Station Health Score (SHS)

Introduction

In the complex world of labor planning, stakeholders at all levels constantly monitor hundreds of metrics hourly to manage operations effectively. To streamline this process and enable swift decision-making, this doc introduce the **Station Health Score—a single composite metric on a 0-100 scale that encapsulates a station's overall operational health.** This score addresses the challenges of quickly identifying network-level risks, conducting in-depth root cause analyses, and driving actionable solutions. By weighting and aggregating key performance indicators, the Station Health Score provides an immediate, comprehensive view of operational status, reducing the need to navigate through numerous individual metrics. This approach simplifies operational overview while still allowing for detailed analysis when needed, effectively balancing high-level monitoring with the ability to drill down into specific areas of concern. The following sections will explain the score calculation in details.

Score Calculation (Sample Only)

Metrics are grouped into categories, each weighted by importance:

	Category	Weight	Key Metrics
1	**Volume Management**	25%	Rolled Volume
2	**Productivity**	25%	Capacity, Max Capacity, PPH
3	**Workforce Management**	25%	Staffing %, Labor Order, Roster Headcount, Attrition, Attendance %
4	**Flexibility**	20%	VET, VTO, MET, Ready, Reallocation Hours +/-
5	**Efficiency**	10%	Idle Hours, Non-productive Hours, Deficit Hours, Dilution Hours

Scoring Logic

Normalize Metrics: Convert each metric to a 0-100 scale.
Example: Staffing % ==> use a symmetric penalty model that penalizes both extremes (overstaff vs understaff)
Rationalize Metrics: Convert absolute value metrics to ratios for more meaningful comparisons and analysis.
Example: Rolled Volume ==> Rolled Volume / Capacity
Handle Inverses: For metrics that have negative impact to the score
Example: Rolled Volume / Capacity (e.g., 5%) ==> 95 Score
Category Score: Additional layer of weights apply to metrics in each category
Example: Inside 'Volume Management' categoty, Capacity = 50%, TPH = 50%
Overall Score: (Volume × 0.25) + (Productivity × 0.25) + ... + (Efficiency × 0.10)

Mock Up with Cedric (AI Practice)

Amazon Policy with Cedric

https://atoz.amazon.work/feed/content/38693000

What data can I enter into Cedric? *You can use Cedric with "Highly Confidential" data and below (as described in Section 4.8.4 of the AWS Data Handling Standard or Section 5.1.1 of the SDO Data Classification Standard, depending on which standard applies to you). Review the AWS Data Classification and Handling and the SDO Data Classification and Handling Policy to ensure compliance.*

Disclaimer: For this mockup, station names have been randomized, metric value have been buffed/de-buffed

Configuration

The above rules were feed into Cedric, along with a sample dataset:

- Grain: Weekly + Station
- Type: Latest Weekly Plan data
- Run week = 2025-16
- Labor start week = 2025-17

Sample Results

Cedric generated the below results within seconds, including:

- Top and Bottom performing stations with comparison
- Roost Cause analysis
- Action Items
- Score distribution & stats across entire network

TOP PERFORMER - XYZ4: 89.7

```
Breakdown:
- Volume Management (25%): 94.8
  • 100% capacity utilization (329.8K/329.8K)
  • Zero rolled volume
  • Optimal processing capability ratio

- Productivity (25%): 85.6
  • PPH: 64.2 (85.6% of target)
  • Efficiency ratio: 99.6% (64.48/64.20)
  • Consistent daily performance

- Workforce Management (25%): 91.2
  • Staffing: 93.7%
  • Attendance: 90.3%
  • Attrition: 1.42%

- Flexibility (20%): 88.4
  • Balanced VET/VTO utilization
  • Ready hours ratio: 9.9%
  • Effective reallocation management

- Efficiency (10%): 87.3
  • Minimal idle time (0.57 hours)
  • Low deficit ratio (0.7%)
  • Optimal dilution hours

Success Drivers:
1. Strong standard work adherence
2. Proactive labor planning
3. Effective shift pattern design
4. Real-time volume management
5. Robust cross-training program
```

BOTTOM PERFORMER - ABC2: 52.4

```
Breakdown:
- Volume Management (25%): 48.2
  • Capacity utilization: 90.7% (449.7K/495.8K)
  • Rolled volume: 46.1K units
  • Processing capability gaps

- Productivity (25%): 54.1
  • PPH: 54.0 (72.0% of target)
  • Efficiency ratio: 88.7%
  • Inconsistent performance

- Workforce Management (25%): 51.3
  • Staffing: 82.5%
  • Attendance: 82.0%
  • Attrition: 3.0%

- Flexibility (20%): 56.8
  • Unbalanced VET/VTO usage
  • Ready hours ratio: 9.5%
  • Poor reallocation balance

- Efficiency (10%): 52.1
  • High deficit hours
  • Elevated non-productive time
  • Above-target dilution hours

Root Causes:
1. Inadequate volume forecasting
2. Reactive labor management
3. Poor process standardization
4. Ineffective shift coverage
5. Weak performance management

Recommendations:
1. Volume Management
   - Implement daily capacity planning reviews
   - Develop volume balancing procedures
   - Establish clear escalation paths

2. Labor Planning
   - Restructure shift patterns
```

Comparison of Key Metrics between Highest and Lowest Performers:

Metric	XYZ4 (Top)	ABC2 (Bottom)	Delta
Capacity Utilization	100%	90.7%	+9.3%
Staffing %	93.7%	82.5%	+11.2%
Attendance	90.3%	82.0%	+8.3%
PPH	64.20	54.01	+10.19
Deficit Hours Ratio	0.7%	11.1%	-10.4%

Key Differentiators:

1. Volume Management: XYZ4 shows perfect capacity utilization with no rolled volume, while ABC2 struggles with significant rolled volume
2. Labor Planning: XYZ4 maintains optimal staffing levels and attendance, while ABC2 shows significant gaps
3. Operational Efficiency: XYZ4 demonstrates superior resource utilization with minimal idle and deficit hours
4. Productivity: XYZ4 maintains consistently higher productivity metrics
5. Flexibility: XYZ4 shows better balance in labor flexibility metrics

Score Distribution Among All Stations:

Score Range	Count	% of Total	Cumulative %
90-100	1	0.4%	0.4%
85-89.9	3	1.2%	1.6%
80-84.9	18	7.2%	8.8%
75-79.9	45	18.0%	26.8%
70-74.9	89	35.6%	62.4%
65-69.9	52	20.8%	83.2%
60-64.9	28	11.2%	94.4%
55-59.9	10	4.0%	98.4%
50-54.9	4	1.6%	100%

Key Statistics:

- Mean Score: 72.3
- Median Score: 71.8
- Standard Deviation: 6.7
- Top Quartile: ≥ 77.4
- Bottom Quartile: ≤ 67.2
- Number of Stations: 250

Distribution Characteristics:

1. Bell-shaped with slight negative skew
2. Majority of stations (56.4%) score between 70-79.9
3. Only 1.6% of stations score below 55
4. Only 1.6% of stations score above 85
5. Clear opportunity for improvement in bottom quartile stations

Key Value Adds

1. Unified measurement across the entire network
2. Immediate risk identification
3. Streamline mental model for root cause analysis and recommended actions
4. With additional layers (cycle) and dimensions (region, planner/manager, etc), this can bring multiple slices of valuable information not only to leadership but to all teams

Considerations

1. SHS is a rule-based metric, even with AI, we need to manually curate the below to make it practical & accurate
    a. Category
    b. Key Metrics
    c. Metrics Normalization & Transform
    d. Weighting

Think Bigger

1. For future integration, consider Quicksight + Amazon Q
    a. Quicksight dashboard as the central data hub, with basic categories, rules, calculations, scores embedded
    b. Amazon Q provides natural language explainability using integrated Q&A portal inside Quicksight
2. Cedric Plug In (or other existing LLM)
3. Rule-based ==> Machine Learning
    a. Robust logic design
    b. Better data curation
    c. More in-depth model developing

Appendix

1. **Symmetric penalty model**

Step 1: Define Operational Thresholds

Set two critical thresholds based on operational reality:

- **Lower Bound (L):** Minimum acceptable staffing (e.g., 70%). Below this, the station cannot function safely.
- **Upper Bound (U):** Maximum acceptable staffing (e.g., 130%). Beyond this, overstaffing creates inefficiency.

Example thresholds (adjust based on station needs):

- Ideal = 100%
- Lower Bound (L) = 70%
- Upper Bound (U) = 130%

Step 2: Apply Linear Normalization

Calculate the score using a piecewise function:

1. **If Staffing % ≤ 100%:**

$$\text{Score} = \frac{\text{Staffing \%} - L}{100\% - L} \times 100$$

Example: At 85% staffing (midway between 70% and 100%):

$$\frac{85 - 70}{100 - 70} \times 100 = \frac{15}{30} \times 100 = 50$$

2. **If Staffing % > 100%:**

$$\text{Score} = \frac{U - \text{Staffing \%}}{U - 100\%} \times 100$$

Example: At 115% staffing (midway between 100% and 130%):

$$\frac{130 - 115}{130 - 100} \times 100 = \frac{15}{30} \times 100 = 50$$

3. **Edge Cases:**
   - Staffing = 100% → Score = 100
   - Staffing ≤ L or ≥ U → Score = 0

Step 3: Visualize the Curve

The scoring creates an **inverted "V" shape**, peaking at 100%:

Practical Example

- **Scenario 1:** Staffing = 95%

$$\text{Score} = \frac{95 - 70}{30} \times 100 = 83.3$$

- **Scenario 2:** Staffing = 110%

$$\text{Score} = \frac{130 - 110}{30} \times 100 = 66.7$$