

Trabalho Final

Data Warehouse: Dados sobre a Exportação do Brasil

Índice

1.	Introdução	3
2.	Fonte de Dados	4
3.	Modelagem Dimensional	6
3.1.	Dimensões	8
3.2.	Fatos	11
4.	Apresentação dos Dados	13

Ferramentas

Ferramenta de ETL: [Apache Hop 2.4.0](#)

Banco de dados: [Maria DB - HeidiSQL 11.0.1](#)

Ferramenta BI: [Power BI](#)

Introdução

A análise de dados desempenha um papel fundamental no entendimento e no direcionamento de estratégias comerciais eficazes. No contexto do comércio internacional, o Brasil possui uma posição de destaque como um importante exportador global. Com o objetivo de obter insights valiosos e orientar decisões informadas no setor de exportação, foi construído um Data Warehouse para consolidar e explorar os dados de exportação do Brasil.

Este relatório técnico apresenta o processo de criação dessa solução de Business Intelligence (BI), focada na análise de dois principais indicadores: peso líquido da exportação e valor exportado FOB (Free on Board). Utilizei o MariaDB como o banco de dados para o armazenamento dos dados consolidados, o Apache Hop como ferramenta de ETL para extração, transformação e carga dos dados brutos e o Power BI como a plataforma de Business Intelligence para a visualização dos resultados.

Os dados brutos de exportação foram obtidos a partir de fontes governamentais confiáveis e atualizadas. No processo de ETL, esses dados foram extraídos, limpos e transformados, mapeando-os para a estrutura do modelo dimensional. As dimensões utilizadas incluem país de destino, estado de origem, bloco econômico de destino, receita federal que exportou, transporte utilizado, produto e tempo, enquanto as medidas consideradas são o peso líquido da exportação e o valor exportado FOB.

A modelagem dimensional adotada no Data Warehouse permite uma organização eficiente dos dados, proporcionando uma visão mais abrangente e facilitando a análise dos indicadores de exportação. Ao explorar as dimensões e medidas relacionadas, os usuários poderão identificar padrões, tendências e insights valiosos para orientar suas decisões comerciais.

A apresentação dos dados é feita por meio do Power BI, uma plataforma avançada de visualização e análise de dados. Com o Power BI, é possível criar painéis interativos, gráficos dinâmicos e relatórios personalizados que permitem aos usuários explorar os dados de exportação de forma intuitiva e visualmente atraente. A interface do Power BI oferece recursos de filtragem, drill-down e interatividade, permitindo uma análise mais profunda e uma compreensão aprimorada dos resultados.

Ao longo deste relatório técnico, serão apresentados exemplos de visualizações geradas no Power BI, demonstrando como os dados de peso líquido da exportação e valor exportado FOB podem ser explorados e interpretados. Essas visualizações fornecem insights valiosos sobre os principais mercados de destino, produtos mais exportados e tendências temporais relevantes,

auxiliando gestores, analistas e demais interessados no planejamento estratégico e na identificação de oportunidades de negócio.

O Data Warehouse para dados de exportação do Brasil, utilizando o MariaDB como banco de dados, o Apache Hop como ferramenta de ETL e o Power BI para visualização, foi desenvolvido para fornecer uma solução robusta e abrangente para análise e monitoramento do desempenho das exportações do país. A seguir, iremos explorar em detalhes as etapas de construção dessa solução de Business Intelligence, oferecendo uma visão aprofundada de sua arquitetura e funcionalidades.

Fontes de Dados

Neste projeto, utilizo o [COMEX Stat](#) como principal fonte de dados para a análise de exportações do Brasil. O COMEX Stat é uma plataforma online disponibilizada pelo Ministério da Economia, que oferece acesso a dados detalhados sobre comércio exterior. Através desta plataforma, é possível obter informações precisas e atualizadas sobre as exportações brasileiras.

Dados brutos:

Os dados brutos de exportação obtidos no COMEX Stat são organizados em tabelas principais e auxiliares, que fornecem informações detalhadas sobre as transações comerciais. A estrutura dos dados brutos inclui as seguintes principais informações:

País de destino: Identifica o país para o qual a exportação foi realizada, classificamos também em blocos econômicos.

Produto(NCM e SH2): Descreve o tipo de produto exportado, utilizando códigos e descrições padronizadas.

Peso líquido da exportação: Indica o peso líquido do produto exportado, fornecendo informações sobre a quantidade física.

Mês/Ano: Refere-se a data em que ocorreu a exportação da mercadoria, foram coletados dados de janeiro de 2000 até março de 2023.

Código da UF de origem: UF significa "Unidade da Federação" e refere-se aos estados brasileiros. Esse código indica o estado de origem da mercadoria exportada, classificamos também por região.

Código da via de transporte: Indica o meio de transporte utilizado para o envio da mercadoria, como marítimo, aéreo, rodoviário, etc.

Código da URF de embarque: URF significa "Unidade da Receita Federal" e representa um local específico, como porto, aeroporto ou ponto de entrada e saída de mercadorias, onde ocorreu o embarque ou desembarque da mercadoria.

Quilograma líquido: É o peso líquido da mercadoria exportada, medido em quilogramas.

Valor dólar FOB (US\$): Representa o valor em dólar da mercadoria no ponto FOB (Free On Board), como mencionado anteriormente. É o valor da mercadoria no momento em que ela é colocada a bordo do meio de transporte no porto de embarque designado.

Falarei mais sobre esses dados posteriormente.

Além dessas informações principais, as tabelas auxiliares fornecem detalhes adicionais, como classificações dos produtos por categorias específicas, códigos harmonizados, informações sobre embalagens, entre outros dados relevantes para uma análise mais detalhada das exportações.

Os dados brutos são disponibilizados em formato estruturado, como arquivos CSV (Comma-Separated Values) ou Excel, que podem ser facilmente importados e processados pelas ferramentas de ETL (Extração, Transformação e Carga) utilizadas no projeto.

Ao realizar a integração dos dados brutos no processo de ETL, é necessário aplicar etapas de limpeza, transformação e mapeamento para estruturar os dados de acordo com o modelo dimensional estabelecido. Essas etapas garantem a consistência e a qualidade dos dados armazenados no Data Warehouse, possibilitando análises precisas e confiáveis sobre as exportações do Brasil.

No próximo tópico, abordaremos em detalhes a modelagem dimensional adotada neste projeto, que visa facilitar a análise e a compreensão dos dados de exportação.

Modelagem Dimensional

A modelagem dimensional adotada no Data Warehouse foi projetada para fornecer uma estrutura organizada e eficiente para a análise dos dados de exportação.

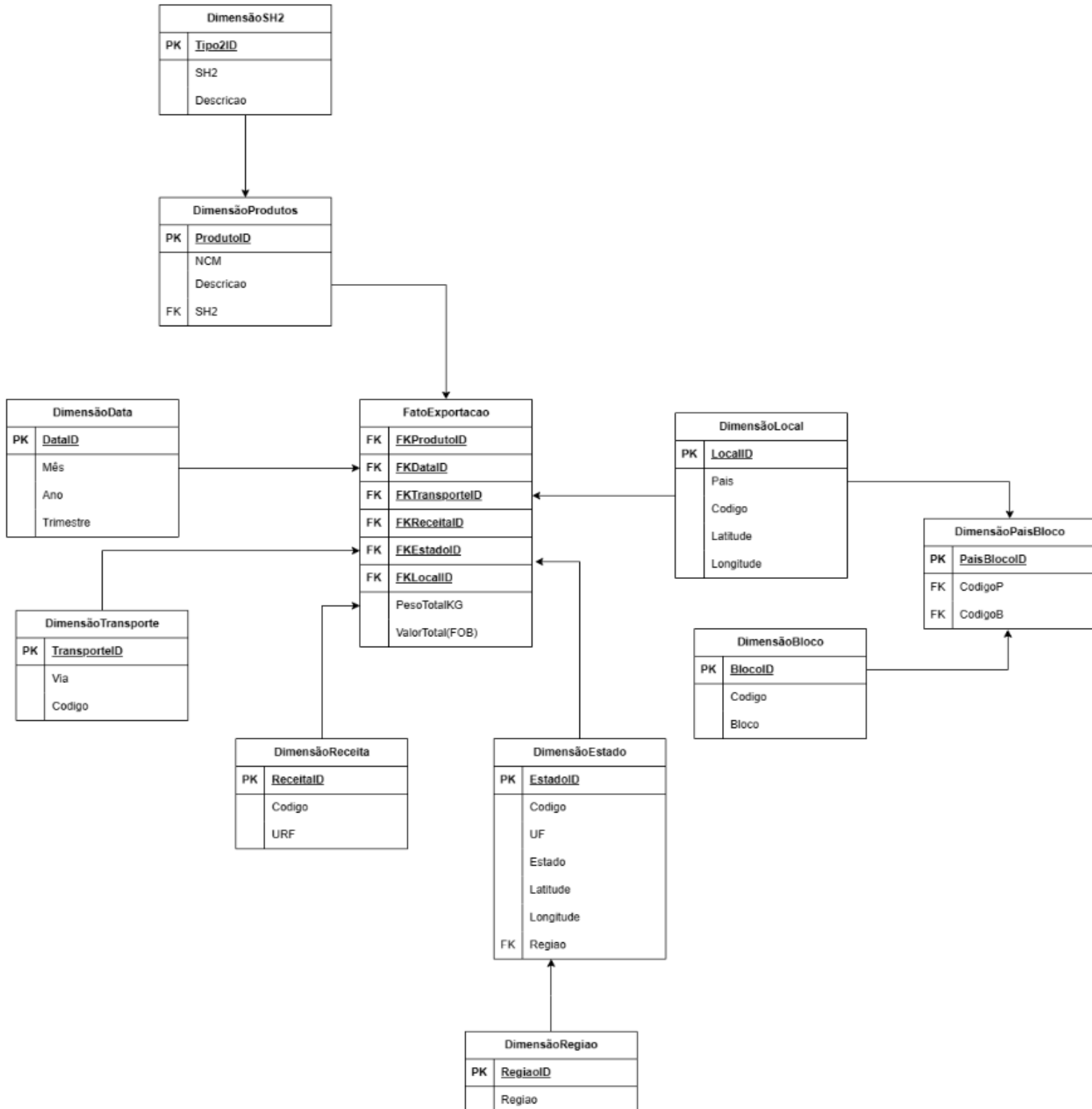
As dimensões permitem uma segmentação e uma análise mais detalhada dos dados, permitindo identificar padrões, tendências e insights relevantes. A tabela de fatos registra as medidas quantitativas, agregando-as em níveis apropriados de granularidade.

Os relacionamentos estabelecidos entre as dimensões e a tabela de fatos permitem a integração dos dados e a realização de análises multidimensionais. As hierarquias foram definidas nas dimensões, facilitando a navegação e a exploração dos dados em diferentes níveis de detalhe.

Além disso, medidas calculadas podem ser derivados da modelagem, fornecendo insights adicionais sobre o desempenho das exportações.

A modelagem dimensional adotada segue os princípios de esquema de **snowflake**, garantindo uma estrutura simples e de fácil compreensão. As chaves primárias e estrangeiras foram definidas para garantir a integridade e a consistência dos dados no Data Warehouse.

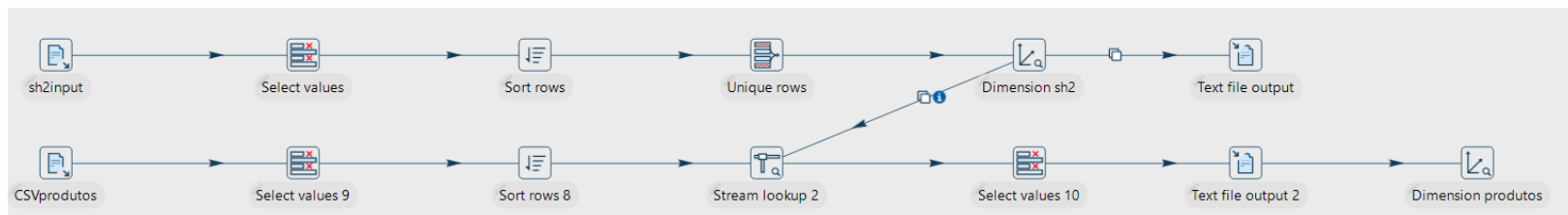
A imagem a seguir representa visualmente a estrutura da modelagem dimensional adotada, destacando as dimensões, a tabela de fatos e os relacionamentos estabelecidos. Esta imagem oferece uma visão geral da organização dos dados e servirá como referência para a análise das exportações no Data Warehouse.



Dimensões

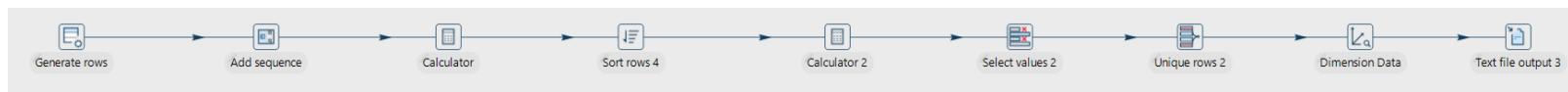
DimensãoProdutos: A DimensãoProdutos é responsável por classificar os produtos exportados pelo seu NCM (Nomenclatura Comum do Mercosul) e fornecer a descrição correspondente. Ela permite a segmentação e a análise dos dados com base nos diferentes tipos de produtos exportados. Os dados contidos nesta dimensão incluem o código NCM (BIGINT) e a descrição (VARCHAR) de cada produto.

DimensãoSH2: A DimensãoSH2 é uma subdivisão da DimensãoProdutos e fornece uma classificação adicional dos produtos com base no Sistema Harmonizado de Codificação e Descrição de Mercadorias. Essa classificação permite uma análise mais detalhada e específica dos produtos exportados. Os dados nesta dimensão incluem o código SH2 (BIGINT) e a descrição (VARCHAR) correspondente.



É imputado dois arquivos, um de referência SH2 e outro com o produto e NCM, onde faço referência para eu ter a que tipo SH2 tal ncm pertence e exparto para o banco de dados e um arquivo.

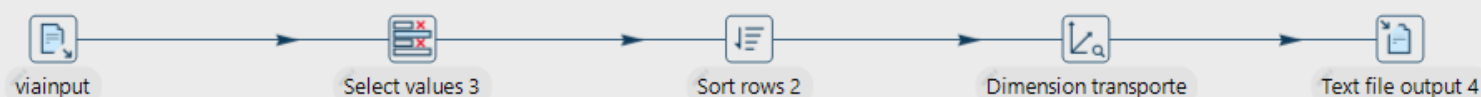
DimensãoData: A DimensãoData representa as informações relacionadas à data das transações de exportação. Ela inclui dados de mês(INT), ano(INT) e trimestre(INT), permitindo uma análise temporal dos dados de exportação. Essa dimensão é essencial para identificar tendências, sazonalidades e variações ao longo do tempo.



É gerado uma sequência de data, com mês, ano e trimestre e exparto para o banco de dados e um arquivo.

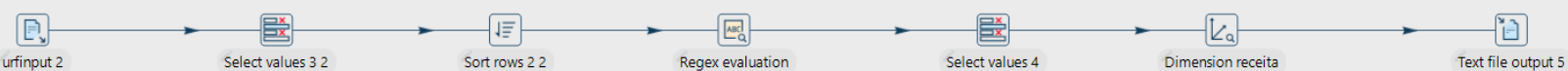
DimensãoTransporte: A DimensãoTransporte registra as informações relacionadas aos meios de transporte utilizados nas exportações. Ela inclui os modos de transporte(VIA-VARCHAR), como aéreo, marítimo, rodoviário, ferroviário, bem como o

código(BIGINT) do transporte. Essa dimensão permite a análise do desempenho de exportação com base nos diferentes meios de transporte.



É imputado um arquivo com os dados dos transportes utilizados, seleciono as colunas descritas acima, ordeno e exporto para o banco de dados e um arquivo.

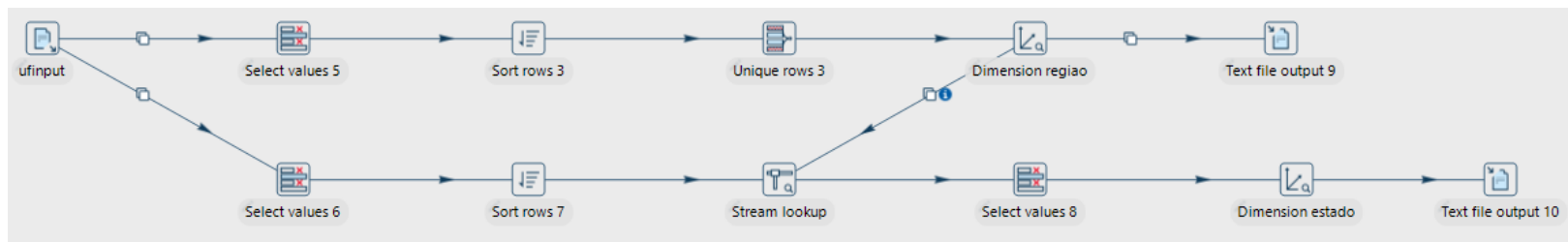
Dimensão Receita: A Dimensão Receita contém informações sobre a receita que realizou as exportações, incluindo seu nome(URF - TINYTEXT) e seu código(BIGINT). Ela permite uma análise detalhada dos aspectos financeiros das exportações, auxiliando na compreensão dos aspectos econômicos e tributários envolvidos nas transações comerciais.



É imputado um arquivo com os dados das URFs utilizadas, seleciono as colunas descritas acima, ordeno, utilizo regex evaluation para retirar o código do nome das URFs já que eles vem juntos e exporto para o banco de dados e um arquivo.

DimensãoEstado: A DimensãoEstado representa os estados brasileiros de origem das exportações. Ela fornece informações sobre o estado emissor das mercadorias, incluindo seu UF(VARCHAR), nome(Estado-VARCHAR), seu código(BIGINT) e um FK para região, permitindo análises geográficas e regionais do comércio exterior brasileiro. Além disso, essa dimensão pode ser relacionada a outras dimensões geográficas.

DimensãoRegiao: Essa dimensão agrupa os estados por regiões geográficas específicas, ela fornece informações sobre a região emissora das mercadorias, incluindo seu nome(Regiao-VARCHAR).

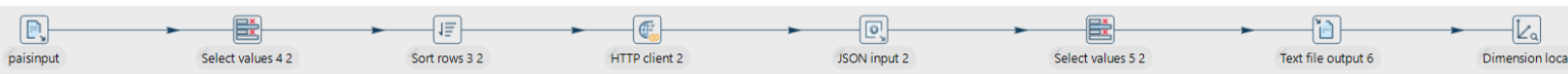


É imputado um arquivo com os dados dos Estados e sua região, seleciono os dados do estado e da região como descritos acima, separo eles, ordeno e faço a referência de região para o estado e exporto para o banco de dados e dois arquivos.

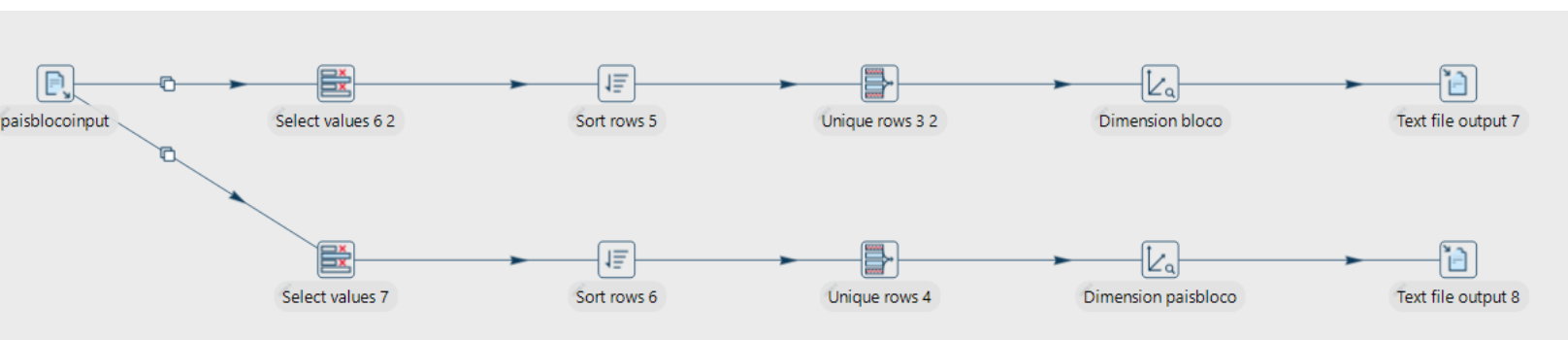
DimensãoLocal: A DimensãoLocal se refere aos locais de destino das exportações, países incluindo seu Nome(Pais-VARCHAR), latitude(TINYTEXT), longitude(TINYTEXT) e seu código(BIGINT). Ela permite uma análise detalhada dos principais mercados de destino das exportações brasileiras. Além disso, a DimensãoLocal pode ser subdividida em outras dimensões.

DimensãoBloco: Agrupa os países em blocos econômicos específicos incluindo seu nome(Bloco-VARCHAR) e seu código(BIGINT).

DimensãoPaisBloco: Associa cada país a um bloco econômico específico incluindo os FKs de bloco(CodigoB-BIGINT) e pais(CodigoP-BIGINT).



É imputado um arquivo com os dados dos Países, seleciono os dados do país como descritos acima, ordeno, faço uma busca por HTTP para coletar as informações de latitude e longitude e exporto para o banco de dados e um arquivo.



É imputado um arquivo com os dados dos Blocos, seleciono os dados do bloco como descritos acima, separo eles em uma dimensão para o bloco e uma para referência entre o código do país e código do bloco, ordeno e exporto para o banco de dados e dois arquivos.

Fatos

FatoExportação:

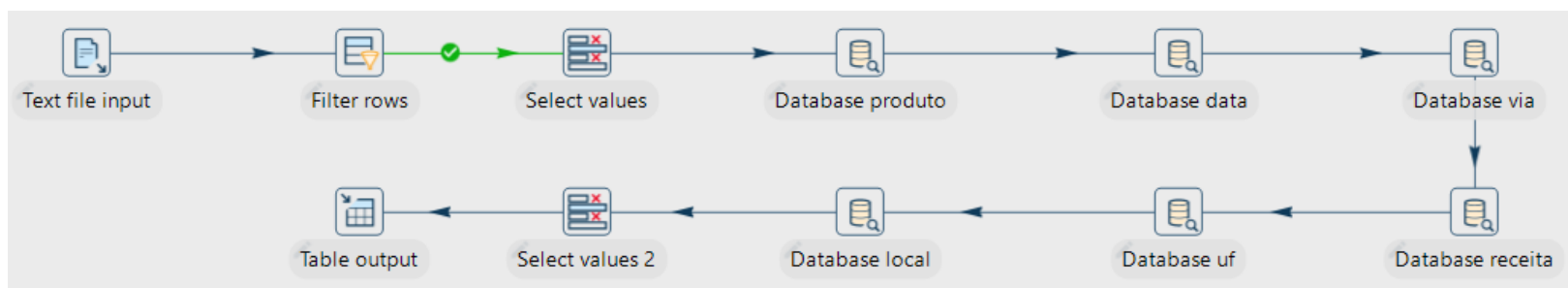
A tabela de fatos é responsável por armazenar as medidas numéricas relacionadas às exportações, como peso líquido da exportação e valor exportado FOB. Essa tabela permite o cálculo de métricas, a realização de análises quantitativas e a comparação de desempenho ao longo do tempo.

As principais informações contidas na tabela de fatos são:

Peso líquido da exportação(BIGINT): Representa o peso total das mercadorias exportadas, descontando-se o peso das embalagens e materiais de suporte. Essa medida é importante para análises logísticas e de transporte, permitindo avaliar volumes de carga e eficiência nas operações de exportação.

Valor exportado FOB (Free on Board)(BIGINT): Refere-se ao valor total das mercadorias exportadas, considerando o preço de venda das mercadorias no local de embarque, sem incluir despesas adicionais de transporte, seguro e impostos. Essa medida é crucial para análises financeiras e de receita, auxiliando no cálculo do valor das exportações e no monitoramento do desempenho econômico.

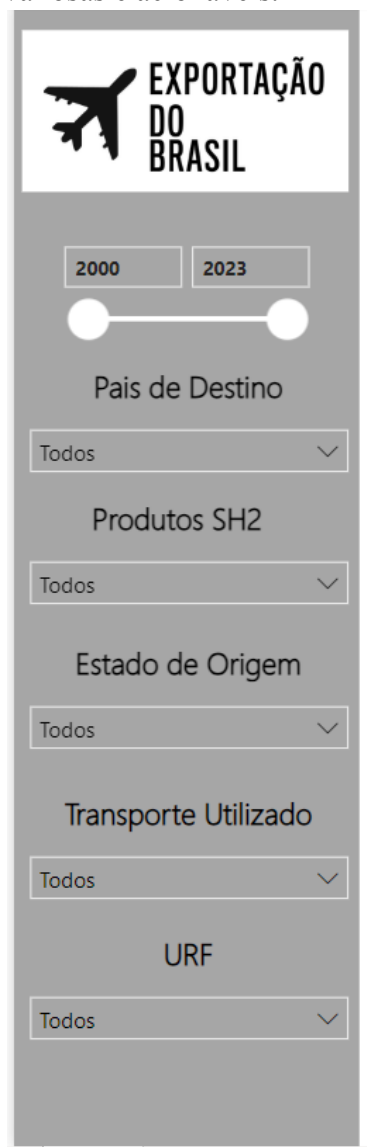
Essas medidas quantitativas são registradas na tabela de fatos, associadas às chaves primárias das dimensões correspondentes, DimensãoProdutos, DimensãoData, DimensãoTransporte, DimensãoReceita, DimensãoEstado e DimensãoLocal. Essa estrutura de relacionamento permite a análise multidimensional e a agregação das medidas em diferentes níveis de granularidade, proporcionando insights valiosos sobre o desempenho das exportações.



É imputado um arquivo com os dados das vendas dos produtos, filtro um certo tempo, seleciono os dados do fato descritos acima, separo eles em um fato e faço referências entre os códigos das dimensões e adiciono o FK das dimensões e exporto para o banco de dados e dois arquivos.

Apresentação dos Dados

A camada de apresentação de dados é uma parte essencial da solução de Business Intelligence, permitindo que você explore e analise os dados de exportação do Brasil de forma interativa e visualmente atraente. Utilizando a poderosa ferramenta Power BI, você pode transformar os dados brutos de exportação, como peso líquido e valor exportado FOB, em informações valiosas e acionáveis.



EXPORTAÇÃO DO BRASIL

2000 2023

País de Destino

Todos

Produtos SH2

Todos

Estado de Origem

Todos

Transporte Utilizado

Todos

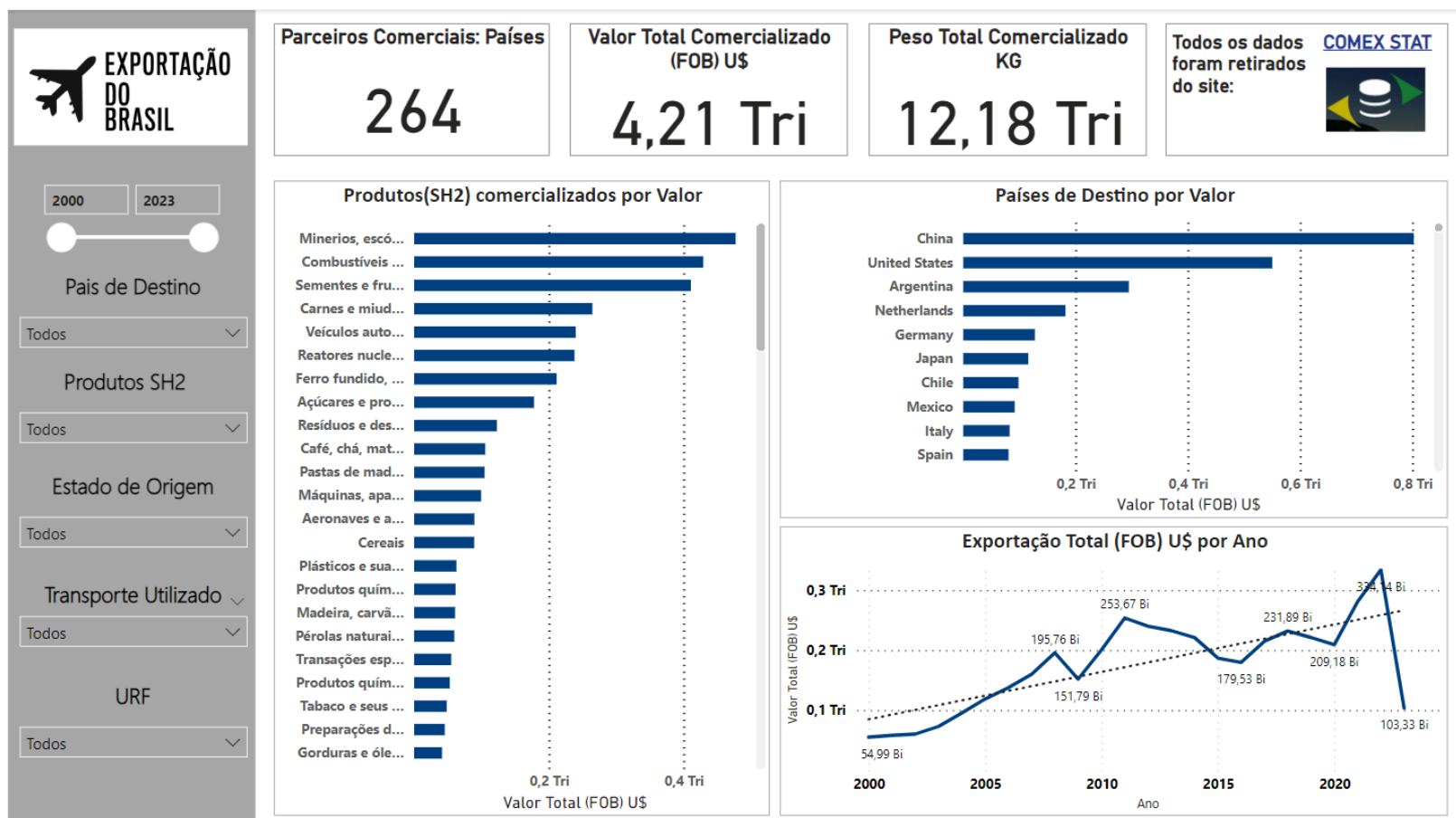
URF

Todos

Em todas as páginas de apresentação, é presente o filtro acima, em que é possível filtrar pelo ano, país de destino do produto, pela categoria de produtos(SH2), pelo estado de origem dos produtos, pelo transporte utilizado para transportá-lo e pelo URF que realizou a exportação.

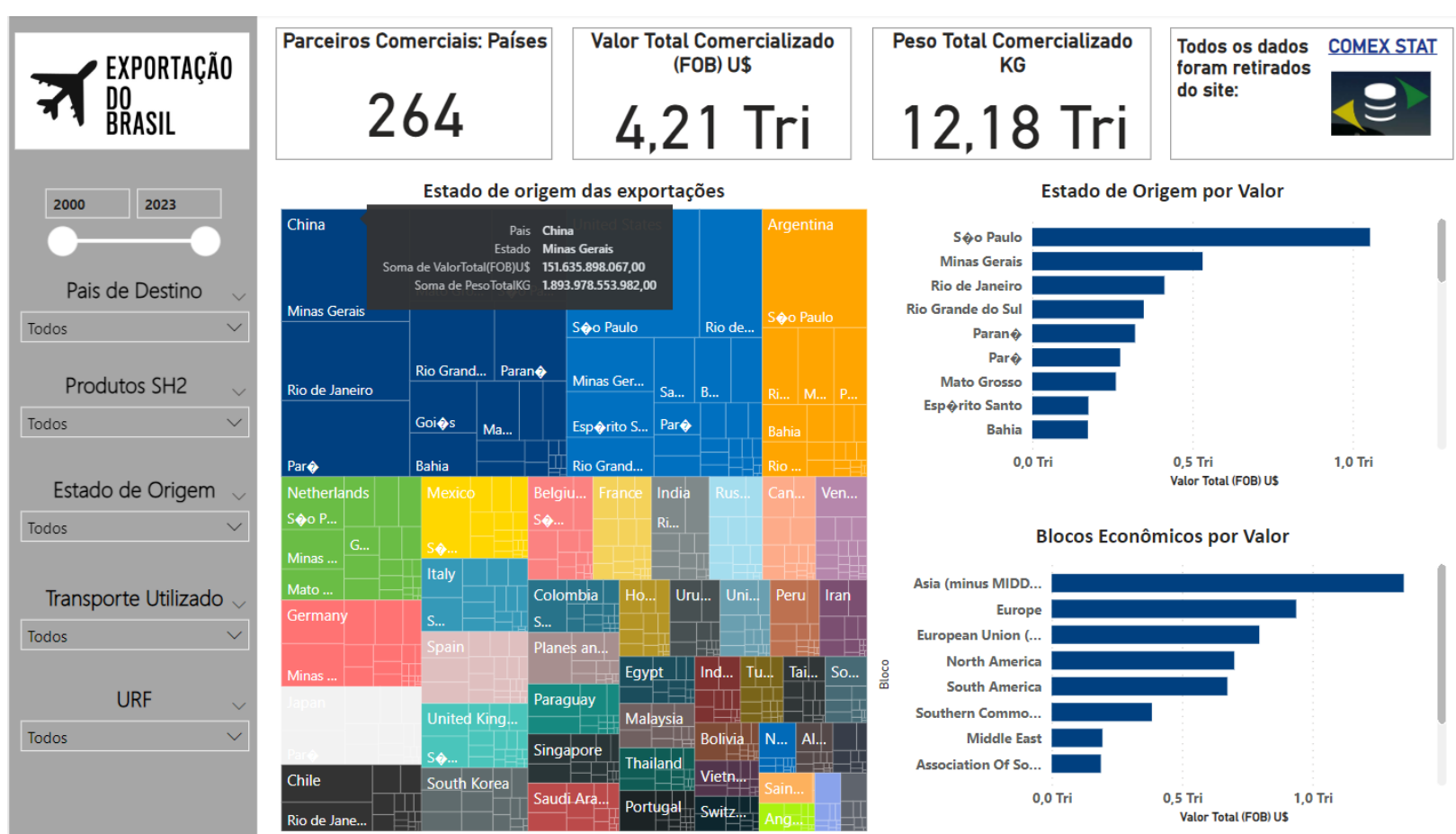
Página 1: Início

Uma visão mais geral dos produtos, em ordem dos mais exportados por valor, os países que mais compram produtos do brasil e o valor exportado por ano.



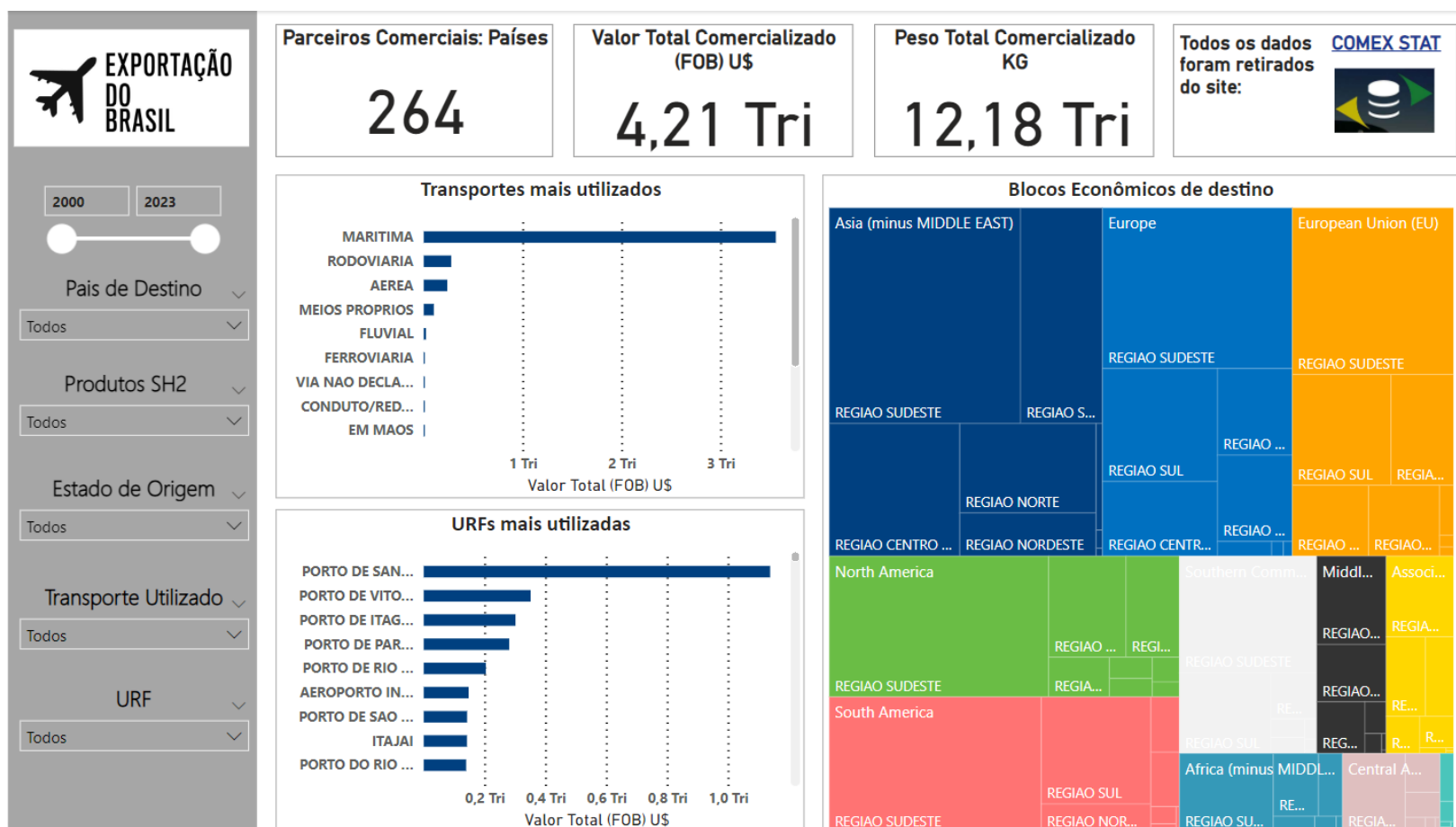
Página 2: Países, Estados e Blocos

Uma visão mais territorial, em ordem dos países que mais compram e de qual estado mais compram, os Estados que mais produzem e blocos econômicos que mais compram do Brasil.



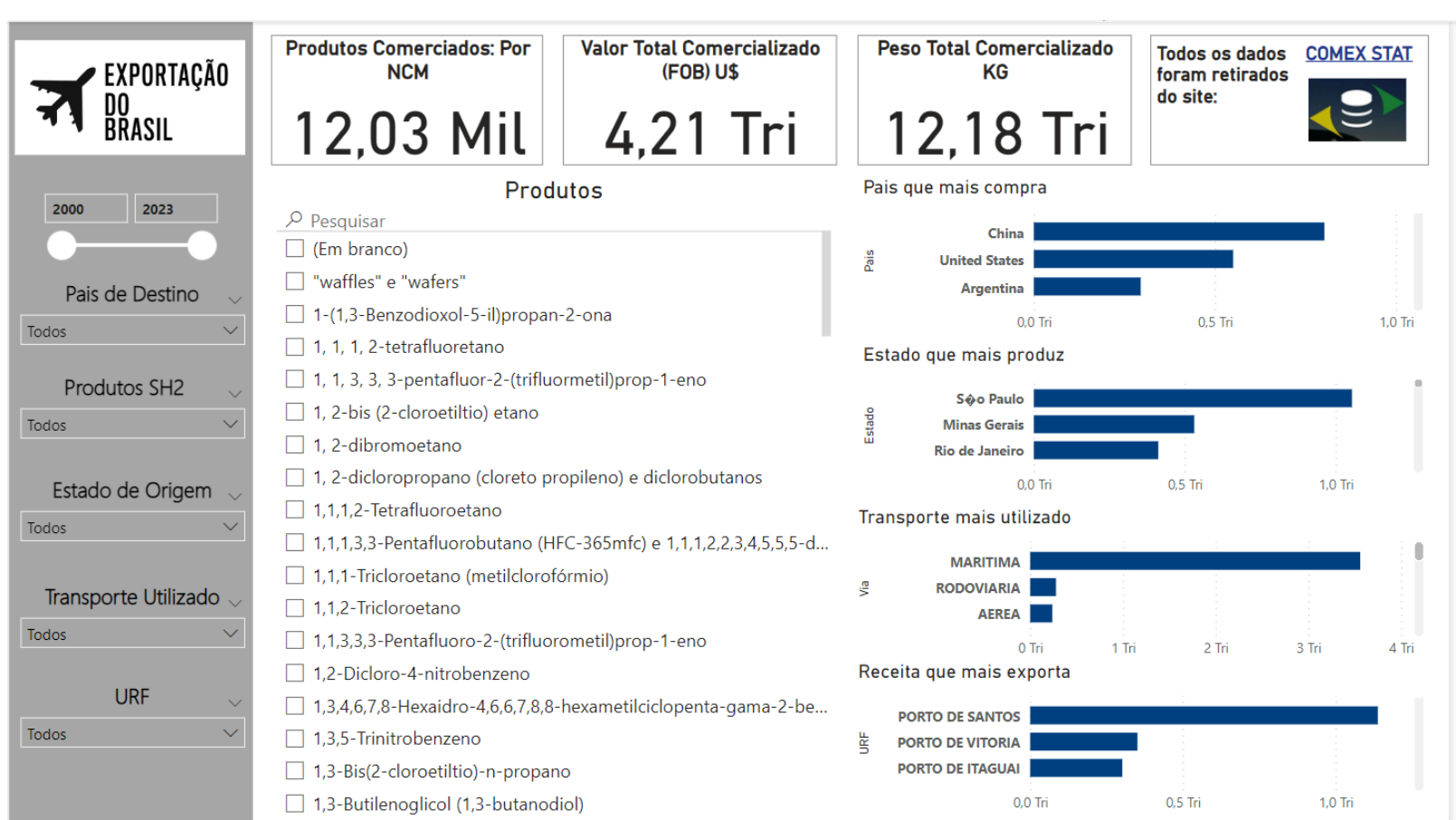
Página 3: URF e Transportes

Uma visão mais de logística, quais URFs mais exportam e qual transporte mais utilizado e um agregado de blocos que mais compram e de qual região mais compram.



Página 4: Produtos

Uma visão mais dos produtos, uma lista de produtos para escolher sobre qual produto quer ver informações, como, o país que mais compra, o estado que mais produz, o transporte que mais utiliza e qual urf mais exporta.



Página 5: Mapas

Uma visão no mapa dos países que mais compram do Brasil.

