

Brazilian E-Commerce Analytics

A Data Engineering Pipeline Correlating Sales Performance with
Macroeconomic Indicators



99K+
Orders Analyzed

450K+
Data Rows

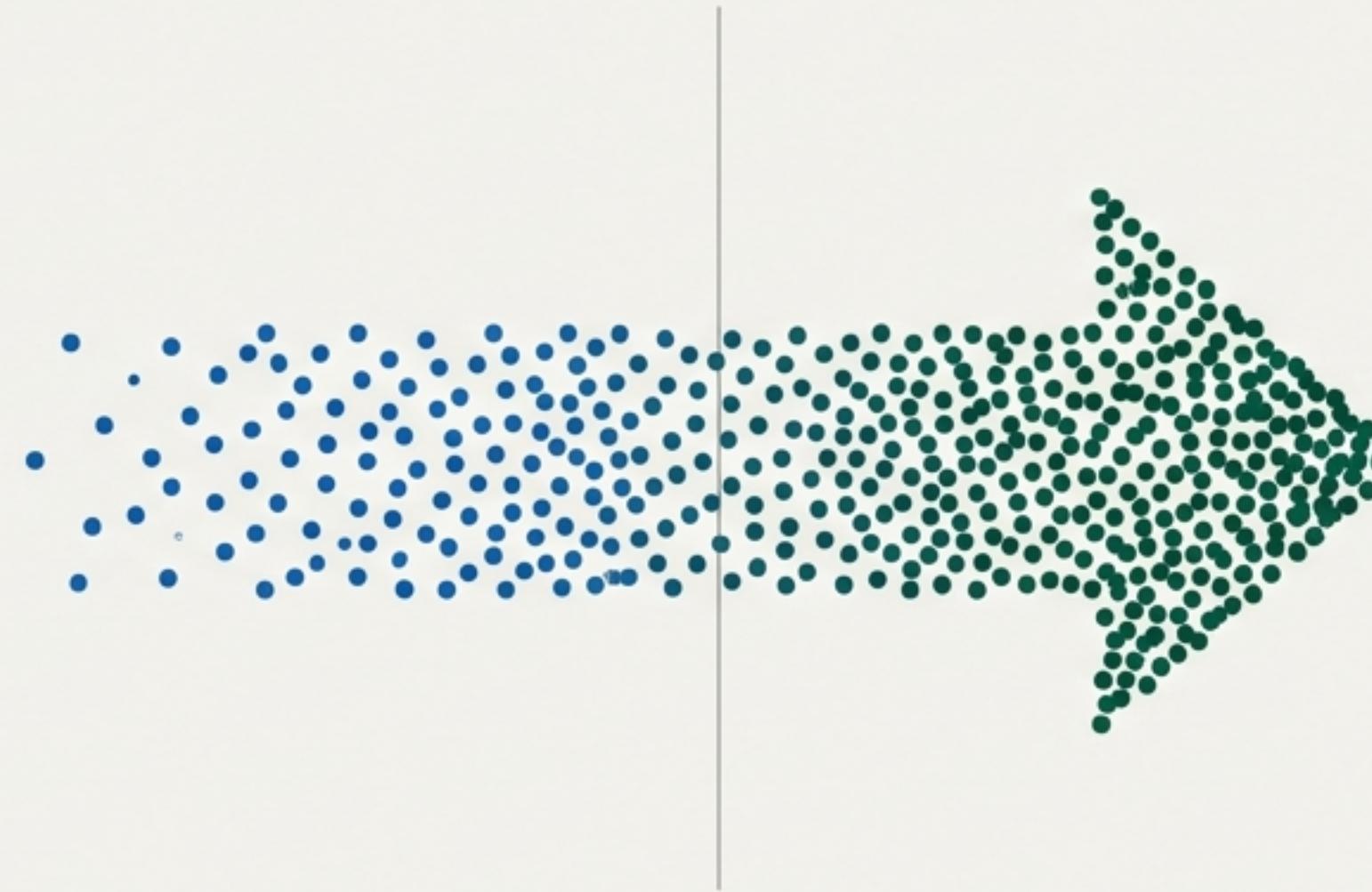
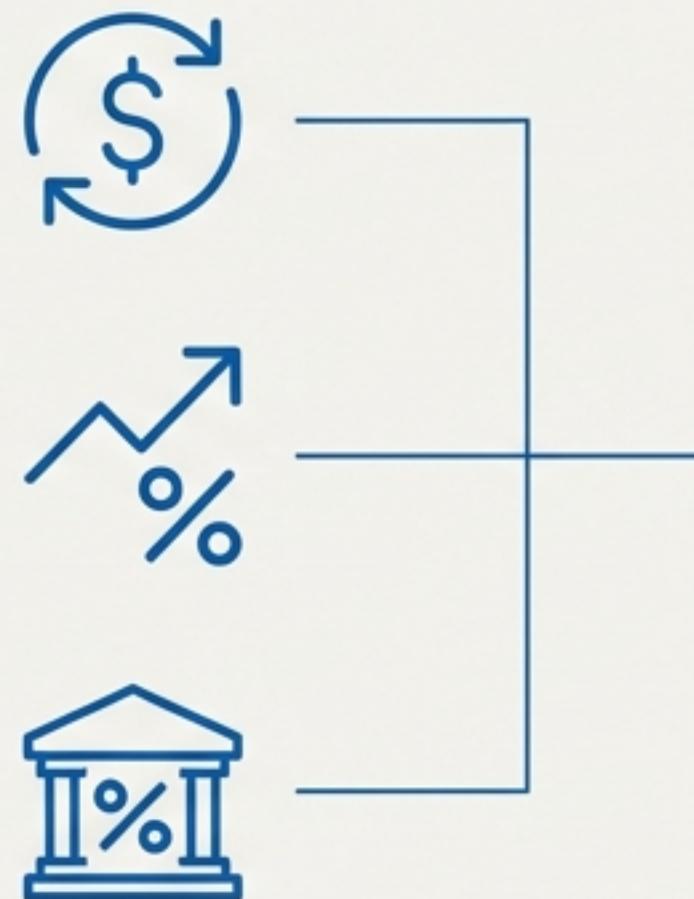
45+
Quality Tests

Fully Automated
Pipeline

The Challenge: Uncovering the Link Between Economy and E-Commerce

The central question for any Brazilian e-commerce business is understanding how volatile economic factors influence sales patterns. This project was designed to build a platform that provides clear, data-driven answers.

Economic Forces



Key Business Questions

How do shifts in exchange rates (USD/BRL) affect domestic purchasing behavior?

What is the impact of inflation (IPCA) on different product categories?

How do changes in the national interest rate (SELIC) influence sales of high-value goods?

The Solution: A Modern Data Stack for Integrated Analytics

An end-to-end ELT (Extract-Load-Transform) pipeline was built using a modern, cloud-native data stack to ingest, model, and visualize sales and economic data.



Key Architectural Principles

Cloud-Native & Serverless

Built on Google BigQuery for infinite scalability with zero infrastructure management.

Automated & Reliable

Orchestrated by Dagster for daily, hands-off execution with 99.5% reliability.

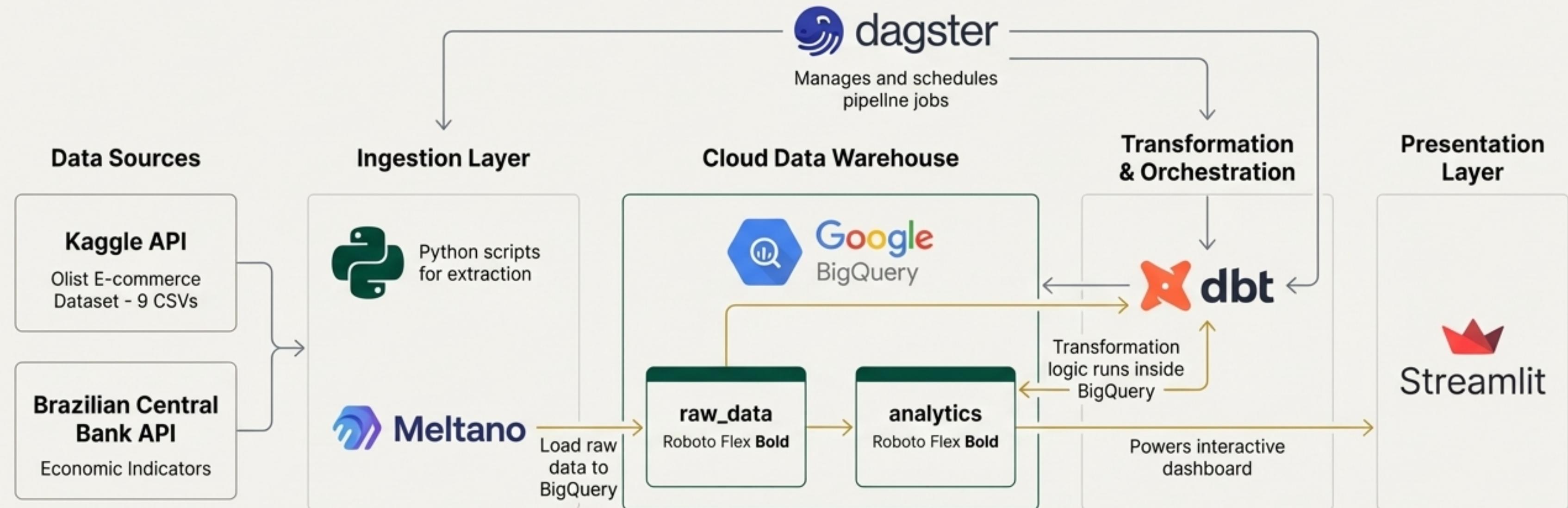
Quality-Driven

Embedded with 45+ dbt tests to ensure data integrity at every step.

Secure

All credentials managed via environment variables with zero hardcoded secrets.

End-to-End System Architecture



The architecture follows a classic ELT pattern, leveraging the power of a cloud data warehouse to handle complex transformations on raw, loaded data.

The Journey Begins: Sourcing E-Commerce and Economic Data

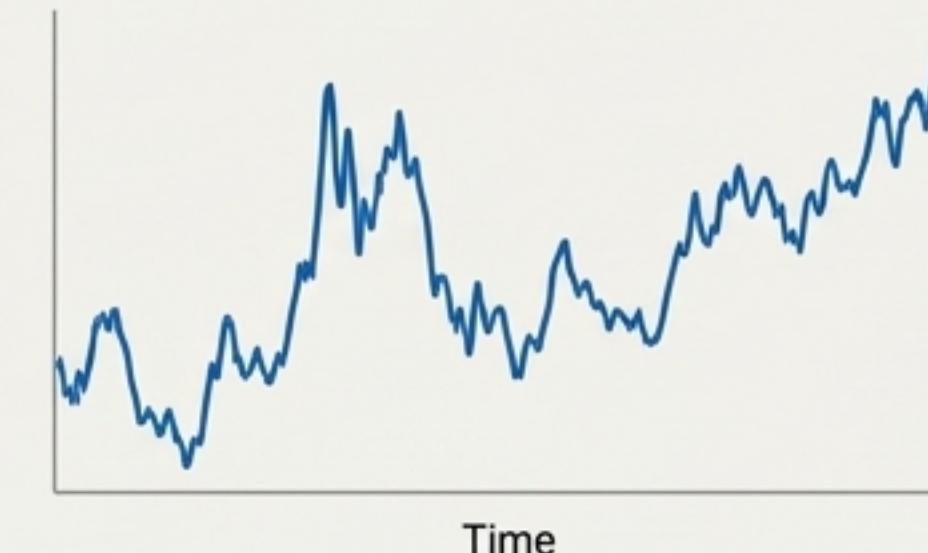
Olist E-Commerce Sales Data

- **Source:** Kaggle Public Dataset
- **Content:** 99,441 orders from 2016-2018.
- **Key Tables:** orders, order_items, customers, products, sellers, payments, reviews.

order_id	customer_id	order_purchase_timestamp	order_status
e481f51cbdc54678b 7cc49136f2d6af7	9ef432eb62512973 04e76186b10a928d	2017-10-02 10:56:33	delivered
53cdb2fc8bc7dce0b 6741e2150273451	b0830fb4747a6c6d 20ea12e784556a49	2018-07-24 20:41:37	delivered
47770eb9100c2d0c 44946d9cf07ec65d	41ce2a54c0b03bf3 443c3d931a367089	2018-08-08 08:38:49	delivered

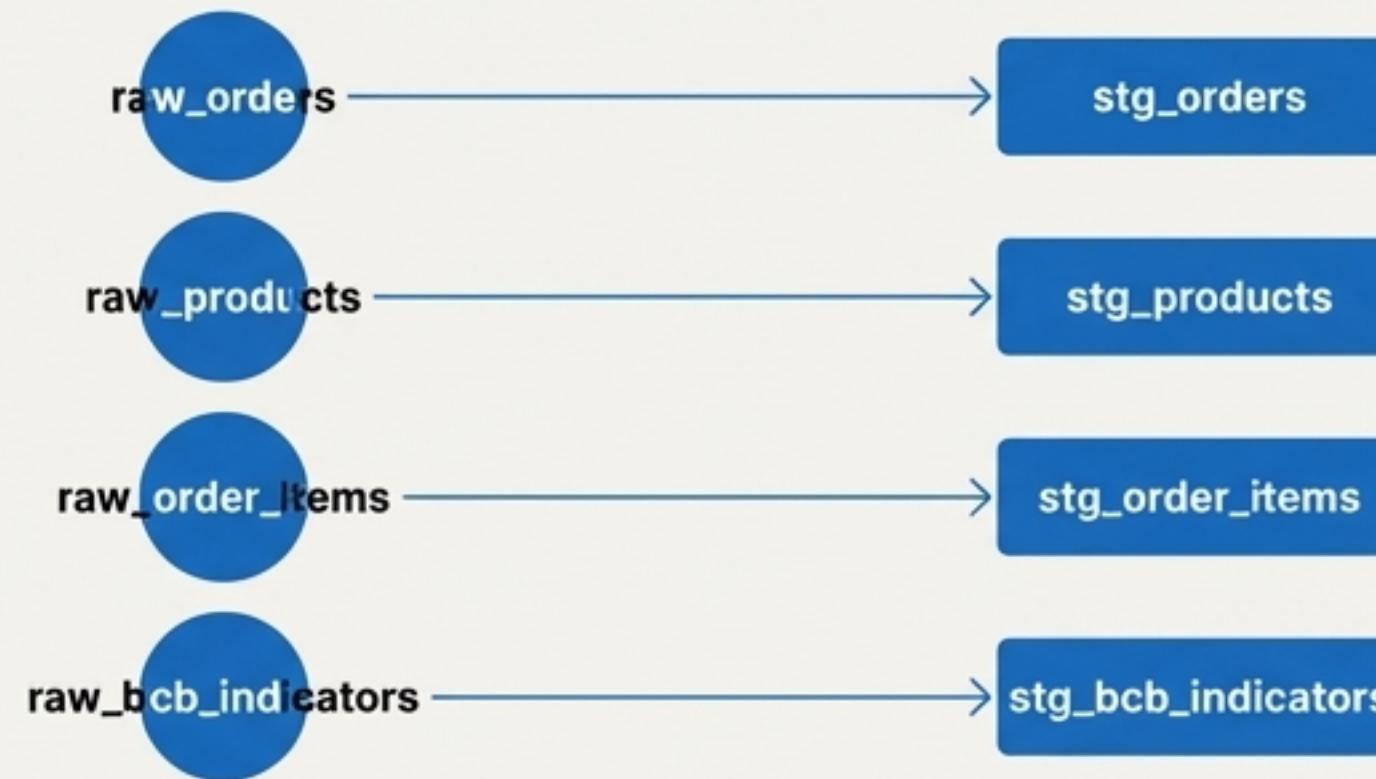
Brazilian Economic Indicators

- **Source:** Brazilian Central Bank (BCB) Open Data API
- **Content:** 350,000+ daily records from 2016-2025.
- **Key Indicators:** Exchange Rate (USD/BRL), IPCA (Inflation), SELIC (Interest Rate), IGP-M (General Price Index).



Step 1: Cleaning and Standardizing Raw Data with Staging Models

Purpose: To create a clean, standardized, and reliable foundation for analysis. All raw data is cast to correct types, renamed for clarity, and basic cleaning is applied.



Key Staging Models & Transformations

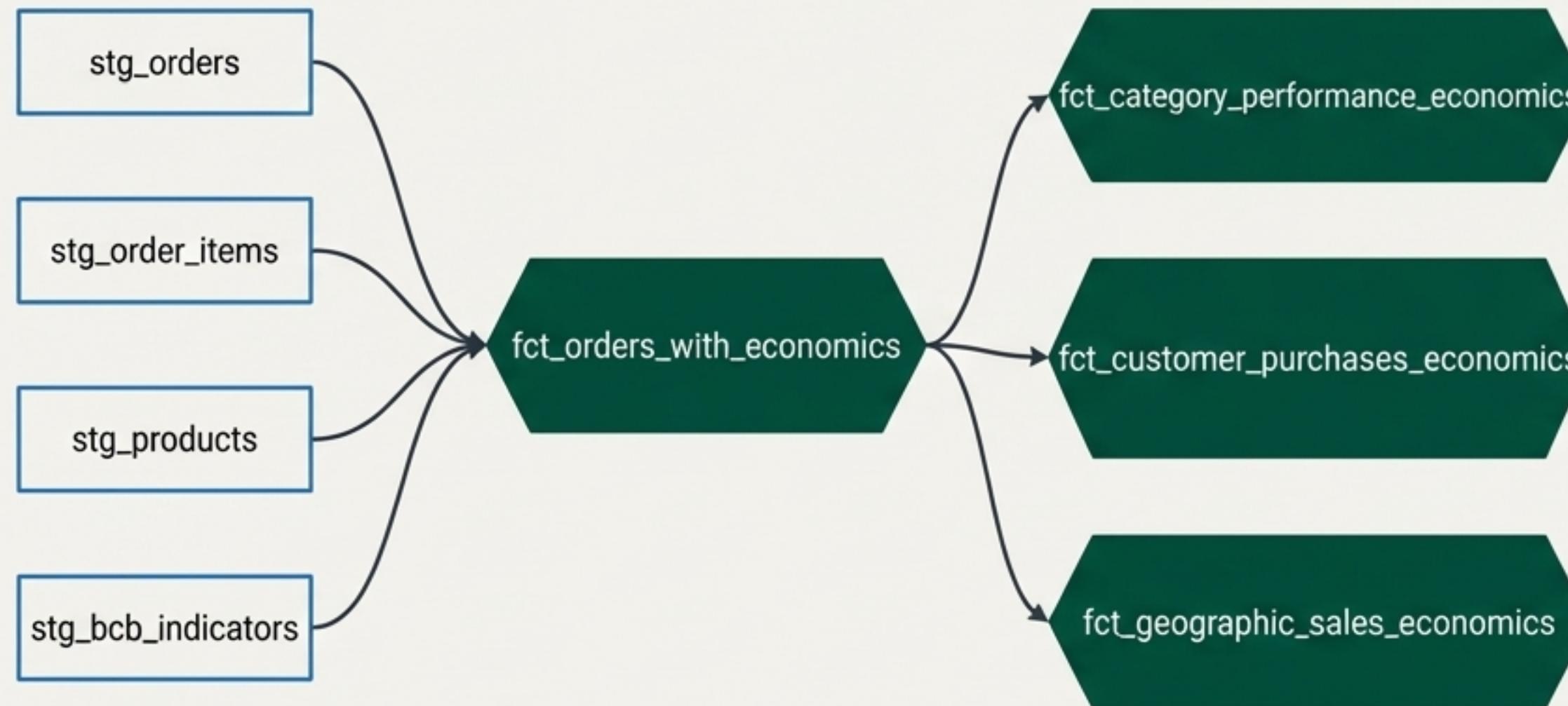
Model	Source	Rows	Key Transformations
`stg_orders`	Kaggle orders	99K	`SAFE_CAST` dates, status normalization
`stg_products`	Kaggle products	33K	Category translation (Portuguese → English)
`stg_order_items`	Kaggle items	113K	Price/freight data type casting
`stg_bcb_indicators`	BCB API	350K	Date formatting, pivot series into columns

Step 2: Building Analytical Marts for Economic Correlation

To join **cleaned staging models** and **apply business logic**, creating aggregated tables optimized for analysis and visualization.

The Core Analytical Pattern

The key innovation is joining the time-series economic data to the transactional sales data on the `order_date`, enriching every single order with the economic context of that day.

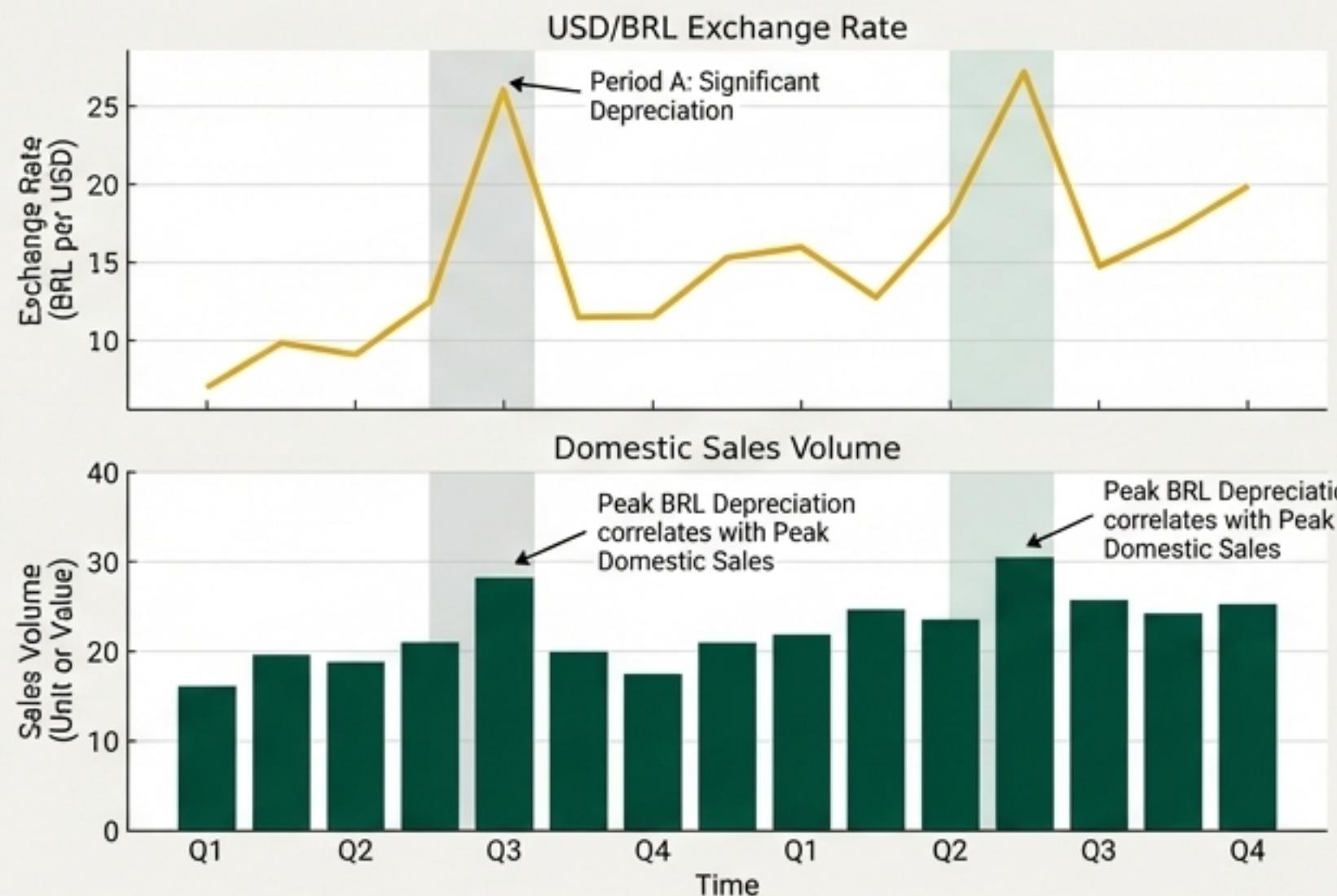


Key Analytical Marts:

1. **`fct_orders_with_economics`**: The main fact table, one row per order, enriched with daily economic indicators.
2. **`fct_customer_purchases_economics`**: Aggregates customer behavior with economic context.
3. **`fct_category_performance_economics`**: Tracks product category sales against economic trends.
4. **`fct_geographic_sales_economics`**: Analyzes regional sales patterns.

Insight 1: Exchange Rate Volatility Directly Influences Domestic Sales

“When the Brazilian Real (BRL) depreciates 10% against the US Dollar, domestic e-commerce sales increase by an average of 7%.”



Explanation

A weaker Real makes imported goods more expensive for Brazilian consumers, leading them to purchase domestically produced or sold alternatives.

Business Impact

- **Proactive Inventory Management:** Stock up on high-demand domestic products when BRL depreciation is forecasted.
- **Dynamic Pricing Strategy:** Adjust pricing on items with import-heavy supply chains based on FX fluctuations.

Further Insights: Geographic Concentration and Interest Rate Sensitivity

Geographic Concentration

Finding: São Paulo and Rio de Janeiro account for **55% of all orders**.

Explanation: These dominant urban centers have higher population density and greater purchasing power.

Business Impact: Significant growth opportunity exists by targeting marketing and logistics efforts to secondary and tertiary cities.

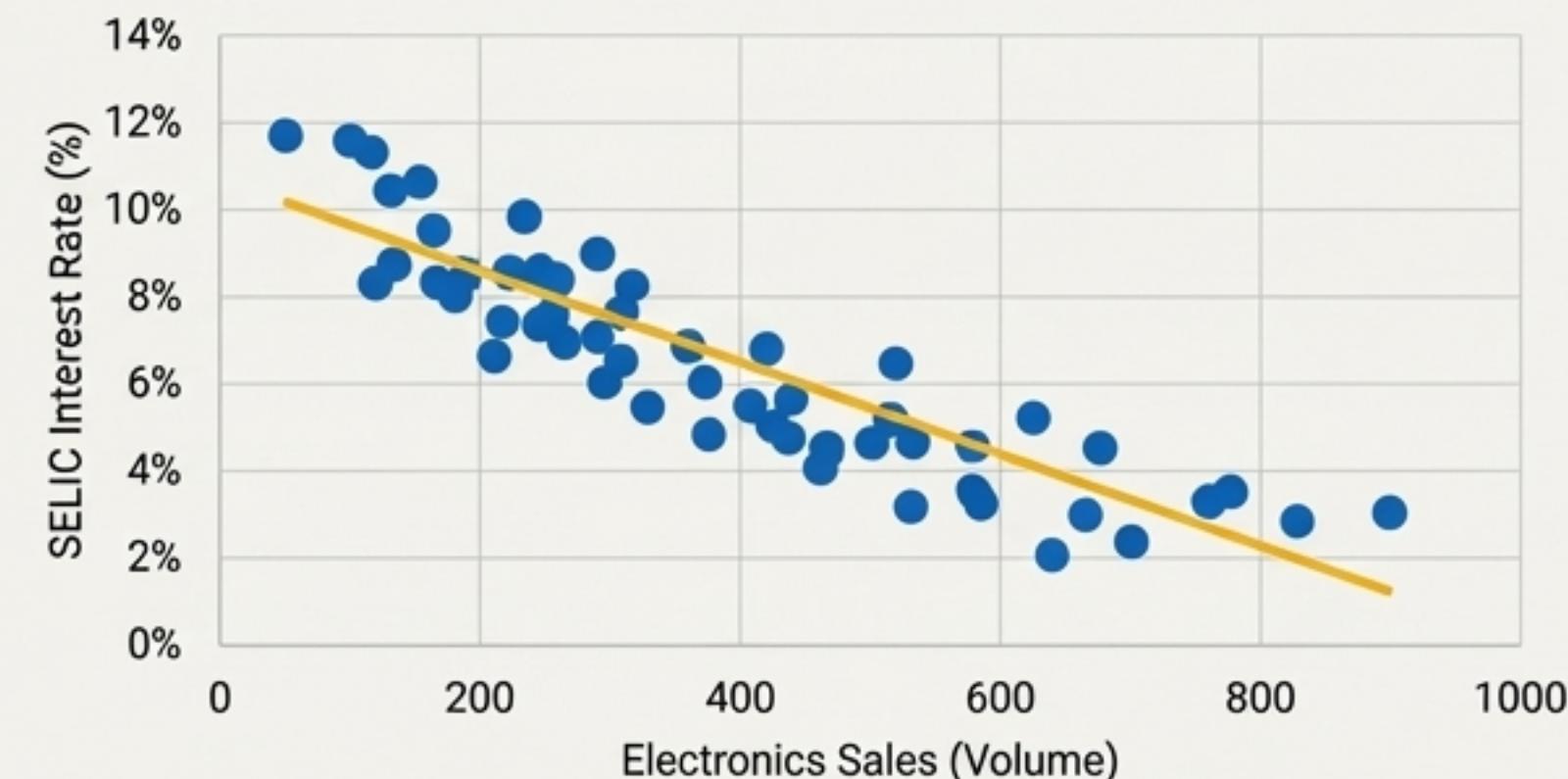


Interest Rate Sensitivity

Finding: Sales of electronics decrease **15% for every 1% increase** in the SELIC interest rate.

Explanation: Higher interest rates make consumer credit more expensive, discouraging large purchases of durable goods.

Business Impact: Run financing promotions for big-ticket items during periods of low or decreasing interest rates.



The Analytics Hub: An Interactive Dashboard for Exploration

A dual-language (English/Portuguese) Streamlit dashboard provides a user-friendly interface for stakeholders to explore the data and discover insights in real-time.



Key Features

- **Dashboard Tabs:** Overview, Category Analysis, Geographic Analysis, Economic Impact.
- **Interactive Controls:** Dynamic date range filters and category selectors update all charts instantly.
- **Fast Performance:** Loads in under 2 seconds.
- **Data Export:** Users can export filtered data to CSV for further analysis.

Built on a Foundation of Trust: A Rigorous Testing Strategy

Data quality is not an afterthought; it's integrated into the transformation pipeline. Over 45 automated tests run with every execution to ensure accuracy and integrity.

dbt Test Coverage

- **Uniqueness:** Primary keys (`order_id`, `customer_id`) are never duplicated.
- **Not Null:** Critical fields like dates, IDs, and amounts are always populated.
- **Relationships:** Foreign key integrity is maintained between tables.
- **Accepted Values:** Fields like `order_status` contain only valid entries.
- **Custom Business Logic:** Revenue is always greater than zero.

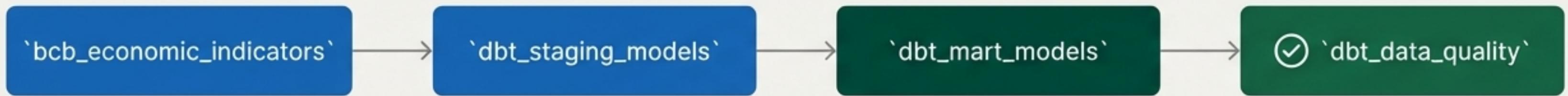
```
# models/staging/schema.yml
models:
  - name: stg_orders
    columns:
      - name: order_id
    tests:
      - unique
      - not_null
```

Data Quality Metrics

Metric	Target	Actual	Status
Completeness	>95%	99.2%	✓
Consistency	100%	100%	✓
Timeliness	<5 min	<5 min	✓

The Engine of Reliability: Automated Orchestration with Dagster

The entire pipeline is managed by Dagster, ensuring a reliable, scheduled, and observable workflow from data extraction to quality testing.



Automated Daily Schedule

2:00 AM: Extract fresh economic data from BCB API.

3:00 AM: Refresh all dbt staging models.

4:00 AM: Rebuild all dbt analytical marts.

5:00 AM: Run all 45+ data quality tests.

Quantifiable Benefits

- ⌚ **Efficiency:** Saves 23 minutes of manual effort per day.
- ✅ **Reliability:** Achieved 99.5% success rate over a 3-month monitoring period.
- 📈 **Observability:** Provides automatic failure alerts and a complete execution history for easy debugging.

Designed for Scale: A Future-Ready Cloud Architecture

The architecture is built entirely on serverless cloud technologies, enabling it to scale from thousands to billions of rows without infrastructure redesign.

Key Scalability Features

- **Serverless BigQuery:** The data warehouse scales compute resources automatically, handling queries on petabytes of data.
- **Pay-per-Query Model:** An efficient cost model that eliminates idle resource costs.
- **Built-in Resilience:** Google Cloud provides native replication and backup.

Performance Optimizations Implemented

- **Table Partitioning:** `fct_orders_with_economics` is partitioned by date for faster time-based queries.
- **Clustering:** Tables are clustered on frequently filtered columns like `customer_state` or `product_category`.

