Product Delivery Efficiency and Delay Analysis: A Case Study on "Deliver on Time"

Prepared by: Adaobi Cynthia Okonkwo

Tools Used: Excel, Data Visualization Techniques

Category: E-Commerce Delivery Data Analytics

1. Outline

This report follows a structured roadmap to present the analysis clearly and logically:

- Introduction
- Story of Data
- Data Splitting and Preprocessing
- Pre-Analysis
- In-Analysis
- Post-Analysis and Insights
- Data Visualizations & Charts
- Recommendations and Observations
- Conclusion
- References & Appendices

2. Introduction

This project analyzes delivery efficiency for the company Deliver, focusing on understanding the factors that influence whether packages arrive on time. The goal is to identify patterns, uncover causes of delays, and provide insights that can improve logistics performance.

The analysis specifically aims to evaluate delivery outcomes across shipment modes, warehouse blocks, product categories, and customer-related factors such as care calls, ratings, product cost, and weight. The core problem being addressed is the persistent delay in product deliveries, which affects customer satisfaction and operational effectiveness.

A structured shipment dataset was used, and Microsoft Excel served as the main analytical tool. Key Excel methods applied include data cleaning, pivot tables, charts, sorting, filtering, and descriptive analysis.

3. Story of Data

- **Data Source:** The data originates from a structured e-commerce shipment dataset, inspired by real-world delivery operations on Kaggle.com.
- **Data Collection:** The data was collected through automated digital logging systems commonly used in e-commerce platforms. Each shipment record captures transaction details, customer interactions, and delivery outcomes.
- **Data Structure:** Rows: Each row represents a single shipment or transaction.

Columns**:** Variables include shipment ID, warehouse block, mode of shipment, customer care calls, customer rating, cost of product, prior purchases, product importance, high-value shipment indicator, gender, discount offered (%), weight (gms), weight category, and whether the product reached on time.

- **Important Features and Their Significance:**
  - Mode of Shipment & Warehouse Block – Helps evaluate operational efficiency and bottlenecks.
  - Customer Rating & Care Calls – Provides insights into service quality and potential delivery issues.
  - Cost, Weight, and Product Importance – Indicates how product characteristics affect delivery speed.
  - Discount Offered – Identifies promotional influences on delivery.
  - Reached on Time – Target variable measuring delivery performance.

Additional engineered features, such as High Value Shipment and Weight category, were added to enrich the pattern

- **Limitations**: No limitations were present

4. Data Splitting and Preprocessing

- **Data Cleaning:** The dataset was reviewed for quality issues, including duplicates, inconsistencies, and incorrect entries. Duplicates were not present, invalid values were not present, and categorical fields were checked to improve reliability.
- **Handling Missing Values:** No null values are present.
- **Data Transformations:**
  - Weight Categories: A new variable, "Weight Category", was created by grouping shipments into Light, Medium, and Heavy based on their weights.
  - High-Value Shipment Indicator: Products exceeding a specified cost threshold, and if the product importance is high, were labeled as "High Value Shipments."
- **Data Splitting:** Variables were separated into:

- Independent Variables: Warehouse block, mode of shipment, customer care calls, customer rating, cost, prior purchases, product importance, high-value shipment, gender, discount offered, weight, and weight category.
- Dependent Variable: Reached on Time, whether the shipment was delivered on schedule.

This separation ensures the analysis clearly identifies which factors influence timely delivery.

- **Industry Context:** The dataset belongs to the e-commerce and logistics industry, where on-time delivery is a critical performance indicator. Timely shipments influence customer satisfaction, retention, and operational efficiency, making this analysis highly relevant to companies managing online order fulfillment.
- **Stakeholders:** Key stakeholders who will benefit from this analysis include logistics and operations teams, customer service teams, Senior Management, and the Marketing & Sales department
- **Value to the Industry:** The analysis provides evidence-based insights that can:
  - Help improve routing, warehouse allocation, and shipment prioritization.
  - Reduce delays to enhance customer trust and loyalty.
  - Enables management to allocate resources effectively and make evidence-based strategic choices.
  - Companies leveraging delivery analytics can outperform competitors by ensuring reliable, timely service.

5. Pre-Analysis

The pre-analysis phase provides an initial understanding of the dataset and uncovers early trends before deeper statistical evaluation.

- **Identify Key Trends:** Early exploration of the shipment data revealed several notable trends:
  - Delivery Performance by Shipment Mode: Certain shipment modes tend to have higher on-time delivery rates compared to other modes.
  - Warehouse Block Influence: Specific warehouse blocks show consistently faster dispatch times, suggesting operational efficiency differences across locations.

- Customer Interaction Patterns: Shipments associated with multiple customer care calls often correlate with delayed deliveries.
- Product Weight and Importance: Heavier or high-priority products show a slight tendency toward timely delivery, likely due to handling protocols.

- **Potential Correlations:** Initial observations suggest possible relationships between variables:
  - Customer Rating vs. On-Time Delivery: Higher ratings generally coincide with shipments delivered on time, indicating a link between service quality and delivery performance.
  - Discount Offered vs. Delivery Timeliness: Shipments with large discounts occasionally experience delays, possibly due to higher order volume during promotions.
  - Prior Purchases vs. Delivery Efficiency: Customers with multiple past purchases appear to receive more reliable delivery, hinting at preferential handling for loyal clients.
- **Initial Insights:**
  - Early analysis indicates that shipment mode, warehouse block, product weight, and customer interactions are likely influential factors in delivery outcomes.
  - Patterns of delay emerge more frequently for certain product categories and lower-value shipments, raising questions for deeper investigation.

These insights provide a roadmap for the detailed analysis, allowing focused exploration of the most significant variables affecting delivery efficiency.

6. In-Analysis

This stage represents the core of the analytical process, where deeper exploration of the dataset reveals verified and emerging patterns. Using Excel-based techniques such as PivotTables, correlation checks, conditional formatting, and lookup functions, several insights were uncovered.

- **Unconfirmed Insights (Hypotheses Under Evaluation):** During initial exploration, several patterns appeared promising but require further validation with statistical tools:
  - Shipment Mode Correlation: Preliminary data suggest that Flight shipments have a slightly higher on-time delivery rate compared to Ship and Road modes, despite smaller volumes.

Flight shipments may benefit from faster transit times and fewer handling points, improving reliability.

- Product Weight and Delivery Timeliness: Heavier shipments appear to have a marginally higher on-time delivery rate than lighter ones. High-value or heavier products may receive priority handling, contributing to faster deliveries.
- Customer Interactions: Shipments associated with more customer care calls show a higher incidence of delays. Customer-initiated inquiries may indicate unresolved issues that affect timely delivery.
- Discounts and Promotions: Shipments during periods of higher discounts occasionally experience delays. Increased order volumes during promotions may strain logistics operations.

- **Recommendations (Preliminary and Based on Emerging Patterns):** Although the analysis is ongoing, early patterns allow for initial recommendations:
  - Optimize Shipment Mode Allocation: Prioritize Flight for high-priority or time-sensitive deliveries to improve overall on-time performance.
  - Enhance Road and Ship Operations: Investigate bottlenecks in Ship and Road deliveries and optimize routes, warehouse handling, or staffing.
  - Monitor Customer Queries: Implement proactive measures for shipments generating multiple customer calls, such as alerts or faster issue resolution.
  - Plan for Promotional Periods: Anticipate higher order volumes during discount periods and adjust resources to prevent delays.

These recommendations will be refined once hypotheses are fully validated.

- **Analysis Techniques Used in Excel:** Excel served as the primary tool for exploring and validating relationships in the dataset. Key functionalities included:
  - Pivot Tables: Used extensively to summarize shipments by mode, warehouse, product type, and delivery outcome for pattern detection.
  - Pivot Charts: Visual forms of pivot outputs helped compare variable distribution and delivery performance across categories.

- Sorting and Filtering: Enabled quick identification of high-delay shipment segments and operational trends.
- Conditional Formatting: Highlighted delayed shipments and patterns in delivery performance.
- Charts and Graphs: Column, bar, and stacked charts visualized delivery efficiency and delays across categories.

7. Post-Analysis and Insights

This stage consolidates the findings from the full analytical process, confirming or rejecting the initial hypotheses and interpreting the patterns observed within the dataset, and providing actionable insights for improving delivery efficiency

- **Key Findings:** Based on the completed analysis, several significant insights emerged:
  - Although Ship mode handles the largest number of on-time deliveries, Flight has the highest on-time delivery rate.
  - While most deliveries are on time, delays among Very Satisfied, Dissatisfied, and Neutral customers highlight service gaps that require targeted improvements to maintain satisfaction.
  - Warehouse F has the highest number of on-time deliveries but also the highest number of delays, likely because it handles the largest volume overall.
  - Heavy shipments dominate delays, making this group the biggest contributor to delays.
  - Most Delayed Deliveries Come from Customers Who Called 3-4 Times.
  - Although most shipments are Standard products, High-value shipments receive slightly better prioritization, resulting in fewer delays proportionally.
  - All delayed orders occur in the 1–10% discount range, and the same discount range has the highest on-time delivery rate.
  - Customers with 3 prior purchases had more on-time deliveries than delayed deliveries.
- **Comparison with Initial Findings**

Comparing the validated results with the early hypotheses reveals several aligned and surprising outcomes:

- Validated Hypotheses: Flight shipments are more reliable, and heavier/high-value products receive better handling. Delays correlate with customer interactions and high-volume discount periods.
- Surprises / Counter-Intuitive Results: Despite higher shipment volumes, Ship mode delays were only slightly worse than Road mode, suggesting that volume alone isn't the main factor. operational processes likely play a bigger role. Road shipments, though smaller in volume, still contributed significantly to delays, highlighting potential inefficiencies in last-mile logistics.
- Insights: Shipment mode, product characteristics, and operational factors are key determinants of delivery success. Proactive planning, prioritization, and monitoring can improve on-time delivery rates. Data-driven interventions targeting high-delay segments (Ship mode, Road shipments, and discount periods) can substantially enhance overall efficiency and customer satisfaction.

8. Data Visualizations & Charts

Visual representations were created in Microsoft Excel to simplify complex data relationships and highlight key patterns related to the delivery status. A summary of the visualizations and their interpretations is presented below.

- **Visuals**
  - **On-Time vs Delayed Deliveries by Rating Category (Bar Chart)**

A Bar chart was used to compare customer ratings with delivery status. **Insight:** Although all customer groups receive more on-time deliveries than delays, Neutral, Satisfied, and Dissatisfied customers experience the highest number of delays, highlighting them as the key segments where delivery reliability needs improvement.

- **Delivery Performance Across Warehouses (Clustered Column Chart)**

This chart visualizes the delivery performance of goods across warehouses.

Insight: Warehouse F dominates total shipments and therefore has the highest on-time deliveries and the highest delays, making it the primary driver of overall performance, while Warehouses A–D handle similar volumes with only small differences in delay rates.

- **Overall delivery performance (Pie Chart)**

This visualization categorizes delivery performance into on-time deliveries and delays.

Insight: Only about 60% of deliveries arrive on time, while a substantial 40% are delayed, signaling a meaningful performance issue that may require operational improvement.

- **Delivery Timeliness by Transport Mode (Clustered Column Chart)**

Delivery status was plotted against Transport Mode.

Insight: All shipment modes perform similarly (around 59–60% on-time delivery), but Ship handles the largest volume and therefore contributes the most to overall delays, while Flight is slightly more reliable and Road is the least reliable of the three.

- **Impact of Shipment Weight on Delivery Performance (Clustered Bar Chart)**

This visualization shows the proportions of weight categories across delivery status outcomes.

Insight: Weight is a major driver of delivery delays, with heavy shipments causing most delays, light shipments performing moderately well, and medium-weight shipments showing near-perfect on-time performance, highlighting that heavier parcels put the greatest strain on operations.

- **Frequency of Complaints Among Delayed Orders (Bar Chart)**

This visualization shows the proportions of customer care calls across delayed status.

Insight: Most delays come from customers who made 3-4 complaint calls, and overall, delays increase as the number of customer care calls rises, indicating unresolved service issues that escalate instead of being addressed early.

- **Delivery Performance by Shipment Value (Clustered Column Chart)**

This visualization shows the proportions of shipment value across delivery status.

Insight: While high-value shipments have a slightly higher on-time rate (~64.5%), standard shipments dominate total delays (94% of all delays) due to their large volume, indicating that standard products are the main source of delivery risk.

- **Delivery Performance by Discount Range (Clustered Column Chart)**

This visualization shows the proportions of discount ranges across delivery status.

Insight: All delivery delays occur in the 1–10% discount range (~53% delay rate), while orders with discounts above 10% are delivered on time, suggesting that higher-discount or promotional orders may receive prioritization or are easier to manage due to smaller volumes.

- **Top Prior Purchases by Delivery Performance (Stacked Column Chart)**

This visualization shows the effect of prior purchases on delivery status.

Insight: Customers with 3 prior purchases form the largest group and contribute most to overall delivery trends, while repeat customers with 4-6 prior purchases experience inconsistent delivery reliability, indicating that even loyal customers face delays that could affect satisfaction.

- **Comparison of Discounts and Average Weight Across Warehouses (Scatter Plot)**

This visualization shows the average discount plotted against the average weight of goods across warehouses.

Insight: All warehouses maintain consistent average discounts (13.09–13.67%) and shipment weights (3,615–3,641 gms), with Warehouse F having the highest average discount and one of the heaviest shipments, while overall shipment sizes and discounts are relatively stable across all blocks.

- **Product Delivery Efficiency and Delay Analysis (Excel Dashboard)**

A consolidated dashboard was developed using slicers, pivot charts, and KPI indicators to allow interactive exploration of: Delivery Timeliness by Transport Mode, Overall Delivery Performance, Delivery Performance by Discount Range, Frequency of Complaints among Delayed Orders, Top Prior Purchases by Delivery Performance, and Delivery Performance Across Warehouses.

Dashboard Insights: The dashboard provides an at-a-glance view of how shipment, customer, and product variables interact with on-time delivery, helping stakeholders quickly identify high-risk segments and factors contributing to delays.

9. Recommendations and Observations

This section translates the findings from the delivery analysis into actionable recommendations that can help improve operational efficiency, customer satisfaction, and overall performance.

- **Actionable Insights**

  - Ship mode handles the largest volume and contributes the most to delays. Prioritize operational improvements in Ship and Road deliveries to reduce the overall delay rate.
  - Heavy shipments account for the majority of delays. Implement better handling protocols, optimize routes for heavier parcels, or consider staggered scheduling to reduce congestion.
  - Repeat customers with 4-6 prior purchases experience inconsistent delivery reliability. Introduce targeted monitoring or priority handling to ensure loyalty is maintained.
  - All delays occur in the 1–10% discount range. Review resource allocation during high-volume standard discount periods to prevent bottlenecks.
  - Delays increase with the number of customer care calls. Establish faster issue resolution for customers reporting multiple complaints to prevent escalation.
  - Warehouse F has the highest shipment volume and delays. Focus on process improvements, staffing, or equipment upgrades in this block to reduce delays.

Optimizations or Business Decisions

- Prioritize high-risk segments (heavy shipments, high-volume warehouses, standard products) for resource allocation.
- Consider predictive scheduling to match shipment mode and product type with delivery reliability.
- Use dashboards to monitor real-time performance and intervene promptly in delayed shipments.

Unexpected Outcomes

- Medium-weight shipments performed nearly perfectly (~99.8% on time), suggesting opportunities to replicate best practices from this category to heavy and light shipments.
- High-value shipments, though fewer in number, have better on-time rates, indicating that prioritization strategies are effective and could be applied more broadly.

10. Conclusion

This section summarizes the key findings, acknowledges limitations, and outlines directions for future analysis, providing closure to the report while highlighting actionable insights.

- **Key Learnings**
  - Ship handles the largest volume, contributing most to delays, while Flight shipments show slightly better reliability.
  - Heavy shipments are the main source of delays, medium-weight shipments perform almost perfectly, and light shipments perform moderately well.
  - Repeat customers and those who call multiple times experience higher delays, indicating service escalation issues.
  - Standard products and orders with 1–10% discounts contribute disproportionately to delays, while high-value shipments and higher discount orders are prioritized effectively.
  - Warehouse F carries the largest volume, making its delays critical to overall KPIs; other warehouses show consistent performance.
- **Limitations**
  - Data Scope: The analysis is limited to the provided dataset; external factors such as weather, traffic, or logistics disruptions were not included.
  - Granularity: Certain operational factors (e.g., driver performance, exact delivery times) were not captured, limiting the depth of some insights.
- **Future Research**
  - Incorporate real-time operational data such as route tracking, traffic patterns, and driver efficiency to improve predictive analysis.
  - Explore customer segmentation in more depth to tailor delivery strategies for repeat and high-value customers.
  - Investigate optimization strategies for heavy and high-volume shipments to reduce delays and improve overall on-time delivery rates.

- Evaluate the impact of promotions and discounts in more detail to ensure resource allocation matches demand peaks.

11. References

This section provides supporting materials and resources for readers who want to explore the analysis in more detail. It ensures transparency, reproducibility and gives credit to sources and tools used.

- Dataset Source: E-Commerce Shipping Data.
- Tools Used: Microsoft Excel (Pivot Tables, Charts, Dashboard).
- External Resources:
    - Literature on delivery logistics and operational efficiency.
    - Research on customer satisfaction and shipment prioritization strategies.
    - Online tutorials and documentation for Excel analytical functions.

- **Appendices**

- Additional Charts and Tables: Detailed visualizations for shipment modes, weights, warehouse performance, customer ratings, and discount impacts.
- Raw Data Samples: Subsets of the dataset used for analysis.
- Formulas and Functions**:** Step-by-step explanations of key Excel formulas used, including:

    IF( ) to append new columns.

    Pivot Table setups for aggregating shipment performance by mode, warehouse, weight, and customer segments.

- Dashboard Screenshots: Illustrations of the consolidated dashboard displaying KPIs and critical metrics for stakeholders.