDA, MARGEM, LCRO_LQDO, RCTA_LQDA, PTRMO,
        cur1.execute(sql, [NOME1, DATA1, CODIGO1, EBIT1, MRGM1, LUCRO1, RECEITA1, PTRMO1, MEDIAANT1])

    cur1.execute("commit")
    cur1.close()
```

Executamos, então, a função EntraDados()

```
In [ ]: EntraDados()
```

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
    FROM dadosanalise
    ORDER BY dadosanalise.nome_empresa, dadosanalise.data_balanco, dadosanalise.cdgo_ngcco_ppl
    LIMIT :Limit;
    #####
    print(result_set)
```

3. Seleccionamos os dados das cotações do dia do balanço até 5 dias posteriores à data de publicação do balanço das empresas e inserimos na tabela 'dados_analise' fazendo um UPDATE. Para isso criamos uma função AlteraDados()

Na função EntraDados() usamos um cursor (cur1) para fazer a alteração dos dados usando o comando SQL UPDATE.

```
In [ ]: def AlteraDados():
    #####
    %sql result_post <<
    SELECT a.nome_empresa AS nome,
           a.data_balanco AS data,
           b.cdgo_ngcco_ppl AS codigo,
           avg(b.preco_medio) AS mediapost
    FROM balancos a,
         cotacoes b
    WHERE a.nome_empresa = b.nome_resmdo_empresa and
          b.data_pregao between (a.data_balanco) and (a.data_balanco + 5)
    GROUP BY a.nome_empresa, a.data_balanco, b.cdgo_ngcco_ppl;
    linhas = len(result_post)
    num = 0
    df = pd.DataFrame(result_post)
    conn = psycopg2.connect(host="localhost", port="5432", dbname="TCC", user="postgres", password="Sucesso8.")
    cur1 = conn.cursor()
    while (num < linhas):
        NOME1 = df.nome[num]
        DATA1 = df.data[num]
        CODIGO1 = df.codigo[num]
        MEDIAPOST1 = df.mediapost[num]
        num = num + 1
        sql = """UPDATE dadosanalise SET media_post= %s WHERE nome_empresa = %s and data_balanco = %s and cdgo_ngcco_ppl = %s"""
        cur1.execute(sql, [MEDIAPOST1, NOME1, DATA1, CODIGO1])

    cur1.execute("commit")
    cur1.close()
```

Executamos, então, a função AlteraDados()

```
In [ ]: AlteraDados()
```

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
    FROM dadosanalise
    ORDER BY dadosanalise.nome_empresa, dadosanalise.data_balanco, dadosanalise.cdgo_ngcco_ppl
    LIMIT :Limit;
    #####
    print(result_set)
```

4. Vamos deletar as tuplas que tem media anterior ou media posterior igual a zero

```
In [ ]: %sql
DELETE FROM dadosanalise
WHERE media_ant = 0 or media_post = 0;
```

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
        FROM dadosanalise
        ORDER BY dadosanalise.nome_empresa, dadosanalise.data_balanco, dadosanalise.cdgo_ngcco_ppl
        LIMIT :Limit;
    #####
    print(result_set)
```

5. Calculamos a diferença entre a média das cotações anteriores e a média das cotações posteriores e atualizamos a tabela 'dados_analise' fazendo um UPDATE. Para isso criamos uma função CalculaDiferenca()

Na função CalculaDiferenca() usamos um cursor (cur1) para fazer a alteração dos dados usando o comando SQL UPDATE.

```
In [ ]: def CalculaDiferenca():
    #####
    %sql result_calc <<
    SELECT nome_empresa AS nome,
           data_balanco AS data,
           cdgo_ngcco_ppl AS codigo,
           media_ant AS mediaant,
           media_post AS mediaapost
    FROM dadosanalise;
    linhas = len(result_calc)
    num = 0
    conn = psycopg2.connect(host="localhost", port="5432", dbname="TCC", user="postgres", password="Sucesso8.",")
    cur1 = conn.cursor()
    df = pd.DataFrame(result_calc)
    while (num < linhas):
        NOME1 = df.nome[num]
        DATA1 = df.data[num]
        CODIGO1 = df.codigo[num]
        MEDIAANT1 = df.mediaant[num]
        MEDIAPOST1 = df.mediaapost[num]
        DIFERENCA1 = MEDIAPOST1 - MEDIAANT1
        num = num + 1
        sql = """UPDATE dadosanalise SET diferenca = %s WHERE nome_empresa = %s and data_balanco = %s and cdgo_ngcco_ppl = %s"""
        cur1.execute(sql, [DIFERENCA1, NOME1, DATA1, CODIGO1])

    cur1.execute("commit")
    cur1.close()
```

Executamos, então, a função CalculaDiferenca()

```
In [ ]: CalculaDiferenca()
```

```
In [ ]: @interact(Limit=widgets.IntSlider(value=5, min=0, max=52,description='Quantidade'))
def Listagem(Limit=5):
    #####
    %sql result_set <<
    SELECT *
        FROM dadosanalise
        ORDER BY dadosanalise.nome_empresa, dadosanalise.data_balanco, dadosanalise.cdgo_ngcco_ppl
        LIMIT :Limit;
    #####
    print(result_set)
```

MBI em Inteligência Artificial e Big Data
Preparação dos dados para o estudo

FIM

Apêndice C – 2-TCC Extração de Padrões

MBI em Inteligência Artificial e Big Data

Trabalho de Conclusão de Curso

1. Análise dos dados

Para esse trabalho assumimos que:

- O **servidor** usado é o SGBD de domínio público **Postgres**,
 - que ele já está instalado,
 - com a base de dados TCC carregada,
 - e que você (o usuário que estabelece a conexão) tem acesso para executar as operações de consultar, criar, atualizar e apagar qualquer objeto.
- O ambiente de execução do **notebook Jupyter** está instalado, com os seguintes módulos disponíveis:
 - `matplotlib` para traçar gráficos,
 - `pandas` e `numpy` para trabalhar com vetores e matrizes,
 - `ipywidgets` para interadores para experimentar com parâmetros das consultas,
 - `io` para operadores de entrada e saída de dados.
 - `create_engine` para criar uma conexão com o banco de dados.
 - `interact` para manipular uma lista com o resultado das consultas SQL SELECT.

```
In [ ]: %pip install matplotlib pandas numpy ipywidgets psycopg2-binary sqlalchemy ipython-sql
```

```
In [ ]: ##### Importar os módulos necessários para o Notebook:
import matplotlib.pyplot as plt
import pandas.io.sql as psql
import numpy as np
import ipywidgets as widgets
import psycopg2
import pandas as pd
import numpy as np
import seaborn as sns
import scipy as scp
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import MaxAbsScaler
from ipywidgets import interact
from sqlalchemy import create_engine
from scipy.stats import pearsonr, spearmanr
```

1.1. Conectar com uma base de dados

Para iniciar qualquer operação de acesso em um servidor de bases de dados, é necessário que a **base de dados** já esteja carregada com todas a(s) **tabela(s)** que armazenam os dados que devem ser manipulados.\

Então, para começar é necessário estabelecer a conexão com o servidor, e indicar a base de dados alvo.

Assumimos aqui que o gerenciador **Postgres** já tem os dados carregados na Base de Dados padrão, chamada **TCC**, e que o usuário tem acesso ao esquema padrão **public** dessa base de dados.

```
In [ ]: ##### Conectar com um servidor SQL ##### --> Postgres
%load_ext sql

# Connection format: %sql dialect+driver://username:password@host:port/database
engine = create_engine('postgresql://postgres:pgadmin@localhost/TCC')
%sql postgresql://postgres:Sucesso8.,@localhost/TCC

### verificar qual é a base de dados que está sendo acessada:
%sql DB << SELECT Current_Database();
print('Conectado à base de dados:\n', DB)
```

Agora estamos conectados à base de dados **TCC**.\

1.3. Recuperar os dados da tabela 'dadosanalise' para uma dataframe Pandas

A carga é feita selecionando os dados da tabela e carregando em um dataframe Pandas

```
In [ ]: %sql result_ant <<
SELECT ebitda, lcro_lqdo, rcta_lqda, ptrmo_lqdo, diferenca
FROM dadosanalise;

#####

df = pd.DataFrame(result_ant)
df

In [ ]: df.info()

In [ ]: df.describe()

In [ ]: indicadores = pd.DataFrame({"EBITDA": df['ebitda'], "Lucro Líquido": df['lcro_lqdo'], "Receita Líquida": df['rcta_lqda'], "Patrimônio Líquido": df['ptrmo_lqdo'], "Diferença": df['diferenca']})
print("Indicadores: dados formato largo")
print(indicadores.head(5))

In [ ]: assimetria = indicadores.skew()
print("Assimetria")
print(assimetria)

In [ ]: curtose = indicadores.kurt()
print("Curtose")
print(curtose)
```

Criamos uma função que plota um histograma de cada um dos índices que serão analisados

```
In [ ]: # plota o gráfico histograma
def plt_hist(df, coluna, title):
    """
    Plota um histograma

    input
    df: dataframe
    coluna: Coluna para ser plotada
    title: título do gráfico
    """
    plt.figure(figsize=(12,8))
    plt.title(title, fontsize=18)
    plt.xlabel(coluna, fontsize=16)
    plt.ylabel("Frequência", fontsize=16)

    plt.hist(df[coluna])

    plt.show()
```

Plotamos um histograma com os dados de cada índice na escala original

```
In [ ]: plt_hist(df, "ebitda", "conjunto de dados na escala original")

In [ ]: plt_hist(df, "lcro_lqdo", "conjunto de dados na escala original")

In [ ]: plt_hist(df, "rcta_lqda", "conjunto de dados na escala original")

In [ ]: plt_hist(df, "ptrmo_lqdo", "conjunto de dados na escala original")

In [ ]: plt_hist(df, "diferenca", "conjunto de dados na escala original")
```

Agora vamos padronizar os dados fazendo a sua normalização usando a função StandardScaler da biblioteca SKLEARN para transformá-los e gerar novos histogramas para comparar

```
In [ ]: colunas = df.columns

In [ ]: # Padronizando os dados
scaler_standard = StandardScaler()
datapad = pd.DataFrame(scaler_standard.fit_transform(df), columns=colunas)
datapad
```

```
In [ ]: plt_hist(datapad, "ebitda", "conjunto de dados padronizado")
```

```
In [ ]: plt_hist(datapad, "lcro_lqdo", "conjunto de dados padronizado")
```

```
In [ ]: plt_hist(datapad, "rcta_lqda", "conjunto de dados padronizado")
```

```
In [ ]: plt_hist(datapad, "ptrmo_lqdo", "conjunto de dados padronizado")
```

```
In [ ]: plt_hist(datapad, "diferenca", "conjunto de dados padronizado")
```

A padronização dos dados usando StandardScaler para transformá-los preservou a sua variância, como podemos observar nas plotagens, porém gerou, mesmo para atributos que tinham todos as linhas com valor positivo, valores negativos, não preservando a sua variância, o que não é interessante para a busca de correlação.

Agora vamos normalizar os dados e gerar novos histogramas para comparar

Fazemos uma nova padronização para normalizar os dados usando a função MaxAbsScaler da biblioteca sklearn porque gera valores entre -1 e 1, preservando a escala dos valores e demais características dos dados.

Como os dados tinham valores negativos e positivos isso é importante para não perder essa referência.

```
In [ ]: # normalizando os dados
scaler_MaxAbsScaler = MaxAbsScaler()
datanor = pd.DataFrame(scaler_MaxAbsScaler.fit_transform(df), columns=colunas)
datanor
```

```
In [ ]: plt_hist(datanor, "ebitda", "conjunto de dados normalizado")
```

```
In [ ]: plt_hist(datanor, "lcro_lqdo", "conjunto de dados normalizado")
```

```
In [ ]: plt_hist(datanor, "rcta_lqda", "conjunto de dados normalizado")
```

```
In [ ]: plt_hist(datanor, "ptrmo_lqdo", "conjunto de dados normalizado")
```

```
In [ ]: plt_hist(datanor, "diferenca", "conjunto de dados normalizado")
```

Após a normalização dos dados geramos os diagramas de dispersão de todos os atributos normalizados para verificar a sua distribuição

```
In [ ]: def ecdf(data):
    """ Calcula a distribuição empírica """
    x = np.sort(data)
    n = x.size
    y = np.arange(1, n+1) / n
    return(x,y)
r,s = ecdf(datanor['ebitda'])
plt.scatter(x = r, y = s);
```

```
In [ ]: def ecdf(data):
    """ Calcula a distribuição empírica """
    x = np.sort(data)
    n = x.size
    y = np.arange(1, n+1) / n
    return(x,y)
r,s = ecdf(datanor['lcro_lqdo'])
plt.scatter(x = r, y = s);
```

```
In [ ]: def ecdf(data):
    """ Calcula a distribuição empírica """
    x = np.sort(data)
    n = x.size
    y = np.arange(1, n+1) / n
    return(x,y)
r,s = ecdf(datanor['rcta_lqda'])
plt.scatter(x = r, y = s);
```

```
In [ ]: def ecdf(data):
    """ Calcula a distribuição empírica """
    x = np.sort(data)
    n = x.size
    y = np.arange(1, n+1) / n
    return(x,y)
r,s = ecdf(datanor['ptrmo_lqdo'])
plt.scatter(x = r, y = s);
```

```
In [ ]: def ecdf(data):
    """ Calcula a distribuição empírica """
    x = np.sort(data)
    n = x.size
    y = np.arange(1, n+1) / n
    return(x,y)
r,s = ecdf(datanor['diferenca'])
plt.scatter(x = r, y = s);
```

Os diagramas de dispersão de todos os atributos demonstram que os dados não seguem uma distribuição linear.

Vamos calcular a correlação existente entre o índice EBITDA e a diferença entre as médias das cotações dos 5 dias antes da publicação do balanço e a média das cotações dos 5 dias depois da publicação, usando método de Pearson e o método de Spearman

```
In [ ]: # Calculando a correlação de Pearson para DataFrame "datanor" com os dados normalizados EBITDA e DIFERENCA
coef_pearson, p_valor = pearsonr(datanor['ebitda'], datanor['diferenca'])

print(coef_pearson)
print(p_valor)
# output: 1.0
```

A correlação existente entre as duas informações é baixa e negativa pelo método de Pearson

```
In [ ]: # Calculando a correlação de Spearman para DataFrame "datanor" com os dados normalizados EBITDA e DIFERENCA
coef_spearman, p_valor = spearmanr(datanor['ebitda'], datanor['diferenca'])

print(coef_spearman)
print(p_valor)
# output: 0.99999999
```

A correlação existente entre as duas informações é baixa e positiva pelo método de Spearman.

Como não encontramos uma correlação entre o EBITDA e a diferença entre a média das cotações anteriores e posteriores à publicação do balanço das empresas, vamos verificar a correlação entre todos os atributos de informações obtidas no balanço das empresas e o atributo diferença.

Vamos calcular e plotar o Mapa de Correlação usando o método de Pearson para todos os atributos

```
In [ ]: # Heatmap
import seaborn as sns
plt.figure(figsize = (7,7))
sns.heatmap(datanor.corr("pearson"), annot = True, cmap = "YlGnBu")
plt.title("Mapa de Correlação dos atributos usando método de Pearson\n", fontsize = 15)
plt.show()
```

Concluimos que, usando o método de Pearson para o cálculo da correlação entre os atributos com dados obtidos no balanço das empresas e o atributo diferença, não foi encontrada nenhuma correlação significativa, sendo que a maior foi com relação ao atributo Lucro Líquido, negativa com valor de -0,42.

Agora vamos plotar o Mapa de Correlação usando o método de Spearman

```
In [ ]: # Heatmap
import seaborn as sns
plt.figure(figsize = (7,7))
sns.heatmap(datanor.corr("spearman"), annot = True, cmap = "YlGnBu")
plt.title("Mapa de Correlação dos atributos usando método de Spearman\n", fontsize = 15)
plt.show()
```

Concluimos que, usando o método de Spearman para o cálculo da correlação entre os atributos com dados obtidos no balanço das empresas e o atributo diferença, não foi encontrada nenhuma correlação significativa, sendo que a maior foi com relação ao atributo Receita Líquida, positiva com valor de 0,18.

MBI em Inteligência Artificial e Big Data

Preparação dos dados para o estudo

FIM

ANEXO A – Layout do Histórico de Cotações de Ações na B3

MODREF - MOEDA DE REFERÊNCIA	MOEDA USADA NA DATA DO PREGÃO	X(04)	53	56
PREABE - PREÇO DE ABERTURA DO PAPEL MERCADO NO PREGÃO		(11)V99	57	69
PREMAX - PREÇO MÁXIMO DO PAPEL MERCADO NO PREGÃO		(11)V99	70	82
PREMIN - PREÇO MÍNIMO DO PAPEL MERCADO NO PREGÃO		(11)V99	83	95

3.2 REGISTRO - 01 - COTAÇÕES HISTÓRICAS POR PAPEL-MERCADO

NOME DO CAMPO / DESCRIÇÃO	CONTEÚDO	TIPO E TAMANHO	POS. INIC.	POS. FINAL
TIPREG - TIPO DE REGISTRO	FIXO "01"	N(02)	01	02
DATA DO PREGÃO	FORMATO "AAAAMMDD"	N(08)	03	10
CODBDI - CÓDIGO BDI	UTILIZADO PARA CLASSIFICAR OS PAPÉIS NA EMISSÃO DO BOLETIM DIÁRIO DE INFORMAÇÕES VER TABELA ANEXA	X(02)	11	12
CODNEG - CÓDIGO DE NEGOCIAÇÃO DO PAPEL		X(12)	13	24
TPMERC - TIPO DE MERCADO	CÓD. DO MERCADO EM QUE O PAPEL ESTÁ CADASTRADO VER TABELA ANEXA	N(03)	25	27
NOMRES - NOME RESUMIDO DA EMPRESA EMISSORA DO PAPEL		X(12)	28	39
ESPECI - ESPECIFICAÇÃO DO PAPEL	VER TABELA ANEXA	X(10)	40	49
PRAZOT - PRAZO EM DIAS DO MERCADO A TERMO		X(03)	50	52

NOME DO CAMPO / DESCRIÇÃO	CONTEÚDO	TIPO TAMANHO E	POS. INIC.	POS. FINAL
PREMED - PREÇO MÉDIO DO PAPEL-MERCADO NO PREGÃO		(11)V99	96	108
PREULT - PREÇO DO ÚLTIMO NEGÓCIO DO PAPEL-MERCADO NO PREGÃO		(11)V99	109	121
PREOFC - PREÇO DA MELHOR OFERTA DE COMPRA DO PAPELMERCADO		(11)V99	122	134
PROFV - PREÇO DA MELHOR OFERTA DE VENDA DO PAPELMERCADO		(11)V99	135	147
TOTNEG - NEG. - NÚMERO DE NEGÓCIOS EFETUADOS COM O PAPEL- MERCADO NO PREGÃO		N(05)	148	152
QUATOT - QUANTIDADE TOTAL DE TÍTULOS NEGOCIADOS NESTE PAPEL- MERCADO		N(18)	153	170
VOLTOT - VOLUME TOTAL DE TÍTULOS NEGOCIADOS NESTE PAPEL- MERCADO		(16)V99	171	188
PREEXE - PREÇO DE EXERCÍCIO PARA O MERCADO DE OPÇÕES OU VALOR DO CONTRATO PARA O MERCADO DE TERMO SECUNDÁRIO		(11)V99	189	201
INDOPC - INDICADOR DE CORREÇÃO DE PREÇOS DE EXERCÍCIOS OU VALORES DE CONTRATO PARA OS MERCADOS DE OPÇÕES OU TERMO SECUNDÁRIO	VER TABELA ANEXA	N(01)	202	202

NOME DO CAMPO / DESCRIÇÃO	CONTEÚDO	TIPO TAMANH O	POS. INIC.	POS. FINAL
DATVEN - DATA DO VENCIMENTO PARA OS MERCADOS DE OPÇÕES OU TERMO SECUNDÁRIO	FORMATO “AAAAMMDD”	N(08)	203	210
FATCOT - FATOR DE COTAÇÃO DO PAPEL	‘1’ = COTAÇÃO UNITÁRIA ‘1000’ = COTAÇÃO POR LOTE DE MIL AÇÕES	N(07)	211	217
PTOEXE - PREÇO DE EXERCÍCIO EM PONTOS PARA OPÇÕES REFERENCIADAS EM DÓLAR OU VALOR DE CONTRATO EM PONTOS PARA TERMO SECUNDÁRIO	PARA OS REFERENCIADOS EM DÓLAR, CADA PONTO EQUIVALE AO VALOR, NA MOEDA CORRENTE, DE UM CENTÉSIMO DA TAXA MÉDIA DO DÓLAR COMERCIAL INTERBANCÁRIO DE FECHAMENTO DO DIA ANTERIOR, OU SEJA, 1 PONTO = 1/100 US\$	(07)V06	218	230
CODISI - CÓDIGO DO PAPEL NO SISTEMA ISIN OU CÓDIGO INTERNO DO PAPEL	CÓDIGO DO PAPEL NO SISTEMA ISIN A PARTIR DE 15-05-1995	X(12)	231	242
DISMES - NÚMERO DE DISTRIBUIÇÃO DO PAPEL	NÚMERO DE SEQÜÊNCIA DO PAPEL CORRESPONDENTE AO ESTADO DE DIREITO VIGENTE	9(03)	243	245

