

57^a REUNIÃO



COGEF

COMISSÃO DE GESTÃO FAZENDÁRIA

13 a 15 de Setembro - 2023

MANAUS - AMAZONAS



Realização:



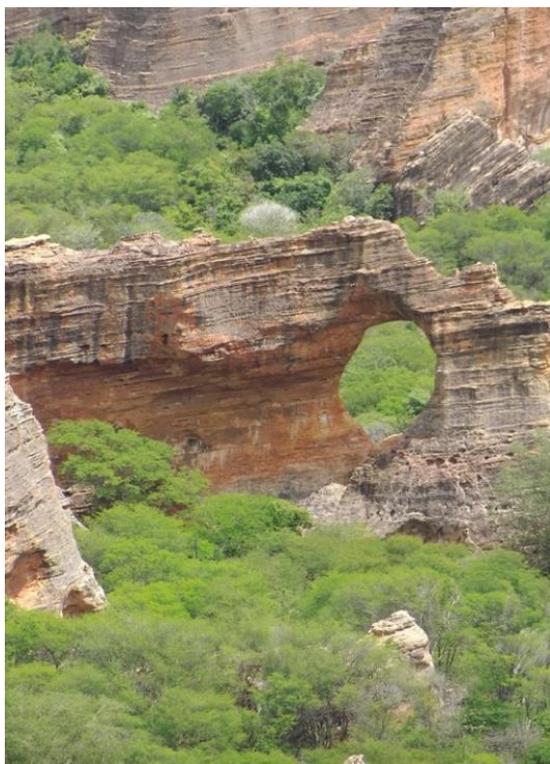
Secretaria de
Cultura e Economia
Criativa





Piauí
GOVERNO DO ESTADO

FAZENDA
Secretaria de Estado
da Fazenda / SEFAZ



SECRETÁRIA DE ESTADO DA FAZENDA DO PIAUI

Pauta Fiscal Automatizada

Big Data Analytics

Projeto:

- Realizado pela empresa EloGroup (<https://elogroup.com.br/>)
- Acompanhamento da Equipe da UNITEC/GEPRO
- Acompanhamento da Equipe UNATRI

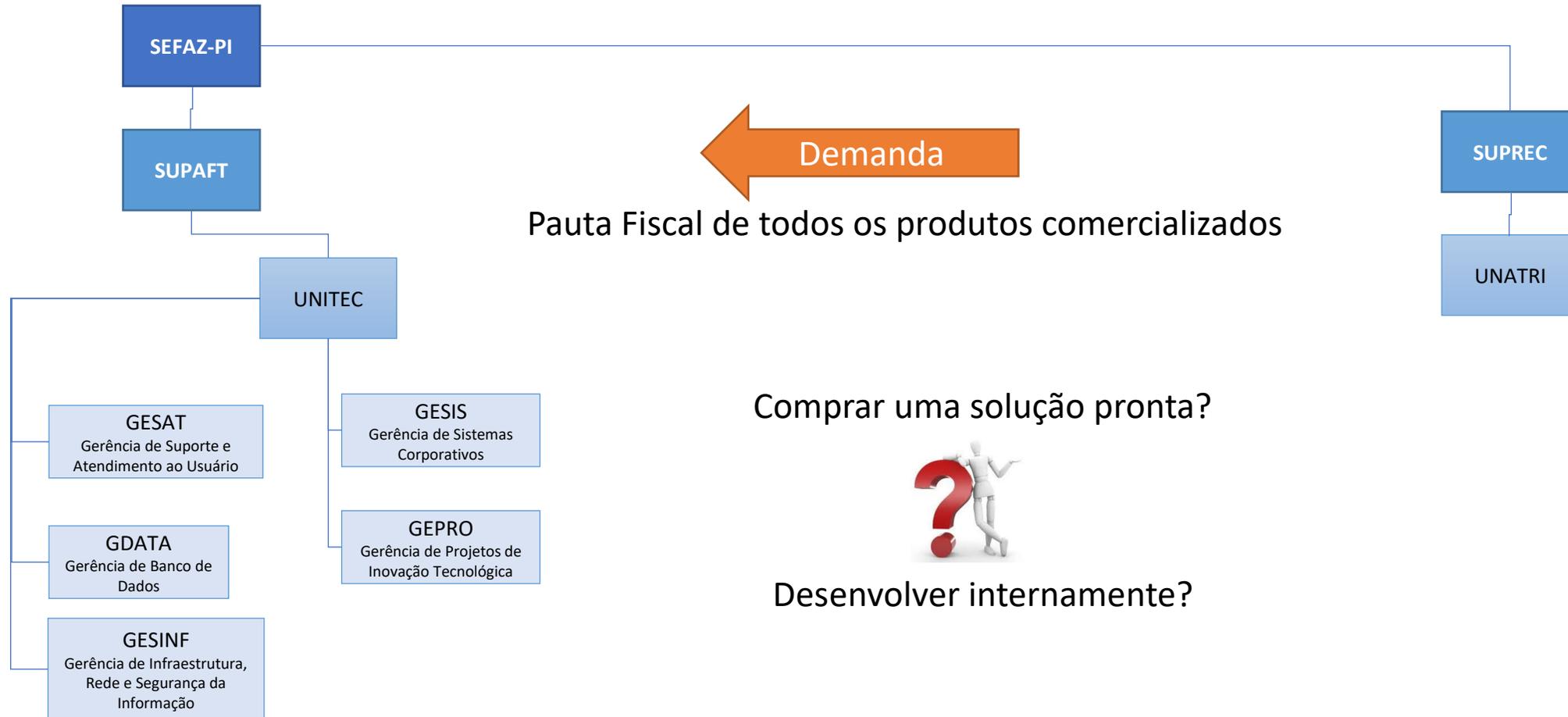
Contextualização do Projeto



Piauí
GOVERNO DO ESTADO

FAZENDA
Secretaria de Estado
da Fazenda / SEFAZ

57ª COGEF
Comissão de Gestão Fazendária



Definição do Projeto



Pauta Fiscal Automatizada

Oportunidade:

Determinar pauta fiscal por meio do uso das informações dos documentos fiscais eletrônicos de venda ao consumidor (NFC-e).

Definido a partir do acompanhamento dos resultados-chave no projeto “**Gestão para Resultados e Riscos**” conduzido pela EloGroup na SEFAZ-PI junto a alta liderança.



Premissas para o Projeto acontecer

- I. **Disponibilidade e acesso aos dados** em ambiente produção & desenvolvimento
- II. **Provisionamento de infraestrutura** (máquinas físicas ou virtuais) seguindo especificação técnica recomendada
- III. **Implantação de uma arquitetura de Big Data na Fazenda** que seria realizado de acordo com o caso de uso (desafio) específico
- IV. As entregas seriam baseadas no **modelo *scrum*** com definição de escopo (desafio) ao longo de cada *sprint*
- V. O projeto seria **construído de forma virtual**
- VI. A **transferência de conhecimento** seria focada na **instalação e suporte da arquitetura com base no caso de uso**



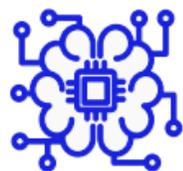
Escopo do Projeto e Produtos



Produto 01

1. Implantação de Arquitetura Big Data

- Foi Implantado ambiente de Big Data com uso de componentes *open source*, para atender ao caso de uso: Elaboração de Pauta Fiscal utilizando-se a Nota Fiscal Eletrônica do Consumidor (NFC-e)



Produto 02

2. Algoritmo para definição de Pauta Fiscal com uso dos DF-e

- Foi desenvolvido algoritmo para definição de Pauta Fiscal de Cervejas e Cervejas sem Álcool por meio de uso de notas fiscais eletrônicas ao consumidor para um produto selecionado
- Houve a transferência de conhecimento em relação à metodologia, modelagem e customização do algoritmo



Abordagem de Implantação da Arquitetura e Solução de Caso de Uso

01.



Visão & Implantação
Arquitetura Big Data

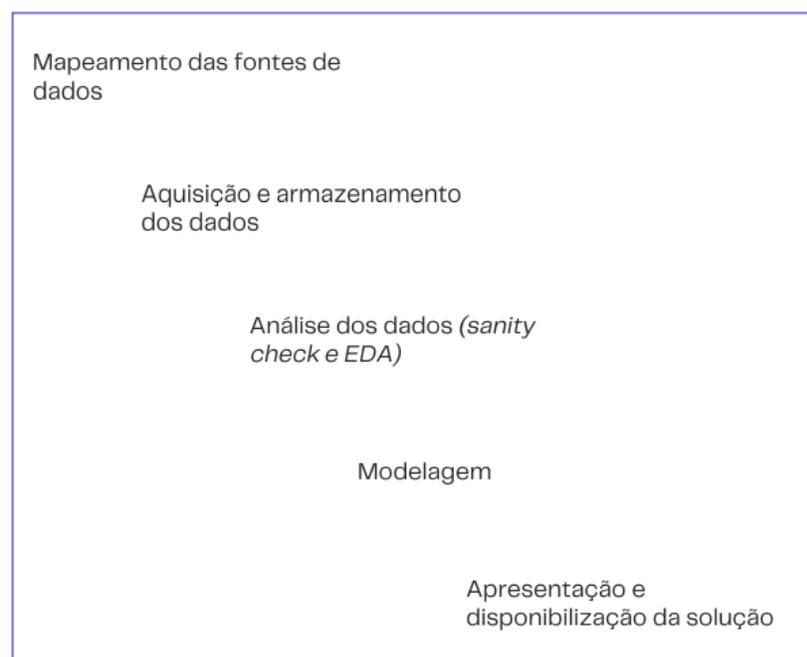
- Diagnóstico de **bases de dados e sistemas**
- Criação **da visão da solução** para guiar as equipes no projeto
- **Implementação e homologação**
- **Desenho e documentação**

02.

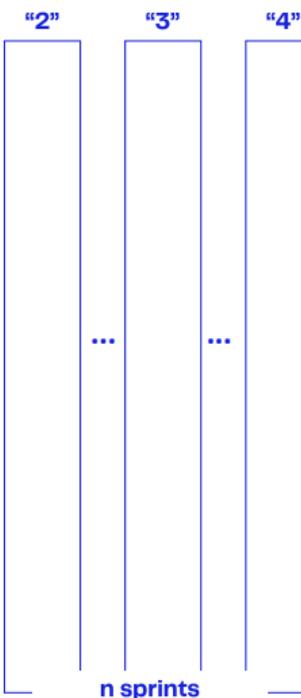


Squad de Delivery

 **Algoritmo para definição de Pauta Fiscal do Produto "A" com uso dos DF-e | Escopo do Projeto: Caso de Uso "1"**



Caso de Uso



03.

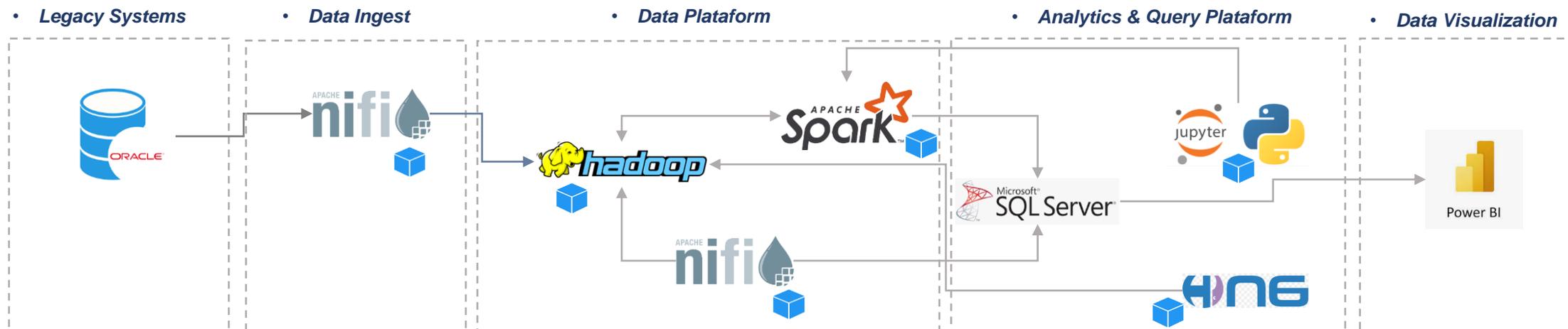


Operação Assistida

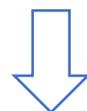
- **Gestão da mudança**
- **Comunicação Interna**
- **Suporte as soluções implementadas**



Arquitetura de Big Data implantada



- **Apache Nifi** – Fluxo de dados (ETL)
- **Apache Hadoop** – Armazenamento Distribuido
- **Apache Saprk** – Processamento Distribuido
- **Jupyter e PySpark** – Interface e linguagem utilizada
- **Hue** – Visualização da estrutura de arquivos Hadoop
- Apache Pinot – Banco relacional para bigdata >>> **SQL Server (DW)**
- Apache Superset – Dashboard >>> **Power BI**



1. Ambiente Desenvolvimento



2. Ambiente Produção

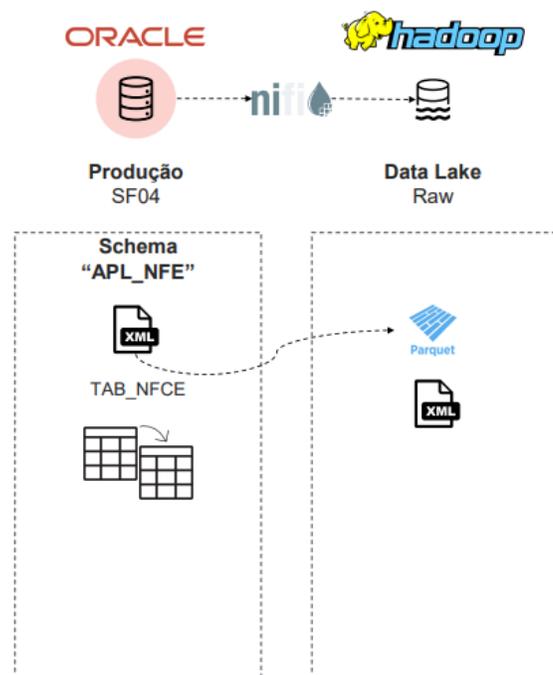


- **Jenkins** – ferramenta de integração contínua de código (pipeline)



Ingestão de Dados das Notas Fiscais (NFC-e)

--- Ingestão de dados XML ---



--- Fluxo de Dados ---



Caso de Uso Pauta de Cervejas

Foi definido a Pauta de **Cervejas e Cervejas Sem Álcool**, através dos NCMs 22030000 e 21910000, por meio do uso de notas fiscais ao consumidor (NFC-e). Houve duas grandes etapas: **Clusterização e Cálculo de Preços dos Produtos**



Na clusterização, foram utilizados principalmente o GTIN e a descrição do produto, para **criar grupos representativos dos produtos**



Construído um dicionário de referência para criação dos grupos



Foram **testadas hipóteses** relativas a precificação



Analizadas **diferenças de preenchimento** entre Bares, Restaurantes e Atacarejo



Com os grupos definidos, na precificação foi utilizado os valores e quantidades unitárias disponíveis na nota para **determinar a Pauta para cada grupo representativo**



Foram **geradas várias versões do modelo** e analisados os resultados



Caso de Uso Pauta de Cervejas

Inicialmente foi carregado um ano de notas fiscais para **Análise Exploratória** dos produtos e sua precificação e decisão quanto a modelagem, tanto de texto quanto de preço. Também definido uma janela dos últimos 3 meses para determinação de cada pauta.

- **Observar o preenchimento do campo GTIN** e entender se o agrupamento era possível



1. GTIN

- **Se a descrição do texto era preenchida de maneira estruturada**, de acordo com o contribuinte que faz a venda para ser possível agrupar por produto



2. Estrutura da Descrição

3. Principais Contribuintes

4. Análise de Preço e Outliers

5. Marcas e Volumes

- **Preciso entender os campos de quantidade, valor e data de emissão** para definir a melhor maneira de fazer o cálculo da Pauta para um período



6. Preço

7. Quantidade

8. Data de Emissão

Tivemos algumas **Descobertas**:

O GTIN cobria uma boa parte das transações

Aproximadamente 2 em cada 3 transações possuía GTIN (68% da base)

Contribuintes com maiores volumes de transações na base preenchem o GTIN

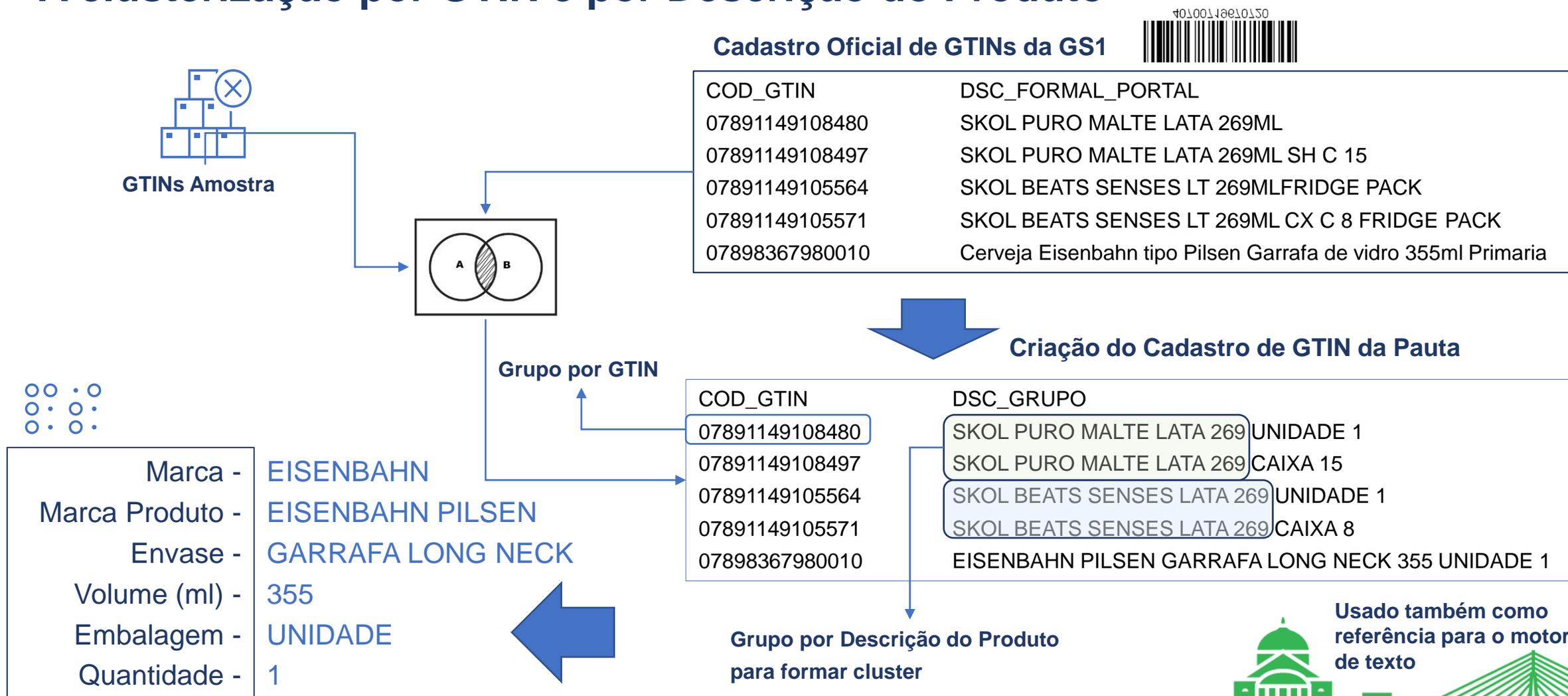
- 639 CNPJs (18,95% dos contribuintes) concentravam 90% das transações
- Existia uma recorrência no preenchimento e no não preenchimento de GTINs
- A distribuição do % de preenchimento era semelhante entre os grandes contribuintes e os demais

A quantidade de grupos gerados via GTIN reduz a base de maneira significativa

897 GTINs concentram 99% das transações



A clusterização por GTIN e por Descrição do Produto



Estrutura final na metodologia de calculo do Pauta Fiscal

Clusterização/Agrupamento

Precificação

Agregação por GTIN

Consumo de tabela referência

Algoritmo de Distribuição

Consumo das notas sem GTIN

Calculo do Pauta



Obtenção de notas com GTIN preenchido.



Identificar notas sem GTIN preenchido



Expurgo de GTINs não-conformes (fora do NCM, prefixo errado)

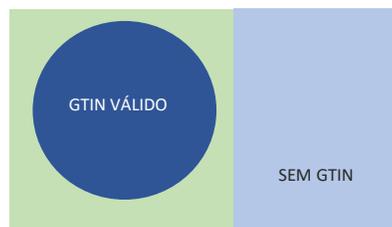


Tabela treino com a descrição correta para cada GTIN e informações complementares como envase, volumetria, marca.



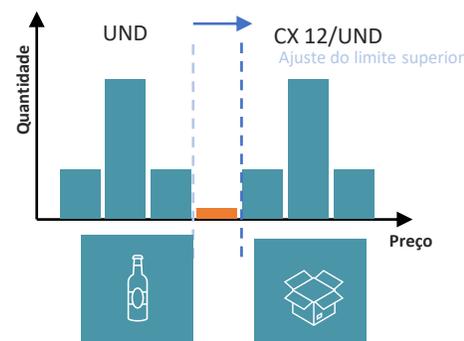
Agregação de GTINs que representam o mesmo produto em unidades diferentes (caixa, item, pacote) a partir da tabela referência



Separa por preço unitário as várias quantidades em grupos individuais.



Corrige o limite superior da quebra da unidade utilizando o dobro do valor da média do SKU unitário.



Inclusão das transações sem GTIN preenchidos nos clusters individuais já identificados utilizando **motor de texto e a lógica fuzzy**.



Criar tabela com transações sem GTIN que não conseguiram ser identificadas (expurgo)



Dentro do grupo analisaremos o volume de vendas e o valor unitário.



Calculo da **Media Ponderada** por GTINs. Campos Quantidade Vendida e Valor Unitário.

$$\bar{x} = \frac{\sum_{i=1}^n (x_i \times p_i)}{\sum_{i=1}^n p_i}$$

GTIN 2131231		R\$5,50
GTIN 2131232		R\$32,50



Metodologia do motor de texto na solução de clusterização para a Pauta

Clusterização

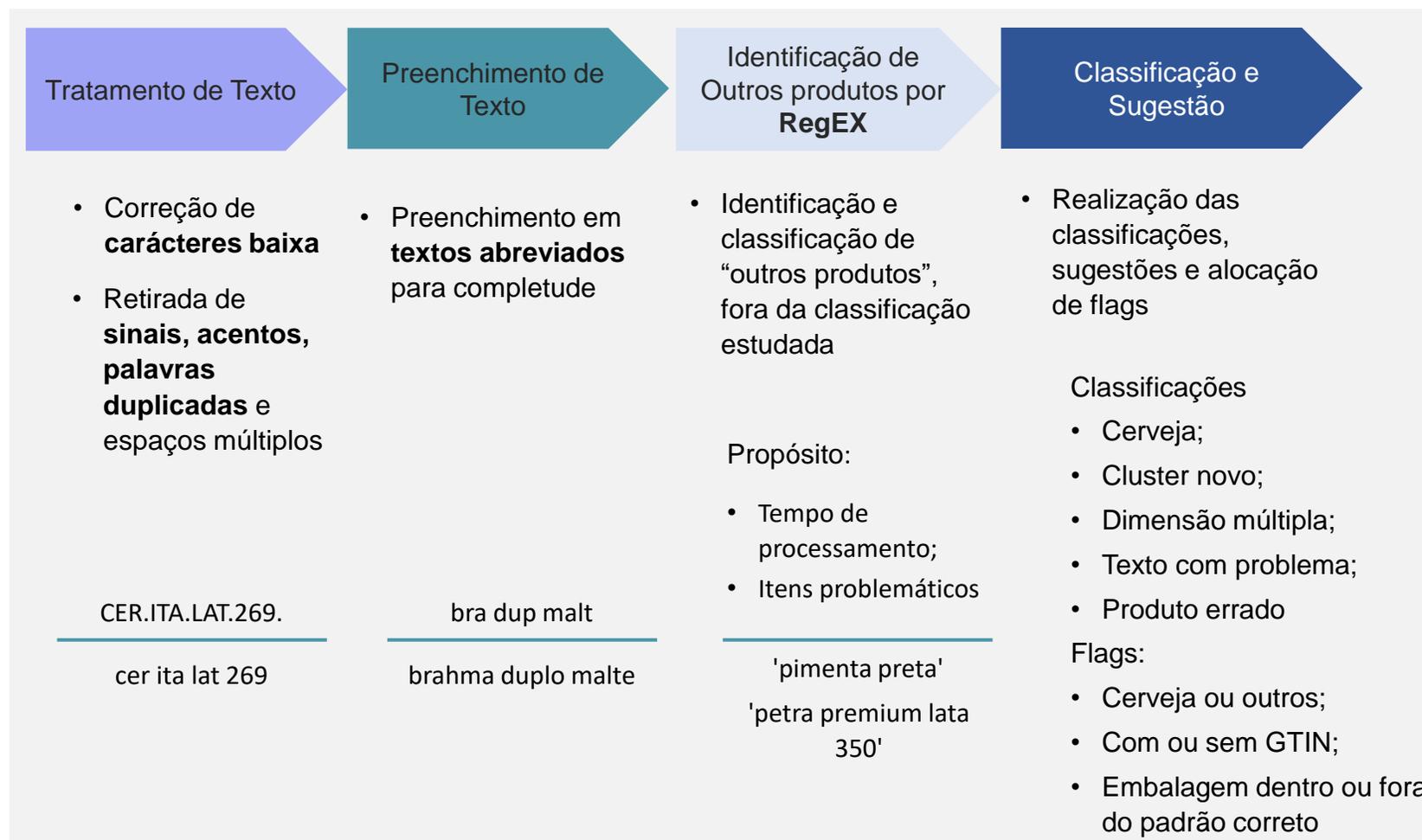
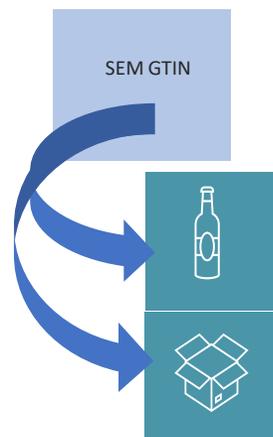
Consumo das notas sem GTIN



Distribuir as notas sem GTIN preenchidos nos grupos individuais já identificados utilizando motor de texto e a lógica fuzzy.



Criar tabela com transações sem GTIN que não conseguiram ser identificadas (expurgo)



RegEx : Expressões regulares



Exemplos de transações sem GTIN incluídas nas amostra e seus clusters

Cluster/Grupo	prod_xProd	sugestao_gtins		
BRAHMA DUPLO MALTE LATA 310 UNIDADE 1	BRA.DUP.MAL.LAT.310.	07891991299084	1	
	BRAHMA DUPLO MALTE 310ML	07891991299084	1	
	BRAHMA DUPLO MALTE LATA 310 mlcx 12	07891991299084	1	
	BRAHMA DUPLO MALTE LATA 310ML	07891991299084	1	
	BRAHMA DUPLO MALTE LATA 310ML CX C/	07891991299084	1	
	BRAHMA DUPLO MALTE LATA 310ML CX C/6 MU 11528	07891991299084	1	
	BRAHMA DUPLO MALTE LATA 310ML CX C/6 MULTPACK	07891991299084	1	
	BRAHMA DUPLO MALTE LATA 310ML CX C6 MULTPACK	07891991299084	1	
	BRAHMA DUPLO MALTE LATA 310ML MULTPACK	07891991299084	1	
	BRAHMA DUPLO MALTE LT 310ML	07891991299084	1	
	CER.BRA.DUP.MAL.LT 310.	07891991299084	1	
	CERVEJA BRAHMA DUPLO MALTE 310ML	07891991299084	1	
	CERVEJA BRAHMA DUPLO MALTE LATA 310ML	07891991299084	1	
	CERVEJA BRAHMA LATA P.MALTE	07891991299084	1	
		dtype: int64		

Em março de 2023, por exemplo, 80% das **transações sem GTIN** foram incorporadas ao modelo de **cálculo da Pauta** pelo motor de texto (1648978 linhas usadas e 408117 não utilizadas)



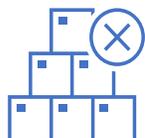
Exemplos de transações sem GTIN não incluídas

prod_xProd	sugestao_gtins			
BEBIDA CERVEJA BRAHMA DUPLO MALTE 3	07891991299084	1	BRAHMA DUPLO MALTEE	07891991299084 1
BRAH DUP MAL/BUDWEIS	07891991299084	1	BRAHM DUPLO MALTE	07891991299084 1
CERV. BRAHMA DUPLO MALTE	07891991299084	1	BRAHMA D.MALTE	07891991299084 1
CERV. DUPLO MALTE BRAHMA	07891991299084	1	BRAHMA DUPLO M.	07891991299084 1
CERV.BRAHMA DUPLO MALTE	07891991299084	1	BRAHMA DUPLO MALTE	07891991299084 1
CERVEJA BRAHAMA DUPLO MALTE	07891991299084	1	BRAHMA DUPLO MALTE UND	07891991299084 1
CERVEJA BRAHMA DUPLO MALT	07891991299084	1	BRAHMA DUPLO MALTE 1 (unid)	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE	07891991299084	1	BRAHMA DUPLO MALTE CX	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE COM 12U	07891991299084	1	BRAHMA DUPLO MALTE LATA	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LAT	07891991299084	1	BRAHMA DUPLO MALTE LATINHA	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LATA	07891991299084	1	BRAHMA DUPLO MALTE LT	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LATA SLE	07891991299084	1	BRAHMA DUPLO MALTTE	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LT C/12 G	07891991299084	1	CERV BRAHMA DUP MALTE LT12	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LT C/12 Q	07891991299084	1	BRAMA DUPLO MALTE	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LT G	07891991299084	1	Brahma Dupla Malte	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LT UND	07891991299084	1	Brahma Duplo Malte	07891991299084 1
CERVEJA BRAHMA DUPLO MALTE LTA.	07891991299084	1	Brahma Duplo Malte Latao	07891991299084 1
CERVEJA BRAHMA LATA CX	07891991299084	1	Brahma Duplo Malte Promo Sexta	07891991299084 1
			Brahma Lata Duplo Malte	07891991299084 1
			Brahma duplo Malte	07891991299084 1
			Brahma duplo malte	07891991299084 1
			CACILDES DUPLO MALTE BRAHMA	07891991299084 1
			CAIXA DE CERVEJA BRAHMA DUPLO	07891991299084 1
			cerveja brahma duplo malte	07891991299084 1

Essas sugestões acima, se aproximam, mas não temos certeza. O principal problema é a **falta de volumetria e envase**, fazendo com que a descrição possua mais de uma sugestão de grupo para agregação.



Na homologação tivemos 3 soluções que agregaram para o resultado da Pauta



Primeira versão: Algoritmo do cálculo da pauta fiscal **somente com transações com GTIN**

(Referência pauta 11/2022)

Amostra total da pauta: 2.25 milhões

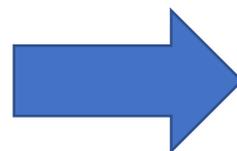
Número de produtos: 391

Preço dos SKUs destacados:

- SKOL 600: R\$6,32 | 18.573
- HEINEKEN 600: R\$10,78 | 66.296
- BRAHMA CHOPP 600: R\$5,83 | 9.059
- ANTÁRCTICA 600: R\$5,81 | 4.325



Segunda versão: Algoritmo do cálculo da pauta fiscal **com inclusão das transações sem GTIN**



Terceira versão: Algoritmo do cálculo da pauta fiscal **com alteração do limite de quebra do grupo unidades (3 X Media)**

(Referência pauta 11/2022)

Amostra total da pauta: 2.8 milhões

Número de produtos: 395

Preço dos SKUs destacados:

- SKOL 600: R\$6,32 | 39.197
- HEINEKEN 600: R\$12,74 | 170.336
- BRAHMA CHOPP 600: R\$5,83 | 11.941
- ANTÁRCTICA 600: R\$5,91 | 5.265

Durante a homologação, o algoritmo teve um acréscimo de 27% da amostra de transações



Exemplos de impacto das novas regras do limite superior

Mudança na distribuição entre caixas e unidades, aumentando o preço das unidades.

SKOL 600

Transações Antigas com GTIN:

36.041

Transações com nova regra
(incluindo as **sem GTINs** e novo
limite superior)

39.149

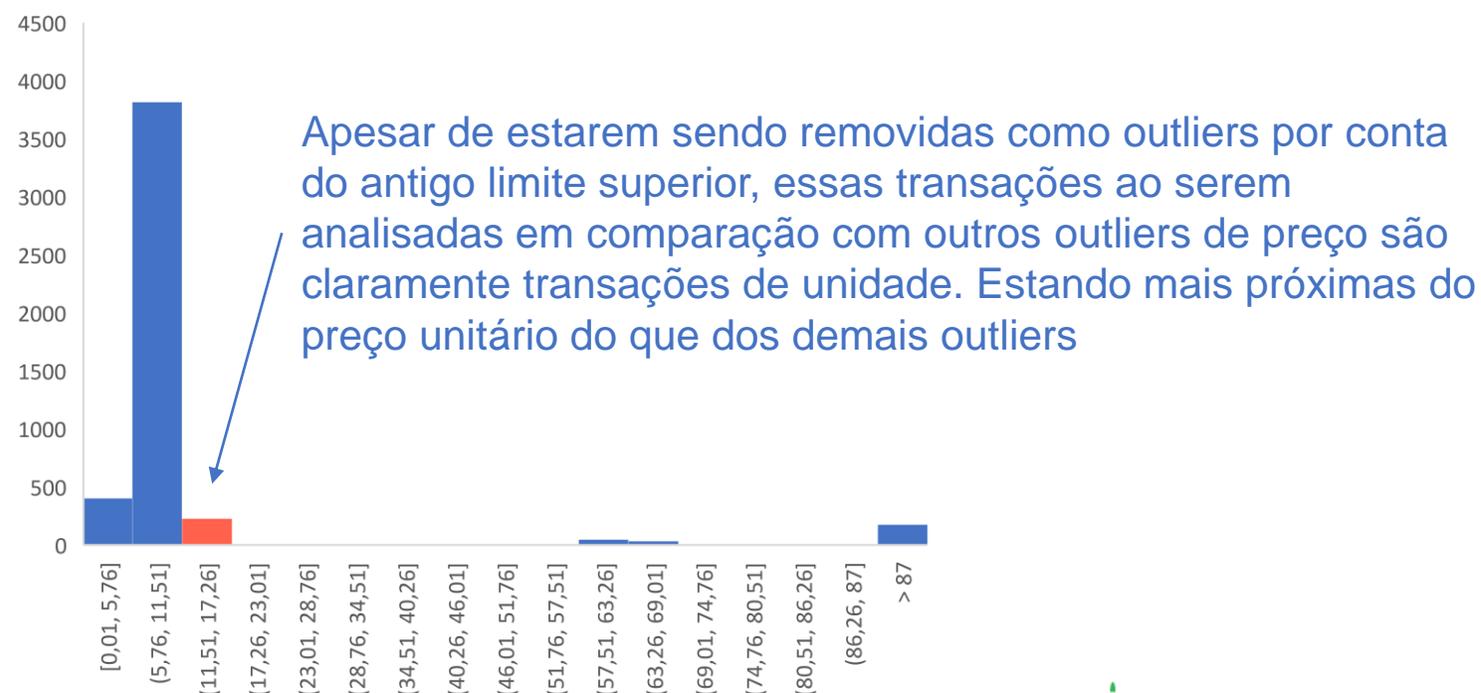
Preço regra anterior:

R\$5,83

Preço nova regra:

R\$6,32

Gráfico de distribuição dos preços



Solução Final de Engenharia de Dados

Implantação do ambiente de **Big Data** para elaboração da Pauta Fiscal utilizando-se a Nota Fiscal do Consumidor Eletrônica (NFC-e) com base no banco de dados disponibilizado pela SEFAZ-PI.



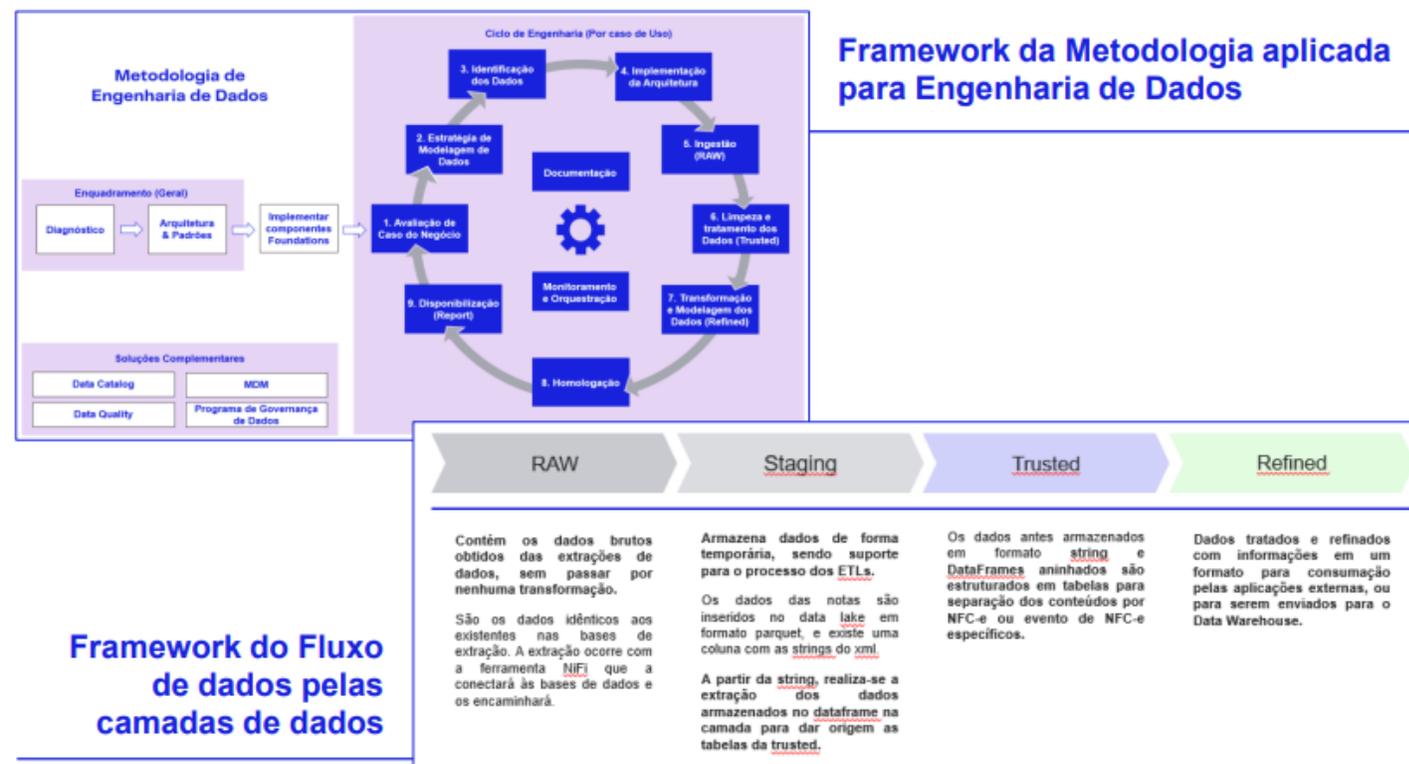
~3 dias e 16 horas para a ingestão de 365 dias de NFC-e na camada **Raw** e camada **Trusted**



Automatização no cálculo do PMPF



+ 277 milhões de notas ingeridas fora de eventos



Solução Final de Ciência de Dados

Desenvolvimento de algoritmo para definição de Pauta Fiscal por meio do uso de NFC-e para produtos **Cerveja** e **Cerveja Sem Álcool**.



+ 2 Milhões de notas únicas analisadas



+ de 405 Produtos em pauta fiscal



Padronização da descrição dos itens



~ 5 minutos para processar novas pautas



Flexibilidade para alterar o número de meses para o cálculo



+ 90% das transações representadas pelos resultados



Capacidade de notas únicas analisadas limitada apenas pela máquina



Solução Final Dashboard da Pauta

Todo Mês temos os dados da **Pauta**, referente às **transações dos Últimos 3 Meses** anteriores a Pauta

Visão Geral de PMPFs



SEFAZ
Secretaria de Estado
da Fazenda do Piauí

Pauta

202308 ▼

GTIN / SKU

Todos ▼

Descrição

Todos ▼

Pauta	GTIN / SKU	Descrição	PMPF
202308	00000078904644	KAISER PILSEN GARRAFA 600 UNIDADE 1	\$4,93
202308	00000078905276	ANTARCTICA PILSEN GARRAFA 600 UNIDADE 1	\$6,15
202308	00000078905344	BOHEMIA PURO MALTE GARRAFA 600 UNIDADE 1	\$7,94
202308	00000078905351	ANTARCTICA ORIGINAL GARRAFA 600 UNIDADE 1	\$10,02
202308	00000078905498	HEINEKEN GARRAFA 600 UNIDADE 1	\$13,49
202308	00000078906709	ITAIPAVA PILSEN GARRAFA 600 UNIDADE 1	\$5,48
202308	00000078909182	ANTARCTICA PILSEN GARRAFA LONG NECK 355 UNIDADE 1	\$3,28
202308	00000078911314	BOHEMIA PURO MALTE GARRAFA LONG NECK 355 UNIDADE 1	\$6,46
202308	00000078912861	CERPA EXPORT GARRAFA LONG NECK 350 UNIDADE 1	\$12,64
202308	00000078916265	SKOL BEATS SENSES GARRAFA LONG NECK 330 UNIDADE 1	\$22,44
202308	00000078929203	CERPA TIJUCA GARRAFA LONG NECK 350 UNIDADE 1	\$12,80
202308	00000078932470	CERPA TIJUCA GARRAFA 600 UNIDADE 1	\$9,00
202308	00000078934115	SOL PREMIUM GARRAFA LONG NECK 330 UNIDADE 1	\$5,53
202308	00000078935495	HEINEKEN GARRAFA LONG NECK 250 UNIDADE 1	\$4,80
202308	00000078936546	CERPA PRIME GARRAFA LONG NECK 350 UNIDADE 1	\$12,94
202308	00000078936683	HEINEKEN GARRAFA LONG NECK 330 UNIDADE 1	\$6,88
202308	00018200001338	MICHELOB ULTRA SUPERIOR LIGHT GARRAFA LONG NECK 355 CAIXA 24	\$6,44

Solução Final Dashboard da Pauta

É possível ver a relação entre os valores das **Pautas Calculadas para cada Mês**


SEFAZ
Secretaria de Estado
da Fazenda do Piauí

Pautas

Todos ▾

GTIN / SKU

Todos ▾

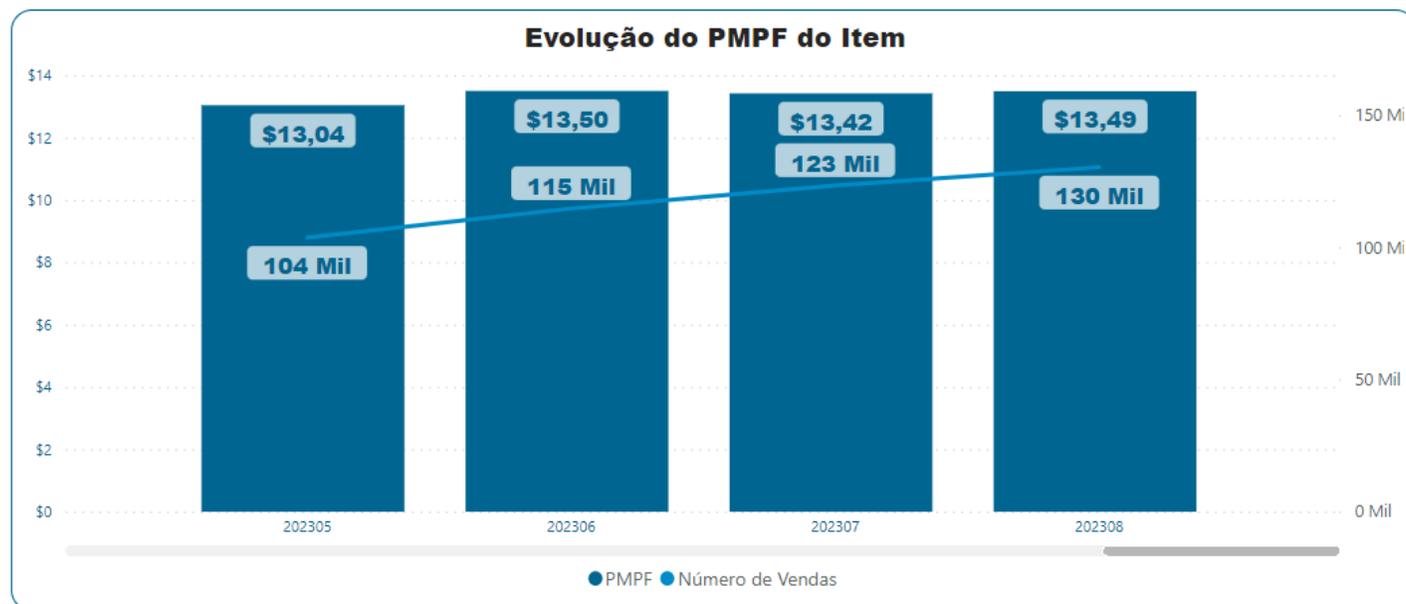
Descrição

HEINEKEN GARRAFA... ▾

← Menu →

Visão Histórica do PMPF do Item

Pauta	GTIN / SKU	Descrição	Vendas	PMPF
202308	00000078905498	HEINEKEN GARRAFA 600 UNIDADE 1	130245	\$13,49
202307	00000078905498	HEINEKEN GARRAFA 600 UNIDADE 1	123366	\$13,42
202306	00000078905498	HEINEKEN GARRAFA 600 UNIDADE 1	114672	\$13,50
202305	00000078905498	HEINEKEN GARRAFA 600 UNIDADE 1	104000	\$13,04



Solução Final Relatórios da Pauta

É possível ver a relação entre a **Pauta vigente** os valores das últimas **Pautas Calculadas** para cada **Mês**



Pauta Fiscal Cobrança

Governo do Estado do Piauí
Secretaria da Fazenda
Unidade de Tecnologia

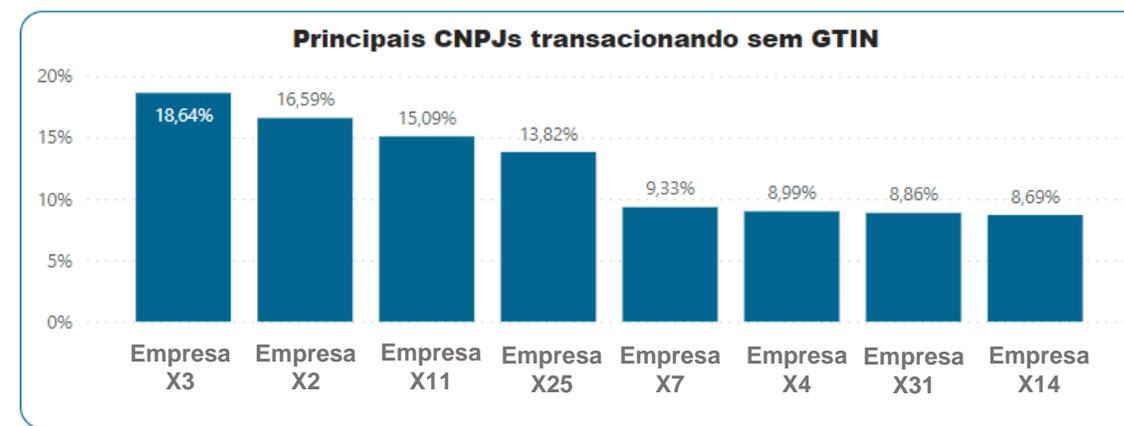
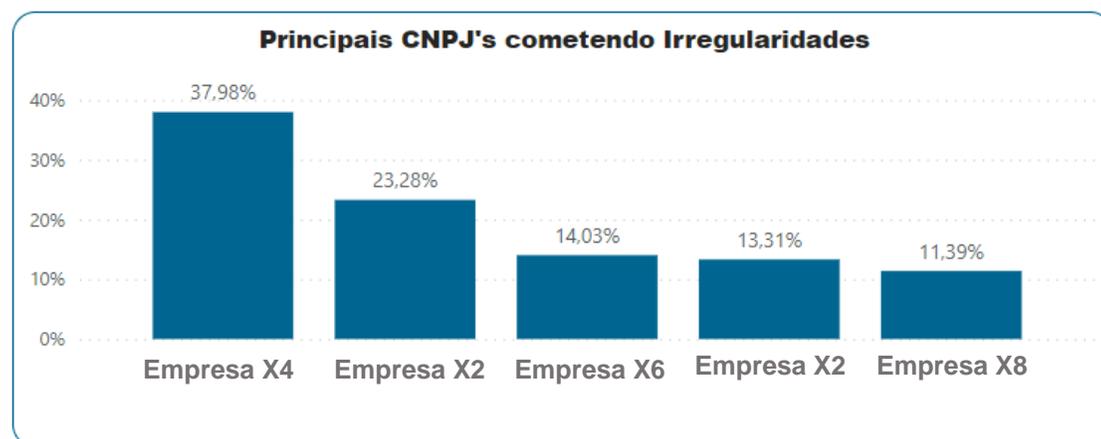
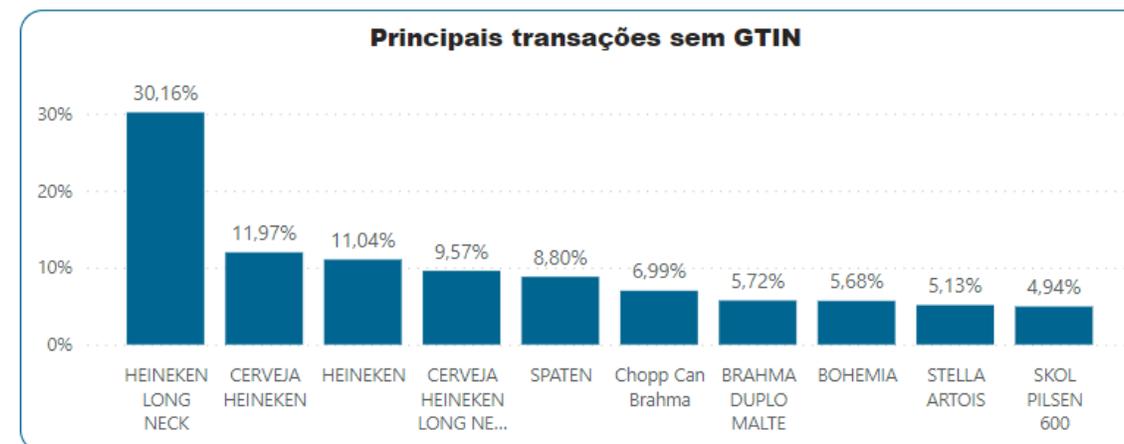
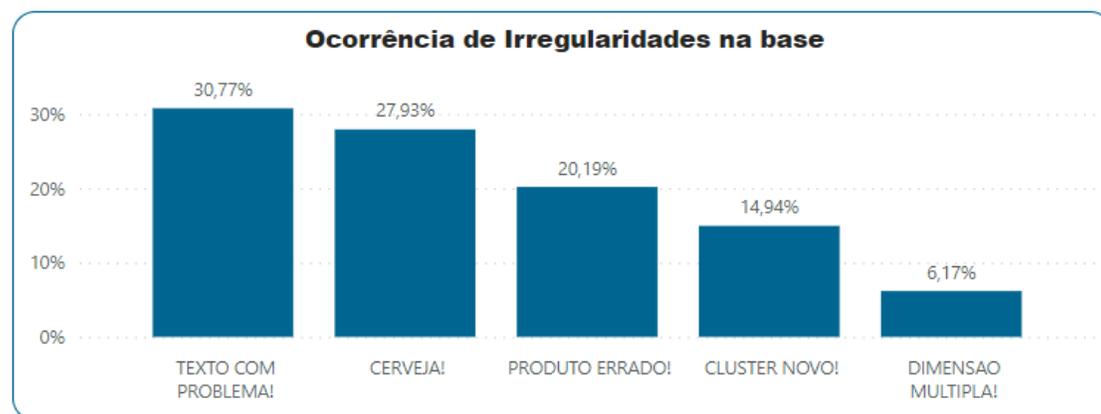
Código GTIN : Natureza da Operação : **OP INTERNA** Descrição do Produto : **CERVEJA HEI** Tipo Embalagem : **UNIDADE**

Dados da Pauta Fiscal Atual				Dados da Pauta Fiscal Automatizada						
Cod. Produto	Redução BC	Descrição Produto	Preço Atual	GTIN	Descrição Grupo	Mes Atual	DIF PMPF	Mes Atual - 1	Mes Atual - 2	Qtd Tran.
14464	0,00	CERVEJA HEINEKEN 0 POR CENTO GARRAFA DESCARTAVEL LONG NECK 330ML	5,83	07896045506040	HEINEKEN ZERO ALCOOL GARRAFA LONG NECK 330 UNIDADE 1	8,16	2,33	8,24	8,11	16287
14465	0,00	CERVEJA HEINEKEN 0 POR CENTO LATA 350 ML	4,43							
10721	0,00	CERVEJA HEINEKEN 0% GARRAFA DESCARTÁVEL LONG NECK 330ML	4,75							
10716	0,00	CERVEJA HEINEKEN 0% LATA 350 ML	4,19	07896045506064	HEINEKEN ZERO ALCOOL LATA 350 UNIDADE 1	6,01	1,82	6,09	5,74	3982
25722		CERVEJA HEINEKEN 0.0% LATA 269ML	3,49							
14466	0,00	CERVEJA HEINEKEN BARRIL 5L	80,26	08712000025649	HEINEKEN BARRIL 5000 UNIDADE 1	15,90	(64,36)	54,95	56,80	360
14467	0,00	CERVEJA HEINEKEN GARRAFA DESCARTAVEL 250ML	4,12	07896045505357	HEINEKEN LATA 250 UNIDADE 1	7,76	3,64	5,33	4,59	487
14468	0,00	CERVEJA HEINEKEN GARRAFA DESCARTAVEL 600ML	10,58	00000078905498	HEINEKEN GARRAFA 600 UNIDADE 1	13,70	3,12	13,49	13,42	123048
14469	0,00	CERVEJA HEINEKEN GARRAFA DESCARTAVEL LONG NECK 330ML	5,83	00000078936683	HEINEKEN GARRAFA LONG NECK 330 UNIDADE 1	6,94	1,11	6,88	7,02	146997
14471	0,00	CERVEJA HEINEKEN GARRAFA RETORNAVEL 600ML	11,10							
14472	0,00	CERVEJA HEINEKEN LATA 250ML	3,76	00000078935495	HEINEKEN GARRAFA LONG NECK 250 UNIDADE 1	4,79	1,03	4,80	4,74	4429
25720		CERVEJA HEINEKEN LATA 269ML	3,49	07896045506590	HEINEKEN LATA 269 UNIDADE 1	3,61	0,12	4,09	4,10	11456
12 Registro(s)										



Solução Final Dashboard de Conformidade

É possível ver quem não está em conformidade para cada Mês de Pauta calculada (expurgo)



Novos encaminhamentos para solução



Mudar origem da ingestão de dados para object storage, uma arquitetura de armazenamento de arquivos, que foi adquirido pela SEFAZPI recentemente.



Atuação fiscal nas empresas identificadas através da ferramenta para solucionar problemas de preenchimento.

O GTIN é referência única do item, sendo sua conformidade imprescindível.



Criação de “Novas metodológicas” do Algoritmo de Cálculo da Pauta Fiscal

Criação das notas metodológicas já prevendo a necessidade de apresentação da solução a sociedade civil (Associação de Bares e Restaurantes, Sindicatos, Representantes do Comércio)



Inclusão de novas pautas fiscais conforme prioridade do negócio.

Começamos trabalhar com a Pauta Fiscal dos Refrigerantes, Energéticos, Isotônicos e Água Mineral.





Equipe de projeto



EloGroup

Gerente Senior

- Marcello Fonseca

Liderança Técnica

- Daniel Micoski

Consultor

- Arthur Leardini

Cientista de Dados

- Lucca Bevilacqua

Engenheiro de Dados Pleno

- Bruno Brandão

Analista de Dados

- Guilherme Laureano
- Rafael Borges
- Pedro Watanabe



SEFAZPI

Diretor da UNITEC

- Alexandre Medrado

Gerente da GEPRO

- Aldenir do Vale

Engenheiro de Dados

- Caubi Sousa
- Melquisedequi

Cientista de Dados

- Rodolfo Melo

Auditor de Negócios

- José Lima Coutinho

