OLIST - MARKETING FUNNEL PERFORMANCE REVIEW

SQL + Python

Bryce Smith

TABLE OF CONTENTS

1. Cover & Executive Summary

Title Olist Marketing Funnel Case Study

One-line Tagline Re-engineering channel mix to double conversion efficiency & ROI

Problem Statement Leadership sought a clear picture of channel efficiency across the funnel (Leads → Contacted → Conversions) in order to rebalance spend, plug attribution gaps, and set next-generation performance targets.

Key Metric Lift / \$ Impact
$138 Rev per Lead from “Unknown” channel - more than 2× the Paid benchmark.

Headline Insights
• “Unknown” leads (16 % vol.) convert at 14 % and yield \$138 Rev/Lead - the hidden gem.
• Paid campaigns under-deliver (6.3 % conv, \$48 Rev/Lead) - immediate optimisation required.
• Seasonality analysis shows revenue peaks Jan-Apr, guiding budget re-timing.

2. Business Context & Objectives

Brief Company Description Olist is a Brazilian e-commerce marketplace that relies on inbound marketing channels to acquire buyers. Unclear attribution tags and sub-optimal budget allocation are eroding ROI and masking true channel performance.

Why This Matters
Inefficient spend wastes cash, slows customer acquisition, and obscures funnel bottlenecks. Tight attribution and optimised budget deployment directly impact revenue growth and CAC.

SMART Objectives
• Elevate Paid Conversion Rate to ≥ 8 % within 3 months.
• Achieve ≥ 95 % lead attribution accuracy by end-Q3 2025.
• Reduce Time-to-First-Response to < 5 minutes for all paid leads by Q4 2025.

Item	Include Here
Source name & public link	Kaggle Olist datasets; GitHub repo (<link-placeholder>)
Time span & row/col counts	Orders 99k rows, Leads 20k rows, 10-15 columns each
Key joins / grain	`marketing_leads ↔ customers ↔ orders` on customer_id

3. Data Overview (high-level)

Source name & public link	Kaggle Olist datasets (Kaggle public dataset)
Time span & row/col counts	Orders 99k rows, Leads 20k rows, 10-15 columns each
Key joins / grain	`marketing_leads ↔ customers ↔ orders` on customer_id

(Detailed SQL scripts, UTM parsing, and raw screenshots reside in Appendix A.)

4. Methodology

Analysis Pipeline
• *ETL in PostgreSQL* - cleaned & joined orders, items, leads, reviews.
• *Impute Channels* - cross-reference session logs to tag “Unknown” leads.
• *Metric Computation* - funnel drop-offs, Conv Rate, Rev/Lead via SQL views.
• *Validation* - replicate KPIs in Python/Pandas; peer dashboard review.
• *Visualisation* - Tableau Story Points for Exec walk-through.

Modelling Techniques
Descriptive funnel metrics and seasonality time-series (12-month moving average). No ML required at this stage; future uplift models planned.

(Code and notebooks linked in GitHub - see Appendix C.)

5. Findings & Visual Evidence

Q1. Which channel delivers the most revenue per lead?

Chart 1 - Rev per Lead by Channel
“Unknown” tops the chart at \$138, dwarfing Paid (\$48) and Organic (\$66).

Q2. Where do we lose prospects through the funnel?
Chart 2 - Funnel Drop-off by Channel
Paid sees the steepest decline from Contacted → Converted (-89 %).

Q3. How does seasonality affect revenue?
Chart 3 - Monthly Revenue per Lead
Revenue peaks Jan-Apr then softens mid-year; campaigns should front-load spend.

Implications
• Scale high-value “Unknown” sources once traced.
• Reallocate Paid budget to high-performers; improve landing pages.
• Time campaigns with seasonal peaks for lift.

6. Recommendations & Impact

Prioritised Actions
1. *Audit & Scale “Unknown”* - trace UTM gaps; replicate source tactics (Marketing Ops, Q3 2025).
2. *Optimise Paid Spend* - shift 20 % budget to top ad sets; A/B landing pages (Paid Media, Q3 2025).
3. *Double-Down on Organic* - expand SEO content; enhance nurture flows (SEO Team, Q4 2025).
4. *Leverage Seasonality* - front-load promos Jan-Apr; light retargeting mid-year (Campaign Team, ongoing).
5. *Close Data Gaps* - enforce UTM + lead timestamps; integrate cost feed (Marketing Ops, Q4 2025).

Expected Lift / Savings & Caveats
Conversion rate uplift on Paid to 8 % drives incremental revenue of ≈ \$0.9 M annual; depends on attribution fix and landing-page performance.

Next-Steps Roadmap
UTM taxonomy finalised → Dashboard alert thresholds → 6-month KPI review.

7. Limitations & Assumptions

• Missing UTM tags create attribution uncertainty (16 % “Unknown”).
• Revenue allocation assumes equal basket value; item-level granularity not assessed.
• Seasonality uses one-year history; COVID-era volatility may distort baselines.
• CRM timestamps sometimes lag actual contact events by minutes.

8. Appendix

A. Data Acquisition & Cleaning
• SQL scripts 01-03 with comments; ER diagram.

B. SQL / Python Highlights
• Channel imputation function.
• Funnel KPI view definitions.

C. Full Code & Notebooks
• GitHub:
https://github.com/thebryce15/olist-marketing-funnel

D. Glossary & KPI Definitions
• *Lead* - marketing form fill.
• *Contacted* - sales touch recorded in CRM.
• *Converted* - order placed within 90 days.
• *Rev/Lead* - Net revenue ÷ total leads.