Service Data Analysis Report

1. Data Cleaning Process - What We Did and Why

The Challenge We Faced

We received two separate datasets with quality issues:

- **Work Order Data**: 500 service records with 319 missing values
- **Repair Data**: 500 repair transactions with 437 missing values

Think of it like having two filing cabinets with incomplete paperwork - some forms were missing dates, some had blank cost fields, and others had different date formats. We needed to standardize everything before we could analyze it properly.

Our Cleaning Steps

Step 1: Fixed the Dates

- Converted all dates to a standard format (Order Date, Invoice Date)
- This ensures we can properly track trends over time

Step 2: Standardized Numbers

- Made sure all financial values (Revenue, Cost, Hours) were in proper number format
- Fixed meter readings that were stored as text
- This allows us to calculate totals and averages accurately

Step 3: Handled Missing Information

- For missing categories (like Complaint, Cause, or Part Description), we labeled them as "Unknown"
- This is better than deleting records - we keep all work orders even if some details are missing
- For financial numbers, we kept them blank if missing (rather than putting zero, which would be misleading)

Results

- **Work Orders**: Cleaned from 319 missing values to 0 missing categorical values
- **Repair Data**: Standardized all financial columns and filled missing descriptions
- **Outcome**: Both datasets are now ready for analysis with consistent formats

---

2. Integration Approach - Connecting the Dots

The Challenge

We had two separate pieces of the puzzle:

- Work orders tell us WHAT failed and WHEN
- Repair data tells us HOW MUCH it cost to fix

We needed to connect these to see the complete picture.

How We Connected Them

Primary Key Selection: "Primary Key" Column

We analyzed three options:

1. **Primary Key**: 500 unique values in work orders, 495 in repairs ✓ SELECTED
2. **Order No**: Only 232 unique values (lots of duplicates)
3. **Segment Number**: Only 19 unique values (not specific enough)

Why "Primary Key"?

- It's specifically designed as a unique identifier (the name itself tells us this!)
- Highest level of uniqueness - almost every work order has a unique key
- Best for mapping repairs back to their original work orders

The Challenge We Found

- Some work orders have multiple repair segments (one problem, multiple fixes)
- This creates a "one-to-many" relationship - like one customer order having multiple line items
- 5 work orders didn't have any repair data (maybe still pending or cancelled)

---

3. Join Type Selection - Why LEFT JOIN?

Our Decision: LEFT JOIN

Think of it like this: Work orders are our primary business records (the boss), and repair costs are additional details (the assistant). We want to keep ALL work orders, even if we don't have cost information yet.

What LEFT JOIN Does:

- Keeps all 500 work orders
- Adds repair cost information where available
- Preserves records even without matching repairs

Our Results:

- Total merged records: **505 rows** (some work orders had multiple repair segments)
- Work orders WITH repair data: **500**
- Work orders WITHOUT repair data: **5**

Why Not Other Join Types?

Join Type	What Would Happen	Why We Didn't Use It
INNER JOIN	Would keep only 495 records	✕ Would lose 5 work orders without repair data

RIGHT JOIN	Would keep all repairs but lose work order context	✕ Repair costs without knowing what failed isn't useful
OUTER JOIN	Would keep everything from both sides	✕ Would include orphan repairs without work orders

Bottom Line: LEFT JOIN ensures we maintain a complete audit trail of ALL service activities, which is critical for business operations and reporting.

---

4. Key Findings - What the Data Tells Us

Trend Analysis (Visualization 1: Monthly Service Orders)

What We Discovered:

- **Peak Demand**: February 2024 had the highest activity with **157 service orders**
- **Lowest Period**: April 2022 had only **5 orders**
- **Average Monthly Volume**: About **39 orders per month**
- **Time Period Analyzed**: 13 months of data

What This Means for You:

- There's significant seasonal variation in service demand
- February 2024 was 30x busier than April 2022
- We need flexible staffing to handle peak periods
- Resource planning should account for monthly fluctuations

Business Impact:

- Consider hiring temporary technicians during peak months
- Stock more parts before high-demand periods
- Schedule preventive maintenance during slower months

Component Failure Analysis (Visualization 2: Top Failures & Costs)

Most Common Failure:

- **"Not Mentioned"** components: 72 occurrences

- This tells us we have a documentation problem - technicians aren't always specifying what failed

Top 5 Failure-Fix Patterns:

1. **Not Mentioned → No Component Mentioned**: 28 times (Documentation gap!)
2. **Error Code - Machine**: 2 occurrences
3. **Fraying - Strap → Limit Strap Replaced**: 2 occurrences
4. **Leak - Axle**: 2 occurrences
5. **Leak - Hose**: 2 occurrences

What This Tells Us:

- **Documentation Issue**: The most "common" problem is that technicians aren't documenting failures properly
- **Real Patterns**: When documented, we see equipment issues like leaks, fraying straps, and error codes
- **Recurring Problems**: Leaks (in axles, hoses, booms) appear multiple times - potential design or maintenance issue

Revenue Insight:

- "Leak - Boom" repairs generate the highest average revenue: **$867.90**
- This suggests boom repairs are complex and costly

Cost Drivers (Visualization 3: Correlation Analysis)

Key Insight: Labor Hours Drive Costs

While we had some data quality issues with financial calculations, the pattern is clear:

- **Actual repair hours** are the primary cost driver
- The more time a repair takes, the more it costs (as expected)
- This suggests labor is our biggest expense category

Most Time-Consuming Repairs:

- All top 5 most expensive repairs took **6.4 hours** each
- These were all from **PASEIH** manufacturer equipment
- Common issues: No Heat in Cab, Not Charging Alternator, Faulty Fan, Oil Loss

Manufacturer Analysis: Average hours by manufacturer:

- **PESSTo**: 29.5 hours per repair (highest!)
- **PASEIH**: 19.3 hours per repair
- **CHR**: 5.6 hours per repair (most efficient)
- **PONDE**: 5.5 hours per repair

Root Cause Investigation

Problem #1: Documentation Gaps

Finding: 28 repairs have no failure or fix component documented

Impact: We can't identify patterns or prevent future issues

Recommendation: Implement mandatory fields in work order system

Problem #2: PASEIH Equipment Issues

Finding: PASEIH equipment appears in all top 5 most expensive repairs

Issues: Heating systems, alternators, fans, oil leaks

Recommendation:

- Review maintenance schedule for PASEIH equipment
- Consider warranty/manufacturer support review
- Evaluate if replacement is more cost-effective than repeated repairs

Problem #3: Leak Issues Across Multiple Components

Finding: Leaks appear in axles, hoses, and booms

Pattern: Suggests potential preventive maintenance opportunity

Recommendation:

- Implement regular seal and hose inspections
- Stock common leak-related parts
- Train technicians on early leak detection

Problem #4: Time-Intensive Repairs

Finding: Some repairs consistently take 6+ hours

Impact: High labor costs, equipment downtime

Recommendation:

- Analyze why these repairs take so long
- Consider pre-stocking complex assemblies
- Evaluate technician training needs

Key Recommendations

Immediate Actions (Next 30 Days)

1. **Fix Documentation Process**: Make failure/fix component fields mandatory
2. **Stock Leak Repair Parts**: Based on recurring leak issues
3. **Review PASEIH Equipment**: Assess maintenance vs replacement costs

Short-Term Actions (Next 90 Days)

4. **Staffing Plan**: Adjust resources for peak demand months (like February)
5. **Technician Training**: Focus on fast, accurate documentation
6. **Preventive Maintenance**: Target leak-prone components

Long-Term Strategy (6-12 Months)

7. **Manufacturer Review**: Evaluate PASEIH equipment reliability
8. **Pricing Analysis**: Review pricing for high-hour repairs
9. **Predictive Maintenance**: Use failure patterns to prevent breakdowns

---

Conclusion

This analysis reveals both operational insights and data quality opportunities. While we've identified clear patterns in service demand and component failures, improving documentation will unlock even better insights. The cleaned, integrated dataset is now ready for ongoing monitoring and deeper analysis.