

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DE DOIS VIZINHOS
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

ABDINARDO MOREIRA BARRETO DE OLIVEIRA

**DESENVOLVIMENTO DE UM SISTEMA WEB PARA AVALIAÇÃO
DO RISCO DE PREÇO DE PRODUTOS FRESCOS: O CASO DAS
CEASAS BRASILEIRAS**

TRABALHO DE CONCLUSÃO DE CURSO

DOIS VIZINHOS
2021

ABDINARDO MOREIRA BARRETO DE OLIVEIRA

DESENVOLVIMENTO DE UM SISTEMA WEB PARA AVALIAÇÃO DO RISCO DE PREÇO DE PRODUTOS FRESCOS: O CASO DAS CEASAS BRASILEIRAS

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Dr. Ives Renê Venturini Pola

DOIS VIZINHOS
2021



4.0 Internacional

Esta licença permite remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

ABDINARDO MOREIRA BARRETO DE OLIVEIRA

**DESENVOLVIMENTO DE UM SISTEMA WEB PARA AVALIAÇÃO
DO RISCO DE PREÇO DE PRODUTOS FRESCOS: O CASO DAS
CEASAS BRASILEIRAS**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de Aprovação: 13/novembro/2021

Ives Renê Venturini Pola
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Pato Branco

Anderson Chaves Carniel
Doutorado
Universidade Federal de São Carlos

Marcelo Teixeira
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Pato Branco

DOIS VIZINHOS
2021

Dedico este trabalho a todos aqueles que, audaciosamente, foram aonde ninguém achava que eles poderiam chegar.

AGRADECIMENTOS

Início os meus agradecimentos à minha família, especialmente à minha mãe (Juerila), meu pai (Abdinardo, *in memorian*) e irmãos (Theófilo e Isabella), por me ajudarem a ser quem sou.

À UTFPR-DV, por ter disponibilizado uma especialização em Ciência de Dados, uma das áreas do saber de destaque para os profissionais do século XXI.

Ao Prof. Dr. Ives Renê Venturini Pola, por ter acreditado e me ajudado no desenvolvimento deste Trabalho de Conclusão de Curso de Especialização.

A todos os professores que lecionaram na Especialização em Ciência de Dados (UTFPR-DV), trazendo seus importantes conhecimentos sobre Ciência de Dados para ser compartilhado entre nós.

A Rozane Marcia Triches, pela incrível mulher que és, e pelo privilégio de ser teu companheiro.

A todos os meus companheiros desta especialização, os quais me ajudaram, cada um ao seu modo, a me fazer uma pessoa melhor.

E a todos aqueles quem não pude citar, saibam que estão no meu coração e lhes presto esta homenagem.

Chi cerca, trova. (DA VINCI, Leonardo, 1505).

RESUMO

OLIVEIRA, Abdinardo Moreira Barreto de. Desenvolvimento de um sistema *web* para avaliação do risco de preço de produtos frescos: o caso das Ceasas brasileiras. 2021. 48 f. Trabalho de Conclusão de Curso – Curso de Especialização em Ciência de Dados, Universidade Tecnológica Federal do Paraná. Dois Vizinhos, 2021.

Na última década, o setor agrário vem se apropriando de tecnologias de *Business Intelligence* (BI) e *Data Warehouse* (DW) para a melhoria de seu processo decisório, especialmente quanto à gestão de risco. Contudo, nem todos os produtos agrícolas são beneficiados plenamente por estas tecnologias, como é o caso dos produtos frescos comercializados em Ceasas brasileiras. Estes, além de não terem acesso aos contratos de *hedging* existentes nos mercados de *commodities*, também não possuem um sistema computacional agregado que auxilie na gestão do risco de preço subjacente. Isto provoca incerteza nos agentes econômicos, prejudicando a qualidade de suas decisões por causa da assimetria de informação. Portanto, este estudo propõe um sistema *web* com abordagem baseada em BI e DW para gerenciar o risco de preço de produtos frescos em Ceasas brasileiras, com as seguintes características: 1) armazenagem e recuperação rápida dos preços, 2) cálculo de métricas avançadas de risco, e 3) apresentação de resultados em formatos intuitivos de tomada de decisão. Esta foi uma pesquisa aplicada (finalidade), do tipo descritiva (objetivo) e quantitativa (abordagem), utilizando-se da pesquisa documental (procedimentos) e da pesquisa bibliográfica (fonte de dados). Os preços mensais (R\$), de todos os produtos frescos comercializados nas Ceasas brasileiras, entre 2006 e 2021, foram coletados no site do Prohort, resultando numa amostra de 499.527 tuplas. A modelagem dimensional ocorreu via DW's ROLAP e NoSQL, com suas implementações feitas no Postgresql 12 e Apache Cassandra 3.11, respectivamente. O desvio-padrão e o CVaR foram definidos como as principais métricas de avaliação tanto para os preços individuais como para as cestas otimizadas de produtos, tendo como base as teorias moderna e pós-moderna do Portfólio. A interface gráfica do usuário foi desenvolvida através do pacote *Shiny*, em conjunto com as demais funcionalidades oferecidas pela linguagem R, considerando os preceitos teóricos e práticos sobre a visualização de dados para avaliação da informação. Portanto, é esperado que os resultados deste estudo contribuam para a diminuição dos efeitos adversos causados pela assimetria de informação de dados agrícolas agregados, reduzindo a incerteza dos agentes econômicos envolvidos e melhorando a eficiência desse mercado.

Palavras-chave: Desenvolvimento de software. Sistema de gerenciamento de dados. Risco de preços agrícolas.

ABSTRACT

OLIVEIRA, Abdinardo Moreira Barreto de. Development of a web system to assess the price risk of fresh products: the case of Brazilian Ceasas. 2021. 48 f. Trabalho de Conclusão de Curso – Curso de Especialização em Ciência de Dados, Universidade Tecnológica Federal do Paraná. Dois Vizinhos, 2021.

In the last decade, the agrarian sector has been appropriating Business Intelligence (BI) and Data Warehouse (DW) technologies to improve its decision-making process, especially regarding risk management. However, not all agricultural products are fully benefited by these technologies, as is the case of fresh products sold in Brazilian Ceasas. These, in addition to not having access to hedging contracts existing in the commodity markets, also do not have an aggregate computer system that helps in managing the underlying price risk. This causes uncertainty in economic agents, harming the quality of their decisions due to information asymmetry. Therefore, this study proposes a web system with a BI and DW-based approach to manage fresh product price risk in Brazilian Ceasas, with the following characteristics: 1) storage and rapid price retrieval, 2) calculation of advanced risk metrics, and 3) presentation of results in intuitive decision-making formats. This was an applied research (purpose), descriptive (objective) and quantitative (approach), using documentary research (procedures) and bibliographic research (data source). Monthly prices (R\$) of all fresh products sold in Brazilian Ceasas, between 2006 and 2021, were collected on the Prohort website, resulting in a sample of 499,527 tuples. The dimensional modeling took place via DW's ROLAP and NoSQL, with their implementations made in Postgresql 12 and Apache Cassandra 3.11, respectively. Standard deviation and CVaR were defined as the main evaluation metrics for both individual prices and optimized product baskets, based on the modern and postmodern theories of the Portfolio. The graphical user interface was developed through the Shiny package, together with the other functionalities offered by the R language, considering the theoretical and practical precepts on data visualization for information evaluation. Therefore, it is expected that the results of this study contribute to reducing the adverse effects caused by the asymmetry of information in aggregated agricultural data, reducing the uncertainty of the economic agents involved and improving the efficiency of this market.

Keywords: Software development. Data management system. Agricultural price risk.

LISTA DE FIGURAS

Figura 1 – Informações de mercado produzidas pelas Ceasas	17
Figura 2 – Exemplo de um hipercubo de dados OLAP	19
Figura 3 – Exemplo de um DW orientado a grafos	20
Figura 4 – Ilustração do CVaR e VaR	22
Figura 5 – Diagrama de visualização de dados	23
Figura 6 – Tendência histórica de popularidade dos SGBD's	27
Figura 7 – Diagrama relacional do DW ROLAP	29
Figura 8 – Tabelas denormalizadas do DW NoSQL	30
Figura 9 – Tela inicial do sistema <i>web</i> de BI para as Ceasas brasileiras	32
Figura 10 – Saída da consulta para a aba <i>Gráfico</i>	32
Figura 11 – Saída da consulta para a aba <i>Mapa de preços</i>	33
Figura 12 – Saída da consulta para a aba <i>Risco</i>	34
Figura 13 – Saída da consulta para a aba <i>Mapa de riscos</i>	35
Figura 14 – Saída da consulta para a aba <i>Portfólios</i>	35
Figura 15 – Saída da consulta para a aba <i>Gráfico de preços</i>	36
Figura 16 – Saída da consulta para a aba <i>Correlações</i>	37
Figura 17 – Saída da consulta para a aba <i>Portfólios (1)</i>	37
Figura 18 – Saída da consulta para a aba <i>Portfólios (1), continuação</i>	38
Figura 19 – Saída da consulta para a aba <i>Portfólios (2)</i>	38
Figura 20 – Saída da consulta para a aba <i>Portfólios (2), continuação</i>	39
Figura 21 – Site do Prohort	40
Figura 22 – Gráfico de linha, resultante de consulta gerada pelo site do Prohort	41

LISTA DE TABELAS

Tabela 1 – Análise comparativa dos trabalhos relacionados	14
Tabela 2 – Quantidade de dados perdidos nas séries de preços de produtos frescos, por Estado	26

LISTA DE ABREVIATURAS E SIGLAS

BI	Business Intelligence
CQL	Cassandra Query Language
CSS	Cascading Style Sheets
CVaR	Valor-em-Risco Condicional
DW	Data Warehouse
ETL	Extract-Transform-Load
GUI	Graphical User Interface
HDD	Hard Disk Drive
HTML	Hyper Text Markup Language
NoSQL	Not only Structured Query Language
OLAP	Online Analytical Processing
Prohort	Programa Brasileiro de Modernização do Mercado Hortigranjeiro
RDS	Relational Database Service
ROLAP	Relational Online Analytical Processing
SGBD	Sistema de Gerenciamento de Banco de Dados
SQL	Structured Query Language
VaR	Valor-em-Risco

SUMÁRIO

1 – INTRODUÇÃO	12
1.1 MOTIVAÇÃO	12
1.2 OBJETIVOS	13
1.2.1 Objetivo Geral	13
1.2.2 Objetivos Específicos	14
1.3 JUSTIFICATIVA	14
1.4 ORGANIZAÇÃO DO TRABALHO	16
2 – FUNDAMENTAÇÃO TEÓRICA	17
2.1 ESTRUTURA INFORMACIONAL DAS CEASAS	17
2.2 MODELAGEM DE DATA WAREHOUSES	18
2.3 MÉTRICAS PARA AVALIAÇÃO DO RISCO DE PREÇO	21
2.4 VISUALIZAÇÃO DE DADOS PARA AVALIAÇÃO DA INFORMAÇÃO	23
3 – MATERIAIS E MÉTODOS	25
3.1 MATERIAIS	25
3.2 MÉTODOS	25
3.2.1 Tratamento de Dados Perdidos	25
3.2.2 Definição dos SGBD's para os DW's ROLAP e NoSQL	26
3.2.3 Cálculo das Métricas de Risco	27
3.2.4 Construção da Interface Gráfica do Usuário	28
4 – RESULTADOS	29
4.1 MODELAGEM DIMENSIONAL DOS DW'S ROLAP e NoSQL	29
4.2 GUI PARA VISUALIZAÇÃO E ANÁLISE DE PREÇOS	31
4.3 GUI PARA VISUALIZAÇÃO E ANÁLISE DO RISCO DE PREÇOS	33
4.4 DISCUSSÃO	39
5 – CONCLUSÃO	42
5.1 LIMITAÇÕES	42
5.2 TRABALHOS FUTUROS	42
5.3 CONSIDERAÇÕES FINAIS	43
Referências	44

1 INTRODUÇÃO

O setor agrário, seja em âmbito nacional ou internacional, vem aumentando o seu interesse por tecnologias de *Business Intelligence* (BI) e *Data Warehouse* (DW) para otimizar seu processo decisório, cuja produção de informações estratégicas úteis vem sendo desenvolvida, majoritariamente, através de soluções tecnológicas gratuitas de código aberto (BRUM; LAMPERT; CAMARGO, 2019).

Uma das formas de uso dessas tecnologias é a abordagem multidimensional de elaboração de DW relacionais. Vários trabalhos propõem a construção de cubos de dados para a armazenagem e rápida recuperação de métricas agrícolas focadas entre o pré-cultivo (ex: área plantada em hectares, tipo de semente) e a colheita (ex: produção em hectares), visando auxiliar nas decisões do usuário final (ex.: produtores rurais) (CHATURVEDI; RAI, 2008; GUPTA; MAZUMDAR, 2013; NIZAR et al., 2021). Numa extensão dos trabalhos anteriores, (NILAKANTA; SCHEIBE; RAI, 2008) também incluem métricas de pós-colheita (ex.: preço), que envolve a comercialização do produto.

Outra possibilidade de aplicação dessas tecnologias é pelo paradigma do *Big Data*. Ngo, Le-Khac e Kechadi (2019) propuseram uma arquitetura de DW alicerçada em bancos de dados NoSQL para lidar com grandes volumes de dados, inerentes da agricultura de precisão, com métricas abrangendo desde o pré-cultivo até a pós-colheita.

1.1 MOTIVAÇÃO

Os estudos descritos acima não trazem quaisquer aplicações de BI/DW para o gerenciamento do risco de preço de produtos agrícolas (i.e., repentina variabilidade de preço num curto espaço de tempo), particularmente para os produtos frescos.

Neste quesito, eles se destacam porque não se beneficiam de mecanismos de *hedging* (proteção) para a mitigação do risco de preço (ex.: contratos futuros, opções), disponíveis em bolsas de mercadorias e de futuros, justamente por não serem considerados *commodities* (AZEVEDO, 1997), diferentemente do que acontece habitualmente com os grãos (ex.: soja, milho, café), os quais possuem várias possibilidades de contratos financeiros destinados para este propósito.

Dada a inexistência de contratos financeiros de *hedging* para a mitigação do risco de preço de produtos frescos, uma das alternativas factíveis para este mercado seria o desenvolvimento de aplicações de BI/DW voltadas para o seu gerenciamento, tendo como referência as pesquisas realizadas para o mercado de *commodities* (BARBAGLIA; CROUX; WILMS, 2020; DAHL; OGLEND; YAHYA, 2020). Logo, um dos argumentos favoráveis para a criação de tais aplicações passa pela quantificação desse risco de preço.

Nessa perspectiva, Wang, Zhao e Huang (2010) foram pioneiros ao calcular o risco de

preço no mercado de frutas na China, classificando-as em três níveis de risco, cuja sazonalidade, capacidade de armazenamento e facilidade no transporte sugerem ser fatores determinantes de seus níveis de risco de preço.

No Brasil, [Oliveira et al. \(2020\)](#) identificaram, com o uso de técnicas financeiras avançadas de *downside risk*, que os produtos frescos comercializados no Estado do Paraná possuem um nível de risco de preço relevante que não deve ser negligenciado pelos produtores rurais, sob pena de permanecerem expostos a ele e terem suas receitas e produção futuras comprometidas no longo prazo.

Todavia, ainda que [Nilakanta, Scheibe e Rai \(2008\)](#) e [Ngo, Le-Khac e Kechadi \(2019\)](#) tenham considerado a métrica *preço* em seus esquemas de DW para produtos agrícolas, e [Wang, Zhao e Huang \(2010\)](#) e [Oliveira et al. \(2020\)](#) tenham mensurado o risco de preço de produtos agrícolas não considerados *commodities*, nenhum desses estudos avançou para a construção de uma proposta de BI que lide com este risco subjacente.

Tal ausência de um sistema computacional que proveja informações claras e sistematizadas tende a provocar incerteza nos agentes econômicos envolvidos quanto ao preço esperado de comercialização de seus produtos frescos, cujas decisões financeiras e de produção são afetadas por vieses heurísticos (i.e., avaliações subjetivas de problemas baseadas em suas experiências passadas). Estes vieses os conduzem a cometerem erros sistemáticos e previsíveis, como o da *Representatividade* (ex.: concepções errôneas de possibilidades, insensibilidade à previsibilidade), o da *Disponibilidade* (ex.: correlações ilusórias, vieses de imaginabilidade) e o de *Ajuste e Ancoragem* (ex.: ajustes insuficientes, ancoragem na avaliação subjetiva das distribuições de probabilidade) ([TVERSKY; KAHNEMAN, 1974](#)). Exemplos desses vieses são vistos em [Oliveira e Santos \(2012\)](#), ao relatar os fatos que culminaram na falência do Banco Barings (1762-1995) através da análise semiótica do filme “A Fraude”.

Nesse cenário de incerteza de informações, um exemplo de como tecnologias de BI/DW poderiam ser melhor aproveitadas no setor agrícola é vista em [Shrivastava, Pal e Walia \(2019\)](#). Ainda que não tenham focado no gerenciamento de risco de preço, eles enfatizam que o preço é a métrica mais importante para os produtores rurais, especialmente para os pequenos. Assim, eles trazem uma versão preliminar de um sistema de *Market Intelligence* para prever o preço futuro de certos produtos agrícolas na Índia (oleaginosas e leguminosas).

1.2 OBJETIVOS

Diante desse contexto, os principais objetivos do trabalho são apresentados a seguir.

1.2.1 Objetivo Geral

Desenvolver um sistema *web* de BI para avaliar o risco de preço de produtos frescos em Ceasas brasileiras.

1.2.2 Objetivos Específicos

1. Elaborar a modelagem dimensional dos Data Warehouses para o sistema *web* de BI
2. Estabelecer as métricas de avaliação do risco de preço para o sistema *web* de BI
3. Produzir uma aplicação *web* de suporte à decisão para o sistema de BI

1.3 JUSTIFICATIVA

Em termos teóricos, uma análise comparativa entre os trabalhos relacionados sobre o uso de BI/DW no setor agrícola com a proposta trazida neste artigo pode ser vista na Tabela 1. Os parâmetros empregados para comparação são: (1) o uso de DW OLAP; (2) o uso de DW NoSQL; (3) se o preço é uma métrica do DW; (4) o uso de alguma medida de risco; (5) se apresenta uma interface gráfica do usuário *web* para BI; (6) se permite a visualização de dados espaciais; (7) se realiza testes de usabilidade; (8) se permite acesso público. A codificação adotada na Tabela 1 é (S) para sim, (N) para não e (P) para parcial.

Tabela 1 – Análise comparativa dos trabalhos relacionados

Autores	1	2	3	4	5	6	7	8
Chaturvedi e Rai (2008)	S	N	S	N	S	S	N	N
Nilakanta, Scheibe e Rai (2008)	S	N	S	N	S	S	S	N
Gupta e Mazumdar (2013)	S	N	N	N	N	N	N	N
Nizar et al. (2021)	S	N	S	N	S	P	N	N
Shrivastava, Pal e Walia (2019)	P	N	S	P	S	N	N	N
Ngo, Le-Khac e Kechadi (2019)	N	S	S	N	N	N	N	N
Esta Monografia	S	S	S	S	S	S	S	S

Chaturvedi e Rai (2008) desenvolveram, através de uma plataforma proprietária, um DW OLAP relacional contendo 13 *data marts*, todos num esquema estrela, destinados para armazenar dados de produtos agrícolas na Índia, entre as etapas de pré-cultivo e colheita, bem como permitindo ao usuário a realização de consultas *online* mediante uma autenticação. Entretanto, ainda que o preço seja armazenado no DW, o seu nível de granularidade é pequeno, já que as consultas nesse caso são feitas apenas no escopo do país. Além disso, nenhuma medida de risco foi identificada para quaisquer métricas citadas.

Nilakanta, Scheibe e Rai (2008) também apresentaram um desenho e a implementação de um DW OLAP relacional via plataforma proprietária para o setor agrícola indiano, num esquema estrela, só que agora no escopo governamental e com a realização de testes de usabilidade. As consultas dos usuários podem ser realizadas via *web* mediante uma autenticação. Ainda que o nível de granularidade do preço tenha aumentado nesse caso (diário, semanal e mensal), novamente nenhuma medida de risco foi localizada para quaisquer métricas citadas.

Gupta e Mazumdar (2013) elaboraram DW's OLAP, nos esquemas estrela e floco-de-neve, cujas tabelas de dimensão estão direcionadas para as características fenológicas de culturas agrícolas (CÂMARA, 2006), isto é, os fatores que afetam as diferentes fases do seu

crescimento e desenvolvimento, derivados do manejo ou, sobretudo, do meio ambiente. Contudo, foram apresentados somente os projetos conceitual e lógico desses DW's.

Nizar et al. (2021) projetaram e construíram um DW OLAP relacional, com seis *data marts* num esquema constelação, para auxiliar no armazenamento e recuperação de dados relacionados à diversificação agrícola, biológica e de mercado das culturas, especialmente as pertencentes dos sistemas alimentares. Normalmente, esses dados estão dispersos entre os pequenos produtores, que muitas vezes não se beneficiam das soluções digitais vistas nos sistemas de produção de grandes produtores. Todavia, apesar da ampla variedade de consultas previstas, nenhuma medida de risco foi identificada para quaisquer métricas citadas, bem como o acesso público ao site estava indisponível durante este levantamento.

Shrivastava, Pal e Walia (2019) propuseram um *Market Intelligence* para o setor agrícola indiano, com a vantagem de realizar previsões de preço e análise de tendências através dos modelos ARIMA e *deep learning*. Os dados seriam armazenados num DW OLAP relacional, num esquema floco-de-neve, dado que a proposta ainda é uma prova de conceito, cujos testes de usabilidade e de acesso público não foram implementados.

Diferentemente dos trabalhos citados anteriormente, Ngo, Le-Khac e Kechadi (2019) utilizaram um DW NoSQL com um projeto lógico híbrido OLAP e utilizando três diferentes bancos de dados (Hive, MongoDB e Cassandra), argumentando que os dados agrícolas possuem todas as características elementares de *Big Data*: volume, variedade, velocidade e veracidade. Embora o desempenho da proposta NoSQL tenha indicativo de comportamento superior ao DW relacional quanto à velocidade das consultas, ela ainda não foi implementada, bem como nenhuma medida de risco foi identificada para quaisquer métricas citadas.

Em termos práticos, é relevante apresentar um panorama das Ceasas na movimentação da economia brasileira: em 2020, elas transacionaram 16,3 milhões de toneladas de hortifrutigranjeiros, avaliados em R\$ 42,3 bilhões, destacando-se a região sudeste, que respondeu por cerca de 51% desses valores (CONAB, 2021a). Ademais, as Ceasas promovem uma reestruturação econômica (ex.: diminuição dos custos de transação), política (ex.: segurança alimentar e nutricional) e cultural (ex.: hábitos alimentares) do território da circulação de hortifrutigranjeiros no país (QUEIROZ, 2018).

Dessa forma, é esperado que os resultados deste estudo contribuam para a diminuição dos efeitos adversos causados pela assimetria de informação de dados agrícolas agregados, conforme mostrado por Coble et al. (2018), Aker, Ghosh e Burrell (2016), que, por conseguinte, reduz a incerteza dos agentes econômicos envolvidos, melhorando assim a eficiência desse mercado. Além disso, intenciona-se preencher as principais lacunas na literatura sobre BI/DW no setor agrícola, a saber: a) diagramas de implementação de DW's tanto na abordagem relacional como na NoSQL; b) métricas consistentes de avaliação do risco de preço; c) visualização de dados espaciais; d) testes de usabilidade; e) acesso público.

1.4 ORGANIZAÇÃO DO TRABALHO

O Capítulo 2 traz a fundamentação teórica utilizada no desenvolvimento do sistema *web* de BI/DW proposto. O Capítulo 3 detalha a metodologia adotada para a implementação do sistema *web* de BI/DW proposto. O Capítulo 4 traz os resultados para os objetivos específicos definidos neste estudo. Finalizando, o Capítulo 5 traz as limitações, os trabalhos futuros sugeridos e as considerações finais deste estudo.

2 FUNDAMENTAÇÃO TEÓRICA

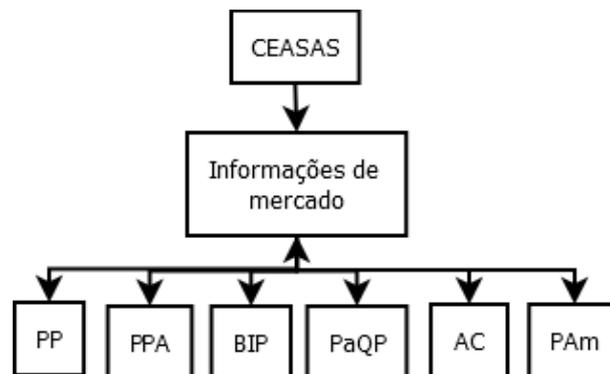
Neste capítulo, são apresentadas as principais referências teóricas que dão suporte ao desenvolvimento do sistema *web* de BI/DW proposto. Aqui são mostradas a estrutura informacional das Ceasas, a modelagem de data warehouses, as métricas para avaliação do risco de preço e os aspectos sobre visualização de dados para avaliação da informação.

2.1 ESTRUTURA INFORMACIONAL DAS CEASAS

As Centrais de Abastecimento S.A. (Ceasas) no Brasil fazem parte de uma rede de distribuição atacadista de produtos frescos, constituindo o espaço de formação de preços, normas de comercialização, informações de mercado e de definição de parâmetros de qualidade e classificação. Sua atividade principal é o abastecimento do mercado de hortigranjeiros, recebendo produtos das mais diversas regiões brasileiras, bem como de fora do país (ABRACEN, 2011).

Outrossim, as Ceasas possuem sistemas de informação capazes de registrar dados relativos aos processos mercantis por elas controlados, informando os números processados para toda a cadeia envolvida na produção/comercialização de hortigranjeiros. A Figura 1 mostra os tipos de informação de mercado disponíveis para consulta, detalhados a seguir.

Figura 1 – Informações de mercado produzidas pelas Ceasas



Fonte: Adaptado de ABRACEN (2011)

A *Procedência de Produtos (PP)* é a informação obtida junto a todos que entram na Ceasa com o propósito de vender ou entregar mercadoria: hortigranjeiros, cereais, produtos diversos alimentícios e não alimentícios, etc. É obrigatório portar um documento de origem da mercadoria, seja nota fiscal ou o romaneio de entrada de mercadoria.

A *Pesquisa de Preços no Atacado (PPA)* capta as mais diversas informações relativas ao mercado, tais como oferta, qualidade do produto presente no mercado, tendências, preferências, etc., realizada por técnicos agrícolas das Ceasas com pelo menos três anos de experiência.

O *Boletim Informativo de Preços (BIP)* é publicado em várias mídias (impresso, e-mail, fax, *website*, etc.) contendo a informação em três níveis (preço mínimo, preço mais comum e preço máximo), sempre considerando as ocorrências para um mesmo produto, variedade, classificação e embalagem, bem como a situação de mercado para cada produto (firme, estável ou fraco).

O *Painel de Quantidade e Preços (PaQP)* é um painel eletrônico ou grandes televisores que mostram, em tempo real, a entrada das principais mercadorias nas Ceasas com a finalidade de informar, tanto aos produtores quanto aos compradores, o volume de entrada no mercado em questão e nos principais mercados alternativos.

A *Análise Conjuntural (AC)* é a interpretação do que ocorreu, ao longo do mês, com os principais produtos de cada mercado, através de análises comparativas em retrospectiva, visando estabelecer perspectivas para o futuro por produto, grupo, subgrupo, setor e até mesmo para o mercado.

Finalmente, as *Pesquisas por Amostragem (PAm)* estabelecem avaliações relativas aos agentes envolvidos nas atividades dos mercados atacadistas (produtores, concessionários e compradores), possibilitando identificar necessidades de ajustes operacionais das Ceasas, bem como orientam o mercado com informações relativas a novas tendências de consumo e de oportunidades de negócios (ABRACEN, 2011).

Na seção seguinte, são apresentadas algumas técnicas sobre modelagem de *Data Warehouses*, fundamentais para o acesso sistematizado dos dados produzidos por várias empresas e instituições, tais como as Ceasas brasileiras.

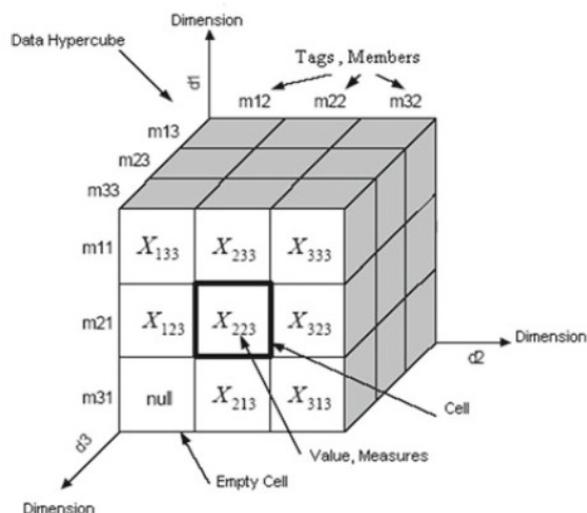
2.2 MODELAGEM DE DATA WAREHOUSES

Os *Data Warehouses (DW)* são grandes repositórios focados no suporte à decisão, cujos dados consolidados, oriundos de diversas fontes, estão armazenados em estruturas que agilizam a sua rápida recuperação para a análise do negócio, visando manter sua vantagem competitiva (VAISMAN; ZIMÁNYI, 2014). Nesse sentido, duas abordagens atendem ao propósito de modelagem de DW's: 1) a OLAP [*Online Analytical Processing*] (CODD; CODD; SALLEY, 1993), que segue o paradigma transacional ACID (*Atomicity, Consistency, Isolation, Durability*) (HAERDER; REUTER, 1983), e 2) a NoSQL [*Not only Structured Query Language*], que segue o teorema CAP (*Consistency, Availability, Partition tolerance*) (GILBERT; LYNCH, 2002).

Os DW's OLAP guardam os dados num espaço n -dimensional chamado hipercubo, definido por fatos e dimensões, conforme a Figura 2 (AYUPOV; ARIFJANOV, 2018). Ele é o modelo conceitual do DW. Os fatos são as células do hipercubo que contêm os valores numéricos, também chamados de medidas, que estão associados com alguma função de agregação (ex.: soma, média, etc.). Já as dimensões, elas são as perspectivas disponíveis no hipercubo (as faces) para analisar os fatos, podendo ter vários níveis de detalhamento (granularidade) para a realização das operações OLAP (*roll-up, drill-down, slice, etc.*).

Para a implementação de um DW OLAP, o enfoque relacional (ROLAP) é o mais usual,

Figura 2 – Exemplo de um hipercubo de dados OLAP



Fonte: [Ayupov e Arifjanov \(2018\)](#)

porque todo o gerenciamento dos dados está fundamentado em DBMS [*DataBase Management System*] relacionais, os quais estão bem padronizados e fornecem grande capacidade de armazenamento ([VAISMAN; ZIMÁNYI, 2014](#)).

Quanto ao modelo lógico, os DW's ROLAP podem ser representados em quatro esquemas ([VAISMAN; ZIMÁNYI, 2014](#)):

- O *star*, que possui uma tabela de fatos central, geralmente normalizada, e um conjunto de tabelas de dimensões geralmente não normalizadas, o que pode causar redundância de dados;
- O *snowflake*, que resolve o problema da redundância de dados pela normalização das tabelas de dimensões. Contudo, o desempenho das consultas é negativamente afetado, dado que mais operações de junção terão que ser realizadas;
- O *starflake*, que é uma combinação dos esquemas *star* e *snowflake*, no qual somente algumas tabelas de dimensões são normalizadas;
- O *constellation*, que possui múltiplas tabelas de fatos, que compartilham suas tabelas de dimensões entre si.

Por fim, para garantir um tempo de resposta adequado às consultas, três técnicas clássicas podem ser aplicadas, as quais representam o modelo físico de um DW ROLAP ([VAISMAN; ZIMÁNYI, 2014](#)):

- As *visões materializadas*, que são consultas *ad hoc* previamente antecipadas e fisicamente armazenadas no banco de dados, cujo objetivo é pré-calcular operações custosas, como junções e agregações;
- A *indexação*, destinada para consultas cujas junções e agregações devem ser calculadas em tempo de execução, valendo-se dos índices *bitmap* ou de junção para esse fim;
- O *particionamento*, que consiste em dividir o conteúdo de uma tabela em vários arquivos,

par chave-valor). Normalmente, eles são empregados para gerenciar dados altamente conectados e realizar consultas complexas através de relacionamentos semanticamente significativos, conforme exemplificado pela Figura 3.

Uma vez apresentadas as possibilidades de modelagem de *Data Warehouses* para a armazenagem sistemática dos dados produzidos pelas Ceasas brasileiras, a próxima seção trata de técnicas para mensurar o risco associado às séries históricas de ativos, cujos preços de produtos frescos podem ser enquadrados nesse quesito.

2.3 MÉTRICAS PARA AVALIAÇÃO DO RISCO DE PREÇO

O início da moderna teoria financeira aconteceu com o surgimento da Teoria do Portfólio (MARKOWITZ, 1952; MARKOWITZ, 1991), que trouxe uma definição matemática para se resolver um importante *trade-off*: a distribuição completa de recursos monetários em ativos financeiros num portfólio diversificado, de modo que i) seu risco fosse o menor possível, com certo retorno esperado, ou ii) seu retorno fosse o máximo possível, com certo risco aceitável.

Num portfólio diversificado, o retorno é definido pelo valor esperado (i.e., sua média), o risco é determinado pela variância e a função utilidade do investidor é definida pela equação (1).

$$f(w, y) = \underbrace{w' \mu}_{\text{retorno}} - \frac{\gamma}{2} \underbrace{w' \Sigma w}_{\text{risco}} \quad (1)$$

$$w = \frac{1}{\gamma} \Sigma^{-1} \left[\mu - \left(\frac{\mathbf{1}' \Sigma^{-1} \mu - \gamma}{\mathbf{1}' \Sigma^{-1} \mathbf{1}} \right) \mathbf{1} \right] \quad (2)$$

Num formato matricial, $w \in \mathbb{R}^n$ é o vetor dos pesos, $\mu \in \mathbb{R}^n$ é o vetor das médias e $\Sigma \in \mathbb{R}^{n \times n}$ é a matriz variância-covariância dos n ativos pertencentes ao portfólio diversificado, cujo *trade-off* retorno-risco do investidor é descrito numa fronteira eficiente do par média-variância em w , balanceado pelo seu coeficiente de aversão ao risco $\gamma \in \mathbb{R}^+$.

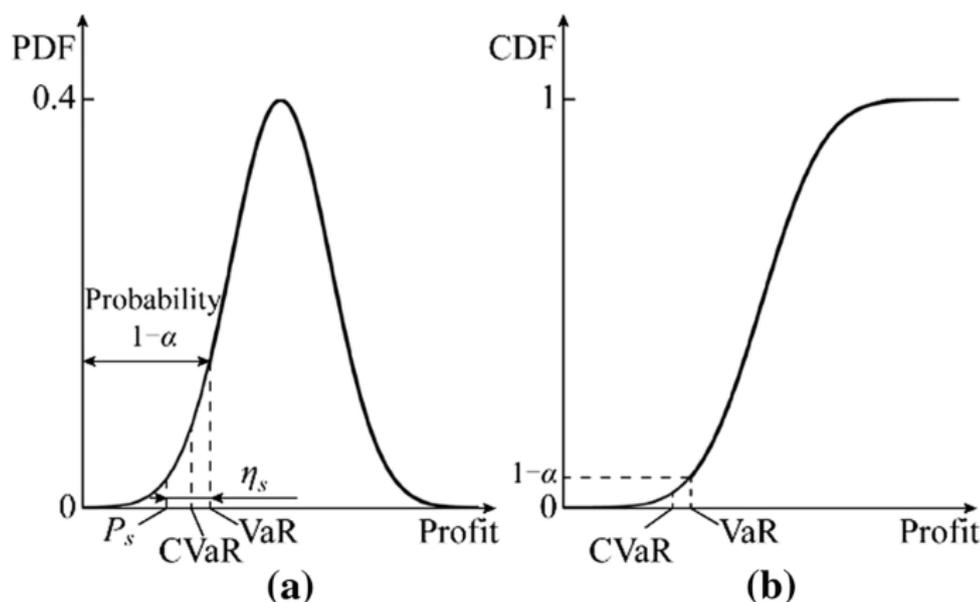
Logo, os pesos que maximizam a equação (1), para certo γ , é obtido analiticamente pela equação (2), cujo vetor $\mathbf{1} = [1, 1, \dots, 1]' \in \mathbb{R}^n$ atrelado à condição $w' \mathbf{1} = 1$ indica que todo o capital disponível deve ser totalmente investido (DAS et al., 2010).

Por outro lado, vale destacar que a solução analítica da equação (2) permite a situação da “venda a descoberto”: os pesos dos ativos podem ser negativos, contrapondo os métodos de programação quadrática ou linear que permitem a restrição de não-negatividade dos pesos. Sua principal vantagem é atenuar a sensibilidade do portfólio aos movimentos do mercado (ELTON et al., 2004), e assim prevenir perda de eficiência com a redução da região viável de portfólios eficientes, dado que basicamente todos os ativos seriam escolhidos (SIMMONS, 1998).

Já a Teoria Pós-Moderna do Portfólio (ROM; FERGUSON, 1994) traz as medidas de *downside risk*, porque reconhece que as reações dos investidores ao risco são assimétricas (a

variância é uma medida simétrica de risco). Dentre elas, o Valor em Risco Condicional (CVaR) mede a expectativa condicional de perdas maiores que certa quantia q . Esta, por sua vez, é a menor perda estimada com probabilidade α de ocorrer em certo tempo t , também chamada de Valor em Risco (VaR). Ambas as medidas estão associadas à ocorrência de eventos extremos (perdas) em séries históricas e em condições normais de mercado, sejam elas de ativos ou de portfólios (JORION, 2002).

Figura 4 – Ilustração do CVaR e VaR



Fonte: Aydogdu, Tör e Güven (2019)

A Figura 4 ilustra, por exemplo, a ocorrência do CVaR/VaR para as funções densidade (a) de probabilidade e (b) acumulada, para certa expectativa de lucros (AYDOGDU; TÖR; GÜVEN, 2019). Contudo, somente o CVaR é considerado uma medida coerente de risco, porque a) atende ao axioma da subaditividade (o risco do portfólio nunca é maior do que a soma dos ativos individuais); b) reconhece concentração de riscos; c) encoraja uma alocação razoável de riscos entre os agentes, já que considera as consequências econômicas dos eventos que deseja controlar; d) é uma função convexa (PFLUG, 2000; ROMAN; MITRA, 2009).

No contexto dos portfólios, encontrar os pesos w que minimizam o CVaR também implica que o VaR mínimo é alcançado, em consonância com a definição de ambas as medidas. A equação (3) calcula o CVaR em função de w , VaR e α , numa aproximação discreta (ROCKAFELLAR; URYASEV, 2000).

$$\widehat{CVaR}(\hat{w}, \widehat{VaR}, \alpha) = \widehat{VaR} + [(1 - \alpha) M]^{-1} \sum_{m=1}^M \max(\hat{w}'r_s - \widehat{VaR}, 0) \quad (3)$$

Os valores \hat{w} , \widehat{VaR} e \widehat{CVaR} , juntamente com uma variável auxiliar u_s , são estimados via programação linear, tendo as seguintes restrições: $w'1 = 1$, $w'r_s \geq \mu_{min}$, $u_s \geq 0$ e

$u_s \geq w'r_s - \widehat{VaR}$. Para prevenir o caso de “vendas a descoberto”, é preciso adicionar a restrição $w \geq 0$. Ademais, a equação (3) dispensa o conhecimento prévio de qualquer distribuição de probabilidade, pois considera em seu cálculo uma amostragem de retornos r_s em $p(\mu)$, de tamanho M .

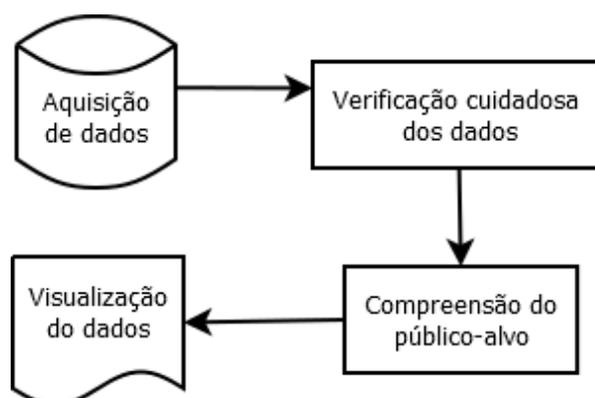
Ainda que tenham sido originalmente desenvolvidas para ativos financeiros, vários trabalhos vêm demonstrando a aplicabilidade da Teoria do Portfólio e do Valor em Risco Condicional para ativos reais, em especial no setor agrícola, conforme levantamento realizado por (OLIVEIRA et al., 2020).

Por fim, com a exposição das métricas para a avaliação do risco de preço de produtos frescos, a seção subsequente trata dos procedimentos para que a informação produzida a partir dos dados das Ceasas brasileiras possa ser apresentada de modo que seus usuários consigam, apropriadamente, utilizá-la em seus processos decisórios.

2.4 VISUALIZAÇÃO DE DADOS PARA AVALIAÇÃO DA INFORMAÇÃO

A visualização de dados pode ser entendida como uma área de estudo que procura estabelecer, a partir de uma ampla variedade de dados qualitativos ou quantitativos disponíveis, um *spectrum* sobre o que o público-alvo quer ver. Ainda que uma teoria robusta de visualização de dados esteja em construção (UNWIN, 2020), ela pode ser resumida em como as pessoas percebem, compreendem e reagem à informação contida em tabelas e gráficos, o que por sua vez, desperta a necessidade de se avaliar a representatividade e o grau de esclarecimento do que está sendo visto, principalmente num contexto de *Big Data* (AZZAM et al., 2013).

Figura 5 – Diagrama de visualização de dados



Fonte: Adaptado de Azzam et al. (2013) e Unwin (2020)

Para que o sucesso da visualização de dados seja alcançado, é importante que as etapas a seguir sejam plenamente atingidas, conforme ilustradas na Figura 5: a aquisição de dados potencialmente visualizáveis, seguindo os preceitos do *Knowledge Discovery in Databases* (KDD) (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996); a sua verificação cuidadosa para garantir a acurácia da informação neles contida, evitando assim possíveis omissões ou super-

representações; e a compreensão do público-alvo, de modo que a figura que lhe é mostrada seja envolvente e, principalmente, elucidativa (AZZAM et al., 2013).

Durante a execução das etapas da Figura 5, decidir como os dados podem ser apresentados para seu público-alvo pode-se tornar uma tarefa desafiadora para os analistas e desenvolvedores de sistemas computacionais, dada a ampla variedade de opções e combinações de gráficos (UNWIN, 2020).

Normalmente, o público-alvo quer acessar os detalhes visualmente, conectar diferentes partes da informação, identificar desvios e padrões interessantes e explorar a informação de múltiplas perspectivas ou níveis, muitas vezes de maneira interativa.

Assim, a análise visual quantitativa tem a capacidade de se aprofundar nos números, mas mantendo o acesso a uma visão panorâmica dos dados em questão, conforme for o propósito da visualização: quantidades, distribuições, proporções, relacionamentos x-y, dados geoespaciais ou incerteza (WILKE, 2019).

Nesse aspecto, os *dashboards* e os infográficos criam representações visuais únicas e de compreensão imediata do desempenho de um projeto/programa/temática em vários indicadores ao mesmo tempo, com a seguinte diferença entre si: enquanto que os *dashboards* focam no público interno de uma organização por trazerem os dados em relatórios mais tradicionais, os infográficos são destinados para o seu público externo, por formatarem os dados numa prosa narrativa (AZZAM et al., 2013).

Ainda nesse âmbito, é importante destacar que, para além da apresentação de resultados, cujo foco é comunicar uma informação conhecida e bem desenhada objetivando a atenção do público-alvo, a visualização de dados também é útil no decorrer da análise exploratória e da mineração de dados, cujo propósito do analista é encontrar novas informações que possam despertar interesse, como tendências, agrupamentos, padrões, discrepâncias, dentre outros (UNWIN, 2020).

3 MATERIAIS E MÉTODOS

Esta monografia, enquanto documento de realização de uma pesquisa científica, foi classificada como uma pesquisa aplicada (finalidade), do tipo descritiva (objetivo) e quantitativa (abordagem), utilizando-se da pesquisa documental (procedimentos) e da pesquisa bibliográfica (fonte de dados). A partir deste preâmbulo, este capítulo detalha os materiais e os métodos utilizados para atender aos objetivos propostos neste estudo.

3.1 MATERIAIS

O Programa Brasileiro de Modernização do Mercado Hortigranjeiro (Prohort) foi instituído pelo Governo Federal por meio da Portaria nº 171, de 29/03/2005, do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), e ampliado em suas funções pela Portaria nº 339, de 11/04/2014, também do MAPA. O Prohort tem por finalidade contribuir para o desenvolvimento e modernização do setor hortigranjeiro nacional, proporcionando, através de seu site, uma maior interação entre os diversos agentes governamentais envolvidos e os integrantes da cadeia de produção e distribuição, além de buscar a melhoria e a ampliação das funções dos mercados atacadistas (CONAB, 2021a).

Nesse contexto, foram coletados no site do Prohort (CONAB, 2021b) os preços mensais, em Reais (R\$), de todos os produtos frescos comercializados nas Ceasas brasileiras, entre 2006 e 2021. Assim, a amostra esperada foi de 508.032 tuplas, resultante da multiplicação de 48 produtos frescos por 56 Ceasas que possuíam *quaisquer* dados disponíveis nos 189 meses entre janeiro/2006 e setembro de 2021. Todavia, a amostra final foi de 499.527 tuplas, tendo em vista que 14 das 56 Ceasas selecionadas não comercializavam todos os 48 produtos identificados neste levantamento. Já a localização geográfica (latitude e longitude) de cada Ceasa foi coletada manualmente pelo site *Google Maps*, utilizando-se "Ceasa <nome da cidade>" como *string* de consulta. Por fim, um algoritmo foi escrito na linguagem R para realizar a etapa de ETL (*Extract-Transform-Load*) desses dados antes de carregá-los nos DW's relacional e NoSQL, incluindo inclusive o tratamento de dados perdidos, visto a seguir.

3.2 MÉTODOS

3.2.1 Tratamento de Dados Perdidos

Ao realizar a coleta dos preços mensais, foi percebida a ocorrência de dados perdidos nas séries históricas dos produtos frescos selecionados, cuja quantidade e percentual, por Estado, são mostrados na Tabela 2. Nela, é percebido que sete [RO,SE,AC,MT,CE,PR,PI] dos 24 Estados que possuem Ceasas apresentaram mais dados perdidos do que dados coletados em suas séries históricas de preços mensais de produtos frescos.

Tabela 2 – Quantidade de dados perdidos nas séries de preços de produtos frescos, por Estado

Estado	Perdido	Contagem	(%)	Estado	Perdido	Contagem	(%)
AC	Sim	6.278	69,20	PB	Sim	12.489	46,53
	Não	2.794	30,80		Não	14.349	53,47
AL	Sim	2.617	28,85	PE	Sim	7.080	45,68
	Não	6.455	71,15		Não	8.418	54,32
BA	Sim	10.269	38,00	PI	Sim	4.828	53,22
	Não	16.758	62,00		Não	4.244	46,78
CE	Sim	19.625	58,01	PR	Sim	25.893	57,56
	Não	14.206	41,99		Não	19.089	42,44
DF	Sim	1.562	17,22	RJ	Sim	109	1,20
	Não	7.510	82,78		Não	8.963	98,80
ES	Sim	160	1,76	RN	Sim	445	5,01
	Não	8.912	98,24		Não	8.438	94,99
GO	Sim	1.587	17,49	RO	Sim	7.008	77,25
	Não	7.485	82,51		Não	2.064	22,75
MA	Sim	4.492	49,51	RS	Sim	4.063	22,39
	Não	4.580	50,49		Não	14.081	77,61
MG	Sim	26.394	41,56	SC	Sim	454	5,00
	Não	37.110	58,44		Não	8.618	95,00
MS	Sim	137	1,51	SE	Sim	6.165	69,40
	Não	8.935	98,49		Não	2.718	30,60
MT	Sim	5.094	64,17	SP	Sim	30.666	24,33
	Não	2.844	35,83		Não	95.397	75,67
PA	Sim	781	8,61	TO	Sim	3.738	41,20
	Não	8.291	91,39		Não	5.334	58,80

Para solucionar essa situação, os dados perdidos foram substituídos pela média mensal dos dados disponíveis. E em caso da média mensal também for um dado perdido, esta foi substituída pela mediana das médias mensais disponíveis.

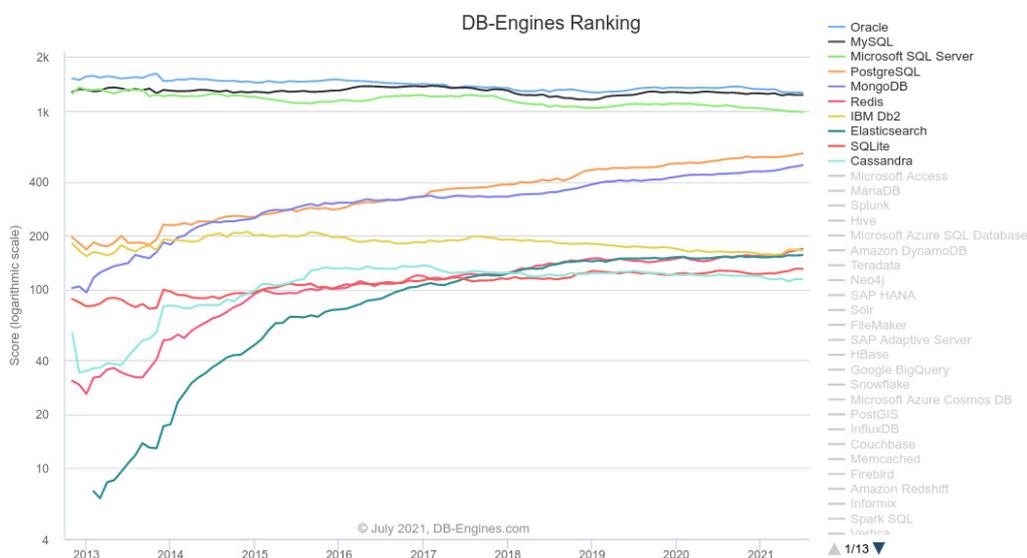
3.2.2 Definição dos SGBD's para os DW's ROLAP e NoSQL

Em relação ao objetivo específico 1, foi modelado um DW ROLAP do tipo *star*, que possui um melhor desempenho nas consultas que o *snowflake* por causa de suas tabelas de dimensões desnormalizadas (VAISMAN; ZIMÁNYI, 2014). A implementação do DW ROLAP via Postgresql, versão 12, seguiu as diretrizes da *Amazon Relational Database Service* [Amazon RDS] (AMAZON WEB SERVICES, 2021), numa instância de nível gratuito.

Esta escolha se deu porque, além de permitir a separação conveniente dos níveis de responsabilidade entre o cliente e o provedor de serviços computacionais configuráveis e rapidamente provisionados em nuvem, a *Amazon Web Services*, em 2020, era a companhia líder enquanto plataforma de serviços e infraestrutura em nuvem (SHIFF, 2020). Quanto ao Postgresql, ainda que ocupe o quarto lugar no ranking de Sistemas de Gerenciamento de Banco de Dados (SGBD) relacional quanto à sua popularidade, ele é o único que apresenta uma

tendência de crescimento significativa desde 2014, conforme visto na [Figura 6](#).

Figura 6 – Tendência histórica de popularidade dos SGBD's



Fonte: [DB-engines \(2021\)](#)

Quanto à modelagem de DW NoSQL, foi utilizado o orientado a coluna, que consegue armazenar quaisquer estruturas de dados, possui alto desempenho nas consultas de agregação e é altamente eficiente na compressão de dados ([DEHDOUH et al., 2015](#); [BOUSSAHOUA; BOUSSAID; BENTAYEB, 2017](#)).

Para a sua implementação, foi utilizado o Apache Cassandra, versão 3.11, por ser o principal SGBD NoSQL colunar desde 2013, por estar listado entre os *top 10* mais populares SGBD's ([DB-ENGINES, 2021](#)) e pela sintaxe do *Cassandra Query Language* (CQL) ser semelhante, em muitos aspectos, ao *Structured Query Language* (SQL) ([CARPENTER; HEWITT, 2016](#)).

O DW NoSQL foi instalado e configurado num computador local, seguindo os preceitos do *Query-first design*: as consultas foram modeladas primeiro e depois os dados foram organizados em torno delas ([CARPENTER; HEWITT, 2016](#)), em *BigFact Tables* ([DEHDOUH et al., 2015](#)).

3.2.3 Cálculo das Métricas de Risco

O pacote *fPortfolio* ([WUERTZ et al., 2009](#); [WUERTZ et al., 2020](#)) da linguagem R foi empregado na escrita do algoritmo usado para a realização do cálculo das medidas de risco dos produtos frescos comercializados nas Ceasas brasileiras, vistas na seção 2.3. O design e a

otimização de portfólios com o pacote *fPortfolio* estão, de uma maneira intuitiva, alicerçados em quatro fundamentos:

1. A *definição* das entradas do portfólio, escrevendo as especificações, carregando os dados dos ativos selecionados e configurando as restrições envolvidas;
2. A *otimização* do portfólio, incluindo o cálculo de portfólios únicos, tais como portfólios viáveis, eficientes, tangência (máximo retorno/risco) ou variância mínima (risco mínimo global), bem como a avaliação de toda a fronteira eficiente;
3. A *geração* de relatórios de portfólio, com a impressão, plotagem e síntese dos resultados encontrados;
4. A *análise* do desempenho do portfólio, incluindo a análise contínua, *backtesting* e *benchmarking*;

Assim, ele contém os processos necessários para resolver problemas de portfólio de média-variância [equações (1) e (2)] ou de média-CVaR [equação (3)], envolvendo tanto a programação quadrática com restrições lineares como a programação linear em grande escala, a programação linear inteira mista e demais problemas relacionados.

3.2.4 Construção da Interface Gráfica do Usuário

A Interface Gráfica do Usuário (GUI) para o sistema de BI/DW proposto para as Ceasas brasileiras foi construída através do pacote *Shiny* (CHANG et al., 2021; WICKHAM, 2021) da linguagem R, seguindo os preceitos vistos na seção 2.4. Ele é um *framework* para a criação de aplicativos para a *web*, que não demanda do programador quaisquer conhecimentos prévios de HTML, CSS ou JavaScript.

Todo aplicativo feito no *Shiny* possui dois componentes-chave: a **UI** (*User Interface*), que define a aparência de seu aplicativo; e a **server function**, que define como seu aplicativo funciona. O *Shiny* usa **programação reativa**, cuja ideia principal é especificar um gráfico de dependências para que, quando uma entrada for alterada, todas as saídas relacionadas sejam atualizadas automaticamente, tornando assim o fluxo de um aplicativo consideravelmente mais simples e compreensível (WICKHAM, 2021).

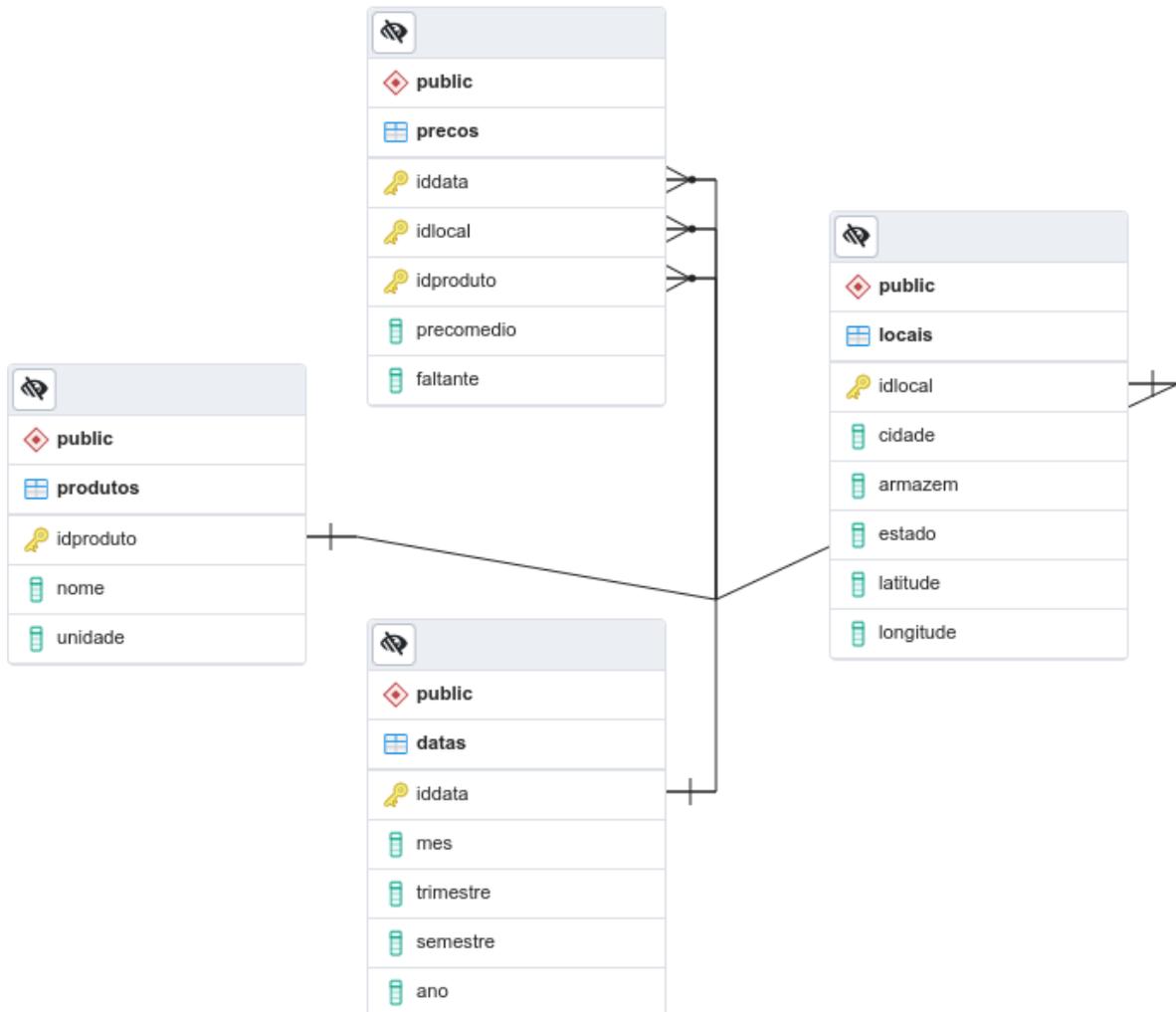
Uma vez desenvolvida no *Shiny*, a GUI foi implantada numa plataforma de autoatendimento em nuvem chamada *shinyapps.io*, que facilita o compartilhamento desses aplicativos, seja para provar alguns conceitos, construir um protótipo, executá-los por um curto período de tempo ou até mesmo apresentá-los como componente principal de ofertas analíticas proprietárias.

4 RESULTADOS

4.1 MODELAGEM DIMENSIONAL DOS DW'S ROLAP e NoSQL

A Figura 7 mostra o projeto lógico para o DW ROLAP proposto para as Ceasas brasileiras, num formato *star* e implantando no Postgresql versão 12, via *Amazon Relational Database Service*. Ele possui uma tabela de fatos (*precos*) e três tabelas de dimensões (*datas*, *locais* e *produtos*). A tabela de fatos *preco* possui duas medidas do tipo não aditivas: o *precomedio*, indicando o preço médio mensal comercializado; e a *faltante*, indicando se aquele preço médio mensal é real ou estimado, para cada *idlocal*, *idproduto*, *iddata* (chave primária composta).

Figura 7 – Diagrama relacional do DW ROLAP



Fonte: Dados do estudo.

O DW ROLAP para as Ceasas brasileiras tem o seguinte esquema de hierarquias para

as tabelas de dimensões:

- datas: (all) ≤ (ano) ≤ (semestre) ≤ (trimestre) ≤ (mes)
- locais: (all) ≤ (armazem) ≤ (cidade) ≤ (latitude, longitude) e (all) ≤ (estado) ≤ (cidade) ≤ (latitude, longitude)
- produtos: (all) ≤ (nome, unidade)

Por fim, o DW ROLAP para as Ceasas brasileiras possui a seguinte tipologia para as suas medidas e atributos:

- *integer*: mes, trimestre, semestre, ano, idlocal, idproduto;
- *character*: iddata, cidade, armazem, estado, nome, unidade, faltante;
- *float*: precomedio, latitude, longitude;

A Figura 8 mostra o projeto lógico para o DW NoSQL proposto para as Ceasas brasileiras, implementado no Apache Cassandra, versão 3.11, num computador local, e que seguiu a abordagem da Desnormalização Lógica, cujas medidas e atributos das tabelas de fatos e de dimensões são identificados pelo mesmo identificador da tabela (*row key*) (DEHDOUH et al., 2015). Nela, são vistas duas *BigFactTables* (*bft_precos1* e *bft_precos2*), cuja *row key* composta é formada por duas partes: a *partition key* (em vermelho) e as *clustering columns* (em ciano).

Figura 8 – Tabelas denormalizadas do DW NoSQL

```

usuario@cqlsh:ceasas> DESCRIBE TABLES;
bft_consulta2 bft_consulta1 bft_precos2 bft_precos1
usuario@cqlsh:ceasas> SELECT * FROM bft_precos1 LIMIT 1;
estado | cidade | nome | iddata | ano | armazem | faltante | mes | precomedio | semestre | trimestre | unidade
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
PE | Caruaru | ABACATE | 2006-01-01 | 2006 | CEASA | Verdadeiro | 1 | 1.1 | 1 | 1 | KG
(1 rows)
usuario@cqlsh:ceasas> SELECT * FROM bft_precos2 LIMIT 1;
nome | iddata | cidade | estado | ano | latitude | longitude | mes | precomedio | unidade
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
PEPINO | 2006-01-01 | Aracaju | SE | 2006 | -10.90879 | -37.06153 | 1 | 1.16 | KG
(1 rows)
usuario@cqlsh:ceasas> SELECT * FROM bft_consulta1 LIMIT 1;
estado | cidade | nome
-----+-----+-----
PE | Caruaru | ABACATE
(1 rows)
usuario@cqlsh:ceasas> SELECT * FROM bft_consulta2 LIMIT 1;
nome | iddata
-----+-----
PEPINO | 2006-01-01
(1 rows)
usuario@cqlsh:ceasas>

```

Fonte: Dados do estudo.

A *partition key* é usada para determinar os nós nos quais as linhas são armazenadas e pode consistir de várias colunas. As *clustering columns* são usadas para controlar como os dados são classificados para o armazenamento em uma *partition key* (CARPENTER; HEWITT, 2016). Tendo em vista que tanto *bft_precos1* como *bft_precos2* foram criadas para

atender às consultas previstas no sistema *web* de BI proposto, o esquema de hierarquias foi assim concebido:

- *bft_precos1*: (all) ≤ (estado) ≤ (cidade) ≤ (nome) ≤ (iddata)
- *bft_precos2*: (all) ≤ (nome) ≤ (iddata) ≤ (cidade) ≤ (estado)

Para contribuir com a rapidez das consultas nas *BigFactTables*, duas tabelas auxiliares foram criadas: *bft_consulta1* e *bft_consulta2*. A principal razão para isto foi seu tamanho final para a execução da cláusula **WHERE**, que, no Apache Cassandra, somente é recomendável para os atributos que compõem a *row key*. Logo, as tabelas *bft_consulta1* e *bft_consulta2* possuem, respectivamente, 2.643 e 9.072 tuplas, as quais atendem todas as possibilidades de consultas previstas para as 499.527 tuplas existentes em ambas *BigFactTables*.

Por fim, o DW NoSQL para as Ceasas brasileiras possui a seguinte tipologia para os seus atributos:

- *integer*: mes, trimestre, semestre, ano;
- *text*: iddata, cidade, armazem, estado, nome, unidade, faltante;
- *float*: precomedio, latitude, longitude;

Vale salientar que, após a modelagem dos DW's ROLAP e NoSQL, os dados foram incluídos utilizando um algoritmo escrito na linguagem R, que realizou a etapa de ETL, considerando também o tratamento de dados perdidos, conforme descrito na seção 3.2.1.

4.2 GUI PARA VISUALIZAÇÃO E ANÁLISE DE PREÇOS

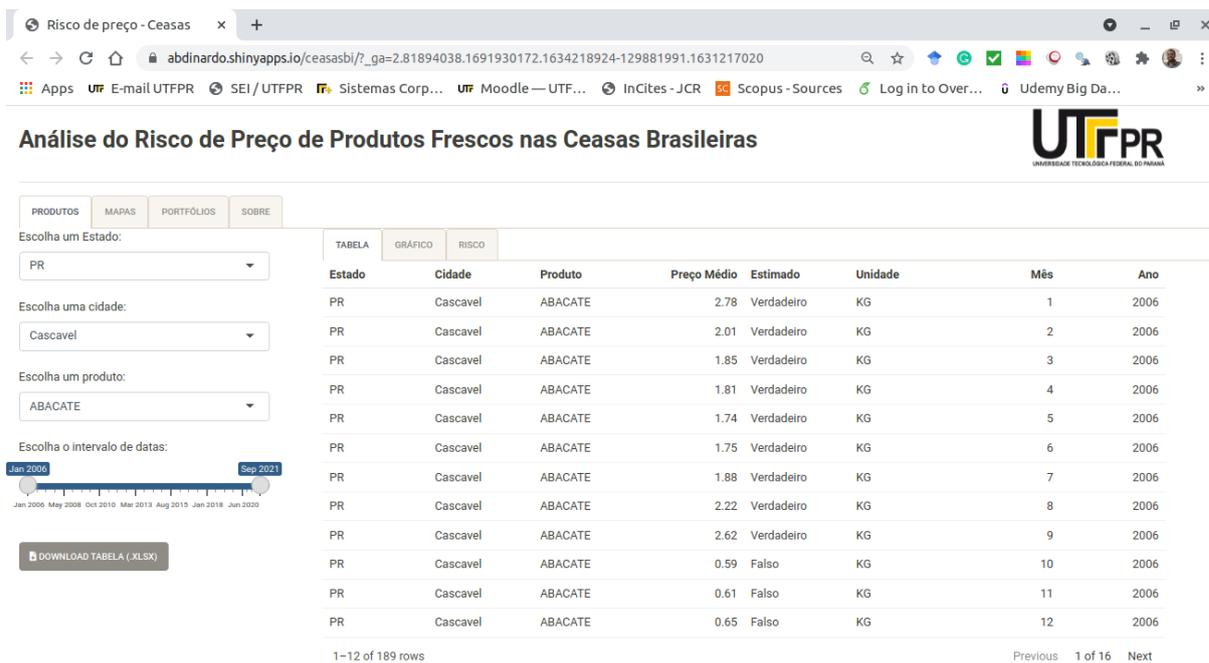
A Figura 9 mostra a tela inicial do sistema *web* de BI para a análise do risco de preço de produtos frescos nas Ceasas brasileiras, com acesso disponibilizado no *shinyapps.io* para o DW relacional. Ela possui quatro abas principais, com suas funcionalidades mostradas a seguir.

Na aba **Produtos**, existe um console à esquerda com as opções para a definição da consulta desejada: uma caixa de seleção única para a escolha do *Estado*, uma caixa de seleção única para a escolha da *cidade*, uma caixa de seleção única para o *produto*, uma barra deslizante para a rápida escolha do intervalo de *datas*, e um botão para *download* dos dados resultantes da consulta especificada, no formato *.xlsx*. À sua direita, aparecem três novas abas, com opções distintas de visualização desses resultados, cuja descrição vem a seguir.

Em **Tabela**, é exibida a tabela decorrente da consulta realizada, mostrando os atributos *Estado*, *Cidade*, *Produto*, *Preço Médio*, *Estimado*, *Unidade*, *Mês*, *Ano*. No canto inferior direito, estão os botões de navegação da tabela, enquanto que no canto inferior esquerdo, são mostradas as linhas exibidas no momento e o total de linhas retornadas pela consulta.

A Figura 10 mostra a saída resultante para a aba **Gráfico**. Nela, um gráfico interativo é exibido com o auxílio do pacote *plotly* (SIEVERT, 2019), no qual é possível fazer operações de *zoom*, *download* da imagem, seleção e comparação de dados *on hover*. Além disso, o título do gráfico traz os atributos *Produto*, *Cidade* e *Estado*, bem como os preços atual, mínimo, máximo, o preço médio do intervalo de *datas* e o percentual de preços estimados na série histórica consultada. Por fim, uma legenda é mostrada no canto superior direito, indicando

Figura 9 – Tela inicial do sistema web de BI para as Ceasas brasileiras



Fonte: Dados do estudo.

se o preço médio mensal, em certa data, é real ou estimado, podendo, inclusive, escolher a visibilidade destes pontos. Quanto à aba **Risco**, sua funcionalidade é detalhada na seção 4.3.

Figura 10 – Saída da consulta para a aba Gráfico

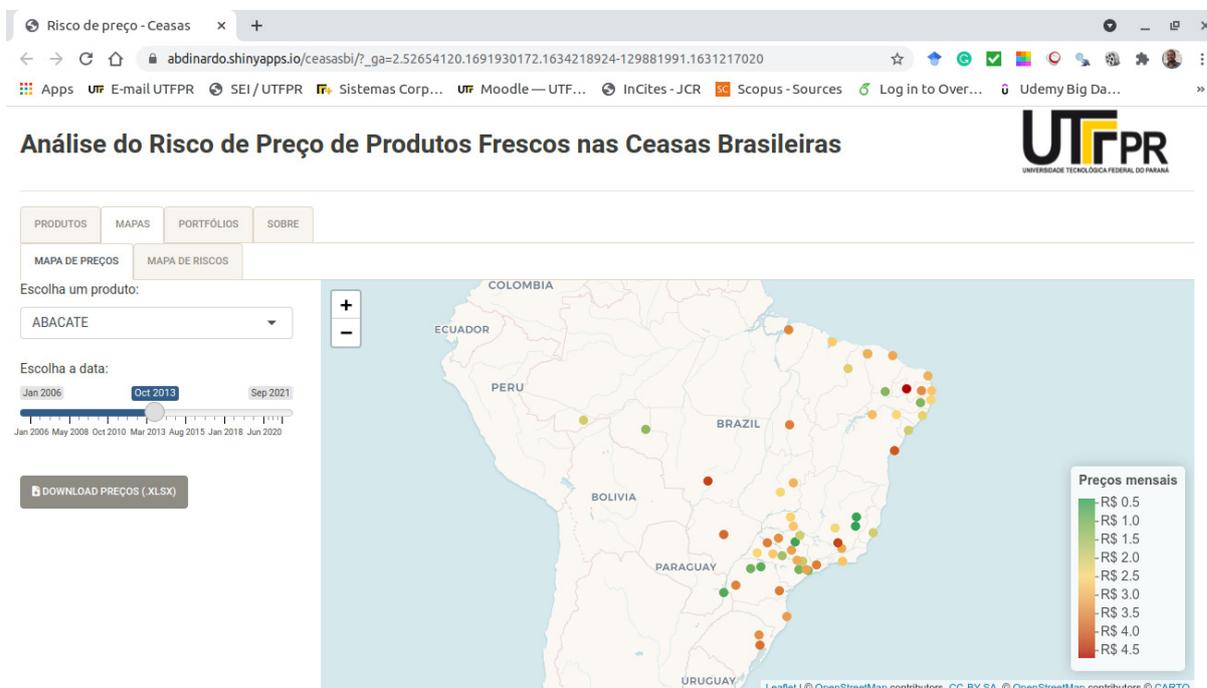


Fonte: Dados do estudo.

A Figura 11 mostra a tela para a aba **Mapas**, que possui duas opções de saída nas

abas **Mapa de preços** e **Mapa de riscos**. Esta última, também tem suas funcionalidades detalhadas na seção 4.3.

Figura 11 – Saída da consulta para a aba *Mapa de preços*



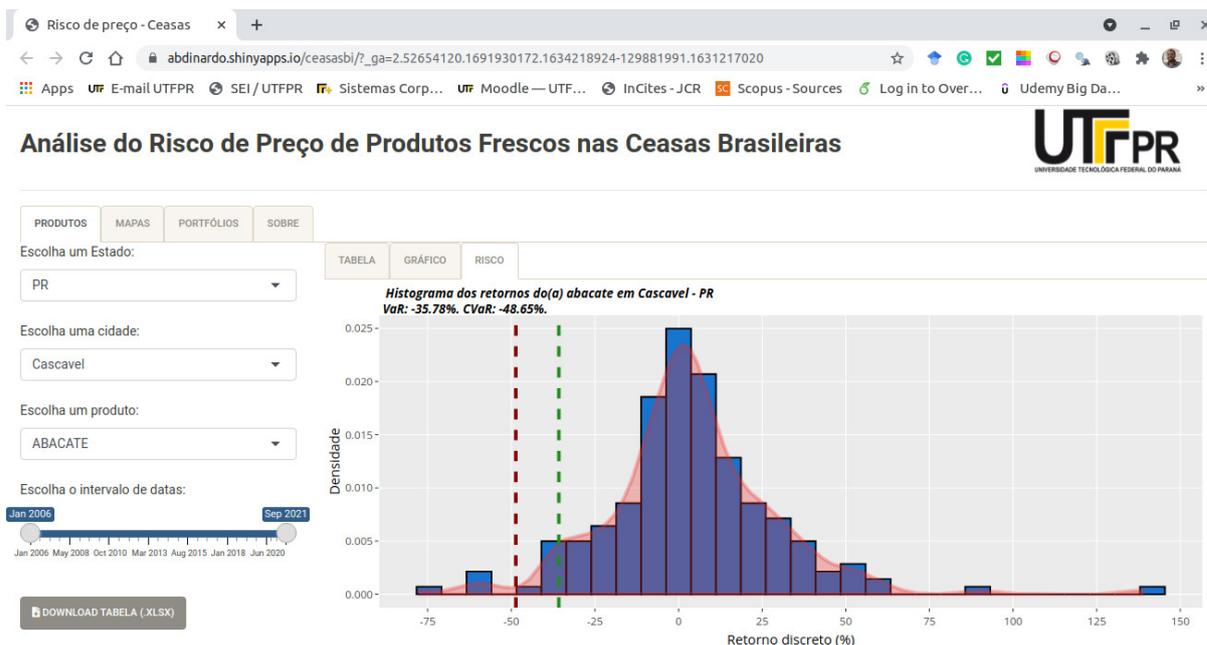
Fonte: Dados do estudo.

Nela, há um console à esquerda com as opções para a definição da consulta desejada: uma caixa de seleção única para a escolha do *produto*, uma barra deslizante para a rápida escolha da *data*, e um botão para *download* dos dados resultantes da consulta especificada, no formato *.xlsx*. À sua direita, aparece um mapa interativo do Brasil, elaborado com o auxílio do pacote *leaflet* (CHENG; KARAMBELKAR; XIE, 2021), mostrando a distribuição espacial do preço mensal do produto selecionado, em certa data, nas Ceasas brasileiras. Ao posicionar o ponteiro do *mouse* num ponto, os atributos *produto*, *unidade*, *cidade*, *estado*, *preço médio* e *data* são mostrados *on hover*. Além disso, os preços mensais estão destacados conforme uma escala contínua de cores, variando entre verde (mais baratos), amarelo (intermediários) e vermelho (mais caros), vistos na legenda à direita.

Finalmente, a aba **Sobre** traz um *link* para um tutorial no Youtube acerca da utilização deste sistema de BI, enquanto que os detalhes para a aba **Portfólios** são tratados a seguir.

4.3 GUI PARA VISUALIZAÇÃO E ANÁLISE DO RISCO DE PREÇOS

A seção 4.2 introduziu as telas iniciais do sistema *web* de BI proposto, com foco na visualização e análise individual de preços dos produtos frescos. Nesta seção, o foco agora é na visualização e análise do risco associado ao *retorno discreto* desses preços, obtidos a partir das métricas vistas na seção 2.3.

Figura 12 – Saída da consulta para a aba *Risco*

Fonte: Dados do estudo.

A Figura 12 mostra a saída para a aba **Risco**, pertencente à aba **Produtos**. Nela, são exibidos um histograma (em azul) e uma curva de densidade de probabilidade (sombreada de vermelho), ambos interativos e construídos com o auxílio do pacote *plotly* (SIEVERT, 2019).

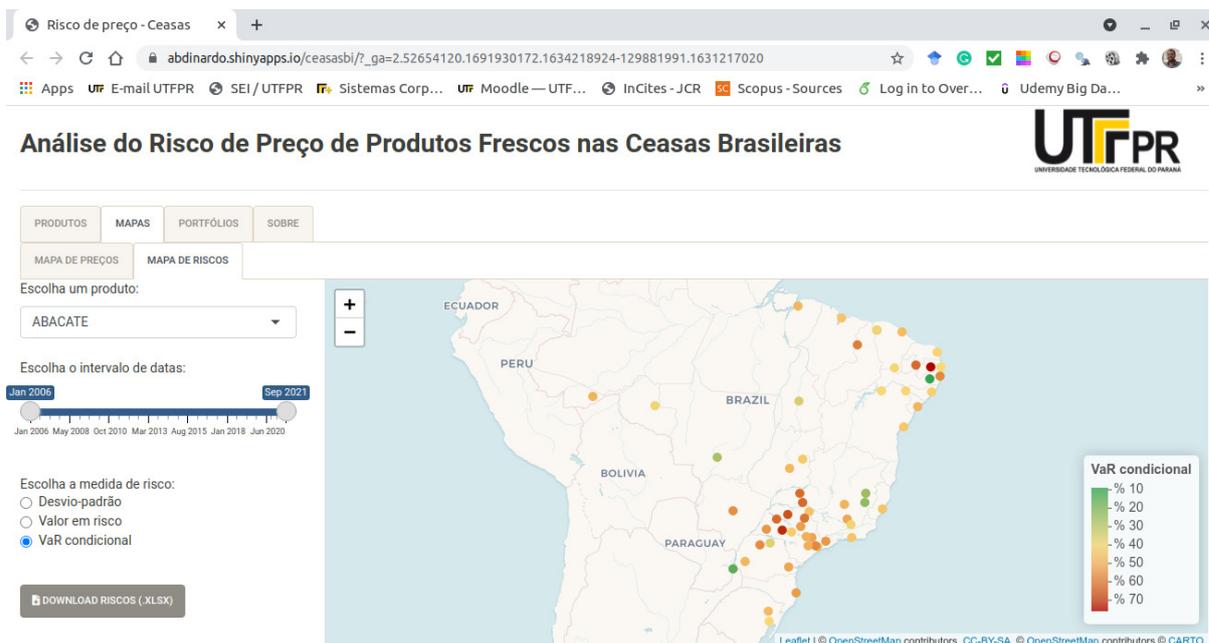
Também são visíveis duas linhas tracejadas verticais, indicando o VaR (em verde escuro) e o CVaR (em vermelho escuro), cujos valores, com probabilidade de ocorrência mensal de 5%, estão dispostos no título do gráfico com os atributos *produto*, *cidade* e *estado*.

A Figura 13 mostra a saída para a aba **Mapa de riscos**, pertencente à aba **Mapas**. Aqui o console à esquerda possui duas modificações em relação ao visto no **Mapa de preços**: é exibida uma barra deslizante para a rápida escolha do intervalo de *datas*, e um painel com botões de opção para a escolha da *medida de risco* a ser visualizada no mapa.

Semelhante ao **Mapa de preços**, à sua direita, aparece um mapa interativo do Brasil, novamente elaborado com o auxílio do pacote *leaflet* (CHENG; KARAMBELKAR; XIE, 2021), expondo a distribuição espacial do risco de preço do produto escolhido, em certo intervalo de datas, nas Ceasas brasileiras. Posicionando o ponteiro do *mouse* em certo ponto, os atributos *produto*, *unidade*, *cidade*, *estado*, *desvio-padrão*, *VaR*, *CVaR*, *data início* e *data fim* são mostrados *on hover*. Ademais, as medidas de risco estão destacadas conforme uma escala contínua de cores, variando entre verde (menos arriscado), amarelo (risco intermediário) e vermelho (mais arriscado), vistos na legenda à direita.

A Figura 14 mostra as abas de consulta para a aba **Portfólios**. À esquerda, está um console com as opções para a definição da consulta pretendida: uma caixa de seleção única para a escolha do *Estado*, uma caixa de seleção única para a escolha da *cidade*, uma caixa de seleção

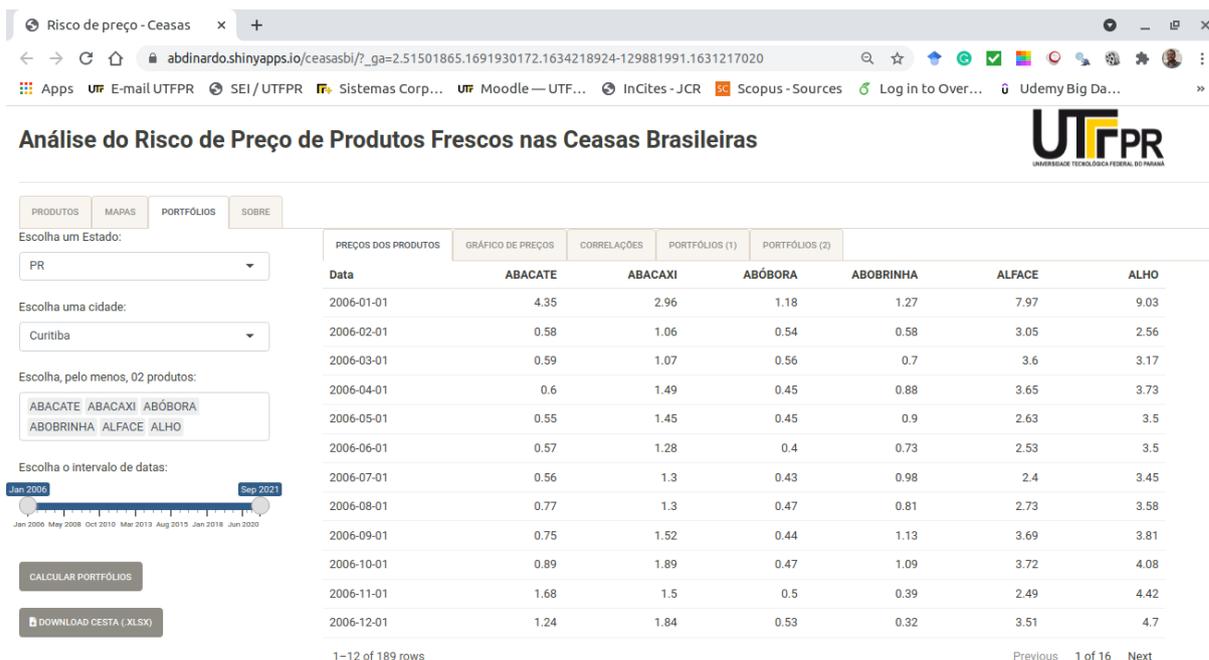
Figura 13 – Saída da consulta para a aba *Mapa de riscos*



Fonte: Dados do estudo.

múltipla para a escolha dos *produtos* (mínimo de dois), uma barra deslizante para a rápida escolha do intervalo de *datas*, e um botão para *download* da cesta de produtos resultante da consulta especificada, no formato *.xlsx*.

Figura 14 – Saída da consulta para a aba *Portfólios*



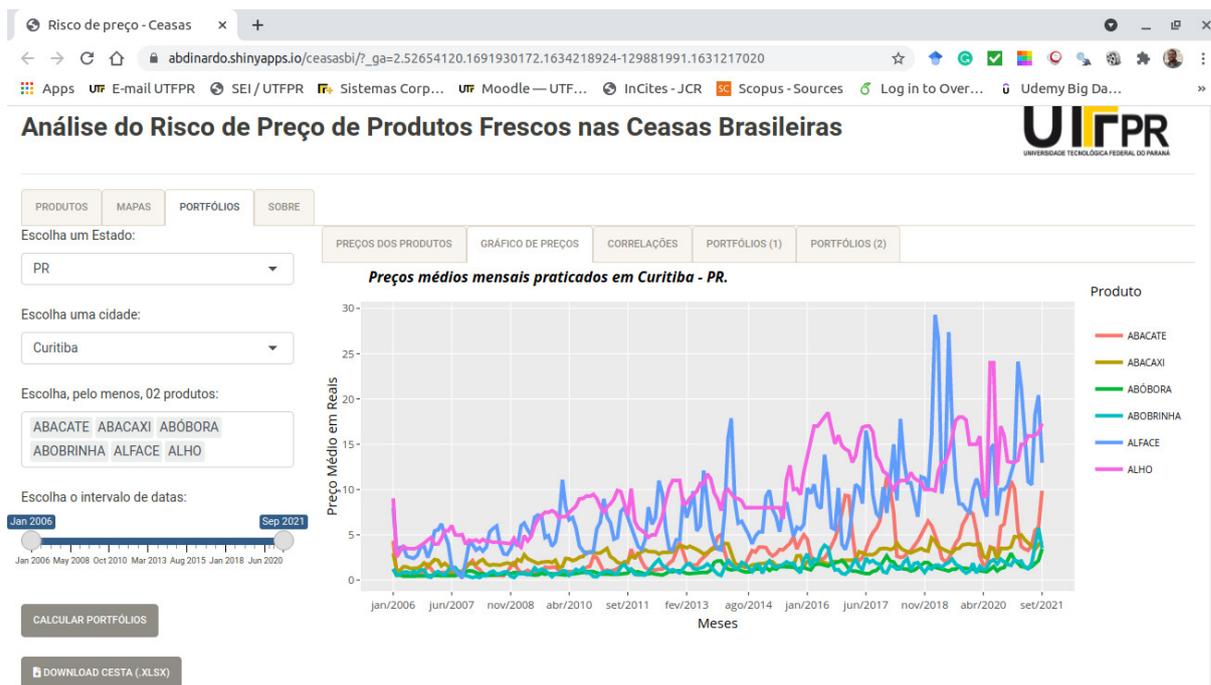
Fonte: Dados do estudo.

Adicionalmente, foi inserido o botão *Calcular Portfólios* para o usuário controlar a programação reativa do *Shiny* em relação a esta ação, evitando assim que o sistema calcule um novo portfólio para cada nova inserção de produto, que por vezes cujo processamento pode ser demorado, a depender do número de produtos selecionados.

Em **Preços dos produtos**, é exibida a tabela decorrente da consulta realizada, mostrando a *Data* e os *Produtos* escolhidos. No canto inferior direito, estão os botões de navegação da tabela, enquanto que no canto inferior esquerdo, estão as linhas exibidas no momento e o total de linhas retornadas pela consulta.

A Figura 15 mostra a saída para a aba **Gráfico de preços**. Nela, um gráfico interativo é exibido com o auxílio do pacote *plotly* (SIEVERT, 2019), no qual é possível fazer operações de *zoom*, *download* da imagem, seleção e comparação de dados *on hover* para as séries históricas de preços médios dos produtos escolhidos, com o título do gráfico trazendo os atributos *Cidade* e *Estado*. Por fim, uma legenda é mostrada no canto superior direito, indicando os produtos mostrados no gráfico, podendo, inclusive, escolher individualmente a sua visibilidade.

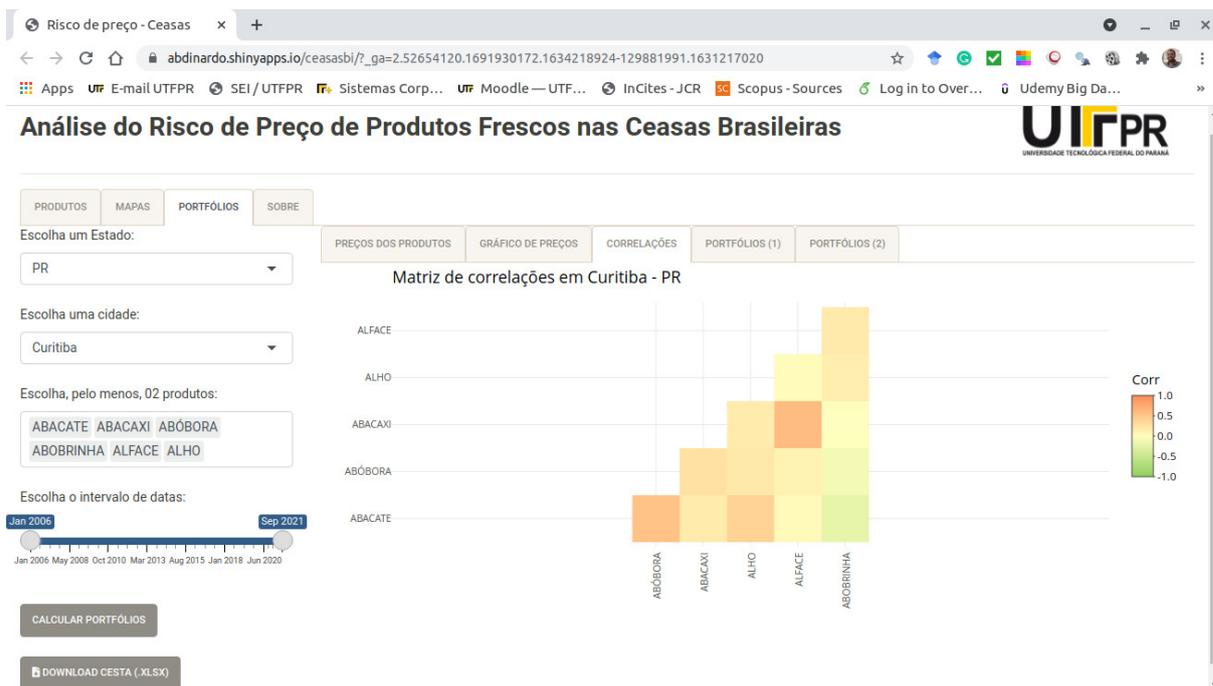
Figura 15 – Saída da consulta para a aba *Gráfico de preços*



Fonte: Dados do estudo.

A Figura 16 mostra a saída para a aba **Correlações**. Nela, é vista a matriz de correlações dos retornos discretos dos produtos selecionados, com uso do pacote *ggcorrplot* (KASSAMBARA, 2019). Posicionando o ponteiro do *mouse* em certo quadrado, o valor da correlação e o par de produtos são mostrados *on hover*, com o título do gráfico trazendo os atributos *Cidade* e *Estado*. Além disso, os valores das correlações estão destacados conforme uma escala contínua de cores, variando entre verde (correlação negativa), amarelo (correlação nula) e vermelho (correlação positiva). Logo, os portfólios com o menor risco possível têm os

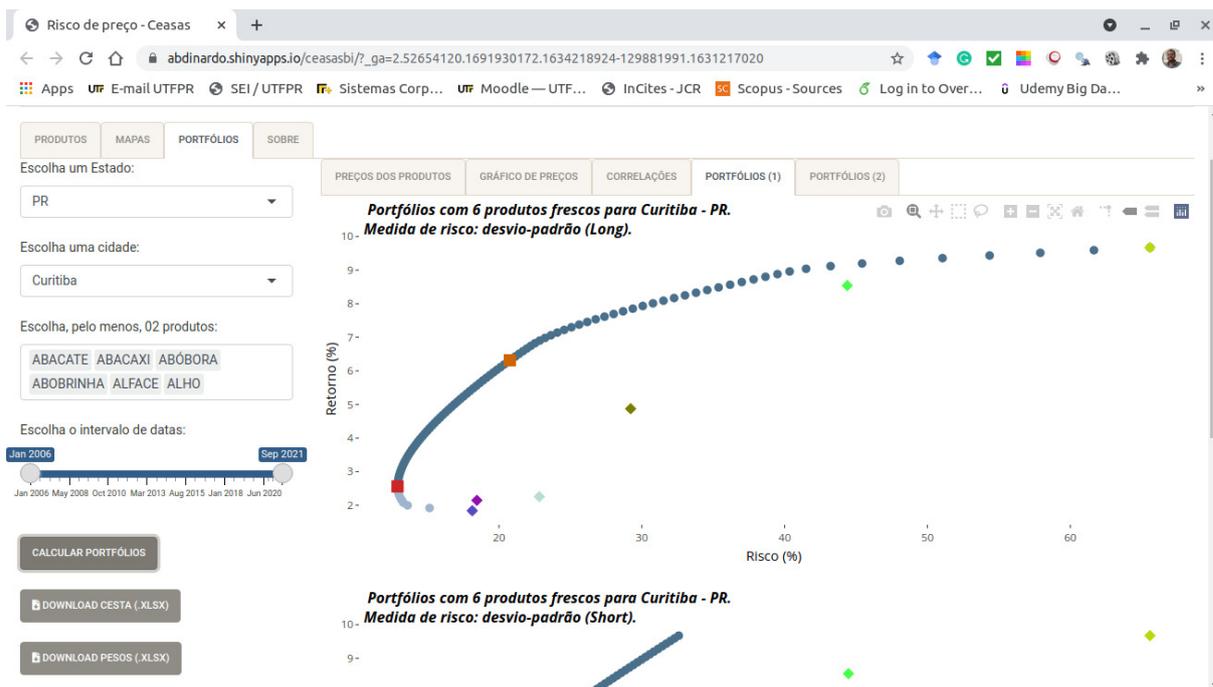
Figura 16 – Saída da consulta para a aba *Correlações*



Fonte: Dados do estudo.

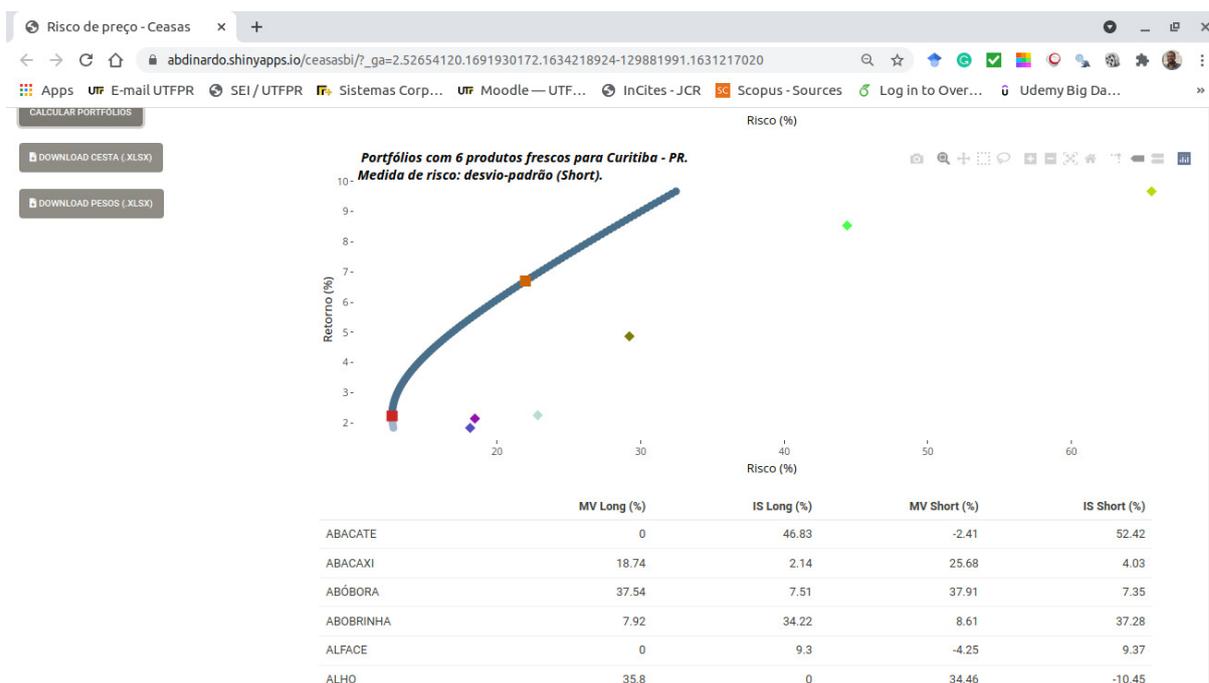
valores das correlações dos retornos discretos dos produtos envolvidos o mais negativo possível, ou seja, bem próximo de -1. Assim, a visualização da matriz de correlações ajuda o usuário a escolher quais produtos devem compor, ou não, o seu portfólio de menor risco desejável.

Figura 17 – Saída da consulta para a aba *Portfólios (1)*



Fonte: Dados do estudo.

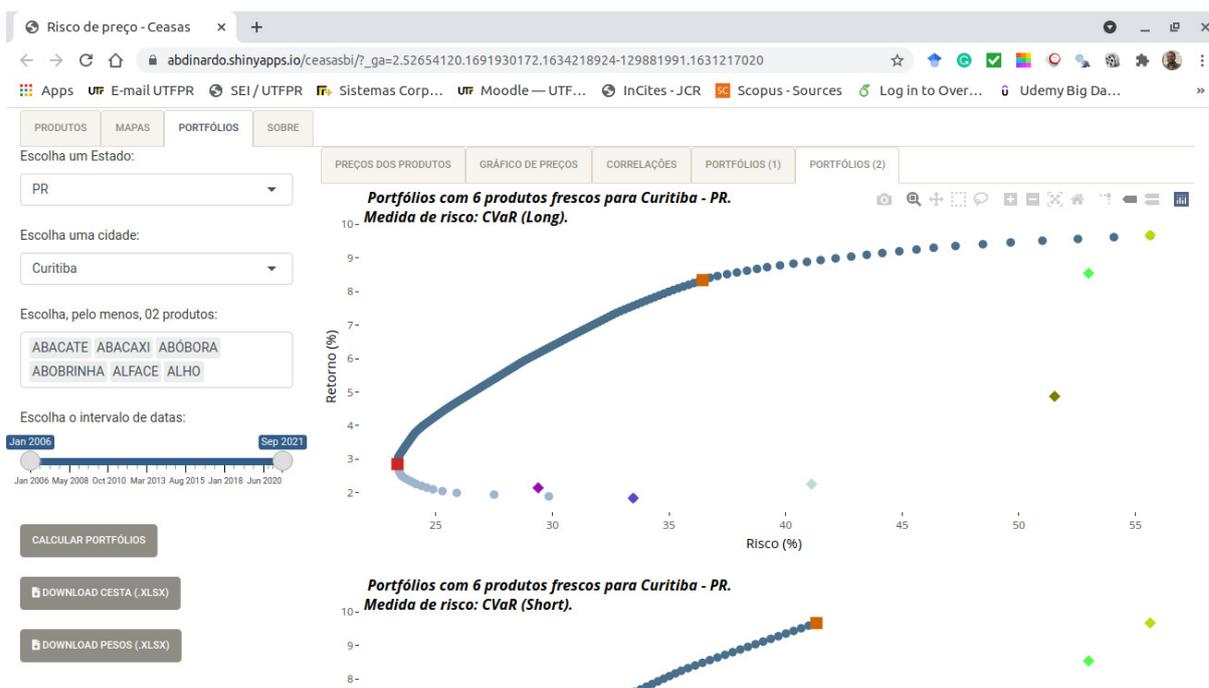
Figura 18 – Saída da consulta para a aba *Portfólios (1)*, continuação



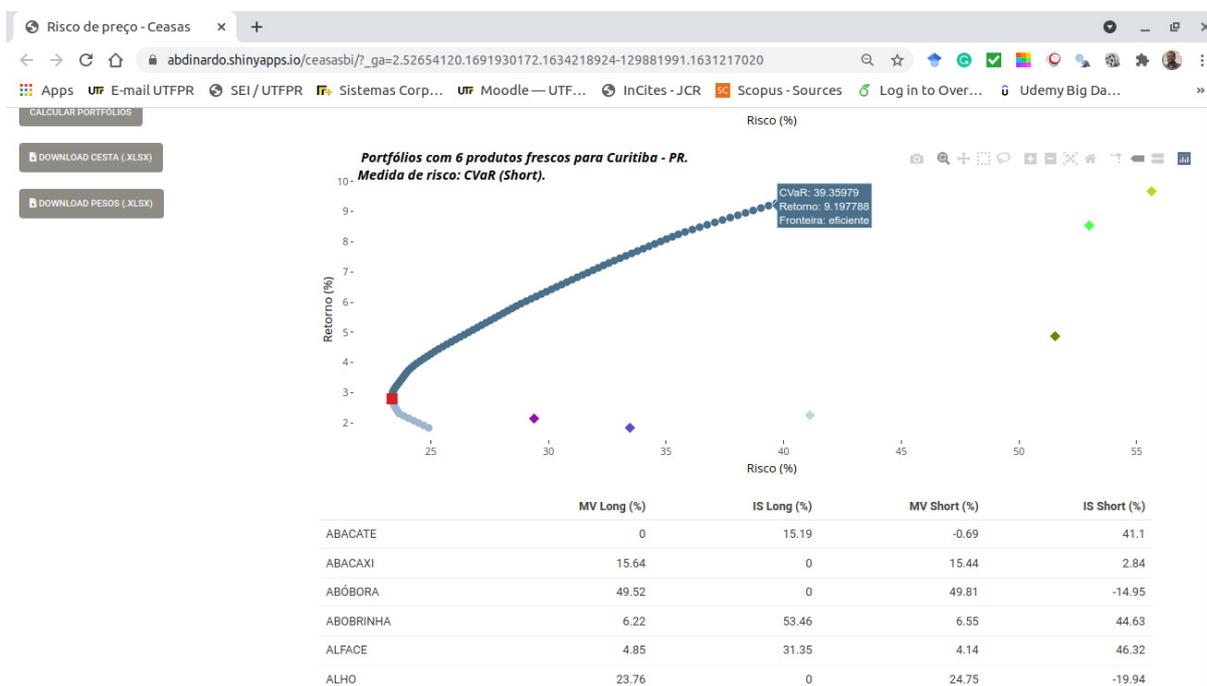
Fonte: Dados do estudo.

As Figuras 17, 18, 19 e 20 mostram as saídas para as abas **Portfólios (1)** e **Portfólios (2)**, que consideram o desvio-padrão e o CVaR como medidas de risco, respectivamente. Elas trazem três saídas: dois gráficos interativos e uma tabela, sendo detalhadas a seguir.

Figura 19 – Saída da consulta para a aba *Portfólios (2)*



Fonte: Dados do estudo.

Figura 20 – Saída da consulta para a aba *Portfólios (2)*, continuação

Fonte: Dados do estudo.

Os gráficos, feitos com o pacote *plotly* (SIEVERT, 2019), exibem as fronteiras eficientes (círculos em azul escuro) e ineficientes (círculos em azul claro) de investimento, os produtos escolhidos (losangos) e uma tabela de pesos, a qual indica a proporção que cada produto tem na composição dos portfólios de variância mínima (MV - quadrado vermelho) e de tangência [máximo retorno/risco] (IS - quadrado laranja) por tipo de operação financeira (*Long ou Short*). Estas saídas são ativadas quando o usuário clica no botão **Calcular Portfólios**, que também habilita o botão **Download Pesos**, gerando um arquivo no formato *.xlsx*.

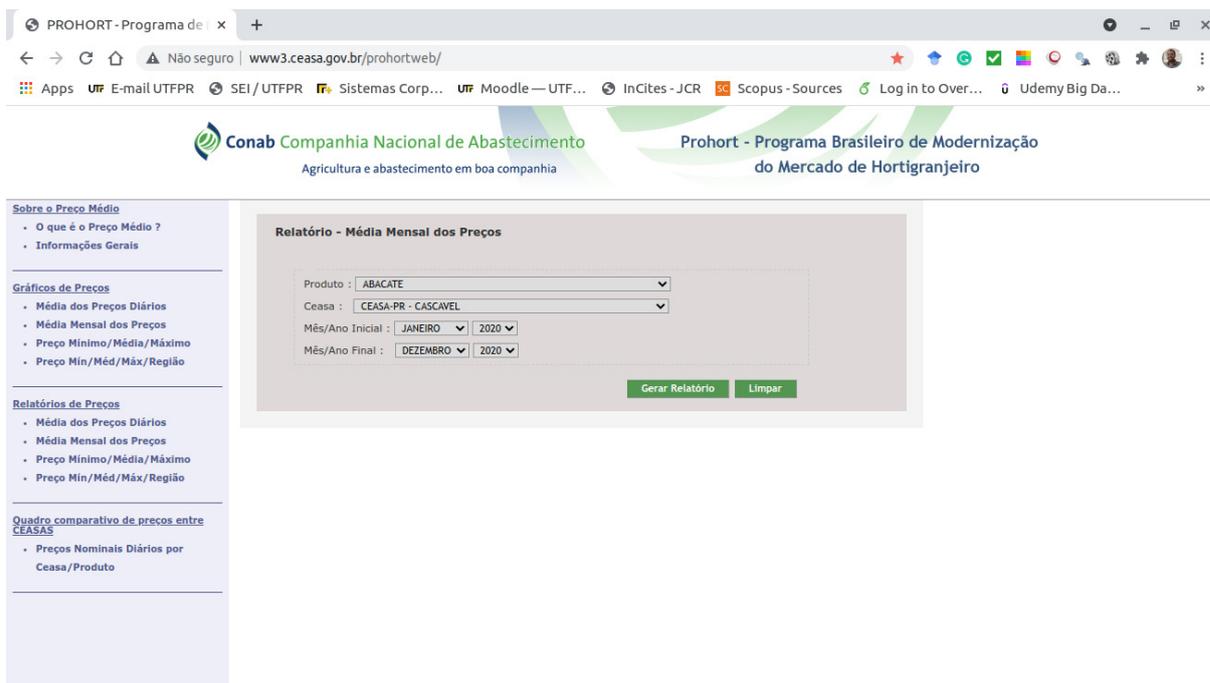
Posicionando o ponteiro do *mouse* em certo ponto, os dados sobre risco, retorno e produto são mostrados *on hover*, bem como se a fronteira é eficiente/ineficiente, com o título do gráfico trazendo a quantidade de produtos selecionados, os atributos *Cidade e Estado*, o tipo de medida de risco e o tipo de operação financeira (*Long ou Short*). Além disso, também é possível fazer operações de *zoom*, *download* da imagem, seleção e comparação de dados.

Por último, é importante ressaltar que o intervalo mínimo entre datas para a realização de todas as consultas referentes ao risco é de 24 meses; do contrário, a mensagem **"Não disponível!"** é exibida em tela. Ademais, todas as telas aqui apresentadas para o DW ROLAP também foram implementadas para o DW NoSQL.

4.4 DISCUSSÃO

Em relação aos aspectos visuais, o sistema *web* de BI proposto para as Ceasas brasileiras segue a lógica do painel de conjunto de guias (*tabset panel*), diferentemente do visto no site

Figura 21 – Site do Prohort



Fonte: (CONAB, 2021b)

do Prohort, que segue a lógica do painel de lista de navegação (*navigation list panel*), como visto na Figura 21.

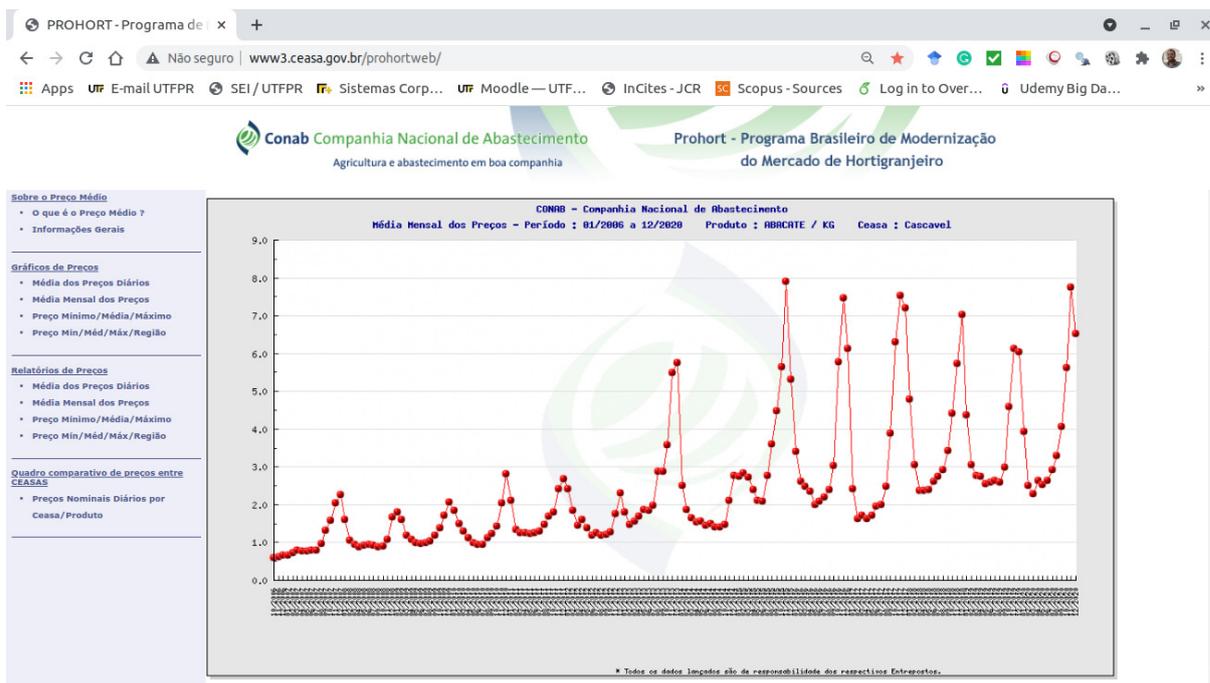
Esta escolha está baseada na hipótese de que os painéis de conjunto de guias são *mais apropriados* para dividir a saída de **uma consulta** em várias seções visualizáveis de forma independente, do que um painel de lista de navegação, cuja função *principal* é fornecer uma lista de *links*, que podem, ou não, apontar para um painel de conjunto de guias.

No site do Prohort, por exemplo, para o usuário obter a tabela e o gráfico de preços mensais, ele precisa realizar a mesma consulta duas vezes, enquanto que no sistema *web* de BI aqui proposto, a consulta só precisa ser realizada uma única vez para se obter os mesmos resultados, cabendo ao usuário tão somente selecionar a saída desejada.

Quanto aos gráficos produzidos, ainda que haja a opção para seleção do tipo de gráfico (barra ou linha), foi verificada a ausência de interatividade destes no site do Prohort, bem como a ausência de formatação adequada no eixo das datas, como visto na Figura 22. Além disso, é importante salientar que o gráfico de barras é mais adequado para visualizar a magnitude de certa quantidade em determinada categoria, sendo o gráfico de linha o *mais indicado* para a visualização de séries temporais (WILKE, 2019), o que é o caso para este tipo de consulta.

Na opção "Quadro comparativo de preços entre CEASAS", embora o site do Prohort permita a consulta dos *atuais* preços diários de vários produtos, para até 10 Ceasas, estes dados são dispostos numa tabela, o que dificulta a realização das comparações desejadas, principalmente no âmbito espacial. Por outro lado, ainda que a consulta seja realizada somente para um produto, o sistema *web* de BI aqui proposto mostra os preços mensais de *todas* as

Figura 22 – Gráfico de linha, resultante de consulta gerada pelo site do Prohort



Fonte: (CONAB, 2021b)

Ceasas, de qualquer período cadastrado, dispostos espacialmente no mapa brasileiro e com uma escala que facilita a sua comparação, conforme já visto na Figura 11.

A respeito dos aspectos analíticos do risco de preço, é visto que o site do Prohort não possui qualquer *link* que trate desse assunto, o qual é suprido pelo sistema *web* de BI aqui proposto, com gráficos interativos e adequadamente selecionados para o tipo de consulta desejada, incluindo a possibilidade de *downloads* das tabelas geradas.

At last, but not at least, é importante comentar sobre a quantidade significativa de dados perdidos identificados nas séries históricas de preços dos produtos frescos nas Ceasas brasileiras, que variou entre 1,51% e 77,25%. Isto evidencia, de certo modo, o grau de compromisso de cada Ceasa para com a atualização dos preços dos produtos frescos no site do Prohort, o que, por sua vez, pode afetar sua credibilidade enquanto fonte de dados para tomada de decisão dos agentes econômicos envolvidos (ver Tabela 2).

Para minimizar esse efeito, o sistema *web* de BI aqui proposto traz uma proposta de estimativa desses preços médios perdidos, sinalizando inclusive para o usuário o *quando* da sua ocorrência.

5 CONCLUSÃO

5.1 LIMITAÇÕES

Os autores deste trabalho reconhecem as seguintes limitações que ocorreram durante o tempo de sua elaboração:

- Não foi possível realizar quaisquer testes de usabilidade com os potenciais usuários do sistema *web* de BI proposto, em função da incerteza de obtenção de uma amostra representativa e colaborativa;
- Ainda que exista um serviço em nuvem equivalente ao Apache Cassandra oferecido pela *Amazon Web Services* (*Amazon Keyspaces*), ele não tem nenhum nível de gratuidade para fins de prototipagem, o que contribuiu para a sua utilização no formato *on-premise*;
- Não foi identificado um pacote funcional e gratuito na linguagem R que permitisse conexão com o Apache Cassandra (AMRUTHNATH, 2018). Esse pacote, nesses termos, foi encontrado para a linguagem Python (DATASTAX, 2021). Assim, foi empregado o pacote *reticulate* (USHEY; ALLAIRE; TANG, 2021) para conectar o R ao Apache Cassandra via Python. Entretanto, esta solução, até onde se sabe, geraria problemas entre a conexão do *shinyapps.io* e o *Amazon keyspaces*, dado que não há na documentação de ambos os serviços orientações de como fazê-la. Logo, esta situação também contribuiu para a utilização do formato *on-premise* para a demonstração do DW NoSQL;

5.2 TRABALHOS FUTUROS

Para a realização de trabalhos acadêmicos e/ou técnico-tecnológicos futuros, o presente trabalho traz as seguintes propostas a serem realizadas:

- Efetuar os testes de usabilidade com os potenciais usuários do sistema *web* de BI proposto;
- Identificar novas técnicas mais efetivas e alternativas à média sazonal para o preenchimento de dados faltantes nas séries de preços mensais dos produtos frescos, com vistas a minimizar o efeito de falsas estimativas;
- Realizar a expansão das opções analíticas do sistema *web* de BI proposto, como, por exemplo, a implementação das funcionalidades disponíveis no pacote *fPortfolio* (WUERTZ et al., 2009; WUERTZ et al., 2020) que permitam ao usuário analisar o desempenho do portfólio selecionado;
- Desenvolver um pacote funcional e gratuito na linguagem R que faça sua conexão com o Apache Cassandra e também com o *Amazon keyspaces*, o que permitiria o acesso às funcionalidades da plataforma de autoatendimento em nuvem do *shinyapps.io*, tal como já acontece com os SGBD's relacionais;

5.3 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo geral desenvolver um sistema *web* de BI para avaliar o risco de preço de produtos frescos em Ceasas brasileiras, tendo em vista a ausência de, até então, quaisquer propostas de implementação de uma ferramenta que lidasse com essa temática. Para alcançá-lo, foi necessário atender a três objetivos específicos, que foram decisivos na construção desta proposta.

Em relação ao objetivo específico 1 (*Elaborar a modelagem dimensional dos Data Warehouses para o sistema web de BI*), foram desenvolvidos e implementados os projetos lógicos para os DW's ROLAP (em nuvem) e NoSQL (*on-premise*), a serem utilizados conforme as necessidades de volume, variedade e velocidade inerentes ao processamento dos dados disponibilizados pelas Ceasas brasileiras.

A respeito do objetivo específico 2 (*Estabelecer as métricas de avaliação do risco de preço para o sistema web de BI*), o desvio-padrão e o CVaR foram definidos como as principais métricas de avaliação tanto para os preços individuais como para as cestas otimizadas de produtos, tendo como base as teorias moderna e pós-moderna do Portfólio.

Sobre o objetivo específico 3 (*Produzir uma aplicação web de suporte à decisão para o sistema de BI*), este foi desenvolvido utilizando o pacote *Shiny*, em conjunto com demais funcionalidades oferecidas pela linguagem R, sendo posteriormente disponibilizado para testes na plataforma de autoatendimento em nuvem *shinyapps.io*.

Em termos teóricos, o presente estudo contribuiu para atender a quase todos os itens elencados na Tabela 1, permanecendo tão somente os testes de usabilidade a serem executados em breve trabalho futuro. Estes, por sua vez, evidenciarão a potencial contribuição prática do sistema *web* de BI aqui proposto na diminuição dos efeitos adversos causados pela assimetria de informação de dados agrícolas agregados, reduzindo assim a incerteza dos agentes econômicos envolvidos, e melhorando, portanto, a eficiência desse mercado.

Referências

- ABRACEN. **Manual Operacional das Ceasas do Brasil**. Belo Horizonte: AD2 Editora, 2011. 122 p. Disponível em: <<http://abracen.org.br/wp-content/uploads/2014/05/manual.pdf>>. Citado 2 vezes nas páginas 17 e 18.
- AKER, J. C.; GHOSH, I.; BURRELL, J. The promise (and pitfalls) of ICT for agriculture initiatives. **Agricultural Economics (United Kingdom)**, v. 47, p. 35–48, 2016. ISSN 15740862. Disponível em: <<http://dx.doi.org/10.1111/agec.12301>>. Citado na página 15.
- AMAZON WEB SERVICES. **Crie e conecte-se a um banco de dados PostgreSQL com o Amazon RDS**. 2021. Disponível em: <<https://aws.amazon.com/pt/getting-started/hands-on/create-connect-postgresql-db/>>. Citado na página 26.
- AMRUTHNATH, N. **Using Cassandra through R**. 2018. Disponível em: <<https://www.r-bloggers.com/2018/12/using-cassandra-through-r/>>. Citado na página 42.
- AYDOGDU, A.; TÖR, O. B.; GÜVEN, A. N. CVaR-based stochastic wind-thermal generation coordination for Turkish electricity market. **Journal of Modern Power Systems and Clean Energy**, v. 7, n. 5, p. 1307–1318, 2019. ISSN 21965420. Disponível em: <<https://doi.org/10.1007/s40565-018-0492-3>>. Citado na página 22.
- AYUPOV, S.; ARIFJANOV, A. Representation and Analysis of Multi-dimensional Data to Support Decision Making in the Management. In: Xu J., Gen M., Hajiyev A., C. F. (Ed.). **Proceedings of the Eleventh International Conference on Management Science and Engineering Management**. Cham: Springer, 2018. v. 2. ISBN 9783319592800. Disponível em: <<https://doi.org/10.1007/978-3-319-59280-0>>. Citado 2 vezes nas páginas 18 e 19.
- AZEVEDO, P. F. Comercialização de produtos agroindustriais. In: BATALHA, M. O. (Ed.). **Gestão Agroindustrial**. 3. ed. São Paulo: Atlas, 1997. cap. 3, p. 49–81. ISBN 9788522445707. Citado na página 12.
- AZZAM, T. et al. Data visualization and evaluation. **New Directions for Evaluation**, v. 2013, n. 139, p. 7–32, 2013. ISSN 1534875X. Disponível em: <<https://doi.org/10.1002/ev.20065>>. Citado 2 vezes nas páginas 23 e 24.
- BARBAGLIA, L.; CROUX, C.; WILMS, I. Volatility spillovers in commodity markets: A large t-vector autoregressive approach. **Energy Economics**, The Author(s), v. 85, p. 104555, 2020. ISSN 01409883. Disponível em: <<https://doi.org/10.1016/j.eneco.2019.104555>>. Citado na página 12.
- BOUSSAHOUA, M.; BOUSSAID, O.; BENTAYEB, F. Logical schema for data warehouse on column-oriented NoSQL databases. In: BENSLIMANE, D. et al. (Ed.). **Database and Expert Systems Applications**. Cham: Springer, 2017. v. 10439, p. 247–256. ISBN 9783319644707. Disponível em: <https://doi.org/10.1007/978-3-319-64471-4_20>. Citado 2 vezes nas páginas 20 e 27.
- BRUM, L. M. L.; LAMPERT, V. N.; CAMARGO, S. S. Business Intelligence and Data Warehouse in Agrarian Sector: A Bibliometric Study. **Journal of Agricultural Science**, v. 11, n. 2, p. 353, 2019. ISSN 1916-9752. Disponível em: <<http://dx.doi.org/10.5539/jas.v11n2p353>>. Citado na página 12.

- CÂMARA, G. M. S. Fenologia é ferramenta auxiliar. **Visão Agrícola**, v. 3, n. 5, p. 63–66, 2006. Disponível em: <<https://www.esalq.usp.br/visaoagricola/sites/default/files/va05-planta-e-ambiente01.pdf>>. Citado na página 14.
- CARPENTER, J.; HEWITT, E. **Cassandra: The Definitive Guide**. 2nd. ed. Boston, MA: O'Reilly Media, Inc., 2016. 343 p. Disponível em: <<https://www.oreilly.com/library/view/cassandra-the-definitive/9781491933657/>>. Citado 2 vezes nas páginas 27 e 30.
- CHANG, W. et al. **shiny: Web Application Framework for R**. 2021. Disponível em: <<https://cran.r-project.org/package=shiny>>. Citado na página 28.
- CHATURVEDI, K.; RAI, A. On-line Analytical Processing in Agriculture using Multidimensional Cubes. **J. Ind. Soc. Agril. . . .**, v. 62, n. 1, p. 56–64, 2008. Disponível em: <<http://js.iasri.res.in/jsp/volume/vol62/Apr-08/Apr-08-7.pdf>>. Citado 2 vezes nas páginas 12 e 14.
- CHENG, J.; KARAMBELKAR, B.; XIE, Y. **leaflet: Create interactive web maps with the JavaScript 'Leaflet' library**. 2021. Disponível em: <<https://cran.r-project.org/package=leaflet>>. Citado 2 vezes nas páginas 33 e 34.
- COBLE, K. H. et al. Big data in agriculture: A challenge for the future. **Applied Economic Perspectives and Policy**, v. 40, n. 1, p. 79–96, 2018. ISSN 20405804. Disponível em: <<http://dx.doi.org/10.1093/aep/ppx056>>. Citado na página 15.
- CODD, E.; CODD, S.; SALLEY, C. Providing OLAP (on-line analytical processing) to user-analysts. **An IT mandate. Technical report**, v. 230, p. 1–20, 1993. Disponível em: <http://www.estgv.ipv.pt/PaginasPessoais/jloureiro/ESI_AID2007_2008/fichas/codd.pdf>. Citado na página 18.
- CONAB. **Comercialização total de frutas e hortaliças**. Brasília, 2021. 21 p. Disponível em: <<https://www.conab.gov.br/info-agro/hortigranjeiros-prohort>>. Citado 2 vezes nas páginas 15 e 25.
- CONAB. **Prohort - Programa brasileiro de modernização do mercado de hortigranjeiro**. 2021. Disponível em: <<http://www3.ceasa.gov.br/prohortweb/>>. Citado 3 vezes nas páginas 25, 40 e 41.
- DAHL, R. E.; OGLEND, A.; YAHYA, M. Dynamics of volatility spillover in commodity markets: Linking crude oil to agriculture. **Journal of Commodity Markets**, v. 20, n. December 2020, p. 100111, 2020. ISSN 24058513. Disponível em: <<https://doi.org/10.1016/j.jcomm.2019.100111>>. Citado na página 12.
- DAS, S. et al. Portfolio Optimization with Mental Accounts. **Journal of Financial and Quantitative Analysis**, v. 45, n. 2, p. 311–334, apr 2010. ISSN 0022-1090. Disponível em: <https://www.cambridge.org/core/product/identifiser/S0022109010000141/type/journal_article>. Citado na página 21.
- DATASTAX. **DataStax Python driver for Apache Cassandra**. 2021. Disponível em: <<https://docs.datastax.com/en/developer/python-driver/3.25/>>. Citado na página 42.
- DB-ENGINES. **DB-Engines ranking - trend popularity**. 2021. Disponível em: <https://db-engines.com/en/ranking_trend>. Citado na página 27.

DEHDOUH, K. et al. Using the column oriented NoSQL model for implementing big data warehouses. In: **International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'15)**. Las Vegas: Worldcomp, 2015. p. 469–475. Disponível em: <<http://worldcomp-proceedings.com/proc/p2015/PDP6237.pdf>>. Citado 3 vezes nas páginas 20, 27 e 30.

ELTON, E. J. et al. **Moderna Teoria de Carteiras e Análise de Investimentos**. São Paulo: Atlas, 2004. 602 p. Citado na página 21.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37–54, 1996. Disponível em: <<https://doi.org/10.1609/aimag.v17i3.1230>>. Citado na página 23.

GILBERT, S.; LYNCH, N. Brewer's Conjecture and the Feasibility of Consistent, Available, Partition-Tolerant Web Services. **ACM SIGACT News**, v. 33, n. 2, p. 51–59, 2002. Disponível em: <<https://doi.org/10.1145/564585.564601>>. Citado na página 18.

GUPTA, A. K.; MAZUMDAR, B. D. Multidimensional Schema for Agricultural Data Warehouse. **International Journal of Research in Engineering and Technology**, v. 02, n. 03, p. 245–253, 2013. ISSN 23217308. Disponível em: <<http://dx.doi.org/10.15623/ijret.2013.0203006>>. Citado 2 vezes nas páginas 12 e 14.

HAERDER, T.; REUTER, A. Principles of transaction-oriented database recovery. **ACM Computing Surveys (CSUR)**, v. 15, n. 4, p. 287–317, 1983. ISSN 15577341. Disponível em: <<https://doi.org/10.1145/289.291>>. Citado na página 18.

JORION, P. **Value at Risk : The New Benchmark for Managing Financial Risk**. 3. ed. New York: McGraw-Hill, 2002. 1–602 p. Disponível em: <<https://link.springer.com/article/10.1007/s11408-007-0057-3>>. Citado na página 22.

KASSAMBARA, A. **ggcorrplot: Visualization of a Correlation Matrix using ggplot2**. 2019. Disponível em: <<https://cran.r-project.org/package=ggcorrplot>>. Citado na página 36.

MARKOWITZ, H. Portfolio Selection. **The Journal of Finance**, v. 7, n. 1, p. 77–91, 1952. ISSN 15406261. Disponível em: <<https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>>. Citado na página 21.

MARKOWITZ, H. Foundations of Portfolio Theory. **The Journal of Finance**, v. 46, n. 2, p. 469–477, 1991. ISSN 0022-1082. Disponível em: <<https://doi.org/10.1111/j.1540-6261.1991.tb02669.x>>. Citado na página 21.

MARQUESONE, R. **Big Data: técnicas e tecnologias para extração de valor dos dados**. São Paulo: Casa do Código, 2016. 245 p. ISBN 978-85-5519-231-9. Disponível em: <<https://www.casadocodigo.com.br/products/livro-big-data>>. Citado na página 20.

NGO, V. M.; LE-KHAC, N. A.; KECHADI, M. T. Designing and implementing data warehouse for agricultural big data. In: Chen K., Seshadri S., Z. L. (Ed.). **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. Springer International Publishing, 2019. v. 11514 LNCS, cap. 1, p. 1–17. ISBN 9783030235505. Disponível em: <<http://dx.doi.org/10.1007/978-3-030-23551-2>>. Citado 4 vezes nas páginas 12, 13, 14 e 15.

NILAKANTA, S.; SCHEIBE, K.; RAI, A. Dimensional issues in agricultural data warehouse designs. **Computers and Electronics in Agriculture**, v. 60, n. 2, p. 263–278, 2008. ISSN 01681699. Disponível em: <<http://dx.doi.org/10.1016/j.compag.2007.09.009>>. Citado 3 vezes nas páginas 12, 13 e 14.

NIZAR, N. M. M. et al. Underutilised crops database for supporting agricultural diversification. **Computers and Electronics in Agriculture**, Elsevier B.V., v. 180, n. April 2020, p. 105920, 2021. ISSN 01681699. Disponível em: <<https://doi.org/10.1016/j.compag.2020.105920>>. Citado 3 vezes nas páginas 12, 14 e 15.

OLIVEIRA, A. M. B. et al. Um panorama sobre o risco de preço de produtos frescos em Ceasas brasileiras: o caso do Estado do Paraná. In: **58^o Congresso da Sociedade Brasileira de Economia, Administração e Sociologia Rural (SOBER)**. Foz do Iguaçu-PR: UNIOESTE, 2020. p. 1–19. ISBN 978-65-88243-90-9. Disponível em: <<https://www.even3.com.br/anais/sober2020/>>. Citado 2 vezes nas páginas 13 e 23.

OLIVEIRA, A. M. B.; SANTOS, J. F. O Uso do filme “a Fraude” para o ensino de finanças comportamentais em cursos de Administração. In: **XV Semead - Desafios da Gestão: Econômico, Social e Ambiental**. São Paulo: FEA - USP, 2012. p. 1–12. ISSN 2177-3866. Disponível em: <<http://sistema.semead.com.br/15semead/resultado/trabalhosPDF/189.pdf>>. Citado na página 13.

PFLUG, G. C. Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk. In: **Nonconvex Optimization and its Applications**. Boston: Springer, 2000. p. 272–281. Disponível em: <<https://link.springer.com/chapter/10.1007/978-1-4757-3150-7>>. Citado na página 22.

QUEIROZ, T. A. N. As Ceasas No Contexto Da Reestruturação Do Território Brasileiro. **Revista Geotemas**, v. 8, n. 2, p. 59–78, 2018. Disponível em: <<http://dx.doi.org/10.33237/geotemas.v8i2.3175>>. Citado na página 15.

ROCKAFELLAR, R. T.; URYASEV, S. Optimization of conditional value-at-risk. **The Journal of Risk**, v. 2, n. 3, p. 21–41, 2000. ISSN 14651211. Disponível em: <<https://doi.org/10.21314/jor.2000.038>>. Citado na página 22.

ROM, B. M.; FERGUSON, K. W. Post-Modern Portfolio Theory Comes of Age. **The Journal of Investing**, v. 3, n. 3, p. 11–17, 1994. ISSN 1068-0896. Disponível em: <<https://doi.org/10.3905/joi.3.3.11>>. Citado na página 21.

ROMAN, D.; MITRA, G. Portfolio selection models: a review and new directions. **Wilmott Journal**, v. 1, n. 2, p. 69–85, 2009. ISSN 17596351. Disponível em: <<https://doi.org/10.1002/wilj.4>>. Citado na página 22.

SELLAMI, A.; NABLI, A.; GARGOURI, F. Transformation of Data Warehouse Schema to NoSQL Graph Data Base. In: ABRAHAM, A. et al. (Ed.). **Advances in Intelligent Systems and Computing**. Cham: Springer, 2020. v. 941, p. 410–420. ISBN 9783030166595. Disponível em: <https://doi.org/10.1007/978-3-030-16660-1_41>. Citado na página 20.

SHIFF, L. **The 2020 Gartner magic quadrant for cloud infrastructure and platform services**. 2020. Disponível em: <<https://www.bmc.com/blogs/gartner-magic-quadrant-cips-cloud-infrastructure-platform-services/#>>. Citado na página 26.

- SHRIVASTAVA, S.; PAL, S. N.; WALIA, R. Market intelligence for agricultural commodities using forecasting and deep learning techniques. In: MADRIA, S. et al. (Ed.). **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. 1. ed. Cham: Springer, 2019. v. 11932 LNCS, cap. 12, p. 193–205. ISBN 9783030371876. Disponível em: <<https://doi.org/10.1007/978-3-030-37188-3>>. Citado 3 vezes nas páginas 13, 14 e 15.
- SIEVERT, C. **Interactive web-based data visualization with R, plotly, and shiny**. 1st. ed. Boca Raton, FL: Chapman and Hall/CRC, 2019. 470 p. Disponível em: <<https://plotly-r.com/>>. Citado 4 vezes nas páginas 31, 34, 36 e 39.
- SIMMONS, G. **Investment science**. Oxford: Oxford University Press, 1998. v. 36. 419–420 p. ISSN 00251747. Citado na página 21.
- TVERSKY, A.; KAHNEMAN, D. Judgment under uncertainty: Heuristics and biases. **Science**, v. 185, n. 4157, p. 1124–1131, 1974. ISSN 00368075. Disponível em: <<http://dx.doi.org/10.1126/science.185.4157.1124>>. Citado na página 13.
- UNWIN, A. Why is Data Visualization Important? What is Important in Data Visualization? **Harvard Data Science Review**, v. 2, n. 1, p. 1–7, 2020. Disponível em: <<https://hdrs.duquduq.org/pub/zok97i7p>>. Citado 2 vezes nas páginas 23 e 24.
- USHEY, K.; ALLAIRE, J. J.; TANG, Y. **reticulate: Interface to Python**. 2021. Disponível em: <<https://cran.r-project.org/package=reticulate>>. Citado na página 42.
- VAISMAN, A.; ZIMÁNYI, E. **Data Warehouse Systems: design and implemetantion**. Berlin: Springer, 2014. 625 p. ISBN 9783642546549. Disponível em: <<https://doi.org/10.1007/978-3-642-54655-6>>. Citado 3 vezes nas páginas 18, 19 e 26.
- WANG, C.; ZHAO, J.; HUANG, M. Measurement of the fluctuation risk of the China fruit market price based on VaR. **Agriculture and Agricultural Science Procedia**, v. 1, p. 212–218, 2010. ISSN 22107843. Disponível em: <<http://dx.doi.org/10.1016/j.aaspro.2010.09.026>>. Citado 2 vezes nas páginas 12 e 13.
- WICKHAM, H. **Mastering Shiny**. 1st. ed. Sebastopol, CA: O'Reilly Media, Inc., 2021. 372 p. Disponível em: <<https://mastering-shiny.org/index.html>>. Citado na página 28.
- WILKE, C. O. **Fundamentals of data visualization: a primer on-making informative and compelling figures**. 1st. ed. Sebastopol, CA: O'Reilly Media, Inc., 2019. 390 p. Disponível em: <<https://www.oreilly.com/library/view/fundamentals-of-data/9781492031079/>>. Citado 2 vezes nas páginas 24 e 40.
- WUERTZ, D. et al. **Portfolio Optimization with R/Rmetrics**. Zurich: Rmetrics Association and Finance Online Publishing, 2009. 455 p. Disponível em: <<https://www.rmetrics.org/ebooks-portfolio>>. Citado 2 vezes nas páginas 27 e 42.
- WUERTZ, D. et al. **fPortfolio: Rmetrics - Portfolio Selection and Optimization**. 2020. Disponível em: <<https://cran.r-project.org/package=fPortfolio>>. Citado 2 vezes nas páginas 27 e 42.