Olist E-Commerce Sales & Fulfillment Analysis

- Driving Business Insight Through SQL, Python, and Visual Analytics
- Presented by: Joe Lam

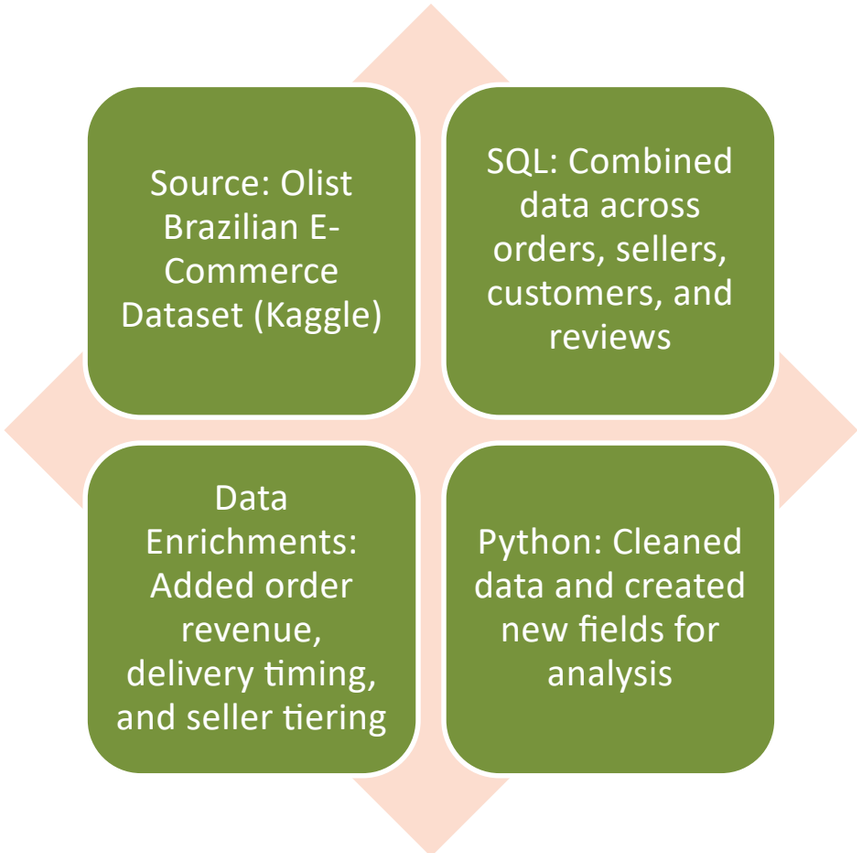
Project Overview

GOAL: ANALYZE OLIST'S ORDER DATA FOR KEY SALES AND FULFILLMENT TRENDS.

TOOLS USED: SQL, PYTHON (PANDAS, SEABORN, MATPLOTLIB), MYSQL, JUPYTER NOTEBOOK

FOCUS AREAS: FULFILLMENT, SELLER & CUSTOMER SEGMENTS, ANOMALIES

Data Sources & Preparation



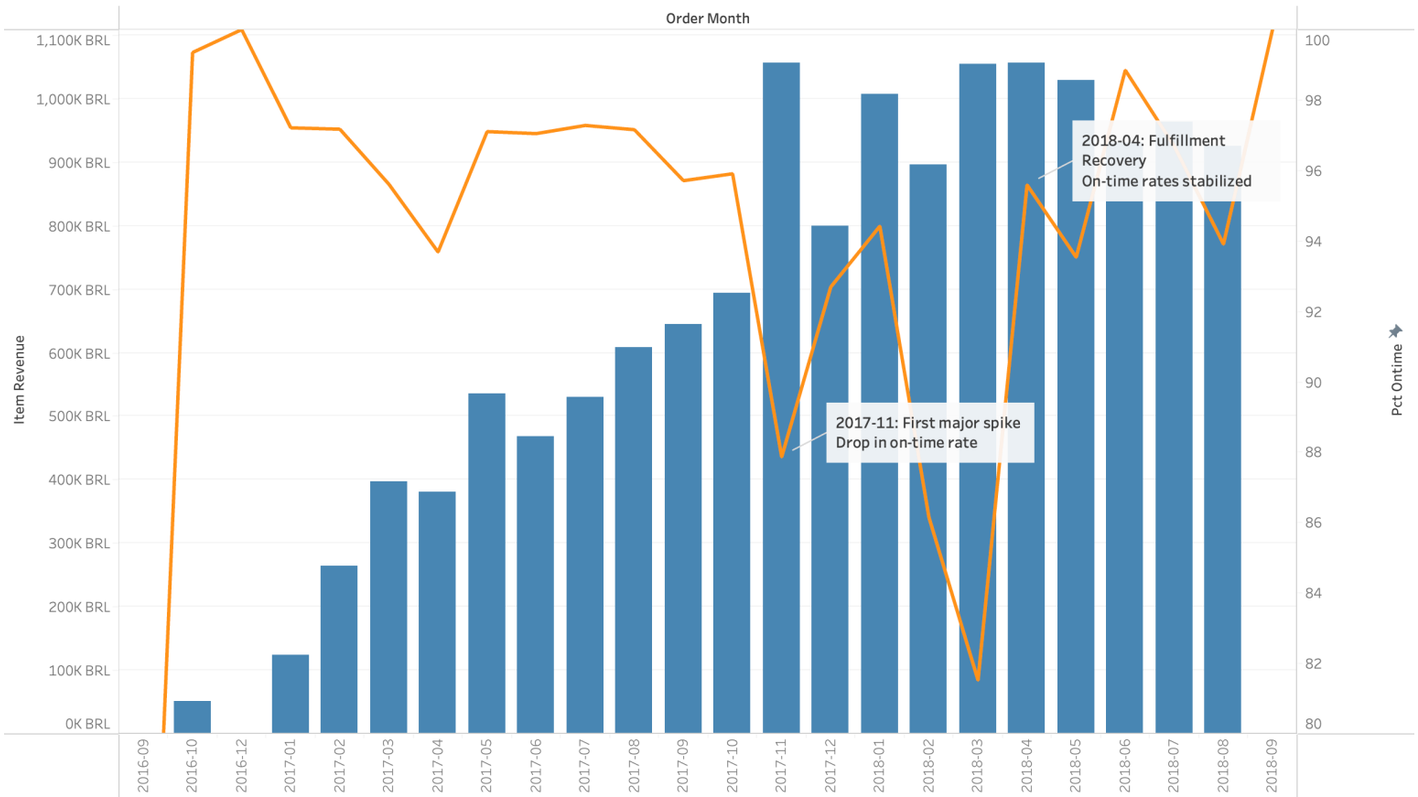
What the Data Reveals: Key Business Insights

- 📉 **Fulfillment Lag During Growth**: From **Nov 2017 to Mar 2018**, on-time delivery rates dropped as monthly revenue surged, suggesting capacity constraints.
- 🛠️ **Operational Recovery**: By **Apr 2018**, Olist recovered fulfillment performance while sustaining growth.
- 🏆 **Seller Concentration**: 17.9% of sellers generate 80% of revenue, roughly confirming the Pareto Principle.
- 🚩 **High-Value Orders More Likely Delayed**: Anomaly detection flagged that large orders were disproportionately affected by fulfillment delays.
- 📦 **Late Deliveries Hurt Satisfaction**: On-time orders averaged a 4.29 review score, while severely delayed ones dropped to 1.73 — quantifying the fulfillment–CX connection.

Operational Strain & Recovery

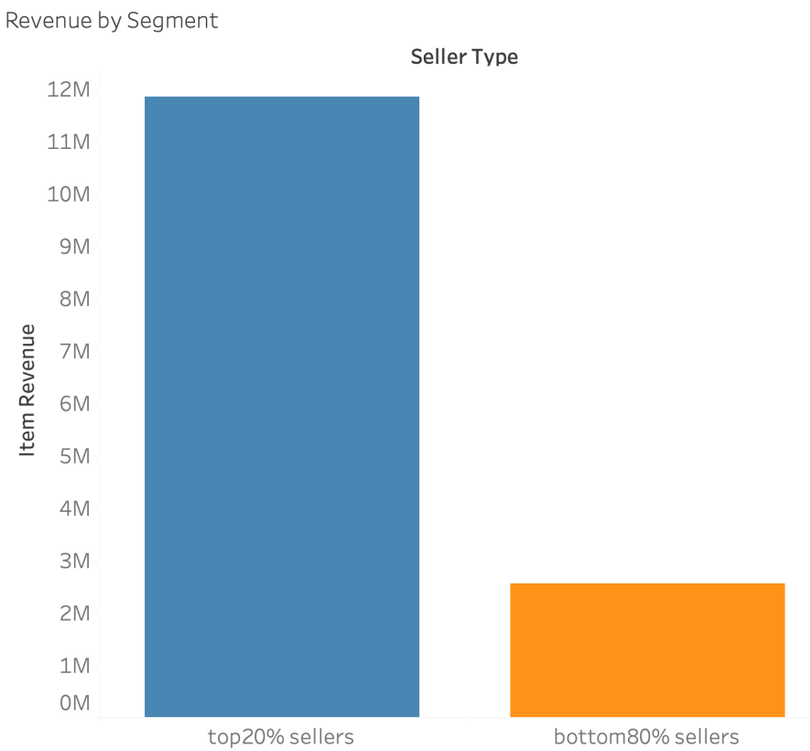
Delivery Delays Amid Rapid Growth

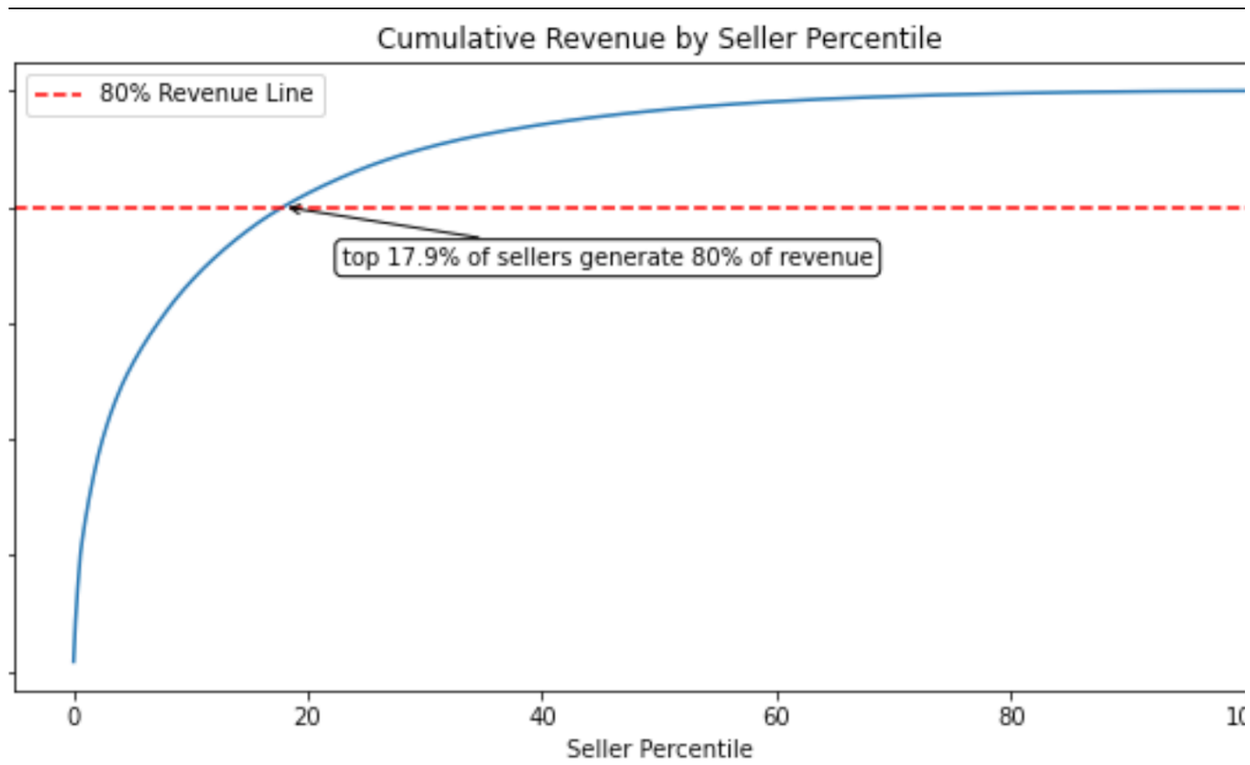
- Fulfillment lag during growth: From Nov 2017 to Mar 2018
- Operational recovery by Apr 2018, likely reflects internal improvements, such as:
  - Scaling fulfillment operations
  - Optimizing delivery logistics
  - Expanding warehouse capacity or third-party delivery support



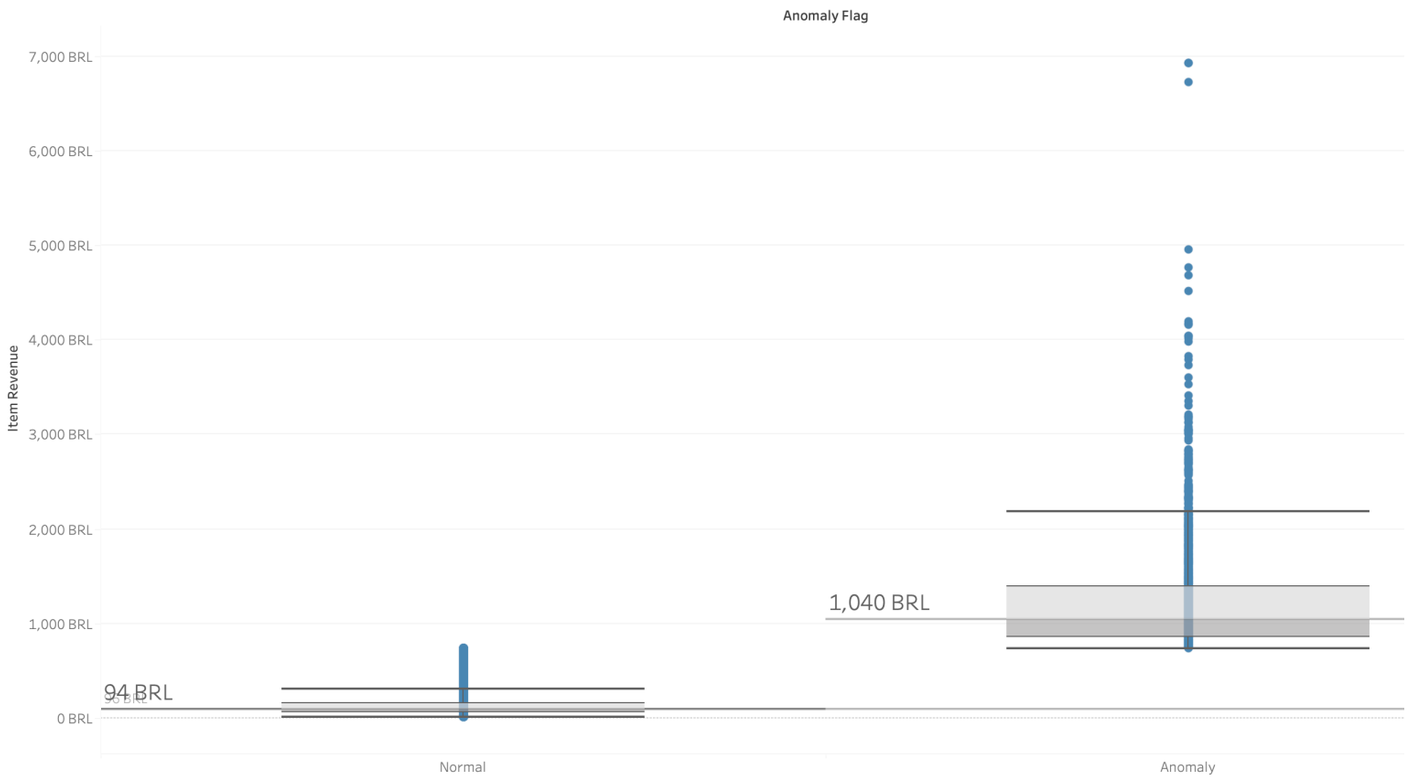
Seller Revenue Concentrated in Top Sellers

- Top 17.9% of sellers drive 80% of revenue — consistent with the Pareto Principle
- Prioritizing top sellers enables high-impact growth without scaling effort across all sellers





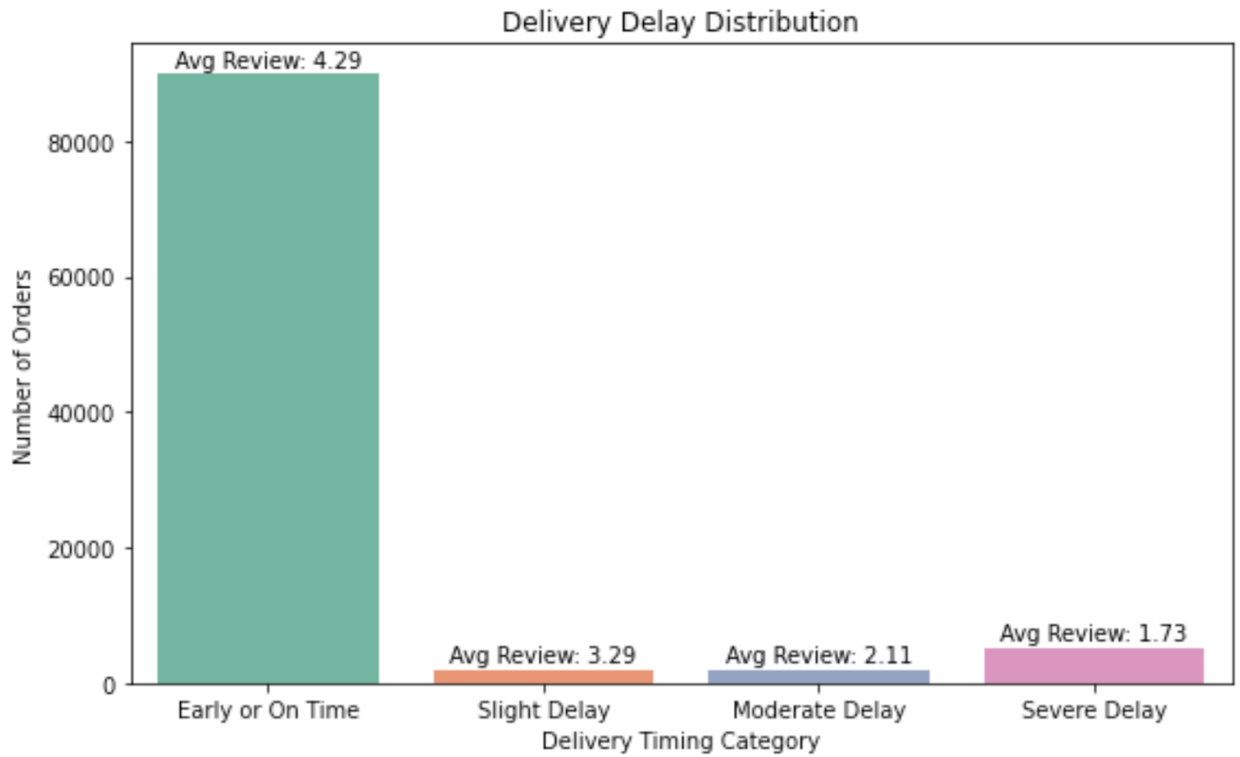
Order Revenue Distribution: Anomalies vs Normal Orders

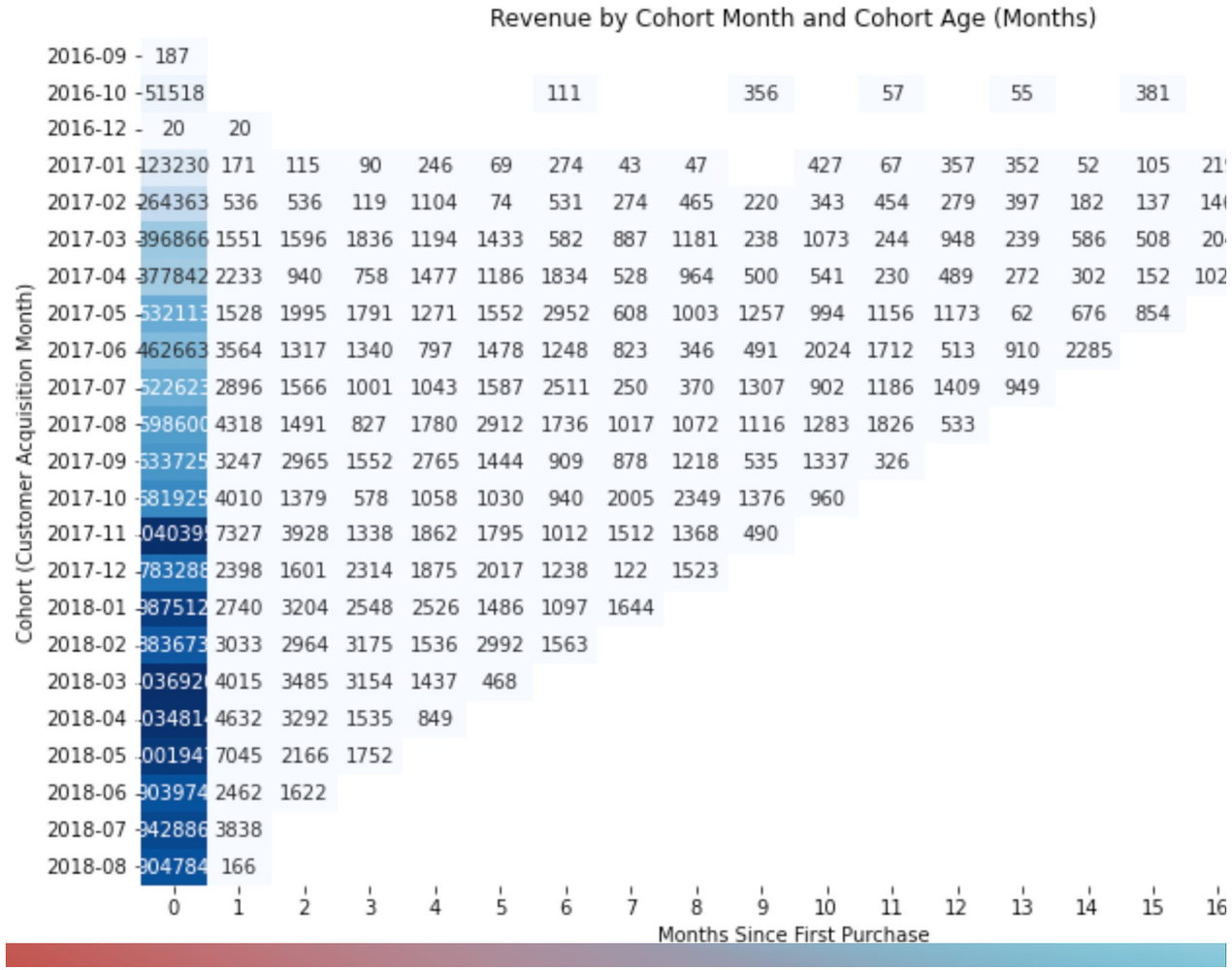


- Anomalies were classified as anything beyond three standard deviations delay
- ❌ High-value orders were disproportionately flagged as anomalies
- 📦 Indicates potential operational strain in fulfilling large orders
- 🧩 Delays may increase risks of cancellations, dissatisfaction, or revenue loss

Delivery Delay & Review Impact

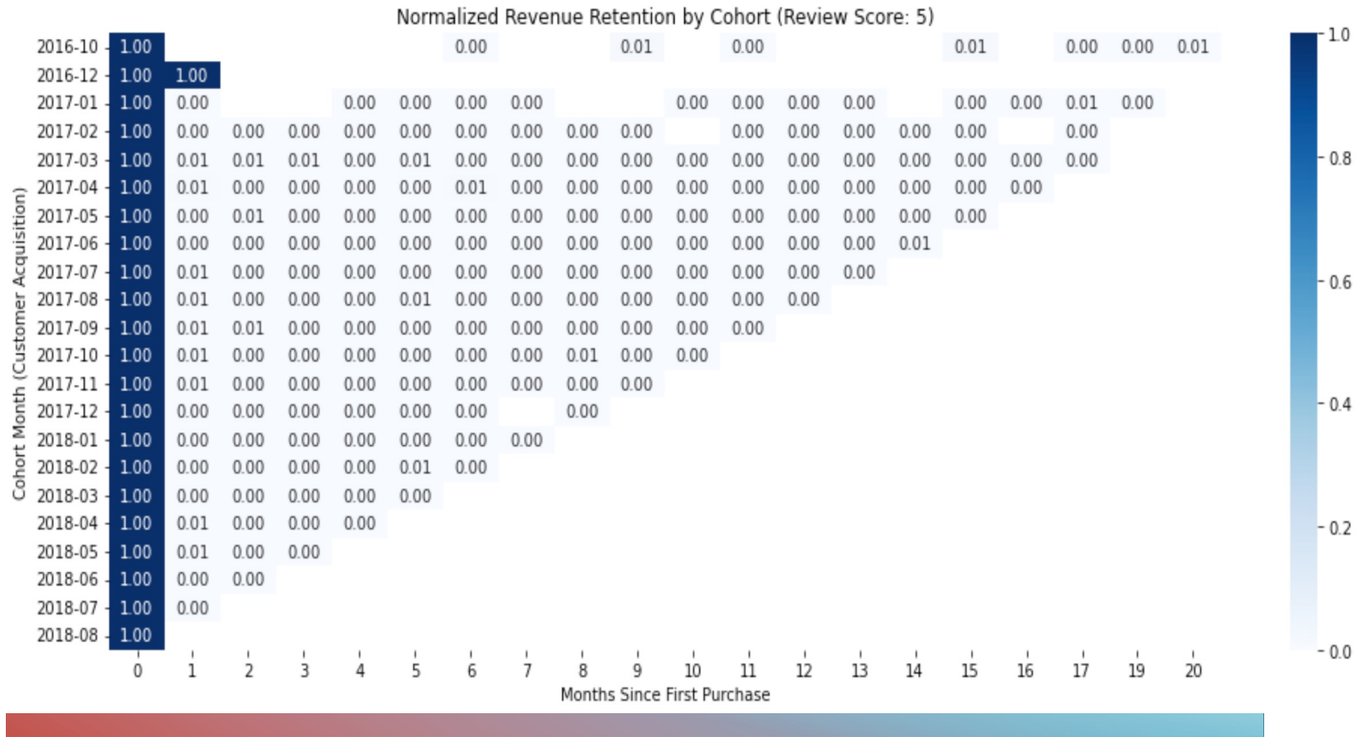
- As expected, Early or On Time deliveries score much higher reviews than late deliveries
- Review ratings drop sharply as delivery delays increase — from 4.29 (on-time) to 1.73 (severe delays), highlighting the direct link between fulfillment and customer satisfaction





Cohort Analysis Reveals One-Time Buyers Dominant

- Most cohorts show strong first-month revenue, but little activity in subsequent months — suggesting minimal repeat purchasing



Cohort Analysis Review score 5

- Review Score 5 ≠ Repeat Customer: Retention Data Inconsistencies
- Likely ID reassignment obscures returning buyer behavior
- This undermines metrics like churn and LTV, and limits customer lifetime analysis

Recommendations

On-time delivery impacts satisfaction

- Improve peak period planning to prevent delays and protect CX

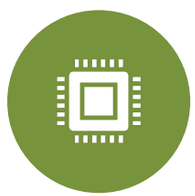
Top 20% sellers drive most revenue

- Launch a top-seller success program to boost high-impact growth

Delayed large orders = Risk

- Audit fulfillment pipeline for high-value order delays

Conclusion



THIS PROJECT SIMULATES A REAL-WORLD BUSINESS SCENARIO:

SQL IS USED TO CREATE CLEAN, JOINABLE DATASETS FROM DISPARATE SOURCES

EDA UNCOVERS FULFILLMENT BOTTLENECKS DURING GROWTH

COHORT ANALYSIS REVEALED UNRELIABLE TRACKING OF REPEAT CUSTOMERS

VISUALIZATION CONNECTS OPERATIONAL ISSUES WITH LOST REVENUE

KEY TAKEAWAY: DATA SCIENCE IS NOT JUST ALGORITHMS — IT'S UNDERSTANDING THE BUSINESS, CHALLENGING ASSUMPTIONS, AND KNOWING WHAT THE DATA CAN AND CANNOT TELL YOU