PREDICTIVE CUSTOMER ANALYTICS

Problem statement-

As analytics professionals at Olist, a leading Brazilian e-commerce platform, we're looking to make our marketing efforts smarter and more impactful. Our goal is to better understand our diverse customer base — how they shop, how satisfied they are, and how likely they are to stay with us.

By segmenting our customers based on their purchasing behavior and satisfaction levels, we aim to uncover patterns that can help us predict each customer's future value. These insights will guide us in designing highly personalized marketing strategies — the kind that not only boost engagement and sales, but also make customers feel truly understood.

Business Objective

Olist, a Brazilian e-commerce platform, wants to better understand its diverse customer base to optimize marketing investments and drive long-term customer value. With data spanning thousands of orders, reviews, and customer interactions, the goal is to:

- Segment customers based on purchasing behavior, satisfaction, and engagement levels.
- Predict future customer value to focus marketing efforts on high-potential individuals.
- Design tailored marketing strategies that align with each segment's behavior and needs.

By doing so, Olist can allocate its marketing budget more efficiently, boost customer retention, and ultimately enhance customer lifetime value (CLTV) — ensuring both growth and sustainability in a competitive marketplace.

Dataset summary-

The dataset provides a comprehensive view of the Olist e-commerce platform's operations in Brazil from 2016 to 2018, capturing the full customer journey — from order placement to final review.

It overall data is in the form of 9 different dataset.

- olist_orders_dataset.csv
- olist_customers_dataset.csv
- olist_order_items_dataset.csv
- olist_products_dataset.csv
- product_category_name_translation.csv
- olist_order_reviews_dataset.csv
- olist_order_payments_dataset.csv
- olist_sellers_dataset.csv
- olist_geolocation_dataset.csv

Key Customer Metrics at a Glance

• Customer Lifetime Value (CLV)

**CLV represents the total revenue generated from a customer over their lifetime.**

Distribution of Customer Lifetime Value (CLV)
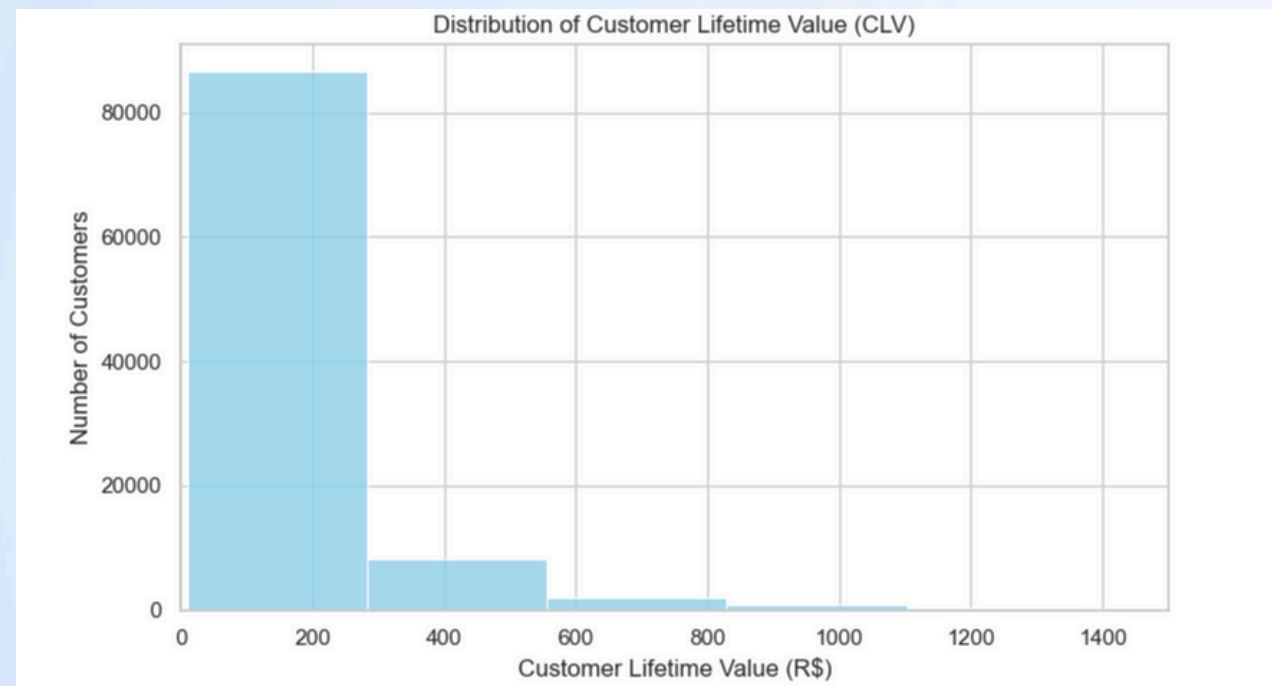
Average CLV: R$ 168.69

Most customers contribute small amounts, but a few account for a large portion of revenue.

Business Takeaway: Focus retention and loyalty programs on the high CLV segment to maximize ROI.

• Customer Satisfaction (Review Score)

Average rating from post-purchase reviews (scale of 1 to 5).

Average Review Score Distribution
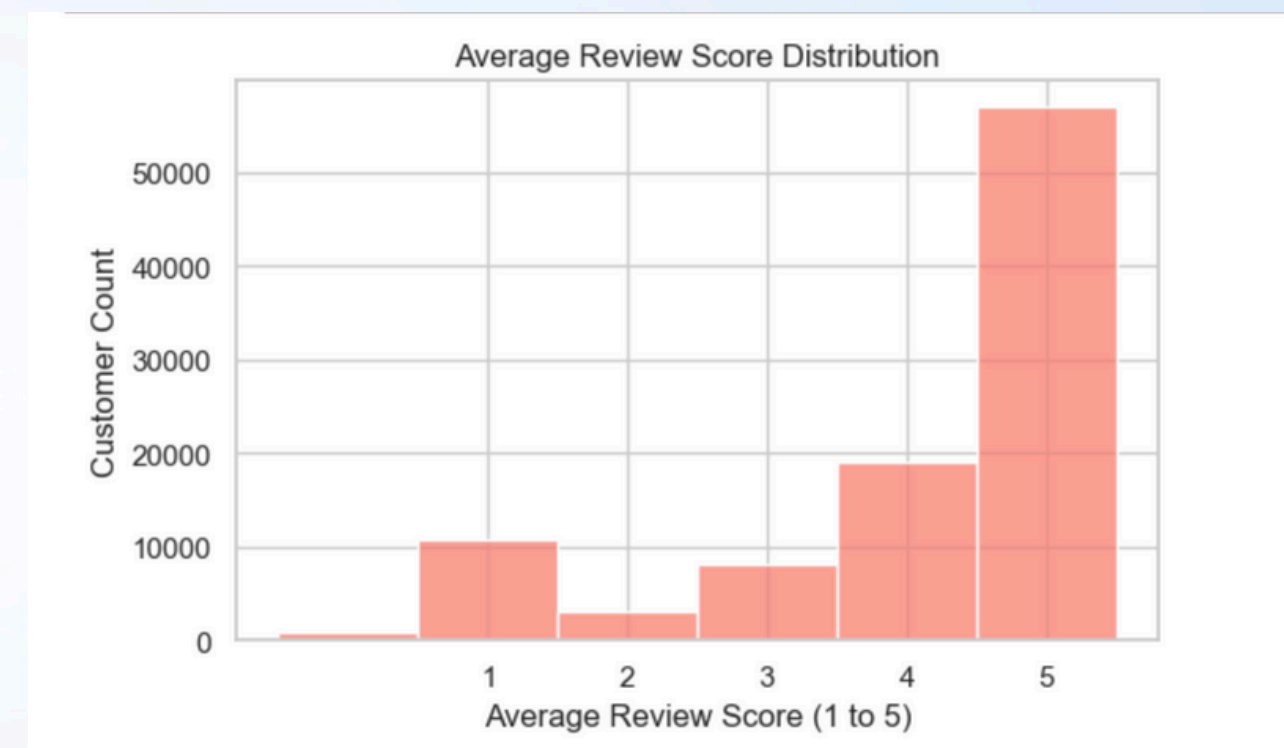
Average Review Score: 4.07

Some 1-star reviews still exist — sentiment analysis suggests common themes (e.g., late delivery, product mismatch).

Business Takeaway: Maintain product quality & delivery speed to keep satisfaction high and reduce churn.

Key Customer Metrics at a Glance

• Delivery Experience

Measures time from purchase to delivery.

Average Delivery Days Distribution
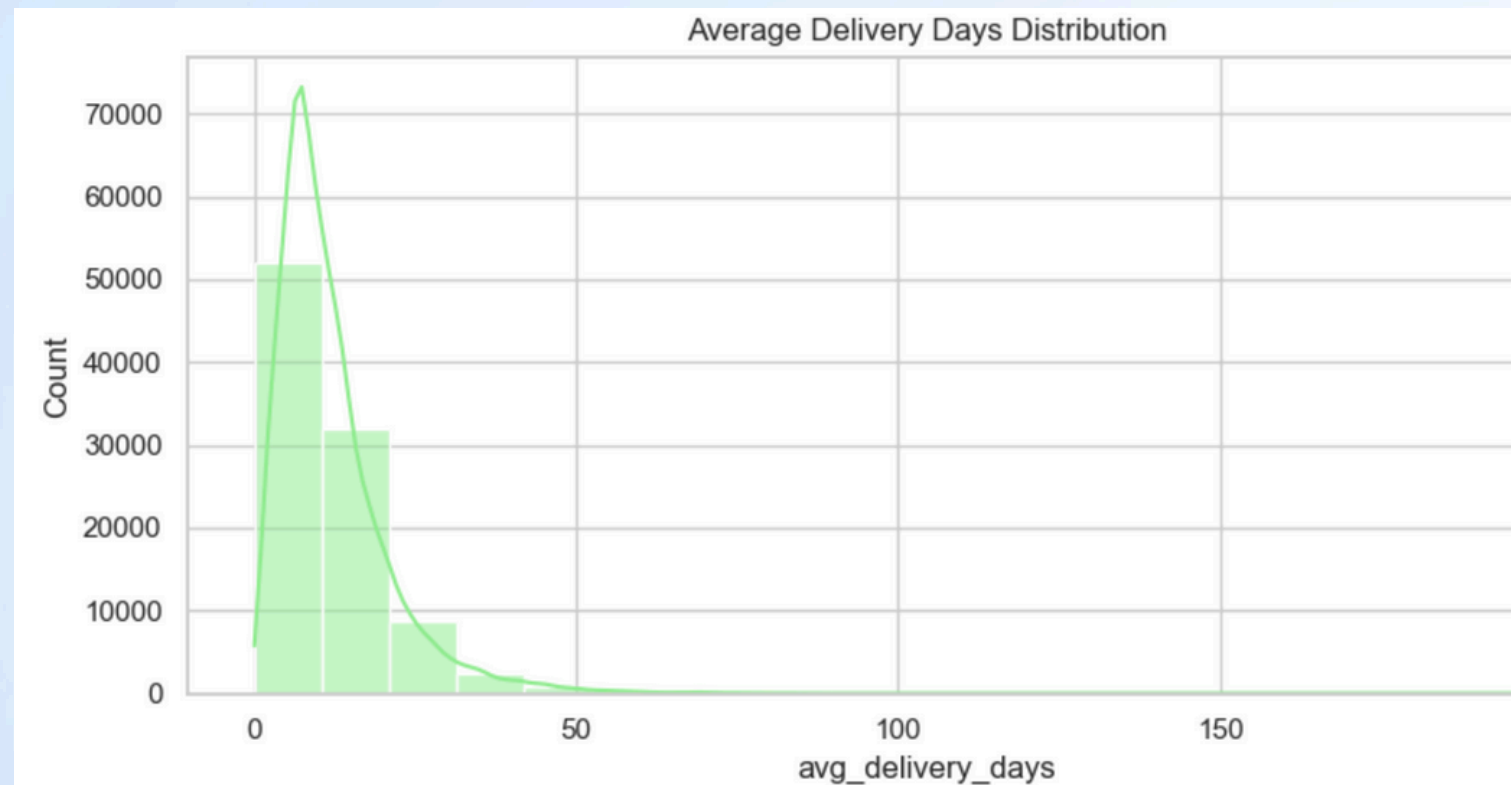
Average Delivery Time: 12.02 days

Long-tail: Some deliveries exceed 20+ days, especially for rural areas.

Business Takeaway: Streamlining last-mile logistics can improve both satisfaction and retention.

• CLV Segment Distribution

Customers were segmented into Low, Medium, and High CLV tiers using RFM analysis.
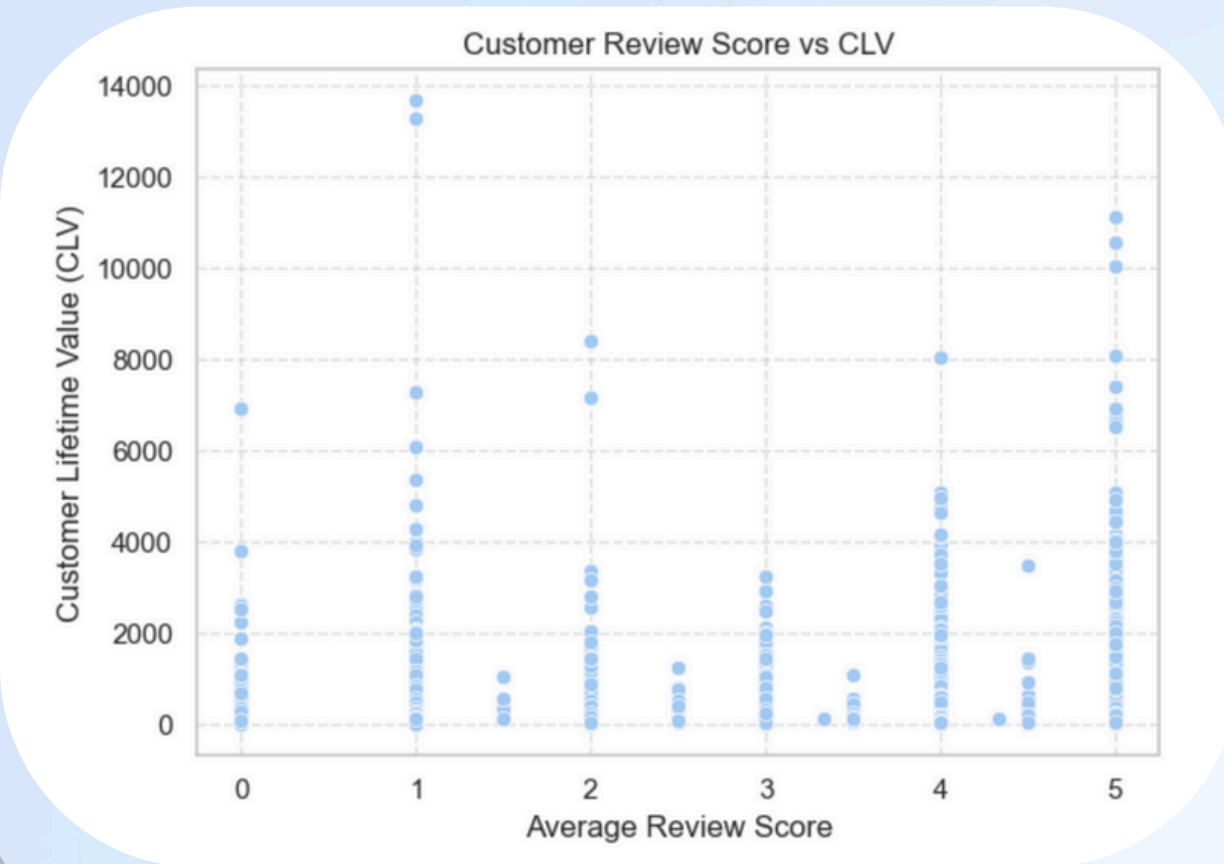

Customer Count per CLV Segment
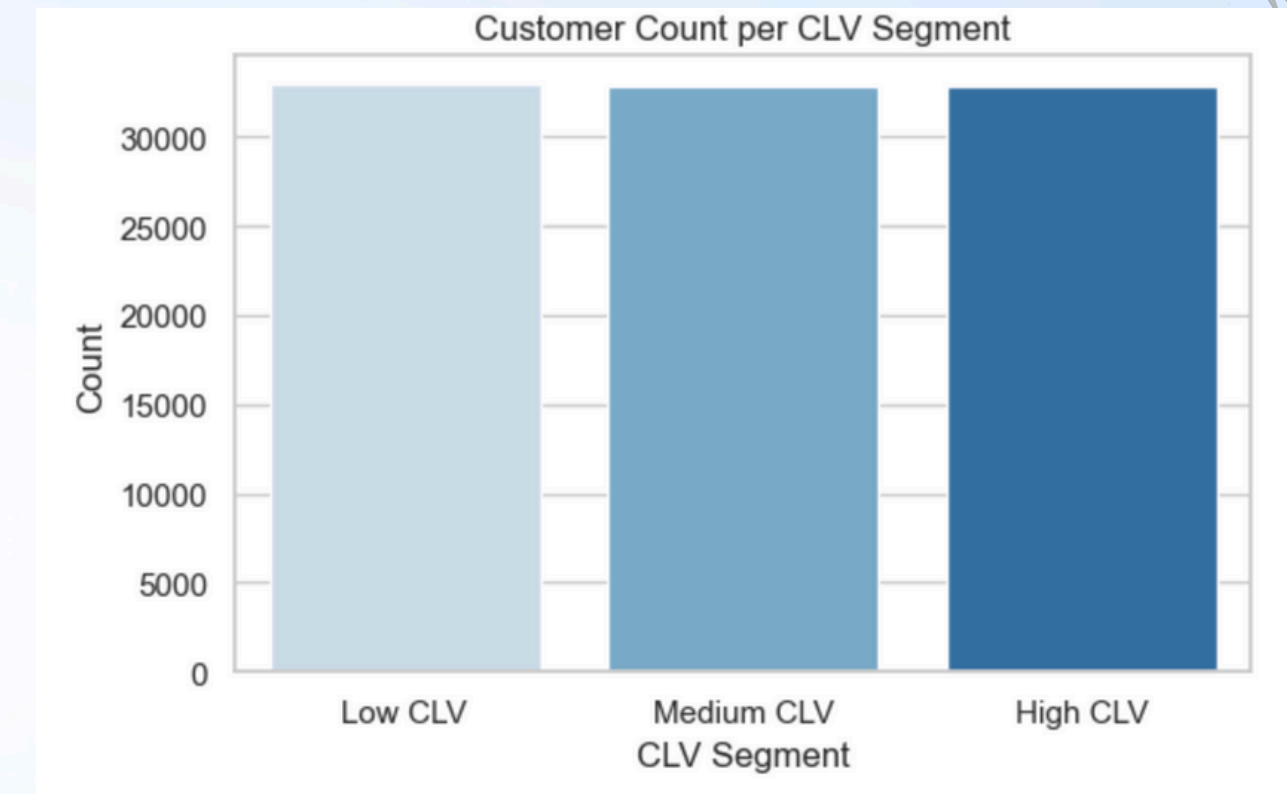
High CLV Customers (%): 33.33%

Each segment accounts for ~33% of customers, but High CLV customers contribute disproportionately to revenue.

Business Takeaway: Implement tailored marketing strategies by segment:
- **High CLV: Exclusive deals, personalized messaging**
- **Medium CLV: Cross-sell popular items**
- **Low CLV: Test onboarding or promo triggers**

What Drives Value & Loyalty?

Customer Review Score vs CLV

Delivery Time vs Customer Review Score

Top Categories by Spend (Avg CLV)

- Customers with review scores above 4.5 tend to have ~20–30% higher CLV on average.

- Positive experiences drive more repeat purchases and higher value per customer."

- Deliveries taking 8+ days consistently get lower reviews.

- Delivery speed is a critical lever for improving satisfaction and loyalty.
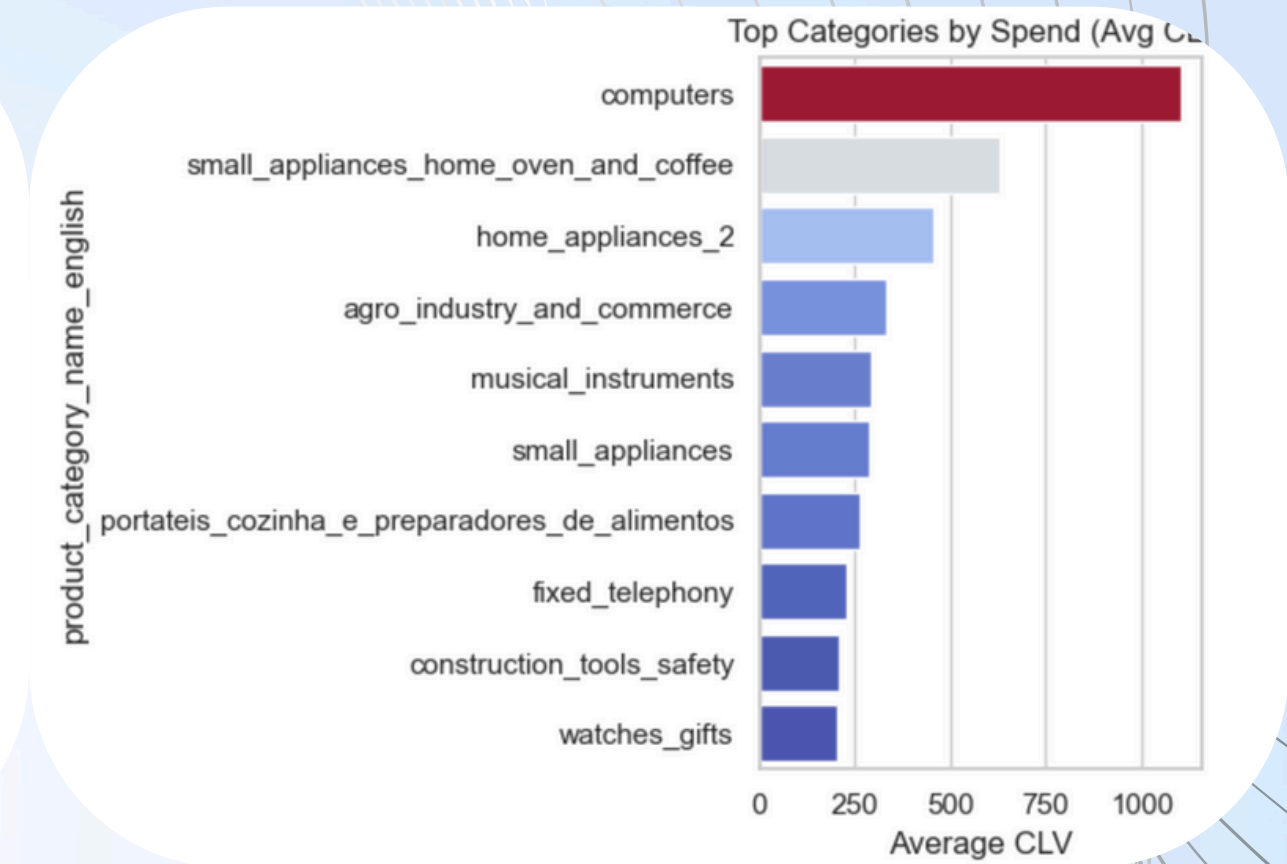
- Categories like computers and small home kitchen appliances have high order value but lower repeat rates and lower review scores.

- These categories may need better post-purchase support and loyalty nudges

Who Are Our Customers?

Customer_segment 1 : High CLV, frequent & satisfied → Loyal Champions

Customer_segment 0: Low CLV, one-time or inactive → At-Risk / Casual Buyers

customer_segment	Avg CLV	Avg Review Score	Avg Items/Order	Customer Count	% of Customers
0	167.92	4.08	1.13	91999	96.4
1	349.40	3.91	1.41	3420	3.6

"Segment 1 contributes more than 7% of total value despite being just 3.6% of our customer base. With an average CLV of ₹349.40, they are our most valuable — and most overlooked — customers."

Customer Segment Distribution

Segment 1

3.6%

96.4%

Segment 0
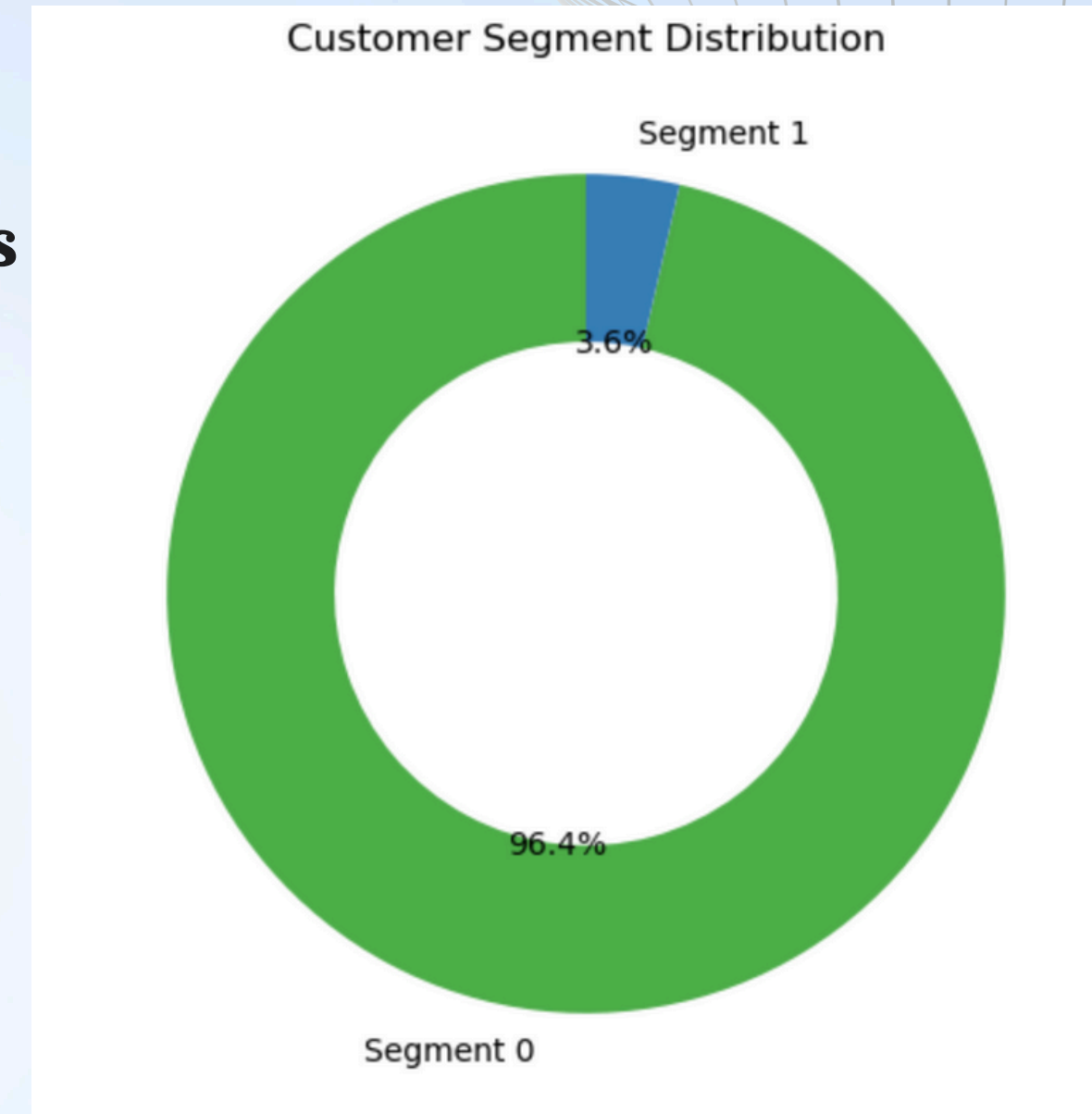
1. Loyal Champions

Profile: High CLV, frequent repeat buyers (≈ 3.6% of base, ≈ 7% revenue)
Insight: Very satisfied, short delivery time, loyal
Strategy:
- Loyalty rewards (points, cashback, birthday offers)
- Early access to sales/new arrivals
- Referral campaigns (incentivize word of mouth)

3. New Buyers / At-Risk

Profile: Low CLV, single-purchase, high churn risk
Insight: Many had long delivery times, neutral/low reviews
Strategy:
- Onboarding email journey (track order, how-to guides)
- Post-purchase surveys to capture issues
- Win-back offers ("We miss you" with coupon)

How Do We Predict Future Customer Value?

Objective:
- Predict 12-month Customer Lifetime Value (CLV) based on customer behavior, purchase patterns, and preferences.
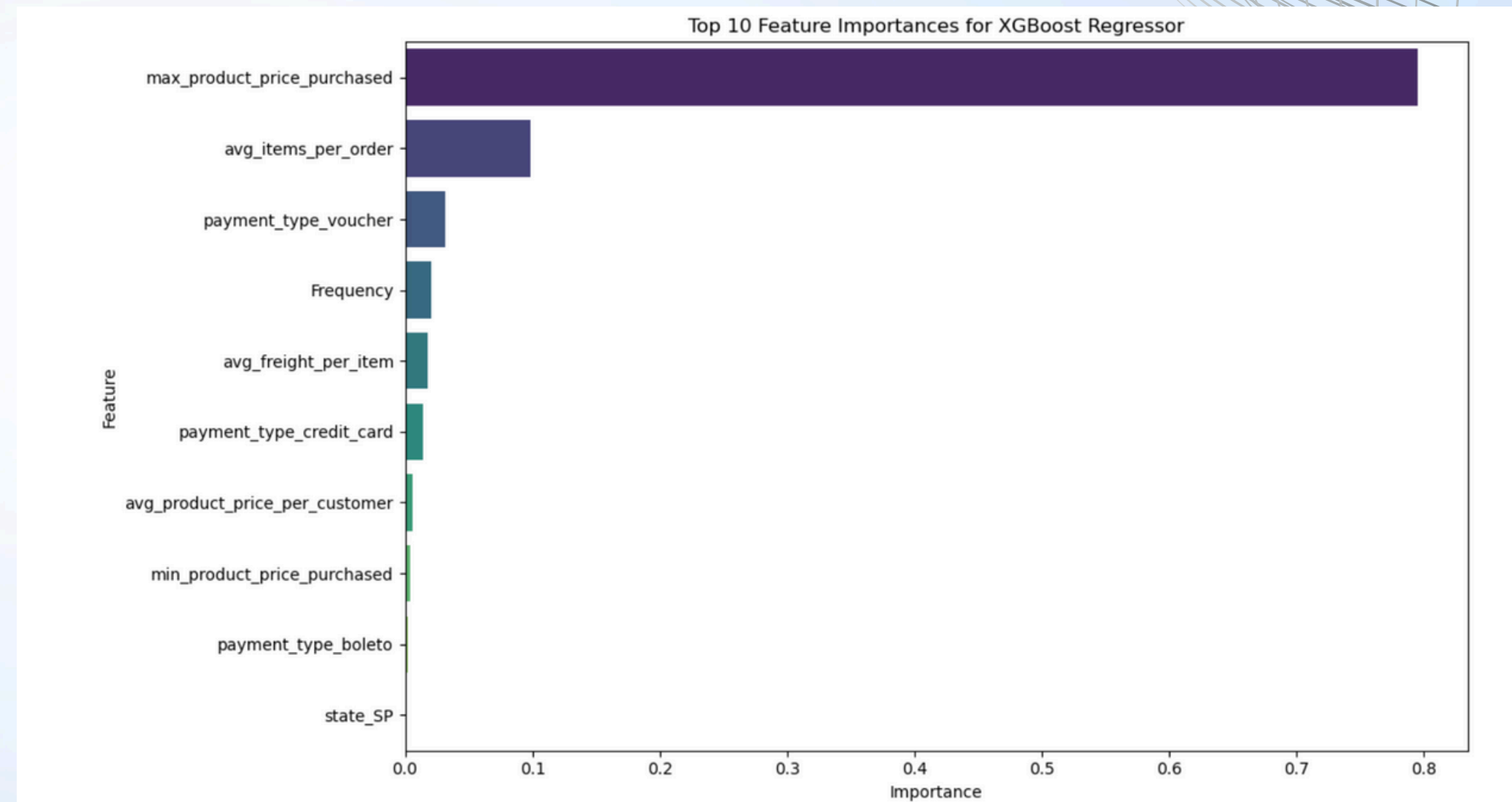
- **Model Used:**
- **Model: XGBoost Regressor**
- **Target Variable: CLV (12-month spend)**
- **Features Used (sample):**
- **Max product price purchased**
- **Avg items per order**
- **Payment type**
- **Purchase frequency**
- **Avg freight cost**
- **Review score, product diversity**

Top 10 Feature Importances for XGBoost Regressor

High CLV customers tend to purchase expensive products, buy in larger quantities, and prefer certain payment methods.

How Accurate is Our CLV Prediction?

```
--- XGBoost Regressor Performance ---
R-squared (Log-transformed): 0.9959
MAE (Log-transformed): 0.0197
RMSE (Log-transformed): 0.0524
R-squared (Original Scale): 0.9290
MAE (Original Scale): 7.52 BRL
RMSE (Original Scale): 73.32 BRL
```

Actual vs. Predicted Monetary Values (XGBoost Regressor)
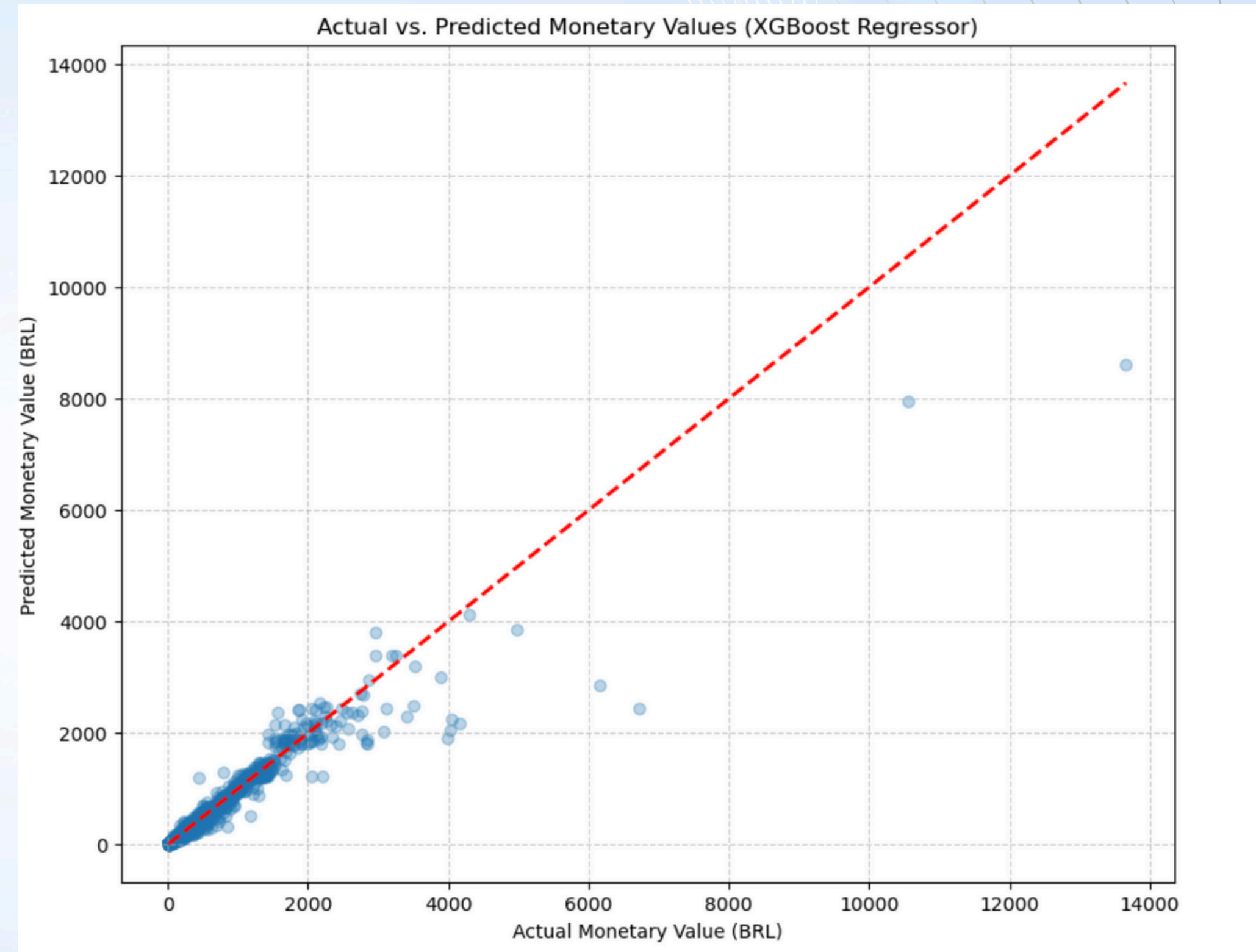
Insight:

The model achieves over 93% accuracy, making it reliable for driving personalized marketing and revenue strategies.

Marketing Strategy by Customer Segment

Segment 1: High CLV Champions
"Top 3.6% customers — drive over 2x average revenue"
Avg CLV: ₹349
Insight: Loyal repeat buyers, happy with experience
Strategy:
Loyalty rewards program
Early product access
Refer-a-friend incentives

Segment 0: Low CLV Mass Buyers
"96.4% of base — low engagement"
Avg CLV: ₹168
Insight: One-time or price-sensitive buyers
Strategy:
Welcome/onboarding campaigns
Feedback surveys to improve satisfaction
Targeted discount or re-engagement offers

"By focusing on Segment 1 for loyalty & referrals, and nurturing Segment 0, we unlock both retention and revenue growth."

THANK YOU

By-YARRA VIVEK SANYU
Aspiring Data Analyst