



E-Commerce Business Performance & Customer Intelligence Report

SQL-Driven Analytics & RFM-Based Segmentation

1. Executive Summary

This project delivers a comprehensive business intelligence analysis of a large-scale Brazilian e-commerce dataset using SQL and Python. The objective was to uncover revenue drivers, evaluate operational efficiency, understand customer purchasing behavior, and design actionable growth strategies through data-driven insights.

Through structured SQL querying and advanced analytical modeling (RFM segmentation), this project identifies key revenue segments, customer retention gaps, geographic concentration risks, and operational bottlenecks.

Key Outcomes:

- 97% delivery success rate (strong operational execution)
- Revenue growth trend from 2016 to 2018
- Loyal customers identified as primary revenue drivers
- High customer churn risk due to low repeat purchase frequency
- Significant delivery time inefficiency impacting retention
- Revenue concentration risk in São Paulo region

This analysis transforms raw transactional data into strategic business insights aligned with executive decision-making.

2. Business Context & Dataset Overview

The dataset represents a multi-category Brazilian e-commerce marketplace containing customer, seller, order, payment, product, and delivery data.

Core Business Metrics

- Total Customers: 96,096
- Total Orders: 99,441
- Total Products: 32,951
- Total Categories: 73
- Total Sellers: 3,095
- Total Revenue: 15,843,553
- Average Order Value (AOV): 160.57

Strategic Interpretation

- Orders ≈ Customers → Majority are one-time buyers.
 - Low AOV → Mass-market positioning.
 - Revenue skewness indicates presence of high-value minority customers.
 - Business operates on scale rather than luxury positioning.
-

3. Operational Performance Analysis

Order Fulfillment Performance

- Delivered: 96,478
- Canceled + Unavailable: < 2%
- Delivery Success Rate: 97%

Executive Insight

Operational reliability is strong. The logistics system is stable and cancellation risk is low. However, reliability does not equal speed — which becomes critical later in retention analysis.

4. Revenue Trend & Growth Trajectory

Time Period Analyzed:

September 2016 – October 2018

Observations:

- Strong scaling from 2017 onward
- 2018 outperformed 2017
- No declining revenue pattern
- End-of-period dip likely due to incomplete data month

Business Interpretation

The business was in a high-growth expansion phase during the analysis window. Market demand was increasing consistently, showing strong product-market fit.

5. Geographic Revenue Concentration

Top Revenue States:

- SP (Dominant)
- RJ
- MG
- RS

- PR

Strategic Risk Identified

Revenue is heavily concentrated in São Paulo (SP), creating geographic dependency risk. Market shocks in a single region could disproportionately impact revenue.

Growth Opportunity

Untapped states represent expansion potential through:

- Targeted marketing
 - Seller onboarding
 - Regional promotions
-

6. Product Category Intelligence

Revenue Leaders:

- health_beauty
- watches_gifts
- bed_bath_table
- sports_leisure
- computers_accessories

Volume Leaders:

- bed_bath_table
- sports_leisure
- furniture_decor

- health_beauty
- housewares

Key Strategic Insight

Revenue drivers ≠ volume drivers.

This reveals:

- Some categories generate margin.
- Others drive scale.
- Portfolio diversification exists between margin and volume products.

This is a positive structural indicator for long-term stability.

7. Payment Behavior & Installment Culture

Payment Preference:

1. Credit Card (dominant)
2. Boleto
3. Voucher
4. Debit Card

Installment Behavior:

- 1–4 installments → Majority
- 5–8 → Moderate
- 9+ → Minimal

Interpretation

The customer base is mid-market.

Customers prefer short-term EMI.

Luxury positioning is limited.

Credit exposure risk is controlled.

8. Seller Ecosystem Concentration

Top Seller Cities:

- São Paulo
- Curitiba
- Rio de Janeiro
- Belo Horizonte

Business Risk

Supply ecosystem is centralized in SP.

Operational shocks in SP could impact fulfillment network.

Strategic diversification of sellers is recommended.

9. Delivery Time Analysis – The Hidden Weakness

Although delivery success rate is 97%, speed is a concern:

Majority deliveries take more than 12 days.

Strategic Impact

Slow delivery:

- Lowers recency score

- Reduces repeat purchases
- Impacts customer satisfaction
- Weakens retention

Logistics speed optimization is the most critical operational improvement area identified in this analysis.

10. Customer Segmentation – RFM Intelligence Model

Customers were segmented using:

- Recency
- Frequency
- Monetary value

Segments Identified:

- Champion
- Loyal
- Regular
- Lost

Revenue Contribution

- Loyal → Highest revenue contribution
- Regular → Growth opportunity
- Champion → High-value but small segment
- Lost → Reactivation potential

Strategic Insight

The business is heavily dependent on Loyal customers.

However:

- Majority customers purchase only once.
- Repeat purchase behavior is moderate.
- Significant churn risk exists.

Retention strategy is underdeveloped.

11. Pareto Revenue Structure

Frequency analysis shows:

- Most customers purchase 1–2 times.
- Small minority contributes disproportionate revenue.

Classic 80/20 structure confirmed.

Retention and LTV optimization could significantly increase profitability without increasing acquisition cost.

12. Key Strategic Insights

1. Business scaling successfully.
2. Operational reliability strong but delivery speed weak.
3. Loyal customers are backbone of revenue.
4. Customer churn risk is high due to low repeat purchases.
5. Revenue geographically concentrated.

6. Portfolio diversification exists between margin & volume categories.
 7. Mid-market positioning confirmed through installment behavior.
-

13. Strategic Recommendations

1. Improve Delivery Speed

Reducing average delivery time can:

- Increase recency scores
- Boost repeat purchase
- Improve customer lifetime value

2. Convert Regular → Loyal

Implement:

- Loyalty programs
- Behavioral targeting
- Discount triggers based on frequency

3. Reactivate Lost Customers

Email marketing + coupons.
Even 10% recovery improves revenue significantly.

4. Increase AOV

- Bundling
- Cross-sell models
- Upselling through installment incentives

5. Reduce Geographic Risk

Expand seller network and marketing outside SP.

14. Technical Implementation

Tools Used:

- SQL (joins, aggregation, grouping, filtering)
- Python (Pandas for data manipulation)
- Matplotlib for visualization
- RFM scoring model
- Heatmap correlation analysis

Project Structure:

- Business Performance Analysis Notebook
 - RFM Customer Segmentation Notebook
 - Structured reporting PDF
-

15. Conclusion

This project demonstrates the ability to:

- Translate raw transactional data into executive-level insights
- Identify structural revenue dependencies
- Evaluate operational bottlenecks
- Build customer intelligence segmentation models

- Provide strategic business recommendations backed by data

The analysis reveals a business with strong growth and operational reliability but significant opportunity in customer retention, delivery optimization, and geographic diversification.

With focused improvements in logistics speed and retention marketing, long-term customer lifetime value can be significantly increased.