Optimizing E-commerce Logistics and Customer Satisfaction

By:

- Jaskirat singh
- Joseph Grande
- Vanshika Bhardwaj
- Suleman Ahmad

Submitted to:
Professor Fernanda Maciel

Background

Dataset is sourced from a Brazilian e-commerce

It contains 97,917 orders made between 2016 and

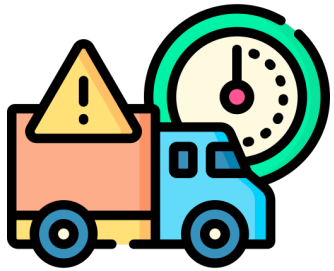
Understanding what drives consumer satisfaction, delivery performance, and cost variability is critical in today's competitive e-commerce landscape. This project explores how product features, pricing, and delivery operations influence customer experience, aiming to provide actionable insights for better logistics and service strategy.

Business Problem

- Untimely deliveries
- Inefficient Logistics Workflow
- Inventory Shortages
- Lack of Real-Time Delivery Tracking

Primary Goal

Improving customer satisfaction



Reducing Delivery Delays

Streamlining Shipping Operations

Standardizing Shipping Costs

Implications

Strategic Growth

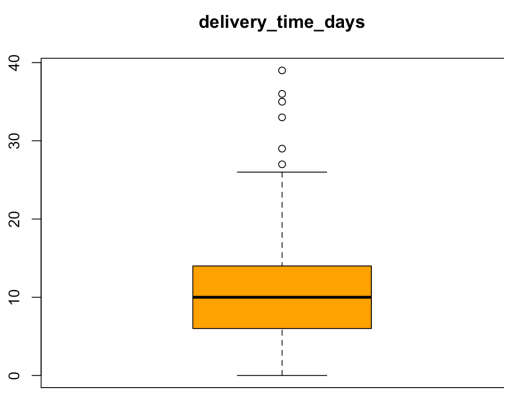
Operational Efficiency

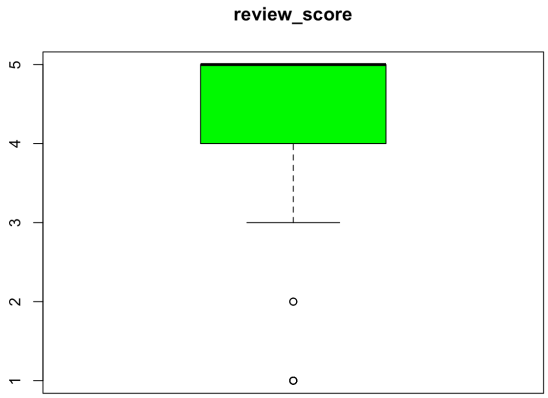
Customer Retention

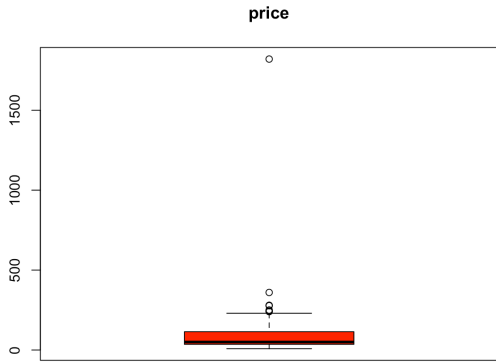
Scalability

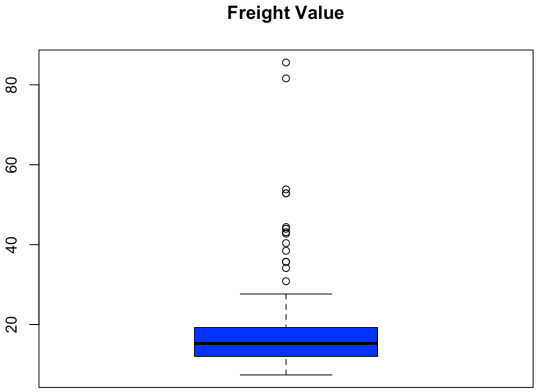
Inferences from the box plot

- Freight values are mostly clustered at lower costs
- There is a long tail of high prices, with significant outliers
- The distribution is heavily skewed towards higher review scores
- The delivery days variable shows a wide range of values, with several noticeable outliers representing significantly delayed deliveries.









Testing freight Determinants in E-commerce

Null hypothesis(H_0)

•There is **no significant relationship** between product freight value and variables such as price, product weight, and delivery time.

Alternative Hypothesis(H_1)

•There is a **significant relationship** between freight value and at least one of the predictors: price, product weight, or delivery time

```
Call:
lm(formula = freight_value ~ product_weight_g + product_length_cm +
    product_height_cm + product_width_cm + delivery_time_days +
    product_photos_qty + price, data = ol)

Residuals:
    Min      1Q  Median      3Q     Max
-51.663  -6.324  -2.257   2.163  74.811

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         1.326e+01  1.307e+00  10.149  < 2e-16 ***
product_weight_g   -6.858e-05  1.419e-04  -0.483   0.6290
product_length_cm   6.321e-02  2.766e-02   2.285   0.0225 *
product_height_cm  -1.898e-02  3.580e-02  -0.530   0.5962
product_width_cm   -5.555e-02  4.186e-02  -1.327   0.1849
delivery_time_days  1.755e-01  4.418e-02   3.971  7.7e-05 ***
product_photos_qty -1.649e-02  2.287e-01  -0.072   0.9426
price               3.623e-02  2.679e-03  13.526  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.02 on 940 degrees of freedom
Multiple R-squared:  0.1789,    Adjusted R-squared:  0.1728
F-statistic: 29.26 on 7 and 940 DF,  p-value: < 2.2e-16
```

Testing Delivery Delay Determinants in E-Commerce

Null Hypothesis (H_o)

• There is no significant relationship between delivery delay and variables such as product weight, freight value, or price.

Alternative Hypothesis (H_1)

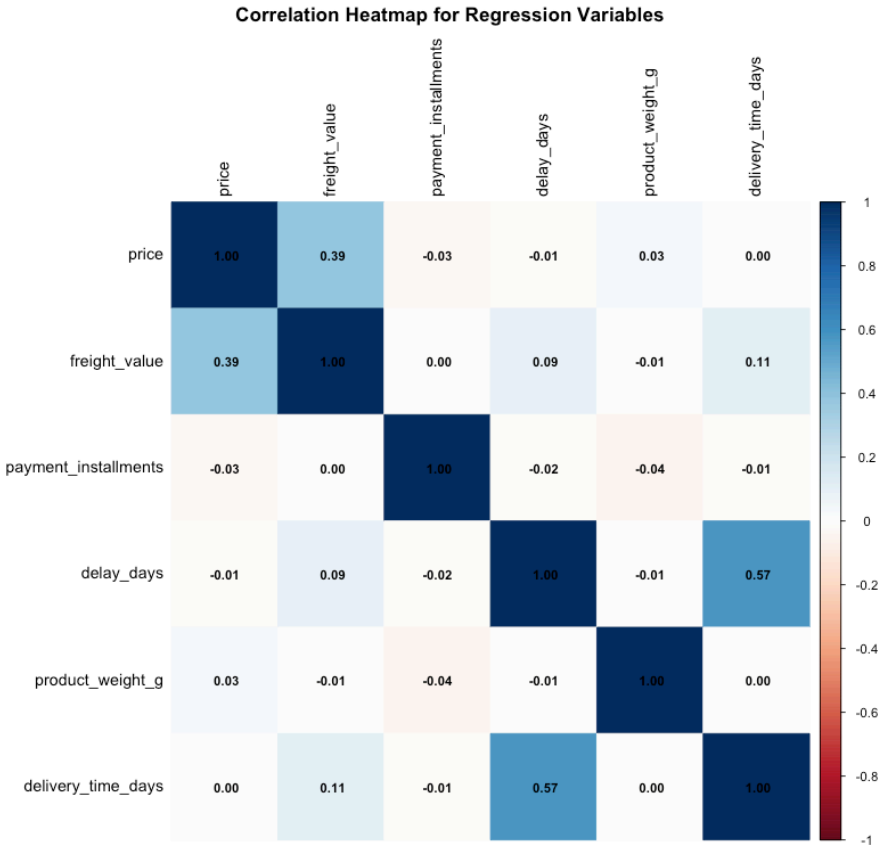
• There is a significant relationship between delivery delay and at least one of the predictors: product weight, freight value, or price

```
Call:
lm(formula = delay_days ~ product_weight_g + product_length_cm +
    product_height_cm + product_width_cm + freight_value + product_photos_qty +
    price, data = ol)

Residuals:
    Min      1Q  Median      3Q     Max
-38.935  -4.902  -0.353   4.770  59.383

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        -1.381e+01  1.060e+00 -13.035  < 2e-16 ***
product_weight_g   -1.109e-04  1.152e-04  -0.963 0.335755
product_length_cm  -6.566e-03  2.252e-02  -0.292 0.770658
product_height_cm   1.577e-02  2.906e-02   0.543 0.587444
product_width_cm    4.378e-02  3.399e-02   1.288 0.198089
freight_value       9.588e-02  2.626e-02   3.651 0.000275 ***
product_photos_qty -2.151e-01  1.857e-01  -1.159 0.246889
price              -4.987e-03  2.373e-03  -2.101 0.035866 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.755 on 940 degrees of freedom
Multiple R-squared:  0.01797,   Adjusted R-squared:  0.01066
F-statistic: 2.457 on 7 and 940 DF,  p-value: 0.01683
```



Regression analysis

Correlation Analysis: Key insights and Their Implications

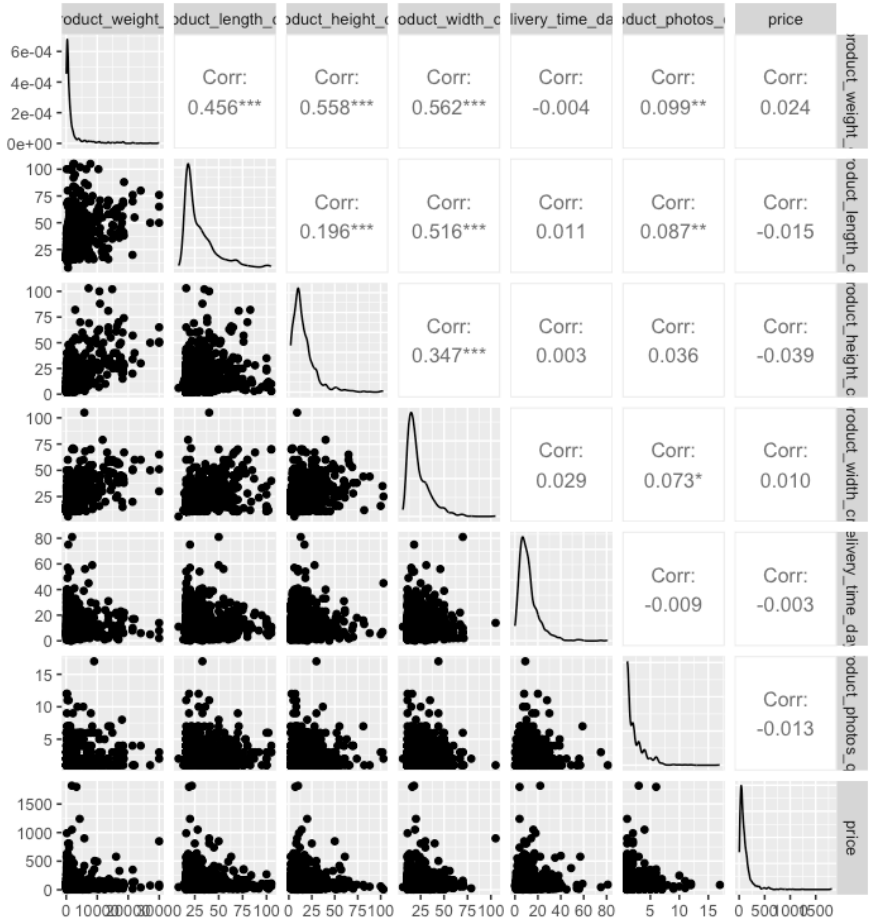
- Leverage delivery time data to identify and control shipping delays.
- Recognize that price moderately influences freight value; adjust accordingly
- • Freight cost is weakly associated with delay days, hinting at operational inefficiencies
- • Product weight and size have little impact on delays; focus elsewhere
- • Encourage simpler logistics flows rather than overemphasizing product features

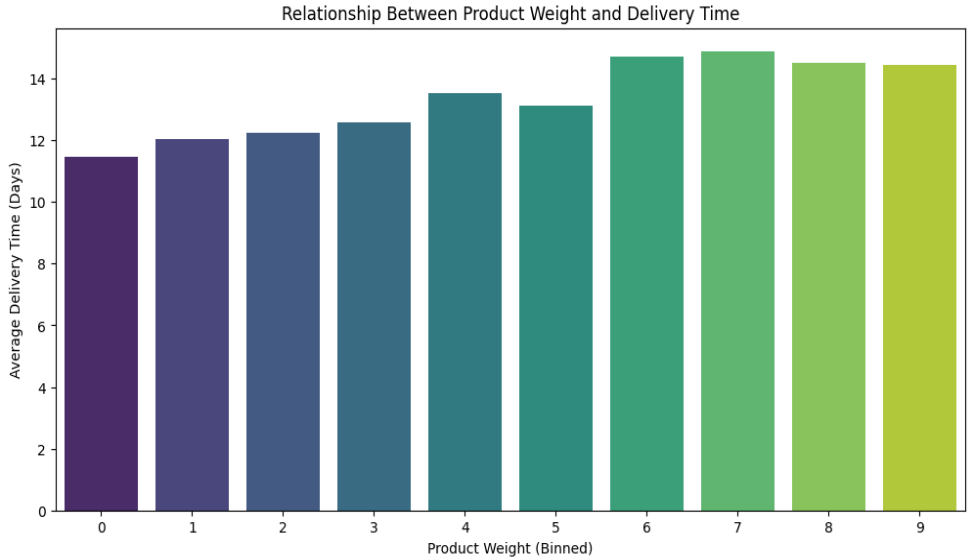
VIF Analysis: Multicollinearity Check

Freight Value Model (Reg):
- Product Weight – 1.96
- Product Length – 1.45
- Product Height – 1.47
- Product Width – 1.62
- Delivery Days – 1.00
- Product Photos – 1.01
- Price – 1.01

Delivery Delay Model (Reg3):
- Product Weight – 1.96
- Product Length – 1.46
- Product Height – 1.47
- Product Width – 1.62
- Freight Value – 1.20
- Product Photos – 1.01
- Price – 1.20

Interpretation:
All VIF values are below 2, well under the common threshold of 5. This indicates that **no multicollinearity** exists among the predictors, and each variable contributes independently to the model.

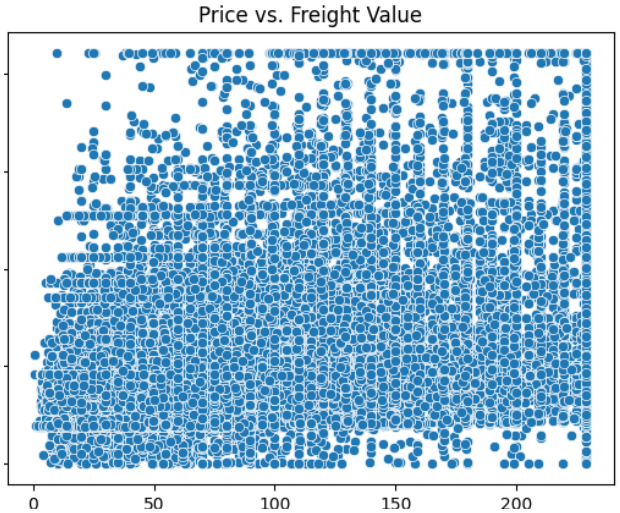




Product Weight Analysis

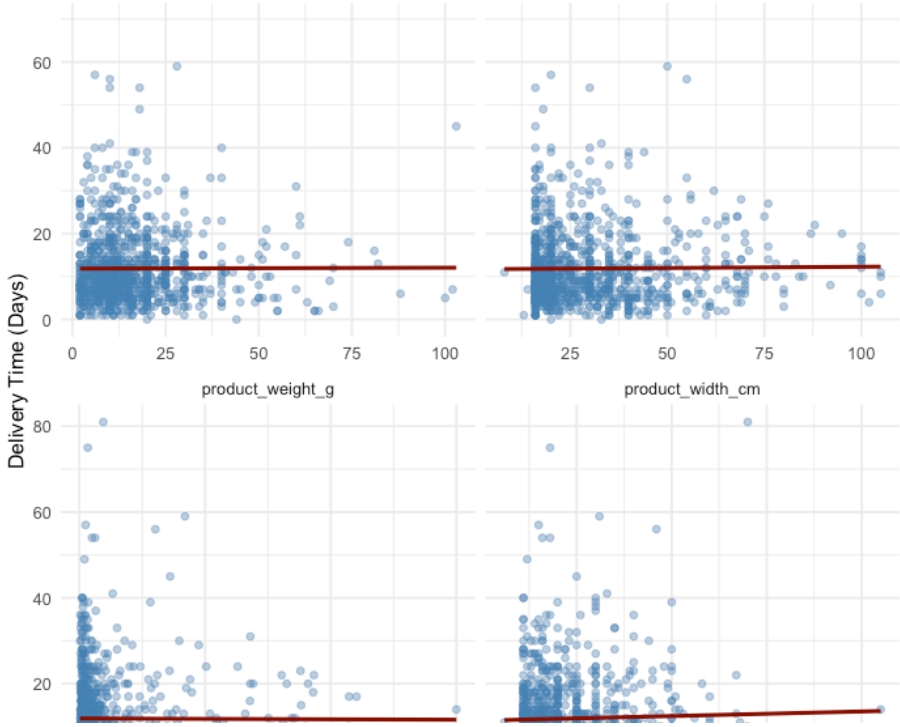
- Heavier products correlate with longer delivery times due to logistical constraints but not significantly contributing to delay
- There is a strong positive correlation between product weight and freight costs.
- Outliers exist in both very heavy and very light products.
- Product weight is not a critical factor influencing shipping costs and delivery efficiency.
- Categories like "Furniture" and "Industrial Goods" have consistently higher weights and longer delivery times.

Price Analysis



- Most products are affordable
- High-value items require special attention.
- Price and shipping costs correlate positively
- Premium items tend to be delivered faster

Product Dimensions and Delivery Time
The scatterplots show the relationship between delivery days and product features like weight, height, length, and width. The trend lines remain mostly flat, indicating **no strong linear association** between product dimensions and delivery duration. This suggests that **external logistics factor** rather than product size or weight may have a greater impact on delivery performance.



Summary of findings

Business Impact

Olist faces challenges related to **delivery delays** and **inconsistent freight costs**, which are not strongly explained by product dimensions or weight. This points to **external logistics inefficiencies** such as third-party carrier performance or fulfillment delays, affecting both cost and customer experience.

Key Takeaways & Strategic Outlook

Our analysis shows that product features have **limited impact** on delivery days and freight value. Significant predictors like **price** and **delivery time** suggest deeper operational dynamics. To address this, Olist could implement a **centralized delivery management system** to improve coordination, reduce delays, and enhance customer satisfaction.

We appreciate your time and attention.
This analysis was aimed at uncovering meaningful insights to help improve delivery performance, optimize logistics, and enhance the customer experience on Olist
We welcome any questions or feedback.
— Team Jaskirat,Vanshika, Joseph, and Suleman