

# Looker Ecommerce Data Warehouse

Business Insights Using SQL Data Warehouse & Power BI

---

End-to-End Analytics Solution

February 2026

*Mustafa Mahmoud Mohamed*

# Project Overview

## Objective

To analyze e-commerce sales data, identify business problems, and provide actionable insights through a comprehensive data warehouse and interactive Power BI dashboard.

## Scope & Tools

Raw e-commerce data (7 tables, 100K+ records) • SQL Server for ETL & DWH • Power BI for visualization • Medallion Architecture (Bronze/Silver/Gold) • Galaxy Schema design

## Project Workflow



# Business Problem & Key Questions

## Why This Project Matters

E-commerce success depends on understanding customer behavior, optimizing product mix, and identifying profitable markets. This project transforms raw transactional data into strategic intelligence.

### Sales & Revenue Questions

- What is the annual profitability trend (2020-2024)?
- How are sales distributed across price ranges?
- Which countries generate highest revenue?
- Which brands have the highest sales volume?
- What is the monthly profit distribution?

### Customer & Marketing Questions

- What is the customer conversion rate?
- Which traffic sources are most effective?
- What is the demographic distribution?
- Which browsers are most commonly used?
- What percentage never completed a purchase?

# Data Collection

## Data Sources

### Kaggle

[Looker Ecommerce BigQuery Dataset](#)

### 7 Core Tables

users, orders, order\_items, products,  
inventory\_items, events, distribution\_centers

### Volume

84,000+ customers, 100,000+ events

## Data Characteristics

Attribute	Value
Total Customers	84,000+
Countries	74
Cities	2,900+
Time Period	2019-2024
Events Tracked	100,000+

## Data Quality Challenges Identified

**Missing Values:** NULL user\_ids (13K+ rows), NULL product names/brands

**Duplicates:** Duplicate users (same email+name), duplicate event sequences per IP

**Inconsistencies:** Mixed data types (float IDs), non-standard values (M/F), string dates

**Business Rule Issues:** Multiple order statuses (Processing, Cancelled, Shipped) requiring filtering

# Exploratory Data Analysis (EDA)

*Systematic investigation to identify data quality issues and patterns*

## EDA Methodology

### Schema Inspection

sp\_help for structure, data type validation

### NULL Analysis

Count NULLs per column, identify critical gaps

### Duplicate Detection

GROUP BY + HAVING, business key analysis

### Distribution Analysis

Distinct counts, statistical summaries (MIN, MAX, AVG)

### Relationship Validation

Foreign key integrity, orphaned records

## Top EDA Insights

### Age Distribution

Average customer age: 40-50 years

### Gender Split

Nearly equal Male/Female distribution

### Geographic Coverage

74 countries, 2,900+ cities worldwide

### Product Categories

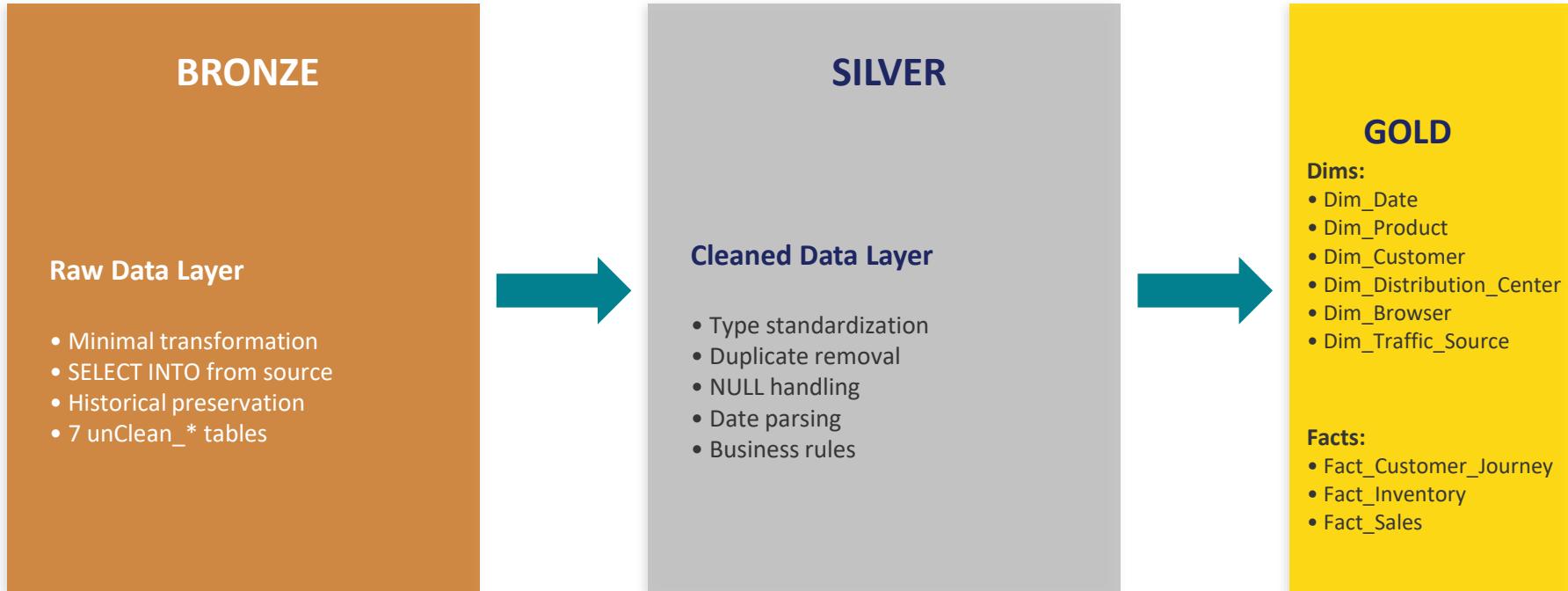
Multiple brands & departments identified

### Data Quality Score

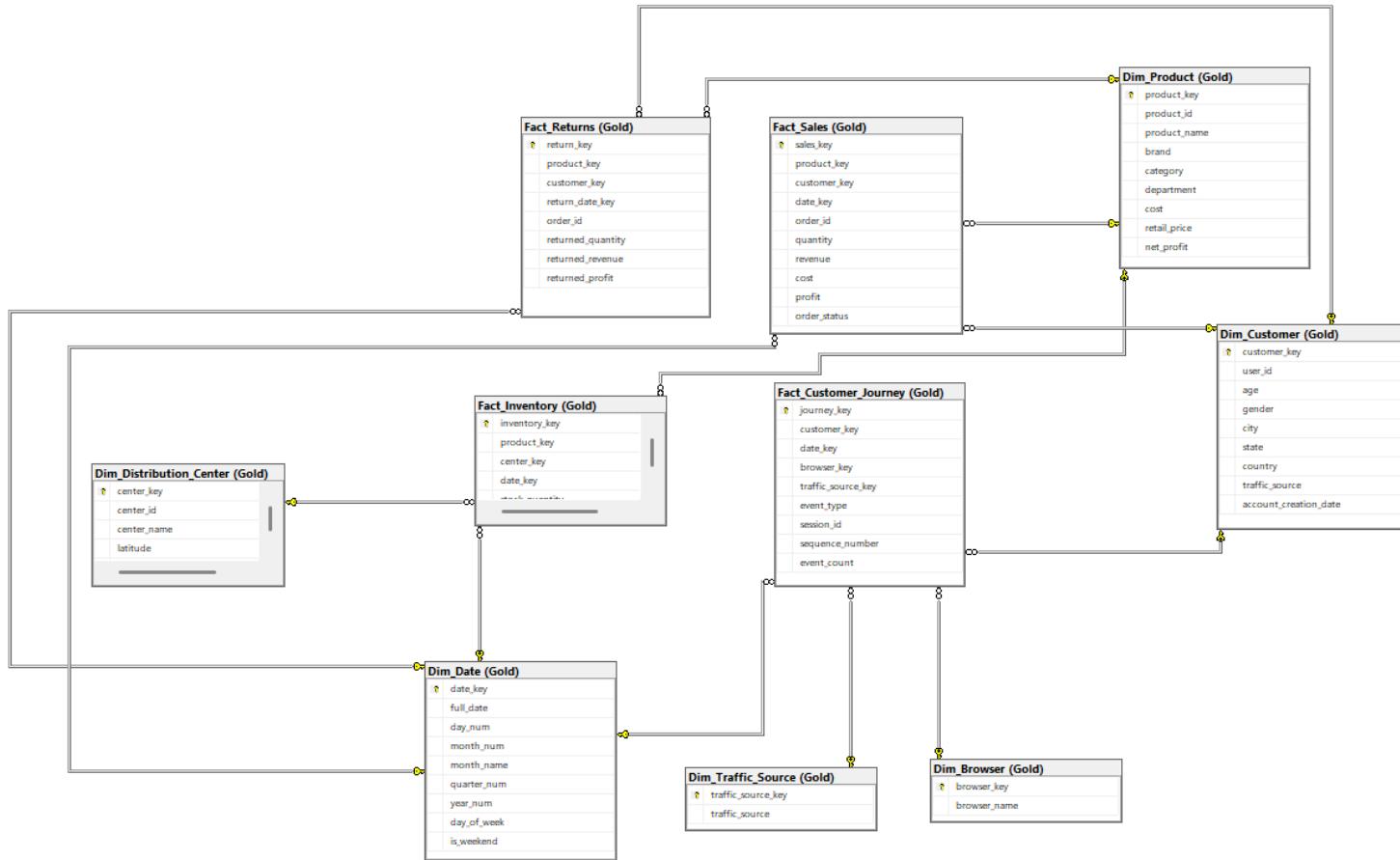
15% of records require cleaning

# Data Warehouse Design

*3-Layer Medallion Architecture: Bronze → Silver → Gold*



# Galaxy schema Design



# ETL Process

*Extraction → Cleaning → Transformation → Loading*



## Implementation Details

### Automation

- T-SQL Stored Procedures: tables\_Bronze, tables\_creation, Load\_data, Goald\_tables
- Automated schema creation and data population

### Data Transformation Techniques

- CAST(CAST(id AS FLOAT) AS INT) for type conversion
- ROW\_NUMBER() OVER (PARTITION BY...) for deduplication
- SUBSTRING() for date component extraction
- Business logic: net\_profit = retail\_price - cost

# Power BI Dashboard Design

## Sales & Revenue

54K

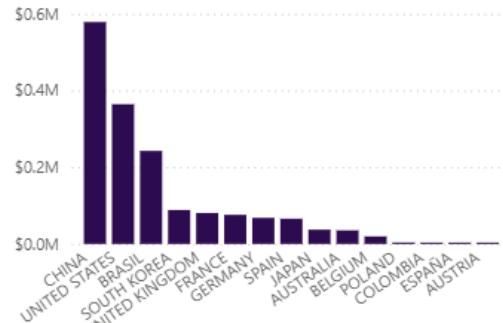
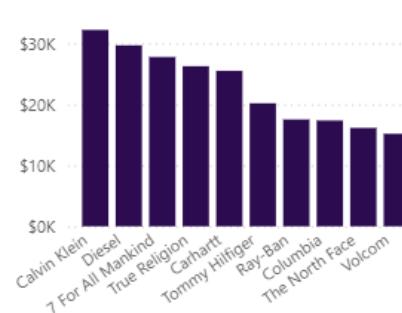
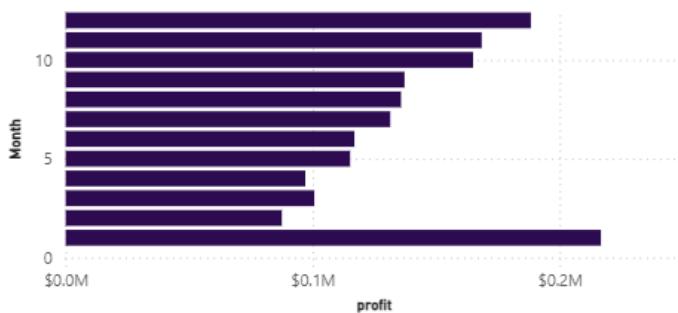
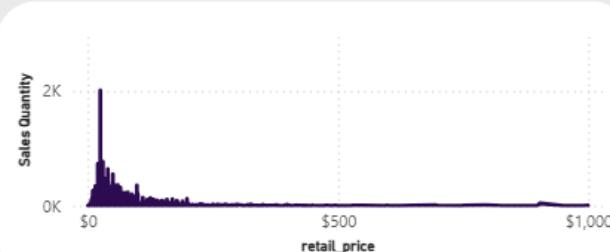
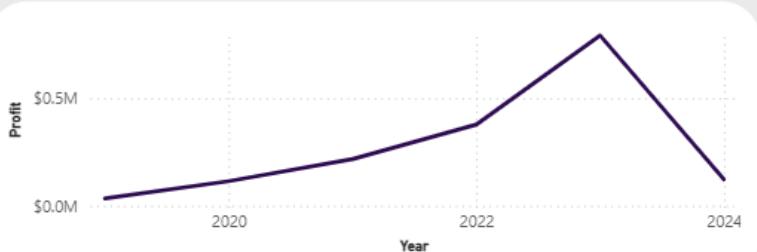
Quantity

\$1.66M

profit

\$3.19M

Revenue



Category

# Power BI Dashboard Design



## Customer Insights

\$1.66M

profit

\$3.19M

Revenue

84K

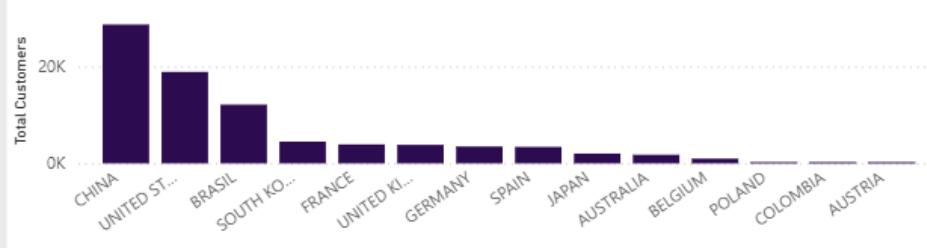
Total Customers

19K

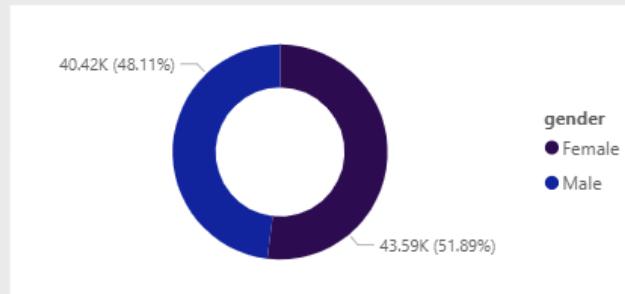
Customers Never Purchased

65K

Customers With Purchase



traffic_source	Chrome	Firefox	IE	Other	Safari	Total
Search	370199	148064	36253	36214	145769	736499
Organic	79573	31523	7787	8339	31782	159004
Facebook	31716	11786	3025	3046	12440	62013
Email	27070	10993	2713	2442	10560	53778
Display	21160	8706	2057	1836	8181	41940
<b>Total</b>	<b>529718</b>	<b>211072</b>	<b>51835</b>	<b>51877</b>	<b>208732</b>	<b>1053234</b>



# Power BI Dashboard Design

## Dashboard Structure & Features

### Navigation & Filters

Slicers for Category

### Visual Selection Rationale

Bar charts for rankings • Line charts for trends • KPI cards for key metrics • Tables for detailed data

## Sales & Revenue Page

KPIs: Quantity, Profit, Revenue

### Visuals:

- Line chart: Profitability trend (2019-2024)
- Histogram: Sales by price range
- Bar chart: Top countries by revenue
- Bar chart: Profit by brand
- Table: Top 5 brands by quantity

## Customer Insights Page

KPIs: Total, Profit, Revenue, Conversion

### Visuals:

- Bar chart: Top countries by customer count
- Bar chart: Average age by country
- Donut chart: Gender distribution
- Table: Traffic source effectiveness
- Table: Browser usage statistics

# Data Analysis & Key Insights

*Answering business questions using Gold layer in Power BI*

## Annual Profitability Trend (2020-2024)

*Visualization: Line chart showing yearly profit*

**Insight:** Steady growth 2019-2022, peak in 2022-2023, decline in 2024

*Action:* Investigate 2024 decline - analyze costs, pricing, competition

## Sales Distribution by Price Range

*Visualization: Histogram/bar chart of price brackets*

**Insight:** Most sales in \$0-\$100 range, peak at \$50-\$75

*Action:* Focus inventory on sweet spot, introduce premium products strategically

## Top Revenue Countries

*Visualization: Bar chart of countries by revenue*

**Insight:** China (\$30K+), United States, South Korea lead

*Action:* Prioritize these markets for marketing spend & expansion

# Customer & Marketing Insights

**\$1.66M**

Total Profit

**\$3.19M**

Revenue

**84K**

Customers

**77%**

Conversion

**70%**

From Search

## Customer Analysis

**Conversion Rate: 77% (65K/84K)**

19K registered but never purchased

### Demographics

- Nearly equal gender split (48% F / 52% M)
- Average age: 40-50 years
- China leads in customer count

### Traffic Sources

- Search: 736K (70%)
- Organic, Facebook, Email follow

### Browser Usage

- Chrome, Firefox, IE (top 3)

## Strategic Actions

### 1. Re-engage Non-Purchasers

Target 19K inactive users with personalized campaigns

### 2. Optimize Search Marketing

70% traffic from search - double down on SEO/SEM

### 3. Geographic Focus

Prioritize China, US, Brazil markets

### 4. Browser Compatibility

Ensure optimal experience on Chrome, Firefox, IE

### 5. Brand Strategy

Strengthen partnerships with top performers (Speedo, Quiksilver, O'Neill)

# Challenges & Solutions

Challenge	Solution	Result
13K+ NULL user_ids in events	DELETE rows + ALTER COLUMN to NOT NULL	Clean event data with referential integrity
Duplicate users (email + name)	ROW_NUMBER() window function deduplication	Unique customer base for analysis
Float IDs requiring integer conversion	CAST(CAST(id AS FLOAT) AS INT)	Proper data types for FK relationships
String timestamps needing parsing	SUBSTRING + date component extraction	Date parts for time-based analysis
Mixed order statuses (incomplete)	Keep only Completed & Returned statuses	Accurate sales & return metrics
No profit calculation available	ADD net_profit = retail_price - cost	Profitability analysis enabled
Complex dimensional model design	Galaxy schema with shared dimensions	Flexible multi-fact analysis
Dashboard clarity & usability	Interactive slicers, clear KPIs, intuitive layout	User-friendly self-service BI

# Summary & Key Takeaways

## Project Accomplishments

- Comprehensive EDA performed
- 3-layer DWH implemented
- Galaxy schema with 3 facts created
- 13K+ records cleaned
- Power BI dashboards built
- Actionable insights delivered

## Key Business Insights

- 77% conversion rate achieved
- 19K re-engagement opportunity
- Search drives 70% of traffic
- \$50-\$75 price sweet spot
- 2024 decline needs investigation
- Focus on China, US, Brazil markets

## Tools & Skills Demonstrated

**Technical Skills:** SQL Server • T-SQL • Stored Procedures • Window Functions • Data Modeling

**Analytical Skills:** EDA • Data Cleaning • ETL Design • Galaxy Schema • Business Intelligence

**Visualization:** Power BI • Dashboard Design • KPI Development • Interactive Reporting

*Thank you for your attention*