

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DE MINAS GERAIS (IFMG)
CAMPUS BAMBUÍ
BACHARELADO EM ENGENHARIA DE COMPUTAÇÃO

Vinícius Tadeu Andrade Costa

**ANÁLISE E INTEGRAÇÃO DE DADOS DO MERCADO FINANCEIRO BRASILEIRO
DISPONIBILIZADOS PELA B3 E CVM**

VINÍCIUS TADEU ANDRADE COSTA

**ANÁLISE E INTEGRAÇÃO DE DADOS DO MERCADO FINANCEIRO BRASILEIRO
DISPONIBILIZADOS PELA B3 E CVM**

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) – *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Marcos Roberto Ribeiro

Vinícius Tadeu Andrade Costa

**ANÁLISE E INTEGRAÇÃO DE DADOS DO MERCADO FINANCEIRO BRASILEIRO
DISPONIBILIZADOS PELA B3 E CVM**

Trabalho de conclusão de curso apresentado ao Curso de Bacharelado em Engenharia de Computação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG) – *Campus* Bambuí para obtenção do grau de Bacharel em Engenharia de Computação.

Aprovado em 18 de Janeiro de 2025 pela banca examinadora:

Marcos Roberto Ribeiro – IFMG – *Campus* Bambuí – (Orientador)

Dedico este trabalho à minha esposa e filhos, incentivadores e fontes inesgotáveis de apoio, amor e compreensão.

AGRADECIMENTOS

Agradeço a toda à minha família, esposa, filhos, pais e minha irmã, por acreditarem em mim e pelo incentivo constante na realização deste trabalho.

Agradeço à minha orientadora, ao meu coorientador e a todos que contribuíram de alguma forma para a realização deste trabalho.

“A tarefa mais importante de uma pessoa que vem ao mundo é criar algo.”
Paulo Freire

RESUMO

Este trabalho é um modelo em \LaTeX utilizando a classe $\text{IFT}_{\text{E}}\text{X}2024$. Tal classe foi desenvolvida com base no manual de normalização de trabalhos acadêmicos do IFMG e nas normas relacionadas da Associação Brasileira de Normas Técnicas. Este modelo apresenta uma estrutura básica com exemplos de elementos pré e pós-textuais. Maiores informações sobre como utilizar a classe podem ser encontradas no manual da classe $\text{IFT}_{\text{E}}\text{X}2024$.

Palavras-chave: $\text{IFT}_{\text{E}}\text{X}2024$. Modelo. IFMG. \LaTeX .

ABSTRACT

This work is a template in \LaTeX using the `IFT \LaTeX 2024class`. This class was developed based on the academic work standardization manual of IFMG and the related norms of the Brazilian Association of Technical Standards. This template presents a basic structure with examples of pre and post-textual elements. Further information on how to use the class can be found in the `IFT \LaTeX 2024class` manual.

Keywords: `IFT \LaTeX 2024`. Template. IFMG. \LaTeX .

LISTA DE FIGURAS

Figura 1 - Esquema lógico do banco de dados para integração de dados da CVM.	23
Figura 2 - Fluxograma do projeto	31

LISTA DE QUADROS

Quadro 1 - Especificações do computador utilizado no TCC	32
Quadro 2 - Comparativo entre conjuntos de dados da CVM	42

LISTA DE TABELAS

LISTA DE SIGLAS

IFMG – Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

TCC – Trabalho de conclusão de curso

B3 – Brasil, Bolsa e Balcão

CVM – Comissão de Valores Mobiliários

DFP – Demonstrações Financeiras Padronizadas

ITR – Formulário de Informações Trimestrais

IPE – Periódicos e Eventuais

FRE – Formulário de Referência

LISTA DE SÍMBOLOS

\mathbb{X} – Variável X

\mathbb{R} – Conjunto dos números reais

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Justificativa	15
1.2	Objetivos	15
1.3	Resultados Esperados	16
2	FUNDAMENTOS TEÓRICOS	17
2.1	Mercado financeiro	17
2.2	Análises para investimento	19
2.2.1	<i>Indicadores de rentabilidade</i>	20
2.2.2	<i>Indicadores de Liquidez</i>	21
2.2.3	<i>Indicadores de Endividamento</i>	21
2.2.4	<i>Valor intrínseco</i>	22
2.3	Modelagem e integração de dados	22
2.3.1	<i>Modelagem dos dados e estruturação lógica</i>	22
2.3.2	<i>Processo de integração de dados</i>	24
2.4	Estado da arte	25
2.4.1	<i>Trabalhos acadêmicos</i>	26
2.4.2	<i>Projetos e plataformas similares</i>	26
2.4.3	<i>Diferencial do trabalho proposto</i>	27
3	METODOLOGIA	29
3.1	Classificação da Pesquisa	29
3.2	Gerenciamento do Projeto	30
3.3	Solução proposta	30

3.4	Materiais e Tecnologias	31
4	DESENVOLVIMENTO	34
4.1	Análise inicial dos dados da CVM	34
4.1.1	<i>Informação cadastral</i>	35
4.1.2	<i>Formulário cadastral (FCA)</i>	35
4.1.3	<i>Periódicos e Eventuais (IPE)</i>	36
4.1.4	<i>Formulário de Referência (FRE)</i>	38
4.1.5	<i>Valores mobiliários negociados e detidos (VLMO)</i>	39
4.1.6	<i>Demonstrativos Financeiros Padronizados</i>	40
4.1.6.1	Informações Trimestrais (ITR)	40
4.1.6.2	Formulário de Demonstrações Financeiras Padronizadas (DFP)	41
4.1.7	<i>Informe do código de governança (ICBGC)</i>	41
4.2	Modelagem e Estruturação dos Dados	43
4.3	Software	44
	REFERÊNCIAS	45
	GLOSSÁRIO	52
	APÊNDICES	53
	APÊNDICE A - Código para baixar os dados da CVM	54
	APÊNDICE B - Código para extrair os dados da CVM	57
	APÊNDICE C - Mapeamento Completo dos Dados da CVM	59
	ÍNDICE	66

1 INTRODUÇÃO

O mercado financeiro brasileiro tem passado por transformações significativas impulsionadas pelo avanço tecnológico e pela crescente demanda por soluções que automatizem a integração e a análise de dados (Dantas, 2020). Esse movimento reflete a busca constante por maior eficiência e acessibilidade nas informações financeiras, fundamentais para investidores e analistas que utilizam a análise fundamentalista como base para a tomada de decisões.

O crescimento desse interesse pode ser observado no relatório anual de 2023 da Brasil, Bolsa e Balcão (B3), que apontou um aumento de aproximadamente 15% no número de investidores em comparação com 2022 (B3, 2023a). Esse cenário reforça a necessidade de garantir o acesso simplificado e estruturado às informações financeiras das empresas, promovendo maior eficiência nas análises e decisões de mercado.

No Brasil, dois órgãos desempenham papéis centrais na disponibilização de dados do mercado de capitais: a B3 e a Comissão de Valores Mobiliários (CVM). A B3, fundada em 1890 e sediada em São Paulo, é a única bolsa de valores do país e fornece dados sobre o histórico de negociações de ativos (B3, 2023c).

Já a CVM, criada em 1976, é a entidade responsável por regulamentar e supervisionar o mercado de capitais brasileiro, disponibilizando publicamente uma ampla gama de informações contábeis e financeiras das empresas listadas na bolsa (CVM, 2009). Essas informações são essenciais para investidores que utilizam a análise fundamentalista, metodologia que avalia a saúde financeira e o desempenho das empresas com base em seus demonstrativos financeiros e outros indicadores econômicos.

Embora a CVM disponibilize esses dados publicamente, seu formato dificulta a manipulação por usuários sem conhecimento técnico avançado. Muitos estão em arquivos complexos que exigem processamento adicional, criando um obstáculo para investidores e analistas que necessitam de informações ágeis e acessíveis. Essa barreira técnica restringe o acesso à informação e limita a capacidade analítica de boa parte do mercado.

A necessidade de soluções mais acessíveis já foi abordada por estudos acadêmicos. (Araujo Gonçalves; Togneri, 2021a) propõem um modelo de dados flexível para análise fundamentalista moderna, estruturando balanço patrimonial, demonstração de resultados e fluxo de caixa em um banco de dados não relacional (MongoDB). Esse tipo de abordagem reforça a importância de alternativas que simplifiquem o acesso e a análise dos dados financeiros.

Diante desse contexto, este trabalho propõe o desenvolvimento de um sistema automatizado para a integração e análise dos dados fornecidos pela CVM, facilitando

tando seu acesso e processamento. A solução busca oferecer uma ferramenta capaz de extrair e estruturar essas informações de forma eficiente, contribuindo para a análise fundamentalista e auxiliando na tomada de decisões. Além disso, pretende otimizar o processo de análise, tornando-o mais acessível e organizado para investidores, analistas e pesquisadores. Com essa abordagem, espera-se reduzir a complexidade no manuseio dos dados e aprimorar a compreensão do mercado.

1.1 Justificativa

A B3 apresentou um crescimento significativo no volume de negócios e de capital movimentado ao longo dos anos. Em 2023, o volume total negociado na B3 alcançou R\$ 7,2 trilhões, representando um aumento de 12,6% em relação a 2022 (B3, 2023c). Além disso, o montante total de dinheiro movimentado na B3 em 2023 foi de R\$ 2,4 trilhões, um aumento de 15,5% em relação ao ano anterior (B3, 2023c).

Embora os dados da B3 não estejam diretamente presentes neste trabalho, seu contexto é fundamental, uma vez que ela representa o principal ambiente de negociação de ativos no Brasil. As informações que embasam os negócios realizados na B3, especialmente aquelas relacionadas às demonstrações financeiras e dados cadastrais das empresas listadas, estão disponíveis na CVM. Dessa forma, a integração dos dados da CVM serve como uma base fundamental para análises que indiretamente impactam o entendimento do mercado da B3.

Atualmente, as informações da CVM estão disponíveis em formatos específicos, fragmentados e sem integração adequada, o que dificulta o acesso e a análise automatizada dos dados. Até o momento, não foram encontrados trabalhos que promovam essa integração de forma estruturada e com disponibilização pública.

A proposta de integrar os dados da CVM visa criar uma base de dados unificada e acessível, que sirva como suporte para futuros trabalhos acadêmicos e aplicações práticas em algoritmos de análise de mercado financeiro. Com isso, espera-se facilitar o desenvolvimento de estudos e ferramentas que contribuam para uma melhor compreensão do mercado de capitais brasileiro (Lindman *et al.*, 2020)

1.2 Objetivos

Este estudo tem como objetivo principal o desenvolvimento de um sistema automatizado capaz de integrar e facilitar o acesso aos dados financeiros disponibilizados publicamente pela CVM. A proposta busca contribuir com a análise fundamentalista e apoiar o processo de tomada de decisão por parte de investidores, pesquisadores e demais interessados no mercado financeiro.

Para atingir esse objetivo, são estabelecidas as seguintes metas específi-

cas:

- analisar a estrutura dos dados públicos fornecidos pela CVM, identificando suas fontes, formatos e padrões;
- propor uma modelagem lógica que permita a integração eficiente e automatizada desses dados;
- projetar e implementar um sistema que realize a extração, o processamento e a disponibilização dos dados de forma acessível e estruturada.

1.3 Resultados Esperados

Espera-se que a ferramenta desenvolvida neste trabalho integre de forma eficiente os dados financeiros públicos disponibilizados pela CVM, permitindo a criação de uma base de dados unificada, estruturada e acessível.

Essa base será construída com foco em um recorte histórico que contemple, no mínimo, o período a partir de 2019, o que proporcionará maior profundidade nas análises e permitirá identificar padrões e variações ao longo do tempo. Tal histórico é essencial para estudos mais robustos e comparativos dentro do contexto da análise fundamentalista.

Embora os dados da B3 não sejam utilizados diretamente neste trabalho, o sistema proposto pode contribuir indiretamente para análises relacionadas à bolsa de valores, uma vez que as informações integradas da CVM servem como base para decisões tomadas por investidores e instituições que atuam no ambiente da B3.

A disponibilização pública dessa base de dados tem o potencial de promover maior transparência, acessibilidade e democratização das informações financeiras no Brasil, beneficiando pesquisadores, investidores e demais interessados no mercado de capitais.

2 FUNDAMENTOS TEÓRICOS

Este capítulo apresenta os fundamentos teóricos que sustentam o desenvolvimento deste trabalho, bem como os principais estudos e iniciativas relacionadas ao tema. A Seção 2.1 aborda o mercado financeiro brasileiro, com ênfase na atuação da CVM e sua função reguladora. A Seção 2.2 discute as diferentes abordagens de análise para investimentos, com foco na análise fundamentalista. A Seção 2.3 descreve a modelagem dos dados e os processos de integração utilizados no projeto. Por fim, a Seção 2.4 apresenta trabalhos acadêmicos e iniciativas similares à proposta deste estudo, destacando seus principais diferenciais.

2.1 Mercado financeiro

O mercado financeiro é um sistema que engloba uma variedade de instituições, instrumentos e atividades relacionadas à gestão de recursos financeiros. Ele abrange diferentes segmentos, incluindo o mercado de capitais, o mercado monetário, o mercado cambial e o mercado de derivativos. O principal propósito do mercado financeiro é facilitar a alocação eficiente de recursos entre poupadores e investidores, fornecendo oportunidades de investimento e financiamento para empresas, governos e indivíduos (Teixeira, 2019; Damodaran, 2012).

No Brasil, o mercado financeiro é composto pelos segmentos de mercado de capitais, mercado monetário, mercado de crédito, mercado de câmbio e mercado de derivativos. O mercado de capitais envolve a negociação de ações, debêntures e outros ativos mobiliários. Esse segmento é fundamental para o financiamento das empresas, permitindo que elas captem recursos por meio da emissão de títulos negociáveis no mercado financeiro (Suno, 2021).

O mercado monetário é o espaço onde ocorrem operações de curto prazo entre instituições financeiras e o Banco Central. Essas operações têm como objetivo controlar a liquidez da economia e a taxa de juros, garantindo o equilíbrio do sistema financeiro (Figueiredo *et al.*, 2021).

O mercado de crédito destina-se ao financiamento de empresas e pessoas físicas por meio de empréstimos e financiamentos bancários. Esse segmento é essencial para a dinamização da economia, permitindo investimentos e consumo a partir da concessão de crédito.

O mercado de câmbio é responsável pelas transações de compra e venda de moedas estrangeiras. Sua função principal é viabilizar o comércio internacional, investimentos estrangeiros e a gestão de reservas cambiais (Góis; Soares, 2019).

O mercado de derivativos compreende a negociação de contratos futuros, opções e *swaps*, instrumentos utilizados para a gestão de riscos financeiros. Esse

segmento permite que investidores e empresas se protejam contra variações adversas nos preços de ativos e taxas de juros (Figueiredo, 2023).

Dentro do mercado financeiro, as ações representam uma das principais modalidades de investimento. Uma ação é uma fração do capital social de uma empresa, conferindo ao seu detentor a propriedade de uma parcela da empresa emissora. Ao adquirir ações, o investidor se torna um sócio da empresa e pode ter direito a participação nos lucros distribuídos e, em alguns casos, a voto nas assembleias de acionistas (Suno, 2021; Gomes, 2007).

A negociação de ações ocorre principalmente em bolsas de valores, como a B3 no Brasil, onde investidores podem comprar e vender ativos. O preço das ações é determinado pela oferta e demanda, ou seja, pela interação entre compradores e vendedores no mercado (Santander, 2024). Além disso, diversos fatores influenciam a valorização ou desvalorização dos papéis, como os resultados financeiros das empresas, a conjuntura econômica e eventos geopolíticos (Damodaran, 2012; Attie, 2013).

Duas instituições fundamentais nesse ecossistema são a B3 e a CVM. A B3 atua como principal plataforma de negociação de ativos financeiros, enquanto a CVM desempenha um papel regulador crucial, garantindo transparência e equidade no mercado de capitais (CVM, 2023a).

Criada pela Lei nº 6.385/1976 (Brasil, 1976), a CVM tem como principal missão proteger os investidores e garantir o funcionamento eficiente do mercado de capitais (CVM, 2009). Entre suas principais atribuições estão (CVM, 2009; Roberto, 2023):

- regulamentação da abertura de capital de empresas e suas obrigações com o mercado;
- supervisão de corretoras, bancos de investimento e outras instituições financeiras;
- fiscalização da negociação de ativos financeiros e combate a fraudes e manipulações de mercado;
- garantia da transparência e divulgação de informações pelas companhias de capital aberto.

A CVM exige que todas as empresas listadas na bolsa de valores publiquem periodicamente suas demonstrações financeiras e outros documentos essenciais, como balanços patrimoniais e informações sobre acionistas relevantes. Essa transparência fortalece a confiança dos investidores e contribui para a estabilidade e o crescimento do mercado financeiro brasileiro (FGV, 2024).

Dessa forma, a regulamentação e a supervisão do mercado financeiro brasileiro desempenham um papel essencial na proteção dos investidores e no desenvolvimento sustentável do setor (Figueiredo, 2023).

Para garantir a ampla disponibilidade dessas informações, a CVM mantém

uma base de dados acessível ao público por meio de seu portal eletrônico¹ e sistemas especializados, como o Sistema Empresas.NET e o CVMWeb. As empresas registradas são obrigadas a enviar seus documentos periodicamente, os quais são disponibilizados em diversos formatos, como PDF, XML e XBRL, permitindo tanto a consulta manual quanto a análise automatizada (CVM, 2009). Essa estrutura facilita o acompanhamento do mercado por investidores, analistas e demais interessados, promovendo maior transparência e acesso à informação.

A base de dados da CVM inclui:

- demonstrações financeiras padronizadas (DFPs);
- informes trimestrais (ITRs);
- relatórios de governança corporativa;
- documentos de ofertas públicas de ações (IPO);
- dados sobre fundos de investimento e gestores de ativos.

2.2 Análises para investimento

A análise de investimentos busca avaliar oportunidades de alocação de capital, utilizando diferentes abordagens para identificar ativos com potencial de retorno (Liaw, 2011). Entre essas abordagens, destacam-se a análise técnica, a análise de sentimento, a análise quantitativa, a análise macroeconômica e a análise fundamentalista.

A análise técnica baseia-se no estudo de padrões gráficos e estatísticos do histórico de preços dos ativos. Os analistas técnicos utilizam indicadores como médias móveis, bandas de Bollinger e índices de força relativa para identificar tendências e pontos de compra ou venda no mercado (Omane-Adjepong; Ababio; Alagidede, 2019).

A análise de sentimento considera o impacto de notícias, redes sociais e opiniões de investidores sobre os ativos financeiros. Essa abordagem busca medir o comportamento do mercado a partir da análise de textos, discursos e volumes de menções relacionadas a determinados ativos, utilizando técnicas de processamento de linguagem natural e aprendizado de máquina (Kearney; Liu, 2014).

A análise quantitativa utiliza modelos matemáticos e estatísticos para prever movimentos de mercado e identificar oportunidades de investimento. Esse método frequentemente emprega algoritmos computacionais, regressões estatísticas e aprendizado de máquina para encontrar correlações e padrões em grandes volumes de dados financeiros (Sahu; Mokhade; Bokde, 2023).

A análise macroeconômica avalia fatores econômicos amplos, como inflação, Produto Interno Bruto (PIB), taxas de juros e política monetária. Essa abordagem busca compreender como variáveis macroeconômicas influenciam os mercados

¹ <https://dados.cvm.gov.br>.

financeiros e os preços dos ativos, sendo fundamental para investidores institucionais e gestores de fundos (Claessens; Kose, 2017).

A análise fundamentalista foca nos fundamentos financeiros das empresas, examinando indicadores como receita, lucro, endividamento e governança corporativa. Esse método busca determinar o valor intrínseco de um ativo, comparando-o ao seu preço de mercado para identificar se está subvalorizado ou supervalorizado. No contexto deste estudo, a análise fundamentalista será o enfoque principal, dada sua relevância na avaliação detalhada do desempenho financeiro e estratégico das companhias.

A análise fundamentalista investiga os dados financeiros e operacionais das empresas para determinar seu valor intrínseco (Mathew, 2024). Para isso, utiliza diversos indicadores que auxiliam na avaliação do desempenho financeiro, da estrutura de capital e da liquidez da empresa. Esses indicadores podem ser agrupados em quatro categorias principais: indicadores de rentabilidade, indicadores de liquidez, indicadores de endividamento e valor intrínseco.

2.2.1 Indicadores de rentabilidade

Os indicadores de rentabilidade medem a eficiência da empresa em gerar lucros em relação a diferentes bases financeiras. Esses cálculos são amplamente utilizados em estudos financeiros e derivam da contabilidade gerencial e da análise de demonstrações financeiras (Kothari, 2001).

O lucro por ação (LPA) indica o montante de lucro líquido atribuído a cada ação ordinária em circulação da empresa. É um indicador essencial para avaliar o desempenho das ações no mercado. A fórmula é dada pela Equação (2.1a). Considerando um lucro líquido de R\$ 10 milhões e 2 milhões de ações em circulação, tem-se o LPA de R\$ 5,00, conforme a Equação (2.1b) (Omane-Adjepong; Ababio; Alagidede, 2019).

$$\text{LPA} = \frac{\text{Lucro líquido}}{\text{Número de ações}} \quad (2.1a) \quad \text{LPA} = \frac{10.000.000,00}{2.000.000,00} = 5,00 \quad (2.1b)$$

O índice preço por lucro (P/L) mede quanto os investidores estão dispostos a pagar pelo lucro da empresa. A fórmula está na Equação (2.2a). Se uma ação custa R\$ 50,00 e o LPA é R\$ 5,00, o P/L será 10,00, indicando que os investidores estão dispostos a pagar 10 vezes o lucro anual por ação (Sahu; Mokhade; Bokde, 2023).

$$\text{P/L} = \frac{\text{Preço da ação}}{\text{LPA}} \quad (2.2a) \quad \text{P/L} = \frac{50,00}{5,00} = 10,00 \quad (2.2b)$$

O *dividend yield* (DY) mede a rentabilidade dos dividendos pagos em relação ao preço da ação. Se a empresa paga R\$ 2,00 de dividendos por ação e o

preço da ação é R\$ 40,00, o cálculo da Equação (2.3b) resulta em um DY de 5,00% (Kearney; Liu, 2014).

$$DY = \frac{\text{Dividendo por ação}}{\text{Preço da ação}} \times 100 \quad (2.3a) \qquad DY = \frac{2,00}{40,00} \times 100 = 5,00\% \quad (2.3b)$$

2.2.2 Indicadores de Liquidez

Os indicadores de liquidez avaliam a capacidade da empresa de honrar suas obrigações de curto prazo (Claessens; Kose, 2017). A Liquidez Corrente (LC) é expressa na Equação (2.4a). Se o ativo circulante é de R\$ 500.000 e o passivo circulante de R\$ 250.000, a liquidez corrente é igual a 2,00, conforme a Equação (2.4b).

$$LC = \frac{\text{Ativo circulante}}{\text{Passivo circulante}} \quad (2.4a) \qquad LC = \frac{500.000}{250.000} = 2,00 \quad (2.4b)$$

2.2.3 Indicadores de Endividamento

Os indicadores de endividamento mostram a dependência da empresa em relação ao capital de terceiros (Kothari, 2001).

Um dos principais indicadores é a dívida bruta, que representa a soma do passivo circulante e do passivo não circulante da empresa, conforme definido na Equação (2.5a). A Equação (2.5a) expressa a fórmula geral da dívida bruta, que é composta pelos valores de obrigações de curto prazo (passivo circulante) e de longo prazo (passivo não circulante). Já a Equação (2.5b) apresenta um exemplo numérico, onde a dívida bruta totaliza R\$ 3.000.000, a partir da soma de R\$ 1.000.000 de passivo circulante e R\$ 2.000.000 de passivo não circulante.

$$\text{Dívida bruta} = \text{Passivo circulante} + \text{Passivo não circulante} \quad (2.5a)$$

$$\text{Dívida bruta} = 1.000.000 + 2.000.000 = 3.000.000 \quad (2.5b)$$

Por exemplo, se a dívida bruta é de R\$ 3.000.000 e o caixa e equivalentes de caixa totalizam R\$ 500.000, a dívida líquida será de R\$ 2.500.000, conforme indicado na Equação (2.6b). A Equação (2.6a) define a dívida líquida, que representa o valor da dívida efetiva da empresa, descontado o montante disponível em caixa ou em aplicações de liquidez imediata. Já a Equação (2.6b) ilustra esse conceito com

valores hipotéticos, demonstrando que, ao subtrair R\$ 500.000 disponíveis em caixa de uma dívida bruta de R\$ 3.000.000, obtém-se uma dívida líquida de R\$ 2.500.000.

$$\text{Dívida líquida} = \text{Dívida bruta} - \text{Caixa e equivalentes de caixa} \quad (2.6a)$$

$$\text{Dívida líquida} = 3.000.000 - 500.000 = 2.500.000 \quad (2.6b)$$

2.2.4 Valor intrínseco

O valor intrínseco é estimado pelo método de Desconto de Fluxo de Caixa (DCF), conforme a Equação (2.7a). Nessa equação, r representa a taxa de desconto e t o período de tempo. Com fluxo de caixa de R\$ 1.000.000 e taxa de desconto de 10%, o valor intrínseco resulta em R\$ 909.090,91, conforme demonstrado na Equação (2.7b). Esse cálculo reflete o valor presente do fluxo de caixa futuro (Figueiredo, 2023).

$$\text{Valor intrínseco} = \sum \frac{\text{Fluxo de caixa esperado}}{(1 + r)^t} \quad (2.7a)$$

$$\text{Valor intrínseco} = \frac{1.000.000}{(1 + 0,10)^1} = \frac{1.000.000}{1,10} = 909.090,91 \quad (2.7b)$$

2.3 Modelagem e integração de dados

A modelagem e integração de dados são etapas fundamentais para a construção de um sistema automatizado de análise fundamentalista. Neste trabalho, essas etapas têm o propósito de estruturar, consolidar e organizar os dados disponibilizados pela CVM, garantindo sua utilização eficiente e integrada. A seguir, são apresentadas as principais abordagens empregadas na modelagem dos dados, bem como os métodos utilizados para sua integração, assegurando a consistência, acessibilidade e interoperabilidade das informações no sistema desenvolvido.

2.3.1 Modelagem dos dados e estruturação lógica

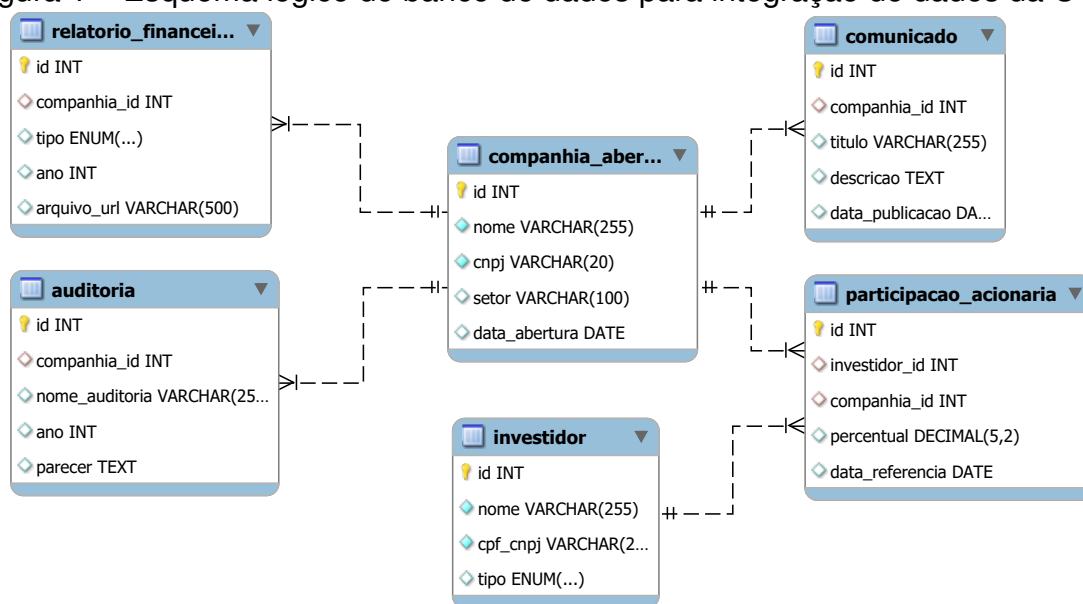
A modelagem de dados é uma etapa fundamental no desenvolvimento de sistemas que manipulam grandes volumes de informações, especialmente em contextos financeiros, onde a integridade, consistência e escalabilidade são requisitos indispensáveis (Elmasri; Navathe; Pinheiro *et al.*, 2005). Neste trabalho, adotou-se uma abordagem centrada na modelagem lógica, com o objetivo de estruturar uma base de

dados relacional robusta, capaz de integrar e armazenar dados públicos oriundos da CVM.

A modelagem lógica traduz os requisitos identificados no domínio do problema em uma estrutura formal que define, de maneira precisa, as tabelas, os atributos, os tipos de dados e as restrições de integridade necessárias para a representação das entidades envolvidas (Connolly; Begg, 2015). Diferente da modelagem conceitual, que geralmente se utiliza de representações como o Diagrama Entidade-Relacionamento (DER), a modelagem lógica já incorpora aspectos técnicos relacionados à implementação em um Sistema de Gerenciamento de Banco de Dados (SGBD), neste caso, o MySQL.

A Figura 1 apresenta o esquema lógico elaborado com base em práticas de engenharia reversa e análise das documentações e estruturas disponibilizadas pela CVM. Essa técnica permite extrair o modelo de dados a partir de fontes existentes, facilitando o redesenho e a normalização da base de dados (Scannapieco, 2006).

Figura 1 – Esquema lógico do banco de dados para integração de dados da CVM.



Fonte: Elaborado pelo autor, 2024.

O modelo proposto é composto pelas seguintes tabelas:

- *companhia_aberta*
- *comunicado*
- *investidor*
- *participacao_acionaria*
- *auditoria*
- *relatorio_financeiro*

A tabela *companhia_aberta* centraliza os dados cadastrais das empresas registradas na CVM, contendo atributos como *nome*, *cnpj*, *setor* e *data_abertura*. O

campo *cnpj* possui uma restrição de unicidade (*UNIQUE*) para garantir a integridade dos dados identificadores.

A tabela *comunicado* armazena registros de publicações oficiais vinculadas às companhias, sendo referenciada pela chave estrangeira *companhia_id*, que estabelece uma relação com a tabela *companhia_aberta*. Esse vínculo viabiliza a rastreabilidade entre os comunicados e as empresas emissores.

A tabela *investidor* registra os agentes que detêm participações acionárias, identificados pelo atributo *cpf_cnpj*, e categorizados segundo o tipo de investidor (*Pessoa Física* ou *Pessoa Jurídica*), utilizando a estrutura *ENUM* para controle de domínio.

A tabela *participacao_acionaria* representa a associação entre investidores e companhias abertas, permitindo o registro da proporção de ações detidas por data de referência. Essa tabela possui chaves estrangeiras para as tabelas *investidor* e *companhia_aberta*, o que possibilita a análise de composição societária ao longo do tempo.

A tabela *auditoria* documenta os pareceres emitidos por empresas auditoras sobre as demonstrações financeiras das companhias, detalhando o ano do parecer, o nome da auditoria e a descrição do conteúdo emitido.

Por fim, a tabela *relatorio_financeiro* concentra os metadados dos arquivos financeiros disponibilizados pela CVM, como os Demonstrativos Financeiros Padronizados (DFP), Informações Trimestrais (ITR), o Formulário de Referência e o IAN. Cada registro inclui o tipo do documento, o ano de referência e o link para o arquivo original.

Essa estrutura relacional foi projetada visando à normalização dos dados, conforme as boas práticas de projeto de bancos de dados relacionais (Elmasri; Navathe, 2016). A modelagem lógica aqui apresentada oferece a base necessária para uma integração eficiente e escalável, compatível com análises automatizadas de caráter fundamentalista. Sua implementação favorece a consistência da base de dados, o desempenho nas consultas SQL e a reutilização de componentes em diferentes módulos do sistema.

2.3.2 Processo de integração de dados

A integração de dados é um procedimento essencial que consolida informações oriundas de diversas fontes, garantindo consistência, qualidade e acessibilidade para análises estratégicas e operacionais. Esse processo transforma dados brutos em informações valiosas para a tomada de decisão e é composto por etapas interdependentes: extração, transformação e carga.

Na etapa de extração, os dados são coletados diretamente de múltiplas fontes, como os repositórios da CVM. Essa fase utiliza técnicas que asseguram a captura completa e precisa das informações, minimizando riscos de perda ou corrupção.

No presente trabalho, essa etapa foi implementada por meio de *scripts* automatizados em Python, que realizam o *download* de arquivos compactados disponibilizados pela CVM, como o `dfp_cia_aberta_2022.zip`. O sistema verifica a existência prévia dos arquivos e sua integridade antes de prosseguir com o processamento, evitando redundâncias e falhas.

Após a extração, os dados brutos passam por um processo de transformação que envolve padronização, limpeza, normalização e aplicação de regras de negócio. Essa etapa é responsável por eliminar inconsistências, corrigir valores ausentes, ajustar formatos e estruturar os dados para análises mais sofisticadas. No contexto deste projeto, foram realizadas conversões de tipos, unificação de formatos de datas e remoção de duplicidades. Informações como o código da conta contábil (`CD_CONTA`) foram associadas às suas descrições completas com base nos planos de contas divulgados pela CVM, garantindo maior legibilidade e utilidade analítica.

Na etapa final, os dados transformados são carregados em um banco de dados relacional estruturado, visando facilitar a manipulação, a consulta e a análise posterior. Para este trabalho, optou-se pelo uso do SQLite como solução leve e eficiente, adequada ao escopo local da ferramenta desenvolvida. As tabelas foram criadas respeitando normas de integridade referencial, com chaves primárias e estrangeiras bem definidas. Dados financeiros, cadastrais e operacionais foram organizados em tabelas como `demonstrativo_financeiro`, `informacao_trimestral` e `empresas`, possibilitando análises temporais, cálculo de indicadores como LPA e P/L, entre outros.

A integração harmônica dessas etapas resultou em uma base de dados confiável e atualizada, que sustenta a automatização dos processos analíticos e permite a construção de modelos quantitativos e preditivos. Essa abordagem se mostra fundamental para a aplicação da análise fundamentalista no contexto do mercado financeiro, proporcionando uma visão estruturada e estratégica para a tomada de decisão (Costa; Ferreira; Resende, 2024; Halevy; Norvig; Pereira, 2006).

2.4 Estado da arte

Nesta seção, são analisados os principais estudos acadêmicos e projetos práticos que fundamentam a abordagem deste trabalho. Inicialmente, apresenta-se uma síntese dos trabalhos existentes, destacando seus objetivos, metodologias empregadas e limitações encontradas. Em seguida, evidencia-se como o presente estudo propõe uma integração inédita de dados e técnicas, superando as deficiências das abordagens anteriores.

2.4.1 Trabalhos acadêmicos

Os estudos acadêmicos voltados à modelagem e à análise de dados financeiros têm se concentrado na automatização da análise fundamentalista e na adaptação de modelos tradicionais ao contexto do mercado brasileiro. Por exemplo, Montoia (2021) desenvolvem um sistema automatizado para a análise de ações, enfatizando a melhoria na rapidez e na eficiência da tomada de decisões. Esse trabalho detalha a aplicação de algoritmos de processamento de dados que reduzem a intervenção manual, permitindo uma análise em tempo real dos indicadores financeiros.

De forma complementar, Araujo Gonçalves e Togneri (2021b) propôs um modelo flexível que adapta os conceitos da análise fundamentalista para refletir as particularidades dos dados nacionais, considerando variáveis específicas do mercado brasileiro. Esse estudo destaca a importância da customização dos parâmetros e a integração de variáveis macroeconômicas na modelagem financeira.

Além disso, Delalibera *et al.* (2023) aprimorou metodologias tradicionais, como o Modelo Rojo, para avaliação de ativos, enfatizando a precisão na quantificação de riscos e oportunidades de investimento. Este trabalho apresenta uma abordagem robusta que combina técnicas estatísticas com algoritmos de *machine learning* para melhorar a acurácia das previsões. Por sua vez, Reis *et al.* (2020) investigou a aplicação prática desses métodos na análise das ações negociadas, proporcionando *insights* sobre a volatilidade do mercado e a relevância de diferentes indicadores financeiros.

Outros estudos, como o de Vieira (2019), analisam estratégias de montagem de carteiras de ações, comparando a performance das carteiras com índices de mercado e ressaltando as dificuldades na diversificação e no balanceamento dos ativos. Da mesma forma, Freitas (2020) explorou a aplicação criteriosa de indicadores no setor bancário, evidenciando os desafios em medir a solidez financeira e a eficiência operacional de instituições financeiras.

2.4.2 Projetos e plataformas similares

No campo dos projetos práticos, diversas iniciativas presentes em repositórios, especialmente no GitHub, vêm se destacando por oferecer ferramentas e bases de dados voltadas para a extração e análise de informações financeiras. Por exemplo, o projeto de Paiva (2025) apresenta um sistema integrado que coleta, processa e organiza dados contábeis de empresas. Essa ferramenta tem como objetivo facilitar a aplicação da análise fundamentalista ao disponibilizar dados de forma estruturada e acessível, promovendo maior transparência e facilidade na análise dos balanços.

De maneira similar, Minas (2025) propõe uma abordagem inovadora para a

análise de balanços empresariais, enfatizando a importância da normalização dos dados e da criação de indicadores personalizados que se ajustem às especificidades de cada setor. Repositórios como os de Fontinele (2025) e Louredo (2025) se destacam ao desenvolver ferramentas específicas para interpretar demonstrações financeiras a partir de dados públicos, permitindo que o usuário identifique rapidamente pontos críticos e oportunidades de investimento. Contudo, grande parte dessas iniciativas se restringe à coleta e ao processamento dos dados, sem oferecer uma análise histórica contínua que possibilite a compreensão das tendências de longo prazo.

2.4.3 Diferencial do trabalho proposto

O diferencial do presente trabalho reside na integração de um histórico abrangente de dados financeiros com um sistema automatizado de acesso e análise das informações públicas disponibilizadas pela CVM. Embora os experimentos realizados tenham utilizado um recorte temporal de cinco anos, a ferramenta não se limita a esse intervalo. Enquanto a CVM mantiver a estrutura dos dados em seus repositórios, o sistema permanecerá funcional e poderá ser continuamente alimentado com novos dados, garantindo sua atualização e longevidade.

A proposta se distingue, ainda, por tratar-se de uma solução de código aberto, o que permite sua auditoria, personalização e extensão por outros pesquisadores e profissionais interessados. Essa abertura favorece a colaboração contínua, possibilitando o aprimoramento da ferramenta e sua adaptação a novos contextos analíticos ou setores econômicos específicos.

A abordagem adotada neste trabalho viabiliza tanto a avaliação contínua e comparativa dos dados quanto a análise contextualizada das informações. A avaliação contínua é sustentada pelo acúmulo de dados históricos, o que permite identificar tendências e padrões relevantes ao longo do tempo. Já a análise contextualizada emerge da capacidade de combinar dados históricos e atuais, ampliando a compreensão sobre ciclos econômicos, sazonalidades e eventos que impactam o mercado.

As principais inovações e contribuições do sistema desenvolvido podem ser resumidas nos seguintes pontos:

- integração de dados históricos financeiros em larga escala;
- processamento automatizado, com foco em escalabilidade e reusabilidade;
- contextualização dos indicadores contábeis e financeiros;
- manutenção da funcionalidade do sistema diante de novas atualizações da CVM, desde que preservada a estrutura de dados;
- disponibilização em formato de código aberto, fomentando a transparência e a colaboração.

Ao superar as limitações de iniciativas anteriores, que se restringem majoritariamente à coleta e estruturação pontual dos dados, a solução apresentada promove uma abordagem sistemática e extensível, capaz de subsidiar análises fundamentalistas robustas e atualizadas. Essa característica torna o sistema especialmente útil para aplicações acadêmicas, institucionais e profissionais voltadas à tomada de decisão baseada em fundamentos econômicos e financeiros.

3 METODOLOGIA

Esta seção apresenta a metodologia de execução do presente trabalho. A Seção 3.1 define o escopo e os objetivos deste trabalho, delineando as principais áreas de investigação. Em seguida, a Seção 3.2 descreve as abordagens e técnicas utilizadas para planejar e conduzir o desenvolvimento da ferramenta proposta. Por fim, a Seção 3.3 detalha os passos do trabalho, incluindo a investigação da estrutura dos dados, a modelagem lógica e o desenvolvimento da ferramenta de integração de dados.

3.1 Classificação da Pesquisa

Para fundamentar a classificação metodológica da pesquisa, foram consultadas as obras de Cervo e Bervian (1983), que apresentam diferentes abordagens científicas, e de Gerhardt e Silveira (2009), que tratam de métodos amplamente utilizados na pesquisa acadêmica.

Quanto à natureza, esta pesquisa é aplicada, pois busca resolver um problema prático relacionado à integração automatizada de dados financeiros públicos disponibilizados pela CVM, promovendo avanços na análise fundamentalista e no apoio à tomada de decisão.

Em relação aos objetivos, a investigação é exploratória e descritiva. A parte exploratória está ligada à compreensão da estrutura dos dados da CVM, enquanto o aspecto descritivo se manifesta na proposição e implementação de uma ferramenta que permita o uso prático dessas informações de forma organizada.

Os procedimentos técnicos combinam pesquisa bibliográfica e documental para levantamento e compreensão teórica dos dados públicos, juntamente com o desenvolvimento de uma solução tecnológica que automatiza sua coleta, integração e disponibilização.

A abordagem metodológica adotada é mista. Aspectos qualitativos estão presentes na análise estrutural e na modelagem dos dados, enquanto os elementos quantitativos se refletem no processamento automatizado e na organização das informações em larga escala.

No contexto da área de Computação, a pesquisa representa o desenvolvimento de uma solução tecnológica voltada à resolução de um problema específico, com potencial de aplicação prática no meio acadêmico e no ambiente de análise financeira (Wazlawick, 2009).

3.2 Gerenciamento do Projeto

O gerenciamento deste projeto foi conduzido com base em práticas ágeis de desenvolvimento de software, adotando a metodologia *Scrum* como estrutura principal (Schwaber, 2004). Como apoio visual, utilizou-se também o quadro de sinalização *Kanban*, contribuindo para a organização e o acompanhamento do fluxo de trabalho (Kanban, 2010).

Ambas as abordagens foram aplicadas de forma adaptada às necessidades do projeto, considerando sua natureza acadêmica e o escopo restrito da equipe envolvida. O *Scrum* foi utilizado de maneira personalizada, com ciclos iterativos de duas semanas, denominados *sprints*, nos quais as tarefas foram definidas, priorizadas e acompanhadas ao longo do desenvolvimento (Sutherland; Sutherland, 2014).

O Kanban complementou esse processo, funcionando como ferramenta visual para monitoramento das atividades, identificação de gargalos e ajustes de fluxo, contribuindo para uma gestão mais eficiente (Hammarberg; Sunden, 2014). Essa combinação flexível e adaptada de métodos ágeis mostrou-se eficaz para garantir a organização, a visibilidade do progresso e a entrega incremental da solução proposta.

3.3 Solução proposta

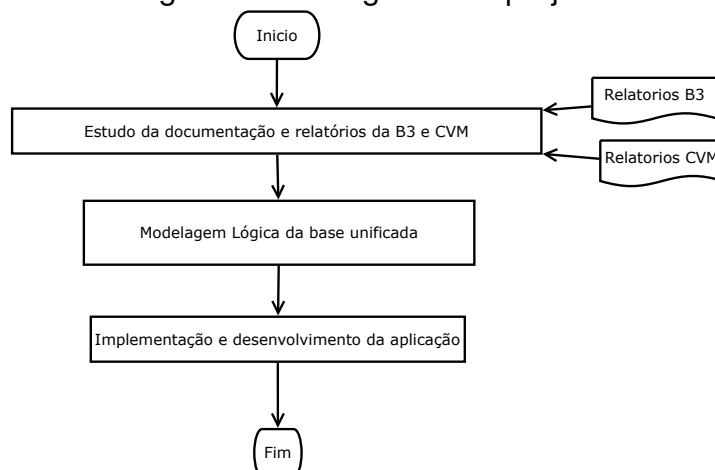
O ponto de partida do projeto consistiu em uma investigação detalhada da estrutura dos dados disponibilizados pela CVM. Essa etapa contemplou a análise dos formatos de arquivos, dos tipos de dados disponíveis e dos métodos de acesso às informações públicas fornecidas por essa instituição.

Embora os dados da B3 não tenham sido utilizados diretamente na construção da ferramenta, sua presença no contexto do mercado de capitais justificou a compreensão de sua estrutura e funcionamento. Dessa forma, foram consultados documentos como o relatório anual da CVM (CVM, 2023b) e referências sobre a estrutura de dados da B3 (B3, 2023b), a fim de mapear desafios e possibilidades associados à integração de dados financeiros.

A Figura 2 ilustra o fluxo geral do projeto. A fase inicial compreendeu a coleta e documentação dos dados da CVM, seguida pela modelagem lógica que estruturou um banco de dados unificado, preparado para integração e análise posterior.

Com base na análise inicial, foi realizada a modelagem lógica da base de dados, com foco na criação de um esquema unificado que contemplasse as diferentes estruturas contábeis e financeiras presentes nos arquivos da CVM. Para essa etapa, foram consideradas boas práticas descritas na literatura sobre modelagem de dados financeiros (DOMINGUES; ALMEIDA, 2020) e integração de grandes volumes de dados (Perlin, 2021).

Figura 2 – Fluxograma do projeto



Fonte: Elaborado pelo autor, 2024.

A etapa seguinte consistiu no desenvolvimento da aplicação de software responsável por automatizar o processo de coleta, tratamento e inserção dos dados na base unificada. Essa ferramenta foi desenvolvida com tecnologias modernas de extração e processamento de dados, garantindo flexibilidade e desempenho.

A aplicação foi capaz de realizar o *download* dos arquivos disponibilizados pela CVM, aplicar os tratamentos necessários para padronização e limpeza dos dados, e armazená-los de forma estruturada. Esse processo assegurou a integridade das informações e facilitou seu uso em análises fundamentalistas, estudos acadêmicos e outras aplicações. A proposta esteve alinhada com práticas consolidadas no desenvolvimento de soluções para automação e análise de dados (Prikladnicki; Willi; Milani, 2014) (NHIMI, 2016).

3.4 Materiais e Tecnologias

Esta seção apresenta os materiais e tecnologias que serão empregados durante o desenvolvimento do trabalho. O projeto será desenvolvido em um computador desktop cujas especificações são apresentadas no Quadro 1.

Durante a modelagem do banco de dados, utilizou-se o MySQL Workbench¹, versão 8.0. Essa ferramenta permitiu a criação visual do esquema lógico, facilitando o planejamento e a organização das entidades e relacionamentos.

Contudo, para o desenvolvimento da aplicação, optou-se pelo uso do SQLite², versão 3.49.1, como solução de banco de dados local. O SQLite foi adotado como banco de dados padrão do sistema, por oferecer maior praticidade na integração com o código desenvolvido. Sua leveza, portabilidade e independência de um

¹ <https://www.mysql.com/products/workbench/>

² <https://www.sqlite.org/>

Quadro 1 – Especificações do computador utilizado no TCC

Componente	Especificação
Processador	AMD Ryzen 5 5600G, 6-Core, 12-Threads, 3.6GHz (4.6GHz Turbo), Cache 19MB, AM4
Memória RAM	2x Kingston Fury Beast, 16GB, 3200MHz, DDR4, CL16
Armazenamento	SSD TGT Seal ST, 240GB, SATA III, Leitura 500 MB/s, Gravação 450 MB/s HD WD Blue, 1TB, 3.5, 5400 RPM, SATA III, Cache 64MB
Sistema Operacional	Windows 11 Pro

Fonte: Elaborado pelo autor, 2024.

servidor o tornam especialmente adequado para o uso do código por estudantes e profissionais da área financeira.

A linguagem de programação utilizada foi o Python, na versão 3.12.9³, pela sua versatilidade, ampla comunidade e bibliotecas especializadas.

As bibliotecas utilizadas no projeto foram:

- Pandas, versão 2.2.1⁴
- Numpy, versão 1.26.4⁵
- BeautifulSoup4, versão 4.12.2⁶
- Lxml, versão 5.1.0⁷
- Requests, versão 2.31.0⁸
- SQLAlchemy, versão 2.0.27⁹
- Mysql-Connector-Python, versão 8.3.0¹⁰

A biblioteca *pandas* foi empregada para manipulação e análise de dados tabulares. Sua estrutura baseada em *DataFrames* proporcionou uma forma eficiente de organizar, filtrar e agrupar informações financeiras.

A biblioteca *numpy* foi utilizada em conjunto com o *pandas* para operações numéricas vetoriais e matriciais, fundamentais na realização de cálculos e no tratamento de grandes volumes de dados.

Para extração de informações de páginas web, foi utilizada a biblioteca *beautifulsoup4*, que permitiu navegar pela estrutura HTML de documentos e extrair dados de forma precisa. A ela foi integrada a biblioteca *lxml*, que serviu como parser rápido e eficiente de HTML e XML.

A coleta dos dados disponibilizados pela CVM foi realizada por meio da

³ <https://www.python.org/downloads/release/python-3129/>

⁴ <https://pandas.pydata.org/>

⁵ <https://numpy.org/>

⁶ <https://www.crummy.com/software/BeautifulSoup/>

⁷ <https://lxml.de/>

⁸ <https://requests.readthedocs.io/en/latest/>

⁹ <https://www.sqlalchemy.org/>

¹⁰ <https://dev.mysql.com/downloads/connector/python/>

biblioteca *requests*, que viabilizou a comunicação com serviços HTTP e o *download* automatizado das informações necessárias.

A interface com o banco de dados foi desenvolvida com o auxílio do *SQLAlchemy*, que forneceu abstração e controle sobre as operações de inserção, consulta e atualização dos dados no SQLite, utilizando o paradigma de mapeamento objeto-relacional.

Durante os testes iniciais com a estrutura modelada no MySQL, foi utilizada a biblioteca *mysql-connector-python*, o conector oficial do MySQL para Python. Essa ferramenta garantiu a compatibilidade entre a modelagem relacional inicial e os scripts desenvolvidos, possibilitando a validação funcional da estrutura antes da migração definitiva para o SQLite.

4 DESENVOLVIMENTO

Este capítulo apresenta o processo de desenvolvimento da ferramenta proposta para a coleta, estruturação e integração de dados financeiros públicos disponibilizados pela CVM, com foco em análises fundamentalistas de companhias abertas brasileiras. A escolha pelo banco de dados SQLite justifica-se por sua leveza, portabilidade e facilidade de integração com sistemas locais, características adequadas a estudos exploratórios e reprodutíveis.

As atividades foram organizadas em três frentes principais, que estruturam as seções deste capítulo. A Seção 4.1 descreve a análise preliminar dos dados disponibilizados pela CVM, com destaque para a identificação de padrões, inconsistências e limitações nos metadados. A Seção 4.2 detalha a modelagem e estruturação da base de dados, com ênfase na padronização relacional e na aplicação de boas práticas de modelagem para contextos financeiros, conforme discutido na literatura (Elmasri; Navathe, 2016). Por fim, a Seção 4.3 apresenta o sistema desenvolvido em Python, incluindo os módulos responsáveis pela coleta, transformação e armazenamento dos dados, bem como os mecanismos de registro de eventos (*logs*).

4.1 Análise inicial dos dados da CVM

A primeira etapa consistiu na compreensão do formato, da frequência de atualização e de outros aspectos relevantes relacionados aos dados disponibilizados pela CVM. Como o foco do trabalho são as Companhias (CIA) Abertas¹, realizou-se uma análise do site como fonte para obtenção dessas informações. A partir disso, foram identificados os conjuntos de dados referentes às Companhias Abertas, organizados nas seguintes pastas:

- informação cadastral;
- formulário cadastral (FCA);
- periódicos e eventuais (IPE);
- formulário de referência (FRE);
- valores mobiliários negociados e detidos (VLMO);
- formulário de informações trimestrais (ITR);
- formulário de demonstrações financeiras padronizadas (DFP);
- informe do código de governança (ICBGC).

As informações cadastrais compõem uma categoria distinta, classificada como cadastro (CAD), enquanto os demais arquivos são agrupados sob a categoria de documentos (DOC). Ao acessar o site de dados da CVM², especificamente a se-

¹ <https://dados.cvm.gov.br>

² <https://dados.cvm.gov.br/dados/>

ção referente às Companhias Abertas, é possível visualizar essas duas categorias de dados.

Será apresentada, a seguir, a descrição de cada pasta referente ao conjunto de dados da CVM da CIA Aberta, com a finalidade de esclarecer o conteúdo e a utilidade de cada uma.

4.1.1 Informação cadastral

O conjunto Informação Cadastral reúne os dados cadastrais das companhias abertas disponibilizados pela CVM. Entre as informações disponibilizadas, destacam-se o número do Cadastro Nacional da Pessoa Jurídica (CNPJ), a data de registro da companhia e a situação atual desse registro. Esses dados são fundamentais para análises regulatórias, econômicas e financeiras, além de servirem como base para estudos acadêmicos e pesquisas de mercado. O dicionário de dados, disponibilizado em formato textual, contém a descrição detalhada das colunas e dos tipos de dados presentes no arquivo principal. Já os dados propriamente ditos, que contêm os registros cadastrais das companhias abertas, são apresentados em um arquivo estruturado.

Este conjunto de dados é de acesso público e está disponível na plataforma de dados abertos da CVM. A base foi originalmente criada em 17 de setembro de 2017, e desde então é atualizada diariamente. Integra a base de dados denominada Cadastro de Regulados, o que reforça seu caráter institucional e sua relevância para fins de controle e transparência regulatória.

4.1.2 Formulário cadastral (FCA)

O FCA é um documento eletrônico cuja entrega periódica ou eventual é regulamentada pela CVM (2022). Este formulário deve ser enviado à CVM por meio do sistema *Empresas.NET*, sendo uma obrigação regulatória destinada às companhias abertas. Sua principal função é atualizar e manter organizadas as informações institucionais e operacionais dessas entidades. O conjunto de dados públicos relacionados ao Formulário de Referência Anual (FCA), disponibilizado pela CVM, apresenta-se organizado em duas categorias principais de informações: os endereços de *download* dos documentos completos submetidos pelas companhias abertas e os conteúdos estruturados das seções que compõem o formulário.

A primeira categoria refere-se ao arquivo `fca_cia_aberta`, que contém os links para o *download* dos documentos completos entregues pelas companhias nos últimos cinco anos. A visualização desses documentos requer a utilização do programa *Empresas.NET*, disponibilizado pela própria CVM. A segunda categoria compreende

os conteúdos integrais do formulário, apresentados em formato estruturado e organizados em arquivos distintos que refletem as diversas seções. Entre os arquivos disponibilizados, destacam-se:

- fca_cia_aberta_geral;
- fca_cia_aberta_pais_estrangeiro_negociacao;
- fca_cia_aberta_canal_divulgacao;
- fca_cia_aberta_endereco;
- fca_cia_aberta_valor_mobiliario;
- fca_cia_aberta_auditor;
- fca_cia_aberta_escriturador;
- fca_cia_aberta_dri;
- fca_cia_aberta_departamento_acionistas.

Cada um desses arquivos aborda uma seção específica do formulário, possibilitando a segmentação e a análise detalhada das informações declaradas pelas companhias abertas. Os arquivos são atualizados semanalmente, refletindo tanto novas entregas quanto reapresentações realizadas pelas companhias reguladas. Além disso, mantém-se um histórico contínuo desde o ano de 2010, incluindo inclusive arquivos que não estão sujeitos à política de atualização vigente. O dicionário de dados correspondente é disponibilizado em formato compactado e oferece descrições detalhadas de todas as colunas e variáveis presentes nos arquivos.

Para fins de análise temporal e organização, os formulários referentes aos anos de 2020 a 2025 encontram-se agrupados em arquivos anuais distintos. Este conjunto de dados integra a base denominada *Documentos Periódicos e Eventuais de Regulados*, cuja criação remonta a 29 de maio de 2020. Sua manutenção ocorre de forma semanal, assegurando a atualização constante das informações disponibilizadas.

4.1.3 Periódicos e Eventuais (IPE)

O conjunto de dados referente às Informações Periódicas e Eventuais (IPE) reúne documentos não estruturados enviados por companhias abertas à Comissão de Valores Mobiliários (CVM), abrangendo os últimos cinco anos. Esses documentos, de caráter regulatório, contemplam tanto as obrigações periódicas quanto as comunicações eventuais exigidas ao longo das atividades societárias e da interlocução com o mercado. Trata-se de um acervo robusto, que reflete a diversidade de exigências legais e normativas aplicáveis às companhias reguladas. O conteúdo do conjunto está organizado em seis grandes categorias documentais:

- governança e estrutura societária;
- relação com investidores e mercado;

- informações econômico-financeiras e contábeis;
- transações e operações societárias;
- companhias em situação especial;
- informações regulatórias específicas.

A categoria de governança e estrutura societária compreende documentos que regulam a organização interna e as diretrizes de conduta das companhias, como Acordos de Acionistas, Estatutos Sociais, Regimentos Internos, Códigos de Conduta, além de Políticas de Governança e Sustentabilidade. Esses registros são fundamentais para garantir a conformidade, a transparência e a integridade das práticas de gestão corporativa. Já os documentos de relação com investidores e mercado incluem comunicados oficiais direcionados ao público investidor e à sociedade, como Comunicados ao Mercado, Avisos a Acionistas e Debenturistas, Fatos Relevantes, Informações prestadas a bolsas estrangeiras e o Calendário de Eventos Corporativos. Tais documentos asseguram a divulgação tempestiva e adequada de informações relevantes, conforme exigido pelas normas da CVM.

No que se refere às informações econômico-financeiras e contábeis, o conjunto abrange dados sobre a situação patrimonial e os resultados das companhias, além de instrumentos financeiros e políticas de distribuição de lucros. Incluem-se nessa categoria Demonstrativos Financeiros, Documentos de Ofertas Públicas, Escrituras de Debêntures, Políticas de Dividendos e de Destinação de Resultados. A categoria de transações e operações societárias, por sua vez, contém registros de eventos relevantes que impactam a estrutura e o capital das empresas, como Comunicações sobre Transações com Partes Relacionadas, Ofertas Públicas de Aquisição (OPA) e Planos de Remuneração Baseados em Ações. Já a documentação relativa a companhias em situação especial reúne informações sobre empresas que enfrentam processos de falência, liquidação ou recuperação judicial/extrajudicial, sendo essenciais para análise de risco e monitoramento da situação jurídico-financeira dessas entidades. Por fim, a categoria de informações regulatórias específicas abrange documentos exigidos por normativos da CVM, como as comunicações sobre negociação de valores mobiliários por pessoas ligadas, políticas de atuação de auditores independentes e outras políticas internas obrigatórias.

Os documentos estão organizados por ano de entrega, reunindo todas as submissões realizadas em cada período. Os arquivos correspondentes ao ano corrente e ao ano imediatamente anterior (A-1) são atualizados semanalmente, incorporando novas submissões e reapresentações. O conjunto conta com um histórico completo desde 2003, incluindo arquivos legados que, embora não estejam sujeitos à política atual de atualização contínua, permanecem acessíveis. As edições anuais estão disponíveis de 2020 a 2025, cada uma distribuída em arquivos compactados contendo os documentos entregues pelas companhias abertas no respectivo ano. A

descrição das colunas e variáveis utilizadas na indexação e organização dos dados está disponível em um dicionário de dados no formato textual. O conjunto foi instituído em 27 de fevereiro de 2021 e segue uma política de atualização semanal, com base nos *Documentos Periódicos e Eventuais de Regulados* recebidos pela CVM.

4.1.4 Formulário de Referência (FRE)

O Formulário de Referência (FRE) é um documento eletrônico previsto na CVM (2022), cuja apresentação à Comissão de Valores Mobiliários (CVM) é obrigatória e periódica (ou, em alguns casos, eventual), sendo realizada por meio do sistema *Empresas.NET*. A principal finalidade do FRE é consolidar e divulgar, de forma padronizada, um conjunto abrangente de informações sobre o emissor, abrangendo sua estrutura societária, situação financeira, práticas de governança, riscos e relação com o mercado.

O conjunto de dados público associado ao FRE disponibiliza tanto os endereços para *download* dos formulários quanto o conteúdo estruturado desses documentos em formato tabular. As duas principais componentes desse conjunto são:

- endereços para *download* dos formulários;
- conteúdo estruturado dos formulários em formato CSV.

O primeiro item refere-se aos links para *download* dos formulários entregues pelas companhias abertas nos últimos cinco anos, disponíveis por meio do conjunto *fre_cia_aberta*. A visualização dos arquivos requer o uso do programa *Empresas.NET*, distribuído gratuitamente pela CVM. Já o segundo item diz respeito ao conteúdo dos formulários em formato estruturado, disponibilizado em arquivos estruturados. Esses arquivos abrangem diversas dimensões informacionais, organizadas nos seguintes grupos:

- informações institucionais e operacionais;
- aspectos financeiros e patrimoniais;
- administração e governança;
- ações e mercado;
- aspectos sociais e de diversidade.

As informações institucionais e operacionais reúnem dados centrais sobre a estrutura e composição da companhia. Entre os principais elementos estão a identificação dos administradores, os dados do auditor independente, a estrutura de participação societária, os ativos imobilizados e intangíveis e o enquadramento da empresa em grupos econômicos. Esses dados permitem compreender a organização corporativa da companhia, sua capacidade operacional e seu posicionamento no mercado.

No que se refere aos aspectos financeiros e patrimoniais, o conteúdo cobre o endividamento da empresa, a política de distribuição de dividendos, a composição e

evolução do capital social, bem como os direitos associados às ações emitidas. Também são informadas as obrigações assumidas, as operações com valores mobiliários e as movimentações relevantes no patrimônio da empresa. Essas informações são fundamentais para a análise da saúde financeira da companhia e sua relação com investidores, credores e o mercado de capitais.

A seção de administração e governança fornece uma visão detalhada da estrutura decisória da companhia, incluindo a composição dos conselhos e comitês, os vínculos entre administradores, as políticas de remuneração — fixas e variáveis — e as práticas de governança corporativa. Também são disponibilizadas informações sobre as políticas internas que regulam a negociação de valores mobiliários por parte dos administradores, o que contribui para a avaliação do grau de transparência e adesão da empresa a boas práticas de mercado. Já a seção de ações e mercado trata da estrutura acionária da companhia, descrevendo a distribuição do capital social, a existência de programas de ADRs e a negociação de ações em bolsas internacionais, permitindo avaliar o grau de internacionalização da empresa.

Introduzidos a partir da versão de 2023 do formulário, os aspectos sociais e de diversidade ampliam o escopo informacional ao incluir dados sobre o perfil dos empregados e da administração com base em critérios como gênero, raça, faixa etária e localização geográfica. A incorporação desses elementos reflete a crescente importância da agenda ESG (ambiental, social e de governança) e reforça o papel da diversidade e da responsabilidade social na estratégia e reputação institucional das companhias abertas.

Os arquivos do FRE são atualizados semanalmente, incluindo tanto novas submissões quanto reapresentações realizadas pelas companhias. O conjunto mantém um histórico contínuo a partir de 2010. Os arquivos estruturados são acompanhados por um dicionário de dados, fornecido em formato compactado, que descreve detalhadamente as variáveis e colunas presentes nos arquivos. Além disso, estão disponíveis edições anuais consolidadas do conjunto, organizadas em pacotes compactados que abrangem os anos de 2020 a 2025, facilitando análises históricas e comparativas. O conjunto foi criado em 3 de novembro de 2019 e sua atualização é feita com frequência semanal, a partir da base de *Documentos Periódicos e Eventuais de Regulados* submetidos à CVM.

4.1.5 Valores mobiliários negociados e detidos (VLMO)

O conjunto de dados VLMO contempla informações de envio obrigatório à CVM, conforme a Resolução CVM (2021). Trata-se de uma obrigação de natureza periódica, cujo cumprimento deve ser realizado pelas companhias abertas por meio do sistema eletrônico *Empresas.NET*. Esse conjunto tem como objetivo registrar e

divulgar, de maneira transparente, a posição e as negociações realizadas com valores mobiliários por administradores, membros do conselho fiscal, controladores e pessoas a eles vinculadas. O conjunto disponibiliza os informes entregues pelas companhias nos últimos cinco anos, organizados em arquivos anuais compactados, contendo os dados estruturados. Esses arquivos reúnem informações como:

- nome e CPF/CNPJ dos declarantes;
- tipos e quantidades de valores mobiliários detidos;
- natureza da operação (compra, venda, bonificação, entre outros);
- data e características das transações realizadas;
- relação do declarante com a companhia emissora.

Os arquivos são acompanhados de um dicionário de dados, também em formato compactado. O conjunto é atualizado semanalmente. O histórico abrange os anos de 2020 a 2025. O conjunto de dados foi criado em 25 de junho de 2023. Ele utiliza como base os *Documentos Periódicos e Eventuais de Regulados*.

4.1.6 Demonstrativos Financeiros Padronizados

Os demonstrativos financeiros padronizados disponibilizados pela CVM, por meio do sistema *Empresas.NET*, reúnem dados contábeis estruturados de companhias abertas brasileiras. Tanto o ITR quanto o DFP seguem um modelo uniforme de reporte, que inclui as principais demonstrações financeiras, informações cadastrais, pareceres e arquivos para *download*.

O conjunto de dados contempla as principais demonstrações financeiras, apresentadas em formato estruturado:

- balanço patrimonial ativo (BPA);
- balanço patrimonial passivo (BPP);
- demonstrações dos fluxos de caixa (DFC-MD e DFC-MI);
- demonstração das mutações do patrimônio líquido (DMPL);
- demonstração do resultado (DRE e DRA);
- demonstração do valor adicionado (DVA).

A principal distinção entre os dois documentos está na periodicidade, o ITR é divulgado trimestralmente, enquanto o DFP é publicado anualmente.

4.1.6.1 Informações Trimestrais (ITR)

O ITR é um documento eletrônico de entrega obrigatória pelas companhias abertas à CVM, conforme a Resolução CVM (2022). Ele reúne informações contábeis elaboradas de forma trimestral, em conformidade com as normas contábeis vigentes, com o objetivo de garantir a transparência e o acompanhamento contínuo do desem-

penho das empresas. Além das demonstrações financeiras, o ITR inclui declarações de responsáveis, dados cadastrais, composição do capital social e *links* para *download* dos arquivos completos. Os registros históricos remontam a 2011, e os dados são acompanhados por um dicionário de variáveis disponível em arquivos compactados. O conjunto foi criado em 29 de novembro de 2019.

4.1.6.2 Formulário de Demonstrações Financeiras Padronizadas (DFP)

O DFP reúne as demonstrações financeiras anuais das companhias abertas, conforme a Resolução CVM (2022), e é submetido exclusivamente por meio do sistema *Empresas.NET*. Esse formulário constitui uma base essencial para a análise da saúde patrimonial e econômica das empresas listadas no mercado de capitais brasileiro. O conjunto de dados é atualizado semanalmente e abrange informações desde 2010. Além das demonstrações financeiras, inclui pareceres de auditoria, declarações de administradores, dados cadastrais, composição acionária e *links* para *download* dos formulários. As demonstrações apresentam tanto contas fixas quanto variáveis, o que permite análises comparáveis e também personalizadas. Um dicionário técnico em formato compactado detalha todas as variáveis e campos disponíveis. O conjunto foi oficialmente disponibilizado em 29 de julho de 2020.

4.1.7 Informe do código de governança (ICBGC)

O ICBGC é um documento eletrônico de envio periódico obrigatório por parte das companhias abertas à CVM, conforme determina a Resolução CVM (2022). O encaminhamento do informe deve ser feito por meio do sistema *Empresas.NET*.

Esse instrumento tem como finalidade central promover a transparência das práticas de governança corporativa adotadas pelas companhias abertas. O conjunto de dados disponibiliza os informes entregues pelas companhias nos últimos cinco anos, organizados por ano-calendário.

Os informes incluem informações sobre:

- estrutura de governança da companhia;
- composição e funcionamento dos órgãos de administração e fiscalização;
- políticas adotadas (remuneração, sustentabilidade, gestão de riscos etc.);
- nível de aderência às recomendações do código brasileiro de governança corporativa;
- justificativas em caso de não adoção de práticas recomendadas (princípio “aplique ou explique”).

Os arquivos estruturados são acompanhados por um dicionário de dados em formato compactado. O conjunto de dados é atualizado semanalmente. O histórico

disponível contempla os anos de 2020 a 2025. O conjunto foi formalmente criado em 31 de março de 2023.

Para conduzir este estudo, desenvolvemos dois scripts em Python, apresentados no Apêndice A e Apêndice B. O Apêndice A realiza a varredura recursiva na pasta de dados abertos da CVM, baixando todos os arquivos estruturados, compactados ou textuais sem aplicação de filtros. Já o Apêndice B é responsável por descompactar em lote os arquivos compactos obtidos, permitindo mensurar o volume e a estrutura dos dados com os quais estamos lidando.

Com os dados baixados e extraídos, construímos o Quadro 2, que resume as principais características de cada conjunto disponibilizado pela CVM.

Quadro 2 – Comparativo entre conjuntos de dados da CVM

Aspecto	ITR	DFP	FRE	FCA	IPE
Tipo de conteúdo	Informações Trimestrais	Demonstrações Financeiras Anuais	Formulário de Referência	Cadastro de Companhia Aberta	Documentos Periódicos/Eventuais
Frequência	Trimestral	Anual	Periódico/Eventual	Periódico/Eventual	Periódico/Eventual
Formato	CSV em ZIP	CSV em ZIP	CSV em ZIP	CSV em ZIP	CSV em ZIP
Volume descompactado	9,62 GB	3,49 GB	941 MB	28,2 MB	261 MB
Estrutura	Demonstrativos contábeis	Demonstrativos consolidados	Informações qualitativas diversas	Dados cadastrais padronizados	Documentos PDF + metadados
Atualização	Semanal	Semanal	Semanal	Semanal	Semanal (A e A-1)
Período histórico	Desde 2011	Desde 2010	Desde 2010	Desde 2010	Desde 2003
Destaques	BPA, DRE, DFC, DMPL, DVA etc.	BPA, DRE, DFC, DMPL, DVA etc.	Capital social, remuneração, ESG, governança	CNPJ, endereço, auditor, canais	Assembleias, fatos relevantes, estatutos, políticas
Complexidade	Alta (parse hierárquico)	Média (consolidação)	Alta (texto livre + variedade de temas)	Baixa	Alta (texto livre, representações)
Utilização	Análise de desempenho trimestral	Análise patrimonial e histórica	Análise estratégica e institucional	Base de entidades (Dim_Companhia)	Monitoramento de eventos corporativos

Fonte: Elaborado pelo autor, 2025.

Em síntese, a exploração dos diversos conjuntos de dados disponibilizados pela CVM permitiu a construção de uma base sólida de conhecimento, fundamental para a compreensão inicial do seu conteúdo. A elaboração de um código capaz de acessar e extrair esses arquivos foi essencial para alcançar um entendimento mais aprofundado acerca das características de cada base de dados, bem como de sua dinâmica de atualização.

A partir desse ponto, direcionamos nossos esforços à análise dos metadados, com o objetivo de transformar esse conhecimento prévio em um embasamento técnico para as etapas seguintes, iniciando-se pelo mapeamento dos dados. Esse mapeamento representa uma etapa fundamental para a modelagem de dados, tema que será abordado com maior profundidade na seção seguinte.

O processo de mapeamento teve início com a identificação dos arquivos de

origem e a verificação de sua periodicidade de atualização. A partir disso, foi possível planejar uma futura tabela de destino, na qual cada campo foi identificado com base nas informações extraídas dos metadados. Essa identificação incluiu a definição das colunas de origem e suas respectivas colunas de destino, bem como a descrição de cada coluna, o tipo de dado, seu domínio e os parâmetros de tamanho e precisão.

O resultado desse mapeamento pode ser consultado no Apêndice C, onde estão destacados, em marcações vermelhas, os campos considerados redundantes e, portanto, passíveis de descarte.

4.2 Modelagem e Estruturação dos Dados

A modelagem inicial foi realizada utilizando o MySQL Workbench, adotando-se uma abordagem relacional para representar os principais tipos de dados extraídos da CVM. Essa fase envolveu a criação inicial de diversas estruturas relacionais, focando nos agrupamentos lógicos, definição de chaves primárias e relações entre as tabelas. Ao longo do processo, houve um esforço consciente para reduzir e simplificar essa estrutura, culminando em um modelo final composto por nove tabelas principais que contêm apenas informações relevantes para análises fundamentalistas.

Após a definição conceitual e lógica, optou-se pela implementação da base utilizando SQLite, devido às suas características de leveza, portabilidade e facilidade de instalação, dispensando a necessidade de servidores externos. Essa escolha estratégica permite que qualquer usuário execute o sistema localmente e replique facilmente a estrutura desenvolvida.

Para operacionalizar o processo de criação do banco de dados, desenvolveu-se um script em Python que gera automaticamente o esquema completo da base, incluindo tabelas, chaves primárias e estrangeiras. Assim, o usuário pode executar o código sequencialmente, obtendo rapidamente uma base funcional e pronta para uso.

As principais estratégias adotadas para consolidar e simplificar a base foram:

- Redução de campos e tabelas desnecessárias;
- Definição clara e consistente dos relacionamentos;
- Unificação dos demonstrativos contábeis em tabelas únicas;
- Criação de índices específicos para acelerar consultas analíticas.

Campos e tabelas com pouca relevância analítica foram removidos, proporcionando uma estrutura mais enxuta, clara e eficiente. Foram definidos relacionamentos claros, garantindo a integridade referencial e facilitando consultas rápidas.

Os demonstrativos financeiros, como Balanço Patrimonial e Demonstração de Resultados, foram consolidados em tabelas únicas, com marcadores específicos

para identificar cada tipo de demonstrativo. Essa simplificação facilitou o acesso e permitiu análises cruzadas mais ágeis.

Além disso, a criação de índices específicos proporcionou maior eficiência nas consultas por período, tipo de demonstrativo e código das empresas, melhorando significativamente o desempenho do banco em análises recorrentes.

4.3 Software

REFERÊNCIAS

ARAUJO GONÇALVES, G. de; TOGNERI, R. M. Modelo de dados flexível para análise fundamentalista moderna: um estudo geral com dados no Brasil. **FGV RIC Revista de Iniciação Científica**, v. 2, 2021a. Disponível em: <https://periodicos.fgv.br/ric/article/view/86064/81092>. Acesso em: 02/02/2025.

ARAUJO GONÇALVES, G. de; TOGNERI, R. M. Modelo de dados flexível para análise fundamentalista moderna: um estudo geral com dados no Brasil. **FGV RIC Revista de Iniciação Científica**, v. 2, 2021b.

ATTIE, P. I. Universidade Estadual de Campinas. **O mercado financeiro e a sustentabilidade: o papel das bolsas de valores**. 2013. Dissertação de Mestrado – Campinas. Disponível em: <https://doi.org/10.47749/T/UNICAMP.2013.919566>. Acesso em: 14/03/2024.

BRASIL, BOLSA, BALCÃO (B3). B3. **Relatório Anual 2023**. B3, a bolsa do Brasil, 2023a. Disponível em: <https://api.mziq.com/mzfilemanager/v2/d/5fd7b7d8-54a1-472d-8426-eb896ad8a3c4/c322dd29-8e26-b81a-53a6-b5e2d70db984?origin=1>. Acesso em: 23/04/2024.

BRASIL, BOLSA, BALCÃO - (B3). B3. **Dados de mercado**. 2023b. Disponível em: https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/dados-de-mercado/. Acesso em: 23/04/2024.

BRASIL. **Lei nº 6.385, de 7 de dezembro de 1976**. 1976. Diário Oficial da União: seção 1, Brasília, DF, 9 dez. 1976. Dispõe sobre o mercado de valores mobiliários e cria a Comissão de Valores Mobiliários. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l6385.htm. Acesso em: 02/02/2025.

BRASIL, BOLSA, BALCÃO (B3). **Número de investidores na B3 cresce 34% em renda fixa e 23% em renda variável em 12 meses**: Número total de investidores chega a 17,6 milhões. Valor sob custódia bate R\$ 2,2 trilhões. Estudo Pessoa Física. 2023c. Disponível em: https://www.b3.com.br/pt_br/noticias/numero-de-investidores-na-b3-cresce-34-em-renda-fixa-e-23-em-renda-variavel-em-12-meses.htm. Acesso em: 14/01/2024.

CERVO, A. L.; BERVIAN, P. A. **Metodologia científica**. São Paulo: McGraw-Hill, 1983. Acesso em: 25/02/2024.

CLAESSENS, S.; KOSE, M. A. Macroeconomic implications of financial imperfections: a survey. BIS Working Paper, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3076410. Acesso em: 10/03/2024.

COMISSÃO DE VALORES MOBILIÁRIOS - (CVM). CVM. **Funções da CVM**. CVM, 2023a. Disponível em: <https://www.gov.br/cvm/pt-br/aceso-a-informacao-cvm/institucional/competencia>. Acesso em: 06/05/2024.

COMISSÃO DE VALORES MOBILIÁRIOS - (CVM). CVM. **O que é a CVM ?** CVM, 2009. Disponível em: <https://www.gov.br/cvm/pt-br/aceso-a-informacao-cvm/servidores/estagio/2-materia-cvm-e-o-mercado-de-capitais>. Acesso em: 25/02/2024.

COMISSÃO DE VALORES MOBILIÁRIOS - (CVM). CVM. Relatório Anual da CVM, 2023b. Disponível em: <https://www.gov.br/cvm/pt-br/centrais-de-conteudo/publicacoes/relatorios/anual>. Acesso em: 25/02/2024.

CONNOLLY, T.; BEGG, C. **Database systems: a practical approach to design, implementation, and management**. Pearson, 2015.

COSTA, T.; FERREIRINHA, P.; RESENDE, R. Integração de dados na construção: Desafios e soluções. *In*: UMINHO EDITORA. 5º Congresso Português de Building Information Modelling Volume 1: ptBIM. 2024. p. 349–360.

DAMODARAN, A. **Investment Valuation: Tools and Techniques for Determining the Value of Any Asset**. 3rd. Hoboken, NJ: Wiley, 2012. ISBN 978-1118011522. Disponível em: <https://books.google.com.br/books?id=E8-30Pga-nkC&printsec=frontcover&hl=pt-BR>.

DANTAS, M. Comportamento da bolsa de valores no Brasil diante das crises globais de 2008 e 2020. Pontifícia Universidade Católica de Goiás, 2020. Disponível em: <https://repositorio.pucgoias.edu.br/jspui/handle/123456789/1251>.

DELALIBERA, W. *et al.* Automatização do Modelo Rojo de Análise Fundamentalista. Universidade Estadual do Oeste do Paraná, 2023.

DOMINGUES, O.; ALMEIDA, G. L. M. **MODELAGEM DE COINTEGRAÇÃO EM ATIVOS FINANCEIROS**, 2020. Disponível em: https://aprepro.org.br/conbrepro/2020/anais/arquivos/09222020_120944_5f6a119ce2bee.pdf. Acesso em: 10/03/2024.

ELMASRI, R.; NAVATHE, S. B. **Fundamentals of database systems seventh edition**. pearson, 2016.

ELMASRI, R.; NAVATHE, S. B.; PINHEIRO, M. G. *et al.* **Sistemas de banco de dados**. Pearson Addison Wesley São Paulo, 2005.

EQUIPE SANTANDER. Santander. **O que é o mercado de capitais?** Atualizado em Fri Feb 16 20:26:43 GMT-03:00 2024. 2024. Disponível em: <https://www.santander.com.br/blog/o-que-e-mercado-de-capitais>. Acesso em: 23/04/2024.

FIGUEIREDO, C. A. d. S. *et al.* **O mercado acionário brasileiro: Possibilidades para investir no setor**. In: ANAIS do V Encontro Internacional de Gestão, Desenvolvimento e Inovação (EIGEDIN). Universidade Federal do Mato Grosso do Sul (UFMS), 2021. Estratégia e Negócios. Disponível em: <https://periodicos.ufms.br/index.php/EIGEDIN/article/view/13996>. Acesso em: 07/07/2024.

FIGUEIREDO, P. N. **Capacidade tecnológica e inovação: desafios para a transição industrial e econômica do Brasil**. Editora FGV, 2023. Disponível em: https://scholar.google.com.br/scholar?hl=pt-BR&as_sdt=0%2C5&authuser=1&q=Capacidade+tecnol%C3%B3gica+e+inova%C3%A7%C3%A3o%3A+desafios+para+a+transi%C3%A7%C3%A3o+industrial+e+econ%C3%B4mica+do+Brasil.&btnG=. Acesso em: 25/02/2024.

FONTINELE, A. **Análise de Dados - Balanços de Empresas CVM Dados Públicos**. 2025. GitHub. Disponível em: https://github.com/Alfredo-Fontinele/Analise_Dados_CVM.

FREITAS, L. C. d. **Uma análise fundamentalista do setor bancário: um estudo de caso com uma seleção de indicadores**, 2020.

FUNDAÇÃO GETULIO VARGAS (FGV). FGV. **Pesquisa da FGV aponta caminhos para transformação econômica do Brasil por meio da inovação**. Rede de Pesquisa FGV, 2024. Disponível em: <https://rededepesquisa.fgv.br/noticia/pesquisa-da-fgv-aponta-caminhos-para-transformacao-economica-do-brasil-por-meio-da>. Acesso em: 07/07/2024.

GERHARDT, T. E.; SILVEIRA, D. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009. Coordenado pela Universidade Aberta do Brasil – UAB/UFRGS e pelo Curso de Graduação Tecnológica – Planejamento e Gestão para o Desenvolvimento Rural da SEAD/UFRGS. Disponível em: <http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>. Acesso em: 20/02/2024.

GÓIS, A. D.; SOARES, S. V. **O efeito da reputação corporativa segundo a transparência contábil no gerenciamento de resultados de empresas listadas na B3**. v. 13. Revista de Educação e Pesquisa em Contabilidade (REPeC), 2019. DOI: 10.17524/repec.v13i2.2229. Disponível em: <https://www.repec.org.br/repec/article/view/2229>. Acesso em: 25/02/2024.

GOMES, F. R. A Bolsa de Valores brasileira como fonte de informações financeiras. **Perspectivas em Ciência da Informação**, v. 2, n. 2, 2007. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/23238>.

HALEVY, P.; NORVIG, P.; PEREIRA, F. The Unreasonable Effectiveness of Data. **IEEE Intelligent Systems**, v. 21, n. 2, p. 8–12, 2006.

HAMMARBERG, M.; SUNDEN, J. **Kanban in action**. Manning Publications Co., 2014. Disponível em: <https://thuvienso.hoasen.edu.vn/handle/123456789/9448>. Acesso em: 20/02/2024.

KANBAN, A. J. D. **Successful Evolutionary Change for Your Technology Business**. Blue Hole Press., 2010. Disponível em: <https://personalpages.bradley.edu/~young/CS690S117old/Kanban.pdf>.

KEARNEY, C.; LIU, S. Textual sentiment in finance: A survey of methods and models. **International Review of Financial Analysis**, Elsevier, v. 33, p. 171–185, 2014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1057521914000295>. Acesso em: 14/03/2024.

KOTHARI, S. P. Capital markets research in accounting. **Journal of accounting and economics**, Elsevier, v. 31, n. 1-3, p. 105–231, 2001.

LIAW, K. T. **The business of investment banking: A comprehensive overview**. John Wiley & Sons, 2011. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=4w3oVc4ZCMsC&oi=fnd&pg=PA1&dq=%22investment%22+AND+%22analysis%22+AND+%22comprehensive%22>

+AND+%22overview%22&ots=pNfZnh8lge&sig=HbbiBmApVV9lgjGlpPIGcMfP0Lw. Acesso em: 14/03/2024.

LINDMAN, sebastian *et al.* market impact on financial market integration: cross-quantilogram analysis of the global impact of the euro. **journal of empirical finance**, elsevier, v. 56, p. 42–73, 2020. DOI: <https://doi.org/10.1016/j.jempfin.2019.10.005>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0927539819300866>. Acesso em: 25/02/2024.

LOUREDO, G. **Códigos para ETL, análise estatística e visualização de dados da CVM**. 2025. GitHub. Disponível em: <https://github.com/GrlouX/AnaDadosFRECIAberta>.

MATHEW, S. An Overview on Capital Investment. Preprints, 2024. Disponível em: <https://www.preprints.org/manuscript/202404.0987>. Acesso em: 14/03/2024.

MINAS, T. **Balanco_Empresas_Dados_CVM - Análise de balanços a partir dos dados da CVM**. 2025. GitHub. Disponível em: https://github.com/thaisminas/Balanco_Empresas_Dados_CVM.

MONTOIA, G. R. Automatização da análise fundamentalista do mercado de ações brasileiro. Universidade Federal de Uberlândia, 2021.

NHIMI, F. Princípios e Práticas em Arquitetura de Software. **Instituto de Gestão em Tecnologia da Informação**, 2016. Disponível em: <https://www.machado.mg.gov.br/fil/es/concursos/1cf11cf161fe4eb688dfeec880d6b4d9.pdf>. Acesso em: 13/03/2024.

OMANE-ADJEPONG, M.; ABABIO, K. A.; ALAGIDEDE, I. P. Time-frequency analysis of behaviourally classified financial asset markets. **Research in International Business and Finance**, Elsevier, v. 50, p. 54–69, 2019. Disponível em: <https://doi.org/10.1016/j.ribaf.2019.04.012>. Acesso em: 14/03/2024.

PAIVA, J. P. R. de. **DadosCVM - Coleta e processamento de dados da CVM**. 2025. GitHub. Disponível em: <https://github.com/joaopedrordepaiva/DadosCVM>.

PERLIN, M. S. **Análise de Dados Financeiros e Econômicos com o R—versão online**. Porto Alegre, 2021. Disponível em: https://msperlin.com/adfer/adfeR_pt_ed03-ONLINE.pdf. Acesso em: 10/03/2024.

PRIKLADNICKI, R.; WILLI, R.; MILANI, F. **Métodos ágeis para desenvolvimento de software**. Bookman Editora, 2014. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=8rQABAAQBAJ&oi=fnd&pg=PR1&dq=%22desenvolvimento%22+AND+%22software%22+AND+%22princ%C3%ADpios%22+AND+%22pr%C3%A1ticas%22&ots=l1qd0qaZI9&sig=xzzbEM0afaCGkpKYFULQah8YeUc>. Acesso em: 13/03/2024.

REIS, L. V. B. *et al.* **Análise fundamentalista aplicada as ações negociadas na bolsa de valores brasileira**. Universidade Federal de Uberlândia, 2020.

REIS, T. Suno. **Negociação de ações nas bolsas de valores**: Tudo sobre o Mercado Financeiro: o que é e como ele funciona? 2021. Disponível em: <https://www.suno.com.br/guias/mercado-financeiro/>. Acesso em: 07/07/2024.

ROBERTO, A. M. **Papel da informação contábil auditada no processo de tomada de decisão: uma pesquisa realizada com futuros gestores**. Mariana: Instituto de Ciências Sociais Aplicadas, Universidade Federal de Ouro Preto, 2023. Disponível em: <http://www.monografias.ufop.br/handle/35400000/6082>. Acesso em: 25/02/2024.

SAHU, S. K.; MOKHADE, A.; BOKDE, N. D. An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: recent progress and challenges. **Applied Sciences**, MDPI, v. 13, n. 3, p. 1956, 2023. Disponível em: <https://www.mdpi.com/2076-3417/13/3/1956>. Acesso em: 10/03/2024.

SCANNAPIECO, M. **Data quality: concepts, methodologies and techniques. Data-centric systems and applications**. Springer, 2006.

SCHWABER, ken. **Agile Project Management with Scrum**. Microsoft Press, 2004. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=6pZCAwAAQBAJ&oi=fnd&pg=PT7&dq=Agile+Project+Management+with+Scrum&ots=kdqNUVaujV&sig=3sxF6tAAVnrNJgtQ1wGo-cJQ-kM>. Acesso em: 20/02/2024.

SUTHERLAND, J.; SUTHERLAND, J. **Scrum: the art of doing twice the work in half the time**. Crown Currency, 2014. Disponível em: <http://elib.vku.udn.vn/bitstream/123456789/3370/1/2014.%20Scrum-The%20Art%20of%20Doing%20Twice%20the%20work%20in%20Half%20the%20time.pdf>. Acesso em: 20/02/2024.

TEIXEIRA, J. C. **Mercado de Capitais: O que é e como funciona?** 2019. Disponível em: <https://fia.com.br/blog/mercado-de-capitais/>. Acesso em: 23/04/2024.

VALORES MOBILIÁRIOS (CVM), C. de. CVM. **Resolução CVM nº 44, de 23 de agosto de 2021**. 2021. *Diário Oficial da União: seção 1*, Brasília, DF, n. 161, p. 21–24. Dispõe sobre a divulgação de informações sobre ato ou fato relevante, a negociação de valores mobiliários na pendência de ato ou fato relevante não divulgado e a divulgação de informações sobre a negociação de valores mobiliários, e revoga as Instruções CVM nº 358/2002, nº 369/2002 e nº 449/2007. Disponível em: <https://conteudo.cvm.gov.br/export/sites/cvm/legislacao/resolucoes/anexos/001/resol044consolid.pdf>. Acesso em: 14/06/2025.

VALORES MOBILIÁRIOS (CVM), C. de. CVM. **Resolução CVM nº 80, de 29 de março de 2022**. 2022. *Diário Oficial da União: seção 1*, Brasília, DF, n. 62, p. 38–42. Dispõe sobre o registro e a prestação de informações periódicas e eventuais dos emissores de valores mobiliários admitidos à negociação em mercados regulamentados de valores mobiliários. Disponível em: <https://conteudo.cvm.gov.br/export/sites/cvm/legislacao/resolucoes/anexos/001/resol080consolid.pdf>. Acesso em: 14/06/2025.

VIEIRA, R. K. d. M. Montagem de carteiras de ações listadas na B3 e comparação com índices de investimentos brasileiros sob a ótica de análise fundamentalista. Universidade Federal da Paraíba, 2019.

WAZLAWICK, R. S. **Metodologia de Pesquisa para Ciência da Computação**. Rio de Janeiro: Elsevier, 2009. Acesso em: 25/02/2024.

GLOSSÁRIO

LaTeX – Linguagem de marcação utilizada principalmente para a composição de documentos técnicos e científicos, fornecendo uma formatação consistente e de alta qualidade.

Modelo / Template – Documento ou conjunto de elementos predefinidos que serve de estrutura base para a criação de outros documentos, permitindo uma formatação consistente e facilitando o trabalho de edição.

APÊNDICES

APÊNDICE A - CÓDIGO PARA BAIXAR OS DADOS DA CVM


```

import requests
from bs4 import BeautifulSoup
import os
import time

# Função para fazer o download do arquivo
def download_file(url, save_path):
    response = requests.get(url)
    with open(save_path, 'wb') as f:
        f.write(response.content)

# Função para fazer uma requisição HTTP com retries
def make_request(url, retries=5, delay=5):
    for i in range(retries):
        try:
            response = requests.get(url)
            response.raise_for_status()
            return response
        except requests.exceptions.RequestException as e:
            print(f"Erro ao acessar {url}: {e}")
            if i < retries - 1:
                print(f"Tentando novamente em {delay} segundos...")
                time.sleep(delay)
            else:
                raise

# URL base inicial
base_url = 'https://dados.cvm.gov.br/dados/CIA_ABERTA/'

# Diretório base de destino
dest_dir_base = 'CIA_ABERTA'

# Lista para armazenar os diretórios a serem explorados, iniciando com o URL base
dirs_to_explore = [base_url]

# Loop para explorar os diretórios
while dirs_to_explore:
    # Remove o primeiro diretório da lista
    current_dir = dirs_to_explore.pop(0)

    # Faz a requisição HTTP para obter o conteúdo da página
    response = make_request(current_dir)
    soup = BeautifulSoup(response.text, 'html.parser')

    # Encontra todos os links na página
    for link in soup.find_all('a'):
        href = link.get('href')

        # Ignora os links que são de volta ('../')
        if href == '../':
            continue

        # Constrói a URL completa do link
        full_url = current_dir + href

        # Se o link for um diretório, adiciona à lista de diretórios a serem explorados
        if href.endswith('/'):
            dirs_to_explore.append(full_url)
        else:
            # Se o link for um arquivo, verifica se é um arquivo CSV, ZIP ou TXT e faz o download
            if href.endswith('.csv') or href.endswith('.zip') or href.endswith('.txt'):
                # Obtém o caminho relativo do arquivo em relação à URL base
                relative_path = os.path.relpath(full_url, base_url)
                # Constrói o caminho completo de destino
                save_dir = os.path.join(dest_dir_base, os.path.dirname(relative_path))
                # Cria os diretórios necessários se não existirem
                os.makedirs(save_dir, exist_ok=True)
                # Constrói o caminho completo do arquivo de destino
                save_path = os.path.join(save_dir, os.path.basename(full_url))
                # Faz o download do arquivo
                download_file(full_url, save_path)
                print("Arquivo baixado:", save_path)

    # Imprime o diretório atual
    print("Diretório atual:", current_dir)

```


APÊNDICE B - CÓDIGO PARA EXTRAIR OS DADOS DA CVM

```
1 import os
2 import zipfile
3
4 # Caminhos de origem e destino
5 pasta_zip = r"CIA_ABERTA"
6 pasta_destino = os.path.join(os.path.dirname(pasta_zip), "CIA_ABERTA_extraida")
7
8 # Percorrer a pasta de origem
9 for root, _, files in os.walk(pasta_zip):
10     for file in files:
11         if file.endswith(".zip"):
12             caminho_zip = os.path.join(root, file)
13
14             # Caminho relativo da subpasta em relação à raiz da pasta_zip
15             caminho_relativo = os.path.relpath(root, pasta_zip)
16
17             # Criar o caminho correspondente na pasta de destino
18             destino_subpasta = os.path.join(pasta_destino, caminho_relativo)
19             os.makedirs(destino_subpasta, exist_ok=True)
20
21             # Extrair o conteúdo do ZIP nesse caminho
22             try:
23                 with zipfile.ZipFile(caminho_zip, 'r') as zip_ref:
24                     zip_ref.extractall(destino_subpasta)
25                     print(f"Extraído: {file} para {destino_subpasta}")
26             except zipfile.BadZipFile:
27                 print(f"Erro: Arquivo ZIP inválido - {caminho_zip}")
```

APÊNDICE C - MAPEAMENTO COMPLETO DOS DADOS DA CVM

DFP - Dados Cadastrais							
Arquivo de Origem		Periodicidade de atualização	Tabela de Destino				
meta_dfp_cia_aberta_bt.zip		Semanal	Dfp				
meta_dfp_cia_aberta.txt							
Colunas Origem		Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	
CATEG_DOC	categoria_doc	Categoria do documento	varchar	Alfanumérico		20	
CD_CVM	codigo_cvm	Código CVM	char	Númerico		6	
CNPJ_CIA	cnpi_companhia	CNPJ da companhia	varchar	Alfanumérico		20	
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanumérico		100	
DT_RECEB	data_recebimento_doc	Data da recebimento do documento	date	AAAA-MM-DD		10	
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD		10	
ID_DOC	id_doc	Identificador do documento	int	Númerico		10	
LINK_DOC	link_doc	Endereço para download do documento	varchar	Alfanumérico		121	
VERSAO	versao	Versão do documento	smallint	Númerico		5	
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD		10	
	mes_doc	Mês do documento (aspecto temporal)	varchar	Númerico		4	
	ano_doc	Ano do documento (aspecto temporal)	varchar	Númerico		4	
Balanco Patrimonial Ativo (BPA)							
Arquivo de Origem		Periodicidade de atualização					
meta_dfp_cia_aberta_bt.zip		Semanal					
meta_dfp_cia_aberta_BPA.txt							
Colunas Origem		Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	Tabela de Destino
CD_CONTA	codigo_conta	Código da conta	varchar	Númerico		18	Dfp
CD_CVM	codigo_cvm	Código CVM	char	Númerico		6	Dfp
CNPJ_CIA	cnpi_companhia	CNPJ da companhia	varchar	Alfanumérico		20	
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanumérico		100	
DS_CONTA	descricao_conta	Descrição da conta	varchar	Alfanumérico		100	
DT_FIM_EXERC	data_fim_exercicio	Data fim do exercício social	date	AAAA-MM-DD		10	Dfp
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD		10	Dfp
ESCALA_MOEDA	escala_monetaria	Escala monetária	varchar	Alfanumérico		100	Dfp
GRUPO_DFP	grupo_dfp	Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado)	varchar	Alfanumérico		206	grupo_dfp
MOEDA	moeda	Moeda	varchar	Alfanumérico		100	Dfp
ORDEM_EXERC	ordem_exercicio	Ordem do exercício social	varchar	Alfanumérico		9	Dfp
ST_CONTA_FIXA	tipo_conta	Indica se é conta fixa ou não	varchar	S/N		1	Dfp
VERSAO	versao	Versão do documento	smallint	Númerico		5	Dfp
VL_CONTA	valor_conta	Valor da conta	decimal	Númerico		29	Dfp
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD		10	Dfp
	mes_doc	Mês do documento (aspecto temporal)	varchar	Númerico		4	Dfp
	ano_doc	Ano do documento (aspecto temporal)	varchar	Númerico		4	Dfp
Balanco Patrimonial Ativo (BPP)							
Arquivo de Origem		Periodicidade de atualização					
meta_dfp_cia_aberta_bt.zip		Semanal					
meta_dfp_cia_aberta_BPP.txt							

Colunas Origem	Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	Tabela de Destino
CD_CONTA	codigo_conta	Código da conta	varchar	Numérico	18	Dfp
CD_CVM	codigo_cvm	Código CVM	char	Numérico	6	Dfp
CNPJ_CIA	cnpj_companhia	CNPJ da companhia	varchar	Alfanumérico	20	
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanumérico	100	
DS_CONTA	descricao_cotna	Descrição da conta	varchar	Alfanumérico	100	
DT_FIM_EXERC	data_fim_exercicio	Data fim do exercicio social	date	AAAA-MM-DD	10	Dfp
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD	10	Dfp
ESCALA_MOEDA	escala_monetaria	Escala monetária	varchar	Alfanumérico	100	Dfp
GRUPO_DFP	grupo_dfp	Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado)	varchar	Alfanumérico	206	grupo_dfp
MOEDA	moeda	Moeda	varchar	Alfanumérico	100	Dfp
ORDEM_EXERC	ordem_exercicio	Ordem do exercicio social	varchar	Alfanumérico	9	Dfp
ST_CONTA_FIXA	conta_fixa	Indica se é conta fixa ou não	varchar	S/N	1	Dfp
VERSAO	versao	Versão do documento	smallint	Numérico	5	Dfp
VL_CONTA	valor_conta	Valor da conta	decimal	Numérico	29	Dfp
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD	10	Dfp
	mes_doc	Mês do documento (aspecto temporal)	varchar	Numérico	4	Dfp
	ano_doc	Ano do documento (aspecto temporal)	varchar	Numérico	4	Dfp
Demonstração de Fluxo de Caixa - Método Direto (DFC-MD)						
Arquivo de Origem		Periodicidade de atualização				
meta_dfp_cia_aberta_txt.zip		Semanal				
meta_dfp_cia_aberta_DFC_MD						
Colunas Origem	Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	Tabela de Destino
CD_CONTA	codigo_conta	Código da conta	varchar	Numérico	18	Dfp
CD_CVM	codigo_cvm	Código CVM	char	Numérico	6	Dfp
CNPJ_CIA	cnpj_companhia	CNPJ da companhia	varchar	Alfanumérico	20	
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanumérico	100	
DS_CONTA	descricao_cotna	Descrição da conta	varchar	Alfanumérico	100	
DT_FIM_EXERC	data_fim_exercicio	Data fim do exercicio social	date	AAAA-MM-DD	10	Dfp
DT_INI_EXERC	data_inicio_exercicio	Data inicio do exercicio social	date	AAAA-MM-DD	10	Dfp
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD	10	Dfp
ESCALA_MOEDA	escala_monetaria	Escala monetária	varchar	Alfanumérico	100	Dfp
GRUPO_DFP	grupo_dfp	Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado)	varchar	Alfanumérico	206	grupo_dfp
MOEDA	moeda	Moeda	varchar	Alfanumérico	100	Dfp
ORDEM_EXERC	ordem_exercicio	Ordem do exercicio social	varchar	Alfanumérico	9	Dfp
ST_CONTA_FIXA	conta_fixa	Indica se é conta fixa ou não	varchar	S/N	1	Dfp
VERSAO	versao	Versão do documento	smallint	Numérico	5	Dfp
VL_CONTA	valor_conta	Valor da conta	decimal	Numérico	29	Dfp
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD	10	Dfp
	mes_doc	Mês do documento (aspecto temporal)	varchar	Numérico	4	Dfp
	ano_doc	Ano do documento (aspecto temporal)	varchar	Numérico	4	Dfp
Demonstração de Fluxo de Caixa - Método Indireto (DFC-MI)						
Arquivo de Origem		Periodicidade de atualização				
meta_dfp_cia_aberta_txt.zip		Semanal				

meta_dfp_cia_aberta_DFC_ML.txt						
Colunas Origem	Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	Tabela de Destino
CD_CONTA	codigo_conta	Código da conta	varchar	Númerico	18	Dfp
CD_CVM	codigo_cvm	Código CVM	char	Númerico	6	Dfp
CNPJ_CIA	cnpi_companhia	CNPJ da companhia	varchar	Alfanúmerico	20	
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanúmerico	100	
DS_CONTA	descricao_conta	Descrição da conta	varchar	Alfanúmerico	100	
DT_FIM_EXERC	data_fim_exercicio	Data fim do exercicio social	date	AAAA-MM-DD	10	Dfp
DT_INI_EXERC	data_inicio_exercicio	Data início do exercicio social	date	AAAA-MM-DD	10	Dfp
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD	10	Dfp
ESCALA_MOEDA	escala_monetaria	Escala monetária	varchar	Alfanúmerico	100	Dfp
GRUPO_DFP	grupo_dfp	Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado)	varchar	Alfanúmerico	206	grupo_dfp
MOEDA	moeda	Moeda	varchar	Alfanúmerico	100	Dfp
ORDEM_EXERC	ordem_exercicio	Ordem do exercicio social	varchar	Alfanúmerico	9	Dfp
ST_CONTA_FIXA	conta_fixa	Indica se é conta fixa ou não	varchar	S/N	1	Dfp
VERSAO	versao	Versão do documento	smallint	Númerico	5	Dfp
VL_CONTA	valor_conta	Valor da conta	decimal	Númerico	29	Dfp
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD	10	Dfp
	mes_doc	Mês do documento (aspecto temporal)	varchar	Númerico	4	Dfp
	ano_doc	Ano do documento (aspecto temporal)	varchar	Númerico	4	Dfp
Demonstração das Mutações do Patrimônio Líquido (DMPL)						
Arquivo de Origem	Periodicidade de atualização					
meta_dfp_cia_aberta_txt.zip	Semanal					
meta_dfp_cia_aberta_DMPL.txt						
Colunas Origem	Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	Tabela de Destino
CD_CONTA	codigo_conta	Código da conta	varchar	Númerico	18	Dfp
CD_CVM	codigo_cvm	Código CVM	char	Númerico	6	Dfp
CNPJ_CIA	cnpi_companhia	CNPJ da companhia	varchar	Alfanúmerico	20	
COLUNA_DF	coluna_df	Coluna da demonstração financeira	varchar	Alfanúmerico	60	Dfp
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanúmerico	100	
DS_CONTA	descricao_conta	Descrição da conta	varchar	Alfanúmerico	100	Dfp
DT_FIM_EXERC	data_fim_exercicio	Data fim do exercicio social	date	AAAA-MM-DD	10	Dfp
DT_INI_EXERC	data_inicio_exercicio	Data início do exercicio social	date	AAAA-MM-DD	10	Dfp
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD	10	Dfp
ESCALA_MOEDA	escala_monetaria	Escala monetária	varchar	Alfanúmerico	100	Dfp
GRUPO_DFP	grupo_dfp	Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado)	varchar	Alfanúmerico	206	grupo_dfp
MOEDA	moeda	Moeda	varchar	Alfanúmerico	100	Dfp
ORDEM_EXERC	ordem_exercicio	Ordem do exercicio social	varchar	Alfanúmerico	9	Dfp
ST_CONTA_FIXA	conta_fixa	Indica se é conta fixa ou não	varchar	S/N	1	Dfp
VERSAO	versao	Versão do documento	smallint	Númerico	5	Dfp
VL_CONTA	valor_conta	Valor da conta	decimal	Númerico	29	Dfp
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD	10	Dfp
	mes_doc	Mês do documento (aspecto temporal)	varchar	Númerico	4	Dfp
	ano_doc	Ano do documento (aspecto temporal)	varchar	Númerico	4	Dfp

[illegible]

Colunas Origem	Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	Tabela de Destino
CD_CONTA	codigo_conta	Código da conta	varchar	Número	18	Dfp
CD_CVM	codigo_cvm	Código CVM	char	Número	6	Dfp
CNPJ_CIA	cnnpj_companhia	CNPJ da companhia	varchar	Alfanumérico	20	
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanumérico	100	
DS_CONTA	descricao_conta	Descrição da conta	varchar	Alfanumérico	100	
DT_FIM_EXERC	data_fim_exercicio	Data fim do exercício social	date	AAAA-MM-DD	10	Dfp
DT_INI_EXERC	data_inicio_exercicio	Data início do exercício social	date	AAAA-MM-DD	10	Dfp
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD	10	Dfp
ESCALA_MOEDA	escala_monetaria	Escala monetária	varchar	Alfanumérico	100	
GRUPO_DFP	grupo_dfp	Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado)	varchar	Alfanumérico	206	grupo_dfp
MOEDA	moeda	Moeda	varchar	Alfanumérico	100	Dfp
ORDEM_EXERC	ordem_exercicio	Ordem do exercício social	varchar	Alfanumérico	9	Dfp
ST_CONTA_FIXA	conta_fixa	Indica se é conta fixa ou não	varchar	S/N	1	Dfp
VERSAO	versao	Versão do documento	smallint	Número	5	Dfp
VL_CONTA	valor_conta	Valor da conta	decimal	Número	29	Dfp
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD	10	Dfp
	mes_doc	Mês do documento (aspecto temporal)	varchar	Número	4	Dfp
	ano_doc	Ano do documento (aspecto temporal)	varchar	Número	4	Dfp

Demonstração de Resultado (DRE)	
Arquivo de Origem	Periodicidade de atualização
meta_dfp_cia_aberta_bt.zip	Semanal
meta_dfp_cia_aberta_DRE.txt	

Colunas Origem	Colunas Destino	Descrição coluna	Tipo de dados	Domínio	Tamanho/Precisão	Tabela de Destino
CD_CONTA	codigo_conta	Código da conta	varchar	Númeroico	18	Dfp
CD_CVM	codigo_cvm	Código CVM	char	Númeroico	6	Dfp
CNPJ_CIA	cnpi_companhia	CNPJ da companhia	varchar	Alfanumérico	20	
DENOM_CIA	denominacao_companhia	Nome empresarial da companhia	varchar	Alfanumérico	100	
DS_CONTA	descricao_conta	Descrição da conta	varchar	Alfanumérico	100	
DT_FIM_EXERC	data_fim_exercicio	Data fim do exercício social	date	AAAA-MM-DD	10	Dfp
DT_INI_EXERC	data_inicio_exercicio	Data início do exercício social	date	AAAA-MM-DD	10	Dfp
DT_REFER	data_referencia_doc	Data de referência do documento	date	AAAA-MM-DD	10	Dfp
ESCALA_MOEDA	escala_monetaria	Escala monetária	varchar	Alfanumérico	100	Dfp
GRUPO_DFP	grupo_dfp	Nome e nível de agregação da demonstração (Mostra se é Individual ou Consolidado)	varchar	Alfanumérico	206	grupo_dfp
MOEDA	moeda	Moeda	varchar	Alfanumérico	100	Dfp

TP_PARECER_DECL.	tipo_parecer_declaracao	Tipo do Parecer/Declaração	varchar	Alfanumérico	101	Dfp_Parecer
TP_RELAT_AUD	tipo_relatorio_auditor	Tipo Relatório do Auditor Independente	varchar	Alfanumérico	19	Dfp_Parecer
TXT_PARECER_DECL.	texto_parecer_declaracao	Texto do Parecer/Declaração	varchar	Alfanumérico	8000	Dfp_Parecer
VERSAO	versao	Versão do documento	smallint	Númérico	5	Dfp_Parecer
	data_doc	Data do documento (aspecto temporal)	date	AAAA-MM-DD	10	Dfp_Parecer
	mes_doc	Mês do documento (aspecto temporal)	varchar	Númérico	4	Dfp_Parecer
	ano_doc	Ano do documento (aspecto temporal)	varchar	Númérico	4	Dfp_Parecer
grupo_dfp						
codigo_grupo_dfp	int	índice da tabela para agrupar os grupos de DPF				
	varchar(255)	texto que mostra se individual ou consolidado de cada grupo				

ÍNDICE