

PUC – Ciência de Dados e Analytics

MVP – Engenharia de Dados

Diego Justen Zilves

Exportações Brasileiras

Objetivo

Analisar exportações brasileiras entre os anos de 2018 e 2022, para saber:

- 1 – Quais são os principais estados exportadores do Brasil?
- 2 – Quais são os principais tipos de produtos exportados pelo Brasil?
- 3 – Quais os principais produtos exportados dos principais estados?

Essa análise permitirá entender melhor como funciona esse braço importante da economia brasileira.

Detalhamento

1 – Busca pelos dados

Os dados foram retirados do site Comex Stat, site do governo federal, no qual reúne todas as informações sobre operações de Comércio Exterior do Brasil.

<http://comexstat.mdic.gov.br/pt/home>





Exportações e Importações Geral

Consultas com dados mensais a partir de 1997 ao ano atual. Filtros e detalhes de países, blocos, UF do produto, NCM e sistema harmonizado (SH6, SH4, Capítulo e Seção), classificação por grandes categorias econômicas e Classificação Uniforme para o Comércio Internacional.

Acessar ➔

2 – Coleta dos dados

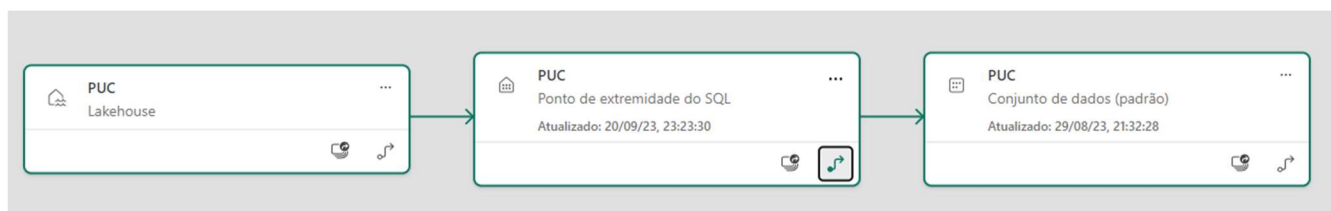
Os dados foram baixados diretamente do site em formato CSV e posteriormente carregados de forma manual para o Workspace denominado PUC, dentro da plataforma Fabric, da Microsoft. O site possui restrição para aplicação de API para realização de Web Scraping, daí a necessidade que seja feito de forma manual todo o processo.

The screenshot shows the Microsoft Fabric workspace 'PUC'. On the left, the 'Gerenciador' (Manager) pane shows a tree view with 'PUC' expanded, containing 'Tables' (with sub-items 'bd', 'cgce', 'ncm', 'uf') and 'Files'. The main area displays a list of files:

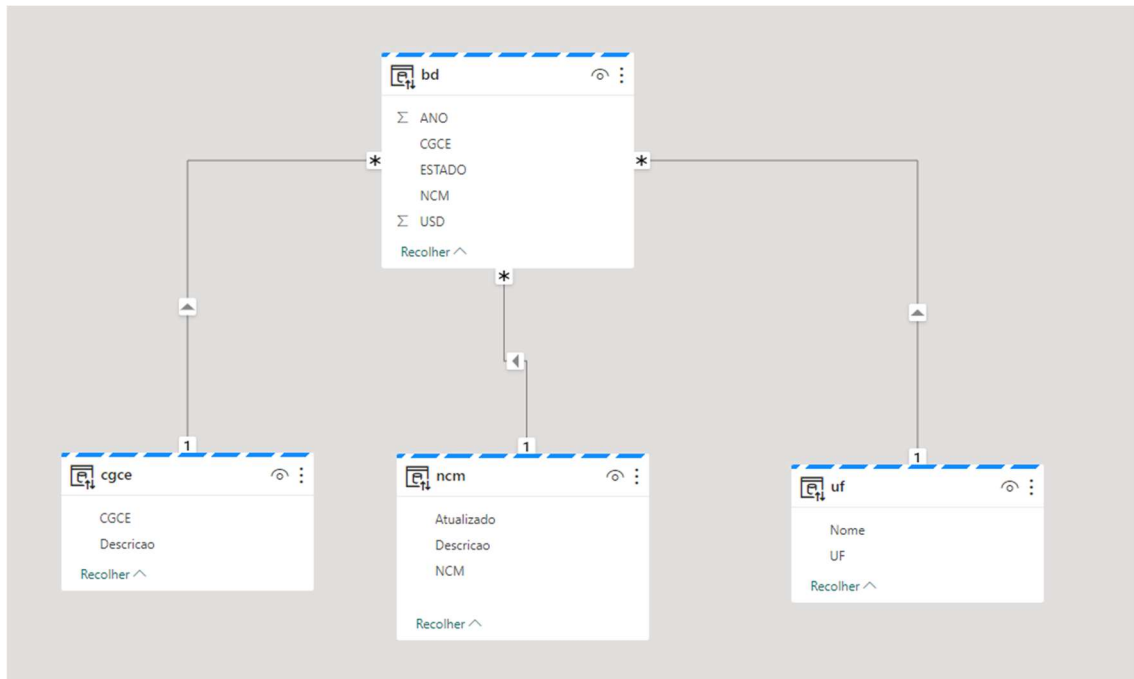
Nome	Data da modificação	Tipo
BD.csv	9/19/2023 9:55:08 PM	CSV
CGCE.csv	9/12/2023 9:41:38 PM	CSV
NCM.csv	9/19/2023 10:32:00 PM	CSV
UF.csv	9/12/2023 9:41:38 PM	CSV

3 – Modelagem

Os dados baixados e carregados de forma manual para dentro do Workspace PUC, criando assim um banco de dados relacional (SGBD). Lá consta o Lakehouse, que seria o Data Lake da plataforma para armazenamento de todos os dados coletados. Depois de carregados, os arquivos CSV foram transformados em tabelas e carregados para dentro do banco de dados do SQL, onde foram trabalhados para identificação do tipo de dados de cada uma das colunas, dando assim origem ao Conjunto de Dados.



Para esse MVP foi considerado a elaboração de um Esquema Estrela, onde temos a tabela fato bd e as tabelas dimensão cgce, ncm e uf alimentando a principal.



As tabelas utilizadas são as seguintes:

- **bd** é uma tabela fato e nela temos as colunas “ANO”, “ESTADO”, “NCM”, “CGCE” e “USD”, com as seguintes características:
 - ANO – ano da exportação realizada. Os dados da coluna são compostos apenas por números com quatro dígitos e valores limitados entre os anos de 2018 e 2022;
 - ESTADO – Unidade Federativa ou cluster exportadora. Os dados são compostos apenas por string com dois caracteres, que representam as siglas dos respectivos estados brasileiros. Essa coluna funcionará como chave estrangeira (FK) para se relacionar com a tabela dimensão “uf”;
 - NCM – Nomenclatura Comum do Mercosul é uma classificação utilizada no Comércio Internacional de cargas. Eles foram criados no intuito de unificar a classificação dos materiais, levando em consideração as características de produção dos mesmos, evitando assim margem de interpretação diferentes pelas pessoas. Além disso, fornecem uma granularidade maior para os tipos de produtos. Os dados são compostos apenas por números com 7 ou 8 dígitos, limitados entre 1000000 e 99999999. Essa coluna funcionará como chave estrangeira (FK) para se relacionar com a tabela dimensão “ncm”;
 - CGCE – Classificação por Grandes Categorias Econômicas é uma representação sintética das grandes categorias econômicas, com níveis hierárquicos para uma melhor entendimento e classificação dos produtos. Os dados são compostos apenas por números com 3 dígitos, limitados entre 110 e 999. Essa coluna funcionará como chave estrangeira (FK) para se relacionar com a tabela dimensão “cgce”;
 - USD – Valor exportado em dólar americano. Os dados são compostos apenas por números e sem valores mínimos ou máximos;

Visualização de dados					Mostrando 1000 linhas
	123 ANO	ABC ESTADO	123 NCM	123 CGCE	12L USD
1	2018	MG	9011110	210	3211912614
2	2018	PR	2071400	324	1741915557
3	2018	SC	2071400	324	1410917489
4	2018	SP	2023000	324	1056769047
5	2018	MT	2023000	324	914960839
6	2018	GO	2023000	324	769441639
7	2018	SC	2032900	220	574817838
8	2018	PR	2071200	324	546449367
9	2018	MG	2023000	324	519965067
10	2018	ES	9011110	210	444741725

- **cgce** é uma tabela Dimensão e nela temos as colunas “CGCE” e “Descrição”, com as seguintes características:
 - CGCE – Classificação por Grandes Categorias Econômicas é uma representação sintética das grandes categorias econômicas, com níveis hierárquicos para uma melhor entendimento e classificação dos produtos. Os dados são compostos apenas por números com 3 dígitos, limitados entre 110 e 999. Essa coluna funcionará como chave primária (PK) para se relacionar com a tabela fato “bd”;
 - Descrição – descrição detalhada do código CGCE. Os dados são uma string, sem que haja restrições;

Visualização de dados		
	123 CGCE	ABC Descrição
1	240	Insumos industriais elaborados
2	322	Bens de consumo nao duraveis
3	110	Bens de capital (exceto equipamentos de transporte)
4	323	Alimentos e bebidas basicos/ destinados principalmente ao consumo domestico
5	120	Equipamentos de transporte industrial
6	312	Automoveis para passageiros
7	230	Insumos industriais basicos
8	321	Bens de consumo semiduraveis
9	313	Equipamentos de transporte nao industrial
10	324	Alimentos e bebidas elaborados/ destinados principalmente ao consumo domestico

- **ncm** é uma tabela Dimensão e nela temos as colunas “NCM” e “Descrição”, com as seguintes características:
 - Atualizado – mostra o ano no qual a NCM passou pela última atualização. Os dados da coluna são compostos apenas por números com quatro dígitos e valores limitados até no máximo o ano atual;
 - NCM – Nomenclatura Comum do Mercosul é uma classificação utilizada no Comércio Internacional de cargas. Eles foram criados no intuito de unificar a classificação dos materiais, levando em consideração as características de produção dos mesmos, evitando assim margem de interpretação diferentes pelas pessoas. Além disso, fornecem uma granularidade maior para os tipos de produtos. Os dados são compostos apenas por números com 7 ou 8 dígitos, limitados entre 1000000 e 99999999. Essa coluna funcionará como chave estrangeira (FK) para se relacionar com a tabela dimensão “ncm”;
 - Descrição – descrição detalhada da NCM. Os dados são uma string, sem que haja restrições;

Visualização de dados

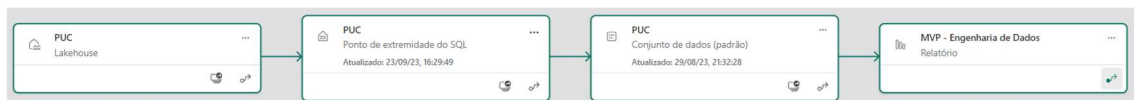
	ABC Atualizado	123 NCM	ABC Descricao
1	2023	1011010	Cavalos reprodutores de raca pura
2	2023	1011090	Animais da especie asinina/muar/ reprodutores/ raca pura
3	2023	1011100	Cavalos reprodutores/ de raca pura
4	2023	1011900	Outros cavalos/ vivos
5	2023	1012000	Asininos e muares vivos
6	2023	1012100	Cavalos reprodutores de raca pura
7	2023	1012900	Cavalos/ exceto reprodutores de raca pura
8	2023	1013000	Animais especie asinina/muar/ reprodutores de raca pura
9	2023	1019000	Animais especie asinina/muar/ exceto reprodutores de raca pura
10	2023	1019010	Cavalos vivos/ exceto reprodutores raca pura

- **uf** é uma tabela Dimensão e nela temos as colunas “UF” e “Nome”, com as seguintes características:
 - UF – Unidade Federativa ou cluster exportadora. Os dados são compostos apenas por string com dois caracteres, que representam as siglas dos respectivos estados brasileiros. Essa coluna funcionará como chave primária (PK) para se relacionar com a tabela fato “bd”;
 - Nome – a descrição ou nome da Unidade Federativa ou cluster exportadora. Os dados são uma string, sendo limitado ao que ela realmente significa e é classificada pelo Governo Federal;

Visualização de dados

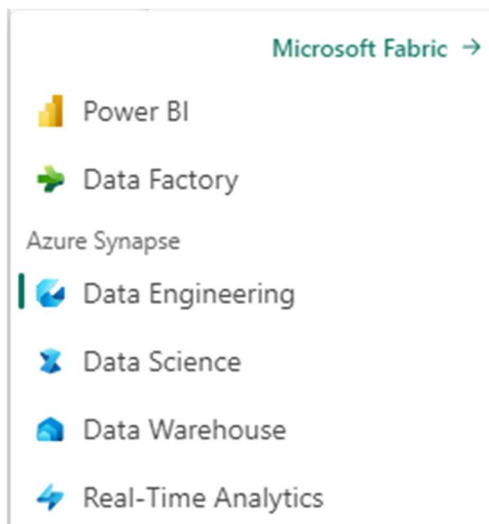
	ABC UF ↑	ABC Nome
1	SC	Santa Catarina
2	SP	Sao Paulo
3	ND	Nao Declarada
4	PR	Parana
5	MA	Maranhao
6	SE	Sergipe
7	RE	Reexportacao
8	BA	Bahia
9	RS	Rio Grande do Sul
10	AL	Alagoas

De modo geral todos os dados presentes são nos relatórios não são muito complexos, basta apenas que se tenha conhecimento do assunto para saber apresentá-los de forma objetiva, gerando valor para os mesmos. Os dados possuem acesso restrito, pois são armazenados num workspace com controle total de acesso por parte do owner do workspace. Esse acesso só é liberado mediante aprovação prévia, garantindo assim uma governança e integridade maior dos mesmos. Por fim, com os dados todos carregados, passando pelo processo de transformação dentro do SQL, originando o conjunto de dados e assim sendo a base para elaboração do relatório final.



4 – Carga

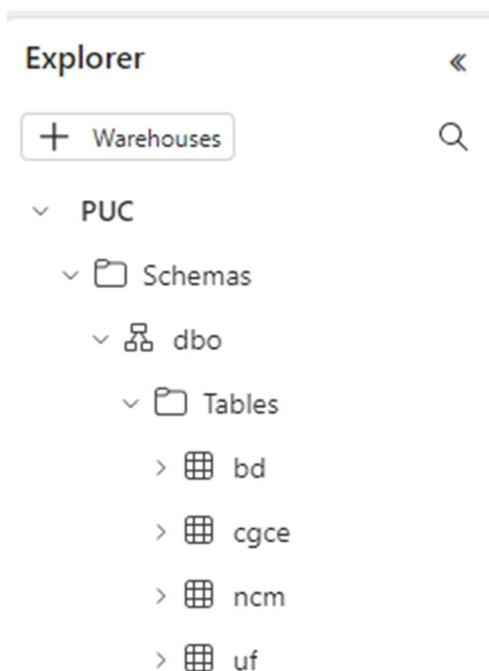
A escolha do Microsoft Fabric se deu pelo fato dela reunir todas as ferramentas num local só, tornando-a mais dinâmica e prática para uso.



Os dados foram carregados do Data Lake (Lakehouse), sendo originalmente em formato CSV para o banco de dados SQL.

Captura de tela da interface de arquivos do Microsoft Fabric. No topo, há uma barra de navegação com 'Início' e 'Lakehouse'. Abaixo, há uma barra de ferramentas com 'Obter dados', 'Novo conjunto de dados do Power BI' e 'Abrir o notebook'. O painel principal é dividido em 'Gerenciador' e 'Files'. O 'Gerenciador' mostra a estrutura de arquivos com 'PUC' selecionado. O 'Files' mostra uma tabela com os seguintes dados:

Nome	Data da modificação	Tipo	Tamanho
BD.csv	9/19/2023 9:55:08 PM	CSV	5 MB
CGCE.csv	9/12/2023 9:41:38 PM	CSV	1 KB
NCM.csv	9/22/2023 10:36:43 PM	CSV	1 MB
UF.csv	9/12/2023 9:41:38 PM	CSV	547 B

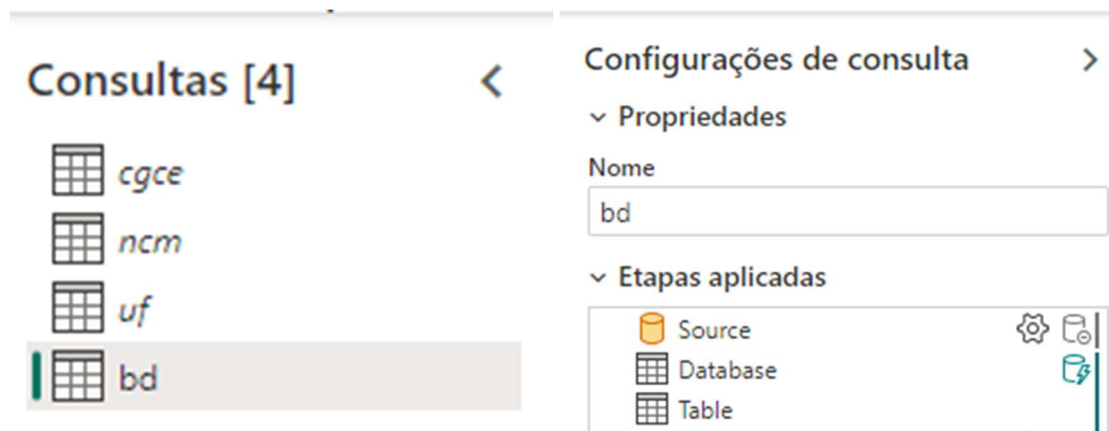


O processo de ETL foi todo feito dentro da própria plataforma, tanto através do Power Query, quanto através do SQL, por meio de consultas. Em ambos os casos o resultado será sempre o mesmo, independente da ferramenta

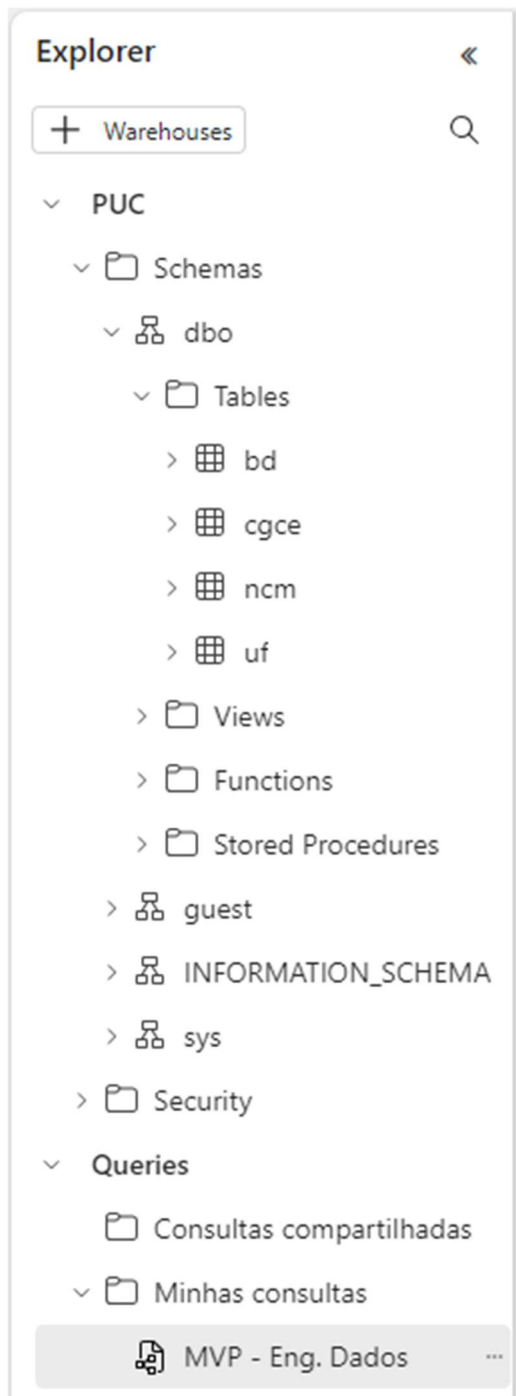
- OBS – Power Query anteriormente era a ferramenta de ETL por trás apenas do Power BI, porém posteriormente ela já foi também introduzida ao Excel e agora no Fabric, como opção ao SQL. Nesse caso serão demonstrados como as etapas acontecem em ambas as ferramentas, haja vista que essa utilização do Power Query como ferramenta geral de ETL é muito recente e tampouco foi demonstrada nas aulas. Além disso ela é uma ferramenta muito mais utilizadas por aqueles que já tem prática no Power BI.

Na etapa de Extração (Extract), os dados foram carregados do Data Lake PUC e carregados em forma de tabela para o banco de dados do SQL no Data Warehouse:

- Power Query



- SQL



Na etapa de Transformação (Transform), o primeiro passo foi utilizar a tabela fato bd e trazer as demais informações que serão utilizadas para análise final das tabelas dimensão cgce, ncm e uf, ficando assim:

- Power Query

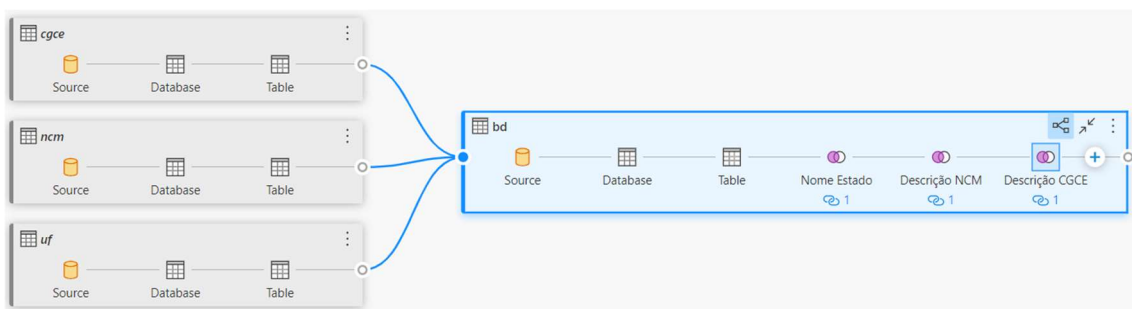
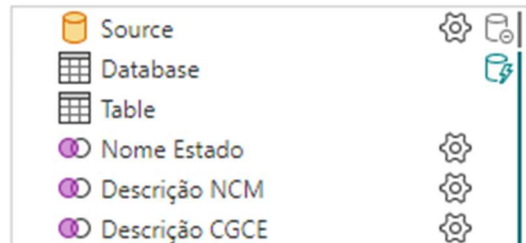
Configurações de consulta >

▼ Propriedades

Nome

bd

▼ Etapas aplicadas



• SQL

```
1 select [$Table].[ANO] as [ANO],
2      [$Table].[ESTADO] as [ESTADO],
3      [$Table].[NCM] as [NCM],
4      [$Table].[CGCE] as [CGCE],
5      [$Table].[USD] as [USD]
6 from [PUC].[dbo].[bd] as [$Table]
```

• Resultado

	123 ANO	AB ESTADO	123 NCM	123 CGCE	123 USD	uf	ncm (2)	cgce (2)
1	2022	SP	27101259	420	201032216	[Table]	[Table]	[Table]
2	2022	RJ	27101259	420	180301562	[Table]	[Table]	[Table]
3	2022	RS	27101259	420	16437837	[Table]	[Table]	[Table]
4	2022	AM	27101259	420	4094965	[Table]	[Table]	[Table]
5	2022	SC	27101259	420	590	[Table]	[Table]	[Table]
6	2022	ND	27101259	420	560	[Table]	[Table]	[Table]
7	2022	PR	27101259	420	443	[Table]	[Table]	[Table]
8	2022	CE	27101259	420	213	[Table]	[Table]	[Table]
9	2022	PA	27101259	420	166	[Table]	[Table]	[Table]
10	2022	ES	27101259	420	161	[Table]	[Table]	[Table]

Como visto no resultado acima, na tabela bd agora temos todos os dados necessários para realização da análise. Todavia, esses novos dados apresentaram alguns problemas, tanto no título das colunas que são pouco intuitivos, como no formato dos dados que foram carregados como tabelas importadas. Sendo assim, mais uma etapa de Transformação precisou ser aplicada, para que fossem corrigidos, ficando assim:

- Power Query

Configurações de consulta >

▼ **Propriedades**

Nome
bd

▼ **Etapas aplicadas**

- Source
- Database
- Table
- Nome Estado
- Descrição NCM
- Descrição CGCE
- Expandido uf
- Expandido ncm (2)
- Expandido cgce (2)
- Colunas renomeadas
- Colunas reordenadas

Visual step navigator showing the sequence of steps: Source, Database, Table, Nome Estado, Descrição NCM, Descrição CGCE, Expandido uf, Expandido ncm (2), Expandido cgce (2), Colunas renomeadas, and Colunas reordenadas.

- SQL

```

1 select [$Outer].[ANO] as [ANO],
2      [$Outer].[ESTADO] as [UF],
3      [$Outer].[Nome] as [ESTADO],
4      [$Outer].[NCM] as [NCM],
5      [$Outer].[Descricao] as [DESCRIÇÃO NCM],
6      [$Outer].[CGCE] as [CGCE],
7      [$Inner].[Descricao2] as [DESCRIÇÃO CGCE],
8      [$Outer].[USD] as [USD]
9 from
10 (
11     select [$Outer].[ANO],
12           [$Outer].[ESTADO],
13           [$Outer].[NCM],
14           [$Outer].[CGCE],
15           [$Outer].[USD],
16           [$Outer].[Nome],
17           [$Inner].[Descricao]
18     from
19     (
20         select [$Outer].[ANO],
21               [$Outer].[ESTADO],
22               [$Outer].[NCM],
23               [$Outer].[CGCE],
24               [$Outer].[USD],
25               [$Inner].[Nome]
26         from [PUC].[dbo].[bd] as [$Outer]
27         left outer join [PUC].[dbo].[uf] as [$Inner] on ([$Outer].[ESTADO] = [$Inner].[UF] or [$Outer].[ESTADO] is null and [$Inner].[UF] is null)
28     ) as [$Outer]
29     left outer join

```

```

30      (
31          select [_.Atualizado] as [Atualizado],
32                 [_.NCM] as [NCM2],
33                 [_.Descricao] as [Descricao]
34          from [PUC].[dbo].[ncm] as [_]
35          ) as [$Inner] on ([Outer].[NCM] = [$Inner].[NCM2] or [$Outer].[NCM] is null and [$Inner].[NCM2] is null)
36      ) as [$Outer]
37  left outer join
38      (
39          select [_.CGCE] as [CGCE2],
40                 [_.Descricao] as [Descricao2]
41          from [PUC].[dbo].[cgce] as [_]
42          ) as [$Inner] on ([Outer].[CGCE] = [$Inner].[CGCE2] or [$Outer].[CGCE] is null and [$Inner].[CGCE2] is null)

```

- Resultado

	1-3 ANO	AR UF	AR ESTADO	1-3 NCM	AR DESCRIÇÃO NCM	1-3 CGCE	AR DESCRIÇÃO CGCE	1-3 USD
1	2019	PR	Parana	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	197301
2	2021	PR	Parana	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	69303
3	2018	PR	Parana	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	540950
4	2020	RS	Rio Grande do Sul	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	157874
5	2021	RS	Rio Grande do Sul	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	497990
6	2022	PR	Parana	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	53700
7	2019	RS	Rio Grande do Sul	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	5357274
8	2020	PR	Parana	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	103170
9	2018	RS	Rio Grande do Sul	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	939270
10	2022	RS	Rio Grande do Sul	1012100	Cavalos reprodutores de raca pura	110	Bens de capital (exceto equipamentos de transporte)	525513

A não necessidade de outros ajustes também se deu muito pelo fato de que os dados coletados são dados oficiais, todos já utilizados pelo Governo Federal para divulgação de relatórios periódicos. Sendo assim, os mesmos não apresentam inconsistências ou mesmo dados nulos/inválidos.

Já na etapa de Carregamento (Load) seria possível utilizar três opções:



- Salvar como exibição – cria uma tabela (view) para que seja posteriormente utilizada por alguma ferramenta de análise;
- Baixar arquivo Excel – automaticamente irá gerar um arquivo em Excel contendo a tabela final;
- Visualizar os resultados – essa opção irá redirecionar e carregar os dados diretamente para a plataforma do Power BI, possibilitando a criação de dashboards e demais análises;

Como as análises e dashboards serão todos feitos diretamente no Power BI, a opção mais simples e rápida seria “Visualizar os resultados”. Todavia, apenas a título de demonstrar os aprendizados ao longo do curso, optou-se por “Salvar como exibição”, pois assim foi criada uma view contendo tudo o que foi realizado e salvo no Data Warehouse:

- SQL

Salvar como exibição



i Isso criará uma tabela com os resultados da consulta visual com Habilitar carregamento aplicado. Para usar uma consulta visual diferente, cancele esta caixa de diálogo e aplique Habilitar carregamento à consulta visual que deseja usar.

Warehouse

PUC

Esquema

INFORMATION_SCHEMA



Nome da exibição *

MVP - Eng. Dados



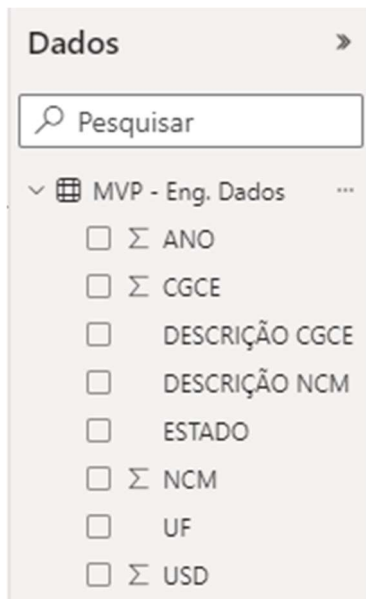
SQL para exibição

```

CREATE VIEW [INFORMATION_SCHEMA].[MVP]
AS
select [$Outer].ANO as [ANO],
       [$Outer].ESTADO as [UF],
       [$Outer].Nome as [ESTADO],
       [$Outer].NCM as [NCM],
       [$Outer].Descricao as [DESCRIÇÃO NCM],
       [$Outer].CGCE as [CGCE],
       [$Inner].Descricao2 as [DESCRIÇÃO CGCE],
       [$Outer].USD as [USD]
from
(
    select [$Outer].ANO,
           [$Outer].ESTADO,
           [$Outer].NCM,
           [$Outer].CGCE,
           [$Outer].USD,
           [$Outer].Nome,
           [$Inner].Descricao
    from
    (
        select [$Outer].ANO,
               [$Outer].ESTADO,
               [$Outer].NCM,
               [$Outer].CGCE,
               [$Outer].USD,
               [$Inner].Nome
        from [PUC].[dbo].[bd] as [$Outer]
        left outer join [PUC].[dbo].[uf] as [$Inner] on
        ([$Outer].ESTADO = [$Inner].UF or [$Outer].ESTADO is null and
        [$Inner].UF is null)
    ) as [$Outer]
    left outer join
    (
        select [_.Atualizado] as [Atualizado],
               [_.NCM] as [NCM2],
               [_.Descricao] as [Descricao]
        from [PUC].[dbo].[ncm] as [_]
    ) as [$Inner] on ([$Outer].NCM = [$Inner].NCM2 or [$Outer].NCM
    is null and [$Inner].NCM2 is null)
    ) as [$Outer]
    left outer join
    (
        select [_.CGCE] as [CGCE2],
               [_.Descricao] as [Descricao2]
        from [PUC].[dbo].[cgce] as [_]
    ) as [$Inner] on ([$Outer].CGCE = [$Inner].CGCE2 or [$Outer].CGCE
    is null and [$Inner].CGCE2 is null)

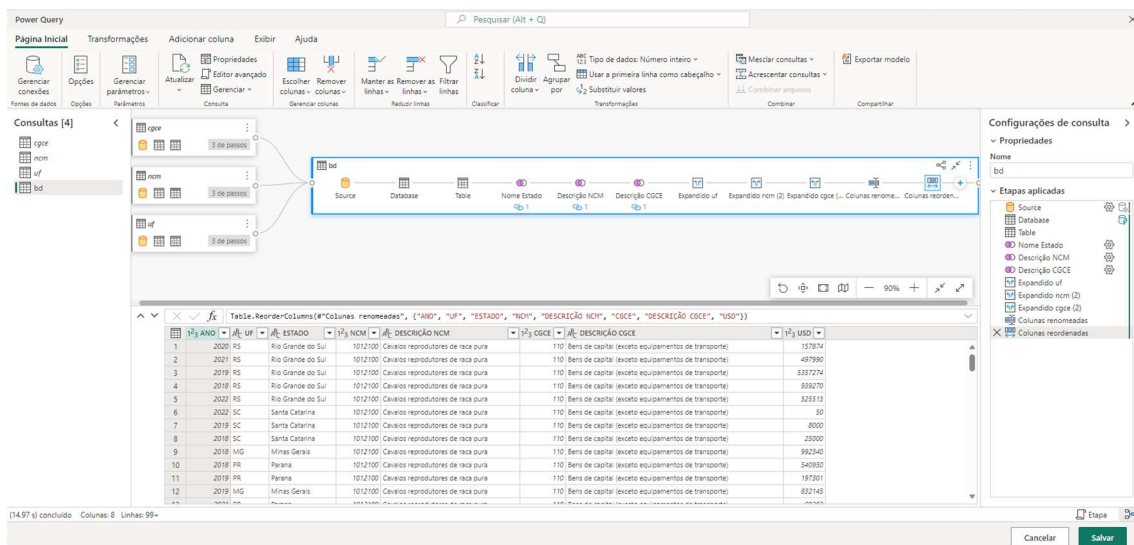
```

- Resultado no Power BI



- OBS – nesse caso foi realizado apenas no SQL, pois no caso do Power Query seria feito através do “Visualizar os resultados”, resultando na mesma view para ambos os casos

Essa é visão geral do ambiente do Power Query, com todas as etapas realizadas:



Uma outra opção dentro do Power Query seria utilizar o Editor Avançado, que seria uma outra opção para criação do processo de ETL e também acompanhamento das etapas realizadas:



Assim ficou o detalhamento final de todo o processo, desde o Data Lake até o Relatório Final.



5 – Análise

A – Qualidade dos dados

Conforme já detalhado anteriormente ao longo das demais etapas, os dados utilizados nesse projeto são dados reais e não apresentaram problemas, uma vez que são dados já utilizados pelo Governo Federal para divulgação de relatórios periódicos, sem que haja inconsistências ou até mesmo dados nulos/inválidos. Pode ser que antes de serem disponibilizados para o público em geral eles passem por diversas etapas de transformação para validação, uma vez que a coleta dos dados é feita através de relatórios que são obrigatoriamente enviados por todas as empresas exportadoras e posteriormente confrontados com os dados do Siscomex, que é a plataforma onde todas as exportações são registradas. Apesar dos campos de preenchimentos dos relatórios serem em sua maioria preenchido de forma automática, essa etapa é feita por seres humanos, sujeitos a falhas. Falo isso com muita propriedade sobre isso, pois sou o responsável da minha empresa para o envio dessas informações.

Caso ao longo do processo fosse detectado algum tipo de inconsistência, em algum dos dados, os mesmos seriam identificados através ao longo do ETL e seriam ajustados ou até mesmo descartados. Isso devido ao fato de que eles poderiam comprometer o que se quer apresentar como solução para os problemas levantados.

B – Solução do Problema

Exportação é um tema muito importante para economia dos países. Desde os primórdios, a “troca de mercadorias” sempre fez parte da história, inclusive o nosso país foi “descoberto” pelos portugueses numa dessas buscas por mercadorias de outros países, como sempre foi relatado nos livros de história.

A exportação é algo importante, pois ajuda a fazer com que o alcance dos produtos seja além das fronteiras locais, aumentando de forma exponencial as possibilidades de vendas. Um aspecto muito importante também é o tipo de produto exportado, pois isso reflete um pouco de como é a economia do país. Uma coisa em comum que as principais economias possuem são que elas são grandes exportadoras, principalmente de produtos com alto valor agregado e não apenas *commodities* (matéria-prima que serve de base para produção).

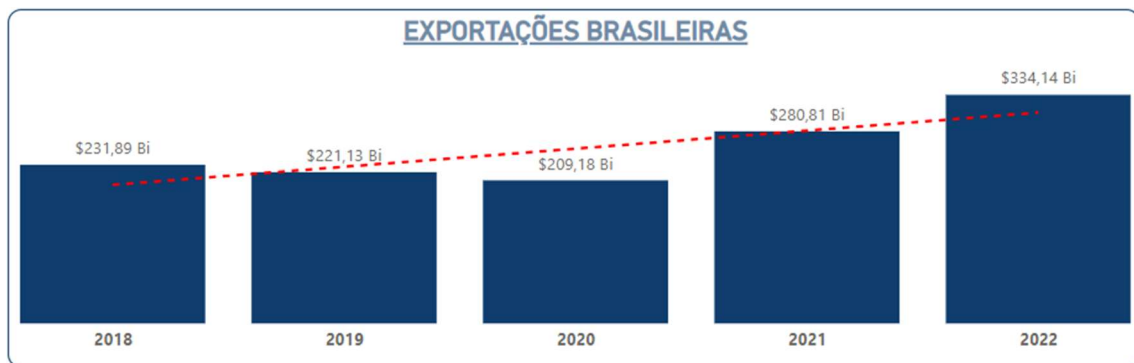
Com isso o principal objetivo é traçar um diagnóstico sobre o cenário brasileiro. O período em questão (2018 – 2022) no qual foi escolhido para essa análise é muito particular, pois em 2020 tivemos a pandemia. Sendo assim, foram escolhidos dois períodos anteriores e dois posteriores para uma análise mais ampla.

Para elaboração das análises e dashboards a ferramenta utilizada foi o Power BI. O arquivo final poderá ser visualizado através do link [MVP - Eng. Dados](#).

- Obs: o arquivo do Power BI só não foi compartilhado apenas pelo fato de não ter sido exigido, porém cadastrando os e-mails dos responsáveis, todos teriam o devido acesso.

Agora considerando os problemas inicialmente apresentados, vamos trabalhar na apresentação e desenvolvimento das respostas.

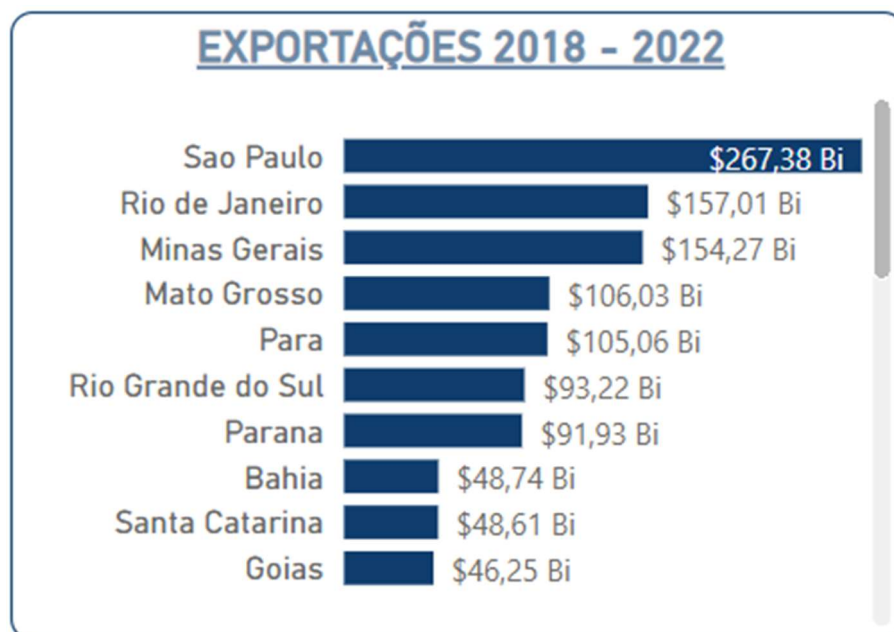
Primeiramente uma visão geral do desempenho das exportações.



- Pré-pandemia as exportações caíram cerca de 9,79%, porém desde então elas vêm crescendo. Comprando o resultado do último ano é possível verificar um crescimento de 59,74% nas exportações. A linha de tendência destacada no gráfico inclusive comprova essa afirmação.

Todos os questionamentos feitos remetem a identificação de quais são os principais estados e produtos exportados. Contudo, esse conceito de “principais” pode ser um pouco vago e com isso foi considerado uma análise Top-N, onde N=10. Com isso, os resultados apresentados levarão em consideração esse tipo de delimitador. Contudo, todos os dados poderão ser visualizados através do link já disponibilizado.

1 – Quais são os principais estados exportadores do Brasil?



- Os 10 estados acima mencionados representam cerca de 90% do total das exportações brasileiras durante o período analisado;
- São Paulo com quase 21%, logo seguido por Rio de Janeiro e Minas Gerais com pouco mais de 12% cada, formam o pódio como os três maiores do país, concentrando cerca de 45% das exportações;
 - Detalhe que esses mesmos estados são as principais economias do país, quando assunto é o PIB;

2 – Quais são os principais tipos de produtos exportados pelo Brasil?

Para análise dos principais produtos teremos dois tipos de classificação, tanto por NCM, quanto pelo CGCE, ambos termos já previamente explicados.

- NCM



- Os produtos acima representam cerca de 52% das exportações brasileiras durante o período analisado;
- O pódio com os principais tipos de produtos é formado por Soja com cerca de 13%, Óleos brutos com 11% e Minério de ferro com 10%, representando assim mais de 1/3 do total das exportações;
- Os produtos de maior destaque são TODOS produtos com baixo ou nenhum valor agregado, basicamente matéria-prima. Isso é muito ruim para economia do Brasil, pois ela é exportada para servir de base para fabricação de algum produto, que posteriormente será exportado para o Brasil;
 - Destacar que o fato de ser produzido em outro país acaba gerando empregos lá e não aqui;

- CGCE



- Os produtos acima representam cerca de 94% das exportações brasileiras durante o período analisado;
 - Destaque é que por ser um número pequenos de classificações possíveis (máximo 19), esse número de 94% já poderia ser esperado;
- O pódio é formado por Insumos industriais elaborados (24%), Alimentos e bebidas básicos (19%) e Insumos industriais básicos (15%), que representam mais da metade dos produtos exportados, com cerca de 58%.
- Destaque para produtos com um pouco mais de valor agregado aparecendo apenas da quarta posição em diante, onde temos Combustíveis (11%) e Alimentos e bebidas para consumo (9%);
 - Destaque para o combustível aparecer nessa posição, pois como destacado no relatório da NCM, onde temos óleo bruto como um dos mais exportados. Com isso, esse material é exportado para algum país,

onde é refinado e volta para o Brasil para que depois ainda consigamos exportar para outros países;

3 – Quais os principais produtos exportados dos principais estados?

- São Paulo
 - NCM



- CGCE



- Rio de Janeiro
 - NCM



- CGCE



- Minas Gerais
 - NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- Mato Grosso
 - NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- Pará
 - NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- Rio Grande do Sul

- NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- Paraná

- NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- Bahia

- NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- Santa Catarina
 - NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- Goiás
 - NCM

EXPORTAÇÕES 2018 - 2022



- CGCE

EXPORTAÇÕES 2018 - 2022



- O que podemos ver que há em comum entre os principais produtos, dos principais estados, todos eles são produtos de baixo valor agregado, assim como já visto. Isso corrobora com o que foi visto na análise dos principais produtos exportados. Em sua maioria são produtos com baixo valor agregado, sendo exportado para que sejam transformados em produtos acabados e retornem ao Brasil como um produto pronto;
- Exportações são motores que movem as principais economias globais e pensando nas próximas etapas, para um desenvolvimento melhor do país seria importante o Brasil investir em tecnologia, para aprimorar sua indústria e passar a exportar bens de capital e produtos com alto valor agregado, assim como ocorre com as principais economias do mundo;

Autoavaliação

Ao final do trabalho foi possível verificar que todos os questionamentos inicialmente levantados foram todos respondidos de forma clara e objetiva. Isso de fato já era esperado, uma vez que os questionamentos feitos eram para fazer uma análise diagnóstica do cenário das exportações brasileiras e não propriamente algo especulativo. O conceito de entender o cenário das exportações é importante, pois traz uma luz para um tema importante, amplamente discutido no meio econômico, que é o tipo de produto exportado pelo país.

Enquanto grandes potências se especializam em trabalhar com produtos tecnológicos, o Brasil cada vez mais está se distanciando disso e se tornando um país que fornece a matéria-prima, justamente para esses países. Não que exportar matéria-prima não seja importante, porém não adianta exportarmos esse tipo de produto “barato”, para depois termos que importar o produto “pronto”. Um bom exemplo é que o segundo tipo de produto mais exportado são óleos brutos de petróleo, enquanto o produto mais importado é óleo diesel. Ou seja, exportamos o óleo bruto, que é refinado e importamos para consumo. Não é que tenhamos que ser autossuficientes em tudo, o que de certo modo seria impossível, mas poderíamos focar num processo melhor e passarmos a industrializar mais nossos produtos, para a partir daí sim exportarmos.

Agora falando das etapas para elaboração do trabalho em si, a maior dificuldade foi definir em qual plataforma realizar o mesmo. Tenho experiência e prática em Power BI, porém uma vez que o ETL não poderia ser feito diretamente por lá, surgiram as dúvidas sobre como proceder. Na empresa trabalho com o SQL Server, conectando diretamente ao banco da empresa, montando as queries conforme necessidade, mas não seria possível utilizar para esse caso, pois teria que ser uma solução mais completa. A Microsoft já vinha com a ideia de unir todas as ferramentas (Azure, Data Factory, Synapse, Power BI etc.) num único lugar, tornando uma experiência mais simples e visual para os usuários. A partir daí veio o Microsoft Fabric, que será algo facilmente escalável, pois é simples, prático e muito direto, auxiliando muito aqueles que não tenham tanta experiência e conhecimento com a parte de programação, pois é intuitivo e direcionado à prática. Por fazer parte do programa de desenvolver Microsoft, tenho um período de avaliação de 60 dias para usar todos os recursos da plataforma. A questão é que como está em Preview e sou usuário como desenvolver, esses “60 dias” são renovados diariamente.

Com a plataforma escolhida, chegou a hora de buscar as bases para que fosse possível responder aos questionamentos realizados na etapa “Objetivo”. Para tal foi necessário acessar o portal onde o governo reúne as informações sobre comércio exterior e extrair os dados nos quais queria se analisar. Além disso, foram necessárias bases auxiliares para servir de suporte a base dados das exportações, pois nela continham informações detalhadas sobre o nome dos estados e descrição, tanto da NCM, quanto do CGCE. Ao longo do trabalho as maiores dificuldades foram apenas entender o que estava sendo solicitado, para justamente conseguir relatar todos os pontos, evitando assim esquecer de relatar e documentar algo. Acredito que dois fatores foram de extrema importância para que não tenha tido tantas dificuldades assim para o trabalho. O primeiro foi a plataforma escolhida, pois já estava bem habituado com a interface da Microsoft e juntando tudo num único lugar, o desenvolvimento foi bem rápido. O segundo fator foi que a base escolhida foi extraída de um portal oficial do governo, onde com toda certeza os dados passam por diversos filtros/camadas/revisões, antes de serem divulgados, pois são dados oficiais do governo. A junção desses dois fatores fez com que as etapas de ETL fossem bem rápidas e sem grandes problemas para solução do problema.

Por atuar já com análise de dados na área de comércio exterior, tem diversos outros tópicos a serem abordados. Na parte de exportação podemos trabalhar para identificar quais seria os principais parceiros comerciais do Brasil, traçando uma análise evolutiva num determinado período. Outro ponto relevante poderia ser analisar também as importações brasileiras e identificar o comportamento delas, assim como nossos principais parceiros. Ao final, poderia juntar ambos e fazer uma análise da balança comercial, que nada mais é que comparar importação x exportação, até mesmo detalhando por país. Enfim, o comércio exterior é algo no qual possibilita inúmeras outras análises diagnósticas, quiçá partir para análise preditivas e prescritivas também.