

Dashboard para Cotações Históricas das Empresas Listadas na B3

André de Araujo Sholl

Laboratório de Inteligência Computacional Aplicada (ICA) – Pontifícia
Universidade Católica do Rio de Janeiro (PUC-Rio) – RJ – Brasil

andresholl@gmail.com

Abstract. This article describes the steps of data collection (which includes downloading CSV files and web scraping from B3 and CVM sites), processing the collected data, creating the database in SQLite, importing the data and graphical visualization in dynamic dashboard.

Resumo. Este artigo descreve as etapas de coleta de dados (que compreende o download de arquivos CSV e web scraping dos sites da B3 e CVM), de tratamento dos dados coletados, da criação do banco de dados em SQLite, importação dos dados e visualização gráfica em dashboard dinâmico.

1. Dados

Os dados utilizados neste artigo foram extraídos dos sites da B3 e CVM:

- Séries Históricas
- Empresas Listadas
- Formulário de Demonstrações Financeiras Padronizadas (DFP)

2. Softwares

No desenvolvimento deste artigo foram utilizados os seguintes softwares:

- Python 3
- Banco de Dados SQLite
- Microsoft Power BI

A linguagem Python é uma linguagem flexível, de código aberto e que possui enorme quantidade de bibliotecas disponíveis para Ciência de Dados, tais NumPy, pandas, scikit-learn, tensorflow, matplotlib dentre outras. Muitos

tutoriais e literatura estão disponíveis na internet e podem ser acessados facilmente.

O SQLite é uma base de dados relacional de código aberto e que dispensa o uso de um servidor para armazenar dados. Ao mesmo tempo em que dispensa um servidor, o SQLite também não demanda nenhum tipo de configuração. Assim, seu uso como base de dados para diferentes tipos de aplicações se torna, além de fácil, mais fluido, dinâmico e leve.

O Microsoft Power BI é uma solução da Microsoft que entrega relatórios e dashboards altamente intuitivos e dinâmicos que permite ao analista obter insights a partir de dados, tomando decisões rápidas e assertivas. Entre as principais vantagens podemos citar: coleta, tratamento e análise de grandes volumes de dados; possibilidade de cruzar informações de várias fontes de dados diferentes; compartilhamento de relatórios na Web; está disponível em diversos dispositivos (computadores, smartphones, tablets) e é altamente customizável.

3. Desenvolvimento

As etapas de coleta de dados (que compreende o download de arquivos CSV e web scraping dos sites da B3 e CVM), de tratamento dos dados coletados e da criação do banco de dados em SQLite foram realizadas através de um jupyter notebook em python 3. Foram utilizadas as seguintes bibliotecas: numpy, time, datetime, pandas, string, wget, zipfile, random, sqlite3 e selenium.

A CVM disponibiliza diversos formulários de demonstrações financeiras desde o ano de 2011 no site <http://dados.cvm.gov.br>. A etapa de coleta de dados no site da CVM consiste em fazer o download dos arquivos e concatená-los para posterior ingestão no banco de dados.

A B3 também disponibiliza informações no site www.b3.com.br. As séries históricas são disponibilizadas em arquivos CSV em séries anuais, mensais ou diárias. Estes arquivos estão formatados por largura de coluna de forma que há necessidade de ajustar estes arquivos conforme layout disponível no site da B3. Após este pré-processamento os arquivos são concatenados para posterior importação no banco de dados. Por outro lado, dados básicos das empresas listadas na B3 não estão disponíveis através de arquivos CSV. Assim, é necessário empregar a técnica de web scraping para extrair informações básicas das empresas (código CVM, CNPJ, razão social, site, código de negociação etc.). O site da B3 emprega muito a linguagem javascript

e iframes dentro de iframes de forma que a extração destas informações não é uma tarefa trivial. Felizmente, o python possui a biblioteca selenium que permite renderizar em um browser, simulando a navegação realizada por ser humano, ao mesmo tempo que coleta as informações. Todas estas informações foram salvas num arquivo CSV.

Após a coleta e tratamentos dos dados, os arquivos CSVs criados foram importados em um banco de dados SQLite. O script notebook está disponível no site https://github.com/asholl/tcc_bi_master.

Com o banco de dados criado, a última etapa é criar uma conexão entre o banco de dados e o Power BI. O Power BI não suporta nativamente o banco de dados SQLite. Para que seja possível a conexão, é preciso baixar um drive ODBC para SQLite disponível no site <http://www.ch-werner.de/sqliteodbc/>.

Para criar a conexão ao banco de dados, é preciso criar uma conexão ODBC no PowerBI, escolher SQLite no menu, informar o local do banco de dados e escolher as tabelas que serão importadas conforme figuras 1 e 2.

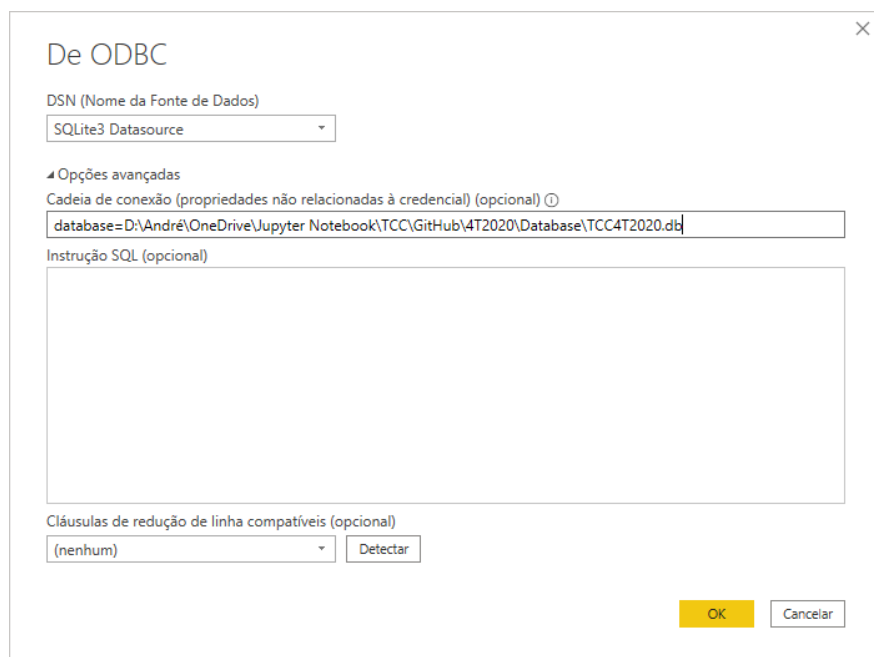


Figura 1. Estabelecendo uma conexão ao banco de dados

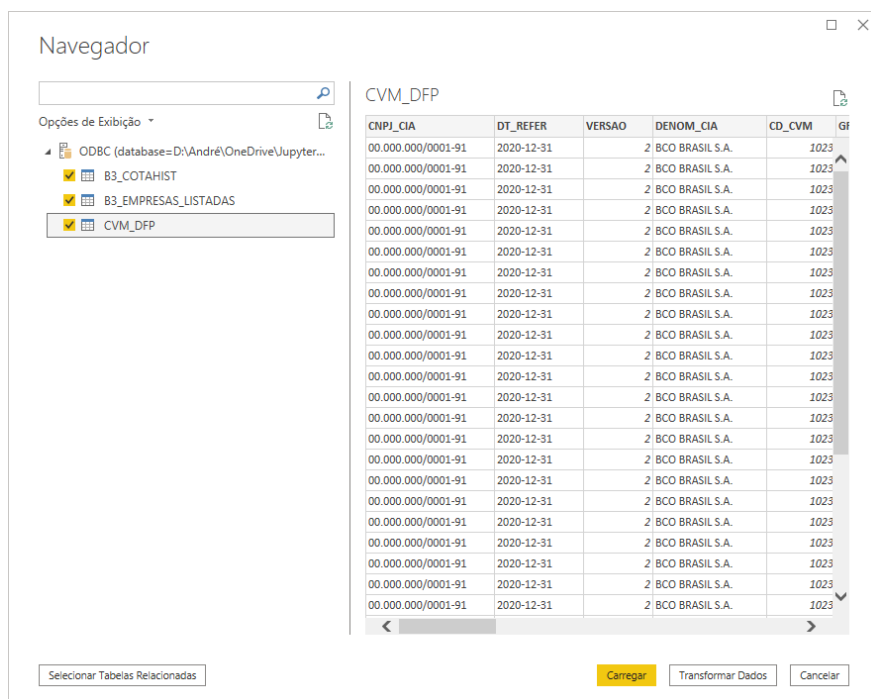


Figura 2. Importando as tabelas do banco de dados

Após a importação das tabelas, são necessários alguns passos: ajustar o tipo de dados para as colunas de datas (definir o padrão de data) e código CVM (alterar de inteiro para texto) e definir os relacionamentos. As tabelas são relacionadas entre si conforme as figuras 4, 5 e 6.

[illegible]

Figura 3. Ajustando os tipos de dados

Editar relacionamento

Selecione tabelas e colunas relacionadas umas às outras.

B3_COTAHIST

TIPREG	DATPRE	CODBDI	CODNEG	TPMERC	NOMRES	ESPECI	PRAZOT	MODREF	PREABE
1	2020-10-01	82	B3SAX546	80	B3SAE /ED	ON NM	0 R\$		3,7
1	2020-10-02	82	B3SAX589	80	B3SAE /ED	ON NM	0 R\$		6,4
1	2020-10-05	82	YDUQX294	80	YDUQE	ON NM	0 R\$		3,6

B3_EMPRESAS_LISTADAS

Razão Social	Links	Nome de Pregão:	
UM ÓLEO E GÁS S.A	http://bvmf.bmfbovespa.com.br/cias-listadas/empresa...	3R PETROLEUM	12
E COM. DE EQUIP. GERACAO DE ENERGIA S/A	http://bvmf.bmfbovespa.com.br/cias-listadas/empresa...	AERIS	12
ENERGIA S.A.	http://bvmf.bmfbovespa.com.br/cias-listadas/empresa...	AES BRASIL	37

Cardinalidade

Muitos para um (*:1)

Direção do filtro cruzado

Único

☒ Ativar este relacionamento
 ☐ Aplicar filtro de segurança em ambos os sentidos

☐ Pressuponha integridade referencial

OK

Cancelar

Figura 4. Definindo o relacionamento entre as tabelas B3_COTAHIST e B3 EMPRESAS LISTADAS

Editar relacionamento

Selecione tabelas e colunas relacionadas umas às outras.

CVM_DFP ▾

CNPJ_CIA	DT_REFER	VERSAO	DENOM_CIA	CD_CVM	
00.001.180/0001-26	2020-12-31	1	CENTRAIS ELET BRAS S.A. - ELETROBRAS	2437	DF Consolidada
00.070.698/0001-11	2020-12-31	1	CIA ENERGETICA DE BRASILIA	14451	DF Consolidada
00.359.742/0001-08	2020-12-31	1	ATOM EMPREENDIMENTOS E PARTICIPAÇÕES S.A.	15423	DF Consolidada

< >

B3_EMPRESAS_LISTADAS ▾

os ISIN3:	Códigos ISIN4:	Códigos ISIN5:	Códigos ISIN6:	Códigos ISIN7:	Códigos ISIN8:	Códigos CVM:
null	null	null	null	null	null	25291
null	null	null	null	null	null	25283
null	null	null	null	null	null	25640

< >

Cardinalidade

Muitos para um (*:1) ▾

Direção do filtro cruzado

Único ▾

☒ Ativar este relacionamento
 ☐ Aplicar filtro de segurança em ambos os sentidos

☐ Pressuponha integridade referencial

OK

Cancelar

Figura 5. Definindo o relacionamento entre as tabelas CVM_DFP e B3_EMPRESAS_LISTADAS

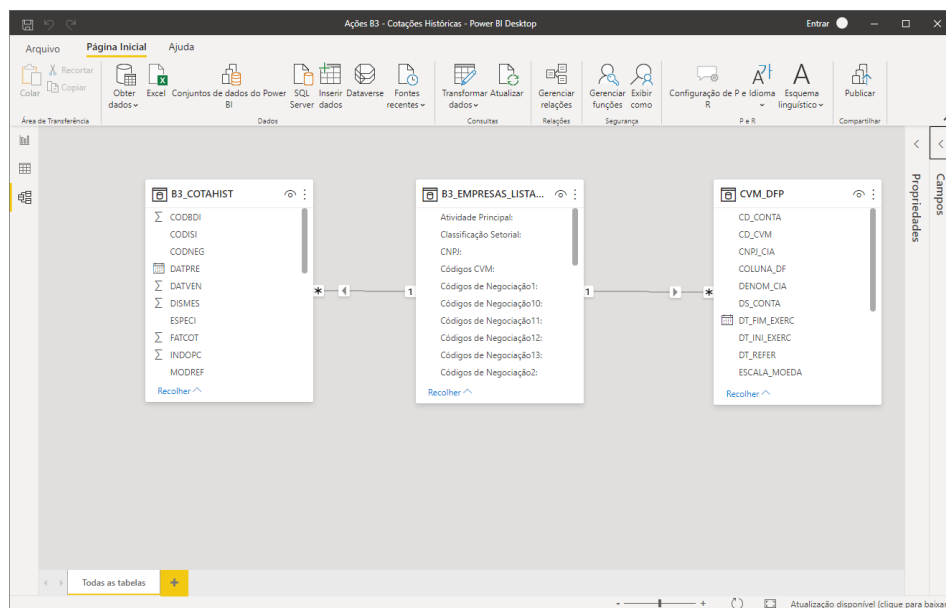


Figura 6. Relacionamentos entre as tabelas

4. Telas Dashboard



Figura 7. Tela de abertura

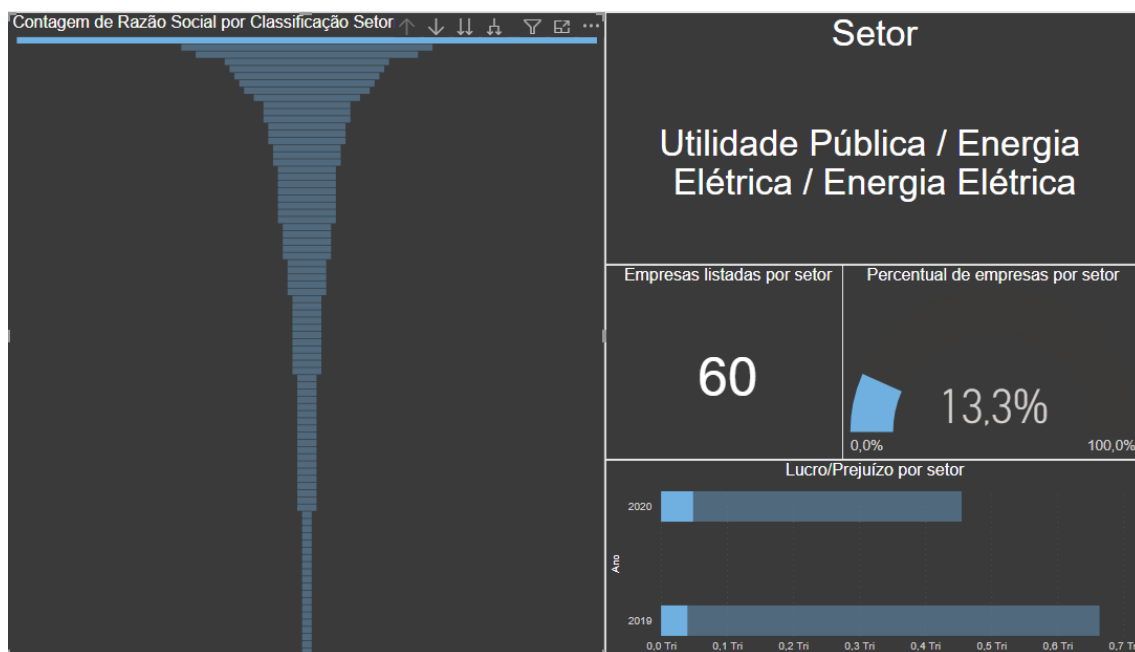


Figura 8. Tela que apresenta a quantidade de empresas listadas, o percentual e o lucro ou prejuízo por setor

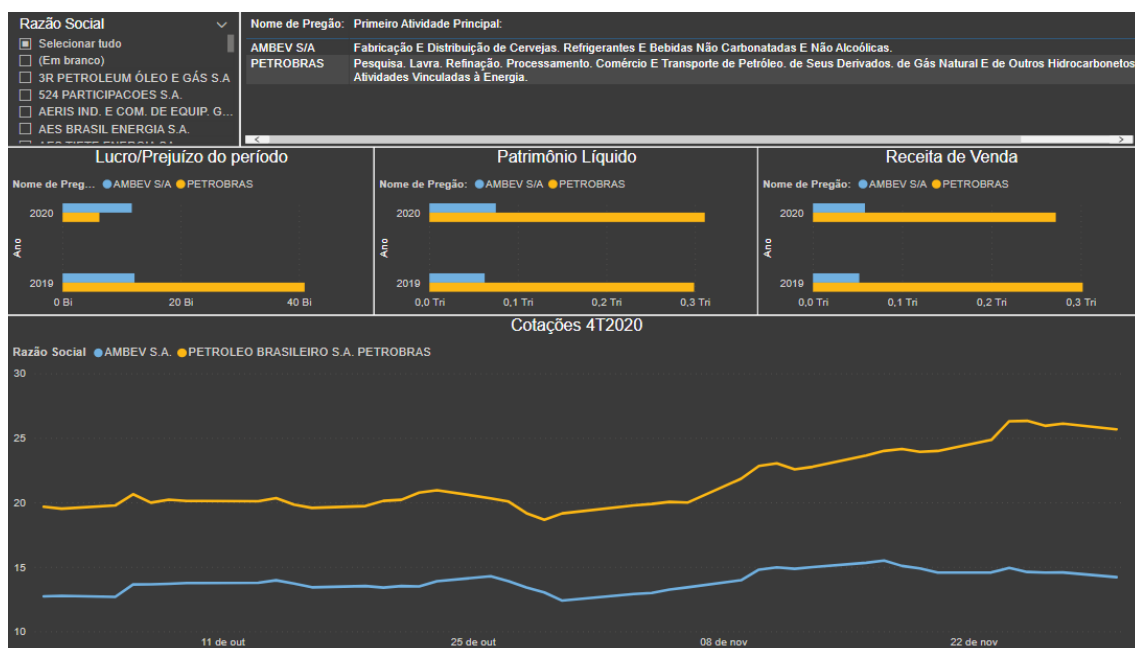


Figura 9. Tela que permite escolher as empresas e apresenta informações básicas sobre nome de pregão, atividade, lucro, prejuízo, patrimônio e receita.

5. Referências

McKinney, Wes (2017), Python for Data Analysis – Data Wrangling with Pandas, NumPy and IPython, OReilly, 2nd edition.

Mitchell, Ryan (2018), Web Scraping with Python, O'Reilly, 2nd edition.

Lawson, Richard (2015), Web Scraping with Python – Scrape data from any website with the power of Python, Packt Publishing, 1st edition.

The Selenium Browser Automation Project. Disponível em: <<https://www.selenium.dev/documentation/en/>>. Acesso em: 16 de Abril de 2021.

Muthukadan, Baiju, Selenium with Python. Disponível em: <<https://selenium-python.readthedocs.io/>>. Acesso em: 16 de Abril de 2021.

Séries Históricas. Disponível em: <http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/series-historicas>. Acesso em: 16 de Abril de 2021.

Empresas Listadas. Disponível em: <http://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/empresas-listadas.htm>. Acesso em: 16 de Abril de 2021.

Formulário de Demonstrações Financeiras Padronizadas (DFP). Disponível em: <http://dados.cvm.gov.br/dataset/cia_aberta-doc-dfp>. Acesso em: 16 de Abril de 2021.