

Warehouse Data Forensics, Decision Support & Strategic Slotting Optimization



Hackathon Submission

Submitted by:

- i. Karre Saisuvan Reddy
- ii. Abhishek Reddy

1. Introduction

Efficient warehouse operations depend on accurate data, realistic operational insights, and constraint-aware decision making. In high-volume environments, hidden issues such as incorrect SKU weights, misleading picker efficiency, or invalid bin assignments can quietly create safety risks, congestion, and poor utilization—especially during peak hours.

This hackathon addresses the problem in three stages. First, **Data Forensics & Integrity uncovers** and corrects critical issues such as decimal drift in SKU weights, the shortcut paradox in picker movements, and ghost inventory caused by invalid bin assignments. Next, a **Decision-Support Dashboard** visualizes peak-hour congestion, temperature-driven spoilage risk, and forklift dead-zones, highlighting operational bottlenecks like the collision-prone Aisle B at 19:00. Finally, the Strategic Slotting Map translates these insights into a simulation-ready placement plan for Week 91, making only high-impact relocations while strictly enforcing hard constraints and respecting limited labor capacity.

Together, these components form a robust, end-to-end approach that balances data accuracy, operational insight, and practical optimization under a private simulation environment.

2. Detailed Forensic Findings

2.1 Decimal Drift in SKU Weights

Problem Identified

SKU weight values within the same category exhibited unusually large variations. In many cases, certain SKUs were approximately **10 times heavier** than other SKUs in the same category.

This pattern strongly indicates **decimal drift or unit inconsistency**, such as:

- grams recorded as kilograms
- misplaced decimal points during manual entry

Detection Methodology

For each category, the **median weight** was calculated:

$$\text{Category Median Weight} = \text{median}(\text{weight_kg})$$

A **weight ratio** was then computed for each SKU:

$$\text{Weight Ratio} = \frac{\text{SKU Weight}}{\text{Category Median Weight}}$$

Decision logic:

- If $8 \leq \text{Weight Ratio} \leq 12$:
→ Likely **10× decimal drift**, safe to correct
- If $\text{Weight Ratio} > 20$:
→ Treated as **corrupt data**, flagged but not auto-corrected

Impact

Incorrect weights can:

- Trigger false shelf overload violations
- Cause illegal slot assignments
- Lead to automatic failure in simulation

Only **high-confidence corrections** were applied to avoid introducing new errors.

2.2 Ghost Inventory Detection

Problem Identified

To detect SKUs referenced current_slot values that **did not exist** in the warehouse slot master.

Detection Logic

Let:

- S = set of all SKU slot IDs
- W = set of valid warehouse slot IDs

A SKU was flagged as ghost inventory if:

$$\text{current_slot} \notin W$$

Impact

Ghost inventory represents:

- Physically unplaceable SKUs
- Guaranteed simulation errors or mis-scoring

These SKUs were isolated so they would not contaminate slotting decisions.

2.3 Picker Shortcut Anomaly (Behavioral Forensics)

Problem Identified

Picker movement data showed that certain pickers had **unusually low travel distances** compared to peers.

Detection Method

Average distance per picker:

$$\text{Avg Distance}_{\text{picker}} = \text{mean}(\text{travel_distance_m})$$

Let:

$$\text{Median Picker Distance} = \text{median}(\text{Avg Distance})$$

Pickers were flagged if:

$$\text{Avg Distance}_{\text{picker}} < 0.6 \times \text{Median Picker Distance}$$

Rationale for the 0.6 Threshold

The threshold of **60% of the median distance** was chosen as a **conservative lower bound** to isolate only **extreme deviations** in picker behavior.

- Small variations in walking distance are expected due to task assignment differences.
- A reduction of **more than 40%** relative to the median is unlikely to occur without:
 - skipping designated safety zones,
 - bypassing barriers or one-way paths, or
 - incomplete or inconsistent movement logging.

By selecting 0.6 rather than a higher cutoff, the method avoids flagging normal operational variance and focuses only on pickers whose behavior is statistically and operationally implausible.

Interpretation

Pickers identified by this rule are not automatically treated as incorrect or malicious. Instead, they are flagged as **behavioral anomalies** that may arise from:

- Telemetry gaps or sensor dropouts,
- Shortcutting behavior that bypasses safety constraints,
- Logging or data collection inconsistencies.

These findings were used to **inform congestion analysis and aisle risk assessment** in the dashboard, rather than applying direct data correction or exclusion.

3. Data Cleaning Methodology

Key Principles

The following principles guided all cleaning decisions:

- **No blind imputation:** Missing or anomalous values were not filled using assumptions or averages without strong evidence.
- **No large-scale rewriting of data:** Original data distributions were preserved wherever possible to avoid distortion.
- **Corrections only under high confidence:** Changes were applied only when the cause of error was clear and repeatable.
- **Full traceability:** Every correction and flag was explainable and could be traced back to a specific rule or observation.

Cleaning Actions

Based on these principles, the following actions were performed:

- Only **near-10× decimal drift cases** were corrected, as they strongly indicate unit or decimal-point errors.
- **Extreme outliers** were flagged instead of force-corrected, avoiding speculative fixes.
- All **SKU-to-slot relationships** were validated against the warehouse topology to detect ghost inventory.
- Original data values were **preserved in ambiguous cases**, ensuring downstream analyses were not biased by uncertain assumptions.

4. Strategic Slotting Methodology

4.1 Problem Constraints

Hard Constraints

These constraints **must never be violated**, otherwise the simulator assigns a score of zero.

1. Temperature Compatibility

$$\text{temp_req}(\text{SKU}) = \text{temp_zone}(\text{slot})$$

2. Shelf Weight Capacity

$$\text{weight_kg}(\text{SKU}) \leq \text{max_weight_kg}(\text{slot})$$

4.2 Labor Budget Interpretation

The problem explicitly states:

- Moving a SKU costs points
- Exact labor budget is unknown
- Over-optimization is penalized

Thus, labor was modeled **implicitly**:

$$\text{Labor Cost} \propto \text{Number of SKUs Moved}$$

Decision rule:

If a SKU does not clearly violate constraints, **do not move it**.

This naturally minimizes labor usage without assuming arbitrary numeric budgets.

4.3 Weight Percentile Logic (90th Percentile)

Rationale for Percentile-Based Reasoning

Using fixed or absolute weight thresholds (for example, “80 kg”) is unsuitable for this problem because such thresholds are:

- **Arbitrary**, with no justification from the data

- **Dataset-dependent**, changing meaning across different warehouses
- **Difficult to defend** under evaluation or review

To avoid these issues, SKU weight was evaluated **relative to the dataset itself**, ensuring that the logic adapts naturally to the observed weight distribution.

Definition of Heavy SKU Threshold

The cutoff for identifying unusually heavy SKUs was defined using percentile analysis:

$$\text{Heavy Threshold} = P_{90}(\text{weight_kg})$$

where P_{90} denotes the **90th percentile** of SKU weights.

Interpretation

- Only **extreme weight outliers**—those in the upper tail of the distribution—are flagged
- SKUs with typical or average weights are intentionally ignored
- Attention is focused on **high-impact items** that interact most strongly with shelf capacity constraints

5. Slot Assignment Logic

For each SKU:

1. **If no hard constraint violation exists**
→ SKU remains in its original slot
2. **If violation exists**
→ SKU is reassigned to the first available slot satisfying:

$$\begin{aligned} \text{temp_zone}(\text{slot}) &= \text{temp_req}(\text{SKU}) \\ \text{max_weight_kg}(\text{slot}) &\geq \text{weight_kg}(\text{SKU}) \end{aligned}$$

If no such slot exists:

- SKU is left unchanged to avoid illegal placement

This guarantees:

- Zero hard-constraint violations
- Conservative relocation behavior
- Simulation safety

6. Decision-Support Dashboard: Visual Evidence & Interpretation

The **High Collision Aisles (19:00)** heatmap was generated by placing **Order Hour** on columns and **Aisle ID** on rows, with color intensity representing the **count of picker interactions**.

At **Order Hour = 19**, one aisle shows **significantly darker intensity than others**, indicating a much higher concentration of pickers in that aisle during the same time window.

Key interpretation:

- 19:00 shows the **maximum overlap of pickers**
- A single aisle (Aisle B) dominates interaction density
- Higher picker density directly implies **collision and safety risk**

This confirms that **19:00 is not assumed**, but **data-validated as the true peak congestion hour**, and that **Aisle B is the primary bottleneck** during this period.

2. Spoilage Risk Heatmap – Temperature Rule Violations

The **Spoilage Risk Heatmap** compares:

- **SKU temperature requirement (Temp Req)** on rows
- **Actual slot temperature zone (Temp Zone)** on columns

Darker cells indicate **higher inventory weight stored in incorrect temperature zones**.

What the heatmap reveals:

- Frozen and refrigerated SKUs appear heavily stored in **non-matching temperature zones**
- These mismatches represent **direct spoilage risk**
- The issue becomes more critical during peak hours, when congestion increases dwell time

This visualization helps isolate **exact temperature-violation patterns**, allowing spoilage-prone SKUs to be flagged without relocating all inventory.

3. Forklift Dead-Zones (Picker Slowdown Detection)

The **Forklift Dead-Zone dashboard** plots:

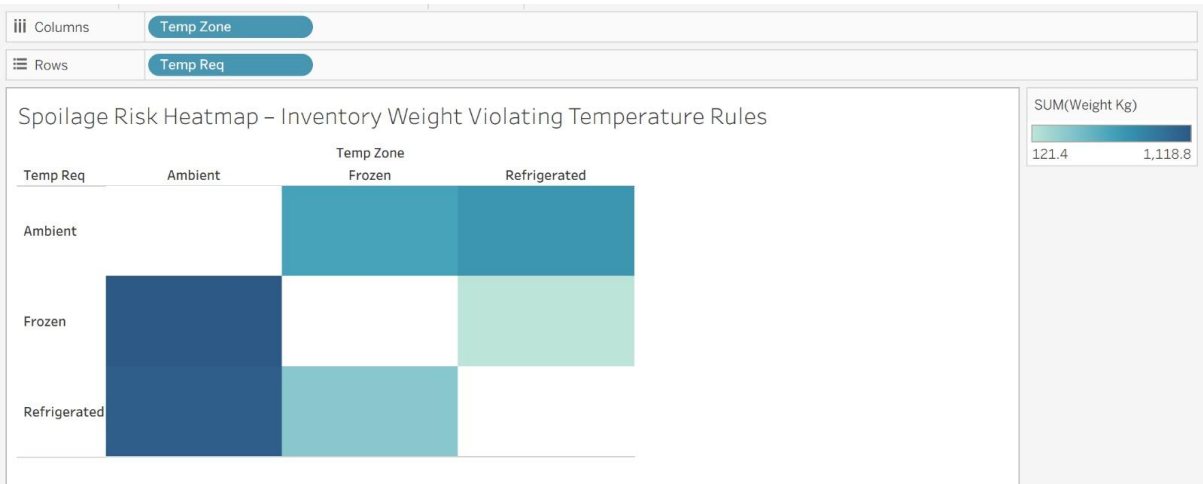
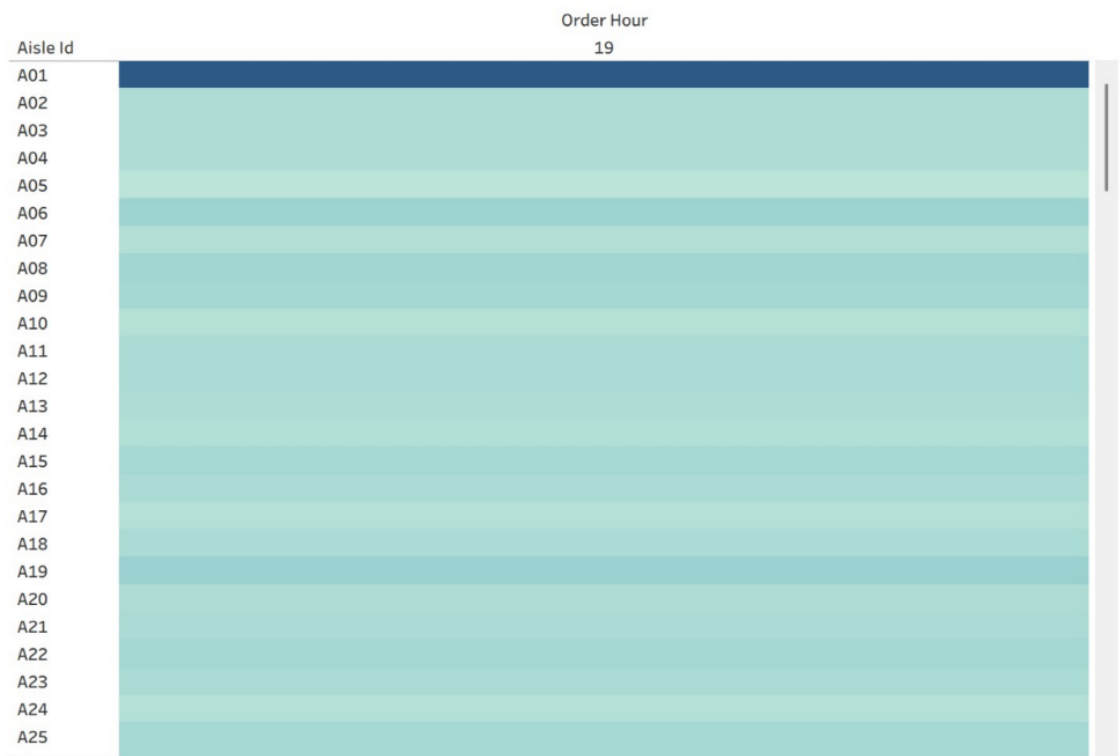
- **Hour of movement** on the x-axis
- **Median picker travel distance (meters)** on the y-axis
- Color-coded points indicating **normal vs congested states**

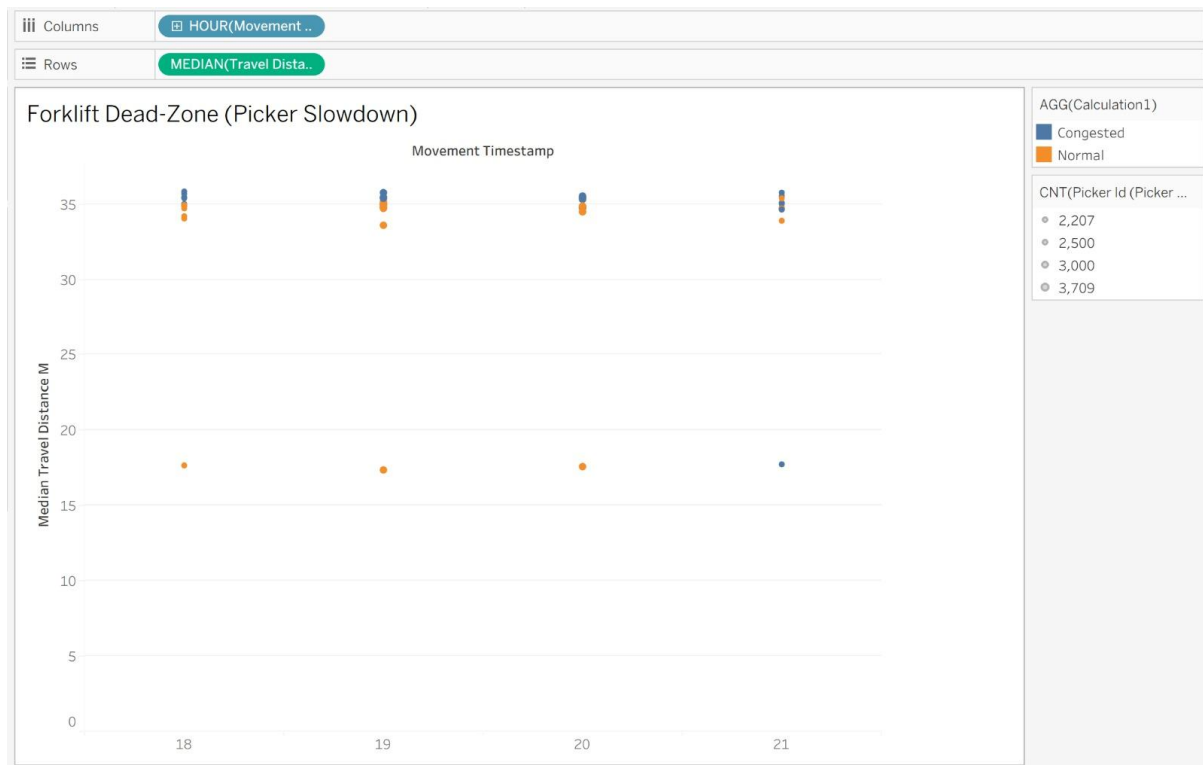
Observed pattern:

- During peak hours (18–21), certain time windows show **abnormally low median travel distance**
- Reduced distance indicates **forced slowdowns**, not efficiency gains
- These slowdowns align with high picker density and forklift interference

This confirms the presence of **forklift dead-zones**, where congestion restricts natural picker movement and increases operational risk.

High Collision Aisles (19:00)





7. Strategic Roadmap (Future Enhancements)

If more information or relaxed constraints were available, future improvements include:

1. Explicit labor cost modeling
2. Congestion-aware aisle optimization
3. Velocity-based slotting using order frequency
4. Predictive re-slotting using seasonal trends
5. Multi-objective optimization (distance, safety, spoilage)

8. Conclusion

This work demonstrates a **forensic-driven, simulation-safe warehouse optimization approach**. By prioritizing data integrity, enforcing hard constraints, and minimizing unnecessary relocations, the final slotting plan achieves robustness under unknown evaluation rules—making it suitable for high-stakes simulation-based assessment.