

**GUILHERME ROCHA MONTOIA**

**AUTOMATIZAÇÃO DA ANÁLISE  
FUNDAMENTALISTA DO MERCADO DE AÇÕES  
BRASILEIRO**

**UBERLÂNDIA - MG**

**2021**



**GUILHERME ROCHA MONTOIA**

**AUTOMATIZAÇÃO DA ANÁLISE  
FUNDAMENTALISTA DO MERCADO DE AÇÕES  
BRASILEIRO**

Trabalho de Conclusão de Curso da Engenharia de Controle e Automação da Universidade Federal de Uberlândia - UFU - Campus Santa Mônica, como requisito para a obtenção do título de Graduação em Engenharia de Controle e Automação

Universidade Federal de Uberlândia – UFU  
Faculdade de Engenharia Elétrica

Orientador: Prof. Dr. Josué Silva De Morais

**UBERLÂNDIA - MG**

**2021**

Montoia, Guilherme Rocha.

Automatização Da Análise Fundamentalista Do Mercado De Ações Brasileiro/ **Guilherme Rocha Montoia. UBERLÂNDIA, 2021-2**  
41p: il. (algumas color.); 30 cm.

Orientador: Prof. Dr. Josué Silva de Morais

Trabalho de Conclusão de Curso - Universidade Federal de Uberlândia - UFU  
Faculdade de Engenharia Elétrica. **2019.**

Inclui bibliografia.

1. Business Intelligence 1. 2. Análise Fundamentalista I. Orientador Prof. Dr. Josué Silva de Morais. II. Universidade Federal de Uberlândia. III. Faculdade de Engenharia Elétrica. IV. Engenharia de Controle e Automação.



## **DEDICATÓRIA**

Dedico este trabalho aos meus pais Luis Carlos e Fátima e minha irmã Nayara, pelos seus esforços, apoio, amor incondicional, tempo, compreensão, dedicação, investimento e sacrifícios que realizaram, dos quais traçaram um caminho próspero que me possibilitou chegar à onde estou.

## **AGRADECIMENTOS**

À Universidade Federal de Uberlândia e à Faculdade de Engenharia Elétrica por todo o aprendizado e qualidade de ensino.

Ao meu professor orientador Dr. Josué Silva de Moraes pelo apoio, incentivo e direcionamento ao longo da graduação e na realização deste trabalho.

Por fim a todos que compartilharam momentos ao longo do curso de Engenharia de Controle e Automação, amigos, professores e profissionais, serei sempre grato pelas oportunidades, conversas e pelas experiências que me proporcionaram.

*“Há mais pessoas que  
desistem, do que pessoas que fracassam. ”  
(Henry Ford)*



## RESUMO

Nos últimos anos houve uma ascensão de novos investidores no mercado de ações brasileiro, porém muitos possuem dificuldades para realização de análises profundas e aplicações efetivas de estratégias de investimento por conta do tempo necessário para consolidar os dados financeiros de forma manual.

Este trabalho, então, define um método de extração de dados financeiros para análise e formatação dos indicadores fundamentalistas, através de estudos de estratégias de investimentos e técnicas de extração de dados. Democratizar, assim, o acesso a aplicações das estratégias de grandes investidores.

O método consiste em aplicação de técnicas de ETL, ferramentas de manipulação de dados e arquiteturas em nuvem, que possibilitam um acesso rápido e simples para a aplicação de diversas estratégias e análises de mercado. O método foi desenvolvido, além do embasamento em literaturas já existentes e também na vivência e experiência de diversos profissionais responsáveis pela disponibilidade de dados para áreas correlatas aos processos de análises financeiras.

**Palavras chaves:** Análise fundamentalista, automação, python, *AWS*, SQL, *business intelligence*.

## **ABSTRACT**

In recent years there has been a rise of new investors in the Brazilian stock market, but many have difficulties in performing in-depth analysis and effective application of investment strategies due to the time needed to manually consolidate financial data.

This paper, then, defines a method of extracting financial data for analyzing and formatting fundamentalist indicators, through studies of investment strategies and data extraction techniques. Thus, democratize access to applications of the strategies of large investors.

The method consists of applying ETL techniques, data manipulation tools and cloud architectures, which enable quick and simple access for the application of different strategies and market analysis. The method was developed, in addition to being based on existing literature and also on the experience and experience of several professionals responsible for the availability of data for areas related to the processes of financial analysis.

**Keywords:** Fundamental analysis, automation, python, AWS, SQL, business intelligence.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Resultado da pesquisa de plataformas BI .....	15
Figura 2 – Diagrama do Processo.....	19
Figura 3 – Obter dados do Amazon Redshift .....	21
Figura 4 – Tela de acesso ao servidor e banco de dados .....	21
Figura 5 – Modelo Star-Schema.....	22
Figura 6 – Tela Inicial da Ferramenta .....	24
Figura 7 – Tela de apresentação dos resultados .....	25

## LISTA DE ABREVIATURAS E SIGLAS

AC/PC – Ativo Circulante / Passivo Circulante

AWS – *Amazon Web Services*

BI – *Business Intelligence*

CAGR – *Compound Annual Growth Rate* (Taxa Composta Anual de Crescimento)

CVM – Comissão de Valores Mobiliários

EBIT – *Earnings Before Interest and Taxes* (Lucro Antes de Impostos e Taxas)

ETL – *Extract Transform Load* (Extração, Transformação e Carregamento)

EV – *Enterprise Value* (Valor da Firma)

HME – Hipótese do Mercado Eficiente

IMF – *Improved Magical Formula* (Formula Mágica Melhorada)

IPO – Initial Public Offering (Oferta Pública Inicial)

P/L – Preço/Lucro

P/VPA – Preço/Valor Patrimonial por Ação

ROIC – *Return On Invested Capital* (Retorno Sobre o Capital Investido)

S.A. – Sociedade Anônima

## SUMÁRIO

<b>FOLHA DE ROSTO .....</b>	<b>3</b>
<b>DEDICATÓRIA.....</b>	<b>6</b>
<b>AGRADECIMENTOS .....</b>	<b>7</b>
<b>EPÍGRAFE.....</b>	<b>8</b>
<b>RESUMO.....</b>	<b>9</b>
<b>ABSTRACT.....</b>	<b>10</b>
<b>LISTA DE ILUSTRAÇÕES.....</b>	<b>11</b>
<b>LISTA DE ABREVIATURAS E SIGLAS .....</b>	<b>12</b>
<b>1. INTRODUÇÃO.....</b>	<b>1</b>
<b>2. REFERENCIAL TEÓRICO.....</b>	<b>4</b>
<b>2.1. ANÁLISE FUNDAMENTALISTA .....</b>	<b>4</b>
<b>2.2. A ESTRATÉGIA DE BENJAMIN GRAHAM VERSUS A ESTRATÉGIA DE JOEL GREENBLATT .....</b>	<b>5</b>
<b>2.3. CRAWLER E WEB SCRAPING .....</b>	<b>9</b>
<b>2.4. ETL .....</b>	<b>10</b>
<b>2.5. PYTHON.....</b>	<b>10</b>
<b>2.6. NODE.JS.....</b>	<b>11</b>
<b>2.7. SQL .....</b>	<b>12</b>
<b>2.8. AWS .....</b>	<b>13</b>
<b>2.9. POWER BI.....</b>	<b>14</b>
<b>3. METODOLOGIA .....</b>	<b>18</b>
<b>3.1. PROCEDIMENTOS DE COLETA E TRANSFORMAÇÃO DOS DADOS.....</b>	<b>18</b>
<b>3.2. CONFIGURAÇÃO DA PLATAFORMA AWS.....</b>	<b>19</b>
<b>3.3. INTEGRAÇÃO POWER BI E AMAZON REDSHIFT .....</b>	<b>20</b>
<b>3.4. MODELAGEM E APLICAÇÃO DAS ESTRATÉGIAS DE <i>VALUE INVESTING</i>.....</b>	<b>22</b>
<b>4. ANÁLISE DOS RESULTADOS.....</b>	<b>24</b>
<b>5. CONCLUSÃO.....</b>	<b>28</b>
<b>REFERÊNCIAS .....</b>	<b>29</b>

## 1. INTRODUÇÃO

O mercado de ações está em ascensão no Brasil, a cada dia muitas pessoas dão o primeiro passo para começar a investir e buscarem a sua sonhada independência financeira, isso ocorre quando os seus rendimentos de aplicações financeiras trazem retornos maiores do que os seus gastos mensais.

Grande parte desses novos investidores, que chamamos de pequenos investidores, não tem conhecimento adequado e não está emocionalmente preparada para a atuação no mercado de ações (BONALDI, 2010). Muitos deles se aventuram a apostar no mercado considerando a um verdadeiro cassino.

Em contraponto, a análise fundamentalista possui maiores garantias de que tenha um possível ganho no mercado de ações. Para Graham (2017, p.37), “(...) uma operação de investimento é aquela que, após análise profunda, promete a segurança do principal e um retorno adequado. As operações que não atendem a essas condições são especulativas”.

Essa abordagem foi determinante para definir a questão de pesquisa: Como possibilitar uma aplicação eficaz da análise fundamentalista de todas as empresas no mercado?

Assim, o presente estudo busca um método de automatização de um sistema de investimento em bolsa de valores, de uma estratégia proposta por Benjamin Graham e também uma estratégia, similar à de Graham, proposta por Joel Greenblatt.

Graham (2003) propõe estratégias de investimento para o pequeno investidor, com o objetivo de minimizar o risco sobre aplicações financeiras, principalmente no mercado de ações, e possibilitar um retorno adequado que permita a multiplicação de patrimônio no longo prazo.

No Brasil têm-se poucos investidores na bolsa de valores, em agosto de 2019 esse número chegou aos 2,2 milhões (EXAME, 2020), isso é cerca de 1% da população brasileira investindo em ações. Comparando com países desenvolvidos, metade da população investe em ações. Isso demonstra que a maioria dos brasileiros não tem conhecimento sobre o assunto, e quando buscam, falam com os gerentes de suas contas bancárias, onde são contratados justamente para vender produtos do próprio banco, buscando com que a instituição financeira maximize seus lucros com o dinheiro das pessoas físicas em sua maioria e estas tenham um rendimento pouco maior daquele oferecido pela poupança.

Com os rendimentos de aplicações de renda fixa muito abaixo do desejado, tornando as menos atraentes, muitas pessoas buscam maiores rendimentos, assim, buscando ativos de rendas variáveis.

Segundo Wiltgen (2013) nem todo mundo que investe por conta própria na bolsa é especialista em economia, negócios ou finanças. Por isso mesmo, os relatórios de analistas de mercado acabam se tornando necessários para ajudar na hora de escolher os melhores papéis e as empresas mais sólidas. Quem busca investimentos com boa rentabilidade no longo prazo – como forma de poupança ou mesmo de formação de uma previdência complementar – provavelmente vai se identificar mais com a análise fundamentalista, escola tradicional dos estudos de mercado. Esta afirmação evidencia a importância que o tema tem e por si só já justifica um estudo mais profundo sobre ele.

Além disso, no cenário atual do mercado brasileiro de ações, há barreiras de acessos aos dados, que apesar de serem públicos, sua consolidação se mostra trabalhosa e com um tempo de trabalho extenso.

O presente trabalho tem como objetivo principal definir um método de extração de dados financeiros para análise e formatação dos indicadores fundamentalistas, através de estudos de estratégias de investimentos e técnicas de extração de dados. Além disso temos como objetivo secundário democratizar o acesso a aplicações das estratégias de grandes investidores.

A *Magic Formula* é um processo de formação de carteiras que segue a filosofia de *Value Investing*, desenvolvida por Joel Greenblatt e primeiramente publicada em 2006 em seu livro *The Little Book that Beats the Market*. Seguindo o mesmo princípio da fórmula de Graham, comprando empresas a um preço irrisório cujo seus múltiplos estivessem mais baixos que os lucros as serem obtidos com o simples fechamento da empresa e a venda dos seus ativos (GREENBLATT, 2010, p.51), ela consiste em identificar, de maneira simples, companhias listadas em bolsa que tenham alto valor e fundamentos sólidos, mas que estejam sendo negociadas a preços mais baixos no mercado. A identificação destas oportunidades é obtida a partir da criação de um ranking.

O ranking baseia-se em escolher ações com altos *ROICs* (*Return on Invested Capital* – retorno sobre capital investido) e *EV/EBIT* (GREENBLATT, 2010). Em outras palavras ranking proposto por Greenblatt é composto, portanto, por empresas de alto valor – medido por meio do retorno sobre o capital – e com preços baixos – mensurado pela relação de lucro sobre o valor de mercado das companhias.

Utilizando um processo simples de escolha de ações, a *Magic Formula* foi capaz de compor portfólios anuais que geraram um *CAGR* acima de 30% no período de 1988 a 2004 nos Estados Unidos, em comparação a 12% do S&P500 (OLIN, 2011). Alguns *back-tests* foram realizados no mercado americano (PERSSON; SELANDER, 2009) e a metodologia foi aplicada em outros mercados, tendo conseguido retornos acima do mercado.

Assim como os resultados alcançados por Graham os resultados da *Magic Formula* contrariam a teoria moderna de finanças, em especial a Hipótese do Mercado Eficiente (HME) difundida por Eugene Fama (1970). Dentre as três formas de eficiência de mercado descritas por Fama (1970) a *Magic Formula* viola a mais fraca, já que retornos superiores ao mercado ajustados pelo risco são consistentemente alcançados apenas utilizando dados históricos.

Para a aplicação destas estratégias de investimentos foi necessário a aplicação de rotinas de extração de dados em conjunto com o armazenamento de dados em estruturas SQL em um ambiente de sistemas distribuídos, possibilitando assim a criação de uma ferramenta com acesso aos dados de um banco de dados em nuvem.

Com o objetivo de consolidar as informações dos indicadores necessários, foi criada uma rotina de Extração, Transformação e Carregamento (ETL) de dados em linguagem Python, utilizando uma técnica de *web scraping* (método de extração de dados da web), para realizar mensalmente a extração de dados fundamentalistas referentes a todas as empresas brasileiras listadas na Bolsa de Valores brasileira.

Após o carregamento os dados são salvos em um banco de dados SQL em nuvem, usando a plataforma *Redshift* da AWS, para viabilizar a ingestão e manipulação em uma ferramenta de *business intelligence*, no caso, o Power BI.



## 2. REFERENCIAL TEÓRICO

Para a elaboração do trabalho foram reunidos diversos conceitos e estudos sobre o tema, a fim de pautar o desenvolvimento do projeto com uma base sólida do assunto abordado. Com isso foi possível desenhar e definir a melhor estratégia e ferramentas para alcançar os objetivos propostos inicialmente.

### 2.1. Análise Fundamentalista

Bazin (2010) aclara que o mercado de capitais permite direcionar a alocação de recursos para projetos que contribuem para o crescimento econômico e geração de valor. Na mesma linha Bartunek, Napolitano, Moreau (2016) informam que a alocação de recursos deve focar na geração de lucros, preços e valores. Nesse ambiente, Graham (2016) exprime que a análise fundamentalista permite selecionar ativos potenciais haja vista que centra no valor do ativo e na consistência dos resultados obtidos pelas companhias.

Segundo Pereira (2013), a análise de investimento se volta para constatar critérios que lhes ajudarão (investidores) a encontrar a melhor decisão para a compra e venda de ações. O sucesso de um investidor depende somente de sua competência em analisar as ações. Conforme sua experiência melhora nas análises, diminuirão as possibilidades de ser influenciado por “boatos ou dicas”.

Nesse sentido emerge a necessidade que norteia esse trabalho: A dificuldade de realizar uma análise completa de maneira manual do mercado de ações afim de identificar as melhores oportunidades de investimentos seguindo o *Value Investing* e minimizando os riscos.

Graham (2016), em seu livro “O Investidor Inteligente”, menciona a necessidade da utilização de métricas e estratégias para a seleção de ativos e propõe um sistema de investimento baseado na análise fundamentalista de ações com o objetivo principal de evitar erros substanciais de seleção e desenvolver técnicas que sejam adequadas aos seus interesses de investimento.

Ainda, nesse mesmo sentido, Graham (2016) é considerado o precursor da estratégia de investimento em valor de ativos, denominada análise fundamentalista. O cerne da estratégia busca adquirir ativos de empresas sólidas com ótimas perspectivas de geração de caixa e mantê-las no portfólio de investimentos por um longo período, de modo a maximizar lucros e valorizar a distribuição de dividendos

O estudo em questão tem como objetivo aplicar o sistema de investimento, proposto por Graham e o modelo proposto por Greenblat, na BM&F Bovespa e criar uma ferramenta que possibilite aos pequenos investidores brasileiros obter rentabilidade acima dos índices de mercado.

## **2.2. A estratégia de Benjamin Graham versus a Estratégia de Joel Greenblatt**

Considerado o precursor da estratégia *value investing* (investimento em valor) em ações, Benjamin Graham formalizou seu método clássico de análise de investimentos no livro *The Intelligent Investor*, no qual propõe um sistema de investimento para o pequeno investidor com o objetivo principal de guiá-lo contra áreas de possíveis erros substanciais e desenvolver estratégias que sejam adequadas aos seus interesses de investimento.

O foco desta estratégia compreende a comprar ações de empresas sólidas com ótimas perspectivas de geração de caixa e mantê-las no portfólio de investimentos por um longo período, objetivando, desta forma, a maximização dos lucros.

De acordo com Graham (2003), em cada compra o investidor buscará obter uma qualidade mínima nos resultados passados e na condição financeira atual da empresa, além de uma quantidade mínima em termos de lucro e patrimônio por cada dólar investido.

Graham (2016) estruturou sua teoria com base em um método puramente quantitativo, focando em encontrar empresas vendidas a múltiplos baixos e nas melhores hipóteses por menos que o seu capital de giro líquido.

Para Graham (2016), investimento é diferente de especulação, posto que aquele se refere a operação que, após a análise de suas premissas, promete a segurança do principal e um retorno adequado. As operações que não se enquadrarem nessa perspectiva são especulativas.

Na composição de uma carteira de ações deve-se esperar a flutuação dos preços, para comprar barato, não se preocupando com grandes quedas, posto que nesse momento encontra-se as grandes oportunidades. Assim, não se deve comprar uma ação porque ela subiu nem vender porque ela caiu, posto que as flutuações do mercado são naturais.

A compra de ações deve ter por base empresas negociadas no mercado a valores não superior a um terço de seus ativos tangíveis líquidos, mediante uma relação P/L satisfatória e uma posição financeira forte e ao menos, uma possibilidade de que os seus lucros serão pelo menos mantidos através dos anos. As flutuações do mercado são constantes, com

variações naturais para o investidor, de modo a ser necessário utilizar essas variáveis para descobrir ações que estejam sendo vendidas abaixo de seu valor intrínseco, ao invés de tentar comprar barato e vender caro, mediante a compra de ações quando os preços foram convenientes e respeitem a margem de segurança. Essas flutuações de mercado são a razão do lucro, seja pela valorização de seus papéis com o tempo, seja mediante a compra e venda a preços vantajosos. Assim, as flutuações de curto prazo são inevitáveis e típicas do mercado.

Ao empregar valores em operações de risco é essencial impor limites e o disciplinamento de um programa regular de investimentos, posto que o investidor deve estar atento às possibilidades de se lucrar com eles.

Graham (2016) sugere sete critérios de quantidade e qualidade Testa (2011) e Testa (2012) para a seleção de ações específicas, adaptados com base nos trabalhos para a realidade brasileira:

i. Tamanho do negócio adequado - empresa com receita anual líquida maior ou igual que R\$ 300 milhões no ano anterior à seleção de ações.

ii. Boa situação financeira - empresa com liquidez corrente (AC/PC) maior ou igual a 1 no ano anterior à seleção de ações;

iii. Estabilidade nos lucros - constância nos lucros, no período de 10 (dez) anos anteriores à seleção das ações, sem prejuízos;

iv. Histórico de dividendos - pagamento de dividendos 10 últimos anos anteriores à seleção de ações;

v. Crescimento do lucro - um crescimento mínimo de pelo menos 30% no lucro, nos últimos 10 (dez) anos anteriores à seleção de ações;

vi. P/L moderado - preço da ação no fechamento não pode ser maior do que 15 vezes o lucro por ação no ano anterior à seleção de ações;

vii. P/L vezes P/VPA moderado - o produto do P/L pelo P/VPA não deve exceder a 22,5 no ano anterior à seleção de ações.

Segundo Graham (2016), esses protocolos eliminarão a grande maioria de ações candidatas ao portfólio, mantendo as ações mais robustas e com maior lucratividade.

Ao investir em ações, com a finalidade de fazer um bom investimento, é preciso analisar a possibilidade de comprar uma ação com P/VPA próximo a 1. Assim, a empresa está valendo um valor perto de seu Patrimônio Líquido. Se a relação P/VPA for menor que 1 o mercado está precificando um risco, nesse caso é um sinal de alerta para uma análise mais profunda.

Um investidor deveria tomar por base o índice P/L satisfatório, uma posição financeira forte, e uma projeção de lucros que tenha condição de se perpetuar por muitos anos. Não é muito exigir de uma ação ser cotada modestamente no mercado, mesmo em situações de mercado em alta. O investidor deve analisar esses aspectos para não comprar múltiplos elevados, a exemplo de crescimento esperado acima da média. O mercado comumente fornece opções de papéis bons e baratos, mesmo que haja uma dificuldade normal de pesquisa para localizar boas empresas, e isto é o objetivo maior do desenvolvimento da ferramenta fruto deste trabalho.

Graham (2016) evidencia que o mercado possui duas vertentes. Na primeira, muito mais frequente que se imaginar, o mercado está errado, sendo uma oportunidade para o investidor obter boas oportunidades nesses erros. A outra, é que maioria das empresas altera suas características e qualidades com o passar dos anos, algumas vezes para melhor, outras para pior, razão pela qual o investidor deve estar atento às performances das companhias.

Conforme destaca Graham (2016), a avaliação de um investimento bem feito depende tanto dos dividendos como da valorização de seus papéis, devendo o investidor testar a efetividade da governança corporativa da companhia. A base de Graham (2016) reside no conceito de margem de segurança que está atrelada ao preço pago. Em preços elevados, não existe segurança para o investimento, ficando destoadas as questões de prudência e discernimento, podendo embutir perigos elevados inerentes as constantes mudanças do mercado. A margem de segurança quando aplicada na análise de ativos subavaliados ou barganhas torna-se relevante.

Conforme vimos, Graham sugere sete critérios de quantidade e qualidade para a seleção de ações específicas, enquanto seu aluno Joel Greenblatt segue somente 3, o ROIC, o P/L e o filtro dos setores de finanças e serviços de utilidade pública.

Ao desenvolver a *Magic Formula*, em sua publicação de 2006, Joel Greenblatt usou-se dos ensinamentos transmitidos por Benjamin Graham com a filosofia de Value Investing.

Neste estudo, Greenblatt (2006 e 2010) pretendeu buscar ações baratas com vantagem competitiva sustentável de longo prazo e identificar tais aspectos com base em indicadores fundamentalistas, com o objetivo de formar um conjunto de ações que gerariam um retorno médio anual maior do que o índice do mercado.

Utilizando os critérios da *Magic Formula*, foram formadas carteiras, segundo o autor, com ganhos superiores ao retorno obtido pelo S&P500 entre 1988 e 2004. As carteiras com

ajustes anuais de suas composições obtiveram rendimento médio anual de 30% contra 12% do S&P500. O desempenho das carteiras utilizando a *Magic Formula* foram superiores ao S&P500 mesmo após descontar o risco. (MILANE, 2016, p.11).

Analisando a estratégia da *Magic Formula* podemos perceber a simplicidade da sua aplicação e replicação, porém Greenblatt adverte sobre o viés de confirmação que temos ao analisarmos dados históricos e a partir deles fazer suposições dos resultados que poderíamos ter tido. Para Greenblatt (2010, p.59), “(...) enquanto uma fórmula computadorizada para escolher ações possa parecer ter gerado retornos teóricos espetaculares, repetir esses resultados no mundo real pode ser bastante difícil. Por exemplo, a “*Magic Formula*” pode acabar escolhendo companhias tão pequenas que apenas poucas pessoas teriam como efetivamente comprá-las. Em geral, companhias pequenas contam com poucas ações disponíveis para compra e até mesmo uma pequena demanda por essas ações pode elevar seu preço. Se for esse o caso, a fórmula pode parecer ótima no papel, mas na vida real os resultados fantásticos não podem ser repetidos. ”

Diante dessas dificuldades, é necessário fazer alguns ajustes na fórmula para adequar ao modelo real, selecionando assim como Graham empresas de grande porte, e removendo os setores de finanças e serviços de utilidade pública (isto é, fundos mútuos, bancos e seguradoras) (GREENBLATT, 2010).

A partir das publicações mais recentes na literatura sobre backtest da “*Magic Formula*”, podemos distinguir dois que adaptaram a fórmula de Greenblatt original (2006 e 2010) para obter melhores resultados.

O lucro bruto na fórmula ROIC (Ebit / Capital total investido) em vez do EBIT foi utilizado por Blackburn e Cakici (2017), que criaram uma fórmula para identificar empresas que possuem vantagem competitiva sustentável no que tange ao Lucro bruto / Capital total investido. Esta nova metodologia foi introduzida por Novy-Marx (2013) e denominada IMF (*Improved Magical Formula*). Blackburn e Cakici (2017) encontraram, nesta publicação, os melhores retornos médios anuais de IMFs de acordo com Novy-Marx (2013) do que na “*Magic Formula*” original desenvolvida por Greenblatt (2006 e 2010), onde os resultados foram significativos em todas as regiões analisadas (América do Norte, Europa, Japão e Ásia).

No Brasil, vale ressaltar que dois estudos retro estimaram a inovadora “*Magic Formula*” proposta por Greenblatt (2006 e 2010) para testar a eficácia da fórmula na obtenção de retornos anormais. Zeidler (2014) combinou uma série de carteiras de dezembro de 2002 a maio de 2014,

combinadas com a quantidade de ações e períodos de swap de ações, e comparou com o índice de ações brasileiro Ibovespa, para que o retorno pudesse ser analisado. Todas as composições de carteiras de investimento formadas pelo método de Greenblatt neste estudo (2006 e 2010) obtiveram retornos significativos, mesmo se ajustados para o risco. Como resultado, a taxa composta de crescimento anual da pior carteira é superior a 13 pontos percentuais. Além disso, a pesquisa mostra que há evidências de que sua forma fraca viola a hipótese de mercado eficiente.

O segundo estudo é o artigo de Milane (2016), originado a partir da “*Magic Formula*”, Milane compôs diversas carteiras de ações com tamanhos distintos e realizou grandes mudanças nas composições entre janeiro de 2000 e junho de 2015, sendo composta por ações do IBrX-100 e comparou com os retornos ao com o retorno do próprio índice IBrX-100 no mesmo período. Nem todas as composições superaram o retorno do índice analisado, segundo o autor:

Obviamente, carteiras com mais ações e períodos de detenção mais longos superam carteiras menores com maior giro. A carteira de investimentos de 10 ações tem prazo permanente de 1 ano e apresenta a maior taxa composta de crescimento anual (17,77%), superando a taxa composta de crescimento anual de 13,17% do IBrX-100 no mesmo período (MILANE, 2016, p.47).

Embora o risco sistêmico de todas as carteiras compostas a partir da publicação Milane (2016) seja inferior ao índice IBrX-100, a conclusão é que a “*Magic Formula*” não aumenta os retornos extraordinários em relação aos retornos do índice de mercado, porém é uma forma válida de investir seguindo a filosofia de *Value Investing*.

### **2.3. Crawler e Web Scraping**

*Web Scraping* é uma técnica de captura de dados a partir de sites. Está intimamente relacionado com a indexação da Web, que indexa mais detalhes sobre a mesma utilizando scripts ou bots (robôs de busca) e é uma técnica mundial adotado pela maioria dos motores de busca. Em contraste, concentram-se mais na transformação de dados não estruturados na internet, geralmente em formato HTML, em dados estruturados que podem ser armazenados e analisados em um banco de dados central, local ou planilha. A extração de dados utilizando este método é realizada pela simulação da navegação de um ser humano, que podem incluir a comparação on-line de preços, monitoramento de dados, detecção de mudanças em Web sites além de pesquisas e integração de dados.

## 2.4. ETL

Abreviatura do termo *Extract Transform Load*, ETL é um processo de manipulação de dados que tem a capacidade de ler diferentes formatos de arquivos e tipos de dados, e transportá-los de um ambiente para outro. O ETL serve para consolidar dados de origens diferentes, muitas vezes não estruturados e formatá-los para atender os requisitos operacionais do sistema de armazenamento e as necessidades de uso.

Ele pode ser dividido em 3 etapas, a extração, a transformação e por fim o carregamento.

Na primeira etapa são definidas as fontes de dados e de acordo com suas estruturas são convertidas para um formato estruturado utilizando de estratégias de automação do processo de extração. Embora haja muitas linguagens de programação que podem ser usadas no processo de ETL, a mais utilizada atualmente é a linguagem Python, pela sua eficiência e agilidade no desenvolvimento (Silva, 2019).

A segunda etapa, é a transformação dos dados para atender a regra de negócio e viabilizar o uso dos dados para o objetivo definido. Ela envolve a aplicação de ações como limpeza de dados, enriquecimento de dados através da junção de informações de uma ou mais fontes de dados distintas, filtros, agregações e outras manipulações necessárias de acordo com os requisitos do projeto.

Por fim, o carregamento, nessa etapa os dados são armazenados em uma estrutura previamente planejada, seja no formato tabular, SQL, quanto no formato não estruturado, NoSQL, que atenda às necessidades posteriores de conexão. Essas estruturas são chamadas de *data warehouse*, *data lake*, *data mart* ou *data store* operacional de acordo com as suas características de relacionamento, tratamento de dados e aplicação. E o uso de dados é realizado através de conexões estabelecidas entre os softwares de consumo de dados com a plataforma de armazenamento escolhida.

## 2.5. Python

Python é uma linguagem de programação, *Open-Source*, orientada a objetos e de alto nível. Sua versatilidade e simplicidade possibilitam desenvolvimento ágil de soluções para diversos campos de estudo. A popularidade da linguagem se deve ao fato da facilidade do aprendizado e sua legibilidade, que tem como consequência positiva o baixo custo na manutenção do código. (PYTHON.ORG, 2020)

A sua escolha se baseia na sua popularidade e facilidade de integração com as plataformas e estratégias usadas no projeto, além da disponibilidade de acesso a bibliotecas que possibilitam ter uma implementação ágil do projeto. (PYTHON.ORG, 2020)

Para realizar a automação das ações necessárias para o ETL, foram usadas as bibliotecas *Selenium Web Driver*, *Beautiful Soup* e *Pandas* respectivamente para navegação e manipulação das páginas da web, extração dos dados das páginas e organização dos dados em modelo tabular.

## 2.6. Node.JS

A linguagem Javascript foi criada em 1995 e se tornou a linguagem padrão para o desenvolvimento de navegadores e clientes da Web.

Desde então, houve várias tentativas de conseguir sua execução *server-side*. Todos falharam porque seu desempenho era extremamente baixo em comparação com as linguagens existentes no mercado (como PHP ou Java).

Porém, com o rápido desenvolvimento da Web nos últimos anos, a linguagem Javascript e seu mecanismo de execução passaram por diversos aprimoramentos, de forma que além de operar páginas HTML, também podem ser executados para outros fins.

Embora o Javascript tenha uma história de mais de 20 anos, seu uso *server-side* só apareceu recentemente. Com essa nova etapa de uso do Javascript, as aplicações *server-side* começaram a ser implementadas e em 2009 foi criado o primeiro ambiente de execução Javascript para esse fim: Node.js.

Portando, Node.js pode ser definido como um ambiente de execução Javascript do *server-side*. Isso significa que, com o Node.js, os scripts Javascript podem ser criados para rodar como um programa independente na máquina, em vez de depender do navegador para rodar como estamos acostumados.

Embora recente, o Node.js tem sido utilizado por grandes empresas do mercado de tecnologia, como Netflix, Uber e LinkedIn.

O principal motivo para sua adoção é sua alta escalabilidade. Além disso, sua arquitetura, flexibilidade e baixo custo o tornam a escolha ideal para soluções rápidas para ambientes



*serverless*. Inclusive os principais fornecedores de produtos e serviços em nuvem já suportaram o uso de Node.js para desenvolver soluções escaláveis.

Em comparação com outras tecnologias mais tradicionais, a criação de um ambiente Node.js e o upload de aplicativos é uma tarefa que não requer muitos recursos de computação. Se usado em combinação com ferramentas como *AWS Lambda*, que já conta com um suporte nativo ao Node.js a velocidade de implantação e replicação da máquina será muito significativa. Em um ambiente escalável, isso significa custos mais baixos e maior eficiência.

## 2.7. SQL

SQL significa "Structured Query Language" ou "Linguagem de Consulta Estruturada" em português. Resumindo, é uma linguagem de programação para lidar com bancos de dados relacionais (baseados em tabelas). Ele foi criado para permitir que vários desenvolvedores acessem e modifiquem simultaneamente os dados da empresa de forma simples e unificada.

O SQL foi originalmente desenvolvido no projeto System R do laboratório da IBM em San Jose no início dos anos 1970 para provar a viabilidade de implementação do modelo relacional proposto por EF Codd. O nome original do idioma era SEQUEL, que era uma sigla para "Structured English Query Language".

A linguagem é um bom padrão de banco de dados. Isso decorre de sua simplicidade e facilidade de uso. Ele difere de outras linguagens de consulta de banco de dados porque a consulta SQL especifica a forma do resultado em vez do caminho para obter o resultado. Ao contrário de outras linguagens procedurais, é uma linguagem declarativa. Isso encurta o ciclo de aprendizagem para iniciantes em idiomas.

Embora o SQL tenha sido originalmente criado pela IBM, vários "dialetos" desenvolvidos por outros fabricantes logo apareceram. Essa expansão levou à necessidade de criar e ajustar padrões para o idioma. Essa tarefa foi realizada pelo American National Standards Institute (ANSI) em 1986 e pela ISO em 1987. Uma de suas variações que mais se destaca é o PostgreSQL, na qual é a base da linguagem da plataforma Amazon Redshift.

Mas apesar disso o Amazon Redshift e o PostgreSQL têm muitas diferenças muito importantes. O Amazon Redshift foi projetado para aplicativos de processamento analítico online (OLAP) e inteligência de negócios (BI) que exigem consultas complexas em grandes

conjuntos de dados. Para atender a requisitos particulares, o esquema de armazenamento de dados dedicado e o mecanismo de execução de consulta usado pelo Amazon Redshift é completamente diferente da implementação do PostgreSQL. Por exemplo, os aplicativos de processamento de transações online (OLTP) geralmente armazenam dados em linhas, enquanto o Amazon Redshift armazena dados em colunas, usando codificação de compressão de dados especializada para uso de memória e E/S de disco otimizado. Já algumas funções do PostgreSQL, adequadas para processamento OLTP em escala menor, como índices secundários e operações eficientes de manipulação de dados de uma única linha, foram omitidas para melhorar o desempenho.

## 2.8. AWS

O ambiente cloud da AWS (*Amazon Web Services*) possui um sistema completo de ferramentas que permitem a criação de uma arquitetura completa de soluções em nuvem. Das quais se destacam os serviços *Amazon Simple Storage Service* (Amazon S3), para armazenamento de dados em formato de arquivos, AWS Lambda, para execução de scripts para automação de processos entre os serviços, e o Amazon Redshift, um banco de dados SQL em *cloud*.

Com um formato amigável, próximo ao ambiente de pastas do sistema operacional Windows, o Amazon S3 é uma ferramenta útil para intermediar os processos de ingestão dos dados na plataforma AWS. Suas características são a escalabilidade, disponibilidade, segurança e performance no acesso dos dados.

A execução dos scripts de carregamento dos dados entre as aplicações cloud fica por conta do AWS Lambda. Sua lógica de dimensionamento de clusters permite a integração e automação do processo de carregamento dos dados no ambiente do Amazon Redshift, gerando agilidade no processo de disponibilização dos dados para consumo dos usuários.

Tendo uma estrutura tabular, atendendo o modelo SQL, o Amazon Redshift se destaca pelo seu desempenho, disponibilidade e segurança dos dados. Graças ao sistema distribuído que o banco de dados possui, a disponibilidade de acesso se torna garantida nas mais diversas ocasiões e sua escalabilidade permite um desenvolvimento sem preocupações com a estrutura do servidor.

Para integração do ETL com o banco de dados, temos presente a biblioteca *boto3*, que permite a integração de dados através da linguagem Python. E para consumo de dados pela plataforma de *business intelligence*, há a presença de um driver JDBC, nativo da ferramenta Power BI, que permite a conexão de forma satisfatória.

## 2.9. Power BI

A necessidade de conexão e visualização dos dados após todos os tratamentos e manipulações realizadas nos leva a escolha de uma ferramenta que facilite a interação do usuário com as informações construídas.

Se baseando em soluções presente nos meios corporativos, as definições das ferramentas de *business intelligence* atendem perfeitamente a nossa necessidade de preparação de dados para análises, possibilitando a criação de relatórios, painéis e visualizações de dados. Permitindo a tomada de decisão por parte do usuário final.

O maior desafio para uma empresa é dominar os dados e informações para analisar e interpretar objetivamente os dados e informações, e obter melhor suporte e auxílio no processo de tomar decisão. As empresas mantêm muitos dados sobre seus negócios, mas têm dificuldade em traduzi-los em informações relevantes, dificultando a tomada de estratégias decisivas.

O Gartner Group (2019) criou o termo *Business Intelligence* (BI) em 1990, mas "este conceito começou antes, e suas raízes estão no sistema de relatórios *GIS-Management Information System* - (do inglês, Management Information System-MIS) na década de 1970 " (TURBAN et al., 2009, p. 27), devido às limitações técnicas da época, o sistema foi projetado para gerar estatísticas e relatórios bidimensionais.

Atualmente o mesmo grupo realiza pesquisas de mercado nas principais plataformas de BI todos os anos para mostrar o posicionamento de cada empresa do setor de BI. Os resultados apresentam a colocação de cada fabricante através de uma matriz denominada "quadrante mágico", onde o eixo "x" representa a abrangência de visão do fabricante e o eixo "y" representa a habilidade de executar a tarefa a que se propõe. Sendo assim na Figura 1 o resultado apresentado na pesquisa realizado em 2019 pela *Gartner Group* (2019). Tendo em destaque a empresa Microsoft, criadora do software Power BI.

Figura 1 - Resultado da pesquisa de plataformas BI



Fonte: GARTNER, 2019.

BI é formado por um conjunto de ferramentas e técnicas que tem por finalidade proporcionar de forma ágil as informações imprescindíveis para as tomadas de decisões (MIKROYANNIDIS; THEODOULIDIS, 2010). Um conceito mais definido Microsoft (2019). Um conceito mais definido Microsoft (2019):

Uma ferramenta de business intelligence (BI) é um software aplicativo que pode coletar e processar grandes quantidades de dados não estruturados de sistemas internos e externos [...]. Essas ferramentas ajudam a preparar dados para análise, permitindo a criação de relatórios, painéis e visualizações de dados. Os resultados permitem que funcionários e gerentes acelerem e aprimorem a tomada de decisões, aumentem a eficiência operacional, descubram o potencial de receita, identifiquem tendências de mercado, apresentem KPIs reais e descubram novas oportunidades de negócios.

Turban e Volonino (2013, p.326) sustentam que o BI agrega valor nas empresas através das extensões de informações em todos os níveis (estratégico, tático e operacional) de forma a maximizar os resultados.

Sua tecnologia visa melhorar a qualidade das informações prestadas aos gestores (AFFELDT; SILVA JUNIOR, 2013). Pode ser definida como uma ferramenta inteligente baseada em informações e monitoramento ambiental, que utiliza dados ou ferramentas técnicas de várias fontes para apoiar as decisões de gestão no empreendimento por intermédio de software (PETRINI; POZZEBON; FREITAS, 2004).

O BI visa integrar todos os tipos de dados, independente de sua fonte ou formato, para que os tomadores de decisão tenham mais informações e dados (TURBAN; VOLONINO, 2013). Bezerra et al. (2014) acrescentam que a ferramenta facilita a integração de dados de fontes múltiplas para fornecer melhores capacidades analíticas, informações inteligentes e atualizadas. Rastreando e simplificando os processos de negócios.

Os principais objetivos do BI são (TURBAN et al., 2009)

- a) Permitir acesso interativo aos dados;
- b) Fornecer operações sobre esses dados e fornecer aos gerentes e analistas de negócios a capacidade de realizar a melhor análise;
- c) Usando dados, condições e desempenho atuais por meio de análise histórica, os tomadores de decisão podem obter insights valiosos que podem ser usados como base para a tomada de decisões e uma melhor visão do cenário.

Para implementar o BI, Abukari e Jog (2003) sugerem que você siga seis passos importantes para alcançar o sucesso em sua implementação:

- a) Determine os requisitos a serem atendidos na solução de BI. A demanda deve estar atrelada aos objetivos e estratégia da empresa;
- b) Identificar fontes de dados já existentes na organização. Já existe uma grande quantidade de informações nos bancos de dados, planilhas e arquivos da organização. Claro, mais informações precisam ser criadas, mas as informações existentes devem ser verificadas primeiro;

c) Extraia, transforme e carregue dados para criar uma base para convergência mútua. Este processo deve garantir que todas as informações relevantes sejam consideradas consistentes;

d) A organização deve selecionar ferramentas para apresentar, visualizar e analisar as informações dos resultados;

e) Cria relatórios padrão, oferece suporte à análise sob demanda e mineração de dados e obtenha uma compreensão profunda dos principais indicadores de desempenho;

f) Planeje uma implantação em toda a empresa para garantir que os tomadores de decisão obtenham informações suficientes quando e onde precisarem.

### 3. METODOLOGIA

O desenvolvimento do projeto foi realizado priorizando, como ponto central, as etapas de extração de dados financeiros para análise e formatação dos indicadores fundamentalistas, sendo esta etapa a maior contribuição do trabalho.

Para alcançar de forma satisfatória o resultado proposto, foram ponderadas técnicas e ferramentas que melhor atendessem às especificações iniciais e limitações de custos. Entretanto, tais limitações não tiveram impactos negativos no resultado final, validando, assim, a escolha da abordagem, técnica e ferramentas, descritas a seguir.

#### 3.1. Procedimentos de coleta e transformação dos dados

Usando unicamente a linguagem Python, nessa etapa foi realizada a automação da rotina de ETL. A primeira etapa consiste na extração dos dados que através do link *site map* da página web “<https://statusinvest.com.br/>”, foi desenvolvida uma rotina de *Crawler* para extração dos links que serviram como *endpoints* para a rotina de *Web Scrapping*.

A extração dos dados das páginas catalogadas pelo *Crawler* foi realizada através das funções de procura da biblioteca *Beautiful Soup*, na qual percorreu em *loop* a lista catalogada extraíndo, dos campos mapeados, as informações necessárias para o desenvolvimento do projeto. No caso dos dados históricos fora necessário o uso da biblioteca *Selenium Web Driver* para a manipulação de objetos *JavaScript* presentes nas páginas consultadas.

Após a extração completa dos dados das páginas foi realizada a segunda etapa do processo de ETL, a transformação dos dados. Primeiro foi transformado o formato lista de dicionários em um formato tabular através da biblioteca *Pandas*, facilitando a limpeza das informações e a transformação dos tipos de dados para atender as necessidades de manipulação e cálculo dos dados do projeto, economizando recursos da plataforma AWS e tempo de processamento das informações no Power BI.

Por fim, o carregamento dos dados foi realizado através da conversão do *dataframe* para o formato CSV e feito o carregamento para o serviço Amazon S3, usando a biblioteca *boto3*.

O diagrama completo da contribuição do trabalho pode ser observado na figura 2, gerando uma visão macro do fluxo desenvolvido ao longo do estudo.

Figura 2 – Diagrama do Processo



Fonte: Autor.

### 3.2. Configuração da plataforma AWS

Para a utilização do serviço Amazon S3, foi necessária a criação de bucket e suas subpastas para armazenar os arquivos coletados. O processo se baseia na execução de rotinas estruturadas que seguem uma ordem definida.

Para a criação do *bucket* no Amazon S3 podemos acessar o AWS Management Console e através da página da ferramenta executar os procedimentos através do botão “Criar *bucket*”. O *bucket* criado serve como um contêiner de objetos, no qual foi utilizado para armazenar as pastas e arquivos utilizados no processo.

Vale ressaltar que o nome do *bucket* deve ser único globalmente, isto é, não é permitido que utilize um nome já utilizado por qualquer usuário da ferramenta do Amazon S3. A região do contêiner é necessária para definir o local de armazenamento desses dados e é recomendável a escolha de uma região próxima para otimizar a latência ou uma região com um custo menor afim de minimizar custos ou atender aos requisitos regulamentares.

Após essa etapa foi necessária a criação do cluster e das tabelas no Amazon Redshift, também, via AWS Management Console.

Semelhante ao processo do Amazon S3, na página inicial da ferramenta Amazon Redshift é possível selecionar a opção “*Create cluster*” e definir o nome do *cluster* e seu plano de assinatura, o cluster serve como um grupo de nós, com coleções de recursos, podendo conter um ou mais bancos de dados.

Para o seu acesso é necessário criar um usuário administrador e a senha. E a criação de tabelas e views pode ser realizada através da ferramenta *Query editor*, nativa do Amazon Redshift, com o uso da linguagem PostgreSQL.



Em seguida, configuramos o AWS Lambda, via AWS Management Console, para realizar de forma automática a integração dos dados do Amazon S3 com o Amazon Redshift, com gatilho configurado para o método POST no Amazon S3.

Na página inicial da ferramenta AWS Lambda há a opção de “criar função”, na qual é necessário definir o nome da função, o idioma e a arquitetura, no nosso caso o idioma Node.js 14.x, por ser a versão mais recente da linguagem e pela facilidade de implementação da função.

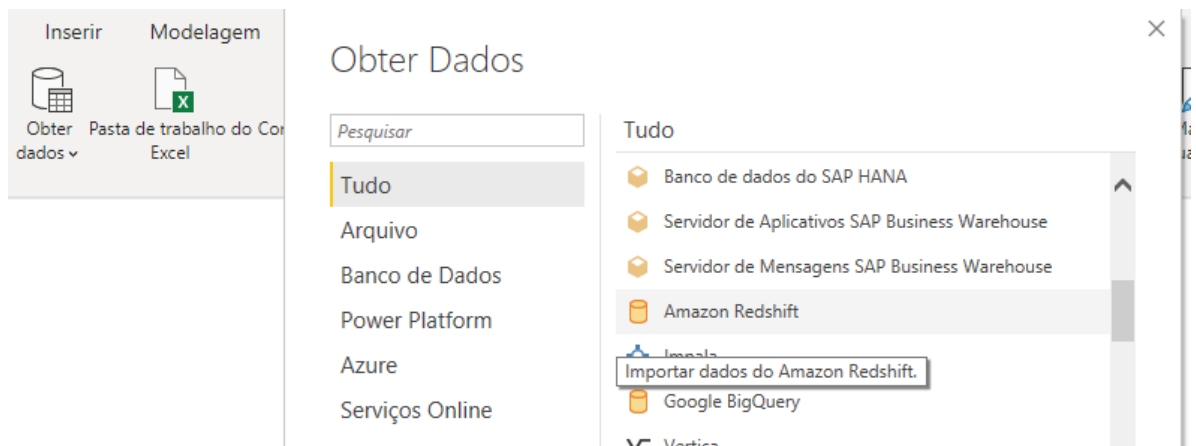
Por fim configuramos a função criada, realizando o carregamento do arquivo zip do script a ser executado e para evitar que a execução do script seja cancelada por limite de tempo, foi configurado o tempo de execução limite de 10 minutos.

Além disso foi configurada a execução do script de forma automática com a criação do arquivo no Amazon s3, criando o gatilho para o método POST e o prefixo da pasta na qual o arquivo será armazenado.

### **3.3. Integração Power BI e Amazon Redshift**

Com os dados presentes no Amazon Redshift, é iniciado o processo de criação do relatório usando o Power BI Desktop. Por permitir uma conexão nativa com o Amazon Redshift, o Power BI nos permite importar as tabelas presentes no banco de dados em nuvem através do botão "Obter dados", conforme figura 3.

Figura 3 – Obter dados do Amazon Redshift



Fonte: Autor.

A seguir é necessário indicar qual é o endereço do servidor do *cluster* que configuramos anteriormente e informar o banco de dados que iremos acessar, conforme os campos “Servidor” e “Banco de Dados” da figura 4.

Figura 4 – Tela de acesso ao servidor e banco de dados

Fonte: Autor.

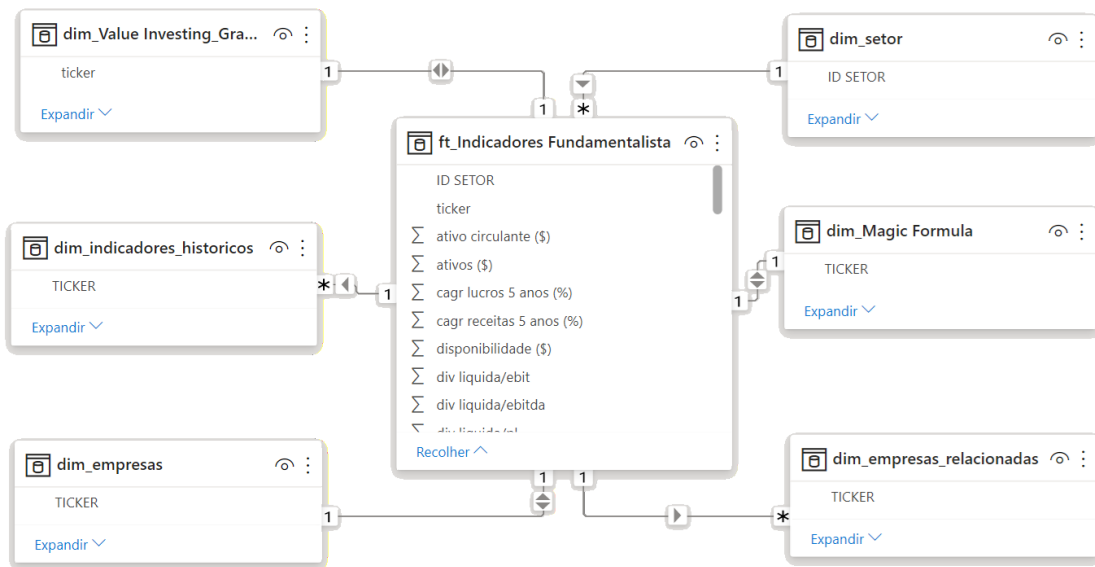
Após a seleção do servidor e banco de dados, informamos as credenciais de acesso ao banco de dados, inserindo o nome de usuário e a senha e escolhemos as tabelas a serem importadas para o Power BI, para início da aplicação das estratégias e modelagem dos dados.

### 3.4. Modelagem e aplicação das estratégias de *Value Investing*

Após a ingestão das tabelas do Redshift no Power BI realizamos a definição de tipos de dados, filtro de valores e aplicação das regras, das estratégias de investimento abordadas, através da ferramenta *Power Query Formula Language* (conhecida como "Linguagem M"). Com o tratamento de dados através do *Power Query Formula Language* reduzimos o volume de dados carregados para aplicação do Power BI, tornando o arquivo leve e garantindo a performance no carregamento dos dados.

Após o carregamento, modelamos os dados seguindo o modelo *Star-Schema*, conforme figura 5, um método de modelagem amplamente usado em bancos relacionais. Contudo para a construção do modelo é necessária a normalização dos dados, otimizando o volume de dados para garantir a performance na visualização dos dados.

Figura 5 – Modelo Star-Schema



Fonte: Autor.

Após a normalização dos dados, segregamos as tabelas geradas em dois tipos, as tabelas dimensões e as tabelas fatos. As tabelas dimensões descrevem a entidade comercial - o item que você modela. As tabelas dimensões do projeto incluem o tipo da ação, o setor, nome da empresa e outras características imutáveis no curto prazo da ação e também datas de calendário. A tabela de dimensão contém uma coluna-chave, para realizar o relacionamento entre a tabela fato.

A tabela fato armazena os valores mutáveis no curto prazo referente a ação, como o *ROIC*, *EV*, *EBIT*, *P/L*, cotação da ação e outros valores que extraímos durante o processo de *Web Scrapping*. A tabela de fatos contém colunas-chave de dimensão relacionadas a tabelas de dimensão e colunas de medidas numéricas.

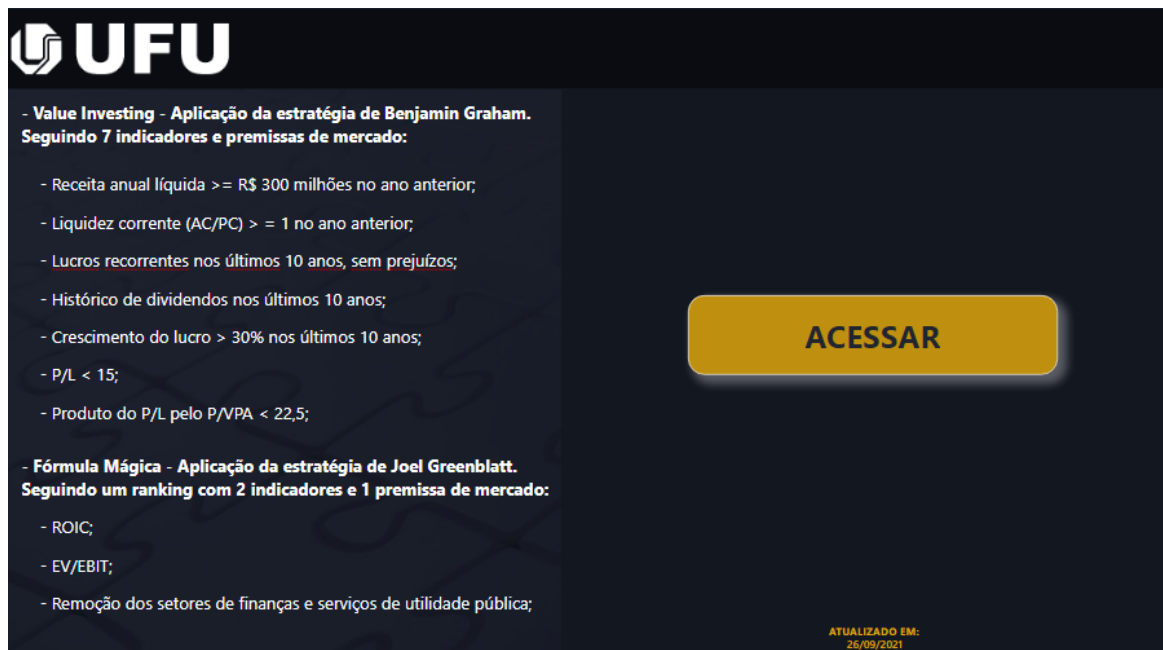
A coluna-chave da dimensão determina a dimensão da tabela de fatos e o valor da chave da dimensão determina a granularidade da tabela de fatos. Por exemplo, considere a tabela de fatos do projeto que armazena os indicadores de análise fundamentalista com duas colunas-chave de dimensão “ID SETOR” e “TICKER”. A tabela tem duas dimensões fáceis de entender. No entanto, a granularidade não pode ser determinada sem considerar o valor da chave da dimensão.

Após a modelagem são aplicadas as regras de negócios faltantes, através de formulas DAX, funções disponíveis dentro da plataforma do Power BI, na qual nos permite criar expressões para calcular o os valores necessários para a nossa aplicação. E por fim iniciamos a criação das visualizações dos dados.

#### 4. ANÁLISE DOS RESULTADOS

Para a visualização de dados foram criadas duas telas, sendo a primeira uma tela inicial na qual explica os métodos de seleção de ações presente no projeto e um botão para acesso a tela principal, conforme figura 6.

Figura 6 – Tela Inicial da Ferramenta

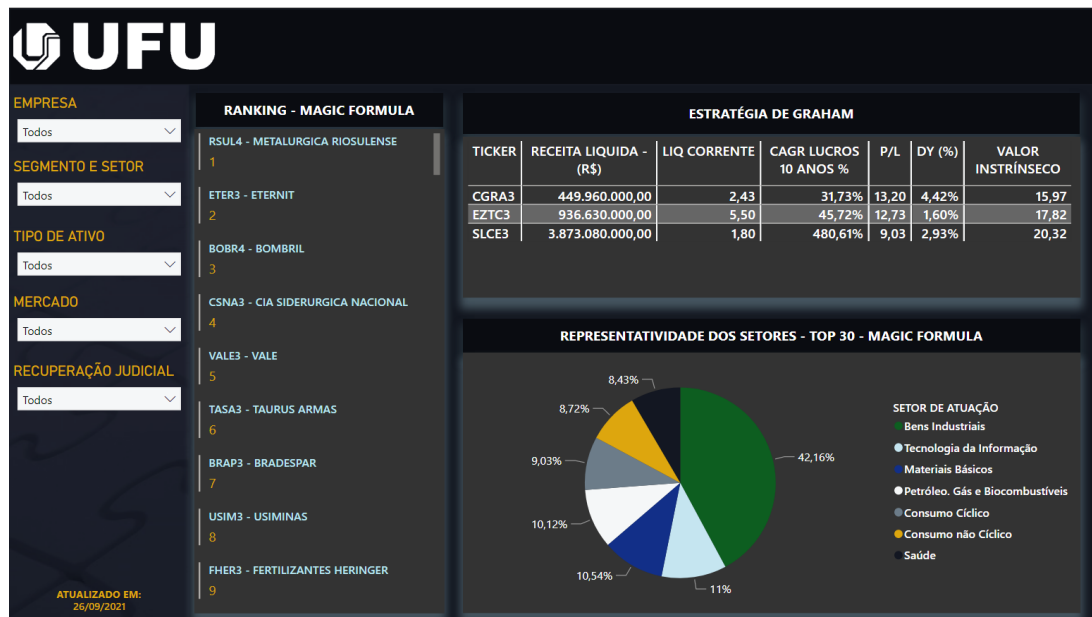


Fonte: Autor.

Na tela principal, conforme figura 7, é possível observar a exposição de dados já coletados pelo programa e compilados como informações às quais o usuário tenha acesso sem necessidade de qualquer outra ação que não a de entrar na página. A partir do acesso pode-se realizar uma série de filtros nos quais conseguirá acesso às informações desejadas.

Através dos filtros o usuário pode refinar sua escolha de investimento com base em suas preferências pessoais, como o tipo do ativo, recuperação judicial, setor e segmento, empresa e mercado. Permitindo o acesso de forma que o usuário fique confortável com sua seleção.

Figura 7 – Tela de apresentação dos resultados



Fonte: Autor.

Os resultados obtidos na aplicação do filtro de Graham são consistentes com o estudo apresentado por Artuso e Chaves Neto (2010), no qual a aplicação do filtro de Graham forma uma carteira com uma quantidade baixa de ações, impactando na diversificação do portfólio. No presente trabalho tivemos apenas 3 ações, sendo elas “CGRA3”, “EZTC3” e “SLCE3”. Neste ponto, os resultados foram satisfatórios na aplicação da estratégia possibilitando uma escolha de investimentos fiéis as estratégias Graham.

Tal resultado comprova as escolhas de métodos, ferramentas e técnicas aplicadas no desenvolvimento.

Contudo, ao discorrer sobre o assunto temos que pontuar os ganhos e perdas que tivemos com as escolhas apresentadas anteriormente. Seguindo o diagrama da figura 2, iniciamos a análise de resultado a partir da fonte de dados, que apesar de ser uma fonte confiável temos limitações que fogem do nosso controle, sendo a principal delas a mutabilidade das páginas HTML, que podem sofrer mudanças nas *tags* e exigir uma manutenção no script desenvolvido. Porém, ao decorrer do desenvolvimento e dos testes não tivemos nenhuma mudança em relação aos campos extraídos, tornando possível a execução fluída e alcançar os objetivos propostos. Vale pontuar que existem outras fontes de dados que podem cumprir tal objetivo, porém há

custos de acesso que fogem do escopo inicial, sendo a principal justificativa para a escolha da fonte de dados utilizada.

A seguir, temos o processo de ETL, na qual utilizamos, dentro da linguagem Python, duas bibliotecas distintas para executar a extração dos dados. Começando pela biblioteca *Beautiful Soap*, que nos permitiu a extração dos dados através dos identificadores presentes na página HTML, porém ao extrair dados que necessitavam de ações de interação *Javascript*, foi necessária a utilização de uma biblioteca complementar para executar tais ações. Com isso utilizamos a biblioteca *Selenium Web Driver*, para simular a navegação na internet e possibilitar o acesso aos dados que antes estavam ocultos pelos scripts do site. Essa escolha trouxe uma redução de velocidade de execução do código, nos limitando ao tempo de carregamento das páginas e da abertura do navegador.

Em terceiro lugar, está a escolha da plataforma de armazenamento dos dados, que foi definida levando em consideração os custos envolvidos. Ao analisar os planos de plataformas *cloud*, a AWS se mostrou uma escolha com o menor custo dentre suas principais concorrentes, graças ao seu período de teste que foi suficiente para executar as demandas previstas para o trabalho sem nenhum custo envolvido. Sua interface gráfica amigável também foi um aspecto positivo no nosso desenvolvimento, permitindo uma implementação rápida do ecossistema utilizado.

Aprofundando um pouco mais na escolha do AWS Lambda em relação a outros serviços da plataforma, podemos citar que não tivemos um custo no uso do serviço, devido ao seu limite de um milhão de solicitações gratuitas por mês e 400.000 GB por segundos de tempo de computação por mês, enquanto em outro serviço como por exemplo o AWS Glue, teríamos um custo estimado de 0,44 USD por execução do processo de transporte dos dados do Amazon S3 para o Amazon Redshift, que apesar de ser um valor baixo, esbarra na limitação proposta de custo zero no desenvolvimento. Sendo assim o AWS Lambda a escolha mais adequada e que melhor atendeu o projeto.

Por fim devemos citar a plataforma de visualização de dados que foi de grande importância para a validação dos nossos resultados. Que apesar de ser uma plataforma líder de mercado, seu uso *off-line* é de forma gratuita e nos oferece um grande leque de funções para criar o nosso modelo de dados.

Mas é necessário pontuar a escolha entre as linguagens presentes no Power BI. A mais utilizada no trabalho foi a linguagem *Power Query Formula Language*, responsável pela

aplicação das estratégias de investimento. Esta escolha foi ponderada pelo volume da base de dados, evitando assim o carregamento de dados que não fossem relevantes para as aplicações propostas e por consequência a redução do tamanho do arquivo gerado.

Ao passo que, na linguagem DAX seria necessário o carregamento de todos os campos que iriam ser utilizados para posteriormente serem aplicadas as transformações, deixando o arquivo pesado para o uso. Mas tendo o benefício de uma atualização mais veloz das bases de dados.



## 5. CONCLUSÃO

Todos os prós e contras apresentados, no capítulo anterior, foram relevantes para o desenvolvimento e atingimento do objetivo central do trabalho. E graças a eles, foi possível realizar a automação do processo e obter uma performance significativa na análise do mercado brasileiro.

Por fim concluímos que o uso da ferramenta se mostra de grande valor, pela redução dos erros de processos manuais, aumentando a segurança na aplicação e na tomada de decisão no dia a dia, atingindo o objetivo central do projeto, com o aumento da democratização do acesso ao mercado brasileiro de ações e a tendo como contribuição principal um método de extração de dados financeiros para análise e formatação dos indicadores fundamentalistas.

Diante do vasto universo de tipos e estratégias de investimentos, diferentes estratégias poderiam ser aplicadas a partir dos dados coletados. Porém, a escolha das estratégias apresentadas foi um fator relevante na comprovação dos resultados obtidos através da automação do processo, que nos permitiu validar e ter confiança no resultado final do projeto.

Vale destacar também que a automação da análise financeira das empresas nos permite ter mais segurança na decisão da composição das carteiras, pela remoção do erro humano e velocidade da análise completa do mercado.

Uma carteira automatizada pode ajudar pessoas que não dispõem de tanto conhecimento do setor financeiro a avaliar os indicadores das empresas e tomar decisões pautadas por estratégias validadas e respeitadas no mercado financeiro.

A principal contribuição deste trabalho, certamente, foi o desenvolvimento de um método de extração de dados financeiros capaz de consolidar todos os dados financeiros de empresas listadas no mercado de ação brasileiro e permitir análises profundas e bem fundamentas sobre os diversos setores que compõem o mercado.

Por fim, todos os objetivos propostos inicialmente no trabalho foram cumpridos e com a contribuição principal, são abertas novas possibilidades de trabalhos futuros sobre o tema, sendo possível aprofundar em análises mais complexas e estratégias mais robustas a partir dos resultados aqui expostos.

Ademais, tal estudo permitiu o autor a realizar de forma ativa o desenvolvimento e transformação da tecnologia em favor das necessidades da sociedade, tal qual prega o Projeto Pedagógico do curso de Engenharia de Controle e Automação desta universidade.

## REFERÊNCIAS

- ABUKARI, K.; JOG, V. Business Intelligence in action. CMA Management, 2003.
- AFFELDT, F. S., SILVA JUNIOR, S. D. Information architecture analysis using business intelligence tools based on the information needs of executives. JISTEM - Journal of Information Systems and Technology Management, São Paulo, V.10, n. 2, 2013.
- BEZERRA, A. A. et al. Business Intelligence: uma perspectiva de soluções aplicadas no contexto da Gestão da Informação. In: ENCONTRO DE ESTUDOS SOBRE TECNOLOGIA, CIÊNCIA E GESTÃO DA INFORMAÇÃO, 5., 2014, Recife. Anais [...], Recife: InFoco Consultoria Júnior/UFPE, 2014.
- Blackburn, Douglas W. and Cakici, Nusret, Overreaction and the Cross-Section of Returns: International Evidence (January 14, 2017). Journal of Empirical Finance, Vol. 42, 2017, Gabelli School of Business, Fordham University Research Paper No. 2800188, Disponível em: <<https://ssrn.com/abstract=2800188>>. Acesso em: 27 set. 2021
- GARTNER GROUP. Key Issues for Analytics, Business Intelligence and Performance Management, 2019. Disponível em:<<https://www.bmc.com/blogs/gartner-magic-quadrant-analytics-business-intelligence/>>. Acesso em: 27 set. 2021
- GRAHAM, Benjamin. O investidor inteligente: O guia clássico para ganhar dinheiro na bolsa. Rio de Janeiro: Casa dos Livros Editora Ltda, 2017.
- GREENBLATT, J. The little book that beats the market. [S.l.]: John Wiley & Sons, 2006.
- GREENBLATT, J. The little book that still beats the market. John Wiley & Sons, 2010.
- MILANE, L. P. Teste de eficiência da magic formula de Value Investing para o mercado brasileiro de ações. Dissertação (Mestrado em Economia) – Fundação Getúlio Vargas - FGV, São Paulo, 2016
- MIKROYANNIDIS, A.; THEODOULIDIS, B. Onto logy management and evolution for business intelligence. International Journal of Information Management. v. 30, 2010.
- NOVY-MARX, R. The other side of value: The gross profitability premium. Journal of Financial Economics, v. 108, n. 1, p. 1-28, 2013.

PEREIRA, Cleverson Luiz. Mercado de Capitais. Curitiba: Editora Intersaberes, 2013.

PETRINI, M.; POZZEBON, M.; FREITAS, M. T. Qual é o papel da inteligência de negócios (BI) nos países em desenvolvimento? Um panorama das empresas brasileiras. In: ENCONTRO DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 28., 2014, Curitiba. Anais [...] Curitiba: ENANPAD, 2004.

PYTHON.ORG. What is Python? Executive Summary. 2020. Disponível em: <<https://www.python.org/doc/essays/blurb/>>. Acesso em: 17 ago. 2021.

SILVA, João E. L., comparando as linguagens de programação FORTRAN e PYTHON através de problemas matemáticos, Cuité - 2019

TESTA, C. H. R.; LIMA, G. A. S. F. O canto da sereia: aplicação da teoria de Graham na Bm&fBovespa. AOS Brazil, v. 1, n. 1, p. 79-93, 2012.

TESTA, Carlos Henrique Rodrigues. Aplicação da estratégia de investimento de Graham à Bm&fBovespa para o pequeno investidor. In: SEMINÁRIOS EM ADMINISTRAÇÃO, 14., 2011. Anais... São Paulo, 2011.

TURBAN, E., VOLONINO, L. Tecnologia da informação para gestão: em busca do melhor desempenho estratégico e operacional. 8. ed. Porto Alegre: Bookman, 2013.

TURBAN, E.; et al. Business Intelligence: um enfoque gerencial para a inteligência do negócio. Porto Alegre, Bookman, 2009.

WILTGEN, Julia. Como funciona a análise fundamentalista de ações. Seu Dinheiro, São Paulo, dez. 2013. Disponível em: <<https://exame.abril.com.br/seu-dinheiro/como-funciona-analise-fundamentalista-acoes-576374/>>. Acesso em: 04 out. 2021.

ZEIDLER, R. G. D. Eficiência da Magic Formula de Value Investing no mercado brasileiro. Dissertação (Mestrado) - Escola de Economia de São Paulo, Fundação Getulio Vargas, São Paulo, p. 0-70, 2014.