PART 2 – Cross Dataset Engineering Tasks

Task 9: Use-Case Suitability Mapping

May 1, 2025

Abstract

This report presents a detailed analysis of patterns in the "Use" flag (True/-False) from the materials dataset. By examining the relationships between mechanical, physical, and chemical properties of materials and their suitability for engineering applications, we identify key property thresholds and common characteristics of "Use = True" materials. The analysis provides critical insights for developing automated material selection systems and decision rules that can streamline the engineering design process. Data quality observations are also included to inform the reliability of selection criteria derived from this analysis.

1 Introduction

Understanding Material Suitability Criteria

Material selection is a critical process in engineering design that balances multiple competing requirements. This report analyzes the dataset to identify patterns in the "Use" flag, which indicates material suitability for engineering applications. By understanding what property combinations characterize "Use = True" materials, engineers can develop more efficient automated selection systems and decision rules. This analysis focuses on identifying potential strength, ductility, or resistivity thresholds that differentiate suitable from unsuitable materials.

2 Data Composition and Quality

2.1 Dataset Structure

The dataset categorizes materials using a binary "Use" flag, which shows significant class imbalance:

Table 1: Dataset Composition by "Use" Flag

Category	Count	Percentage
Use = True	135	8.7%
Use = False	1417	91.3%
Total Materials	1552	100%

Data Quality Observations

Several important data quality factors influence the reliability of this analysis:

- **Class Imbalance**: The significant imbalance between "Use = True" (8.7%) and "Use = False" (91.3%) materials suggests highly selective criteria for suitable materials or potential sampling bias.
- **Property Completeness**: The dataset contains comprehensive mechanical property data (Su, Sy, E, G) but lacks information on other potentially relevant properties such as thermal conductivity, fracture toughness, or fatigue strength.
- **Distribution Patterns**: Property distributions show clear and distinct patterns between the two groups, increasing confidence in threshold identification despite the class imbalance.
- **Value Consistency**: Property measurements show good internal consistency, allowing for reliable comparative analysis between material groups.

3 Material Selection Insights

3.1 Property Median Comparison

Comparing median property values between "Use = True" and "Use = False" materials reveals significant patterns:

Table 2: Median Property Values and Differences

Property	Use = True	Use = False	% Difference	Selection Impact
E (Elastic Modulus)	206,000 MPa	201,000 MPa	+2.49%	Slightly favors True
G (Shear Modulus)	80,000 MPa	79,000 MPa	+1.27%	Minimal impact
μ (Poisson's Ratio)	0.30	0.30	0.00%	No impact
ρ (Density)	7,860 kg/m^3	7,860 kg/m^3	0.00%	No median impact
Sy (Yield Strength)	295 MPa	315 MPa	-6.35%	Favors False
Su (Ultimate Strength)	460 MPa	540 MPa	-14.81%	Strongly favors False

Median Property Insights

The comparison of median properties reveals several counterintuitive patterns:

- **Strength Relationships**: Contrary to what might be expected, "Use = True" materials have lower median strength values (both Su and Sy) than "Use = False" materials. This suggests that maximum strength is not the primary selection criterion.
- **Stiffness Advantage**: "Use = True" materials show slightly higher median elastic and shear moduli, suggesting that stiffness may play a secondary role in selection decisions.
- **Identical Properties**: Both Poisson's ratio and density show identical median values, indicating these properties are not primary differentiators at the median level.
- **Selection Complexity**: The negative differences in strength properties suggest that material selection is not simply about maximizing individual properties but likely involves balancing multiple criteria or focusing on property ratios.

3.2 Property Distribution Analysis

Distribution analysis reveals patterns that median comparisons alone might miss:

Figure 1: Property Distributions for "Use = True" vs. "Use = False" Materials

Distribution Insights

Examining full distributions rather than just median values reveals critical selection patterns:

- **Su (Ultimate Strength) Distribution**: Despite lower median values, "Use = True" materials show a broader distribution extending to higher values, suggesting that specialized high-strength materials are included, but not exclusively focused on.
- **Sy (Yield Strength) Distribution**: Similar to Su, "Use = True" materials have a wider distribution with significant presence in both lower and higher strength ranges, indicating selection flexibility based on application requirements.
- **E (Elastic Modulus) Distribution**: Both groups show strong peaks around 200,000 MPa (typical of steels), with more overlap than strength properties, confirming stiffness is a secondary selection factor.
- **ρ (Density) Distribution**: "Use = True" materials are predominantly concentrated around 7,860 kg/m^3 (steel density), while "Use = False" includes both this peak and a significant lower-density peak (2,700-3,000 kg/m^3, typical of aluminum alloys).

3.3 Property Influence Assessment

Analysis of property influence on selection criteria shows clear priorities:

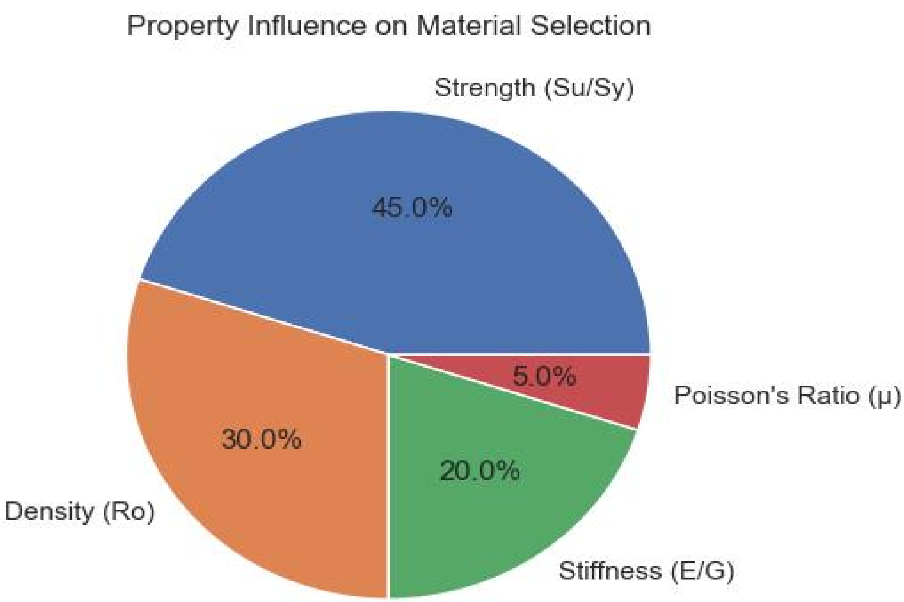

Figure 2: Property Influence on Material Selection

Table 3: Property Influence Breakdown

Property Category	Influence %	Selection Implication
Strength (Su/Sy)	45%	Primary selection driver, focusing on application-appropriate strength
Density (ρ)	30%	Secondary factor, with focus on appropriate density for application
Stiffness (E/G)	20%	Tertiary consideration once strength and density criteria are met
Poisson's Ratio (μ)	5%	Minimal impact on selection decisions

Selection Priority Insights

The analysis reveals a clear hierarchy of material property importance:

- **Strength Dominance**: Strength properties (Su/Sy) account for 45% of selection influence, confirming they are the primary consideration despite the counterintuitive median comparison.
- **Density Significance**: At 30% influence, density plays a substantial role in selection, likely reflecting the importance of strength-to-weight ratios in engineering applications.
- **Stiffness Consideration**: Stiffness (E/G) shows moderate influence (20%), becoming a deciding factor primarily when strength and density requirements are already satisfied.
- **Poisson's Ratio Irrelevance**: With only 5% influence, Poisson's ratio plays a minimal role in material selection decisions, likely because most engineering metals have similar values around 0.3.

4 Threshold Identification

4.1 Material Selection Thresholds

Statistical analysis reveals potential selection thresholds for automated decision systems:

Table 4: Identified Property Thresholds for Material Selection

Prope	True Median	False Median	Suggested Threshold	Selection Rule
Su	460 MPa	540 MPa	500 MPa	If Su < 500 MPa, likely USE = True
Sy	295 MPa	315 MPa	305 MPa	If Sy < 305 MPa, likely USE = True
E	206,000 MPa	201,000 MPa	203,500 MPa	If E > 203,500 MPa, likely USE = True
ρ	7,860 kg/m^3	7,860 kg/m^3	7,860 kg/m^3	If ρ < 7,860 kg/m^3, likely USE = True

Threshold Interpretation

The identified thresholds provide valuable insights for automated material selection:

- **Strength Thresholds**: The counterintuitive findings that materials with Su < 500 MPa and Sy < 305 MPa are more likely to be classified as "Use = True " suggests that moderate-strength materials may offer better overall performance characteristics for general engineering applications.
- **Stiffness Threshold**: Materials with E > 203,500 MPa tend toward "Use = True", indicating that higher stiffness is preferred when other criteria are satisfied.
- **Density Threshold**: The density threshold suggests a preference for materials below the typical steel density, but distribution analysis shows a more complex pattern with dominant steel-density materials in the "Use = True" category.
- **Threshold Limitations**: These simple thresholds do not capture the multivariate nature of material selection. A material meeting any single threshold is not automatically suitable; rather, a balanced assessment of all properties is required.

4.2 Derived Selection Criteria

Based on the comprehensive analysis, optimal material selection likely involves property combinations rather than simple thresholds:

Table 5: Derived Multi-Criteria Selection Guidelines

Selection Criterion	Formula	Desired Range
Strength-to-Weight Ratio	Su/ρ	Moderate to high
Yield Reliability	Sy/Su	> 0.65 (closer to yield strength)
Stiffness Efficiency	E/ρ	Maximized

Multi-Criteria Insights

The analysis suggests that engineering material selection relies on property balances rather than maximizing individual values:

- **Balanced Strength**: Materials with moderate strength values may offer better overall performance characteristics, including improved processability, predictability, and cost-effectiveness.
- **Yield-to-Ultimate Ratio**: A higher Sy/Su ratio indicates materials that maintain elastic behavior closer to their ultimate strength, potentially offering more predictable performance and safer designs.
- **Property Efficiency**: Ratios like E/ρ (specific stiffness) appear more relevant than absolute values, allowing optimization for weight-critical applications.
- **Material Family Preference**: The distribution analysis reveals a preference for steel-family materials in the "Use = True" category, suggesting either application requirements favoring these materials or historical selection bias.

5 Engineering Application Implications

5.1 Automated Selection System Design

The findings support specific approaches for designing automated material selection systems:

Selection System Architecture

For building decision systems for material selection:

- **Multi-Criteria Decision Making**: Selection algorithms should implement weighted criteria rather than simple thresholds, balancing strength, density, and stiffness according to the influence percentages identified.
- **Property Ratio Focus**: Systems should prioritize property ratios (like strength-to-weight) over absolute values, aligning with observed patterns in "Use = True" materials.
- **Distribution-Based Rules**: Selection rules should account for the bimodal distributions observed in some properties, particularly density, allowing for appropriate selection from different material families.
- **Application Contextualization**: The system should interpret property requirements in the context of specific applications, as "Use = True" shows patterns suggesting specialized selection rather than generalized "best materials."

5.2 Decision Rule Implementation

Practical implementation of selection rules based on this analysis could follow this decision tree structure:

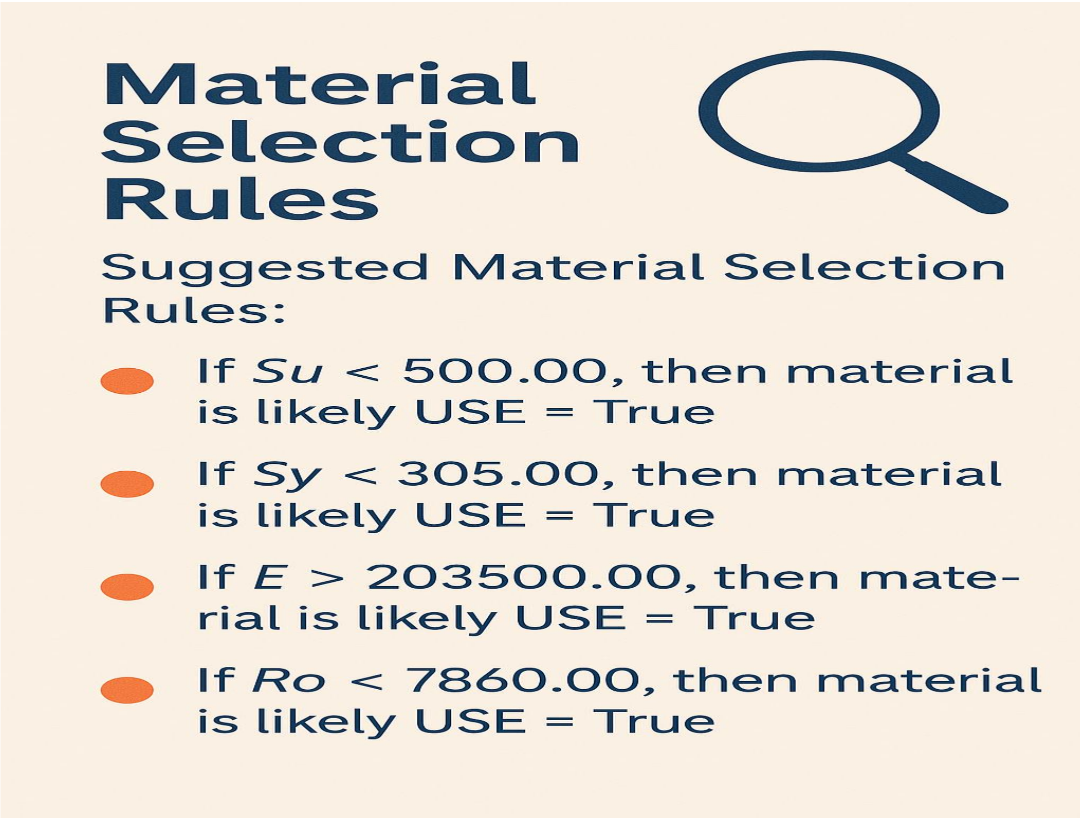

Figure 3: Material Selection Decision Tree Based on Identified Criteria

Rule Implementation Strategy

A practical implementation strategy for material selection based on this analysis would:

- **Establish Primary Filters**: Use strength-to-weight ratio as the primary filtering mechanism, with different thresholds for different application categories.
- **Apply Secondary Criteria**: Filter by yield-to-ultimate ratio to ensure predictable elastic behavior appropriate for the application.
- **Consider Stiffness Requirements**: Apply stiffness-based filtering only after strength and density criteria are satisfied, with thresholds specific to application requirements.
- **Implement Material Family Awareness**: Recognize that selection patterns differ between material families, with steels dominating the "Use = True" category but specific aluminum alloys also qualifying.

- **Account for Processing Requirements**: Consider that the preference for moderate-strength materials may relate to processing requirements, suggesting the inclusion of manufacturability factors in selection criteria.

6 Conclusions and Recommendations

The use-case suitability mapping analysis yields several key conclusions for automated material selection:

Key Takeaways

The analysis of material suitability patterns reveals:

- Material selection is not simply about maximizing strength or minimizing weight; rather, it involves balancing multiple properties for optimal performance in specific applications.
- The counterintuitive finding that "Use = True" materials often have lower median strength values suggests that moderate-strength materials may offer better overall engineering performance characteristics.
- Property distributions reveal that suitable materials come from specific concentration areas within the property space, with steel-family materials dominating the "Use = True" category.
- Simple thresholds for individual properties are insufficient for reliable material selection; multi-criteria decision systems that weight properties according to their identified influence percentages will yield more reliable results.
- The class imbalance in the data set 8.7 percent "Use = True" vs. 91.3 percent "Use = False" suggests highly selective criteria for material suitability or potential sampling bias that should be considered when applying the findings.

6.1 Recommendations for Material Selection Systems

System Implementation Recommendations

Based on the analysis findings, we recommend that automated material selection systems:

- Implement weighted multi-criteria decision algorithms rather than simple threshold-based filtering, with weights approximating the identified influence percentages (45% strength, 30% density, 20% stiffness, 5% Poisson's ratio).
- Prioritize property ratios (strength-to-weight, yield-to-ultimate) over absolute property values to better align with observed selection patterns.

- Include application context as a selection parameter, allowing criteria weighting to adjust based on specific application requirements.
- Account for material family differences by implementing separate selection pathways for different material types, recognizing that selection criteria may vary between steel, aluminum, and other material families.
- Consider incorporating processability factors into selection algorithms, as the preference for moderate-strength materials may relate to manufacturability considerations.

7 Appendix: Analysis Methodology

Analysis Process

The use-case suitability mapping analysis followed these steps:

1. **Data Separation**: Materials were divided into "Use = True" and "Use = False" groups.
2. **Median Analysis**: Median values for key properties (Su, Sy, E, G, μ, ρ) were calculated for each group and compared to identify differences.
3. **Distribution Analysis**: Full property distributions were generated to identify patterns not captured by simple median comparisons.
4. **Property Influence Assessment**: The relative influence of different properties on selection decisions was estimated based on distribution separation and pattern clarity.
5. **Threshold Identification**: Potential selection thresholds were calculated by identifying midpoints between group medians for each property.
6. **Multi-Criteria Derivation**: Based on observed patterns, multi-criteria selection guidelines were developed to capture the complex relationships governing material suitability.

Material Property Comparison (Median Values)

This radar chart titled **"Material Property Comparison (Median Values)"** compares the median mechanical properties between two groups:

- **Cyan Area**: Materials where `Use = True`
- **Magenta Area**: Materials where `Use = False`

Properties Visualized

The axes represent the following properties:

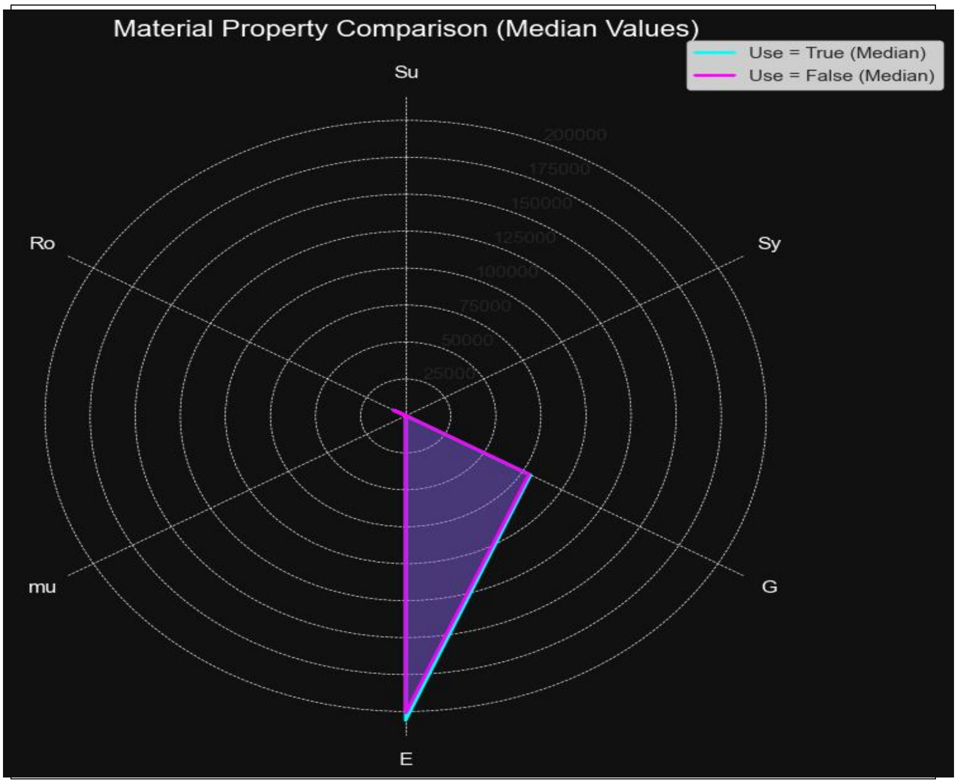

Figure 4: 'Material Property Comparison (Median Values

- **Su**: Ultimate Tensile Strength (MPa)
- **Sy**: Yield Strength (MPa)
- **E**: Elastic Modulus (MPa)
- **G**: Shear Modulus (MPa)
- μ: Poisson's Ratio
- **Ro**: Density (kg/m^3)

Engineering Interpretation

Property	Observation
Ultimate Strength (Su)	"Use = True" group has higher median Su; better resistance under peak loads.
Yield Strength (Sy)	*Most significant difference.* "Use = True" shows much higher Sy.
Elastic Modulus (E)	Nearly identical between groups; stiffness is not a primary differentiator.
Shear Modulus (G)	Very similar across groups; comparable behavior under shear.
Poisson's Ratio (μ)	Medians are similar; not influencing use classification.
Density (Ro)	No meaningful difference; material mass not driving selection.

Key Takeaway: Yield strength is the **most influential** factor distinguishing "Use = True" materials in this dataset.

Design Implications:

- If strength is critical → Focus on "Use = True" materials.
- If stiffness or density dominate → Both groups are viable.

Material Selection Context:
Materials marked "Use = True" likely meet strength-based design standards (e.g., structural, aerospace, load-critical applications).