

INSTITUTO TECNOLÓGICO DE AERONÁUTICA



Pedro Henrique Ferreira Cavalcante

**DESENVOLVIMENTO DE BIBLIOTECA DE CAPTURA
DE INFORMAÇÕES DE EMPRESAS LISTADAS NA B3
EM CÓDIGO R E SUAS APLICAÇÕES PRÁTICAS**

Trabalho de Graduação
2021

Curso de Engenharia da Computação

Pedro Henrique Ferreira Cavalcante

**DESENVOLVIMENTO DE BIBLIOTECA DE CAPTURA
DE INFORMAÇÕES DE EMPRESAS LISTADAS NA B3
EM CÓDIGO R E SUAS APLICAÇÕES PRÁTICAS**

Orientador

Prof. Dr. Elton Felipe Sbruzzi (ITA)

ENGENHARIA DA COMPUTAÇÃO

SÃO JOSÉ DOS CAMPOS
INSTITUTO TECNOLÓGICO DE AERONÁUTICA

Dados Internacionais de Catalogação-na-Publicação (CIP)
Divisão de Informação e Documentação

Ferreira Cavalcante, Pedro Henrique

Desenvolvimento de biblioteca de captura de informações de empresas listadas na B3 em código R e suas aplicações práticas / Pedro Henrique Ferreira Cavalcante.

São José dos Campos, 2021.

46f.

Trabalho de Graduação – Curso de Engenharia da Computação– Instituto Tecnológico de Aeronáutica, 2021. Orientador: Prof. Dr. Elton Felipe Sbruzzi.

I. Instituto Tecnológico de Aeronáutica. II. Título.

REFERÊNCIA BIBLIOGRÁFICA

FERREIRA CAVALCANTE, Pedro Henrique. **Desenvolvimento de biblioteca de captura de informações de empresas listadas na B3 em código R e suas aplicações práticas**. 2021. 46f. Trabalho de Conclusão de Curso (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: Pedro Henrique Ferreira Cavalcante

TÍTULO DO TRABALHO: Desenvolvimento de biblioteca de captura de informações de empresas listadas na B3 em código R e suas aplicações práticas.

TIPO DO TRABALHO/ANO: Trabalho de Conclusão de Curso (Graduação) / 2021

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias deste trabalho de graduação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte deste trabalho de graduação pode ser reproduzida sem a autorização do autor.

Pedro Henrique Ferreira Cavalcante
Rua Santa Justina, 277, apto 132
04.545-041 – São Paulo–SP

DESENVOLVIMENTO DE BIBLIOTECA DE CAPTURA DE INFORMAÇÕES DE EMPRESAS LISTADAS NA B3 EM CÓDIGO R E SUAS APLICAÇÕES PRÁTICAS

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação

Pedro Henrique Ferreira Cavalcante

Autor

Elton Felipe Sbruzzi (ITA)

Orientador

Prof. Dr. Johnny Marques

Coordenador do Curso de Engenharia da Computação

São José dos Campos, XX de XXX de 2021.

A meus pais, que sempre se esforçaram para que eu pudesse chegar onde quisesse. A minha família, que sempre me apoiou em todas as situações que precisei e aos meus amigos, por todo o apoio que me permitiu chegar até aqui.

*'Maturity is the ability to reject good alternatives
in order to pursue even better ones.'*

— RAY DALIO

Resumo

Este trabalho apresenta a criação de uma biblioteca em R com o objetivo de capturar informações referentes a empresas listadas na bolsa de valores brasileira (B3). Além de apresentar todo o processo de obtenção e estruturação de dados para que assim possa servir de inspiração e guia para outros projetos semelhantes. Por fim, serão apresentados exemplos de como esses dados podem ser utilizados por um cientista de dados para que seu objetivo possa ser alcançado.

Abstract

This work presents the creation of a library in R with the objective of capturing information referring to companies listed on the Brazilian stock exchange (B3). In addition to presenting the entire process of obtaining and structuring data so that this work can serve as an inspiration and guide for other similar projects. Finally, examples will be obtained of how these data can be used by a data scientist to achieve their goal.

Lista de Figuras

FIGURA 3.1 – Estrutura banco de dados	20
FIGURA 3.2 – Classificação Setorial B3	21
FIGURA 3.3 – Fonte de dados CVM	22
FIGURA 3.4 – Arquivos dentro da pasta compactada ITR 2020	22
FIGURA 3.5 – Dados BPA 2020	23
FIGURA 3.6 – Fonte de dados DRE	24
FIGURA 3.7 – Site Fundamentus	26
FIGURA 3.8 – Número de ações PETR4	27
FIGURA 4.1 – Função Construir Banco de Dados	29
FIGURA 4.2 – Crawler CVM	30
FIGURA 4.3 – Função Depara	31
FIGURA 4.4 – Estruturação TBL_DRE	31
FIGURA 4.5 – DRE Petrobrás	32
FIGURA 4.6 – Balanço Patrimonial Petrobrás	33
FIGURA 4.7 – Estruturação TBL_DIVULGACOES	34
FIGURA 4.8 – Estruturação TBL_COTACOES	34
FIGURA 4.9 – Função de atualização da ponta	35
FIGURA 4.10 – Função de evolução dre	35
FIGURA 4.11 – Receita Bruta Petrobrás	36
FIGURA 4.12 – Ativo Total, Passivo Total e Patrimônio Líquido da Petrobrás	37
FIGURA 4.13 – Métricas por Empresa	38
FIGURA 4.14 – Métricas por Empresa	39

FIGURA 4.15 –Dashboard	39
FIGURA 4.16 –Body RShinyboard	41
FIGURA 4.17 –Server RShinyboard	41
FIGURA 4.18 –Back RShinyboard	41
FIGURA 4.19 –GIT	42
FIGURA 4.20 –documentacao <i>library</i>	43
FIGURA 4.21 – <i>help</i> função construir_bd	43
FIGURA 4.22 –funcoes auxiliares <i>library</i>	44

Lista de Tabelas

Lista de Abreviaturas e Siglas

CVM	Comissão de Valores Mobiliários
BP	Balanco Patrimonial
BPA	Balanco Patrimonial Ativo
BPP	Balanco Patrimonial Passivo
DRE	<i>Demonstrativo de Resultados</i>
HTTP	<i>Hyper Text Transfer Protocol</i>
SQL	<i>Standard Query Language</i>
B3	<i>Brasil, Bolsa, Balcão</i>
DFP	<i>Demonstração Financeira Padronizada</i>
ITR	<i>Informações Trimestrais</i>

Sumário

1	INTRODUÇÃO	14
1.1	Contexto	14
1.2	Objetivos	15
1.2.1	Objetivo Geral	15
1.2.2	Objetivo Específicos	15
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	CVM	16
2.2	Companhias de Capital Aberto	16
2.2.1	Balanço Patrimonial	16
2.2.2	Demonstrativo de Resultados	16
2.2.3	Individual ou Consolidado	17
2.2.4	Ações	17
2.2.5	Preferencial e Ordinária	17
2.3	Variáveis Macroeconômicas	17
2.3.1	Dólar	18
2.3.2	PIB	18
2.3.3	Selic	18
3	FONTE DE DADOS	20
3.1	Divisão Setorial	21
3.2	CVM	22
3.2.1	Balanço Patrimonial	23
3.2.2	Demonstrativo de Resultados	24

3.3	Outros Dados	25
3.3.1	Data de Referência e Data de Publicação	25
3.3.2	Ação e Nome da Empresa	26
3.4	Número de Ações	27
3.5	Cotação	27
3.6	IPEA	28
3.7	SIDRA	28
4	DESENVOLVIMENTO	29
4.1	Construção BD	29
4.1.1	Série CVM	29
4.1.2	TBL_EMPRESAS	30
4.1.3	TBL_DRE	31
4.1.4	TBL_BPA e TBL_BPP	32
4.1.5	TBL_DIVULGACOES	33
4.1.6	TBL_COTACOES	34
4.2	Atualização	34
4.3	Construção de Métricas	35
4.3.1	Evolução do DRE	35
4.3.2	Evolução do BP	36
4.3.3	Métricas por Empresa	37
4.3.4	Métricas via Preço	38
4.4	Visualização	39
4.4.1	Construção	40
4.5	Documentação e GIT	42
5	CONCLUSÃO	45
	REFERÊNCIAS	46

1 Introdução

1.1 Contexto

Investimento financeiro consiste na aplicação de capital por meio de empreendimentos rentáveis como fundos de investimentos, títulos públicos, ações e entre outros, com o intuito de obter no futuro um retorno superior ao investido, podendo-se ter um horizonte de rendimento com curto, médio ou longo prazo a depender do perfil do investidor (CARDOZO *et al.*, 2019).

No Brasil, até em torno de 2019, a população brasileira não investia seu dinheiro ou aplicavam apenas em renda fixa (poupança, tesouro direto e etc). Contudo, com a queda de juros e o advento da *Internet* e cursos de investimento, tem ocorrido uma migração para o investimento em empresas listadas em bolsa.

Essas empresas, por receber investimentos de pessoas físicas, possuem alguns deveres em relação à CVM, entidade autárquica que possui a finalidade de disciplinar, fiscalizar e desenvolver o mercado de valores mobiliários. (CVM, 2020)

Entre esses deveres, as empresas listadas em bolsa devem divulgar trimestralmente seu Balanço Patrimonial (BP) e seu Demonstrativo de Resultados (DRE), que são informações de suma importância para que se possa entender a saúde da empresa e como foi o seu desenvolvimento ao longo do tempo.

Apesar de esses dados serem divulgados publicamente, muitas vezes são difíceis de serem encontrados, ou estão resumidos demais para que um cientista de dados se aprofunde em análises mais complexas, ou então esses dados estão tratados em plataformas e sistemas que cobram um preço caro demais para que uma pessoa física possa fazer um estudo inicial.

O objetivo deste projeto é o processo de criação de uma *library* que junte o maior número de informações apenas com dados disponíveis na web.

1.2 Objetivos

1.2.1 Objetivo Geral

Este trabalho de conclusão de curso possui o objetivo de desenvolver um banco de dados com o maior número de informações possíveis sobre as empresas listadas na B3, assim como captar dados auxiliares que possuem algum poder preditivo sobre a sua cotação. Além desenvolver toda a rotina de atualização para que os usuários possam utilizar essas informações continuamente.

1.2.2 Objetivo Específicos

Para o desenvolvimento deste banco de dados será necessário entender como e quais os dados que um investidor utiliza ao tomar um processo de decisão sobre alguma empresa. Aqui estão apresentados alguns dos objetivos específicos que foram realizados para o desenvolvimento desta *library*:

- Entendimento sobre técnicas captação de dados de fontes distintas.
- Estudo sobre o processo de modelagem de banco de dados com intuito de criar estruturas de armazenamento que permitem a associação e resgate de informações de maneira padronizada.
- Pesquisar quais dados públicos são importantes para o processo de investimento, assim como entender quais são suas fontes e a periodicidade de suas atualizações.

2 Fundamentação Teórica

Neste capítulo serão explicadas alguns tópicos do mercado financeiro que serão essenciais para entender a *library* como um todo.

2.1 CVM

A CVM (Comissão de Valores Mobiliários) é uma entidade autárquica vinculada ao ministério da economia que junto a outros órgãos tem a finalidade de disciplinar, fiscalizar e desenvolver o mercado de valores mobiliários. Em especial para o assunto do trabalho de conclusão de curso em questão, as empresas de capital aberto. (CVM,2020).

2.2 Companhias de Capital Aberto

São as companhias que possuem ações que são negociadas na bolsa de valores. Essas ações podem ser compradas por pessoas físicas a partir de 18 anos. Por conta disso, essas empresas possuem responsabilidades reguladas pela CVM. Entre elas está a divulgação do balanço patrimonial e do demonstrativo de resultados.

2.2.1 Balanço Patrimonial

O balanço patrimonial representa todos os bens, direitos e obrigações que uma empresa possui. A sua composição é feita por ativo, que corresponde a todas aplicações de recurso de uma empresa. Passivo, que corresponde a todas obrigações da empresa. O patrimônio líquido, que se refere a todos os recursos próprios da empresa(NETO, 2012).

2.2.2 Demonstrativo de Resultados

A demonstração de resultados tem o objetivo de fornecer os resultados (lucros ou prejuízos) de uma empresa durante o exercício social. Esses resultados vão ser transferidos

diretamente para o patrimônio líquido (NATALIA, 2015).

2.2.3 Individual ou Consolidado

Algumas empresas, por terem investimento em outras empresas, divulgam duas demonstrações de resultado, um referente à empresa maior, a qual é chamado de resultado **individual**, e outra referente a ela e todas suas participações em outras empresas, a qual é chamado resultado **consolidado**. Para o trabalho de curso em questão será considerado apenas a divulgação de resultado da empresa no que se refere ao **consolidado**.

2.2.4 Ações

As ações representam uma fração da companhia a ser vendida, no Brasil elas são negociadas na B3. O preço dela é definido pelo próprio mercado e é relacionado ao seu lucro que dá e sua saúde financeira. Informações essas que apresentadas a partir do Demonstrativo de Resultados (DRE) e do Balanço Patrimonial (BP).

2.2.5 Preferencial e Ordinária

Quando uma mesma empresa é cotada a partir de dois *tickers* geralmente é porque uma se refere a ação ordinária e a outra se refere a ação preferencial.

Enquanto a ação ordinária o investidor tem poder de decisão sobre a empresa (proporcional ao número de ações que possui) e seus dividendos são ganhos a partir da valorização da ação, as ações preferenciais são mais como título de dívida, de tal forma que a retribuição dos dividendos tende a ser fixa e o investidor não tem poder de voto sobre a empresa (GITMAN *et al.*, 2010).

2.3 Variáveis Macroeconômicas

Até esta seção, muito se falou sobre resultados financeiros das empresas a partir de um aspecto endógeno, isto é, como foi a evolução do seu lucro líquido ao longo dos anos, ou o quanto bem a empresa conseguiu diminuir seus custos ao longo dos trimestres para conseguir entregar uma margem de lucro maior e como o preço da ação reflete o quanto bem o mercado prospecta o sucesso desta empresa ao longo dos anos.

Apesar de a análise endógena ser de suma importância para entender a saúde financeira de uma empresa, é necessário entender também que ela presta um serviço ou vende um

produto que está sujeito a fatores exógenos, isto é, que podem alterar seu preço ou suas unidades vendidas sem que a empresa tenha controle sobre ela.

Algumas variáveis que merecem destaque neste paper são o PIB (Produto Interno Bruto), a relação real-dólar e a taxa de juros, pois elas influenciam, direta ou indiretamente, todas as empresas.

2.3.1 Dólar

A relação real dólar mede justamente o grau de valorização (ou desvalorização) da moeda brasileira frente a moeda norte-americana. Dado que o Brasil é uma economia aberta e ainda hoje o dólar é a principal moeda de troca entre os países, é necessário entender bem o quanto que está variável influencia os resultados financeiros de uma empresa.

De fato, dependendo do grau de exposição de uma empresa frente ao mercado externo essa variável pode influenciar mais ou menos os resultados da empresa, seja ela positiva ou negativamente. No que tange à exportação o quanto mais o real se desvaloriza em relação ao dólar mais benéfico é para empresa. Contudo, deve-se entender que a realidade brasileira hoje é que ela é muito dependente de importações, seja de insumos ou de maquinário. Por isso é necessário entender o grau de exposição cambial para que se possa realizar melhores previsões sobre o resultados de uma empresa ao longo dos próximos trimestres.

2.3.2 PIB

O PIB (Produto Interno Bruto) é uma variável trimestral que ilustra o grau de crescimento econômico de um país. No caso do Brasil, empresas que possuem um mercado interno são diretamente correlacionadas a essa variável. Isto é, excluindo fatores endógenos, quão mais o país se desenvolve, mais é esperado que a empresa possua resultados melhores. Apesar disso, algumas empresas são mais ou menos expostas a essa variável. Em se tratando de mercado, quão maior é a perspectiva de crescimento do país, mais as empresas expostas ao mercado interno são valorizadas e em tempos de recessão econômica de um país, empresas menos expostas são mais valorizadas, pois seus resultados são poucos ou menos afetados.

2.3.3 Selic

Pode-se correlacionar a taxa de juros a partir de 2 aspectos diferentes mais correlacionados:

De acordo com o Banco Central do Brasil, "A Selic é a taxa básica de juros da economia. É o principal instrumento de política monetária utilizado pelo Banco Central (BC) para controlar a inflação. Ela influencia todas as taxas de juros do país, como as taxas de juros dos empréstimos, dos financiamentos e das aplicações financeiras."(BCB, 2021). Isto é, no âmbito de contas financeiras quanto maior a SELIC, mais as empresas endividadas terão de pagar de juros, o que prejudica a saúde financeira da empresa por conta dos juros, e mais os bancos são beneficiados, pois geralmente eles são os que emprestam dinheiro. Além disso, como a renda fixa é um substituto para o mercado de renda variável, quanto maior a taxa de selic, mais os investidores optam pelo mercado de renda fixa.

3 Fonte de Dados

Uma vez entendida a importância de cada dado que irá compor a biblioteca, será apresentado neste capítulo suas fontes e como ele é inserido no banco de dados, além de entender como funciona o seu relacionamento entre as outras variáveis. A figura 3.1 ilustra a composição final no banco de dados para que o leitor já consiga entender sua estrutura final e possa consultar caso tenha dúvidas ao longo deste capítulo.

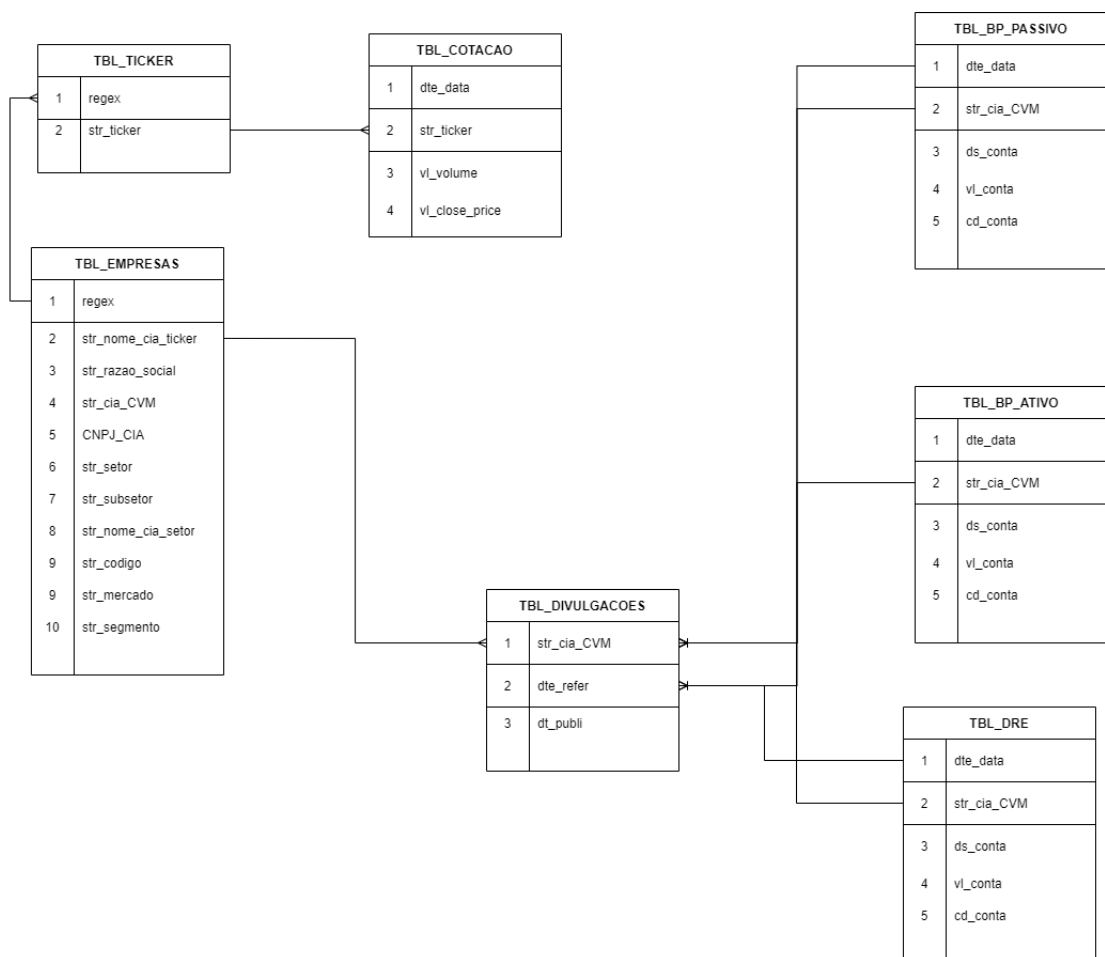


FIGURA 3.1 – Estrutura banco de dados

3.1 Divisão Setorial

Uma vez que cada empresa está inserida em um setor, subsetor e segmento, é necessário capturar essas informações a partir dos dados do site (B3, 2021). Essa informação é útil porque apresenta uma análise mais aprofundada de quais outras empresas são competidoras neste segmento. Neste sentido, a figura 3.2 exemplifica como algumas empresas são divididas a partir de sua atuação. Estes dados serão importantes para popular a **TBL-EMPRESAS**, que tem como objetivo apresentar as informações únicas de cada empresa. As variáveis advindas destas fontes de dados e que fazem parte da **TBL-EMPRESAS** são:

- `str_setor`: Informação referente ao setor econômico da empresa.
- `str_subsetor`: Informação referente ao subsetor econômico da empresa.
- `str_segmento`: Informação referente ao segmento que a empresa atua.
- `str_mercado`: Informação referente ao mercado em que a empresa atua.
- `str_nome_cia_setor`: Informação referente ao nome da empresa de acordo com a B3.
- `str_codigo`: Informação referente ao código da empresa de acordo com a B3.

Essas informações, por serem únicas para cada empresa, estão inseridas na **TBL-EMPRESAS**. Além disso, a variável `str_codigo` é importante porque o *ticker* em que a empresa é negociada na bolsa de valores é composto por esta variável acrescentada de uma numeração (que será explicada na subseção **Cotação**).

SETOR ECONÔMICO	SUBSETOR	SEGMENTO	LISTAGEM	
			CÓDIGO	SEGMENTO
Petróleo, Gás e Biocombustíveis	Petróleo, Gás e Biocombustíveis	Exploração, Refino e Distribuição		
		3R PETROLEUM	RRRP	NM
		COSAN	CSAN	NM
		DOMMO	DMMO	
		ENAUTA PART	ENAT	NM
		PET MANGUINH	RPMG	
		PETROBRAS	PETR	N2
		PETROBRAS BR	BRDT	NM
		PETRORECSA	RECV	NM
		PETRORIO	PRIO	NM
		ULTRAPAR	UGPA	NM
		Equipamentos e Serviços		
		LUPATECH	LUPA	NM
		OCEANPACT	OPCT	NM
		OSX BRASIL	OSXB	NM
Materiais Básicos	Mineração	Minerais Metálicos		
		BRADSPAR	BRAP	N1
		CSNMINERACAO	CMIN	N2
		LITEL	LTEL	MB
		LITELA	LTLA	MB
		MMX MINER	MMXM	NM
		VALE	VALE	NM
		Siderurgia e Metalurgia		
		Siderurgia		
		FERBASA	FESA	N1
		GERDAU	GGBR	N1
		GERDAU MFT	GDAU	N1

FIGURA 3.2 – Classificação Setorial B3

3.2 CVM

No site (CVM, 2021) é possível importar as informações referentes ao BP e ao DRE, em que cada uma dessas fontes de dados possui suas particularidades e devem ser tratadas separadamente. A figura 3.3 ilustra como os dados são apresentados.



A screenshot of the CVM Portal Dados Abertos website. The browser address bar shows 'dados.cvm.gov.br/dados/CIA_ABERTA/DOC/'. The page title is 'Portal Dados Abertos CVM'. Below the title is a table with two columns: 'Name' and 'Last modified'. The table lists several folders: 'Parent Directory', 'DFP/', 'FCA/', 'FRE/', 'IPE/', and 'ITR/'. Each folder entry includes a folder icon and a timestamp.

Name	Last modified
Parent Directory	
DFP/	2020-09-07 14:04
FCA/	2021-02-11 22:03
FRE/	2021-02-11 18:12
IPE/	2021-02-27 11:23
ITR/	2019-11-29 23:01

FIGURA 3.3 – Fonte de dados CVM

Desta fonte de dados da CVM as pastas que possuem informações do DRE e do BP são as pastas **ITR** e **DFP**. Em que o ITR se refere ao formulário que contém as demonstrações contábeis trimestrais das empresas, e a DFP se refere ao formulário de demonstrações contábeis anuais das entidades (ALMEIDA, 2013).

Nessas pastas são apresentados um conjunto de arquivos compactados separados por ano e que possuem informações necessárias para a utilização do código. A figura 3.4 ilustra como estão apresentados os dados dentro desses arquivos compactados.

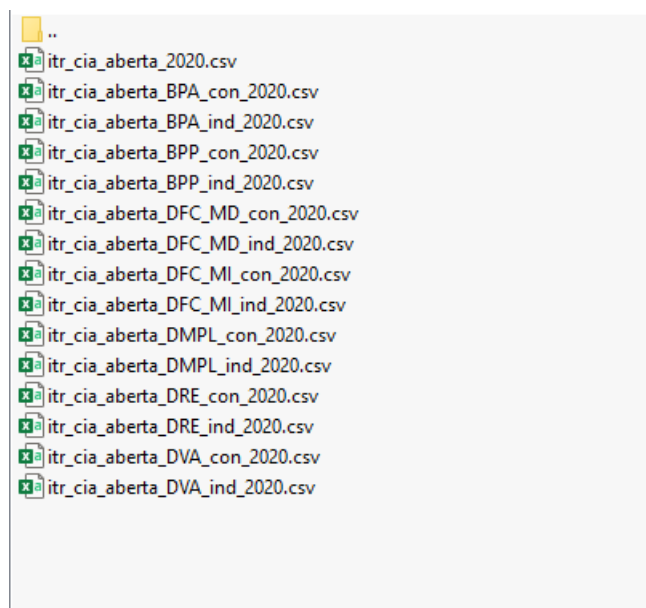


FIGURA 3.4 – Arquivos dentro da pasta compactada ITR 2020

Desse conjunto de arquivos, só serão utilizados os seguintes arquivos:

- `itr_cia_aberta_2020.csv`: Contém o dia específico em que o DRE e o BP foram divulgados.
- `itr_cia_aberta_BPA_con_2020.csv`: Informações do Balanço Patrimonial Ativo no que tange as empresas no âmbito consolidado.
- `itr_cia_aberta_BPP_con_2020.csv`: Informações do Balanço Patrimonial Passivo no que tange as empresas no âmbito consolidado.
- `itr_cia_aberta_DRE_con_2020.csv`: Informações do Demonstrativo de Resultados no que tange as empresas no âmbito consolidado.

Ressalte-se que os dados **DFP** apresentam a mesma estrutura do **ITR**. Com a diferença que o DFP só contém o consolidado do fim do ano. Sua importação é necessária porque o ITR não apresenta os dados do 4^o trimestre das empresas.

3.2.1 Balanço Patrimonial

Uma vez entendido como os dados brutos do BP são baixados, é necessário entender como eles são tratados para serem utilizados como fonte de dados. Ao analisar os dados brutos, seja do BPA ou BPP, é encontrada a seguinte estrutura, conforme ilustrado na figura 3.5

CNPJ_CIA	DT_RE	VER	DENOM_CIA	CÉ	GRUPO_DFP	MOI	ESCALA_MC	ORDEM	DT_FIM	CD_CON	DS_CONTA	VL_CONTA	ST_CONTA_F
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1	Ativo Total		9.260.110.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1	Ativo Total		9.716.450.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01	Ativo Circulante		1.121.010.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01	Ativo Circulante		1.635.620.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.01	Caixa e Equivalentes de Caixa		297.140.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.01	Caixa e Equivalentes de Caixa		803.820.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.02	Aplicações Financeiras		35.800.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.02	Aplicações Financeiras		33.460.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.02.01	Aplicações Financeiras Avaliadas a Valor Justo através do Resultado		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.02.01	Aplicações Financeiras Avaliadas a Valor Justo através do Resultado		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.02.01.01	Títulos para Negociação		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.02.01.01	Títulos para Negociação		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.02.01.02	Títulos Designados a Valor Justo		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.02.01.02	Títulos Designados a Valor Justo		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.02.02	Aplicações Financeiras Avaliadas a Valor Justo através de Outros Resi		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.02.02	Aplicações Financeiras Avaliadas a Valor Justo através de Outros Resi		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.02.03	Aplicações Financeiras Avaliadas ao Custo Amortizado		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.02.03	Aplicações Financeiras Avaliadas ao Custo Amortizado		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.03	Contas a Receber		151.640.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.03	Contas a Receber		158.660.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.03.01	Clientes		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.03.01	Clientes		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.03.02	Outras Contas a Receber		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.03.02	Outras Contas a Receber		0.000000000	\$
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.04	Estoque		330.090.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	ULTIMO	31/03/2020	1.01.04	Estoque		312.360.000.000.000 \$	
33.000.167/0001	31/03/2020	1	PETROLEO BRASILEIRO S.A. PETROBR	9512	DF Consolidado - Balanço Pat REAL	MIL	PENULTIMO	31/12/2019	1.01.05	Ativos Biológicos		0.000000000	\$

FIGURA 3.5 – Dados BPA 2020

Apesar de o BPA e o BPP possuírem exatamente a mesma estrutura, eles foram separados em tabelas diferentes para que o usuário tenha maior facilidade em tratar quais contas do resultado financeiro se referem ao Balanço Patrimonial Ativo e ao Passivo.

O dados advindos deste *xlsx* que servem para popular a **TBL_BP_ATIVO** e **TBL_BP_PASSIVO** são:

- DT_REFER: referente ao trimestre de exercício social da publicação.
- DENOM_CIA: referente ao nome da empresa dada atribuída pela CVM.
- DS_CONTA: referente a descrição de qual conta do Balanço Patrimonial se refere. Por exemplo, no caso da Balança Patrimonial Ativo se é Ativo Circulante, Ativo Não Circulante, Contas a Receber. No caso da Balança Patrimonial Passivo, se é Contas a Pagar, Passivo Circulante, Patrimônio Líquido e etc.
- CD_CONTA: referente ao código de DS_CONTA.
- VL_CONTA: referente ao valor referente ao CD_CONTA em milhares de reais.

Além disso, as variáveis DENOM_CIA, CNPJ_CIA são inseridas na **TBL_EMPRESAS** por se tratar de informações únicas e importantes das empresas da CVM.

3.2.2 Demonstrativo de Resultados

A estrutura do DRE é bem semelhante ao BP, contudo, existem algumas diferenças referentes à metodologia da divulgação do resultado das empresas. Enquanto o trimestre de referência do BP ilustra o quanto a empresa possui de Ativo, Passivo e Patrimônio líquido no trimestre em questão, o DRE ilustra os lucros e prejuízos aferidos ao longo do período analisado. Portanto, analisar o DRE entre 1 e 2 trimestres faz bastante diferença.

Dito isto, é necessário entender não só as colunas que existem na fonte de dados do DRE, mas também como elas são divulgadas ao longo dos trimestres. A figura 3.6 ilustra as colunas do xlsx.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
CNPJ_CIA	DT_REFER	VERSO_DENOM_CIA	DENOM_CIA	CD_CVM	GRUPO_DFP	MODA	ESCALA	ORDEN	EXERC	DT	IN	EXERC	FIM	EXERC	CD	CONTA	DS_CONTA	VL_CONTA	ST_CONTA	FLUXO	ST_FLUXO	ST_CONTA_FIM	ST_FLUXO_FIM	ST_CONTA_FIM	ST_FLUXO_FIM
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.01	Receita de Venda de Bens e/ou Serviços	64.656.570.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.01	Receita de Venda de Bens e/ou Serviços	69.556.360.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.02	Custo dos Bens e/ou Serviços Vendidos	-12.892.520.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.02	Custo dos Bens e/ou Serviços Vendidos	-17.509.950.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.02.01	Energia comprada para revenda	-4.345.320.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.02.01	Energia comprada para revenda	-6.462.200.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.02.02	Encargos sobre o uso da rede elétrica	-2.096.730.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.02.02	Encargos sobre o uso da rede elétrica	-4.464.590.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.02.03	Combustível para produção de energia elé	-5.300.030.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.02.03	Combustível para produção de energia elé	-4.678.980.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.02.04	Construção	-1.125.440.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.02.04	Construção	-1.903.180.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.03	Resultado Bruto	51.764.050.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.03	Resultado Bruto	52.046.410.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.04	Despesas/Receitas Operacionais	-26.536.630.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.04	Despesas/Receitas Operacionais	-28.712.220.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.04.01	Despesas com Vendas	0.000000000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.04.01	Despesas com Vendas	0.000000000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.04.02	Despesas Gerais e Administrativas	-29.969.790.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.04.02	Despesas Gerais e Administrativas	-30.004.870.000.000.000	5														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.04.02.01	Pessoal, Material e Serviço	-18.231.960.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.04.02.01	Pessoal, Material e Serviço	-16.954.810.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.04.02.06	Remuneração de ressarcimento	0.000000000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.04.02.06	Remuneração de ressarcimento	0.000000000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.04.02.07	Depreciação e amortização-imobilizado e	-4.072.650.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.04.02.07	Depreciação e amortização-imobilizado e	-4.470.080.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	PENÚLTIMO	01/01/2019	31/03/2019	3.04.02.08	Amortização	-206.920.000.000.000	N														
00.001.188.31/03/2020	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437 DF Consolidado - Demonstração do REAL	MIL	ÚLTIMO	01/01/2020	31/03/2020	3.04.02.08	Amortização	-222.570.000.000.000	N														

FIGURA 3.6 – Fonte de dados DRE

O dados utilizados para popular a **TBL_DRE** são:

- DT_FIM_EXERC: referente ao trimestre de exercício social da publicação.
- DENOM_CIA: referente ao nome da empresa dada atribuída pela CVM.

- **CD_CONTA**: referente a qual conta do demonstrativo de resultados. Por exemplo, se está se tratando do custo total, do Resultado Financeiro, do Lucro Líquido e entre outros.
- **DS_CONTA**: referente a descrição do que exatamente se refere o **CD_CONTA**.
- **VL_CONTA**: referente ao valor referente ao **CD_CONTA** em milhares de reais.

Para que a **TBL_DRE** possua a mesma estrutura da **TBL_BP_ATIVO**, é necessário padronizar que os lucros (ou prejuízos) registrados na **TBL_DRE** sejam ao longo de um trimestre. Desta forma, caso o usuário queira ver o lucro anual, ele terá de somar os 4 trimestres do ano em questão.

3.3 Outros Dados

Uma vez que os dados da CVM e da B3 foram populados, serão agora importados dados auxiliares que acrescentam informações que enriquecem mais o banco de dados.

3.3.1 Data de Referência e Data de Publicação

Uma empresa costuma divulgar o resultado trimestral depois de 2 a 3 meses do fechamento do trimestre em questão, uma vez que a empresa possui todo um trabalho operacional para averiguar as contas e aferir os resultados. Por conta disso, um analista de dados, ao realizar um estudo sobre os dados presentes nesta *library* deve considerar quais informações estavam disponíveis no dia analisado em questão.

Realizar esta tarefa para uma empresa parece ser fácil, mas ao realizar isso para várias pode ocasionar uma certa problemática, pois esses dados podem ser mais difíceis de serem obtidos. Felizmente, na base de dados da CVM existe este tipo de informação. Esses dados são compilados e são populados na **TBL_DIVULGACOES**. Em que seus componentes são:

- **str_cia_CVM**: nome da companhia dada pela CVM.
- **dte_refer**: data do exercício social.
- **dte_publi**: data exata de quando as informações financeiras são publicadas, referente ao exercício social do **dte_refer**.

3.3.2 Ação e Nome da Empresa

Conforme fora explicado no tópico **Individual ou Consolidado**, as empresas possuem *tickers* correspondentes a sua ação na bolsa de valores. Dito isto, as empresas podem ter mais de um *ticker* negociado. Por exemplo, a empresa farmacêutica Raia Drogasil possui o *ticker* RADL3, enquanto a empresa distribuidora de gasolina e outros componentes Petrobrás possui os *tickers* PETR3 e PETR4. Isto porque a PETR3 é uma ação ordinária e PETR4 é uma ação preferencial, diferenças essas ilustradas no tópico **2.2.5**.

Por conta disso, é necessário buscar em uma outra fonte de dados uma maneira de interligar o nome da companhia listada na CVM com o seu *ticker* da B3. Isso é feito a partir do site (FUNDAMENTUS, 2021), como ilustra a figura 3.7

Selecione o Papel

Papel	Nome Comercial	Razão Social
AALR3	ALLIAR	CENTRO DE IMAGEM DIAGNOSTICOS S.A.
ABCB3	ABC Brasil	BANCO ABC BRASIL S/A
ABCB4	ABC Brasil	BANCO ABC BRASIL S/A
ABEV3	AMBEV S/A	AMBEV S.A.
ABRE3	SOMOS EDUCA	SOMOS EDUCAÇÃO S.A.
ABYA3	ABYARA	ABYARA PLANEJAMENTO IMOBILIARIO S.A.
ACES3	ARCELORMITTAL INOX BRASIL	ARCELORMITTAL INOX BRASIL S.A.
ACES4	ARCELORMITTAL INOX BRASIL	ARCELORMITTAL INOX BRASIL S.A.
ADHM3	ADVANCED-DH	ADVANCED DIGITAL HEALTH MEDICINA PREVENTIVA S.A.
AEDU11	ANHANGUERA EDUCACIONAL PARTICIPAÇÕES SA	ANHANGUERA EDUCACIONAL PARTICIPAÇÕES S.A
AEDU3	ANHANGUERA EDUCACIONAL PARTICIPAÇÕES SA	ANHANGUERA EDUCACIONAL PARTICIPAÇÕES S.A
AELP3	AES ELPA	AES ELPA SA
AERI3	AERIS	AERIS IND. E COM. DE EQUIP. GERACAO DE ENERGIA S/A
AESB3	AES BRASIL	AES BRASIL ENERGIA S.A.
AESL3	AES SUL DISTRIB. GAÚCHA DE ENERGIA S.A.	AES SUL DISTRIB. GAÚCHA DE ENERGIA S.A.
AESL4	AES SUL DISTRIB. GAÚCHA DE ENERGIA S.A.	AES SUL DISTRIB. GAÚCHA DE ENERGIA S.A.
AFLT3	AFLUENTE T	AFLUENTE TRANSMISSÃO DE ENERGIA ELÉTRICA S/A
AFLU3	AFLUENTE	AFLUENTE GER.E TRANSM.ENERG.ELETR. S.A.
AFLU5	AFLUENTE	AFLUENTE GER.E TRANSM.ENERG.ELETR. S.A.
AGEI3	AGRE EMP IMOB	AGRE EMPREENDIMENTOS IMOBILIÁRIOS S.A.
AGEN33	Agrenco	AGRENCO LTD.

FIGURA 3.7 – Site Fundamentus

A tabela deste site é populada nas **TBL_TICKERS** e **TBL_EMPRESAS**. Isto porque enquanto na primeira ocorre o armazenamento de todos os papéis de empresa (str_ticker),

a segunda guarda as informações do nome comercial e a razão social da empresa que vem do (FUNDAMENTUS, 2021). A relação entre essas tabelas é de 1:n, uma vez que enquanto a **TBL_TICKERS** armazena *tickers* diferentes de uma mesma empresa, a outra tabela guarda informações únicas de cada empresa. Em geral, a razão social é o mesmo DENOM_CIA que aparece nos dados oriundos da CVM, exceto com algumas diferenças no que tange à escrita que serão corrigidas via código. Além disso, a variável papel deste *site* é, sem o número no fim da palavra, exatamente a variável código dos dados oriundos da B3 e que tá como str_codigo na **TBL_EMPRESAS**. Desta forma, é possível estabelecer uma completa relação entre as 3 fontes de dados utilizadas (CVM, B3, FUNDAMENTUS).

3.4 Número de Ações

Conforme explicado no item 2.2.4, a ação é uma fração de uma empresa, isto que significa que a multiplicação do número de ações de um papel pela sua cotação equivale ao valor desta empresa de acordo com o mercado. A fonte de dados que possui o número de ações de um *ticker* é (FUNDAMENTUS, 2021), conforme ilustra a figura 3.8.

fundamentus.com.br/detalhes.php?papel=PETR4

Página Inicial	Investimento Consciente	Mais Opções	Entre em Contato	Fundamentus Mobile	Detalhes	Gráficos	Dados Históricos	Histórico de Cotações
Papel	PETR4	Cotação	29,03					
Tipo	PN	Data últ cot	15/06/2021					
Empresa	PETROBRAS PN	Min 52 sem	18,06					
Sector	Petróleo, Gás e Biocombustíveis	Max 52 sem	30,11					
Subsector	Exploração, Refino e Distribuição	Vol \$ méd (2m)	1.921.510.000					
Valor de mercado	378.682.000.000	Últ balanço processado	31/03/2021					
Valor da firma	711.544.000.000	Nro. Ações	13.044.500.000					

FIGURA 3.8 – Número de ações PETR4

Uma vez que uma empresa pode realizar o processo de aumentar ou diminuir o número de ações é preferível que ela não seja inserida no banco de dados, mas que exista uma função de acesso caso o usuário tenha a necessidade de acessar esta variável.

3.5 Cotação

Para se ter acesso a cotação histórica das empresas foi utilizada uma *library* existente no R BatchGetSymbols (PERLIN, 2020), sua cota histórica é populada na **TBL_COTACAO** a partir das seguintes variáveis:

- `str_ticker`: Referente ao ticker da empresa.
- `dte_data`: Referente a data da cotação analisada.
- `vl_close`: Referente ao preço de fechamento da ação do dia correspondente.
- `vl_voluma`: Referente ao volume negociado no dia correspondente.

3.6 IPEA

O Instituto de Pesquisa Econômica Aplicada (IPEA) é uma fundação vinculada ao Ministério da Economia. Suas atividades de pesquisa fornecem suporte técnico e institucional às ações governamentais. (IPEA, 2021)

Para a captura de dados macroeconômicos é utilizado a library `ipeadatar` (IPEA_R, 2021), em que nele é possível capturar informações de juros e cotação do dólar.

3.7 SIDRA

O sidra é um banco de dados mantido pelo Instituto Brasileiro de Geografia e Estatística (IBGE), em que nele é possível armazenar os dados de pesquisa feitos pelo mesmo. (SIDRA, 2021)

Para a captura de dados do sidra é utilizado a library `sidrar` (SIDRAR, 2021), em que nele é possível capturar informações de Inflação e PIB

4 Desenvolvimento

Uma vez que foi apresentada toda a fundamentação teórica dos dados e suas fontes, este capítulo apresentará como esses dados são de fato capturados, tratados e modelados via código R. Também será ilustrada toda a rotina de atualização destes dados.

4.1 Construção BD

Uma vez que o processo de captura de dados passa pelo *download* de diversos dados, além de alguns arquivos precisarem ser descompactados, torna-se viável exportar esses dados para caso seja necessário acessá-los novamente. A função ilustrada na figura 4.1 apresenta as rotinas utilizadas neste processo.

```
construir_bd<-function(anos,diretorio){  
  bulk=serie_CVM(anos)  
  tbl_dre=construir_DRE(bulk)  
  tbl_bp_ativo=construir_ativo(bulk)  
  tbl_bp_passivo=construir_passivo(bulk)  
  depara=construir_depara(bulk)  
  tbl_empresas=depara$tbl_empresas  
  tbl_ticker=depara$tbl_ticker  
  tbl_divulgacoes=construir_release(bulk)
```

FIGURA 4.1 – Função Construir Banco de Dados

Essa função tem como *input* os anos desejados na exportação dos dados da CVM, que começam a partir de de 2012, enquanto a variável diretório guarda a pasta em que será exportado esses arquivos.

4.1.1 Série CVM

Esta função é um *crawler* que captura os arquivos compactados da CVM. Existe uma facilidade em capturar estes dados pois o link para download é padronizado, com a alteração apenas do ano em que se deseja capturar o dado. Depois de guardar as

informações essenciais para o projeto (DRE, BPA, BPP e informações gerais) os outros arquivos são deletados para não gastar memória. A figura 4.2 ilustra o código de captura desses arquivos da CVM.

```
serie_LVM<-function(anos){
  sufixos_desejaveis=c("BPA_con", "BPP_con", "DRE_con", "DRE_ind", "") #VOU ANALISAR SOH OS CONSOLIDADOS
  lista_itr=excel_por_ano(anos,sufixos_desejaveis,"itr")
  lista_dfp=excel_por_ano(anos,sufixos_desejaveis,"dfp")
  return(list("itr"=lista_itr,"dfp"=lista_dfp))
}
excel_por_ano<-function(anos,sufixos_desejaveis,itr_dfp){
  lista=list()
  for(ano in anos){
    temp<-tempfile()
    #Baixa o arquivo por ano
    site_base=paste('http://dados.cvm.gov.br/dados/CIA_ABERTA/DOC/ITR/DADOS/', 'itr_cia_aberta_',ano, ".zip", sep="")
    site_base=str_replace(site_base, "ITR", toupper(itr_dfp))
    site_base=str_replace(site_base, "itr", tolower(itr_dfp))
    download.file(site_base,temp)
    for(sufixo in sufixos_desejaveis){
      nome_sufixo=paste("itr_cia_aberta_",sufixo,sep="")
      nome_sufixo=str_replace(nome_sufixo,"itr",itr_dfp)
      nome_sufixo_ano=paste(nome_sufixo,"_",ano,sep="")
      nome_sufixo_ano=str_replace(nome_sufixo_ano,"_", "_")
      arquivo_sufixo=paste(nome_sufixo_ano, ".csv", sep="")
      #Baixa o arquivo pelo sufixo desejado
      unzip(temp,arquivo_sufixo)
      df=read.csv(arquivo_sufixo,sep=";")
      lista[[nome_sufixo]]=rbind(lista[[nome_sufixo]],df)
      unlink(arquivo_sufixo)
    }
    unlink(temp)
  }
  return(lista)
}
```

FIGURA 4.2 – Crawler CVM

A variável `sufixos_disponiveis` corresponde aos arquivos desejáveis de captura. Caso o usuário deseje capturar os dados na empresa no âmbito individual (conceito explicado no item 2.2.3), deve-se trocar "BPA_con" para "BPA_ind". Além disso, a variável `anos` é passada aqui justamente para capturar todos os anos desejáveis pelo usuário. O *output* dessa função é uma lista correspondendo todos os arquivos capturados, que servirão de insumo para a construção da **TBL_DRE**, **TBL_BPA_ATIVO**, **TBL_BPA_PASSIVO**.

4.1.2 TBL_EMPRESAS

Essa tabela possui informações essenciais de todas as empresas listadas na bolsa, como CNPJ e setor atuante, também é a partir desta tabela que se conecta as 3 principais fontes de dados (CVM, B3 e Fundamentus).

A figura 4.3 ilustra o código de construção dos *tables*.

```

ticker_empresa<-function(){
  html="https://www.fundamentus.com.br/detalhes.php?papel="
  depara=(read_html(html)%>%html_nodes(xpath="//*[@id='test1']")%>%html_table()[[1]])
  colnames(depara)=c("str_ticker", "str_nome_cia_ticker", "str_razao_social")
  #Tira os Espaços
  depara$str_razao_social=stri_trans_general(depara$str_razao_social, "Latin-ASCII")
  #Tira o Traço
  depara$str_razao_social= gsub("-", "", as.character(depara$str_razao_social))
  return (depara)
}

construir_depara<-function(bulk){
  tb1_CVM=data.frame(unique(bulk$itr$itr_cia_aberta)%>%select(str_cia_CVM=DENOM_CIA,CNPJ_CIA))
  tb1_ticker_empresa=ticker_empresa()
  tb1_setores=buscar_setores()

  tb1_empresas=merge(tb1_ticker_empresa%>%mutate(regex=str_replace_all(str_razao_social, "[^[:alnum:]]", ""),
    regex=str_replace_all(regex, "BANCO", "BCO")),
    tb1_CVM%>%mutate(regex=str_replace_all(str_cia_CVM, "[^[:alnum:]]", ""),
    regex=str_replace_all(regex, "BANCO", "BCO")),by="regex"%>%select(-regex)
  tb1_empresas=merge(tb1_empresas%>%mutate(regex=str_replace_all(str_ticker, "[^[:alnum:]]", ""),
    regex=gsub("[0-9]+", "", regex)),
    tb1_setores%>%mutate(regex=str_codigo),by="regex")
  tb1_ticker=tb1_empresas%>%summarise(str_ticker,regex)
  tb1_empresas=tb1_empresas%>%select(-str_ticker)
  tb1_empresas=tb1_empresas[!duplicated(tb1_empresas),]
  return(list("tb1_ticker"=tb1_ticker,"tb1_empresas"=tb1_empresas))
}

```

FIGURA 4.3 – Função Depara

Primeiramente dos dados da CVM captura o nome da companhia dado pela CVM e o CNPJ. Após isso ocorre a captura da tabela do *site* (FUNDAMENTUS, 2021) que possui informações do papel, nome comercial e razão social. Acontece que o nome da empresa dado pela CVM corresponde a razão social do FUNDAMENTUS, com algumas diferenças no que tange a acentuação e espaçamento. Neste sentido, aplica uma função de *regex* que retira esses caracteres especiais, para que enfim essas duas fontes de fato se conectem. Além disso, o nome da empresa dado pela B3 é igual ao *ticker* da empresa com a retirada da numeração, desta forma também é possível conectar os dados da B3 ao do FUNDAMENTUS e da CVM.

4.1.3 TBL_DRE

Uma vez capturado os dados da CVM, pode-se criar a **TBL_DRE**. A figura 4.4 ilustra o processo da criação desta tabela.

```

construir_DRE<-function(bulk){
  #Captura os 3 primeiros trimestre a partir do ITR
  dre=bulk$itr$itr_cia_aberta_DRE_con
  dre_arual=bulk$dre$dre_cia_aberta_DRE_con
  dre=dre%>%filter(ORDEM_EXERC=='ÚLTIMO')%>%filter(month(DT_INI_EXERC)==month(DT_FIM_EXERC)-2)
  #Faz o acumulado dos 3 trimestre para depois subtrair do DFP
  dre_arual=dre_arual%>%filter(ORDEM_EXERC=='ÚLTIMO')%>%rename(VL_ANUAL=VL_CONTA)
  dre_3tri=dre%>%mutate(ano=year(DT_FIM_EXERC))%>%group_by(ano,CD_CONTA,DENOM_CIA)%>%summarise(tres_tri=sum(VL_CONTA,na.rm=TRUE))
  dre_arual=merge(dre_arual%>%mutate(ano=year(DT_REFER)),dre_3tri,by=c("ano","CD_CONTA","DENOM_CIA"))
  dre_arual$VL_CONTA=dre_arual$VL_ANUAL-dre_arual$tres_tri
  dre=rbind(dre%>%dplyr::select(DENOM_CIA,CD_CONTA,VL_CONTA,DT_FIM_EXERC,DS_CONTA),
    dre_arual%>%dplyr::summarise(DENOM_CIA,CD_CONTA,VL_CONTA,DT_FIM_EXERC,DS_CONTA))
  colnames(dre)=c("str_cia_CVM","cd_conta","vl_conta","dte_data","ds_conta")
  return(dre)
}

```

FIGURA 4.4 – Estruturação TBL_DRE

Primeiramente, captura as informações essenciais da fonte de dados trimestrais (ITR). Depois disso, uma vez que esta fonte de dados só apresenta os dados dos primeiros 3

trimestres, realiza-se o acumulado dos 3 primeiros trimestres, agrupado por empresa e por descrição de conta (Lucro Líquido, Resultado financeiro e etc). Em seguida subtrai estes valores da base de dados DFP, uma vez que ela corresponde ao acumulado anual. A figura 4.5 ilustra como a **TBL_DRE** para a Petrobrás é apresentada.

str cia CVM	cd conta	vl conta	dte data	ds conta
PETROLEO BRASILEIRO S.A. PETROBRAS	3.01	8.6174e+07	2021-03-31	Receita de Venda de Bens e/ou Serviços
PETROLEO BRASILEIRO S.A. PETROBRAS	3.02	-4.2141e+07	2021-03-31	Custo dos Bens e/ou Serviços Vendidos
PETROLEO BRASILEIRO S.A. PETROBRAS	3.03	4.4033e+07	2021-03-31	Resultado Bruto
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04	-1.0129e+07	2021-03-31	Despesas/Receitas Operacionais
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.01	-5.1980e+06	2021-03-31	Despesas com Vendas
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.02	-1.4960e+06	2021-03-31	Despesas Gerais e Administrativas
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.03	0.0000e+00	2021-03-31	Perdas pela Não Recuperabilidade de Ativos
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.04	0.0000e+00	2021-03-31	Outras Receitas Operacionais
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05	-4.4540e+06	2021-03-31	Outras Despesas Operacionais
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05.01	-5.8100e+05	2021-03-31	Tributárias
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05.02	-6.3900e+05	2021-03-31	Custo com Pesquisa e Desenvolvimento Tecnológico
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05.03	-1.1960e+06	2021-03-31	Custo Exploratório para Extração de Petróleo e Gás
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05.04	0.0000e+00	2021-03-31	Participação nos Lucros ou Resultados
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05.05	-1.5300e+06	2021-03-31	Outras Despesas/Receitas Operacionais Líquidas
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05.06	0.0000e+00	2021-03-31	Baixa de gastos adicionais capitalizados indevidamente
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.05.07	-5.0800e+05	2021-03-31	Reversão/Perdas no valor de recuperação de ativos - Impair...
PETROLEO BRASILEIRO S.A. PETROBRAS	3.04.06	1.0190e+06	2021-03-31	Resultado de Equivalência Patrimonial

FIGURA 4.5 – DRE Petrobrás

A figura ilustra que a Receita Líquida da Petrobrás ao longo do primeiro trimestre de 2021 foi de 8.6 Bilhões.

4.1.4 TBL_BPA e TBL_BPP

As tabelas envolvendo o balanço patrimonial passivo e balanço patrimonial ativo seguem a mesma estrutura, com a diferença que o valor de uma conta é justamente o quanto a empresa possui no fim daquele exercício social. Por conta de o ITR não possuir os dados do último trimestre das empresas é necessário capturar essas informações do DFP. A figura 4.6 ilustra o balanço patrimonial da Petrobrás no fim do primeiro trimestre de 2021.

str cia CVM	cd conta	vl conta	dte data	ds conta
PETROLEO BRASILEIRO S.A. PETROBRAS	1	998283000	2021-03-31	Ativo Total
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01	153972000	2021-03-31	Ativo Circulante
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.01	68155000	2021-03-31	Caixa e Equivalentes de Caixa
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02	3299000	2021-03-31	Aplicações Financeiras
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.01	0	2021-03-31	Aplicações Financeiras Avaliadas a Valor Justo através do Re...
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.01.01	0	2021-03-31	Titulos para Negociação
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.01.02	0	2021-03-31	Titulos Designados a Valor Justo
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.02	0	2021-03-31	Aplicações Financeiras Avaliadas a Valor Justo através de Ou...
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.03	0	2021-03-31	Aplicações Financeiras Avaliadas ao Custo Amortizado
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.03	13432000	2021-03-31	Contas a Receber
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.03.01	0	2021-03-31	Clientes
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.03.02	0	2021-03-31	Outras Contas a Receber
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.04	39730000	2021-03-31	Estoques
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.05	0	2021-03-31	Ativos Biológicos
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.06	8552000	2021-03-31	Tributos a Recuperar
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.06.01	8552000	2021-03-31	Tributos Correntes a Recuperar
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.06.01.01	2237000	2021-03-31	Imposto de Renda e Contribuição Social correntes

FIGURA 4.6 – Balanço Patrimonial Petrobrás

A tabela indica que a petrobrás possui um ativo total 998 bilhões no primeiro trimestre de 2021.

A estrutura do balanço patrimonial passivo é semelhante ao balanço patrimonial ativo.

4.1.5 TBL_DIVULGACOES

Esta informação já está disponível na fonte de dados da CVM e possui estrutura semelhante ao *xlsx*. O código ilustrado na figura 4.7 apenas captura as colunas essenciais para popular esta tabela.

str cia CVM	cd conta	vl conta	dte data	ds conta
PETROLEO BRASILEIRO S.A. PETROBRAS	1	998283000	2021-03-31	Ativo Total
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01	153972000	2021-03-31	Ativo Circulante
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.01	68155000	2021-03-31	Caixa e Equivalentes de Caixa
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02	3299000	2021-03-31	Aplicações Financeiras
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.01	0	2021-03-31	Aplicações Financeiras Avaliadas a Valor Justo através do Re...
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.01.01	0	2021-03-31	Titulos para Negociação
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.01.02	0	2021-03-31	Titulos Designados a Valor Justo
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.02	0	2021-03-31	Aplicações Financeiras Avaliadas a Valor Justo através de Ou...
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.02.03	0	2021-03-31	Aplicações Financeiras Avaliadas ao Custo Amortizado
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.03	13432000	2021-03-31	Contas a Receber
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.03.01	0	2021-03-31	Clientes
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.03.02	0	2021-03-31	Outras Contas a Receber
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.04	39730000	2021-03-31	Estoques
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.05	0	2021-03-31	Ativos Biológicos
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.06	8552000	2021-03-31	Tributos a Recuperar
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.06.01	8552000	2021-03-31	Tributos Correntes a Recuperar
PETROLEO BRASILEIRO S.A. PETROBRAS	1.01.06.01.01	2237000	2021-03-31	Imposto de Renda e Contribuição Social correntes

FIGURA 4.7 – Estruturação TBL_DIVULGACOES

4.1.6 TBL_COTACOES

Uma vez que a **TBL_TICKERS** guarda as informações dos papéis de todas as empresas, é necessário apenas utilizar a função **BatchGetSymbols** para capturar as informações do preço de fechamento das empresas, assim como o seu volume negociado, como mostra a figura 4.8

```
get_acoes<-function(tickers,data_inicio=as.Date('2000-01-01'),data_fim=Sys.Date()){
  sss=BatchGetSymbols(paste(tickers,".SA",sep=""),first.date=data_inicio,
    last.date=data_fim, bench.ticker="ABVSP")
  acoes=sss$df.tickers
  acoes$ticker=substr(acoes$ticker,1,str_length(acoes$ticker)-3)
  acoes=acoes%>%summarise(dte_data=ref.date,str_ticker=ticker,vl_close=price.close,
    vl_volume=volume)
  return(acoes)
}
```

FIGURA 4.8 – Estruturação TBL_COTACOES

4.2 Atualização

Uma vez que já foi criado e importado todo o banco de dados, é necessário criar uma rotina de atualização para que não seja necessário importar todos os dados.

Para atualizar os dados da CVM nas datas mais recentes é necessário realizar o *download* de todo o arquivo compactado de 2021 e atualizar os dados mais recentes a partir da função **anti_join** da *library dplyr*. Esta função realiza a tarefa de não sobrepor nenhum dado, apenas acrescentar dados novos. Para os dados da **TBL_COTACAO** também se utiliza a função **anti_join** para apenas acrescentar dados novos, conforme ilustra a figura

4.9.

```

atualizar_ponta<-function(diretorio){
  bulk=serie_CVM(2021)

  tbl_dre_ponta=construir_DRE(bulk)
  tbl_bp_ativo_ponta=construir_ativo(bulk)
  tbl_bp_passivo_ponta=construir_passivo(bulk)
  depara_ponta=construir_depara(bulk)
  tbl_empresas_ponta=depara_ponta%>%tbl_empresas
  tbl_ticker_ponta=depara_ponta%>%tbl_ticker
  tbl_divulgacoes_ponta=construir_release(bulk)
  tbl_cotacoes_ponta=get_acoes(tbl_ticker_ponta%>%str_ticker,as.Date('2021-01-01'))

  tbl_dre=read.csv(paste(diretorio,"tbl_dre.csv",sep=""),sep="")
  tbl_bp_ativo=read.csv(paste(diretorio,"tbl_bp_ativo.csv",sep=""),sep="")
  tbl_bp_passivo=read.csv(paste(diretorio,"tbl_bp_passivo.csv",sep=""),sep="")
  tbl_empresas=read.csv(paste(diretorio,"tbl_empresas.csv",sep=""),sep="")
  tbl_ticker=read.csv(paste(diretorio,"tbl_ticker.csv",sep=""),sep="")
  tbl_divulgacoes=read.csv(paste(diretorio,"tbl_divulgacoes.csv",sep=""),sep="")
  tbl_cotacoes=read.csv(paste(diretorio,"tbl_cotacoes.csv",sep=""),sep="")

  tbl_dre=rbind(tbl_dre,anti_join(tbl_dre_ponta,tbl_dre))
  tbl_bp_ativo=rbind(tbl_bp_ativo,anti_join(tbl_bp_ativo_ponta,tbl_bp_ativo))
  tbl_bp_passivo=rbind(tbl_bp_passivo,anti_join(tbl_bp_passivo_ponta,tbl_bp_passivo))
  tbl_empresas=rbind(tbl_empresas,anti_join(tbl_empresas_ponta,tbl_empresas))
  tbl_ticker=rbind(tbl_ticker,anti_join(tbl_ticker_ponta,tbl_ticker))
  tbl_divulgacoes=rbind(tbl_divulgacoes,anti_join(tbl_divulgacoes_ponta,tbl_divulgacoes))
  tbl_cotacoes=rbind(tbl_cotacoes,%>%mutate(dte_data=as.Date(dte_data)),anti_join(tbl_cotacoes_ponta,tbl_cotacoes,%>%mutate(dte_data=as.Date(dte_data))))

  write.csv(tbl_dre,paste(diretorio,"tbl_dre.csv",sep=""),row.names=FALSE)
  write.csv(tbl_bp_ativo,paste(diretorio,"tbl_bp_ativo.csv",sep=""),row.names=FALSE)
  write.csv(tbl_bp_passivo,paste(diretorio,"tbl_bp_passivo.csv",sep=""),row.names=FALSE)
  write.csv(tbl_empresas,paste(diretorio,"tbl_empresas.csv",sep=""),row.names=FALSE)
  write.csv(tbl_ticker,paste(diretorio,"tbl_ticker.csv",sep=""),row.names=FALSE)
  write.csv(tbl_divulgacoes,paste(diretorio,"tbl_divulgacoes.csv",sep=""),row.names=FALSE)
  write.csv(tbl_cotacoes,paste(diretorio,"tbl_cotacoes.csv",sep=""),row.names=FALSE)
}

```

FIGURA 4.9 – Função de atualização da ponta

4.3 Construção de Métricas

Uma vez que todos os bancos de dados estão estruturados e possuem rotinas de atualização diárias, é possível analisar a evolução das contas das empresas ao longo do tempo, assim como é possível comparar empresas de mesmos setores e afins.

4.3.1 Evolução do DRE

Uma vez que todas as empresas possuem DRE estabelecidos é possível analisar como as principais contas de uma empresa evoluiu ao longo dos trimestres. A função `evolucao_dre_por_empresa` mostra como as contas do Demonstrativo de Resultados de uma empresa evoluiu ao longo do tempo, conforme ilustra a figura 4.10

```

quais_mostrar=list("3.01"="Receita_Bruta",
                  "3.02"="Custo",
                  "3.03"="Lucro_Bruto",
                  "3.04"="Despesa_Operacional",
                  "3.06"="Resultado_Financeiro",
                  "3.07"="Ebitda",
                  "3.08"="IR_e_Outros",
                  "3.09"="Lucro_Liquido"
)

empresa_analisada=(merge(tbl_ticker,%>%filter(str_ticker == ticker),
                        tbl_empresas,%>%dplyr::select(regex,str_cia_CVM))%>%summarise(str_ticker,str_cia_CVM))$str_
dre_empresa=tbl_dre%>%filter(str_cia_CVM==empresa_analisada)
dre_empresa=dre_empresa%>%
  group_by(cd_conta)%>%
  arrange(dte_data)%>%
  mutate(vl_valor_acumulado12M=rollapply(vl_conta,FUN=sum,width=4,align='right',partial=FALSE,fill=NA),
         variacao_3M=100*(vl_conta/lag(vl_conta,1)-1),
         variacao_anual=100*(vl_valor_acumulado12M/lag(vl_valor_acumulado12M,1)-1))

info=dre_empresa%>%filter(cd_conta %in% names(quais_mostrar))%>%mutate(str_conta=quais_mostrar[cd_conta])
info=info%>%ungroup()%>%select(dte_data,variacao_3M,variacao_anual,vl_conta,vl_valor_acumulado12M,str_conta)%>%
  gather(variable,value,-c(dte_data,str_conta))
info=info%>%dcast(paste(str_conta,variable)~dte_data)
colnames(info)[1]=ticker
x=t(info)
colnames(x)=x[1,]
x=x[-1,]
x=data.frame(x)

```

FIGURA 4.10 – Função de evolução dre

A função realiza esta análise para uma empresa só, a partir das datas definidas pelo usuário. A variável `quais_mostrar` define quais contas da empresa vão ser analisadas. Essa variável pode ser customizada pelo usuário, e é altamente recomendado que ele o faça, uma vez que cada empresa possui contas específicas que devem ser analisadas mais profundamente. A figura 4.11 evidencia a evolução da Receita Bruta da Petrobrás ao longo dos trimestres.

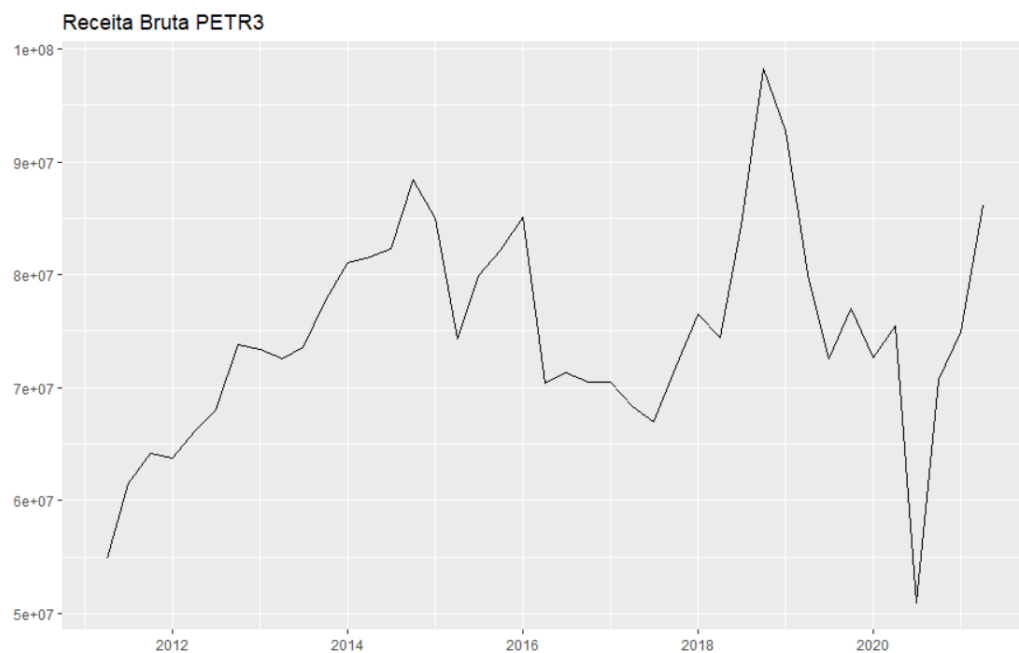


FIGURA 4.11 – Receita Bruta Petrobrás

4.3.2 Evolução do BP

Semelhante a seção anterior, a função `evolucao_bp_por_empresa` mostra como as contas do balanço patrimonial de uma empresa evoluiu ao longo do tempo. A figura 4.12 ilustra como o Ativo Total, Passivo Total e Patrimônio Líquido da Petrobrás evoluíram ao longo do tempo.

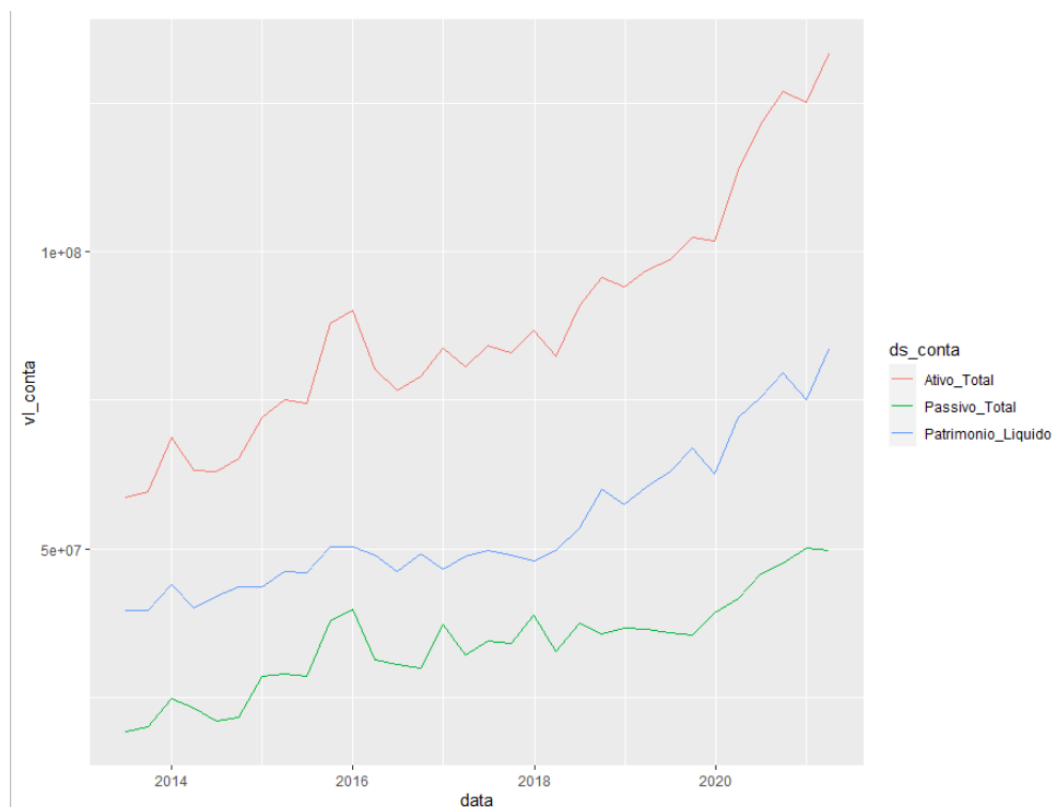


FIGURA 4.12 – Ativo Total, Passivo Total e Patrimônio Líquido da Petrobrás

4.3.3 Métricas por Empresa

Além de entender como o Lucro Líquido de uma certa empresa evoluiu ao longo do tempo, também se faz necessário realizar comparações a partir de outras métricas. Por exemplo: O quanto uma empresa faturou em relação ao quanto que ela gastou, ou de tudo que a empresa vendeu o quanto ela de fato lucrou (depois de ser abatido todos os custos e despesas). Variáveis como essas são necessárias para entender a saúde de uma empresa, por conta disso a função `evolucao_metrica_por_empresa` evidencia justamente essas métricas. Para criar essas métricas, muitas vezes combina-se os dados do DRE e do BP. O fato de essas métricas terem sido construídas evidencia a robustez em que o processo de modelagem foi realizado. A figura 4.13 ilustra quais métricas estão sendo criadas. O usuário possui liberdade para criar mais métricas entre diferentes empresas.

```

indicadores=rbind(ativo_df,
                  passivo_df,
                  dre_anual)%>%group_by(dte_data)%>%
  dcast(dte_data~cd_conta,value.var="vl_conta",fun.aggregate=sum)
numero_de_acoes=buscar_num_acoes(ticker)
indicadores$vl_acao=numero_de_acoes$vl_acao
info=indicadores%>%
  summarise(dte_data,
            #Ativo Circulante/Passivo Circulante
            patrimonio_liquido=(`2.03`-`2.03.09`),
            lucro_liquido_12M=`3.11.01_acumulado`,

            indice_liquidez=`1.01`/`2.01`,
            #Lucro Líquido/Receita Líquida
            margem_liquida=100*`3.11.01`/`3.01`,
            #Ebit: Ebit/Receita Líquida
            margem_ebit=100*(`3.05`-`3.04.06`-`3.04.05`-`3.04.04`)/(`3.01`),
            #Margem Bruta: Lucro Bruto/Receita Líquida
            margem_bruta=100*`3.03`/`3.01`,
            #Giro Ativo: Receita Líquida (acumulado 12 meses)/Ativo Total
            giro_ativo=(`3.01_acumulado`/`1`),
            #Divida bruta / patrimonio liquido
            divida_bruta_patrimonio=((`2.01.04`+`2.02.01`)/patrimonio_liquido),
            #ROE: Lucro Líquido (acumulado 12 meses)/patrimonio liquido
            ROE=100*lucro_liquido_12M/patrimonio_liquido,
            #EBIT (acumulado 12 meses)/ativo
            ebit_ativo=100*(`3.05_acumulado`-`3.04.06_acumulado`-`3.04.05_acumulado`-`3.04.04_acumulado`)/(`1`),

            lucro_liquido_acao=1000*lucro_liquido_12M/vl_acao,
            valor_patrimonial_acao=1000*patrimonio_liquido/vl_acao)

```

FIGURA 4.13 – Métricas por Empresa

4.3.4 Métricas via Preço

As métricas utilizadas até agora utilizavam apenas informações financeiras. Contudo, existem métricas que são referenciadas pela cotação de uma empresa. Isto é, o quanto a empresa vale hoje de acordo com o mercado, que é calculada pela cotação da empresa X o número de ações dela. Essas variáveis servem para medir o grau de expectativa do mercado em relação àquela empresa e aos seus lucros futuros. Como as informações de cotação da empresa, lucro líquido dela e até mesmo quando o resultado dela foi divulgado estão presentes no banco de dados, é possível montar a série dessa expectativa para os anos anteriores, a figura 4.14 ilustra a expectativa de Preço/Lucro Líquido anual das empresas Raia Drogasil e Ambev de 2017 até o dia 24/06/2021.

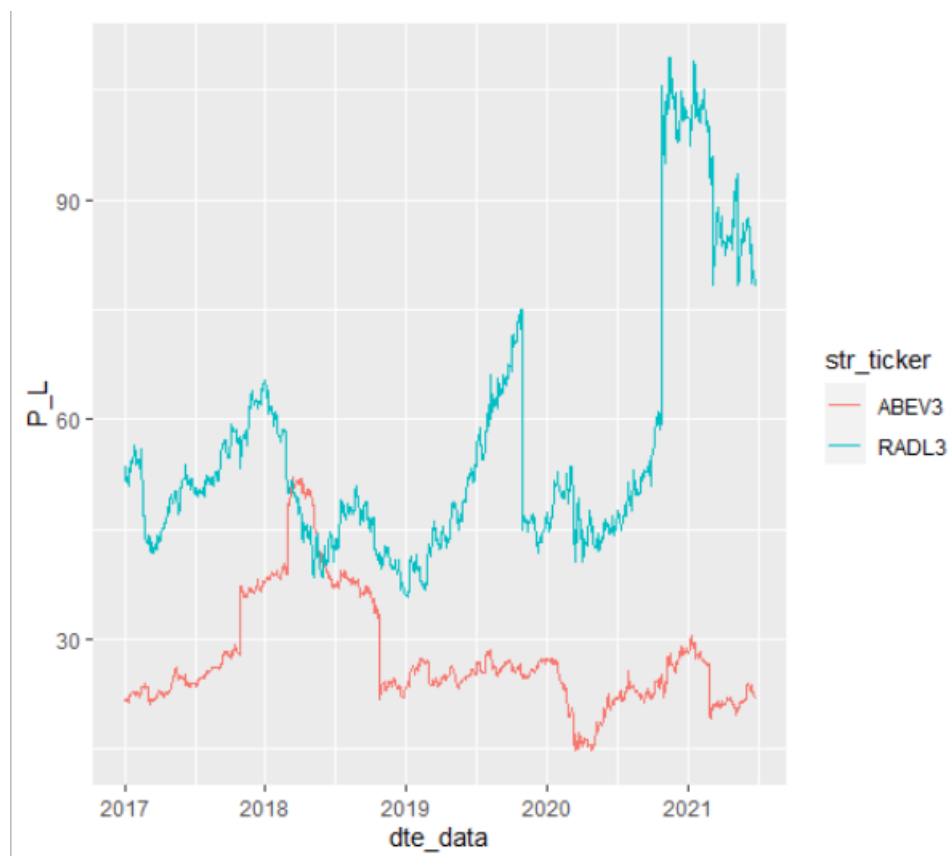


FIGURA 4.14 – Métricas por Empresa

4.4 Visualização

Uma vez feita toda a captura e manipulação dos dados, foi construído um *dashboard* para que se possa visualizar melhor os dados. A figura 4.15 ilustra o painel de visualização.

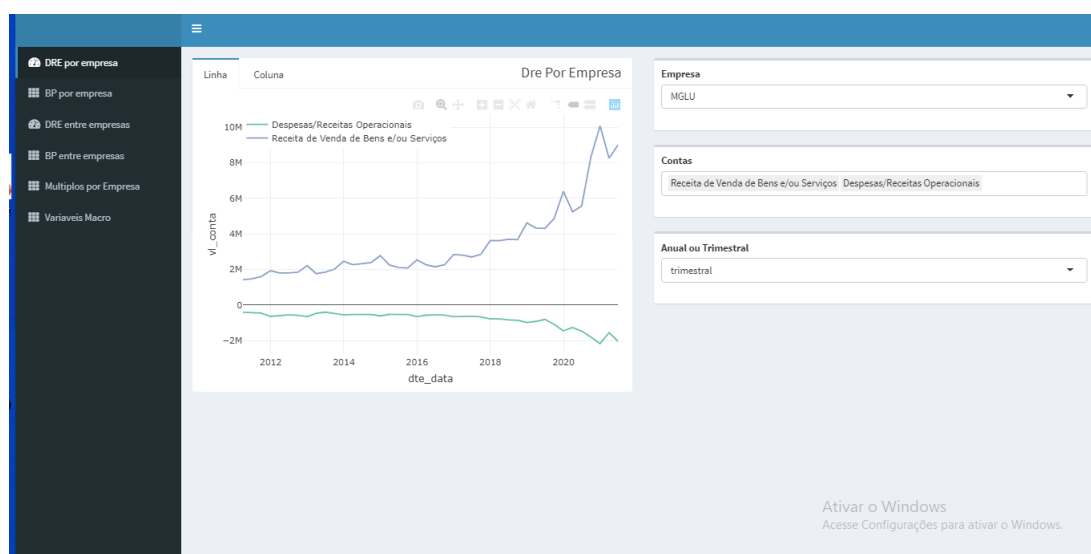


FIGURA 4.15 – Dashboard

Para a construção do *dashboard* foi utilizado o Rshiny, que é uma *library* que facilita a criação de página web. (RSHINY, 2021)

Além disso foi utilizado o ShinyDashboard, que facilita ainda mais a criação de dashboard a partir de funções prontas oriundas do shiny (RSHINYDASHBOARD, 2021)

E para a visualização de gráficos foi utilizado o plotly, pois ele permite criar gráficos iterativos com o intuito de que os usuários possam realizar análises mais aprofundadas sobre os dados financeiros das empresas. (PLOTLY, 2021)

Como *template* pré pronto foi disponibilizado algumas abas para que os usuários possam realizar análises iniciais dos dados. Ressalte-se aqui que todos esses dados são manipulados após a construção do banco de dados que foi explicado nas seções anteriores, entre eles são:

- DRE por Empresa: Analisar as diversas contas do Demonstrativo de Resultados por cada empresa, essa visualização pode ser trimestral ou anual, e pode ser mostrada em formato de coluna ou de linha.
- BP por Empresa: Analisar as diversas contas do Balanço Patrimonial por cada empresa, essa visualização é feita em formato de coluna ou de linha.
- DRE entre empresas: Analisar diversas empresas a partir de uma mesma métrica presente no Demonstrativo de Resultados, como Lucro Líquido, Resultado bruto e entre outros.
- BP entre empresas: Analisar diversas empresas a partir de uma mesma métrica presente no Balanço Patrimonial, como Ativo Total, Passivo Total e entre outros.
- Múltiplo entre Empresas: Analisar os múltiplos e margens das empresas a partir de uma mesma métrica, como Margem Bruta, Margem Líquida, Margem EBITDA e entre outros. Além de ser possível visualizar a cotação das empresas ao longo dos anos.
- Variáveis Macro: Analisar alguns dados macroeconômicos da economia brasileira, como PIB, inflação e entre outros.

4.4.1 Construção

Este tópico será útil apenas para quem deseja costumar ainda mais o *dashboard*. Para quem só deseja o banco de dados em si, este tópico não o lhe será útil.

Caso o usuário queira adicionar, ou remover algum gráfico ele só precisará se preocupar com 3 coisas:

- Body: Corresponde ao front-end do painel
- Server: Responsável por receber os inputs e criar os outputs (no dashboard em questão, a renderização do gráfico).
- Função: "Backend" do painel, onde receberá os inputs do server e realizará operações para criar o output.

Conforme ilustra a figura 4.16, a linguagem utilizada pelo *rshinyboard* lembra o *bootstrap* do html. Além disso, utiliza-se a função **selectInput** para criar uma janela suspensa em que o usuário pode definir o valor atribuído a uma variável. Por fim, a função **plotlyOutput** printa o gráfico que foi renderizado pelo server

```
variaveis_macro= tabItem(tabName="variaveis_macro",
  fluidRow(tabBox(title="Variavel Macro",id="tabset1",height="250px",
    tabPanel("Coluna",plotlyOutput("variaveis_macro"))),
    box(selectInput("input_variavel_macro", "Variavel Macro",
      choices = c("PMS_RECEITA", "abate", "juros", "dolar",
        "inflacao", "pib", "PIM", "PMS_VOLUME",
        "PMS_RECEITA")))))
```

FIGURA 4.16 – Body RShinyboard

Um outro campo importante é o server, que captura o valor das variáveis registradas no body e renderiza o gráfico que é retornado no backend do código, conforme ilustra a figura 4.17.

```
server <- function(input, output) {
  #Linha
  output$dre_por_empresa_linha <- renderPlotly({
    montar_DRE_por_empresa(tbl_dre=dfs$tbl_dre,
      tbl_ds_conta=dfs$tbl_ds_conta,
      empresa_selec = input$input_empresa_unique_dre,
      anual_trimestral=input$input_anual_trimestral,
      contas_selec=input$input_contas_multiple_dre,
      linha_coluna="linha"))}
}
```

FIGURA 4.17 – Server RShinyboard

Por fim, é no back do código em que ocorre de fato a criação do gráfico a partir do banco de dados já tratado, o retorno da função é o gráfico, mas poderia ser uma tabela, figura e entre outros, conforme a necessidade do usuário, ilustrado na figura 4.18.

```
#variavel_macro="pib"
montar_variavel_macro<-function(variavel_macro){
  print(variavel_macro)
  if(variavel_macro %in% c("dolar","juros")){
    df=importar_sgs(variavel_macro)
    colnames(df)=c("dte_data", "vl_valor")
  }else{
    df=importar_sidra(variavel_macro)
    colnames(df)=c("dte_data", "vl_valor")
  }
  return(plot_ly(df,
    x=~dte_data,
    y=~vl_valor,
    type='scatter',
    mode='lines'))
}
```

FIGURA 4.18 – Back RShinyboard

4.5 Documentação e GIT

Com o intuito de democratizar o acesso a esses dados, a *library* foi disponibilizada no git com o link: https://github.com/Folksz/projeto_B3_crawler, nela está presente o todos os códigos, a documentação deles e este relatório, para caso os usuários queira saber mais sobre as fontes de dados que originaram esta *library*, também está disponível o código que permita visualizar esses dados em um *dashboard*, para que iniciantes possam se familiarizar com os dados, a figura 4.19 ilustra os arquivos que estão na página até a presente data (GIT, 2021).

ksz / projeto_B3_crawler Public

Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags

Go to file Add file Code

Folksz Add files via upload ae69e65 2 hours ago 9 commits

File/Folder	Commit Message	Time Ago
R	atualizacao	3 hours ago
man	atualizacao	3 hours ago
.Rbuildignore	atualizacao	3 hours ago
.gitignore	atualizacao	3 hours ago
DESCRIPTION	atualizacao	3 hours ago
Documentacao_projeto_B3_crawler....	Add files via upload	2 hours ago
LICENSE	Add files via upload	5 hours ago
LICENSE.md	Add files via upload	5 hours ago
NAMESPACE	atualizacao	3 hours ago
crawlerB3.Rproj	Add files via upload	5 hours ago
shiny_backend.R	atualizacao	3 hours ago

FIGURA 4.19 – GIT

Conforme ilustrado, além dos códigos existe a documentação para que usuários possam visualizar e entender melhor as funções que compõem esta *library*, conforme ilustra a figura 4.20.

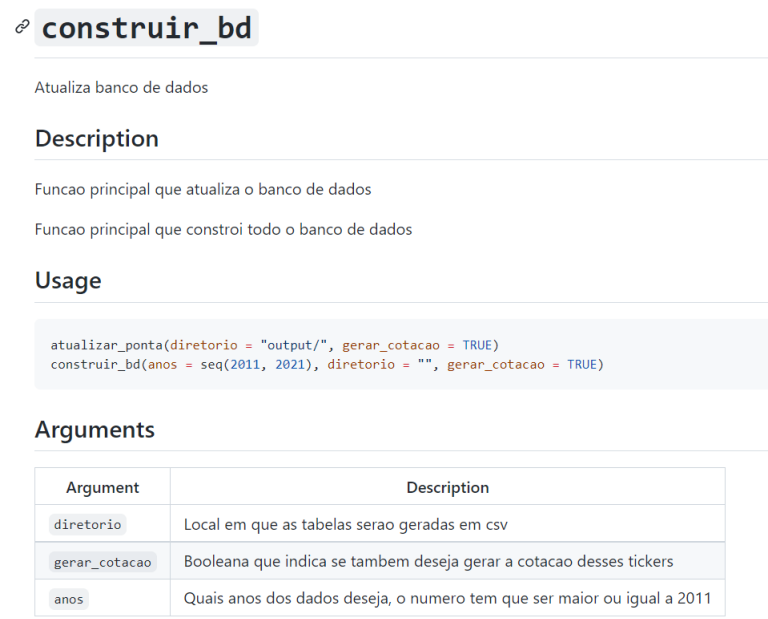


FIGURA 4.20 – documentacao *library*

Ademais, ao importar devidamente a library, as informações sobre as funções que compõem o projeto podem ser facilmente acessadas a partir do comando *help*, conforme ilustrada na figura 4.21

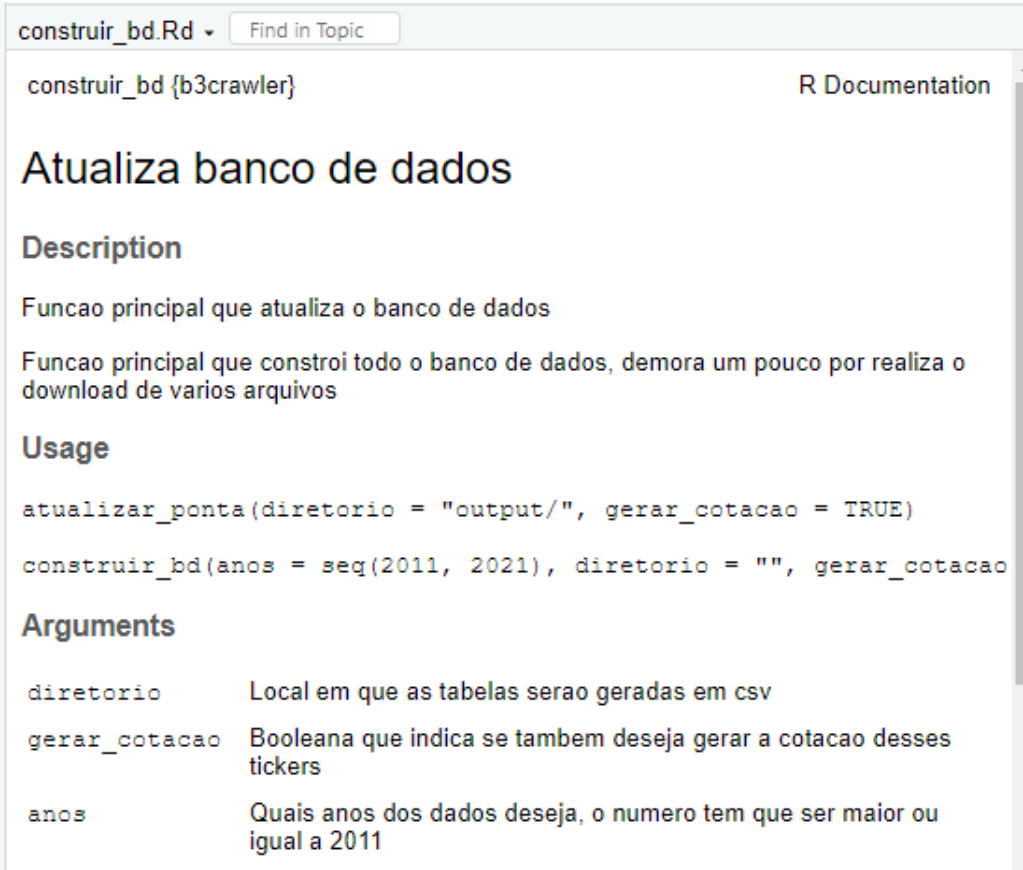


FIGURA 4.21 – help função construir_bd

Por fim, a criação deste projeto só foi possível por conta de diversas *libraries* já prontas que de alguma forma viabilizou a confecção deste trabalho de maneira mais rápida e eficiente. As *libraries* que foram utilizadas neste projeto em algum momento está disponível na figura 4.22

```
Package: b3crawler
Title: Crawler B3
Version: 0.0.0.9000
Authors@R:
  person(given = "Pedro",
         family = "Cavalcante",
         role = c("aut", "cre"),
         email = "pedrohen.fc@gmail.com")
Description: Crawler dados da B3.
License: MIT + file LICENSE
Encoding: UTF-8
LazyData: true
Roxygen: list(markdown = TRUE)
RoxygenNote: 7.1.1
Imports:
  stringr,
  dplyr,
  rvest,
  xml2,
  BatchGetSymbols,
  xlsx,
  zoo,
  stringi,
  sidrar,
  ipeadatar,
  utils,
  lubridate,
  plotly,
  reshape2,
  shinydashboard,
  shiny
```

FIGURA 4.22 – funcoes auxiliares *library*

5 Conclusão

Criar e modelar um banco de dados como o que fora feito neste trabalho passa por diversos processos complexos e trabalhosos, principalmente quando se tratar de dados de diferentes fontes e que são capturados via diferentes técnicas de *crawler*.

Para chegar até este passo, foi necessário entender a fundamentação teórica (que envolve conceitos intermediários de economia), diferentes técnicas de captura (uma vez que foram diferentes fontes de dados) e, por fim, foi necessário consultar a literatura para que se pudesse alocar todas estas informações maneira ótima e eficiente respeitando os aspectos de inserção, busca e remoção de dados.

Todos esses passos foram árduos e exigiram uma certa técnica e esforço que não seria possível sem a ajuda de diversos contribuintes, seja dos colaboradores que mantêm esses dados disponíveis na internet, seja dos mais diversos usuários que criaram funções e libraries (ilustrados na figura 4.22) que possibilitaram o uso de atalhos para o desenvolvimento deste projeto e, principalmente, de professores e orientadores que me ensinaram não só a desenvolver este projeto como também a me tornar um profissional melhor.

Espero que de alguma forma este trabalho possa auxiliar alguns indivíduos, seja investidores iniciantes ou cientista de dados, a alcançar seus objetivos profissionais ou financeiros, assim como fui auxiliado a alcançar os meus.

Referências

ALMEIDA, A. P. d. O impacto da concentração da evidenciação contábil no mercado acionário. 2013.

B3. 2021. [Online; accessed 10-June-2021]. Available at: http://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/acoes/consultas/classificacao-setorial/.

BCB. 2021. [Online; accessed 06-November-2021]. Available at: <https://www.bcb.gov.br/controleinflacao/taxaselic>.

CARDOZO, T. T. M.; MODESTO, N. L. P.; MAGALHÃES, N. P.; FONSECA, R. V. S.; POLICARPO, R. V. S. Análise do perfil de investidores brasileiros. 2019.

CVM. 2021. [Online; accessed 10-June-2021]. Available at: http://dados.cvm.gov.br/dados/CIA_ABERTA/DOC/.

FUNDAMENTUS. 2021. [Online; accessed 10-June-2021]. Available at: <https://fundamentus.com.br/consciente.php>.

GIT. 2021. [Online; accessed 08-November-2021]. Available at: https://github.com/Folksz/projeto_B3_crawler.

GITMAN, L. J. *et al.* Princípios de administração financeira. Pearson Education do Brasil, 2010.

IPEA. 2021. [Online; accessed 06-November-2021]. Available at: https://www.ipea.gov.br/portal/index.php?option=com_contentview=articleid=1226Itemid=68.

IPeA_R. 2021. [Online; accessed 06 – November – 2021]. Available at : <https://cran.r-project.org/web/packages/ipeadatar/ipeadatar.pdf> .

NATALIA, D. Análise das demonstrações financeiras. SESES, 2015.

NETO, A. A. **Estrutura e análise de balanços: um enfoque econômico-financeiro: comércio e serviços, indústrias, bancos comerciais e múltiplos.** [S.l.]: Atlas, 2012.

PERLIN, M. Batchgetsymbols: Downloads and organizes financial data for multiple tickers. **R package version**, v. 2, n. 7, 2020.

PLOTLY. 2021. [Online; accessed 06-November-2021]. Available at:
<https://plotly.com/r/>.

RSHINY. 2021. [Online; accessed 06-November-2021]. Available at:
<https://shiny.rstudio.com/tutorial/>.

RSHINYDASHBOARD. 2021. [Online; accessed 06-November-2021]. Available at:
<https://cran.r-project.org/web/packages/shinydashboard/shinydashboard.pdf>.

SIDRA. 2021. [Online; accessed 06-November-2021]. Available at:
<https://sidra.ibge.gov.br/home/pimpfbr/brasil>.

SIDRAR. 2021. [Online; accessed 06-November-2021]. Available at:
https://cran.r-project.org/web/packages/sidrar/vignettes/Introduction_to_sidrar.html.