

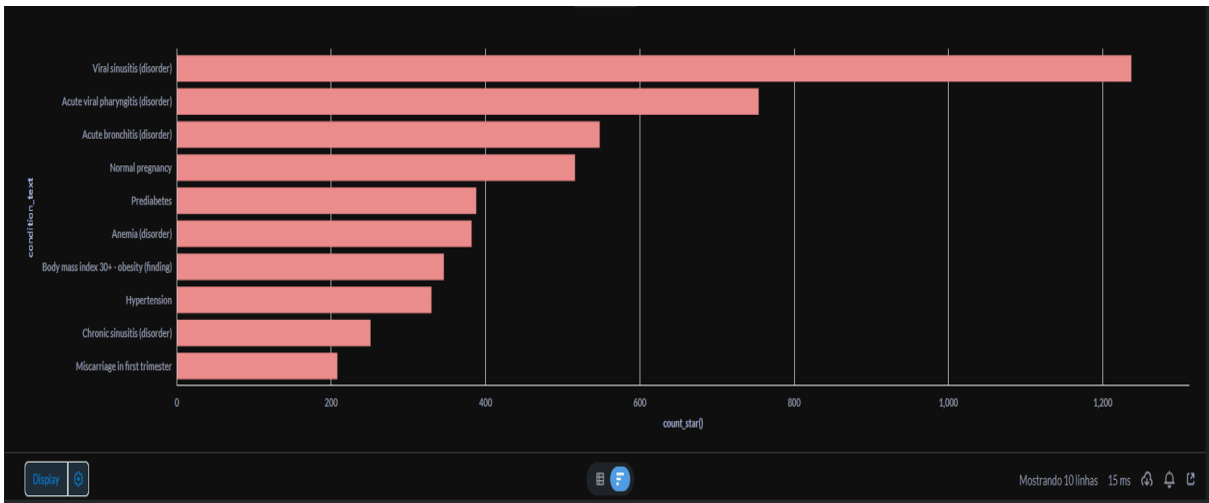
Resultados e Visualizações

Os dados processados foram carregados no DuckDB e visualizados no Metabase, gerando os seguintes insights:

Condições Mais Frequentes

A condição mais comum é "Viral sinusitis (disorder)" com 1.237 casos, indicando uma prevalência significativa de infecções respiratórias.

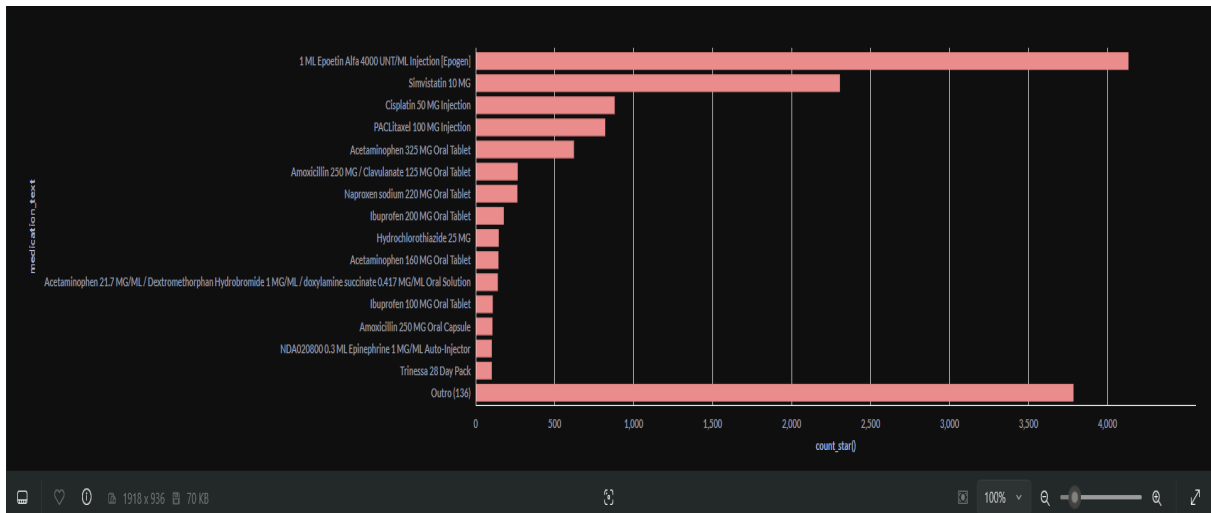
Outras condições frequentes incluem "Acute pharyngitis" e "Acute bronchitis", sugerindo um padrão de doenças respiratórias agudas.



Medicamentos Mais Prescritos:

"1 ML Epoetin Alfa 4000 UNT/ML Injection [Epogen]" lidera com mais de 3.500 prescrições, possivelmente relacionadas ao tratamento de anemia em pacientes com condições crônicas.

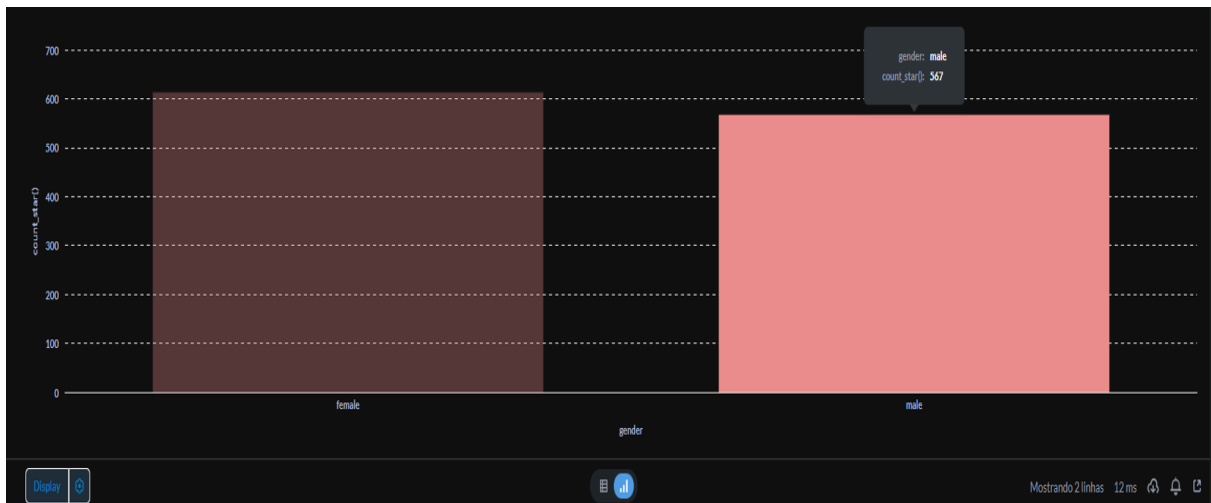
Medicamentos como "Clopidogrel" (antiplaquetário) e "Acetaminophen" (analgésico) também são amplamente prescritos, refletindo necessidades comuns de manejo de dor e prevenção cardiovascular.



Distribuição de Gênero

A distribuição de gênero é equilibrada, com 567 pacientes masculinos.

Isso sugere que a amostra de dados não apresenta viés de gênero significativo.



Execução no Airflow

O DAG `process_health_data` foi executado com sucesso em 19/03/2025 às 19:25:43 UTC.

A tarefa `run_spark_processing` processou os dados e salvou no DuckDB, pronta para visualização no Metabase.

localhost:8080/dags/process_health_data/grid?tab=code

apresentação - Doc...Data Pipelines com...Inscrição e matricul...IBO_2022 Descrigh...Daniel de Souza Mir...Introduction to SQL...Weather Data in Big...Demo de RestoreBig Data - Foundati...Certification ProfileLogin Decree | Aces...>>Todos os favoritos

AirflowDAGsCluster ActivityDatasetsSecurityBrowseAdminDocs19:28 UTC

DAG: process_health_dataDAG para processar dados de saúde com Spark e salvar no DuckDBSchedule: 1 day, 0:00:00Next Run ID: 2025-03-19, 00:00:00 UTC

19/03/2025 19:25:05All Run TypesAll Run StatesClear FiltersAuto-refresh25

Press (shift) + ⌘ for Shortcutsdeletedfailedqueuedremovedrestartingrunningcheduledshutdownskippedsuccessup_for_rescheduleup_for_retryupstream_failedno_status

DAGprocess_health_data

DetailsGraphGanttCodeAudit LogRun DurationCalendar

Duration00:01:0600:00:3300:00:00

run_spark_processing

Parse at: 2025-03-19, 19:24:52 UTC

```
1 from airflow import DAG
2 from airflow.operators.python import PythonOperator
3 from datetime import datetime, timedelta
4 import subprocess
5
6 # Função para executar o script Spark
7 def run_spark_processor():
8     result = subprocess.run(
9         ['python', '-I', 'airflow/spark_processor.py'],
10        capture_output=True,
11        text=True
12    )
13    if result.returncode != 0:
14        raise Exception(f'Erro ao executar spark_processor.py: {result.stderr}')
15    print(result.stdout)
16
17 # Configuração padrão do DAG
18 default_args = {
19     'owner': 'airflow',
20     'depends_on_past': False,
21     'email_on_failure': False,
22     'email_on_retry': False,
23     'retries': 1,
24     'retry_delay': timedelta(minutes=5),
25 }
```

Toggle Wrap