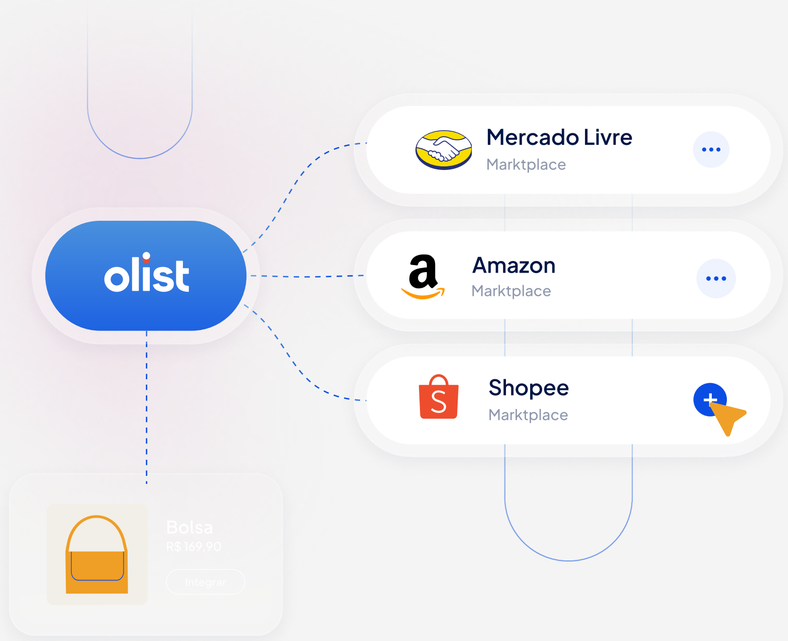
olist

IGNACIO MUÑOZ GOMEÑUKA

Decision Science Analytics



Objective

To understand and predict what drives customer satisfaction, integrating data from operations, payments, and logistics to uncover actionable insights that improve the end-to-end customer experience.

Sources for the analysis:

Analysis Approach

1. Analyze key operational factors — delivery time, freight cost, payment method, and product type — and their relationship with customer reviews.

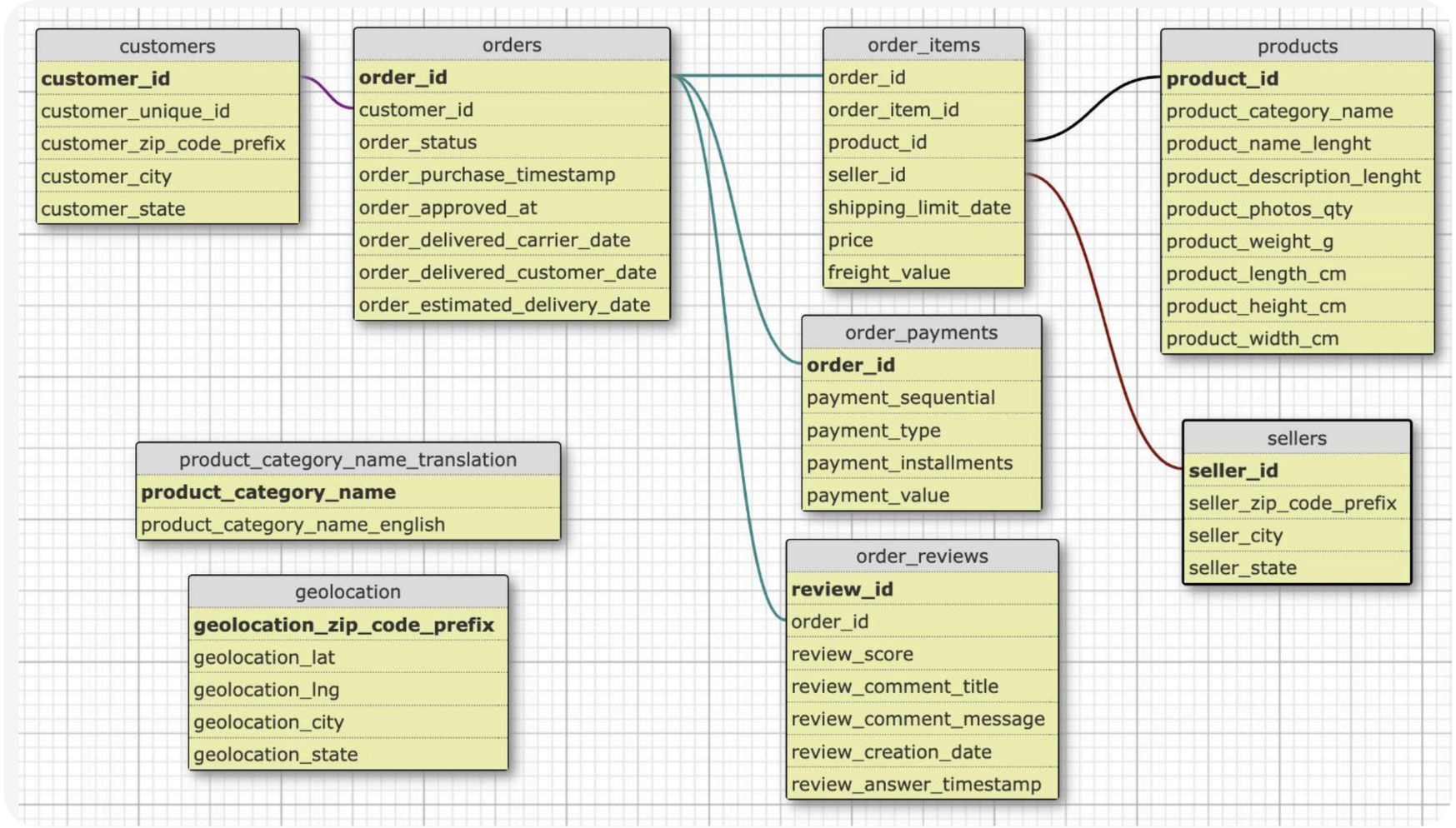
2. Identify geographic and logistic bottlenecks by mapping delivery delays and satisfaction across Brazilian states.

3. Build predictive models to:
   - Classify orders likely to receive negative reviews (≤ 2 stars).
   - Estimate the expected review score for any given order.

4. Segment customers by purchase behavior (frequency, spend, satisfaction) to support retention and loyalty strategies.

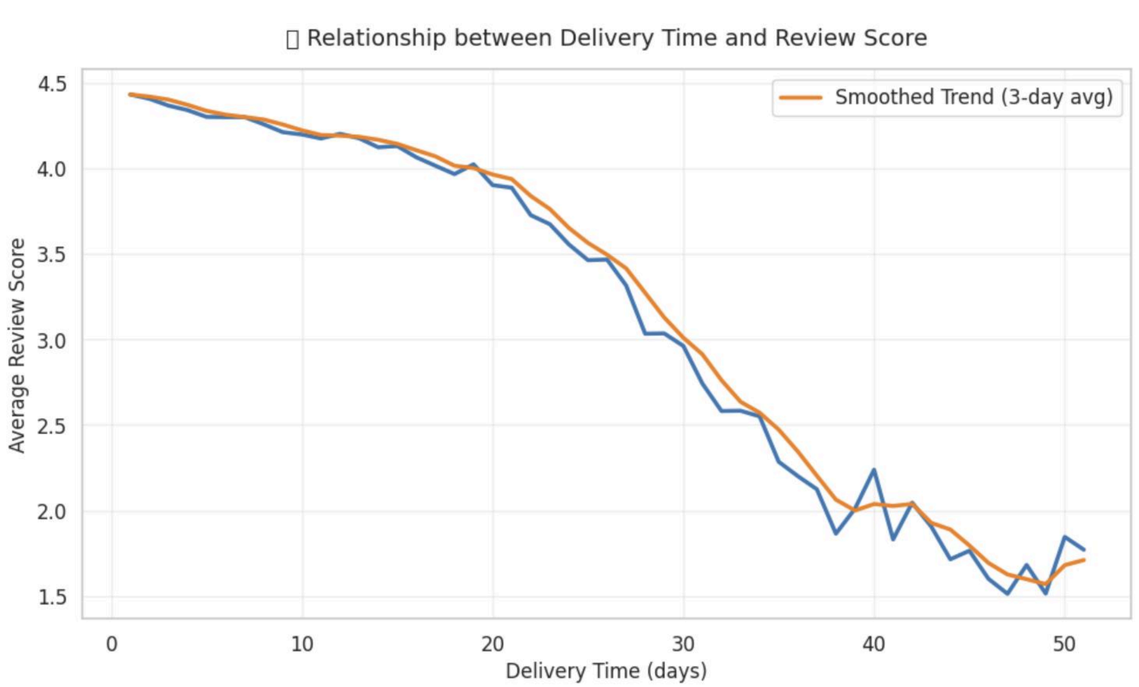
The dataset

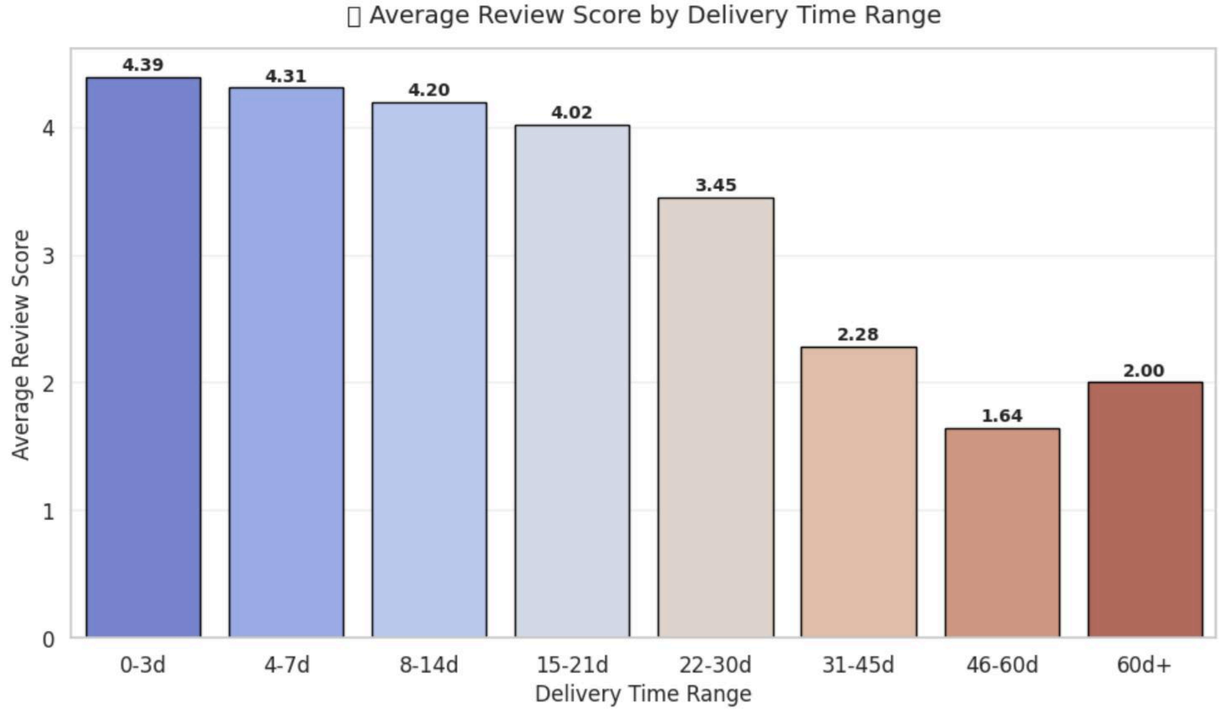
This is an overview of all the datasets that were merged to perform this analysis and to run the machine learning models.



Key Findings

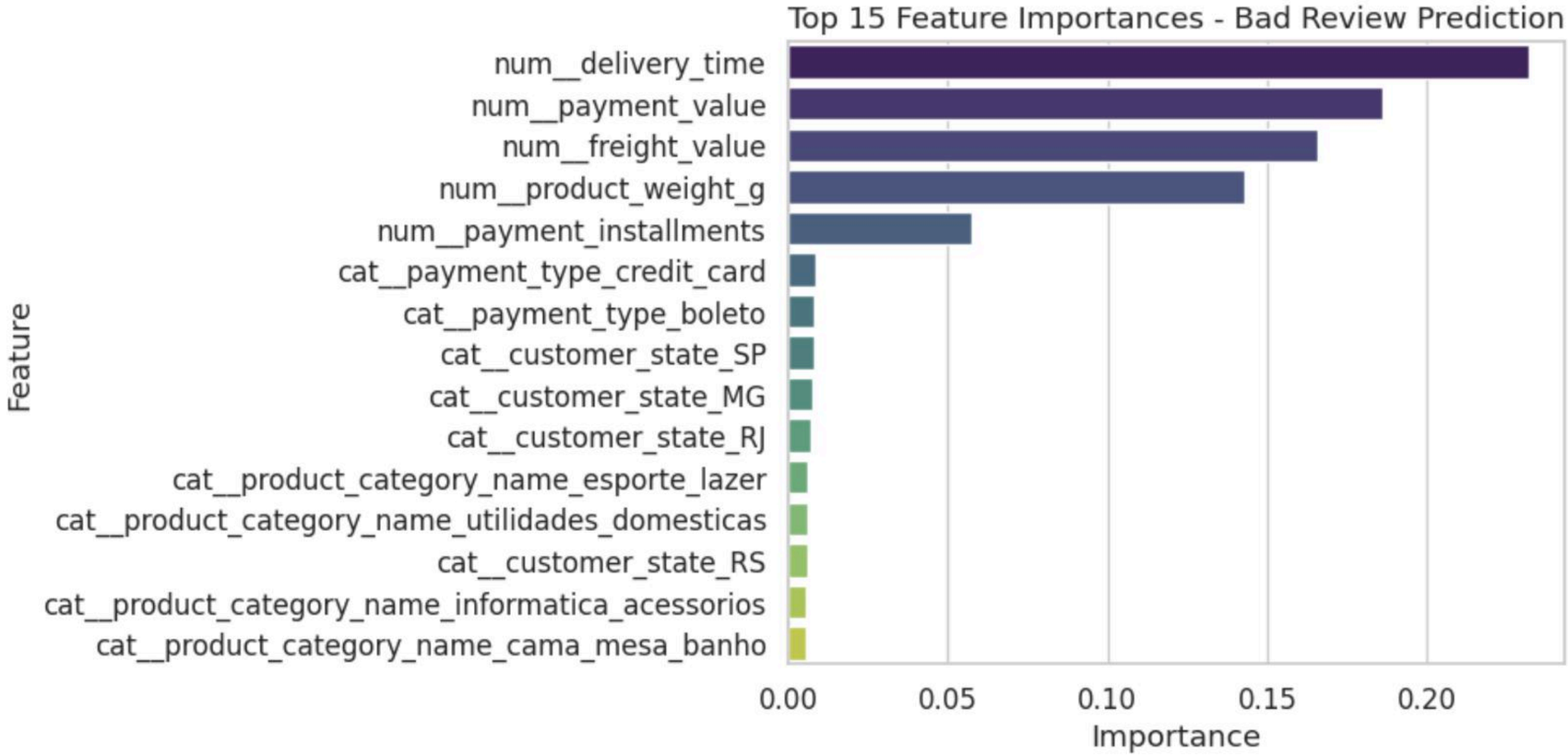
This charts shows the strong negative relationship between delivery time and customer satisfaction. As delivery times increase, average review scores steadily decline—from around 4.5 stars for fast deliveries to below 3 stars once orders take more than 30 days to arrive. The trend highlights **how longer delivery delays significantly reduce customer satisfaction**, making delivery performance one of the most influential factors in overall user experience.





What drives bad reviews?

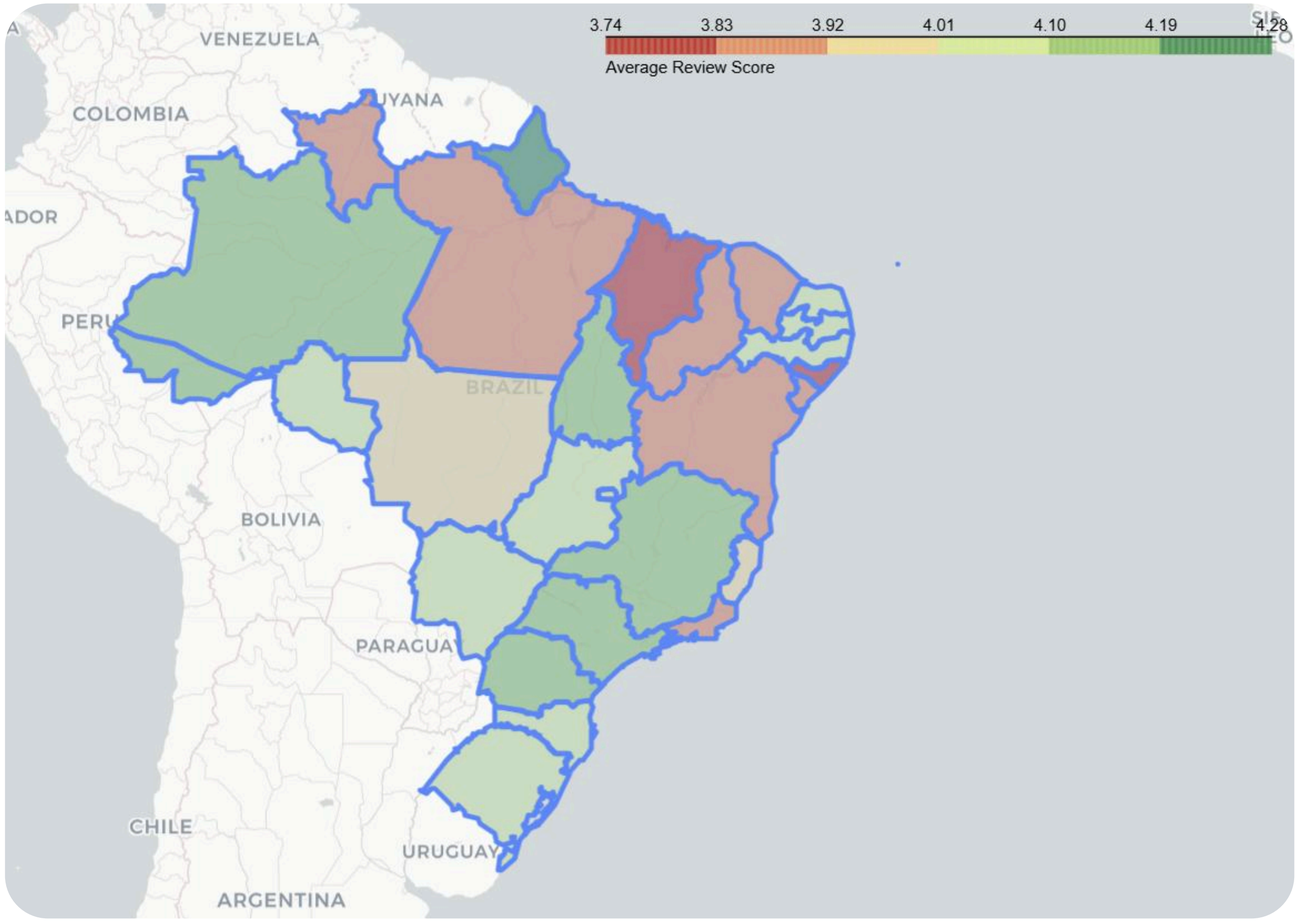
The model reveals that delivery time is the most influential factor behind negative reviews, confirming that longer or delayed deliveries have a strong impact on customer satisfaction. Payment value and freight cost follow as the next most important drivers, suggesting that customers who pay higher totals — especially when shipping costs are significant — tend to have higher expectations and lower tolerance for issues.



How logistics affect review scores and time deliveries

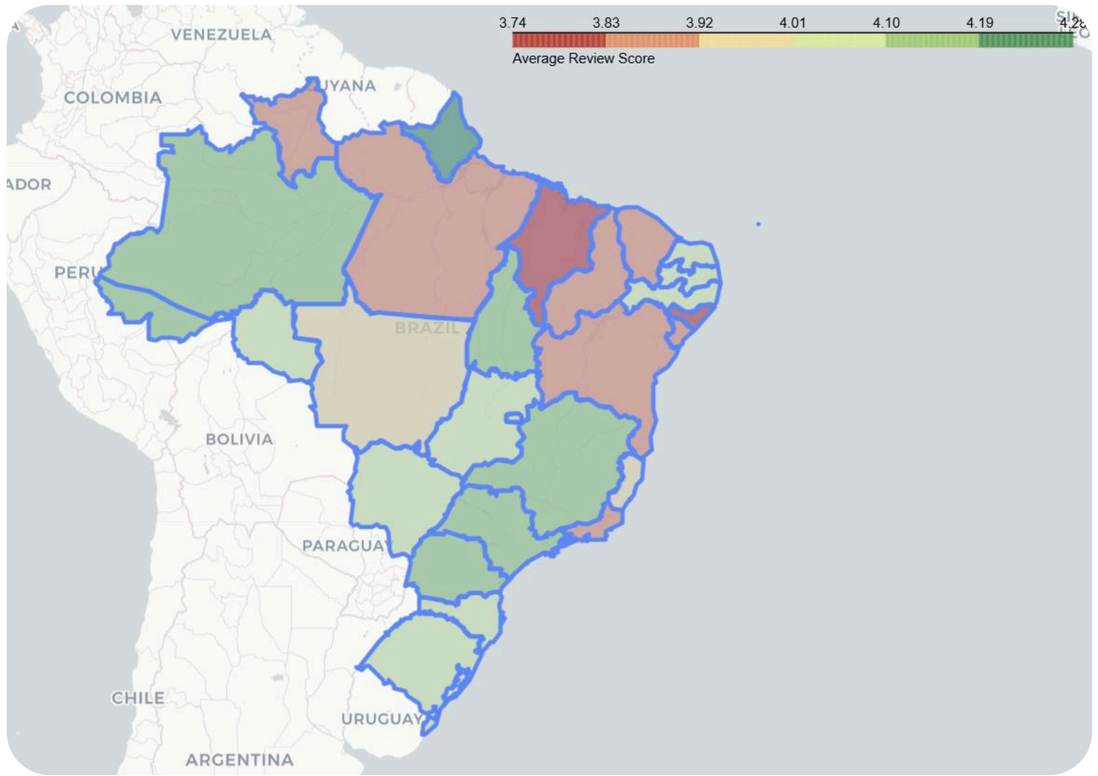
This map shows the geographic distribution of average customer review scores across Brazil, revealing clear regional differences in satisfaction. States in the South and Southeast, such as São Paulo, Paraná, and Rio Grande do Sul, have the highest ratings, while Northern and Northeastern states—particularly Maranhão—show significantly lower averages.

Maranhão's poor performance could be explained by longer delivery times and more complex logistics networks due to limited infrastructure and greater distances from major distribution centers. These factors likely lead to delayed shipments, higher freight costs, and unmet customer expectations, resulting in a concentration of negative reviews in that region.

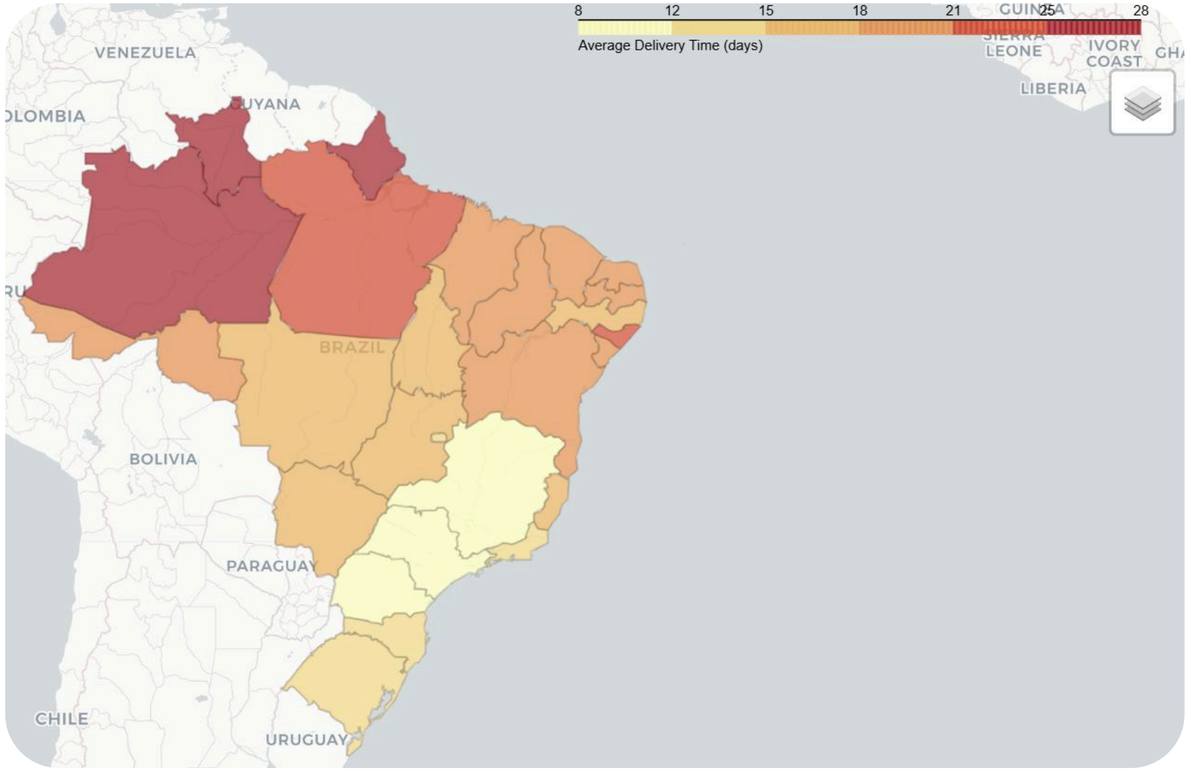


How logistics affect review scores and time deliveries

If we compare this two maps one with the average review score per state and the other with the average dealy time per state we can see how the top state with average delay hasnt the worst average review score.



Average Review Score per State



Average Delay Time (days)

ML Models to understand behaviour

Logistic Regression

This model estimates **the probability that an order receives a bad review** (≤2 stars), given:

- delta_z: How much slower/faster the order was compared to the median delivery time of that state (standardized).
- high_expect_region: If the order is from a fast region (e.g. São Paulo, Paraná, etc.).
- delta_z * high_expect_region: Checks if being late has a different impact in fast regions.
- freight_value, payment_value, payment_installments: Other controls related to cost and payment complexity.

Mixed Effects Model

This model explains review score (1–5) as a function of delivery time and other variables, but allows each state to have its own relationship (slope) between delivery time and satisfaction.

- Fixed effects: general influence across Brazil.
- Random effects: how each state differs from that national pattern.

Model Conclusions

Customers judge delivery speed relative to local norms, not in absolute days.
Delays hurt satisfaction more in high-expectation regions (urbanized, fast-shipping states like SP, PR, RJ).
In slower states like Amazonas, long deliveries are “normal”, so users don’t penalize them as harshly.

The models together reveal two key behavioral dimensions:

1. Relative expectations drive satisfaction.
   Customers compare against their local experience, not an absolute standard.
2. Regional tolerance matters.
   In northern, remote regions (Amazonas, Amapá, Roraima), customers are delay-tolerant.
   In southern and southeastern Brazil (São Paulo, Rio, Santa Catarina), customers are time-sensitive — a few extra days can drastically lower ratings.

State	Slope	Interpretation
MS, RJ, SC	~ -0.14 to -0.15	Strongly penalize delays → very time-sensitive customers
MA, PE, MG	~ -0.07 to -0.05	Moderate penalty for delays
AC, PR, AM, RR, AP	~ 0.00 to +0.39	No penalty or even positive effect — people tolerate or expect long deliveries