

Avaliação de Séries Temporais de Produtos da Cadeia Produtiva do Agronegócio Fluminense

Pedro Vieira Cruz¹, Brena Geliane Fernalda¹, Sergio Manuel Serra da Cruz^{2,3}

¹ Escola Superior de Agricultura Luiz de Queiroz – Universidade de São Paulo (USP)
Piracicaba – SP – Brasil

² Departamento de Computação - Universidade Federal Rural do Rio de Janeiro
(UFRRJ) – Seropédica, RJ – Brasil

³ Programa de Educação Tutorial PET-SI - Universidade Federal Rural do Rio de Janeiro
(UFRRJ) – Seropédica, RJ – Brasil

pedrovieira.br@gmail.com, serra@ufrrj.br

Abstract. *Brazil stands out as one of the world's leading food producers, encompassing everything from agribusiness commodities to basic items from family farming. To understand the intricate food markets, analysts turn to vast datasets, exploring the growth and evolution of consumption and revenue. This article develops multidimensional models and time series analyses on the commercialization of agricultural products at CEASA-Rio, covering the period from 2005 to 2021. As an architectural data management model, we opted for the data lakehouse associated with the analysis tool, Power BI, due to its efficiency and low cost. We conducted data treatments and detailed historical analyses of prices and weights of the main products at CEASA-Rio during the period, as well as developed dynamic and georeferenced dashboards.*

Resumo. *O Brasil destaca-se como um dos principais produtores de alimentos do mundo, englobando desde commodities do agronegócio até itens da cesta básica provenientes da agricultura familiar. Para compreender os intricados mercados alimentícios, os analistas recorrem a vastos conjuntos de dados, explorando o crescimento e a evolução do consumo e faturamento. Este artigo elabora modelos multidimensionais e análises de séries temporais sobre a comercialização de produtos agropecuários na CEASA-Rio, abrangendo o período de 2005 a 2021. Como modelo arquitetural de gestão de dados optamos pelo Data Lakehouse associado com ferramenta de análise, Power BI devido à sua eficiência e baixo custo. Realizamos tratamentos de dados e análises de longas históricas detalhadas dos preços e pesos dos produtos na CEASA-Rio durante o período mencionado, além de desenvolvermos dashboards dinâmicos e georreferenciados.*

1. Introdução

A produção agrícola no Brasil representa um amplo conjunto de atividades que abrangem todos os elos da cadeia produtiva do agronegócio. Segundo Cunha e Belik (2012), as relações entre o atacado de alimentos e os elos a montante e a jusante têm sido objeto de estudos em um contexto mais amplo. Nos países desenvolvidos, grande

parte da função do atacado está desaparecendo devido ao encurtamento da cadeia produtiva e à redução das margens de comercialização. Por outro lado, as atividades logísticas emergem com cada vez mais força, absorvendo parte das margens que antes eram apropriadas por outros setores produtivos. No entanto, no Brasil o amplo uso de dados históricos para apoiar tais estudos ainda é limitado devido aos problemas relacionadas a baixa a qualidade de dados, sua dispersão em silos de dados heterogêneos e ausência de metadados descritivos (Cruz, 2023).

De acordo com a Embrapa (2022), o Brasil destaca-se como um dos maiores produtores de alimentos do mundo com ampla variedade de produção que vai desde commodities do agronegócio até alimentos da cesta básica produzidos pela agricultura familiar. Segundo os indicadores oficiais da CONAB (2022), CEASA-Rio (2022) e EMATER-RIO (2022), alguns produtos têm maior destaque na produção e comercialização fluminense, tais como aipim, banana, quiabo, entre outros.

As CEASA são a principal instituição de comercialização de hortifrutigranjeiros no território fluminense, com 48 anos de existência (CEASA-Rio, 2022). A central é a segunda maior do Brasil e está localizada na Zona Norte da cidade do Rio de Janeiro, atualmente conta com outros cinco mercados situados em São Gonçalo, na Região Metropolitana; Nova Friburgo, na Região Serrana; Paty do Alferes, no Vale do Paraíba; São José de Ubá e Itaocara, no Norte e Noroeste do estado, respectivamente.

Durante o ano de 2022, entre os meses de janeiro e julho, a CEASA-Rio comercializou quase 1 milhão e 700 mil toneladas de alimentos, gerando um volume de movimentação financeira de R\$ 3,6 bilhões. Apesar da forte presença no Estado e da elevada movimentação financeira, o processo de coleta de dados na CEASA ainda enfrenta problemas relacionados à falta de padronização e à ausência de metadados. (Silva, 2021 e Cruz, 2023).

Para compreender as dinâmicas dos mercados, é fundamental unir longas séries de dados históricos de diversas fontes para produzir conjuntos de dados de qualidade antes de analisá-los ou mesmo oferecê-los às tomadoras de decisão. A partir de novos conjuntos de dados, é possível estimar o crescimento e a evolução das curvas de consume, produção e faturamento. Para esse fim, diferentes técnicas podem ser utilizadas. Tradicionalmente, os analistas recorrem ao uso de planilhas eletrônicas e gráficos simples. No entanto, esses recursos são de uso individualizado, não padronizados e frequentemente utilizam dados de qualidade insatisfatória.

A cadeia produtiva do agronegócio Fluminense necessita explorar grandes conjuntos de dados de diferentes fontes. Um padrão arquitetônico que vem emergindo nos últimos anos é o *Data Lakehouse* (Inmon et al, 2021). Ele combina baixo custo, escalabilidade e flexibilidade de um repositório na nuvem com as amplas possibilidades analíticas da estrutura de um data warehouse.

Data Lakehouses ainda são pouco estudados no contexto da cadeia do agronegócio brasileiro e global (Madera et al., 2017), mas vêm se destacando na capacidade de gerenciar a complexidade dos dados dos ecossistemas agrícolas e centralizar todas as fontes de dados para encontrar novas correlações e análises de dados.

Neste trabalho, apresentamos uma plataforma de Data Lakehouse baseada no Power BI capaz de analisar dinamicamente longas séries de dados da cadeia produtiva do agronegócio do estado do Rio de Janeiro. Os resultados das análises indicam a vocação agrícola em muitos municípios do estado do Rio de Janeiro. A plataforma também ingere e trata grandes volumes de dados, agregando qualidade aos dados e ampliando as possibilidades analíticas através de dashboards interativos com dados curados e georreferenciados.

O presente trabalho está organizado da seguinte forma: além desta introdução, a seção 2 apresenta os materiais e métodos adotados na pesquisa. A seção 3 descreve os conjuntos de dados e a modelagem multidimensional do Data Lakehouse. A seção 4 ilustra a capacidade analítica da solução e, por fim, a seção 5 apresenta as conclusões e trabalhos futuros.

2. Materiais e Métodos

A elaboração deste trabalho utilizou séries históricas de dados brutos produzidos pela CEASA-Rio e disponibilizados por Silva (2021). Os conjuntos de dados brutos foram tratados para atender ao formato aberto com 3 estrelas (segundo a classificação de Berner-Lee (2024)) e estão disponíveis em um repositório no Silva (2021). Esses conjuntos de dados contêm informações sobre mais de 100 tipos de produtos comercializados pela CEASA-Rio durante os anos de 2005 a 2021, incluindo seus respectivos preços e quantidades (em quilogramas) no estado do Rio de Janeiro.

Para o tratamento dos dados, foi adotada a metodologia proposta por Ferrari e Russo (2016), utilizando o Microsoft Power BI em um ambiente de data lake na nuvem da AWS. Essa ferramenta é amplamente difundida no agronegócio e é adequada para Data Lakehouses (ASPIN, 2018; POWER BI, 2022).

O Power BI limpa, enriquece e processa dados usando um repositório escalonável em nuvem para *Business Intelligence* corporativo. Além disso, o Power BI oferece a possibilidade de conectar diferentes tipos de fontes de dados, como bancos de dados SQL (locais ou em nuvem), APIs, serviços de nuvem, páginas da web e arquivos em diversos formatos (.csv, .xls, .json, .xml).

A transformação dos dados brutos foi necessária devido a problemas de falta de padronização, como diferentes grafias nos quantitativos de unidades e na nomenclatura dos municípios do estado do Rio de Janeiro, especialmente entre os anos de 2005 e 2010. Como parte do tratamento de qualidade dos dados, foram verificados valores ausentes, duplicados, inconsistências e dados considerados outliers, entre outros.

Complementarmente, foi adicionada uma coluna com o ano de cada arquivo para tratar os dados dos produtos de acordo com o ano e correlacionar a evolução ao longo do tempo. Essas informações foram relacionadas para realizar a junção das tabelas no programa Power BI, sendo necessário ter uma coluna com o ano para realizar análises gráficas subsequentes.

Em seguida, utilizou-se o mapa territorial do estado do Rio de Janeiro com seus municípios, através dos dados e arquivos disponibilizados pelo IBGE (2022), para a construção do mapa de distribuição geográfica dos produtos agrícolas. Neste artigo, foram escolhidas as 10 culturas agrícolas de maior relevância para a produção e consumo no estado do Rio de Janeiro, de acordo com o estudo da EMATER (2022), a

saber: abacaxi, abobrinha, alface, caqui, cenoura, espinafre, milho verde, pepino, quiabo e repolho.

3. Descrição dos Datasets

Foram ingeridos 46 conjuntos de arquivos de dados brutos do tipo .CSV referentes aos anos de 1976 a 2021 contendo as 10 culturas de maior relevância para o estado do Rio de Janeiro (Silva, 2021). Aos dados, além dos tratamentos, foram inseridos atributos identificadores únicos, os resultados desse processo de normalização dos dados possibilitaram a construção indicadores capazes de sistematizar e analisar as informações históricas (Figura 1).

Destaca-se que os arquivos texto não permitiriam análises inteligentes de forma prática para a extração de informações referentes a comercialização de produtos pois estes estavam em unidades distintas (Kg, maço, molho, penca, cacho, caixa, etc). Cada produto possui um peso em Kg distinto. Logo, padronização em termos de massa em Kg é essencial nos processos analíticos. A desorganização dos dados brutos ocorreu devido à falta de padronização e coleta manuais dos dados brutos ao longo do tempo na CEASA-Rio.

A Figura 1 ilustra um diagrama estrela, criado no Data Lakehouse indicando os relacionamentos entre as fontes de dados, segundo a metodologia de Ferrari e Russo (2016). Para a construção do indicador, foram relacionadas às tabelas dimensões, as quais contém os valores únicos para criar as ligações como chaves primárias ao conjunto de dados fixados nas tabelas fato com isso, realizando as ligações de 1 para muitos, com isso sendo possível criar as ligações e a construção de dashboards dinâmicos e as análises descritas na próxima seção.

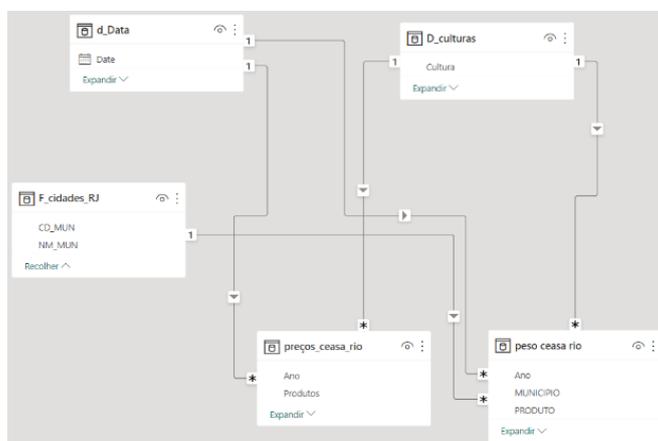


Figure 1. Diagrama estrela ilustrando o relacionamento entre as dimensões do indicador. Fonte Autores.

4. Análises Dinâmicas de Dados no Data Lakehouse

A construção de indicadores dinâmicos em ambiente de Data Lakehouse proporcionou agilidade ao processo de análise de dados com baixo custo. Segundo CEASA (2021), os principais produtos comercializados em 2021 correspondem à categoria dos hortigranjeiros com cerca de 85% dos produtos comercializados, destacando-se principalmente as frutas, hortaliças e ovos. É importante ressaltar que a

comercialização de produtos agrícolas no Rio de Janeiro constitui importante movimentação econômica, ocupando diversas etapas da cadeia produtiva.

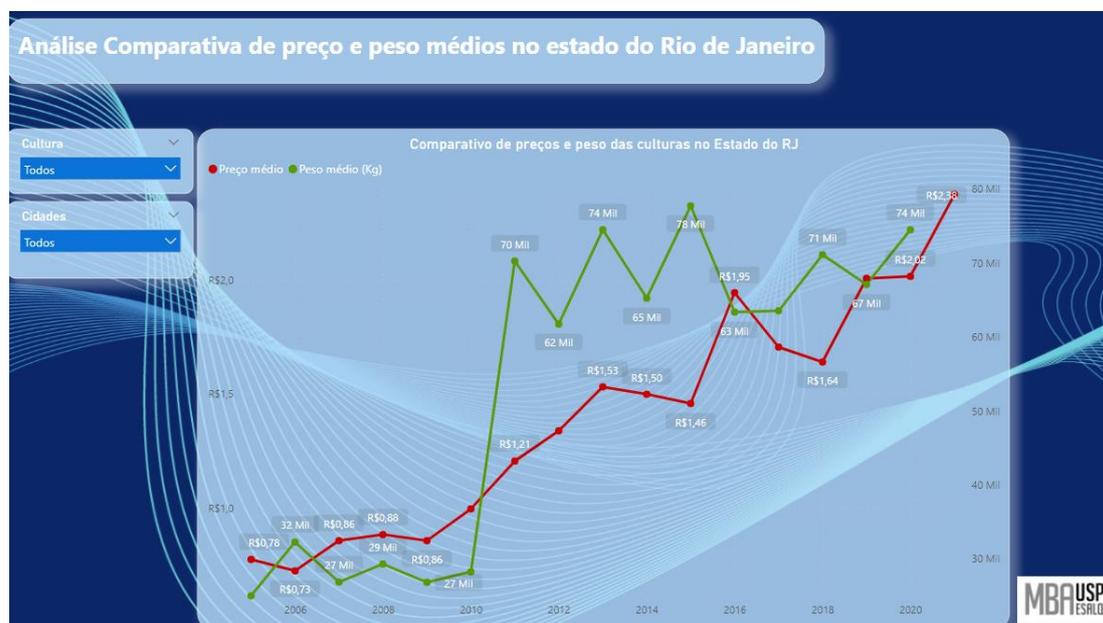


Figura 2 - Análise comparativa de preços e pesos médios dos principais produtos agrícolas produzidos e comercializados no estado do Rio de Janeiro entre os anos de 2005 e 2021. Fonte Autores.

Ao avaliar as séries históricas observou-se aumento considerável ao longo dos anos, no preço médio dos alimentos comercializado no CEASA-Rio a partir do ano de 2009. Nos anos posteriores observou-se um aumento nos valores assim como a quantidade média de peso dos produtos comercializados (Figura 2).

Verificou-se que ocorreu um aumento de produtos comercializados e produzidos no estado do Rio de Janeiro a partir do ano de 2011 quando comparado ao ano anterior, onde se observava um valor de aproximadamente 27 mil kg para 70 mil kg em 2011 representando um aumento de 350% na quantidade de produtos comercializados. No mesmo período também se notou o aumento do preço médio desses produtos, passando de R\$1,00/kg em 2010 para 1,21/kg em 2011, representando um aumento de 121 %.

Os aumentos tornam-se ainda mais expressivos quando se realiza a comparação do preço médio por kg em 2005, no valor de R\$0,78/kg, com o valor de R\$2,30/kg em 2021. Neste caso observou-se um aumento de 294,87 % em todo o período analisado. Algumas hipóteses para esse grande aumento podem ser relacionadas com o aumento dos custos de produção, por ser um estado onde possui pequenas áreas de produção encarecendo suas atividades assim destacado em CEPEA (2020).

Ainda dentro das funcionalidades do indicador, elaboram-se mapas com os dados para aumentar as possibilidades de análises a serem realizadas a partir desta ferramenta. A Figura 3 exibe um mapa dinâmico da distribuição dos produtos agrícolas no estado onde está representado cada município produtor e o total de produtos agrícolas produzidos e comercializados no CEASA-Rio. Neste dashboard do indicador, é possível utilizar de forma dinâmica os filtros como: anos de comercialização, o mês e a cultura de interesse.

Desta forma realizou-se uma análise comparativa com a realizada por Emater (2020) (Figura 4) onde foram apresentados os municípios destacados em função de seus faturamentos. Ao comparar as Figuras 3 e 4, através do uso do indicador, de forma rápida e eficiente podemos ter acesso a dados quantitativos no mapa exibido na Figura 2 em contraposição a dados estáticos e qualitativos exibidos na Figura 3, em relação aos principais produtos produzidos pelos municípios do estado.

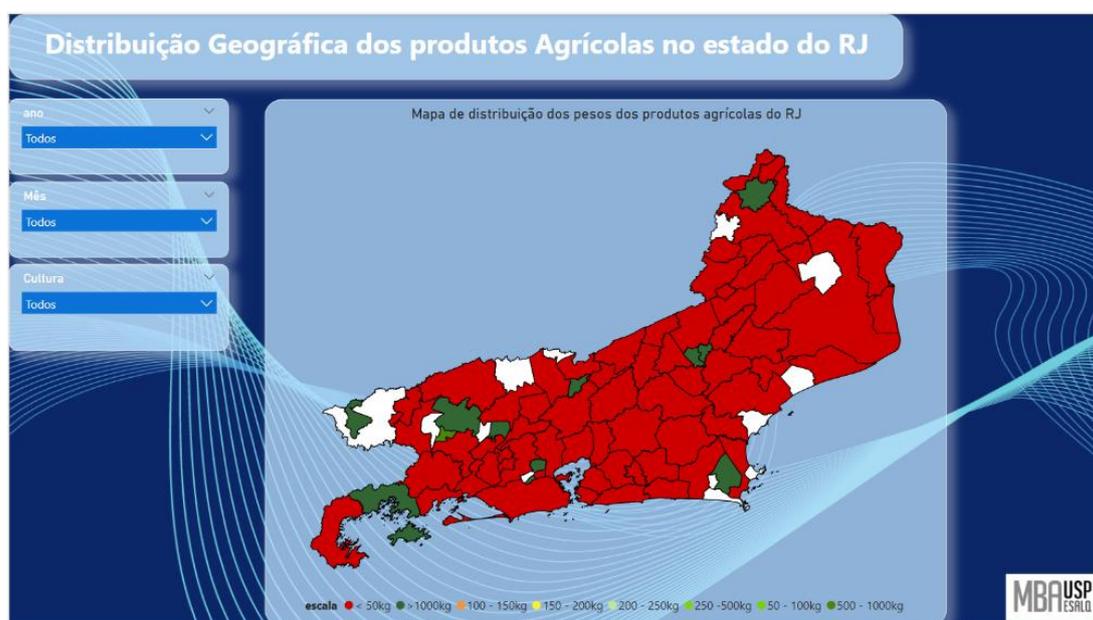


Figura 3 – Mapa elaborado através do indicador exibindo a distribuição geográfica dos produtos Agrícolas no Estado do RJ. Fonte Autores.

A Figura 5 ilustra a distribuição dos preços comercializados no CEASA-Rio. Há dois tipos principais de análises dinâmicas. Na primeira, a análise é temporal, é possível demonstrar a variação dos preços médios ao longo dos anos de 2005 a 2021, como destacado anteriormente, na variação total de 294,87 % de aumento dos preços praticados na central de abastecimento.

Na parte inferior da Figura 5 é possível observar a segunda análise dinâmica, através das comparações entre os preços médios de cada cultura avaliada no presente trabalho, onde a cultura que apresentou o maior preço médio no período analisado foi a cultura do quiabo e o menor preço foi apresentado pela cultura do milho verde.

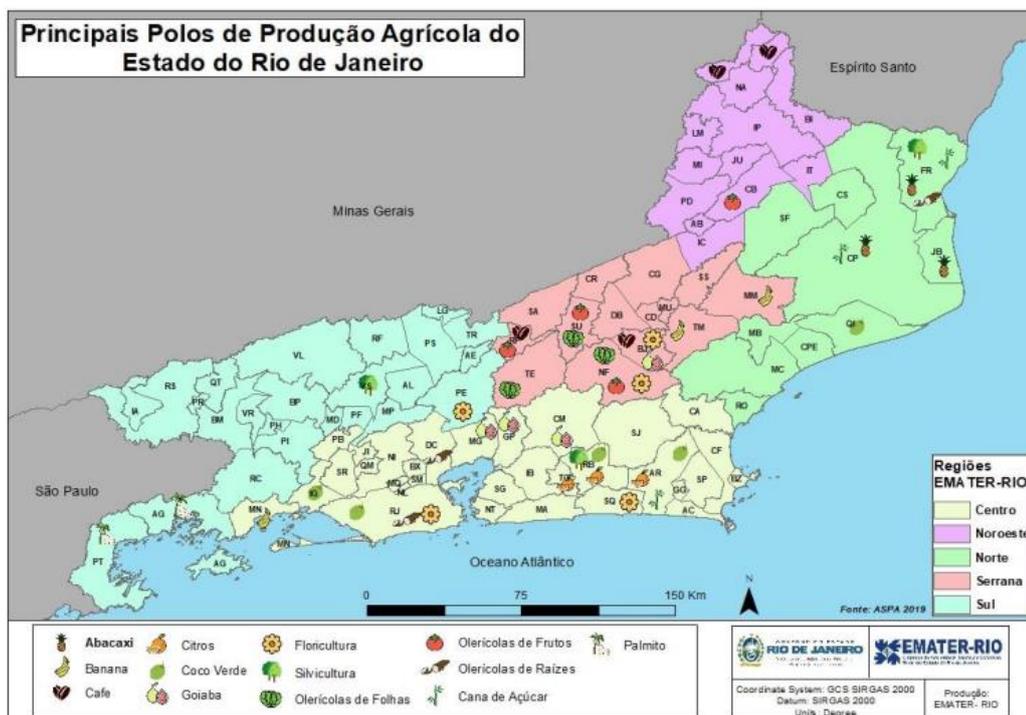


Figura 4 - Principais polos de produção agrícola no estado do RJ. Fonte: Emater (2020).

Disponibilizou-se um dashboard dinâmico com a avaliação dos pesos médios em kg dos produtos agropecuários no estado do Rio de Janeiro, onde cada peso demonstra o total comercializado do produto com o seu centro de origem em cada município do estado (Figura 6).



Figura 5 - Distribuição dos preços médios comercializados no CEASA-Rio. Fonte Autores.

Em uma primeira análise ao longo de 2005 a 2021, na parte superior da Figura, é possível identificar quais são os 10 principais municípios produtores das diferentes culturas, destacando-se o município de São Francisco de Itabapoana como o principal produtor do estado do Rio de Janeiro tanto das culturas supracitadas quanto em relação à sua comercialização no CEASA-Rio ao longo dos anos analisados. Na parte inferior da Figura, pode-se observar a evolução média dos pesos dos diferentes produtos comercializados ao longo da série histórica.

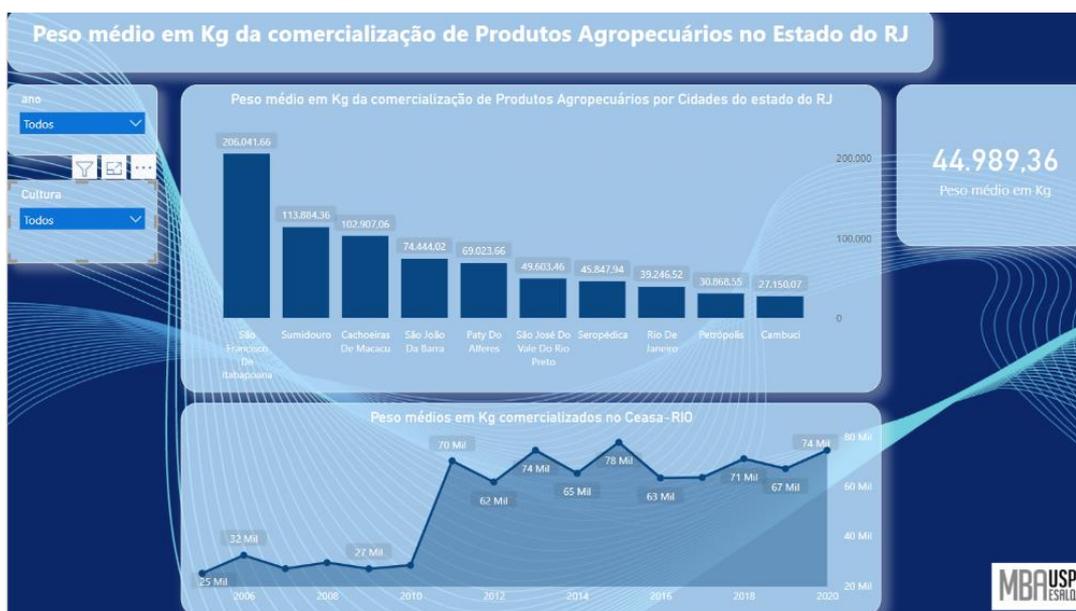


Figura 6 - Peso médio em Kg da comercialização dos produtos no estado do RJ. Fonte Autores.

5. Conclusão

Neste trabalho, a plataforma de Data Lakehouse permitiu combinar os benefícios de um Data Warehouse e dos repositórios de dados brutos e curados na nuvem. O Data Lakehouse combina o baixo custo, a escalabilidade e a flexibilidade de um repositório na nuvem com o gerenciamento e a estrutura de dados de um Data Warehouse.

Do ponto de vista investigativo, o presente trabalho reforça, através da análise de séries de dados históricos, que há uma vocação agrícola em muitos municípios do estado do Rio de Janeiro, fruto do fortalecimento da agricultura familiar através de diferentes frentes de ação. Os volumes de produtos comercializados pela principal central de abastecimento do estado demonstram a importância dos produtos agrícolas na movimentação da economia do estado do Rio de Janeiro.

Verificou-se que ainda existem poucos trabalhos na literatura científica que correlacionam Data Lakehouses e as cadeias produtivas do agronegócio no Brasil. Através das análises, foi possível destacar que o PowerBI se constituiu uma forma dinâmica e eficiente de apresentação dos dados, favorecendo a comparação temporal de diferentes parâmetros, o que pode ser de grande utilidade para gestão e planejamento dos setores produtivos e de comercialização de produtos agrícolas

No presente trabalho poderá ser usado em políticas públicas que auxiliam a sociedade a terem melhor localidade dos centros de distribuição e a diminuição das distâncias entre produtor e mercado consumidor. Centralizando os produtos e assim facilitando o consumo e comércio dos produtos agrícolas, criando formas de conexão e deixando a cada vez mais interligado e otimizado o processo da cadeia produtiva.

A plataforma permitiu de forma eficiente as análises históricas de preço comercializado e quantidade produzida ao longo dos anos de 2005 a 2021 na CEASA Rio. Como trabalhos futuros se pretende ampliar a plataforma e ingerir as novas series de dados do estado do Rio de Janeiro e possivelmente de outros estados do Brasil. Se pretende se interconectar o Data Lakehouse com os dados pedológicos das plataformas OpenSoils (Cruz et al., 2018) e FAIRCHAIN (Marinho et al, 2022) através da sua rede Blockchain.

Agradecimentos

Os autores agradecem à CAPES (código 001), ao CNPq (processos 306115/2021-2 e 400044/2023-4) e ao Fundo Nacional para o Desenvolvimento da Educação/PET-SI-UFRRJ.

Referencias

- Aspin, A. Pro Power BI Desktop. 2. ed. Staffordshire, Reino unido: Apress, 2018. 761 p. ISBN 978-1-4842-3210-1
- Centrais de Abastecimento do Estado do Rio de Janeiro [CEASA-RIO]. (2021). Relatório anual de comercialização - 2021. Disponível em: < https://www.ceasa.rj.gov.br/sites/site_ceasa/files/arquivos_paginas/Relato%CC%81rio%20de%20Comercializac%CC%A7a%CC%83o%20Anual%20referente%20ao%20ano%20de%202021_0.pdf >. Acesso em: 31 mar. 2024.
- Centrais de Abastecimento do Estado do Rio de Janeiro [CEASA-RIO]. (2022). Ceasa/RJ completa 48 anos de atividade. Disponível em < <https://www.ceasa.rj.gov.br/> > Acesso em: 31 mar. 2024
- Centro de Estudos Avançados em Economia Aplicada [CEPEA]. (2020) PIB DO AGRONEGÓCIO DO RIO DE JANEIRO: Ano base de Evolução 2017 -2020. Disponível em:https://www.cepea.esalq.usp.br/upload/kceditor/files/Relat%C3%B3rio%20PIB%20do%20Agroneg%C3%B3cio%20do%20Rio%20de%20Janeiro_FINAL_dez21.pdf. Acesso em: 31 mar. 2024.
- Cruz, S. M. S. et al. (2018). Towards an e-infrastructure for Open Science in Soils Security.
- Cruz, P. V. C. (2023). Validação Dinâmica E Inteligente De Séries Temporais De Produtos Comercializados Pelo CEASA-Rio. Monografia ESQLE-USP.
- Companhia Nacional de Abastecimento [CONAB]. (2022). Portal de Informações Agropecuárias. Disponível em: < <https://portaldeinformacoes.conab.gov.br/produtos-360.html> >. Acesso em: 31 mar. 2024.
- Cunha, A. R. A. de A.; Belik, W (2015). A produção agrícola e a atuação das Centrais de Abastecimento no Brasil. Segurança Alimentar e Nutricional, Campinas, SP, v.

- 19, n. 1, p. 46–59, 2015. DOI: 10.20396/san.v19i1.8634668. Disponível em: < <https://periodicos.sbu.unicamp.br/ojs/index.php/san/article/view/8634668> > Acesso em: 31 mar. 2024.
- Empresa Brasileira de Pesquisa Agropecuária [EMBRAPA]. (2022). Visão do Futuro do Agro Brasileiro Disponível em: < <https://www.embrapa.br/visao-de-futuro/trajetoria-do-agro> >. Acesso em: 31 mar. 2024.
- Empresa De Assistência Técnica e Extensão Rural Do Estado Do Rio De Janeiro [EMATER-RIO]. (2022). Acompanhamento Sistemático da Produção Agrícola - ASPA. Disponível em: < <http://www.emater.rj.gov.br/tecnica.asp> >. Acesso em: 31 mar. 2024.
- Ferrari, A.; Russo, M. (2024) Introducing Microsoft Power BI. Washington, Estados Unidos da America: Microsoft Press, 2016. 384 p. ISBN 978-1-5093-0228-4. Disponível em: < https://download.microsoft.com/download/0/8/1/0816F8D1-D1A5-4F60-9AF5-BC91E18D6D64/Microsoft_Press_ebook_Introducing_Power_BI_PDF_mobile.pdf >. Acesso em: 31 mar. 2024.
- Inmon, B., Levins, M., Srivastava, R. (2021). Building the Data Lakehouse. 1a ed., Technics Publications.
- Instituto Brasileiro De Geografia E Estatística [IBGE]. (2022) Malha Municipal. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=acesso-ao-produto>. Acesso em: 31 mar. 2024.
- Lee, T. B. (2024) 5 ★ Open Data. Disponível em: <https://5stardata.info/en/>. Acesso em: 31 mar. 2024.
- Madera, C., Laurent, A., Rouge, T.L., Miralles, A., (2017). How can the data lake concept influence information system design for agriculture? In: 11th European conference dedicated to the future use of ICT in the agri-food sector, bioresource and biomass sector (EFITA 2017), pp. 181–182.
- MARINHO, Élton Carneiro; SCHMITZ, Eber Assis; CRUZ, Sérgio Manuel Serra da. Provenance, Blockchain, and Smart Contracts as a Traceable Soil Mapping Solution. In: Workshop De Teses E Dissertações (WTDBD) - Simpósio Brasileiro De Banco De Dados (SBBDD), 37. , 2022, Búzios. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 89-97.
- Power BI (2022). Power BI Microsoft. Site. Disponível em: < <https://powerbi.microsoft.com/pt-br/what-is-power-bi/> > Acesso em: 31 mar. 2024.
- Silva, S. A. 2021 GitHub - Repositórios de dados. Disponível em < <https://github.com/SeignerAfonso?tab=repositories> >. Acesso em: 31 mar. 2024.