Enhancing Shipping Speed and Predictability

Insights, Recommendations, and Model Analysis
By Shaked Markovich
Dec 2024

Key Insights from Data Analysis

Sellers with longer average delivery times and higher approval-to-delivery days contribute significantly to delays.

Freight value and product price show a strong inverse correlation with delivery delays:

- Higher-priced items and higher freight value orders tend to have shorter delivery times.

Orders shipped to distant locations tend to take longer.
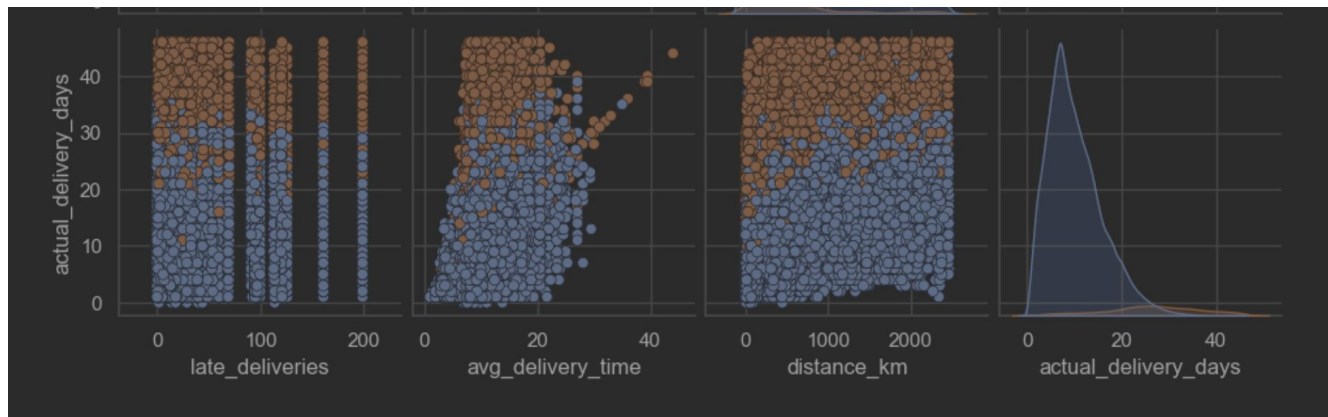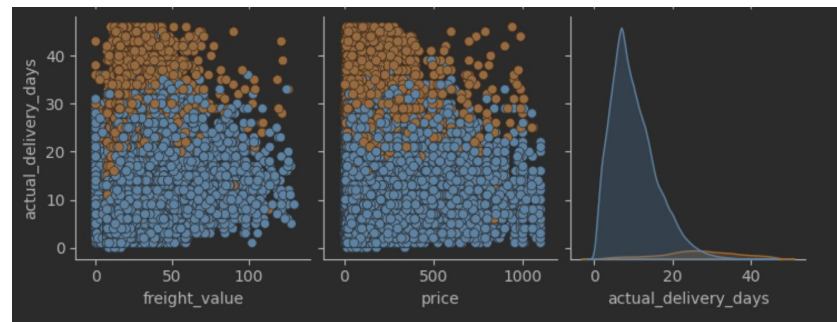Product category and seller performance significantly impact delivery time:

- Specific categories like *casa_conforto_2* show a higher rate of late deliveries (17%).
- Some sellers consistently perform better than others.

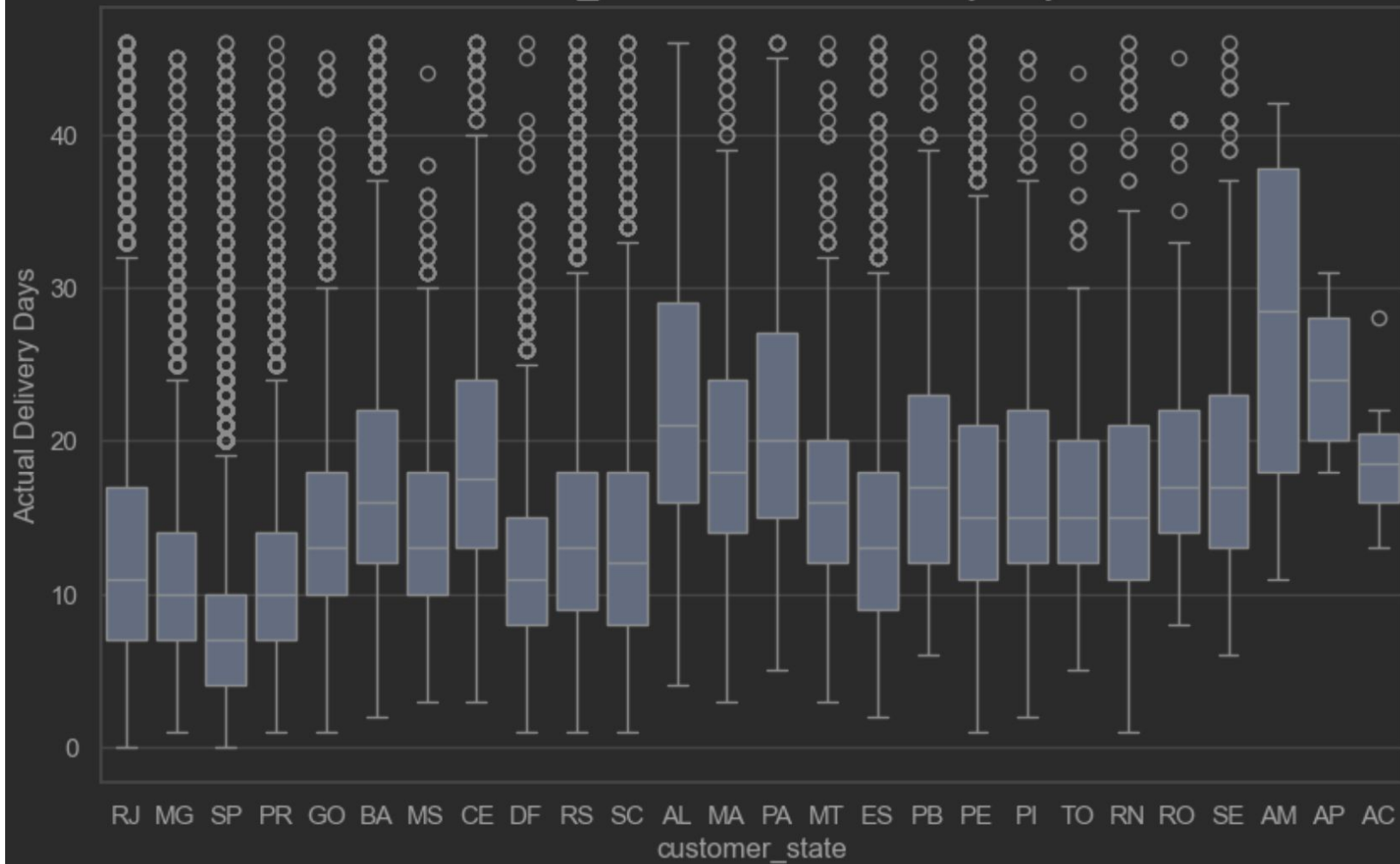A 20-day threshold appears critical:

- Non-late orders are typically delivered within this time, while late orders exceed it.
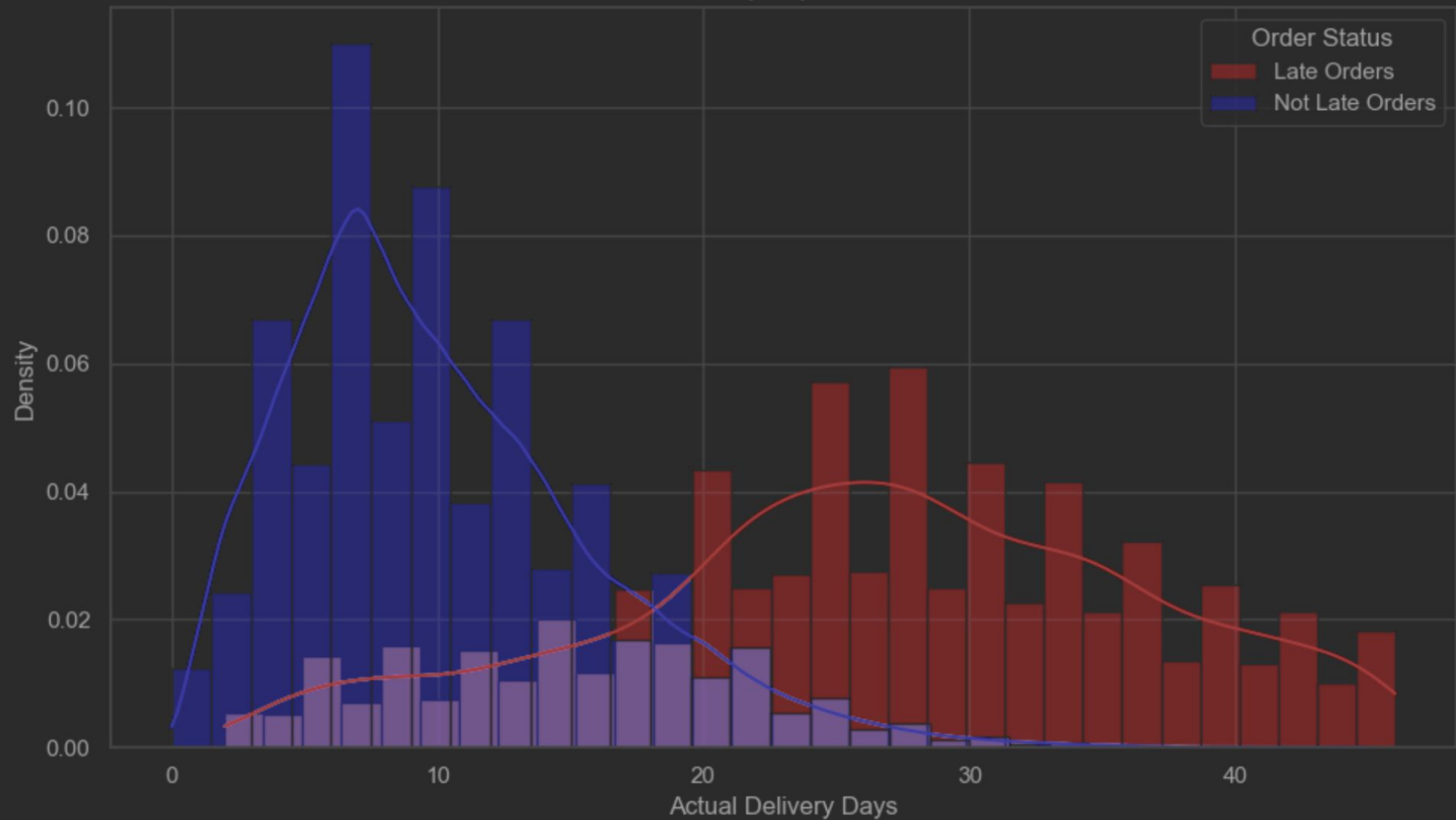
Correlation Heatmap
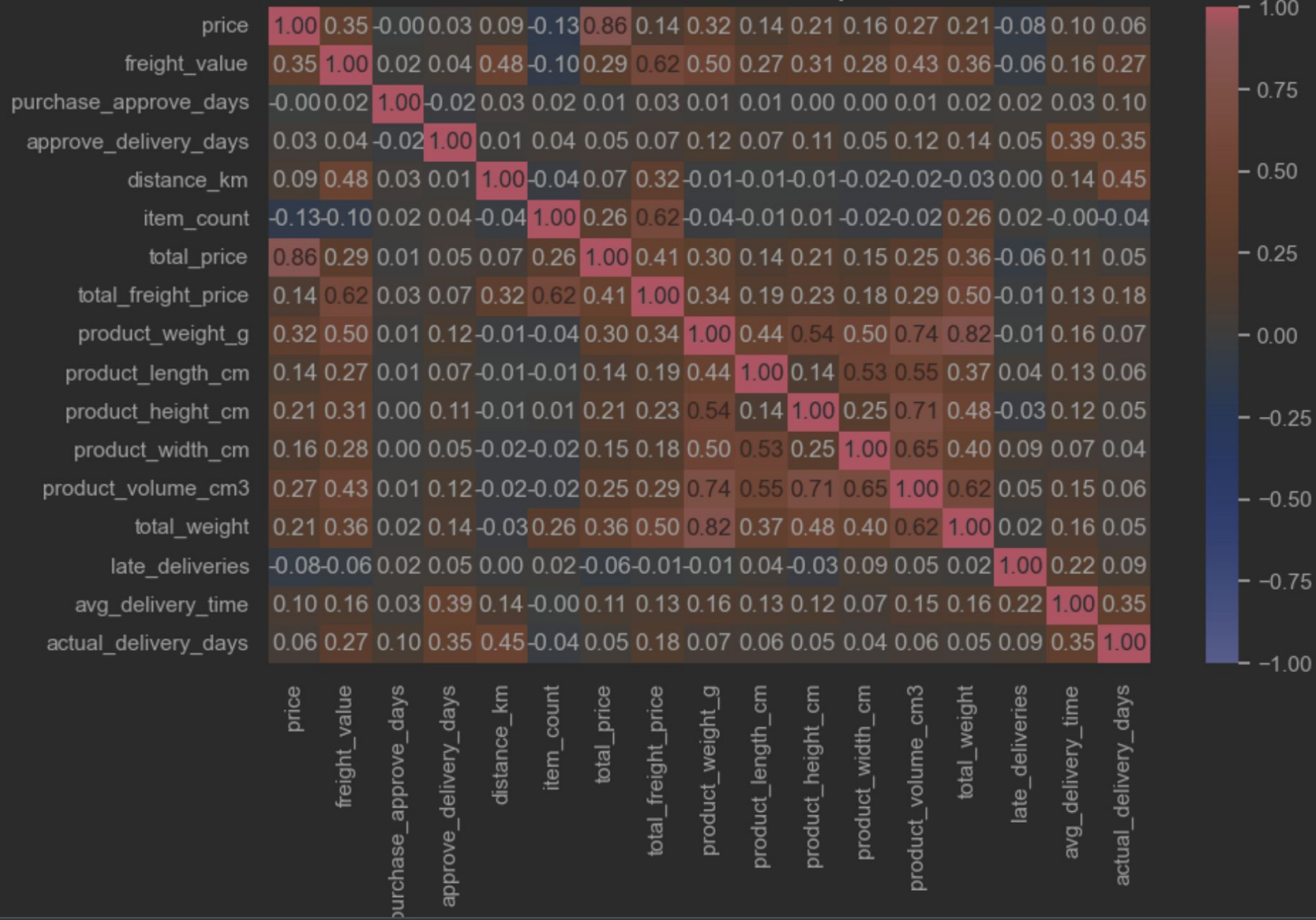
customer_state vs Actual Delivery Days

Distribution of Actual Delivery Days for Late vs Not Late Orders

Model Analysis

Performance:

- RMSE of 5.8 days vs. baseline of 14.39 days (60% improvement).

Feature Importance:

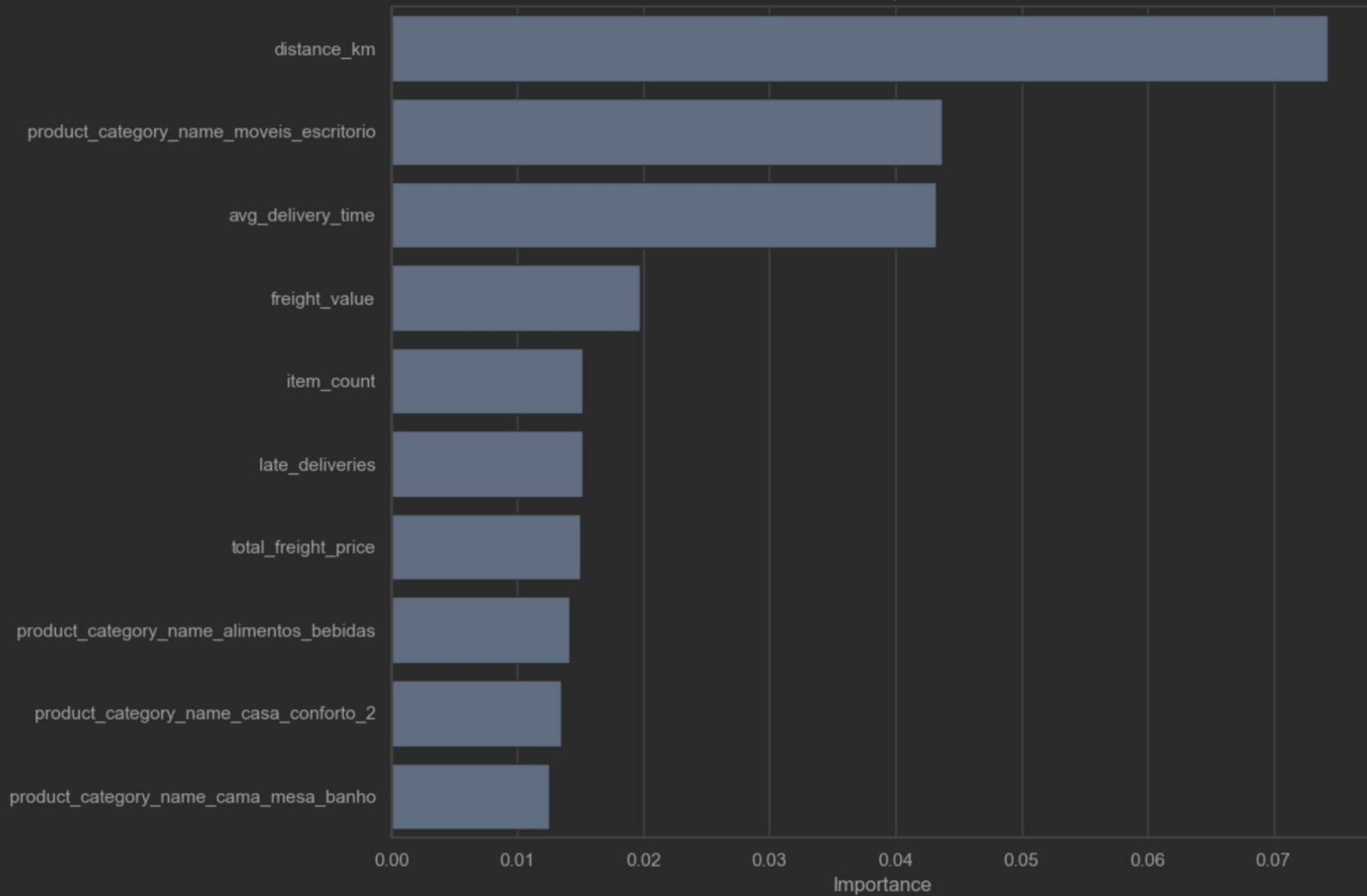- Top predictors: distance_km, product_category_name_moveis_escritorio, avg_delivery_time, freight_value

Strengths:

- Captures complex relationships effectively.

Limitations:

- Sensitivity to data quality and feature engineering.

Feature Importance

Questions for the Product Manager

- What is shipping_limit_date? can we use this to make more features like time untill this date?

- Do we have any benchmarks for seller performance in terms of delivery times? Are there actions we're taking to help sellers with longer delivery times improve?

- How do we handle logistics for distant locations? Are there specific factors (like carrier limitations or warehouse locations) that contribute to longer delivery times in those areas?

- Can we set some targets for the "approval-to-delivery" duration? Is there an ideal range for this metric?

- Are there any strategies to optimize freight costs? Do we have any goals around the relationship between product price, freight value, and delivery performance?

- Are there any pricing strategies in place to encourage faster delivery times?

- Are we prioritizing certain sellers based on their delivery times or service levels? Do we have KPIs in place to track and support sellers that need help?

- Can we track the reasons for orders that are consistently late (over 20 days)?

Recommendations

- We could focus on identifying sellers who consistently have higher average delivery times.

- Consider optimizing logistics for sellers who frequently ship to distant locations. This could involve partnering with more efficient shipping carriers or offering discounts for expedited deliveries to high-priority regions.

- It might be useful to identify specific types of products (especially large or fragile items) that consistently lead to delays.

- Investigating the causes behind late orders (e.g., delays at the warehouse, customs issues, or carrier problems) could provide valuable insights for improving delivery times.

What to Expect in Production

- Model Deployment:
  After training, the XGBoost model will be deployed to make predictions on new data.

- Data Handling:
  The data pipeline will handle incoming data from multiple sources (like sellers or logistics systems). We'll need to create automated ETL.

- Model Monitoring:
  We'll track its accuracy (RMSE in this case), check for any changes in the input data (data drift), and make sure it stays reliable. If it starts underperforming, we may need to retrain it or tweak it.

- Feedback Loop:
  We'll compare the model's predictions to actual outcomes (e.g., was the delivery late or on time?). This feedback will help us improve the model over time.

- Performance Optimization and Scalability:
  Depending on the setup (cloud, on-premise), we might need to optimize the system to make predictions quickly, especially if we're doing real-time predictions for a lot of orders.

Plan for Recommending Alternatives to Shorten Shipping Times

Step 1: Add to EDA and Model Evaluation

- Group by Sellers and Regions:
  - Identify sellers and regions with consistently faster delivery times.
  - Use avg_delivery_time and is_late for sellers, cities, and states.
- Product Analysis:
  - Identify categories with shorter delivery times using product_category_name.
  - Suggest faster products in the same category and region.
- Feature Importance:
  - Analyze key drivers like seller_state, distance_km, avg_delivery_time, and product_category_name.

Step 2: Recommendation Logic for Long Delivery Time Predictions:

- Suggest alternative products from sellers with lower avg_delivery_time.
- Prioritize nearby sellers (seller_city, seller_state) and faster regions.
- Balance cost (price, freight_value) and delivery performance.
- Flag orders predicted to exceed 20 days for actual_delivery_days and suggest alternatives.

Step 3: Evaluate Recommendations

- Measure success by comparing delivery times before and after recommendations.
- Track customer acceptance and refine the logic over time.

Example Use Case

- Current Order:
  - Product: "casa_conforto_2"
  - Seller: Slow Seller A
  - Predicted Delivery Time: 25 days
- Recommendation:
  - Suggest alternatives in the "casa_conforto_2" category with faster sellers (avg_delivery_time < 15 days).
  - Highlight products from nearby locations for quicker logistics.

Conclusion

Insights:

- Clear correlation between features like seller performance, price, freight value, and delivery times.
- Data reveals actionable opportunities for improving shipping speed and predictability.

Next Steps:

- Refine logistics operations using insights from the model.
- Implement a recommendation system for alternative products to optimize shipping timelines.

Q&A

Thank you for your attention!

Let's discuss any questions or feedback.