

**Universidade Tecnológica Federal do Paraná - Campus Dois Vizinhos**

**Especialização em Ciência de Dados**

# **RELATÓRIO PROJETO INTEGRADOR**

**Dois Vizinhos - PR**

**2021**

---

**Título: Análise de preços de medicamentos por estado, princípio ativo, tipo e tarja.**

**Área: Saúde**

**Grupo 1 - Autores: André Lucas Silva; Flávio Rôvere Perseguini; Israel de Freitas Gonçalves Dias; Lucas Menezes Moreira; Thiago Dias**

---

### **Objetivo geral do trabalho:**

A cobrança de diferentes alíquotas de ICMS em cada estado para cada medicamento gera uma distribuição de valores cobrados ao consumidor final. Neste trabalho analisamos a dispersão desses valores, seja por localidade, princípio ativo, laboratório, nome fantasia para apresentar de uma forma clara e objetiva para o consumidor.

### **1. Análise e Modelagem de Banco de Dados:**

#### **(i) Descrição dos requisitos de banco de dados:**

**Fonte de Dados:** A criação do banco de dados foi construído através de informações disponíveis no site da Anvisa: (<https://www.gov.br/anvisa/pt-br/assuntos/medicamentos/cmed/precos>).

A base de dados contém muitas informações, que vão desde o preço dos medicamentos até informações mais técnicas e específicas sobre o medicamento. Entretanto, todas as informações de um medicamento específico é contida em um registro único contendo diversos atributos. Muitos deles, principalmente os que envolvem características técnicas e específicas de medicamentos e laboratórios, não têm muito valor para a nossa análise.

Além disso, algumas informações apresentam uma certa redundância, como a questão dos preços por alíquota por ICMS. A Anvisa define um preço único por medicamento, válido em todo o território nacional, e o preço final do medicamento para cada estado varia conforme a alíquota do ICMS que cada estado aplica. Dessa forma, uma base de dados não faria sentido armazenar o valor final do medicamento, visto que este pode ser alterado, caso um estado altere sua alíquota, entretanto deve armazenar o preço único e as alíquotas utilizada para cada medicamento.

#### **(ii) Modelagem do Diagrama (MER)**

A Imagem 1 abaixo apresenta o Diagrama Entidade-Relacionamento (MER) utilizado em nossa aplicação, aplicada sobre a base de dados, conforme os dados que vamos armazenar e analisar. Primeiramente, da base de dados fornecida pela ANVISA, para cada medicamento, foi realizada uma filtragem, removendo os atributos que não são de interesse de nossa análise.

Dessa forma, a base de dados irá conter uma entidade denominada medicamento, contendo atributos como um identificador, GGREM (Gerência Geral de Regulação Econômica, um código relacionado à apresentação de um medicamento), nome do medicamento, apresentação (texto descritivo do medicamento), a sua classe terapêutica, a informação de sua tarja (preta, vermelha, etc), o tipo de medicamento (genérico, fitoterápico, etc) e o seu preço unitário (denominado de Preço Final na base de dados).

Além dessa entidade, foi criada a entidade Laboratório, que armazena as informações do CNPJ e o nome do laboratório em específico, e a entidade Estado, que armazena o nome de cada estado e a alíquota ICMS que este estado está adotando no momento.

As relações em geral são de cardinalidade 1:N em relação ao medicamento. Por exemplo, um medicamento contém uma única Tarja, porém um grupo específico de tarja pode conter diversos medicamentos. O mesmo vale para o princípio ativo, o tipo de medicamento e a classe terapêutica e laboratório. Um laboratório pode fabricar diversos medicamentos, mas um medicamento pode ser fabricado por um único laboratório.

Por fim, a relação entre medicamento e estado é mapeada com cardinalidade N:N, pois um medicamento pode ser vendido em distintos estados, bem como um estado pode vender diversos medicamentos.

### **(iii) Figura do projeto lógico do Banco de Dados (MR)**

A Imagem 2 abaixo apresenta o modelo lógico do banco de dados, criado baseado no MER da Imagem 1. Dessa forma, cada entidade foi mapeada em uma tabela, contendo os seus atributos. Normalmente, os IDs são utilizados como chaves primárias, salvo exceções como o laboratório que utilizou o CNPJ e o estado que utilizou a sigla. A relação N:N entre medicamento e estado fez surgir uma terceira tabela, denominada medicamento\_estado. A tabela medicamento, além de conter os seus atributos, armazena as chaves estrangeiras das suas relações 1:N.

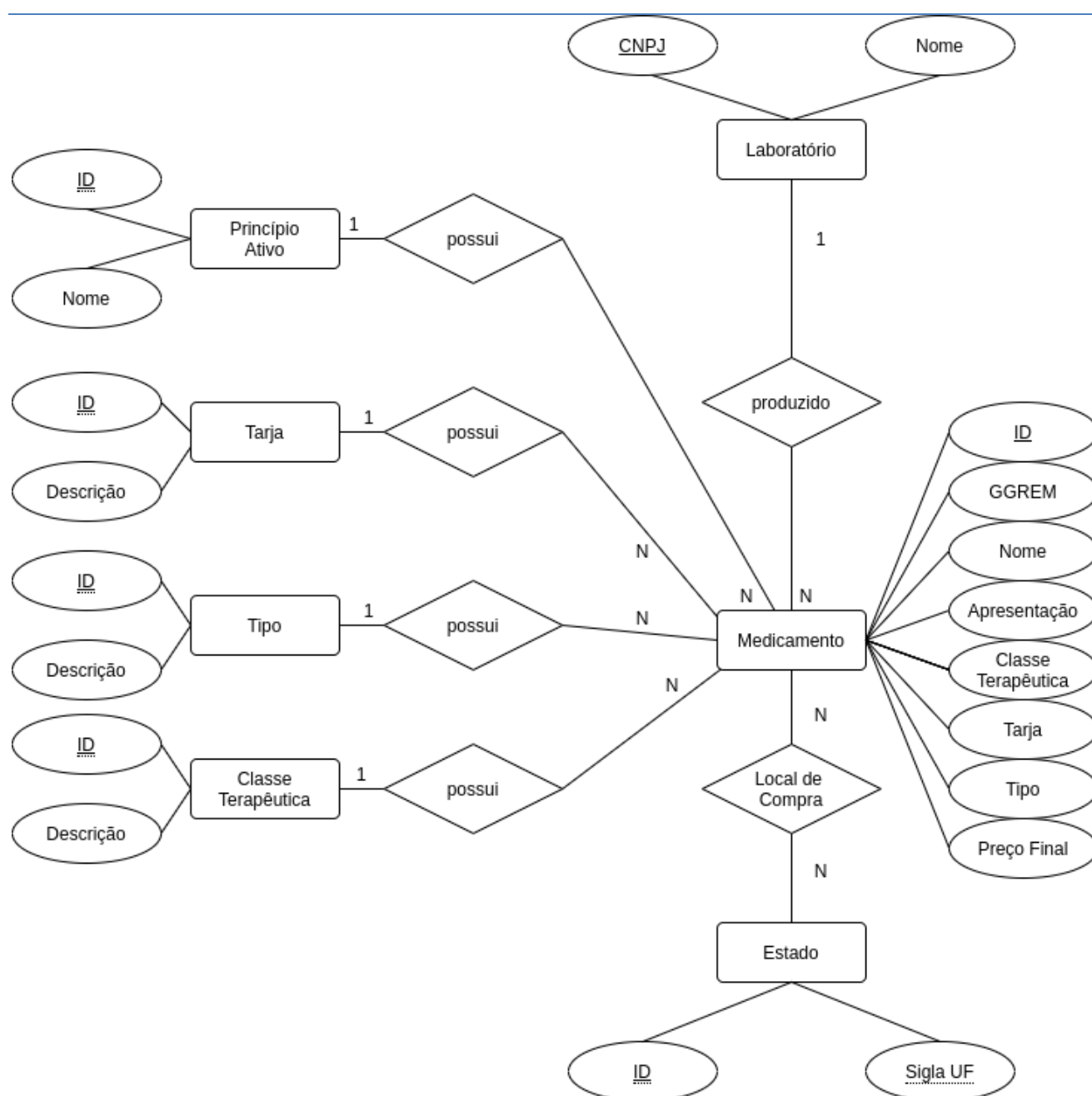


Imagem 1 - Modelo Entidade-Relacionamento (MER)

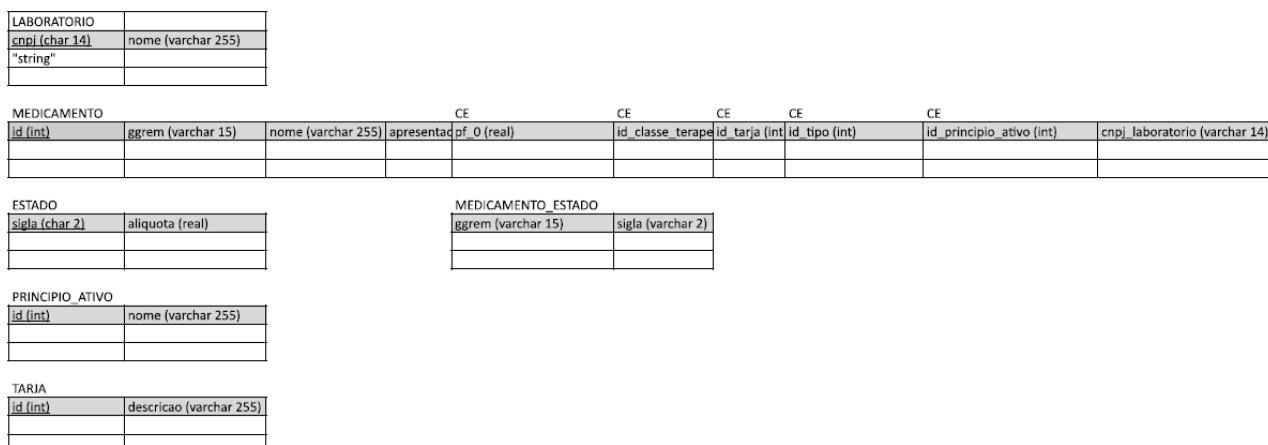


Imagem 2 - Projeto lógico do banco de dados

## 2. Introdução ao Gerenciamento de Banco de Dados:

### (i) Um script chamado `cria_banco.sql` com todos os CREATE TABLE

Este script se encontra no arquivo `criar_banco.sql` dentro do diretório IGBD enviado em anexo.

### (ii) e (iii) script de consulta de banco sql com todos os SELECT's definidos pelo grupo e seus comentários

Este script se encontra no arquivo `consulta_banco.sql` dentro do diretório IGBD enviado em anexo.

No mesmo diretório IGBD, no arquivo `read_insert-db.R` está o script em R que foi utilizado para ler o arquivo .csv com os dados, conectar-se com o banco de dados do PostgreSQL e inserir os dados na base de dados.

## 3. Introdução à Análise e Ciência de Dados:

### (i) gráficos de barras, linhas e colunas

Os arquivos Projeto Integrador - Análise Q1.R e Projeto Integrador - Análise Q3.R dentro do diretório IACD enviado em anexo contém análise realizada sobre os valores máximo, médio e mínimo cobrado por princípio ativo de medicamento, filtrando apenas os 10 primeiros valores, devido ao elevado número de princípios ativos. A Imagem 3, 4, 5 abaixo apresenta os resultados obtidos.

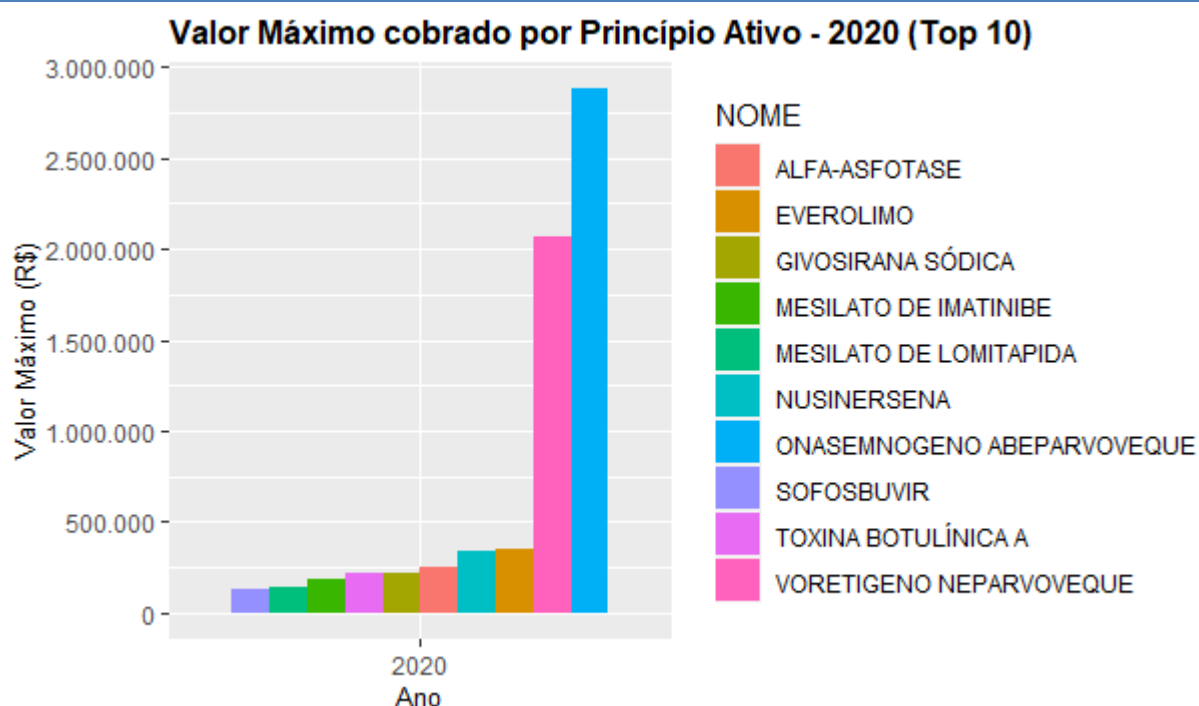


Imagem 3 - Valor máximo por princípio ativo

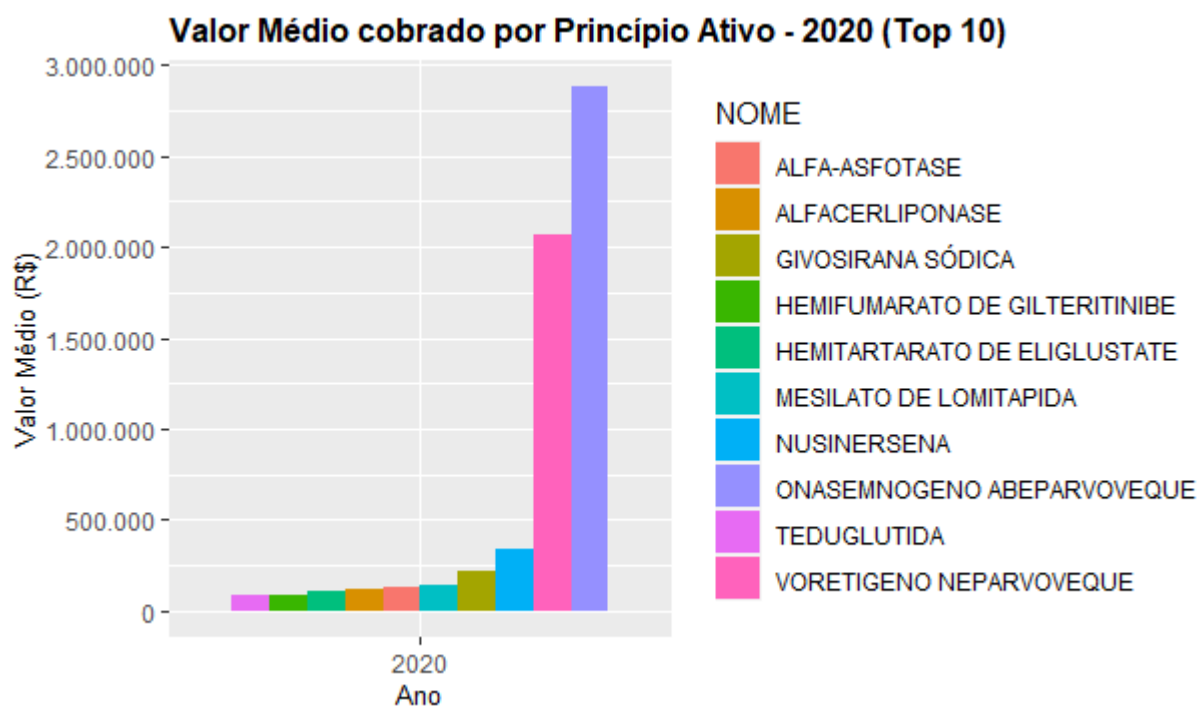
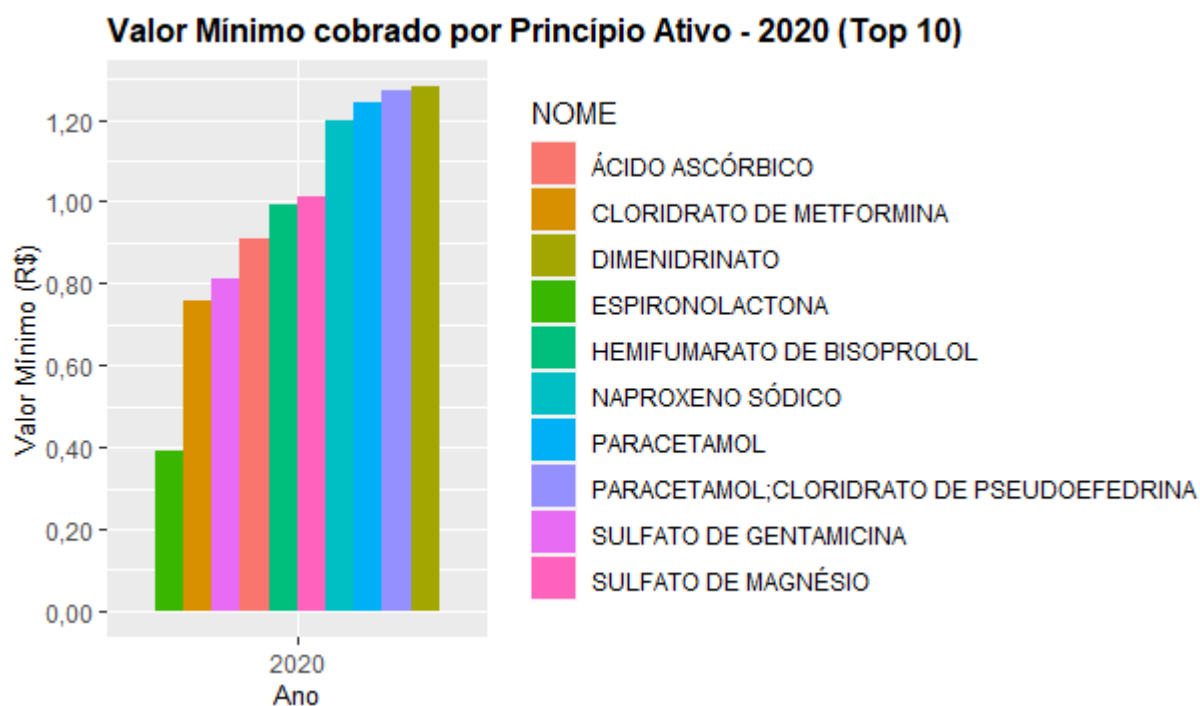


Imagem 4 - Valor médio por princípio ativo



**Imagem 5 - Valor mínimo por princípio ativo**

O arquivo q2.R dentro do diretório IACD enviado em anexo contém análise realizada sobre os valores médio cobrado por laboratório, filtrando apenas os 10 primeiros, devido ao elevado número de laboratórios. A Imagem 6 abaixo apresenta os resultados obtidos.

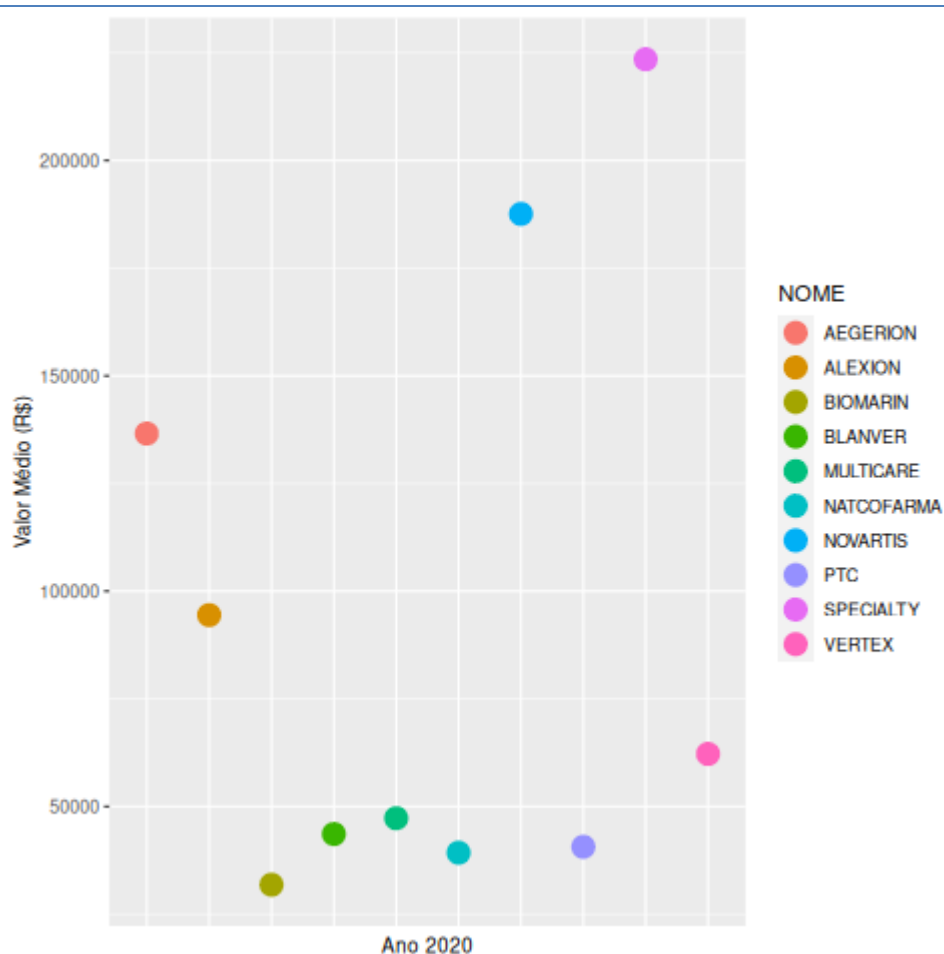


Imagem 6- Valor médio por laboratório

## (ii) tabelas

O arquivo Quant\_Principio\_Ativo.R dentro do diretório IACD apresenta como resultado uma tabela contendo a quantidade de medicamento por princípio ativo, filtrando os 10 primeiros resultados. A Imagem 7 abaixo apresenta os resultados obtidos. A coluna Quantidade apresenta a quantidade de medicamentos do princípio, e a coluna PrincipioAtivo apresenta o nome do princípio ativo.

O arquivo Quant\_Tarja.R dentro do diretório IACD apresenta como resultado uma tabela contendo a quantidade de medicamento por tarjas. A Imagem 8 abaixo apresenta os resultados obtidos. A coluna Quantidade apresenta a quantidade de medicamentos da tarja, e a coluna Descricao apresenta o nome da tarja.



	▲ Quantidade ▼	PrincipioAtivo ▼
1	442	COLECALCIFEROL
2	223	PARACETAMOL
3	220	OLANZAPINA
4	204	PREGABALINA
5	197	OXALATO DE ESCITALOPRAM
6	191	NI/NC
7	190	ROSUVASTATINA CÁLCICA
8	184	HEMIFUMARATO DE QUETIAPINA
9	178	CITRATO DE SILDENAFILA
10	176	IBUPROFENO

Imagem 7 - Tabela com os princípios ativos com mais medicamentos

▲	Quantidade ▼	Descricao ▼
1	10891	- (*)
2	9485	Tarja Vermelha
3	3642	Tarja Vermelha sob restrição
4	1405	Tarja Sem Tarja
5	353	Tarja Preta

Imagem 8 - Tabela com a quantidade de medicamento por tarja