Marketing Funnel Analysis

Cleaned Olist E-commerce Data
Type: Product & Data Analytics Case Study
Dataset: Olist Marketing Funnel (Cleaned)
Tools: Python (pandas)

Author: Shashwat Chauhan
Institution: BITS Pilani, Goa Campus
Date: December 2025

> Analyzing how marketing-qualified leads progress into closed deals, and where drop-offs occur across the funnel.

PRODUCT CONTEXT & PROBLEM

Context

E-commerce platforms depend on structured marketing funnels to convert leads into paying customers. While acquisition volumes may be high, inefficiencies in qualification and conversion stages often limit overall revenue impact.

Understanding how leads move through the funnel is essential for improving product strategy, sales readiness, and growth efficiency.

Problem Statement

This project aims to understand:

- How marketing-qualified leads progress toward closed deals
- Where the largest drop-offs occur in the funnel
- Which stages present the highest opportunity for product or process improvement

The analysis prioritizes **data readiness and funnel integrity** as the foundation for reliable product insights.

DATA SOURCES & FUNNEL SCOPE

Datasets Used

Marketing Qualified Leads - 8,000 unique leads
- First contact date, landing page, acquisition origin

Closed Deals - 842 confirmed conversions
- Business segment, lead type, behavior profile, revenue

Funnel Scope

This study tracks:

First Contact → Marketing Qualified Lead → Closed Deal

Data Ingestion

```
cd = pd.read_csv('olist_closed_deals_dataset.csv')
mql = pd.read_csv('olist_marketing_qualified_leads_dataset.csv')
```

> Ensured multi-table funnel data ingestion for lead-level analysis.

DATA CLEANING & PREPARATION

Why Cleaning Was Necessary

Raw funnel data contained high null density, inconsistent date formats, unused operational fields, and categorical inconsistencies. Without cleaning, funnel analysis would produce unreliable conclusions.

Key Cleaning Actions

Datetime Standardization

```
cd['won_date'] = pd.to_datetime(cd['won_date'])
mql['first_contact_date'] = pd.to_datetime(mql['first_contact_date'])
```

> Standardized timestamps to enable funnel sequencing.

Removing Non-Product Columns

```
cd.drop(
    columns=['seller_id','sdr_id','sr_id',
             'has_company','has_gtin',
             'average_stock','declared_product_catalog_size'],
    inplace=True
)
```

Removed operational noise to focus on product-relevant behavior.

Handling Missing Values

```
cd.fillna('Unknown', inplace=True)
mql.fillna('unknown', inplace=True)
```

Stabilized categorical fields for consistent funnel analysis.

CLEANED DATA OVERVIEW

Dataset Shape Validation

```
cd.info()
mql.info()
```

Validated analysis-ready schema and column integrity.

Lead-Level Uniqueness Check

```
cd[['mql_id','won_date']].duplicated().sum()
mql[['mql_id','first_contact_date']].duplicated().sum()
```

Confirmed one row per lead per funnel stage.

Key Metric

```
conversion_rate = len(cd) / len(mql)
```

≈ 10.5% MQL → Closed Deal conversion

FUNNEL INSIGHTS

Funnel Volume Anchor

```
len(mql), len(cd)
```

Compared top-of-funnel vs conversion volume.

Conversion Rate

```
conversion_rate
```

Computed overall MQL → Closed Deal conversion.

Observations

- Large top-of-funnel drop-off
- Conversion varies by lead type, business segment, and behavior profile
- Revenue distribution is highly skewed

PRODUCT INTERPRETATION & RECOMMENDATIONS

Lead Type Distribution

```
cd['lead_type'].value_counts()
```

Compared conversion distribution across lead types.

Acquisition Channel Overview

```
mql['origin'].value_counts()
```

Analyzed acquisition mix to inform funnel segmentation.

Recommendations

- Strengthen early segmentation and scoring
- Tailor follow-up flows by behavior profile
- Test qualification criteria before sales handoff

LIMITATIONS & CONCLUSION

Revenue Skew Check

```
cd['declared_monthly_revenue'].describe()
```

Observed heavy skew and sparsity in revenue reporting.

Conclusion

Clean, well-structured funnel data enables meaningful product and growth decisions. Improving **lead quality and mid-funnel decision points** offers higher impact than increasing acquisition volume alone.

FINAL APPENDIX — SELECTED CODE SUMMARY

```
# Schema validation
cd['business_type'].unique()
cd['lead_behaviour_profile'].unique()
mql['origin'].unique()

# Final structures
cd.shape
mql.shape
```

Project Report
Author: Shashwat Chauhan
Program: BITS Pilani, Goa Campus
Project: Marketing Funnel Analysis — Olist Dataset
Date: December 2025