

## Consolidação e Análise de Dados de Comércio Exterior - Projeto COMEX BI

Jackson David Rodrigues Penaforte ( [www.linkedin.com/in/jackson-penaforte-53901089](https://www.linkedin.com/in/jackson-penaforte-53901089) )

Vitor Amaral (linkedin)

Stela Garcia (linkedin)

Geovanna Lara (linkedin)

Marcos Pacheco (linkedin)

Professor M2 ou Orientador: Prof. Carlos Bastos

Professor P2: Prof. Marcus Nascimento

### Resumo do projeto:

O projeto COMEX BI tem como objetivo desenvolver uma API voltada à consolidação e análise de dados de comércio exterior, a partir de bases públicas disponibilizadas pelo Ministério do Desenvolvimento, Indústria, Comércio e Serviços (MDIC) e pela Fundação SEADE. A proposta visa solucionar a dificuldade de acesso e tratamento de grandes volumes de dados, transformando-os em informações úteis para análise logística. O desenvolvimento utilizou Python (com bibliotecas como Pandas) para limpeza e normalização das bases, integrando os resultados ao Power BI para criação de dashboards interativos. O projeto foi conduzido sob a metodologia ágil SCRUM, dividido em três sprints: automação de dados (MVP1), ajustes de precisão e documentação (MVP2), e publicação final (MVP3). Os resultados obtidos incluem a automação completa da compilação de dados de exportação e importação, eliminação de duplicidades, padronização textual e integração com dashboards dinâmicos. O sistema desenvolvido permite filtrar informações por município, estado e período, facilitando comparações entre regiões e auxiliando na tomada de decisão estratégica. O projeto contribui para o aprimoramento das práticas logísticas e evidencia o potencial do uso de ferramentas de ciência de dados em ambientes acadêmicos e corporativos.

Palavras-Chave: Comércio Exterior; Power BI; Python; Dados Abertos; Logística.

## Abstract:

The COMEX BI project aims to develop an API focused on the consolidation and analysis of foreign trade data, based on public databases provided by the Brazilian Ministry of Development, Industry, Commerce and Services (MDIC) and the SEADE Foundation. The project addresses the challenge of accessing and processing large data volumes, transforming them into actionable insights for logistics analysis. The development process used Python (Pandas library) for data cleaning and normalization, integrating the results into Power BI to create interactive dashboards. Following the Agile SCRUM methodology, the project was divided into three sprints: data automation (MVP1), precision adjustments and documentation (MVP2), and final publication (MVP3). The results achieved include full automation of data compilation for export and import, removal of duplicates, standardized formatting, and integration with dynamic dashboards. The developed system allows filtering by municipality, state, and period, enabling regional comparisons and supporting strategic decision-making. The project contributes to the improvement of logistics practices and highlights the potential of data science tools in both academic and corporate contexts.

Keywords: Foreign Trade; Power BI; Python; Open Data; Logistics.

## 1. Contextualização do projeto

O projeto foi proposto pelo cliente CADI, representado pelo Prof. Marcus Nascimento, com o desafio de desenvolver uma solução capaz de acompanhar o desempenho comercial dos municípios do estado de São Paulo, utilizando dados abertos de comércio exterior. O tema está alinhado ao contexto logístico brasileiro e à necessidade de transformar informações públicas em conhecimento estratégico. O projeto busca oferecer uma ferramenta que apoie análises regionais e comparativas entre estados e municípios, identificando tendências de exportação, importação e variações econômicas no período de 2023 a 2025.

## 2. Objetivos do projeto

O objetivo geral é criar uma API de consolidação de dados de comércio exterior, integrando informações de diferentes bases públicas e disponibilizando-as em um formato limpo e padronizado para análise no Power BI. Os objetivos específicos incluem: automatizar a coleta e limpeza de dados no Google Colab, eliminar duplicidades e inconsistências, gerar arquivos CSV consolidados, e disponibilizar dashboards interativos com filtros e comparativos regionais.

## 3. Fundamentação dos métodos analíticos e das tecnologias utilizadas

O projeto foi desenvolvido com base na metodologia ágil SCRUM, permitindo entregas iterativas e validadas junto ao cliente. A linguagem Python foi empregada para o tratamento e análise dos dados, utilizando bibliotecas como Pandas e NumPy. O Power BI foi adotado como ferramenta de visualização, permitindo a criação de dashboards dinâmicos e acessíveis. O GitHub serviu como repositório de versionamento e publicação dos artefatos do projeto.

Tabela 1 – Referências Técnicas

Autor(es)/Fonte	Métodos/Ferramentas	Aplicação no Projeto	Conclusões Relevantes
McKinney (2017)	Biblioteca Pandas (Python)	Limpeza e manipulação de dados	Automatização eficiente do tratamento de grandes volumes de dados.
Power BI Docs (2024)	Power BI Service e Desktop	Criação de dashboards interativos	Permite análises comparativas entre municípios de forma intuitiva.
Schwaber & Sutherland (2020)	Metodologia SCRUM	Gestão das sprints do projeto	Favorece entregas iterativas e validação constante pelo cliente.

#### 4. Coleta e descrição dos dados utilizados

Os dados foram obtidos a partir do portal de Estatísticas de Comércio Exterior do MDIC e da Fundação SEADE, abrangendo os anos de 2023 a 2025. O tratamento foi realizado no Google Colab, utilizando Python para leitura, concatenação, mesclagem e limpeza das tabelas. Foram removidas duplicidades com a função `drop_duplicates`, e padronizados os nomes de municípios e unidades federativas. As bases resultantes, `finalex.csv`, `finalimp.csv`, `finalexSP.csv` e `finalimpSP.csv`, foram integradas ao Power BI para visualização interativa.

#### 5. Resultados esperados

Os resultados esperados incluíam a automatização do processo de compilação de dados e a geração de relatórios analíticos acessíveis. Esses objetivos foram alcançados com sucesso: os dashboards Power BI oferecem filtros dinâmicos e indicadores detalhados sobre a balança comercial paulista. O projeto também entregou documentação técnica e código aberto no GitHub, permitindo futuras expansões e manutenções.

#### Referências

McKinney, W. (2017). Python for Data Analysis. O'Reilly Media.  
Schwaber, K.; Sutherland, J. (2020). The Scrum Guide: The Definitive Guide to Scrum. Microsoft. Power BI Documentation. Disponível em: <https://learn.microsoft.com/power-bi/>.

Ministério do Desenvolvimento, Indústria, Comércio e Serviços (MDIC). Estatísticas de  
Comércio Exterior em Dados Abertos.  
Fundação SEADE. Indicadores Econômicos e de Comércio Exterior.