Mini E-Commerce Logistics Project Report

Analyst: Sifanele Jula

Date: 24 November 2025

Data Source: Kaggle Dataset, Mini E-Commerce Logistics by Revati Borkhade

https://www.kaggle.com/datasets/revatiborkhade03/mini-e-commerce-logistics-dataset-delivery-times (Accessed: 23 November 2025)

Introduction

Purpose of Project:

The objective of this project is to analyze delivery performance within a mini e-commerce logistics dataset. The focus is on validating data quality, cleaning the dataset, creating performance KPIs, analyzing carrier efficiency, and building a dashboard that demonstrates my Excel skills in data cleaning, analysis and visualization.

Business Context:

Logistics performance, especially delivery speed and on-time performance, is central to customer satisfaction and supply chain efficiency. This project assesses carrier reliability and delivery patterns to identify improvement opportunities in transport and fulfillment.

Dataset Description

This data set contains transactional delivery records, including:

- Order Identifier
- Order Date
- Shipping Date
- Delivery Times in days
- Warehouse Used
- Carrier Used
- On-time confirmation

This dataset reflects typical e-commerce logistics flows, from order receipt to delivery.

Data Cleaning Process

Date Type Validation:

To ensure all fields were recognized correctly, I created a helper column using: *=isnumber(cell)*

Result: all values in Ship_Date and Delivery_Date are TRUE, confirming they were stored as valid Excel serial numbers.

Cleaning Log:

Date	Issue	Action Taken	Notes
23/11/25	Verified if date values were numeric	Applied ISNUMBER ()	All values are valid

Standardizing Date Formats

Although dates were valid, they were not consistently formatted. I used: Data ⇨ Text Columns ⇨ Finish. Which prompts Excel to reinterpret the data strictly as dates.

I then applied the following format to both date columns:

DD-MMM-YYYY

(e.g., 02-Feb-2025)

Why this matters:

A consistent date format is crucial for downstream calculations like carrier efficiency and SLA checks.

Cleaning Log Entry:

Date	Issue	Action Taken	Notes
23/11/25	Inconsistent date formatting	Used Text to split then changed date format	Order and Ship dates are aligned and standardized

Delivery Duration Validation

I checked Delivery_TTime_Days column by comparing it to a helper column with the formula: [S*hip_Date – Order_Date*].

This confirmed delivery duration values matched the date difference and were numerically correct.

Cleaning Log Entry:

Date	Issue	Action Taken	Notes
23/11/25	Confirm delivery duration accuracy	Compared with dynamic formula	Values confirmed correct

Missing Values/Blanks Validation

I checked if the dataset contained any missing values or blank spaces that may affect the analysis or model performance.

Actions taken:

1. Visual scan: Performed a visual scan using the heading drop down of each column.
   - No visible blanks were found
2. Applied the COUNTBLANK () test on the entire dataset.
   - Result: 0

The conclusion is that no missing values are detected.

Duplicate Values Validation

A check for duplicates was then carried out using the Remove duplicates on the Data tab.

The test returned a message stating no duplicates were found.

Feature Engineering

I created a new column titled **O*n_Time_Flag*** to convert the Yes/No values from the O*n_Time* column into 1/0, making it easier to quantify and analyze carrier and warehouse performance.

Pivot Tables Created

- **Carrier On-Time%**
  Rows: Carrier
  Values: On_Time_Flag; Average; Percentage
  Purpose: Measure OTIF performance by carrier

- **Average Delivery Days**
  Rows: Carrier
  Values: Delivery_Time_Days; Average; One Decimal
  Purpose: Identify delivery speed differences between carriers

- **Warehouse by Carrier Performance**
  Rows: Warehouse
  Columns: Carrier
  Values: On_Time_Flag; Average; Percentage
  Purpose: Compare carrier performance across warehouse locations

- **Warehouse Performance**
  Rows: Warehouse
  Values:
    - Count of Order_Id
    - On_Time_Flag; Average; Percentage
    - Delivery_Time_Days; Average; One decimal

  Purpose: evaluate each warehouses volume, reliability and delivery speed

Analysis Suggestions

Carrier On-Time%

Overall carrier on-time performance, also known as On Time In-Full (OTIF), is **significantly below the industry expectations**. Most organizations target 90% - 98%, with SLA levels typically set at 95%. The organization's **current performance ranges from 58.2% to 62.8%**, with DHL and UPS tied as the highest performing carriers at 62.8%, while FedEx is the lowest performing at 58.2%.

The narrow performance gap indicates that **delays may not be driven by the choice of carrier but rather by internal process inefficiencies prior to dispatch**. Further investigation into internal processes such as order processing, staging, and pickup scheduling is recommended.

Average Delivery Time

The average delivery time across all carriers is approximately 3 days, which indicates that **there is a consistent delivery cycle within the network**. DHL and UPS both deliver slightly faster, at 2.9 days, while FedEx averages 3.0 days, showing a marginal difference of only 0.1 days.

The small variation suggests that **delivery speed is not significantly influenced by carrier selection**, as the organization maintains a consistent delivery speed regardless of the carrier. Since performance differences are minimal, improving overall delivery speed will therefore require process optimization at warehouse or operational level, rather than switching carriers.

Warehouse vs Carrier Performance

Warehouse-level carrier performance shows clear operational differences across the network. **Warehouse A performs the strongest overall**, with an average on-time rate of 67.5%, partially driven by better results from DHL (73.5%) and UPS (72.7%).

In contrast, Warehouse B (59.0%) and Warehouse C (57.7%) underperform across carriers, indicating that issues are likely linked to internal warehouse processes instead of carrier reliability. DHL and UPS show significant performance drops when moving from Warehouse A to Warehouse B and C, while FedEx consistently remains low across all three sites. This suggests that FedEx may not be the optimal carrier for time-sensitive deliveries.

Overall, although Warehouse A is the better performing site, **all sites perform extremely poorly when compared to industry norms**. This warrants a prompt operational investigation to get to the root cause of this poor performance.

Warehouse Performance

Across the network, a total of 548 orders were processed with an overall on-time performance of 61.1% and an average delivery time of 3.0 days.

- Warehouse A
  Orders: 169 orders

OTIF: 67.5%
Avg Delivery: 2.8 days
Insights: Strongest performance indicates higher efficiency and stable capacity

- Warehouse B
  Orders: 178
  OTIF: 59.0%
  Avg Delivery: 3.1 days
  Insights: Underperformance is not linked to volume, suggesting workflow or timing inefficiencies.

- Warehouse C
  Orders: 201
  OTIF:57.7%
  Avg Delivery: 3.0 days
  Insights: Low OTIF despite normal delivery speed points to dispatch bottlenecks such as staging congestion or inefficient pick-ups.

Overall, the comparison indicates that warehouse performance is volume sensitive, but the volume is not the determining factor of underperformance. The difference in the on-time rate between the best (67.7%) and worst (57.7%) performing warehouse is approximately 10%, which suggests the issue is beyond volume and is likely systemic, which requires urgent operational assessment.

Insights and Recommendations

1. **Overall Delivery Performance is Significantly Below Standard**

   Insight:
   Network wide performance is 61.1%, far below the typical SLA level of 95%. Delivery speed, at 3.0 days, is consistent, which means the issue is not carrier related.

   Recommendations:
   - Conduct an end-to-end process review, reviewing everything from order processing, staging to dispatch.

- Introduce daily cut-off times for staging and release.
- Implement short-term corrective actions, such as:
  - Earlier order printing and digital pick releases
  - Improved staging readiness before carrier arrival
  - Confirming pick up windows with carriers

2. **Carrier Choice Does Not Meaningfully Affect Performance**

   Insight:

   All carriers perform within 4% of each other (58.2% - 62.8%) and have nearly identical delivery times (2.9 - 3.0 days). This shows that switching carriers will not solve the OTIF issue.

   Recommendations:

- Focus on internal process improvement, not carrier replacement.
- Maintain current carrier mix but:
  - Review FedEx performance for time sensitive deliveries
  - Renegotiate expectations and pickup schedules

3. **Warehouse A Outperforms B and C Despite Lower Volume**

   Insight:

   Warehouse A achieves 67.5% OTIF and the fastest delivery time of 2.8 days. Warehouse B and C underperform regardless of carrier, proving the root cause is likely site specific not external.

   Recommendations:

- Use Warehouse A as a benchmark site to identify best practices in:
  - Picking accuracy
  - Cut off times
  - Staffing patterns
  - Staging layout
- Conduct process mapping in Warehouse B and C to identify:
  - Bottlenecks in order release
  - Congestion/staging delays
  - Shift timing misalignment
- Implement standardized operating procedures (SOPs) across all warehouses.

4. **Warehouse C Shows Signs of Capacity or Flow Breakdown**

   Insight:
   Warehouse C has the highest order volume (201) but the lowest OTIF (57.7%), while maintaining normal delivery time (3.0 days). This indicates delays occur before shipment, not during transit.

   Recommendations:
   - Review:
     - Labor allocation per shift
     - Picking-to-dispatch cycle time
     - Staging space and order congestion

   - Introduce:
     - Zone and Batch picking
     - Priority order tagging
     - Real time order visibility

5. Data Quality and Monitoring Need Strengthening

   Insight:
   Current pivots provide a strong baseline, but performance cannot be sustainably managed without ongoing monitoring.

   Recommendations:
   - Establish a monthly OTIF and delivery time dashboard monitoring:
     - Each warehouse
     - Each carrier
     - And order type
   - Track additional metrics over time:
     - Late reason codes
     - Planned vs actual pickup time
     - Order release timestamp
   - Implement monthly review meetings with all stakeholders to ensure consistent improvement.

Conclusion

This analysis provided a full view of the logistics performance across carriers and warehouses. While delivery speed across the network is stable at around 3 days, the overall on-time performance of 61.1% falls well below the industry standard of 90–98%. The warehouse-level analysis shows that there are internal operational challenges, particularly at Warehouses B and C. The analysis suggests that warehouse performance is the primary driver of delayed deliveries, rather than carrier performance alone.

The dashboard developed alongside this report allows stakeholders to quickly identify bottlenecks, compare carrier and site performance, and monitor KPIs that matter to customer experience. With further investigation into warehouse workflows, dispatch timing, and carrier allocation strategy, the organization can meaningfully improve delivery reliability and move closer to industry benchmarks.