BRAZIL E-COMMERCE ANALYTICS DASHBOARD

Technical Decisions & Rationale

PROJECT OVERVIEW

This document outlines the technical decisions made during the development of an interactive Power BI dashboard analyzing $13.6M in e-commerce revenue across 100,000+ transactions from a Brazilian online marketplace (2016-2018).

DATA MODEL DESIGN

Star Schema Architecture

Decision: Implemented a star schema with Order_Items as the central fact table surrounded by dimension tables for Products, Orders, Customers, and Date.

Rationale:

- Optimal query performance for analytical workloads
- Intuitive structure following Kimball dimensional modeling principles
- Enables multi-dimensional analysis (time, geography, product category)
- Smaller model size through normalized dimensions and denormalized facts

Fact Table Structure

Decision: Order_Items table contains transaction-level granularity with price, freight_value, and foreign keys to dimensions.

Rationale:

- Transaction-level detail supports all aggregation scenarios
- Maintains referential integrity through proper foreign keys
- Enables flexible slicing and filtering at any level

Relationship Configuration

Decision: All relationships configured as one-to-many (*:1) from dimensions to facts with single cross-filter direction.

Rationale:

- Ensures correct cardinality for star schema pattern
- Single direction prevents ambiguous filter propagation
- Optimizes query performance by reducing relationship complexity

DAX IMPLEMENTATION STRATEGY

Time Intelligence Approach

Decision: Used direct YEAR() calculations on order timestamps instead of Date table time intelligence functions.

Rationale:

- Timestamp data type compatibility issues with Date table relationships
- Simpler DAX formulas that are easier to maintain and troubleshoot
- Direct calculations on actual order dates ensure 100% data accuracy
- Pragmatic solution that delivers business value without complexity

Error Handling

Decision: Implemented DIVIDE() function with 0 fallback instead of standard division operator.

Rationale:

- Prevents division-by-zero errors that break visualizations
- Gracefully handles edge cases in filtered contexts
- Maintains dashboard stability across all user interactions

Calculated Columns vs Measures

Decision: Created calculated columns only for Order Year and Order Quarter; used measures for all aggregations.

Rationale:

- Calculated columns needed for filtering and slicing (year, quarter)
- Measures preferred for aggregations to leverage filter context
- Minimizes model size by avoiding unnecessary calculated columns
- Enables dynamic calculations that respond to user selections

POWER QUERY TRANSFORMATIONS

Category Translation Strategy

Decision: Merged Portuguese-to-English category translation during data preparation stage in Power Query.

Rationale:

- Better performance than DAX-based RELATED() lookups at query time
- Translation happens once during refresh, not repeatedly during queries
- Results in smaller, cleaner data model without separate translation table

Sorting Helper Columns

Decision: Created numeric helper columns (Month Number, Quarter Sort) in Power Query for proper chronological sorting.

Rationale:

- Text-based month names sort alphabetically (incorrect: April, August, December...)

- Numeric columns ensure proper calendar order (correct: January=1, February=2...)
- Power Query is the appropriate layer for data preparation tasks

Data Type Optimization

Decision: Handled all data type conversions in Power Query rather than using DAX type conversion functions.

Rationale:

- Power Query transformations execute once during refresh
- DAX calculations execute repeatedly during user interactions
- Proper data types enable better compression and faster queries

VISUALIZATION DESIGN CHOICES

KPI Cards for Executive Summary

Decision: Placed five large KPI cards at the top of the dashboard showing Total Sales, Total Orders, Total Customers, Average Order Value, and Total Products Sold.

Rationale:

- Provides immediate executive summary visible without scrolling
- Large numbers with distinct colors improve scannability
- Follows established BI dashboard best practices

Chart Type Selection

Decision: Used bar charts for category/state rankings and line chart for quarterly trends.

Rationale:

- Bar charts: Best for comparing discrete categories (top-to-bottom ranking is intuitive)
- Line chart: Best for showing trends over continuous time periods
- Data labels on bars eliminate need to read axis scale

Conditional Formatting

Decision: Applied color gradients to bar charts (light to dark blue/orange based on values).

Rationale:

- Creates visual hierarchy highlighting top performers
- Reduces cognitive load by pre-attentively directing attention
- Maintains clean design without adding clutter

Interactive Features

Decision: Implemented bookmarks for preset views (Full Dataset, 2018 Only, Health & Beauty) with navigation buttons.

Rationale:

- Enables one-click access to common analysis scenarios
- Improves user experience for non-technical stakeholders
- Demonstrates advanced Power BI interactivity capabilities

PERFORMANCE OPTIMIZATION

Relationship Optimization

Decision: Configured all relationships as one-to-many from dimensions to facts with single cross-filter direction.

Rationale:

- Reduces query complexity and execution time
- Prevents ambiguous filter propagation paths
- Follows Power BI best practices for star schema designs

DAX Efficiency

Decision: Avoided row context where unnecessary; used filter context and aggregation functions instead.

Rationale:

- Row context forces evaluation row-by-row (slow on large datasets)
- Filter context enables bulk operations (faster)
- Simple aggregation functions (SUM, DISTINCTCOUNT) leverage engine optimizations

Column vs Measure Trade-offs

Decision: Minimized calculated columns; used only when necessary for filtering/sorting.

Rationale:

- Calculated columns consume memory (stored in model)
- Measures compute on-the-fly (no storage cost)
- Order Year and Order Quarter columns essential for slicing (worth the memory cost)

CHALLENGES & SOLUTIONS

Date Table Relationship Issues

Challenge: Standard Date table time intelligence functions did not work reliably due to timestamp-to-date type mismatch.

Solution: Created calculated columns extracting year and quarter directly from order timestamps, bypassing Date table dependency.

Learning: *Pragmatic solutions that work are better than theoretically perfect approaches that fail. Direct date extraction proved more reliable than complex relationship configurations.*

Text Sorting in Chronological Order

Challenge: Month names (January, February, March) sorted alphabetically instead of chronologically in visualizations.

Solution: Created numeric helper column (Month Number: 1-12) and configured 'Sort by Column' to use it for proper ordering.

Learning: *Power BI's 'Sort by Column' feature is essential for maintaining logical order of text-based time periods.*

Anonymized Data Constraints

Challenge: Dataset contained anonymized product and customer identifiers without descriptive names.

Solution: Focused analysis on available categorical data (product categories, geographic states, time periods) rather than individual products/customers.

Learning: *Working with real-world anonymized data mirrors enterprise scenarios where privacy and compliance restrict detailed analysis. This constraint demonstrated adaptability and business understanding.*

KEY TAKEAWAYS

- **Pragmatic solutions that work are preferable to theoretically perfect approaches that fail in practice**
- **Data preparation in Power Query (one-time cost) performs better than DAX calculations (repeated cost)**
- **Star schema design principles deliver measurable performance benefits and maintainability advantages**
- **User experience improvements (bookmarks, tooltips, conditional formatting) differentiate good dashboards from great ones**
- **Problem-solving and adaptability matter more than perfect execution of textbook patterns**
