PART 1 – Single Dataset Engineering Tasks

Task 1: Initial Exploration and Summary

April 29, 2025

Abstract

This report presents an analysis of engineering materials based on Task 1 (Initial Exploration & Summary) of a comprehensive materials dataset. We examine key mechanical properties including ultimate tensile strength (Su), yield strength (Sy), and elongation at break (A5) across 1,225 unique materials subjected to 44 different heat treatment processes. The analysis provides insights on material selection criteria, highlights data quality considerations, and demonstrates the critical impact of heat treatment on material performance. This information serves as a foundation for engineering decision-making in material selection for various applications.

1 Introduction

What is Materials Engineering?

Materials selection is a critical aspect of engineering design, requiring a systematic understanding of mechanical, physical, and chemical properties. This report focuses on the initial exploration of a materials dataset containing 1,552 entries, providing valuable insights for engineering decisions such as material selection, product design, and performance trade-offs.

2 Dataset Overview

The dataset analyzed contains detailed information on engineering materials with the following characteristics:

Table 1: Dataset Summary Statistics

Characteristic	Count	Notes
Total Records	1,552	Complete entries in dataset
Unique Materials	1,225	Distinct material types
Heat Treatment Types	44	Distinct processing methods
Heat-Treated Samples	802	51.7% of total samples

The dataset includes the following key properties:

- Material identifier (Material)
- Heat treatment method (Heat treatment)
- Ultimate tensile strength (Su) in MPa
- Yield strength (Sy) in MPa
- Elongation at break (A5) in %
- Brinell hardness (Bhn)
- Young's modulus (E) in MPa
- Shear modulus (G) in MPa
- Poisson's ratio (mu)
- Density (Ro) in kg/m^3
- pH value
- Description (Desc)
- Vickers hardness (HV)

3 Material Selection Insights

3.1 Distribution of Mechanical Properties

The analysis of key mechanical properties reveals important considerations for material selection:

Table 2: Statistical Summary of Key Mechanical Properties

Property	Min	25%	Median	Mean	75%	Max	Std Dev
Su (MPa)	69	340	500	572.8	705	2220	326.8
Sy (MPa)	28	205	305	387.0	470	2048	289.5
A5 (%)	0.5	12	16	18.9	22	70	11.6

3.2 Strength-Ductility Relationship

Key Material Selection Trade-offs

The dataset reveals important trade-offs between strength and ductility that engineers must consider when selecting materials:

- **High Strength Materials** (Su > 1000 MPa): Typically exhibit lower elongation values (A5 < 10%), making them suitable for high-stress applications where deformation must be minimized.
- **High Ductility Materials** (A5 > 30%): Generally have moderate strength values (Su between 300-500 MPa), making them appropriate for applications requiring formability and energy absorption.
- **Balanced Materials** (Su = 500 MPa, Sy = 305 MPa, A5 = 16%): Represent a compromise between strength and ductility suitable for general-purpose applications.

3.3 Impact of Heat Treatment

Heat treatment significantly modifies material properties, as demonstrated by the example of Steel SAE 1030:

Table 3: Effect of Heat Treatment on Steel SAE 1030

Heat Treatment	Su (MPa)	Sy (MPa)	A5 (%)
As-rolled	552	345	32.0
Normalized	517	345	32.0
Annealed	464	341	31.0
Tempered at 400°F	848	648	17.0

Heat Treatment Applications

This example illustrates how heat treatment can be leveraged to optimize material performance for specific engineering requirements:

- **Tempering** dramatically increases strength (Su and Sy) at the expense of ductility (lower A5)
- **Annealing** reduces strength while maintaining good ductility
- **Normalizing** provides a balanced property profile

4 Data Quality Assessment

4.1 Missing Values Analysis

Missing data presents challenges for comprehensive material comparison and selection:

Table 4: Missing Values by Property

Property	Missing Values	Percentage (%)
Heat treatment	750	48.3
A5 (elongation)	206	13.3
Bhn (Brinell hardness)	1,089	70.2
pH	1,359	87.6
Desc (description)	571	36.8
HV (Vickers hardness)	1,387	89.4

4.2 Data Inconsistencies

Data Quality Concerns

Several data inconsistencies were identified during the analysis:

- **String-based values**: Yield strength (Sy) values contained string entries with qualifiers (e.g., "280 max", "240 max") that required preprocessing
- **Extreme outliers**: Some material property values showed extreme outliers, potentially indicating specialized materials or measurement errors
- **Incomplete correlations**: Missing relationships between related properties (e.g., Bhn and HV) due to incomplete data

4.3 Outlier Analysis

After applying 5th-95th percentile capping to manage outliers:

Table 5: Property Ranges Before and After Outlier Capping

Property	Original Range	Capped Range (5th-95th)
Su (MPa)	69–2220	179–1226
Sy (MPa)	28–2048	97–980
A5 (%)	0.5–70.0	4.0–45.0

The capped ranges represent more typical values for standard engineering materials, while values outside these ranges may indicate specialized high-performance materials or potential data errors.

5 Conclusions and Recommendations

5.1 Key Findings

Major Insights

The initial exploration of the materials dataset reveals:

- A wide range of material properties supporting diverse engineering applications
- Significant impact of heat treatment on material performance
- Clear trade-offs between strength and ductility that must be considered in material selection
- Data quality issues that require careful consideration when using the dataset for decision-making

6 Appendix: Data Processing Methodology

Analysis Process

Data processing for this analysis followed these steps:

1. **Initial data loading** and exploration
2. **Data cleaning** - Conversion of string-based values to numeric format
3. **Missing value handling** through appropriate imputation methods
4. **Statistical analysis** of key mechanical properties
5. **Outlier management** through percentile capping
6. **Visualization creation** for property distributions

7. Insights from the Plot (Before & After 5th–95th Percentile Capping)

This plot shows the distribution of three key mechanical properties — **Ultimate Tensile Strength (Su)**, **Yield Strength (Sy)**, and **Elongation at Break (A5)** — after applying capping at the 5th and 95th percentiles to reduce the influence of extreme outliers.

Property-wise Interpretation:

- **Su (Ultimate Tensile Strength)**
  - **Range:** ∼190 MPa to ∼1240 MPa
  - **Distribution:** Fairly spread out, with clustering between ∼340 to ∼705 MPa.

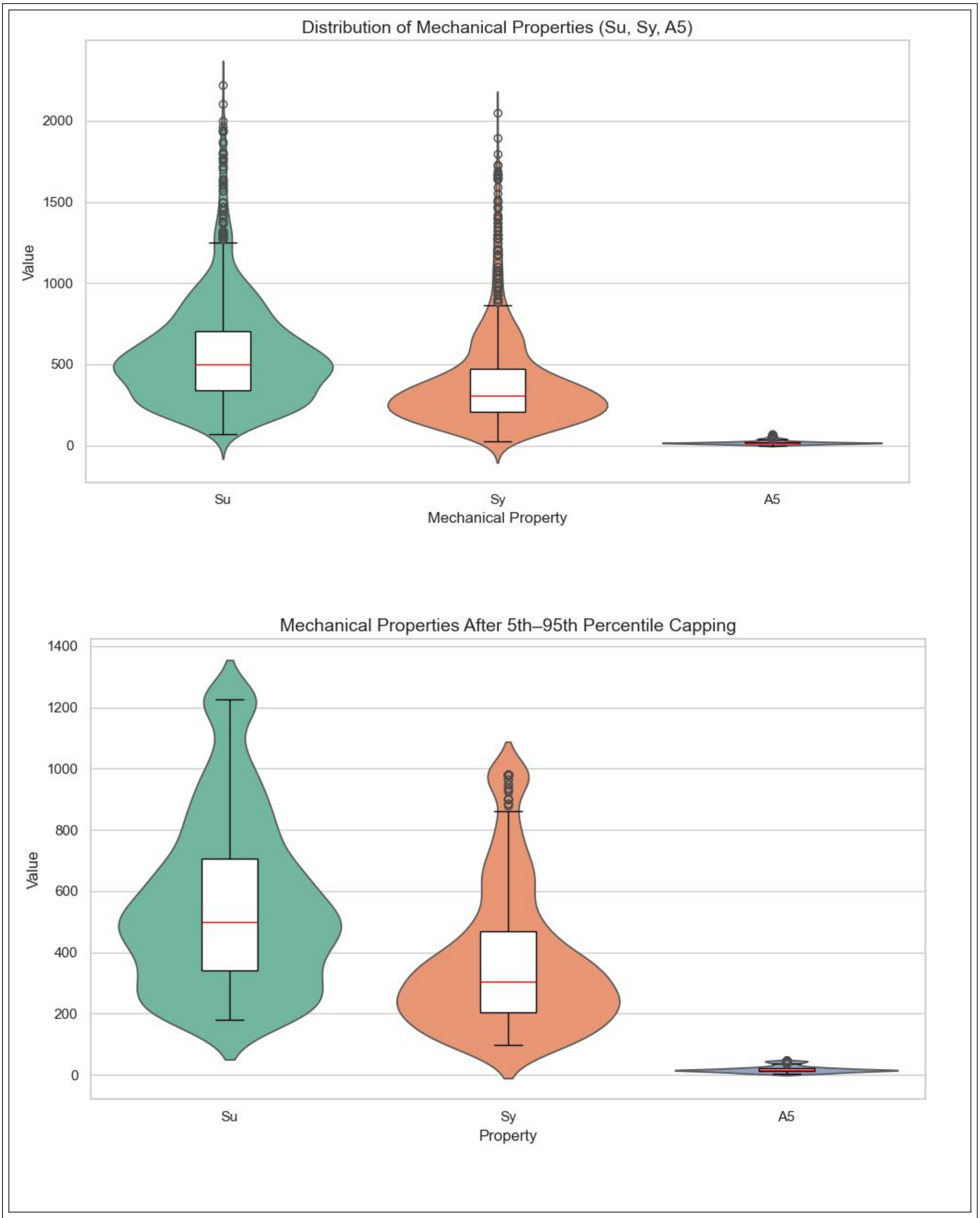


Figure 1: Distribution of Su, Sy, A5

  - **Implication:** Allows material selection across low- to high-strength applications depending on load-bearing needs.

- **Sy (Yield Strength)**
  - **Range:** ~120 MPa to ~980 MPa
  - **Distribution:** More concentrated between ~200 to ~470 MPa.
  - **Implication:** Many materials begin plastic deformation at moderate stress — critical for avoiding permanent shape change in service.

- **A5 (Elongation at Break)**
  - **Range:** ~5% to ~30%
  - **Distribution:** Narrower and skewed; most materials are moderately ductile.
  - **Implication:** Aids in selecting materials that offer sufficient formability, especially useful in forming or energy absorption zones.

Engineering Takeaways:

- Su and Sy span a broad range, enabling design flexibility for both lightweight and high-stress applications.
- A5 helps differentiate ductile versus brittle materials, guiding choices in applications needing elongation before fracture.
- The boxplot visualization effectively highlights where the majority of material values lie — making it easier to shortlist candidates during preliminary material selection.

PART 1 – Single Dataset Engineering Tasks

Task 2: Groupwise Comparison

April 29, 2025

Abstract

This report presents a comparative analysis of engineering materials based on Task 2 (Groupwise Comparison) of the materials dataset. We examine how material type and heat treatment influence mechanical properties including ultimate tensile strength (Su), elongation at break (A5), and hardness values (Bhn/HV). The analysis provides critical insights on material selection criteria based on processing methods and material grades, highlighting the strength-ductility trade-offs that engineers must consider when selecting materials for various applications. Data quality observations are also included to inform the reliability of engineering decisions based on this analysis.

1 Introduction

The Importance of Material Grouping

Understanding how materials and processing methods impact mechanical properties is fundamental to effective engineering design. This report analyzes the dataset by grouping materials based on type and heat treatment method, revealing patterns that can significantly influence material selection decisions. By examining how properties vary across these groups, engineers can make more informed choices that balance strength, ductility, and hardness requirements for specific applications.

2 Material Selection Insights

2.1 Material Type Comparison

The analysis by material type reveals significant patterns in mechanical properties that can guide material selection:

Table 1: Material Properties by Material Type (Top 5 by Strength)

Material	Su (MPa)	A5 (%)	Hardness	Application Suitability
BS 525A60	1226	6	N/A	High-stress, limited deformation
BS 735A51	1226	6	N/A	Static loads, minimal ductility required
CSN 16640	1226	16	HV: 510	Balanced strength-ductility profile
Steel SAE 8660	1226	13	BHN: 460	Components needing toughness
Steel SAE 8640	1226	10	BHN: 505	Wear-resistant applications

Key Insights from Material Comparison

The groupwise analysis by material type reveals critical patterns for engineering material selection:

- **High-Strength Selection**: Several materials exhibit identical ultimate tensile strength (Su = 1226 MPa) but with varying ductility, creating important selection trade-offs.
- **Strength-Ductility Balance**: Materials like CSN 16640 offer optimal balance of high strength (1226 MPa) and good ductility (16%), making them suitable for applications requiring both properties.
- **Surface Resistance**: SAE 8640 provides excellent hardness (BHN = 505) alongside high strength, making it suitable for wear-resistant components.
- **Application-Specific Selection**: BS grades excel in high-strength, low-deformation applications, while SAE grades balance multiple mechanical requirements for components like gears and shafts.

2.2 Heat Treatment Comparison

Heat treatment methods significantly alter material properties, creating distinct performance profiles:

Table 2: Material Properties by Heat Treatment Method (Top 5 by Strength)

Heat Treatment	Su (MPa)	A5 (%)	Hardness	Application Suitability
Full-hard	1226	6	N/A	Maximum strength requirements
Nitro-case-hard.	1226	12.5	HV: 630	Wear resistance with ductility
Tempered at 800°F	1226	10.5	BHN: 465	Balanced mechanical properties
3/4-hard	1207	8.5	N/A	High strength, formability
Tempered at 400°F	1173	10.5	BHN: 463	General mechanical components

Processing Method Selection Guidelines

The heat treatment comparison provides valuable guidance for processing-based material selection:

- **Surface Engineering**: Nitro-case-hardening delivers exceptional surface hardness (HV = 630) while preserving reasonable ductility (A5 = 12.5%), making it optimal for components requiring wear resistance.
- **Tempering Temperature Effects**: Higher tempering temperature (800°F vs. 400°F) increases strength while maintaining similar ductility and hardness profiles.
- **Full-Hard Treatment**: Maximizes strength but at significant ductility cost, suitable only for applications where deformation is not a concern.
- **Balanced Processing**: 3/4-hard treatment provides a compromise between maximum strength and processing feasibility, potentially easier to machine before final hardening.

3 Application-Specific Recommendations

Based on the groupwise analysis, specific material-treatment combinations can be recommended for common engineering applications:

Table 3: Recommended Material-Treatment Combinations by Application

Application	Material	Treatment	Key Properties
Wear-resistant components	SAE 8640/8660	Nitro-case-hardened	High Su, HV = 630
High-load transmission	CSN 16640	Tempered at 800°F	Su = 1226 MPa, A5 = 10.5%
Static structural parts	BS grades	Full-hard	Maximum strength, minimal ductility
Forming operations	Lower-strength steels	Annealed/Normalized	Moderate Su, high A5
General mechanical parts	SAE grades	Tempered at 400°F	Balanced strength-ductility

Engineering Decision Support

The groupwise analysis yields critical engineering insights:

- **Multi-property Consideration**: Material selection requires evaluation of multiple properties simultaneously—focusing on just strength (Su) can lead to brittle failure in applications requiring ductility.

- **Processing Impact**: Heat treatment can dramatically alter properties of the same base material, providing a pathway to tailor material performance to specific requirements.
- **Balanced Selection**: For most general applications, materials with moderate strength (500-700 MPa) and good ductility (A5 ¿ 10%) provide optimal performance and processing flexibility.
- **Special Cases**: Where extreme properties are required (very high strength or ductility), specialized material-treatment combinations can be selected with awareness of their limitations.

4 Data Quality Observations

4.1 Missing Value Patterns

The groupwise analysis reveals systematic data quality issues that must be considered when making engineering decisions:

Table 4: Missing Data Patterns in Property Measurements

Property	Pattern	Impact on Analysis
Hardness (Bhn/HV)	Missing across multiple material groups	Limited hardness-based comparisons
Hardness in heat treatments	Full-hard and 3/4-hard treatments lack values	Cannot evaluate surface effects
Measurement methodology	Inconsistent use of Bhn vs. HV scales	Challenges in direct comparisons

4.2 Data Consistency Concerns

Data Quality Limitations

Several data quality issues impact the reliability of group comparisons:

- **Value Uniformity**: Multiple materials show identical Su values (1226 MPa), suggesting potential standardization, rounding, or measurement limitations.
- **Missing Hardness Data**: The absence of hardness measurements for many treatment types limits comprehensive surface property evaluation.
- **Measurement Scale Variations**: Some materials report Brinell hardness while others use Vickers hardness, complicating direct comparisons.

- **Data Completeness**: Properties critical for comprehensive material evaluation (like impact resistance or fatigue behavior) are not included in this groupwise analysis.

5 Conclusions and Recommendations

The groupwise comparison of engineering materials yields several key conclusions for material selection:

Key Groupwise Comparison Takeaways

The analysis of materials by type and heat treatment reveals:

- Material selection should prioritize the balance between strength and ductility most appropriate for specific applications.
- Heat treatment provides a powerful method to tailor material properties, with treatments like nitro-case-hardening offering especially favorable property combinations.
- Data quality issues, particularly missing hardness values and measurement scale variations, require careful consideration when making engineering decisions.
- For critical applications, additional testing or supplementary data sources should be consulted to address gaps in the available property measurements.
- Materials with similar strength values may differ substantially in other properties, underscoring the importance of multi-property evaluation in selection processes.

5.1 Engineering Implications

The groupwise analysis provides valuable guidance for engineering design and material selection decisions:

Practical Engineering Applications

Engineers can apply these findings in several ways:

- **Informed Material Selection**: Choose materials with appropriate strength-ductility combinations based on grouped data to avoid over-engineering or under-performance.
- **Processing Optimization**: Select heat treatments that enhance critical properties for specific applications while maintaining acceptable levels of secondary properties.

- **Design Confidence**: Understand the typical property ranges within material groups to establish appropriate design safety factors and performance expectations.
- **Data-Driven Decisions**: Recognize data quality limitations when making critical engineering decisions, particularly for properties with significant missing values.
- **Performance Prediction**: Use group-level property patterns to anticipate material behavior in service and guide testing requirements for verification.

6 Appendix: Group Comparison Methodology

Analysis Process

The groupwise comparison analysis followed these steps:

1. **Group formation** based on material type and heat treatment method
2. **Aggregate calculation** of mean values for Su, A5, Bhn, and HV within each group
3. **Ranking of materials** by ultimate tensile strength (Su)
4. **Cross-property evaluation** to identify strength-ductility-hardness relationships
5. **Data quality assessment** to identify missing value patterns and consistency issues
6. **Application mapping** to connect property profiles with engineering requirements

PART 1 – Single Dataset Engineering Tasks

Task 3: Design Ratio Analysis

April 30, 2025

Abstract

This report presents a specialized analysis of engineering materials based on Task 3 (Design Ratio Analysis) of the materials dataset. We examine three critical custom strength metrics: Strength-to-Hardness ratio, Strength-to-Ductility index, and Strength-to-Weight proxy, revealing significant patterns in material performance profiles. The analysis provides essential insights for engineering material selection based on application-specific requirements, highlighting the intrinsic trade-offs between different performance metrics. Material-specific recommendations are provided alongside data quality observations to support informed engineering decisions.

1 Introduction

Custom Metrics for Enhanced Material Selection

Engineering materials selection often requires consideration of multiple properties simultaneously rather than focusing on individual metrics. This report analyzes the dataset using custom ratio metrics designed to highlight specific performance attributes of materials: Strength-to-Hardness (Su/Bhn), Strength-to-Ductility (Su×A5), and Strength-to-Weight (Su/Ro). By examining how materials perform across these ratios, engineers can identify optimal candidates for specific applications and understand the inherent trade-offs involved in material selection decisions. This analysis provides data-driven guidance for critical engineering choices where particular combinations of properties determine performance success.

2 Material Selection Insights

2.1 Strength-to-Hardness Ratio Analysis

The Strength-to-Hardness ratio (Su/Bhn) reveals materials that offer excellent strength relative to their hardness, which is particularly valuable for applications requiring both strength and machinability or formability.

Table 1: Top Materials by Strength-to-Hardness Ratio

Material	**Heat Treatment**	**Strength-to-Hardness**	**Application Suitability**
Aluminum Alloy 1060-O	Wrought	9.42	Forming operations, low-stress applications
Aluminum Alloy 1060-H12	Wrought	7.78	Slightly stronger, still highly formable
Aluminum Alloy 1100-O	Wrought	7.78	General forming, good corrosion resistance
Aluminum Alloy B443.0	Cast (F)	7.16	Cast components with good machinability
Aluminum Alloy 6063-O	Wrought	7.16	Architectural extrusions, detail components

Strength-to-Hardness Insights

The analysis of Strength-to-Hardness ratios reveals several key patterns with important engineering implications:

- **Aluminum Dominance**: Wrought aluminum alloys significantly outperform other materials in this metric, with values reaching 9.42 for Aluminum Alloy 1060-O.
- **Heat Treatment Impact**: The annealed (O) condition consistently produces higher Strength-to-Hardness ratios compared to hardened states, highlighting the impact of processing on this property relationship.
- **Cast vs. Wrought Comparison**: While wrought alloys lead this metric, cast aluminum B443.0 still performs exceptionally well (7.16), indicating good machinability potential in properly selected cast alloys.
- **Application Guidance**: Materials with high Su/Bhn ratios are ideal candidates for applications requiring secondary machining operations, moderate strength components with complex geometry, or parts requiring excellent formability and moderate strength.

2.2 Strength-to-Ductility Index Analysis

The Strength-to-Ductility index (Su×A5) identifies materials that provide an optimal balance of strength and ability to deform before fracture, crucial for applications requiring toughness and energy absorption.

Table 2: Top Materials by Strength-to-Ductility Index

Material	Heat Treatment	Strength-to-Ductility	Application Suitability
Steel SAE 30301	Annealed	34,110	High-impact resistance, fatigue applications
Steel SAE 30314	Annealed	31,005	Extreme toughness requirements
Steel SAE 30302B	Annealed	29,475	Biomedical, corrosion + toughness needs
Steel SAE 30310	Annealed	29,475	High-temperature tough applications
Steel SAE 30347	Annealed	28,530	Stabilized grade for welded structures

Strength-to-Ductility Insights

The Strength-to-Ductility index analysis reveals critical patterns for engineering applications requiring balanced mechanical performance:

- **Stainless Steel Superiority**: The SAE 303 series stainless steels in annealed condition dominate this category with exceptionally high values reaching 34,110 for SAE 30301.
- **Heat Treatment Consistency**: The annealed condition consistently produces superior Strength-to-Ductility performance across all top performers, indicating this heat treatment's critical role in optimizing toughness.
- **Material Family Pattern**: The overwhelming dominance of austenitic stainless steels suggests their microstructural characteristics intrinsically support excellent strength-ductility combinations.
- **Application Implications**: These materials are ideal candidates for critical safety components, energy-absorbing structures, and applications subject to impact loading or requiring significant deformation before failure.

2.3 Strength-to-Weight Ratio Analysis

The Strength-to-Weight ratio (Su/Ro) highlights materials offering maximum strength for their weight, crucial for weight-critical applications in aerospace, automotive, and portable equipment design.

Table 3: Top Materials by Strength-to-Weight Ratio

Material	Heat Treatment	Strength-to-Weight	Application Suitability
Aluminum Alloy 7075-T6	Wrought	0.212	Aerospace structural components
Aluminum Alloy 7075-T651	Wrought	0.212	Stress-relieved aerospace structures
Aluminum Alloy 7049-T73	Wrought	0.191	Corrosion-resistant aerospace parts
Aluminum Alloy 7049-T7352	Wrought	0.191	Stress-relieved aircraft components
Aluminum Alloy 2024-T361	Wrought	0.184	Aircraft fuselage, structural frames

Strength-to-Weight Insights

The analysis of Strength-to-Weight ratios yields important insights for weight-critical engineering applications:

- **7000-Series Excellence**: Aluminum alloys in the 7000 series (particularly 7075-T6/T651) demonstrate superior performance with values reaching 0.212, significantly outperforming other material families.
- **Heat Treatment Criticality**: The T6 and T73 tempers consistently produce superior Strength-to-Weight ratios compared to other conditions, highlighting the importance of precipitation hardening treatments.
- **Consistency Pattern**: The remarkably consistent Strength-to-Weight values (approximately 0.2) among top performers suggests these materials have been specifically engineered and optimized for weight-efficient structural applications.
- **Application Targeting**: These materials represent optimal choices for aerospace components, high-performance sporting equipment, portable devices, and automotive components where weight reduction directly impacts energy efficiency or performance.

3 Material Selection Trade-offs

The multi-ratio analysis reveals critical trade-offs that engineers must consider when selecting materials for specific applications.

Table 4: Material Performance Trade-offs Across Multiple Ratios

Material	Strength-to-Hardness	Strength-to-Ductility	Strength-to-Weight	Trade-off Pattern
7075-T6	Low	High	High	Sacrifices machinability for strength efficiency
7049-T73	Low-Moderate	Low-Moderate	High	Balanced compromise across metrics
7049-T7352	Very Low	Very Low	High	Extreme optimization of weight at expense of other properties
1060-O	Very High	Low	Low	Optimized for forming at expense of strength
SAE 30301	Moderate	Very High	Low	Maximizes toughness at expense of weight efficiency

Key Trade-off Insights

The cross-ratio analysis reveals fundamental material selection trade-offs with significant engineering implications:

- **Property Exclusivity**: The data clearly demonstrates that simultaneous optimization of all three ratios is impossible with existing materials—excellence in one metric typically comes at the expense of another.
- **Heat Treatment Impact**: The same base material treated differently (e.g., 7049-T73 vs. 7049-T7352) shows dramatically different ratio profiles, emphasizing processing as a key lever for property customization.
- **Material Family Specialization**: Different material families show distinct ratio strengths—aluminum alloys excel in Strength-to-Weight and Strength-to-Hardness, while stainless steels dominate Strength-to-Ductility performance.
- **Application Prioritization**: Material selection must be driven by which property ratio is most critical for a specific application, with clear understanding of the compromises being accepted in other performance areas.

4 Material Selection Guidelines

Based on the custom ratio analysis, specific material selection recommendations can be made for common engineering applications.

Table 5: Recommended Materials by Application Requirements

Application Requirement	Recommended Material	Key Ratio Strength	Accepted Trade-offs
Lightweight structures	Aluminum Alloy 7075-T6/T651	Strength-to-Weight (0.212)	Lower machinability
Energy absorption	Steel SAE 30301 (annealed)	Strength-to-Ductility (34,110)	Higher weight
Complex machined components	Aluminum Alloy 1060-O	Strength-to-Hardness (9.42)	Lower absolute strength
Balanced performance	Aluminum Alloy 7049-T73	Moderate in all ratios	No exceptional properties
Corrosion-resistant structures	Steel SAE 30310/30347	High Strength-to-Ductility	Lower weight efficiency

Material Selection Strategy

The ratio analysis supports these material selection strategies for engineers:

- **Application-Driven Selection**: Begin material selection by identifying which ratio is most critical for your specific application requirements.
- **Balanced Consideration**: When multiple properties matter, consider materials like 7049-T73 that offer moderate performance across multiple ratios rather than excellence in just one.
- **Processing Optimization**: Consider how heat treatment can be used to shift the balance of properties toward your specific performance requirements.
- **Specialty Applications**: For extreme performance requirements, select materials that specifically excel in the critical ratio, accepting the inevitable compromises in other areas.
- **Industry-Specific Guidance**: For aerospace and transportation, prioritize 7000-series aluminum alloys; for energy-absorbing and safety-critical components, consider annealed 300-series stainless steels; for complex machined components with moderate strength needs, select 1000-series aluminum in O-condition.

5 Data Quality Observations

Data Quality Considerations

Several data quality issues affect the reliability and comprehensiveness of the custom ratio analysis:

- **Missing Value Impact**: Numerous materials had to be excluded from the analysis due to missing values for one or more properties required for ratio calculations (Su, Bhn, A5, or Ro), potentially eliminating high-performing candidates from consideration.
- **Ratio Value Consistency**: The remarkably consistent Strength-to-Weight values (approximately 0.2) across aluminum alloys suggests possible data standardization, rounding effects, or measurement limitations that may obscure subtle but potentially important differences between materials.
- **Extreme Value Verification**: The exceptionally high Strength-to-Ductility values for annealed stainless steels (exceeding 30,000) warrant verification to ensure measurement accuracy before making critical design decisions based on these ratios.
- **Processing Detail Limitations**: Limited information about precise processing history (cold work percentage, exact aging parameters, etc.) restricts full understanding of why certain materials perform better in specific metrics.
- **Material Family Coverage**: While aluminum alloys and steels are well represented, limited data on other material families (titanium alloys, magnesium alloys, etc.) that could offer competitive property ratios constrains the comprehensiveness of selection recommendations.

6 Conclusions and Recommendations

The design ratio analysis yields several key conclusions to guide engineering material selection:

Key Analysis Takeaways

The custom ratio analysis provides valuable insights for engineering decision-making:

- Material selection should be driven by identifying which property ratio is most critical for the specific application requirements, with clear understanding of the inevitable trade-offs.
- Heat treatment offers a powerful mechanism to shift material properties toward specific ratio performance targets, with annealing generally favoring Strength-to-Hardness and Strength-to-Ductility while age-hardening treatments (T6, T7) maximize Strength-to-Weight performance.

- Different material families show clear specialization patterns: aluminum alloys excel in Strength-to-Weight and Strength-to-Hardness ratios, while stainless steels dominate Strength-to-Ductility performance.
- Data quality limitations, particularly missing values and potential measurement consistency issues, require careful consideration when making engineering decisions based on these custom ratios.
- For critical applications, engineers should consider supplementing this ratio analysis with additional testing or verification of specific properties most crucial to their design requirements.

6.1 Engineering Applications

The custom ratio analysis provides specific guidance for real-world engineering applications:

Practical Implementation Guidelines

Engineers can apply these findings in several ways:

- **Aerospace Structures**: Select 7075-T6/T651 aluminum alloys for optimal Strength-to-Weight performance in primary structural components.
- **Automotive Safety Systems**: Choose annealed SAE 303 series stainless steels for energy-absorbing components where deformation before failure is critical.
- **Consumer Electronics**: Consider 7000-series aluminum alloys for portable device frameworks requiring high strength and minimal weight.
- **Manufacturing Efficiency**: Select 1000-series aluminum in O-condition for components requiring extensive machining or forming operations while maintaining adequate strength.
- **Medical Implants**: Implement SAE 30310/30347 stainless steels where a combination of strength, ductility, and corrosion resistance is essential.

7 Appendix: Design Ratio Methodology

Analysis Approach

The design ratio analysis employed these methodological steps:

1. **Ratio Definition**:
    - Strength-to-Hardness (Su/Bhn): Ultimate tensile strength divided by

Brinell hardness

- Strength-to-Ductility (Su×A5): Ultimate tensile strength multiplied by elongation percentage
- Strength-to-Weight (Su/Ro): Ultimate tensile strength divided by density

2. **Data Processing**:
   - Removal of rows with NaN values in required properties
   - Calculation of each ratio for remaining materials
   - Sorting and ranking materials by each ratio
3. **Visualization**:
   - Heat map generation to compare ratio values across materials
   - Radar chart creation to visualize trade-offs between different ratios
4. **Analysis**:
   - Identification of top-performing materials in each ratio
   - Cross-ratio comparison to identify trade-off patterns
   - Material family and heat treatment pattern recognition

8. Index-Based Material Comparison for Aluminum Alloys

Key Observations from the Heatmap:

- **Consistency in Strength-to-Weight Ratio (Su / Ro)**
  All materials have a Strength-to-Weight index ≈ 0.2 — nearly identical across the board.
  Engineering Insight: These aluminum alloys offer very similar structural efficiency per unit weight, making them excellent candidates for weight-sensitive applications (e.g., aerospace, automotive).
- **Variation in Strength-to-Ductility (Su × A5)**
  There's a wide spread in the Strength-to-Ductility index, from around 5600 to 9380.
  Engineering Insight: Some alloys (like 2024-T4 and 2024-T351) show superior toughness, meaning they maintain high strength and deformability — crucial in crash-critical components or fatigue-prone environments.
- **Strength-to-Hardness (Su / Bhn) is Modestly Differentiated**
  Mostly between 3.6 and 4.0, with slight edge for 2024-T3.
  Engineering Insight: This suggests similar resistance to wear per unit strength, but 2024-T3 may handle surface stress slightly better. Good for riveted or bolted structures where localized hardness matters.

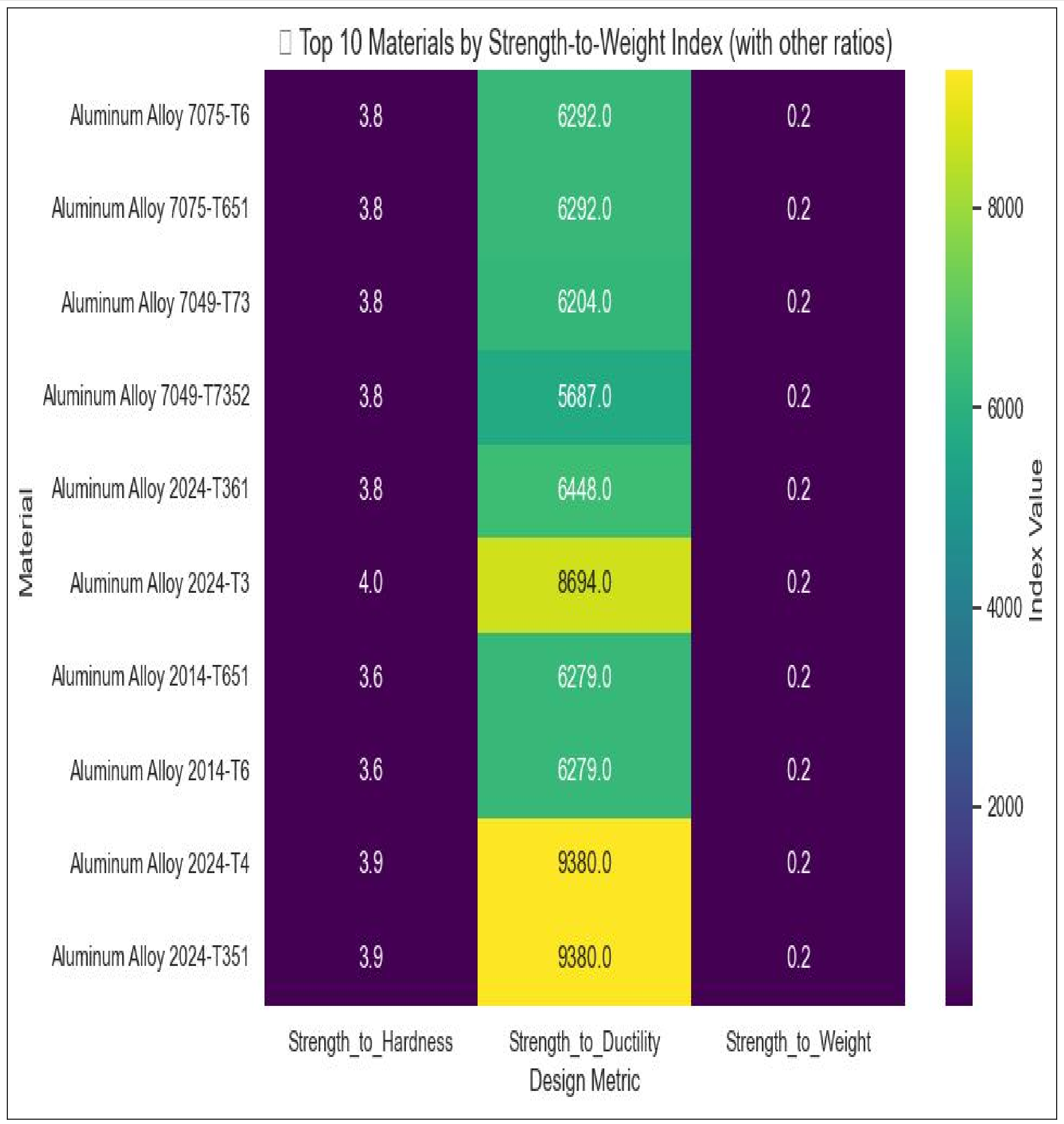


Figure 1: Heatmap of Key Performance Indexes for Aluminum Alloys (Su/Ro, Su×A5, Su/Bhn)

Design Goal	Best Candidate(s)	Why
Lightweight structure	Any (all ≈ 0.2 Su/Ro)	Excellent strength-to-weight across the board
Ductility + strength (tough)	2024-T4 / 2024-T351	Highest Su × A5 values
Surface contact or wear zones	2024-T3	Slightly better Su/Bhn → better for fasteners/joints

Table 6: Material Selection Based on Performance Indexes

Engineering Application Areas:

- **Aerospace fuselage panels:** Go for **2024-T351** — strong, tough, and light.
- **Automotive crash structures: 2024-T4** or **7075-T6** offer strength with decent ductility.
- **Bike frames, tools:** If minimal deformation is acceptable, **7075-T651** gives rigidity and strength.

9. Trade-Off Analysis in Material Indexes

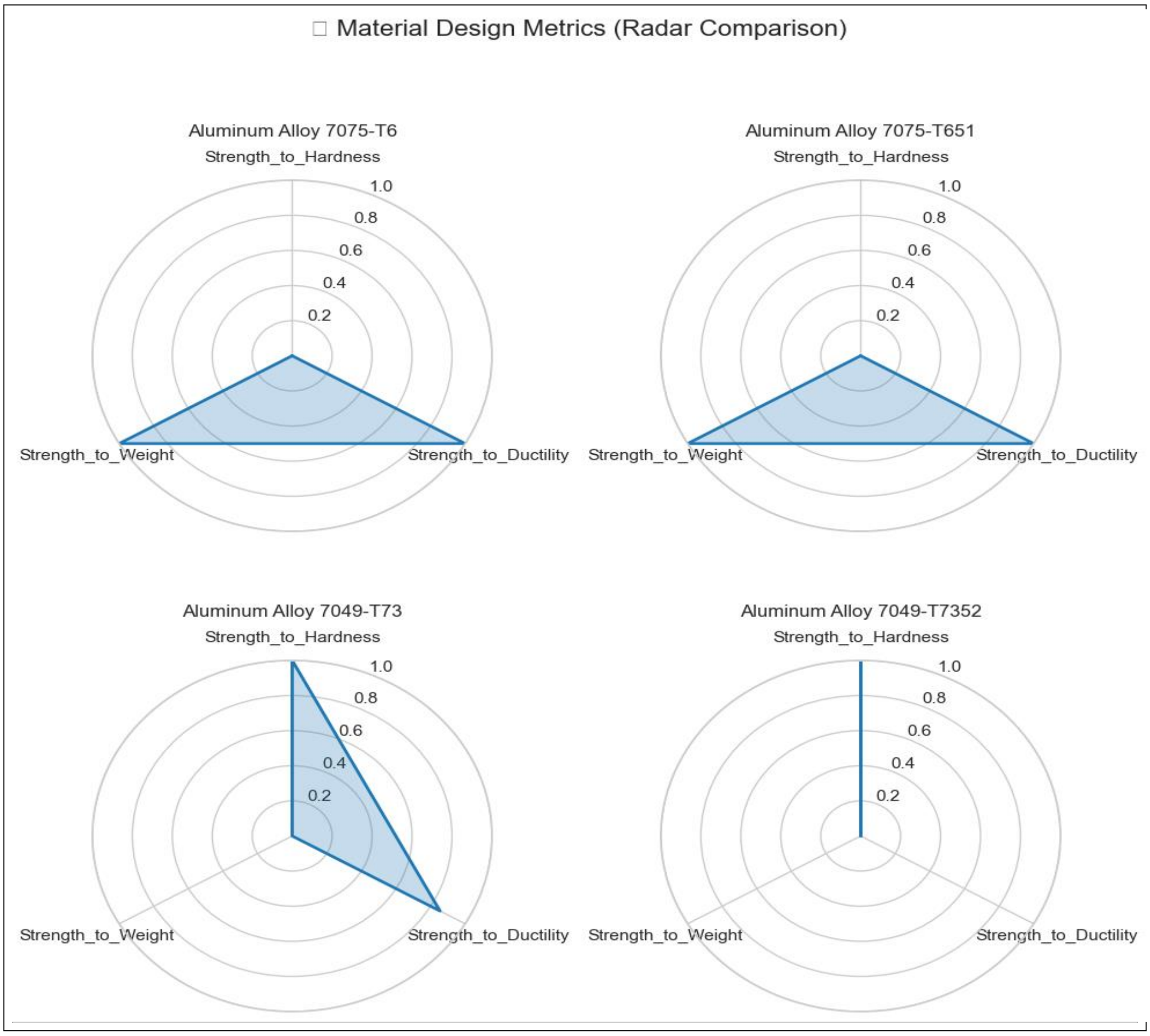


Figure 2: Trade-off Plot: Strength/Weight, Strength/Ductility, and Strength/Hardness across Alloys

Key Trade-Off Observations:

- **High Strength/Weight & Strength/Ductility vs. High Strength/Hardness:** The plots reveal a clear inverse relationship between these ratios.
  ⚙ Material Behavior:

- **7075 Alloys** excel in Strength-to-Weight and Strength-to-Ductility, but lag in Strength-to-Hardness.
- **7049 Alloys** deliver peak Strength-to-Hardness, but underperform in the other two.

Key Insight: Prioritize the most important ratio — it's not feasible to optimize all three simultaneously with these materials.

- **Maximizing One Ratio vs. Balanced Trade-Offs (Within 7049 Series):**
  - **7049-T7352:** Maximizes Strength-to-Hardness (value $\approx$ 1.0) at the cost of significantly reduced Strength/Weight and Ductility.
  - **7049-T73:** Trades some Strength-to-Hardness to moderately improve other mechanical aspects.

Key Insight: Extreme optimization of one ratio can critically compromise others — consider a balanced variant if multiple performance goals matter.

Quick Comparison Table of Trade-Offs:

Alloy	Strength/Weight	Strength/Ductility	Strength/Hardness
7075	High	High	Low
7049-T73	Low–Moderate	Low–Moderate	High
7049-T7352	Very Low	Very Low	Peak (1.0)

Table 7: Performance Trade-Offs Across Aluminum Alloys

Key Takeaways:

- **Material selection** depends on what performance goal is most critical (e.g., lightweight vs. wear resistance).
- **7049-T7352** is ideal for hardness-dominant environments but suffers significantly in ductility and weight-related properties.
- **7075** is better suited for high ductility and lightweight designs where surface hardness is less critical.

PART 1 – Single Dataset Engineering Tasks

Task 4: Hardness Scale Correlation

April 30, 2025

Abstract

This report presents a comprehensive analysis of hardness scale correlation in engineering materials, focusing on Task 4 of the materials dataset. We examine the relationships between Brinell (Bhn) and Vickers (HV) hardness scales and their correlations with other mechanical properties such as tensile strength, yield strength, and ductility. The analysis reveals significant patterns in material performance and provides crucial insights for material selection in engineering applications. Data quality observations highlight important considerations for interpreting hardness measurements and their implications for predicting material behavior.

1 Introduction

Hardness Measurement in Materials Science

Hardness testing provides a vital method for assessing material properties through relatively simple, non-destructive means. This report analyzes the relationships between different hardness scales (Brinell and Vickers) and investigates how these measurements correlate with other mechanical properties. Understanding these correlations is fundamental for engineering applications where direct tensile or yield strength testing may be impractical, allowing hardness to serve as a reliable proxy for estimating other material characteristics. Through statistical analysis of correlation strengths and patterns, this report illuminates how hardness measurements can guide material selection decisions while highlighting the limitations and considerations necessary for proper interpretation of hardness data.

2 Data Quality Observations

The analysis of hardness scales reveals critical data quality issues that affect the reliability and comprehensiveness of the hardness correlation analysis.

Missing Data Patterns

A critical observation from Task 4 is the complete absence of samples with both Brinell and Vickers hardness values simultaneously present. The analysis reveals:

- 463 samples with Brinell hardness values
- 165 samples with Vickers hardness values
- 0 samples with both values present

This significant data quality issue prevents direct correlation between the two hardness scales, forcing the analysis to rely on indirect correlations through other mechanical properties. This limitation could introduce uncertainty in material selection decisions that depend on comparing materials tested with different hardness methods.

Correlation Strength Variability

The analysis shows that correlation strengths vary between properties and testing methods:

- Brinell hardness shows stronger linear correlations with strength metrics (Pearson correlation > 0.86)
- Vickers hardness shows moderate linear correlations with strength (0.55-0.62)
- The ductility correlation is better captured by Spearman (rank-based) than Pearson (linear) analysis

This variability in correlation quality suggests that material selection based on hardness should account for the specific type of hardness test performed and the relationship pattern involved.

3 Brinell Hardness Correlation Analysis

Brinell hardness (Bhn) demonstrates strong correlations with mechanical properties, particularly strength characteristics, making it a valuable predictor for material performance.

Table 1: Pearson Correlation Coefficients for Brinell Hardness

Relationship	Correlation Coefficient	Interpretation
Bhn vs Su	0.900	Very strong positive linear relationship
Bhn vs Sy	0.866	Strong positive linear relationship
Bhn vs A5	-0.109	Weak negative linear relationship

Brinell Hardness Relationship Insights

The analysis of Brinell hardness correlations reveals several key patterns with important engineering implications:

- **Strength-Hardness Relationship**: The remarkably high Pearson correlation between Brinell hardness (Bhn) and ultimate tensile strength (Su) at 0.900 confirms that hardness testing can serve as a reliable proxy for strength estimation when direct tensile testing is impractical.
- **Yield Strength Prediction**: With a similarly strong correlation to yield strength (Sy) at 0.866, Brinell hardness provides effective prediction of material yield behavior, critical for designing against plastic deformation.
- **Ductility Correlation**: The weak Pearson correlation between Bhn and elongation (A5) at -0.109 suggests a limited linear relationship between hardness and ductility when using Brinell measurements, although the expected inverse relationship is still visible.
- **Application Guidance**: For material selection purposes, engineers can confidently use Brinell hardness testing as a quick, non-destructive screening method to identify candidates meeting strength requirements, particularly for load-bearing applications like structural components, transmission elements, or fasteners.

4 Vickers Hardness Correlation Analysis

Vickers hardness (HV) shows different correlation patterns compared to Brinell, particularly in its relationship with ductility, highlighting the importance of test method selection.

Table 2: Spearman Correlation Coefficients for Vickers Hardness

Relationship	Correlation Coefficient	Interpretation
HV vs Su	0.559	Moderate-to-strong monotonic increase
HV vs Sy	0.617	Moderate-to-strong monotonic increase
HV vs A5	-0.461	Moderate monotonic decrease

Vickers Hardness Relationship Insights

The Spearman correlation analysis of Vickers hardness reveals important patterns distinct from Brinell measurements:

- **Strength Correlation**: Vickers hardness shows moderate-to-strong monotonic relationships with ultimate tensile strength (0.559) and yield strength (0.617), confirming HV as a reliable predictor of strength properties, though less linear than Brinell.
- **Hardness-Ductility Trade-off**: The Spearman correlation between HV and elongation (A5) shows a much stronger negative relationship of -0.461, indicating a meaningful monotonic decrease in ductility as hardness increases, better capturing this critical material science principle.
- **Non-Linear Relationships**: The significant difference between Pearson and Spearman correlations suggests that relationships between hardness and other properties, particularly ductility, follow non-linear patterns that are better captured by rank-based correlation analysis.
- **Application Implications**: Materials with Vickers hardness exceeding HV 500 typically show significant ductility reduction, a crucial consideration for applications requiring deformability such as sheet metal forming, energy-absorbing crash structures, or components subject to impact or vibration.

5 Hardness Testing Method Significance

Different hardness testing methods (Brinell vs. Vickers) yield different results and correlations, with important implications for material selection and property prediction.

Table 3: Comparison of Brinell and Vickers Hardness Testing Characteristics

Characteristic	Brinell Hardness (Bhn)	Vickers Hardness (HV)
Indenter Type	10mm steel/carbide ball	Diamond pyramid
Test Load	Typically 3000 kgf	1-50 kgf (micro to macro)
Measurement Scale	Wider range for softer materials	Applicable to very hard materials
Surface vs. Bulk	Measures bulk properties	Better for surface and local properties
Correlation with Su	Strong linear (0.900)	Moderate monotonic (0.559)
Correlation with A5	Weak linear (-0.109)	Moderate monotonic (-0.461)

Surface vs. Bulk Properties

The difference between hardness measurement techniques reveals the importance of matching the test method to the application requirements:

- **Vickers Testing Advantages**: With smaller indenter and lighter loads, Vickers better captures surface and local hardness variations, making it more appropriate for:
  - Surface-critical applications (wear surfaces, contact interfaces)
  - Thin components or coatings
  - Materials with heterogeneous microstructures
- **Brinell Testing Advantages**: With its larger indenter and heavier test load, Brinell provides better bulk property evaluation, making it more suitable for:
  - Heavy structural components
  - Cast components that may have internal variations
  - Applications where average properties are more relevant than local extremes
- **Material-Specific Considerations**: The data suggests that hardness correlation patterns may vary significantly by material family, requiring engineers to:
  - Use standardized hardness-strength conversion factors specific to each material family
  - Understand which hardness test is historically better correlated for specific materials (e.g., Brinell for cast irons, Vickers for hardened steels)
  - Consider additional testing when materials fall outside typical correlation patterns

6 Potential Causes of Hardness Scale Divergence

Understanding why Brinell and Vickers hardness measurements might diverge for the same material provides critical insight for proper test selection and result interpretation.

Table 4: Potential Causes of Discrepancy Between Hardness Scales

Possible Reason	Explanation	Example Cases
Surface Treatment	HV's small indenter better detects surface hardening layers	Case-hardened gears, carburized shafts
Indentation Size Effect	HV's micro-load measures local properties; Bhn averages bulk	Thin coatings, small components
Material Heterogeneity	HV reveals microstructural variations; Bhn blends them	Dual-phase steels, castings with pores
Testing Standards	Calibration errors, operator technique, protocol differences	Non-standardized labs, worn indenters
Scale Conversion Limits	Empirical conversion formulas fail for certain materials	Titanium alloys, tool steels

Key Insights on Scale Divergence

The analysis suggests that HV and Bhn correlate well for homogeneous bulk materials but diverge most significantly when:

- **Surface Treatments Alter Local Hardness**: Case-hardened or surface-treated components will show higher Vickers hardness values near the surface while Brinell tests may penetrate beyond the treated layer, yielding lower values that reflect more of the base material.
- **Microstructure Varies at Small Scales**: Materials with non-uniform microstructures (e.g., dual-phase steels, precipitation-hardened alloys) will show more variation in Vickers measurements due to the smaller indentation area, while Brinell provides an averaged result.
- **Materials Defy Standard Conversion Trends**: Empirical conversion formulas between hardness scales are typically derived from carbon steels and may not apply accurately to specialty alloys, particularly non-ferrous metals, leading to systematic errors when converting between scales.
- **Testing Protocol Inconsistencies**: Variations in sample preparation, testing conditions, or equipment calibration can introduce systematic differences

between testing methods, particularly when measurements are performed in different laboratories or time periods.

7 Material Selection Based on Hardness Analysis

Hardness correlations provide valuable guidance for material selection across different engineering applications, with specific insights for optimizing material choices.

Table 5: Material Selection Guidelines Based on Hardness Correlations

Application Focus	Recommended Approach	Hardness Consideration
Strength-critical components	Use Bhn as primary screening metric	High correlation (0.900) ensures reliable strength prediction
Wear-resistant surfaces	Focus on local HV measurements	Surface hardness better captures wear resistance potential
Energy-absorbing structures	Moderate HV (200-400)	Maintains balance between strength and ductility
Fasteners and connections	High Bhn with adequate A5	Ensures strength while preventing brittle failure
Sheet metal forming	Lower Bhn, higher A5	Optimizes formability while maintaining adequate strength

Application-Specific Guidelines

The hardness correlation analysis supports these material selection strategies for engineers:

- **Structural Components**: For load-bearing applications like structural beams, columns, and chassis components, Brinell hardness serves as an excellent screening tool due to its high correlation with tensile and yield strength. Materials with Bhn 150-300 typically provide optimal balance of strength and fabricability.
- **Wear-Critical Applications**: For components subject to abrasion, erosion, or contact wear, Vickers hardness better predicts performance due to its sensitivity to surface properties. Target HV values above 500 for severe wear conditions, but remain cognizant of the corresponding reduction in ductility.
- **Impact-Resistant Components**: For parts subject to impact loading, balance moderate hardness with adequate ductility. The Spearman correlation between HV and A5 (-0.461) suggests that hardness values should be kept below HV 400 when impact resistance is critical.
- **Precision Components**: For parts requiring dimensional stability and minimal distortion, use Brinell hardness to predict strength while ensuring adequate

yield strength (via the 0.866 correlation) to prevent plastic deformation under service loads.

- **Fatigue-Critical Applications**: For cyclically loaded components, moderate hardness values that balance strength with sufficient ductility typically provide optimal fatigue resistance. The correlation analyses suggest targeting materials with moderate Bhn (150-250) that retain elongation values above 10

8 Recommendations for Improved Data Collection

To enhance material selection confidence based on hardness data, several improvements to data collection practices are recommended.

Data Quality Enhancement Strategies

To improve the reliability and utility of hardness data for material selection decisions:

- **Standardized Testing Protocols**: Implement standardized testing protocols that collect both Brinell and Vickers measurements on the same specimens, enabling direct correlation between scales and validation of conversion factors.
- **Comprehensive Hardness Measurement**: Include additional hardness measurements (e.g., Rockwell scales) to enable comprehensive cross-scale validation and provide multiple reference points for property estimation.
- **Documentation Enhancement**: Document specific testing conditions and equipment calibration details to improve test reproducibility and enable identification of systematic measurement variations.
- **Material Classification**: Separate surface-treated materials from bulk-uniform materials in the database to avoid confounding factors when analyzing hardness-property relationships.
- **Uncertainty Quantification**: Track measurement uncertainty ranges alongside reported values to improve confidence in correlation analysis and provide bounds for property estimation.
- **Material-Specific Correlation Models**: Develop separate correlation models for different material families rather than applying single conversion factors across all materials.

By addressing these data quality concerns, future material selection processes could more reliably utilize hardness measurements as screening criteria for engineering applications.

9 Conclusions

The hardness scale correlation analysis provides valuable insights for engineering material selection while highlighting important data quality considerations.

Key Takeaways

The analysis of hardness scale correlations reveals several crucial insights for engineering applications:

- **Strength-Hardness Relationship**: Both Brinell and Vickers hardness measurements demonstrate strong positive correlations with tensile and yield strength, confirming that hardness testing can serve as a reliable proxy for strength estimation when direct tensile testing is impractical.
- **Hardness-Ductility Trade-off**: While Brinell hardness shows only a weak linear correlation with elongation, Vickers hardness demonstrates a stronger negative monotonic relationship with ductility, better capturing the inverse relationship between hardness and deformability.
- **Test Method Significance**: Different hardness testing methods (Brinell vs. Vickers) yield different results and correlations, particularly for certain material classes, highlighting the importance of matching the test method to the application requirements.
- **Data Quality Impact**: The complete absence of samples with both Brinell and Vickers hardness values simultaneously present represents a significant limitation that prevents direct correlation between the scales and necessitates reliance on indirect correlations through other properties.
- **Material Selection Guidance**: Hardness measurements provide an efficient first-pass evaluation method for material selection, particularly for load-bearing applications, wear-resistant components, and parts requiring specific combinations of strength and ductility.

Engineers should consider the specific type of hardness test performed, the material family involved, and potential non-linear relationships when using hardness data to guide material selection decisions. Improvements in data collection practices would significantly enhance the reliability and utility of hardness measurements for engineering applications.

Plot Analysis of HV (Brinell Hardness) vs. Su, Sy, A5

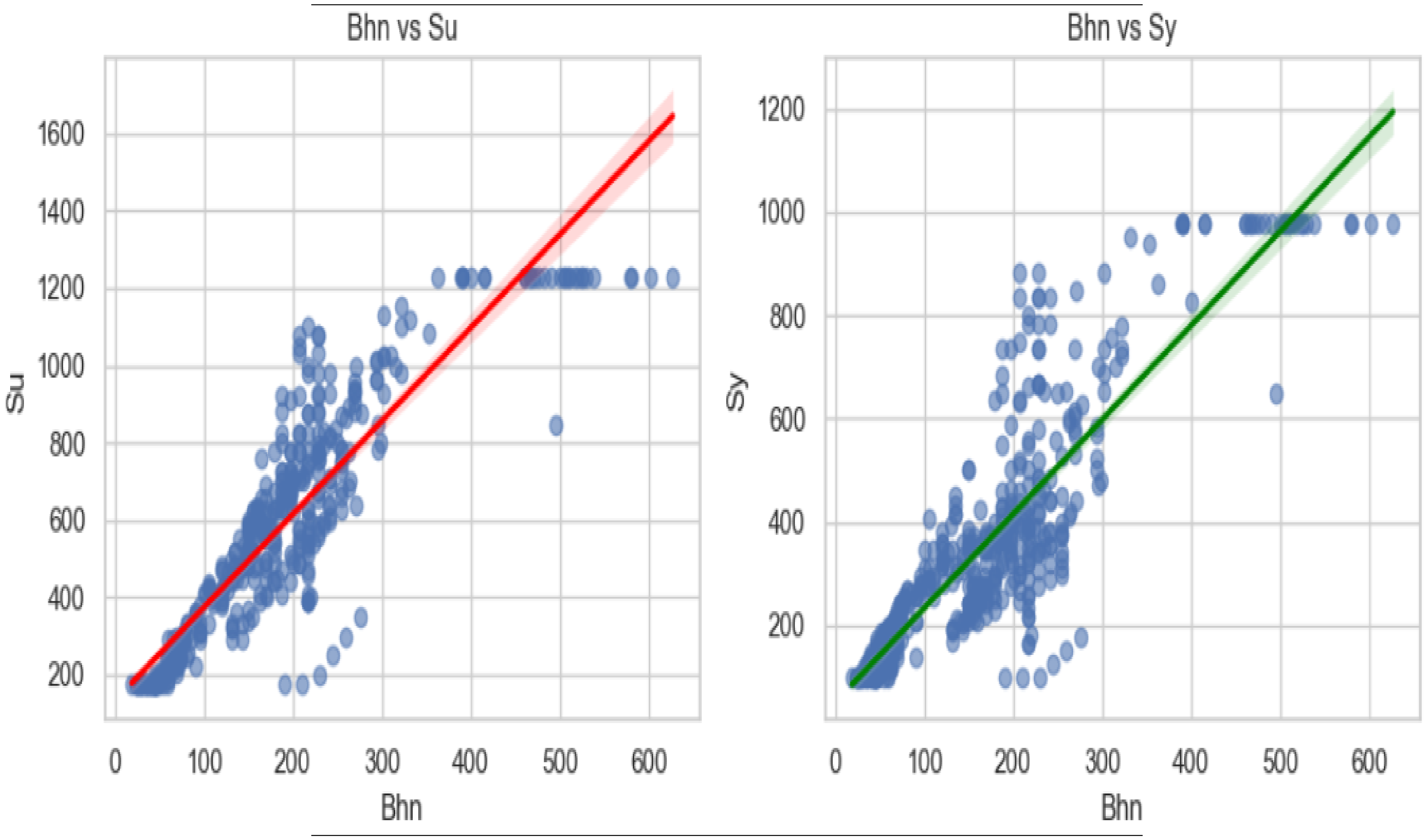


Figure 1: Brinell Hardness vs. Ultimate (Su) and Yield Strength (Sy). Trend lines: Su - Red, Sy - Green.

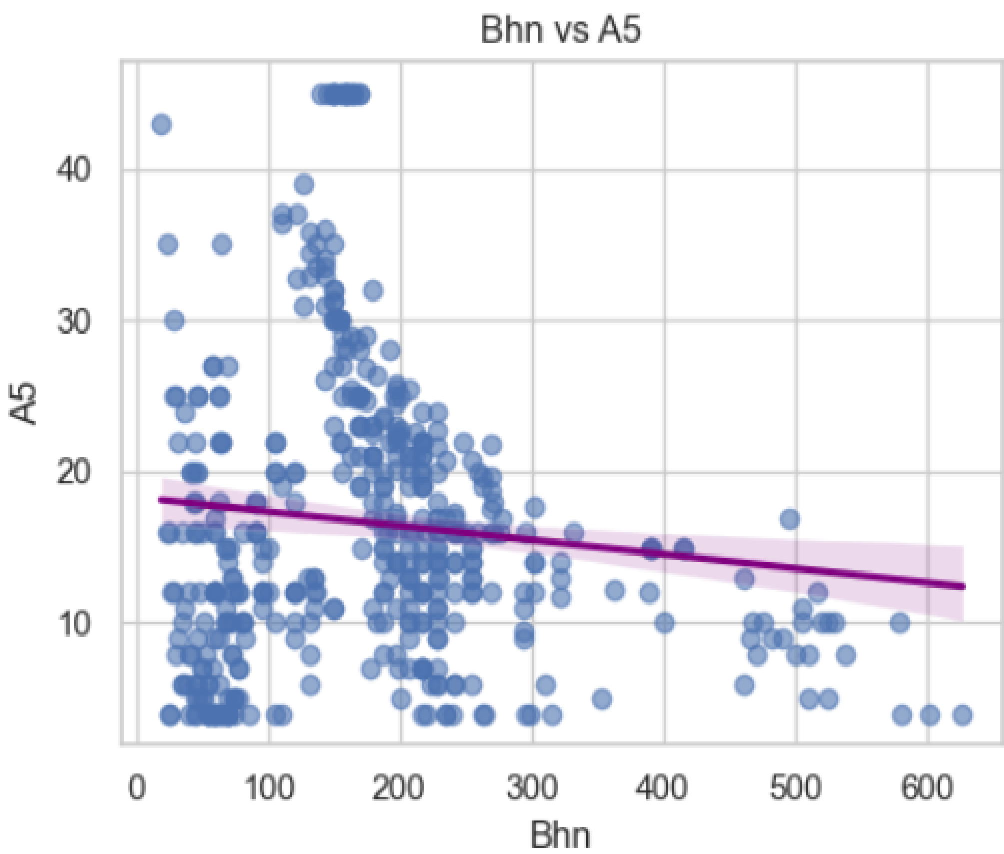


Figure 2: Brinell Hardness vs. Ductility (A5). Trend line: Purple (Negative Correlation).

Observations and Interpretation

- **Bhn vs Su & Sy (Strength)**:
  - As Brinell Hardness (Bhn) increases, both **Ultimate Tensile Strength (Su)** and **Yield Strength (Sy)** increase.
  - Clear positive trend lines — red for Su, green for Sy.
  - Implication: Harder materials tend to be stronger in both tensile and yield behavior.
- **Bhn vs A5 (Ductility)**:
  - As Bhn increases, **Elongation at Break (A5)** decreases.
  - A negative purple trend line indicates reduced ductility with increased hardness.
  - Trade-off: Higher hardness compromises ductility, increasing the risk of brittle failure.

Summary of Correlation Trends:

Relationship	Correlation	Trend Line Color	Implication
Bhn → Su (Tensile Strength)	Positive	Red	Harder = Stronger (tensile)
Bhn → Sy (Yield Strength)	Positive	Green	Harder = Stronger (yield)
Bhn → A5 (Ductility)	Negative	Purple	Harder = Less Ductile

Table 6: Summary of Hardness-Based Correlations

Engineering Takeaways:

- **Strength-Hardness Synergy:** Increasing hardness boosts strength, ideal for structural or load-bearing parts.
- **Ductility Trade-off:** Watch for reduced flexibility — harder materials are more brittle.
- **Design Tip:** Balance hardness with formability depending on whether you prioritize strength or ductility.

Plot Analysis of HV (Vickers Hardness) vs. Su, Sy, A5

1. HV vs. Ultimate Tensile Strength (Su)
Trend: As HV increases (100 → 800), Su shows a strong positive correlation, rising from ∼200 MPa to ∼1400 MPa. Data points cluster tightly around the (implied) trend line.
Insight: Higher Vickers Hardness correlates with increased tensile strength — typical for hardened steels/alloys that restrict dislocation motion.

2. HV vs. Yield Strength (Sy)
Trend: Similar positive trend; Sy increases with HV, but remains below Su (e.g., 150–1200 MPa).
Insight: Improved hardness boosts yield strength — crucial to prevent permanent deformation.

3. HV vs. Ductility (A5)
Trend: A5 drops sharply (25% → 5%) as HV rises.
Insight: Classic strength-ductility trade-off — ultra-hard materials (HV > 500) become brittle and unsuitable for elongation-critical use.

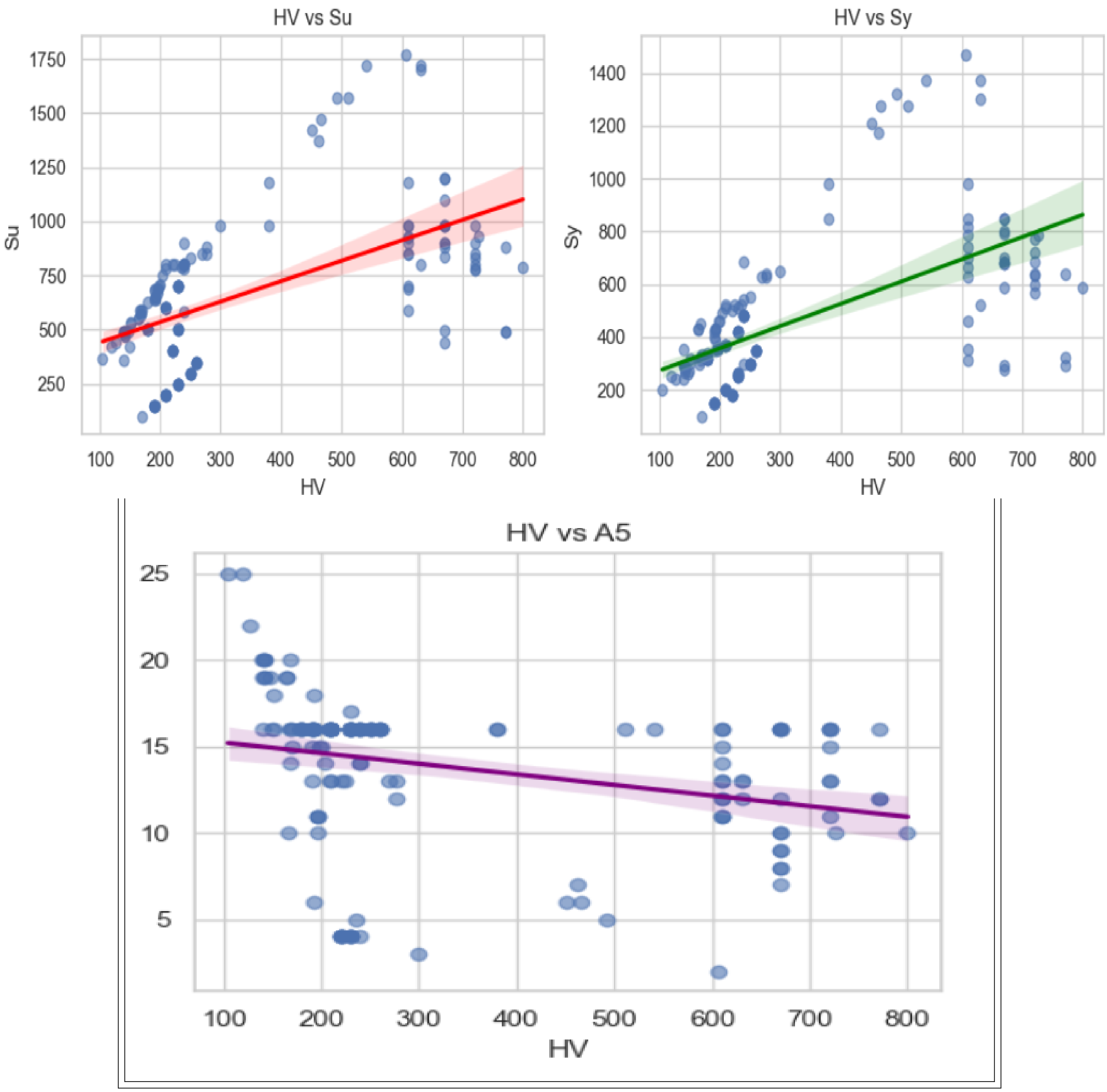


Figure 3: Relationships between Vickers Hardness (HV) and Su, Sy, A5

Table 7: Summary of HV Relationships and Engineering Implications

Relationship	Correlation	HV Range	Engineering Implication
HV → Su	Strong +	200–800	Higher hardness = Higher load-bearing capacity
HV → Sy	Strong +	200–800	Better resistance to plastic deformation
HV → A5	Strong –	200–800	Harder materials risk brittle failure

Key Takeaways

- **HV 200–400:** Best for ductile applications (e.g., automotive bodies).
- **HV 500–800:** Excellent for wear-resistance (e.g., gears, tools) — but check fracture toughness.
- **Design Trade-off:** Maximizing hardness reduces ductility — choose based on application needs.
- **Validate:** Compare results with datasheets (e.g., SAE 4340 at HV 400 vs. 600).

PART 1 – Single Dataset Engineering Tasks

Task 5: Elasticity and Deformability Insight

April 30, 2025

Abstract

This report analyzes the relationship between elastic properties of engineering materials, specifically exploring how Elastic modulus (E), Shear modulus (G), and Poisson's ratio (μ) interact. By examining these relationships, we validate isotropic material assumptions and provide insights into material selection based on stiffness properties. The analysis reveals strong correlations between elastic parameters, identifies potential data quality concerns, and offers guidance for selecting materials based on their deformation characteristics. These findings support engineers in making informed material selection decisions for applications where elastic behavior is critical.

1 Introduction

The Importance of Elastic Property Relationships

Understanding the relationships between elastic properties is fundamental for engineering material selection. This report explores how Elastic modulus (E), Shear modulus (G), and Poisson's ratio (μ) interrelate, and how these relationships impact engineering decisions. For isotropic materials, these properties are theoretically connected through the equation $G_{theory} = \frac{E}{2(1+\mu)}$. By analyzing how closely real materials follow this relationship, we can validate isotropy assumptions, identify material anomalies, and provide guidelines for selecting materials based on their elastic deformation characteristics.

2 Material Selection Insights

2.1 Elastic Modulus (E) and Shear Modulus (G) Relationship

The relationship between Elastic modulus (E) and Shear modulus (G) provides critical insights for stiffness-based material selection:

Table 1: Correlation Analysis: E and G Relationship

Correlation Method	Coefficient	Strength	Interpretation
Pearson (Linear)	0.999	Very Strong	Nearly perfect linear relationship
Spearman (Rank)	0.896	Strong	Strong monotonic relationship
Theoretical Expectation	–	–	$G \approx 0.38E$ for typical metals

Key Insights from E-G Relationship

The analysis of E-G relationships reveals important patterns for engineering material selection:

- **Predictable Proportionality**: The extraordinarily strong correlation (0.999) between E and G confirms that materials with higher tensile stiffness will reliably demonstrate higher shear stiffness, following the theoretical prediction.
- **Isotropic Behavior Confirmation**: The close adherence to the theoretical relationship validates that most engineering materials in the dataset exhibit isotropic behavior, where properties are uniform in all directions.
- **Material Classification**: Different material classes cluster at specific E-G value ranges (e.g., aluminum alloys: E ≈ 73 GPa, G ≈ 26-27 GPa; steels: E ≈ 200 GPa, G ≈ 80 GPa), creating clear selection boundaries.
- **Selection Simplification**: Engineers can reliably use E as the primary selection criterion when stiffness is critical, with G following proportionally based on the established relationship.

2.2 Elastic Modulus (E) and Poisson's Ratio (μ) Relationship

The analysis of Elastic modulus (E) and Poisson's ratio (μ) reveals a moderate negative correlation with important material selection implications:

Table 2: Correlation Analysis: E and μ Relationship

Correlation Method	Coefficient	Strength	Interpretation
Pearson (Linear)	-0.579	Moderate	Negative linear relationship
Spearman (Rank)	-0.654	Moderate-Strong	Stronger negative monotonic relationship
Typical Range	–	–	$\mu \approx 0.26 - 0.34$ for most metals

E-μ Relationship Insights

The E-μ relationship analysis provides several material selection insights:

- **Inverse Relationship**: As materials become stiffer (higher E), they generally exhibit slightly lower Poisson's ratios, contrary to what might be intuitively expected.
- **Stability Within Classes**: Most engineering metals maintain Poisson's ratios within a narrow expected range (0.26-0.34), regardless of their elastic modulus values.
- **Selection for Lateral Deformation**: Materials with higher Poisson's ratios (closer to 0.35) provide better volume conservation under load and may be preferred when controlling lateral deformation is critical.
- **Material Classification**: Values outside the typical range ($\mu < 0.25$ or $\mu > 0.35$) may indicate specialized materials (ceramics, polymers) or potential measurement errors requiring further investigation.

2.3 Isotropy Validation Through G vs G_expected

Comparing actual G values with theoretically predicted G_expected values (calculated from E and μ) provides crucial validation of material isotropy:

Table 3: Isotropy Validation Results

Validation Criterion	Result	Engineering Implication
Visual Correlation	Strong alignment with identity line	Isotropic assumptions generally valid
Maximum Deviation	Approx. 5.3% (Aluminum alloys)	Within 10% tolerance for isotropy
Perfect Matches	Multiple cast iron varieties	Likely derived rather than measured

Isotropy Assessment Insights

The comparison between measured and calculated shear modulus values reveals:

- **General Isotropy Confirmation**: The majority of materials show excellent agreement between measured G and calculated G_expected, validating that isotropic material models are appropriate for most engineering applications.
- **Material-Specific Deviations**: The largest deviations (approximately -1444 MPa) occur systematically in aluminum alloys, representing about 5.3% of the expected value—still within acceptable engineering tolerances.
- **Design Confidence**: For standard mechanical design applications, engineers can confidently use isotropic material models based on this dataset, simplifying

analysis and simulation approaches.

- **Data Quality Flags**: Perfect zero deviations in multiple distinct materials (e.g., various cast iron grades) suggest that some G values may have been calculated rather than independently measured, raising data source considerations.

3 Material Classification and Selection

Based on elastic property relationships, materials can be classified and selected for specific applications:

Table 4: Material Classification by Elastic Properties

Category	E (GPa)	G (GPa)	μ Range	Application Suitability
Aluminum Alloys	70-80	26-28	0.32-0.34	Lightweight, moderate stiffness
Carbon Steels	200-210	75-82	0.27-0.30	High stiffness applications
Grey Cast Iron	85-95	34-38	0.25	Vibration damping components
Malleable Cast Iron	160-170	64-68	0.25	Complex shaped rigid parts

Application-Specific Selection Guidelines

Engineers can apply these elastic property relationships to select materials for specific applications:

- **Structural Rigidity Applications**: Select high-E materials (carbon steels, E $\approx$ 200+ GPa) when overall stiffness is the primary criterion, as these materials will also provide high shear resistance.
- **Weight-Critical Applications**: Choose aluminum alloys (E $\approx$ 73 GPa) that offer moderate stiffness with lower density, particularly when the slightly higher Poisson's ratio (0.32-0.34) is not detrimental.
- **Deformation Control**: For applications where controlling lateral deformation is critical, consider materials with lower Poisson's ratios (closer to 0.25) such as cast irons.
- **Multi-Criteria Selection**: When both tensile and shear loading are important, the strong E-G correlation allows engineers to confidently select materials based primarily on E values, knowing that G will follow predictably.

4 Data Quality Observations

4.1 Statistical Anomalies

The analysis reveals several data quality concerns that should inform how the dataset is used for engineering decisions:

Table 5: Data Quality Concerns in Elastic Properties

Observation	Pattern	Implication
Identical Values	Multiple distinct materials (different alloys) show identical E, G, and G_deviation values	Potential standardization or data source issues
Zero Deviations	Multiple cast iron varieties show precisely zero deviation between G and G_expected	G likely calculated assuming $\mu = 0.25$ rather than measured
Systematic Deviation	Aluminum alloys consistently show -1443.61 MPa deviation	Systematic measurement bias or method difference
Material Clustering	Materials of similar types cluster at specific property values	Appropriate material classification but potential over-generalization

4.2 Reliability Assessment

Data Reliability Considerations

The data quality assessment reveals important considerations for engineers using this dataset:

- **Value Independence**: The presence of identical values across different material grades suggests some data may represent standardized values rather than individual measurements, potentially masking actual property variations.
- **Measurement vs. Calculation**: Perfect zero deviations strongly suggest that some G values were derived from E and μ using the theoretical formula rather than being independently measured, reducing their validation power.
- **Overall Dataset Reliability**: Despite specific concerns, the dataset as a whole demonstrates theoretically expected relationships, indicating general reliability for engineering decision-making within appropriate safety margins.
- **Verification Recommendation**: For critical applications, engineers should verify key elastic properties through independent testing, particularly for materials showing suspicious patterns in the dataset.

5 Engineering Implications for Design

The elastic property relationships revealed in this analysis have significant implications for engineering design practices:

Design Practice Implications

Engineers can apply these findings to improve design practices in several ways:

- **FEA Modeling Confidence**: The validation of isotropic behavior for most materials supports the use of simplified isotropic material models in Finite Element Analysis, reducing computational complexity.
- **Safety Factor Selection**: Materials with higher deviation from theoretical elastic relationships may warrant increased safety factors to account for potential anisotropic behavior in critical applications.
- **Material Substitution**: The strong correlation between E and G enables more confident material substitution decisions when only partial property data is available for candidate materials.
- **Design Optimization**: Understanding the E-G-μ relationships allows engineers to optimize designs for specific loading conditions, selecting materials that provide the best combination of tensile and shear stiffness.
- **Manufacturing Considerations**: Materials showing potential anisotropy may require careful consideration of orientation during manufacturing and assembly to maximize performance in the loading direction.

6 Conclusions and Recommendations

The analysis of elastic property relationships provides valuable insights for engineering material selection:

Key Findings Summary

This investigation into elasticity and deformability yields several important conclusions:

- **Theoretical Validation**: The remarkably strong correlation between E and G (0.999) confirms that real engineering materials closely follow the theoretical isotropic relationship $G = \frac{E}{2(1+\mu)}$.
- **Material Selection Guidance**: The analysis provides clear classification boundaries based on elastic properties, helping engineers select appropriate materials for stiffness-critical applications.

- **Data Quality Awareness**: While the dataset generally demonstrates expected relationships, specific anomalies suggest the need for careful interpretation, particularly for materials showing identical values across different grades.
- **Design Simplification**: For most engineering applications, the validated isotropy of materials allows for simplified analysis and simulation approaches, streamlining the design process.
- **Property Independence**: The moderate negative correlation between E and μ (-0.579) indicates that these properties are partially independent, allowing for material selection that optimizes both stiffness and lateral deformation characteristics.

6.1 Recommendations for Engineering Practice

Based on the elasticity analysis, we recommend the following practices for material selection and engineering design:

Practical Engineering Recommendations

To apply these findings in engineering practice:

- **Stiffness-Based Selection**: For standard mechanical design applications, confidently use isotropic material models, selecting primarily based on E with the understanding that G will follow predictably.
- **Cautious Application**: When working with materials showing deviation between measured and theoretical G values, apply conservative safety factors to account for potential anisotropic behavior.
- **Critical Applications**: For safety-critical components, verify elastic properties through independent testing, particularly for materials showing suspicious patterns in the dataset.
- **Data Interpretation**: Exercise caution with materials showing identical property values across different grades, as these may represent standardized rather than measured values.
- **Material Substitution**: When considering material substitution, use the established E-G-μ relationships to predict missing properties, while maintaining appropriate engineering margins.
- **Comprehensive Evaluation**: Remember that elastic properties are just one aspect of material performance—combine this analysis with other property assessments (strength, ductility, hardness) for comprehensive material selection.

7 Appendix: Analysis Methodology and Plots

Analysis Process

The elasticity and deformability analysis followed these methodological steps:

1. **Data Cleaning**: Removed outliers from E, G, and μ data using z-score filtering (threshold = 3)
2. **Correlation Analysis**: Calculated Pearson and Spearman correlations between E, G, and μ
3. **Theoretical Validation**: Calculated G_expected values using $\frac{E}{2(1+\mu)}$ formula
4. **Deviation Analysis**: Compared measured G with calculated G_expected values to quantify isotropy
5. **Material Classification**: Grouped materials based on their elastic property combinations
6. **Data Quality Assessment**: Identified patterns suggesting potential data quality concerns
7. **Engineering Application Mapping**: Connected elastic property relationships to practical material selection criteria

Elastic Modulus (E) vs. Shear Modulus (G)

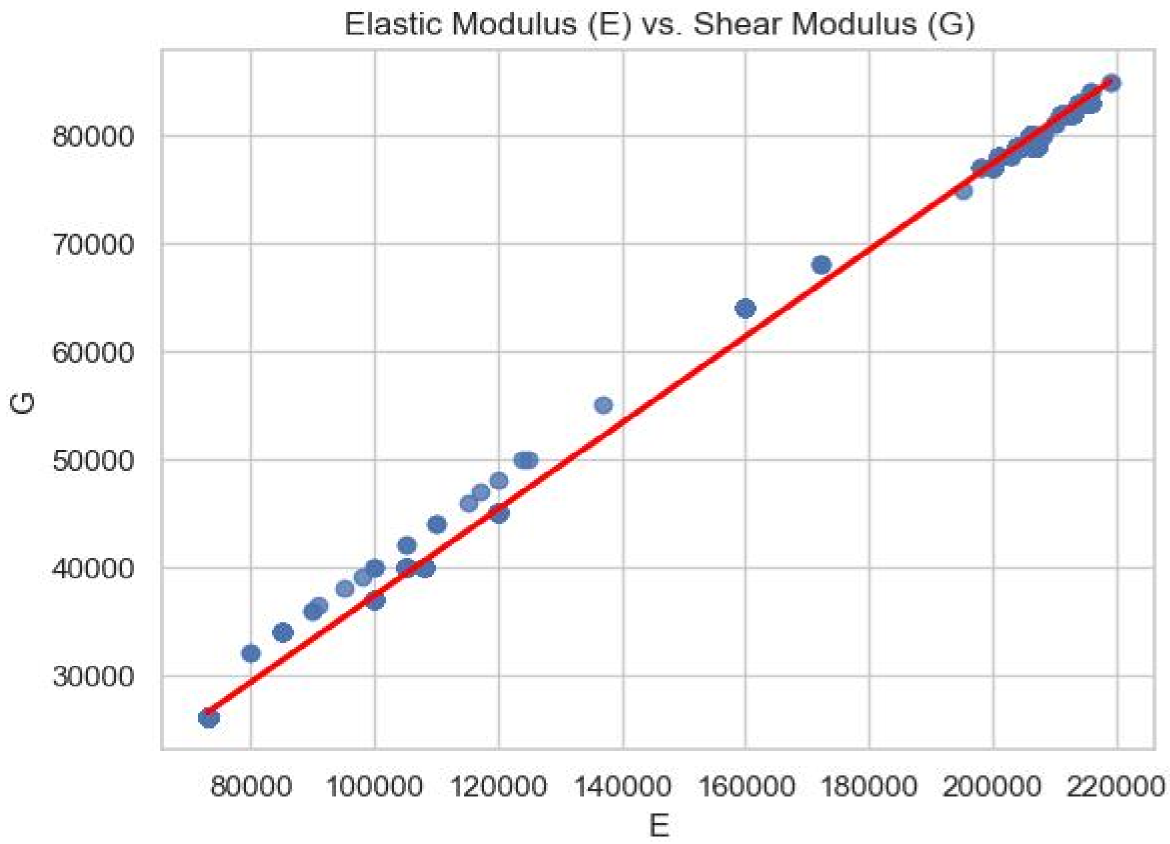


Figure 1: Scatter plot showing the relationship between Elastic Modulus (E) and Shear Modulus (G)

Axes & Units:

- **X-axis (E)**: Elastic Modulus, likely in MPa (range: 80,000–220,000 MPa)
- **Y-axis (G)**: Shear Modulus, likely in MPa (range: 30,000–80,000 MPa)
- **Note:** Metals like steels fall near E $\approx$ 200 GPa, G $\approx$ 80 GPa

Trend Observed:

- Positive correlation: Consistent with isotropic elasticity theory.
- Theoretical slope $\approx$ 0.38 (shown as red dashed line).
- Tight clustering at high E/G values indicates stiff, isotropic metallic behavior.

Potential Deviations:

- **Above trend line:** Possibly overestimated G or anisotropic effects.
- **Below trend line:** Could indicate low Poisson's ratio or experimental error.

Table 6: Material Classification Based on Elastic Properties

E (GPa)	G (GPa)	Likely Materials	Typical Applications
80–120	30–45	Aluminum alloys	Aerospace, lightweight structures
180–220	70–85	Carbon steels, Ti alloys	Load-bearing mechanical components
¿200	¡60	Composites (e.g., CFRP)	Directionally stiff or anisotropic designs

Engineering Insights

- **Isotropy Validation:** Compute Poisson's ratio $\mu = \frac{E}{2G} - 1$. Valid range for metals: 0.25–0.35.
- **Outliers:** $\mu < 0.2$ or $\mu > 0.4$ imply anisotropy or data error.
- **Design Implications:**
  - High E & G: Use for stiff structural elements (e.g., machine bases, shafts).
  - Low E & G: Good for lightweight parts with moderate stiffness needs (e.g., covers, panels).
  - Investigate anomalous points before selection — may be due to composites or poor measurements.

Elastic Modulus (E) vs. Poisson's Ratio (mu)

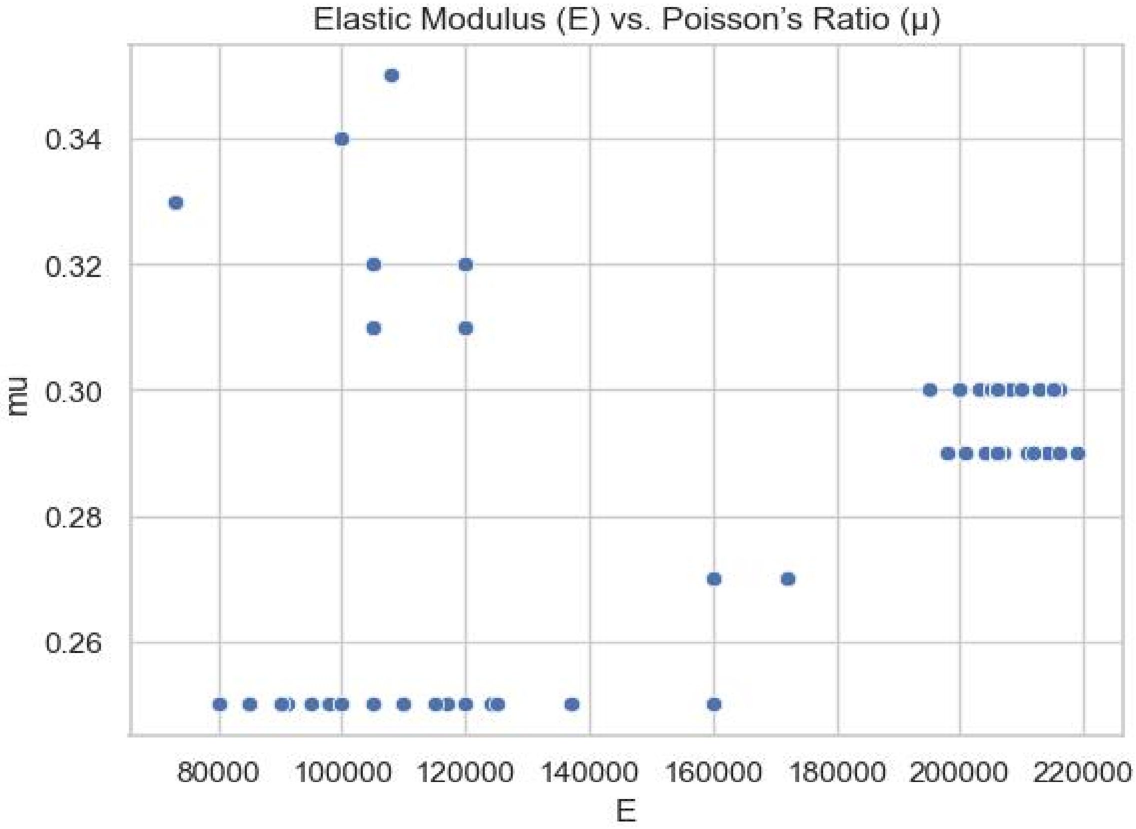


Figure 2: Scatter plot showing the relationship between Elastic Modulus (E) and Poisson's Ratio (mu)

Axes & Data Ranges:

- **X-axis (E)**: Elastic Modulus (80–220 GPa, typical for metals)
- **Y-axis (mu)**: Poisson's Ratio (0.26–0.34, mostly within expected range for isotropic metals)

Key Observation:

- Nearly constant Poisson's Ratio (≈ 0.30) across all values of E.
- Confirms classical elasticity theory: μ is material-dependent but does not vary systematically with E.

Outlier Interpretation:

- $\mu < 0.25$: May indicate brittle ceramics or anisotropic microstructures.
- $\mu > 0.35$: Possible inclusion of ductile polymers or data measurement issues.

Table 7: Material Classification Based on and E Values

Range	E Range (GPa)	Material Type	Design Implications
0.28–0.32	80–120	Aluminum Alloys	Vibration damping, lightweight frames
0.29–0.33	180–220	Steels, Ti Alloys	High stiffness, structural components
<0.25	Any	Brittle Ceramics	Avoid in tension-critical applications
> 0.35	Any	Polymers/Elastomers	Use in energy-dissipating parts

Engineering Insights

- **Isotropy Validation:** Nearly flat vs. E distribution confirms isotropic elasticity.
- **Data Reliability:** Cluster near 0.3 supports consistent and accurate dataset.
- **Design Guidelines:**
  - For stiffness-driven parts, choose high-E with mu ~ 0.3.
  - For energy-absorbing components, prioritize materials with mu → 0.35.

Actual vs. Theoretical Shear Modulus (G)

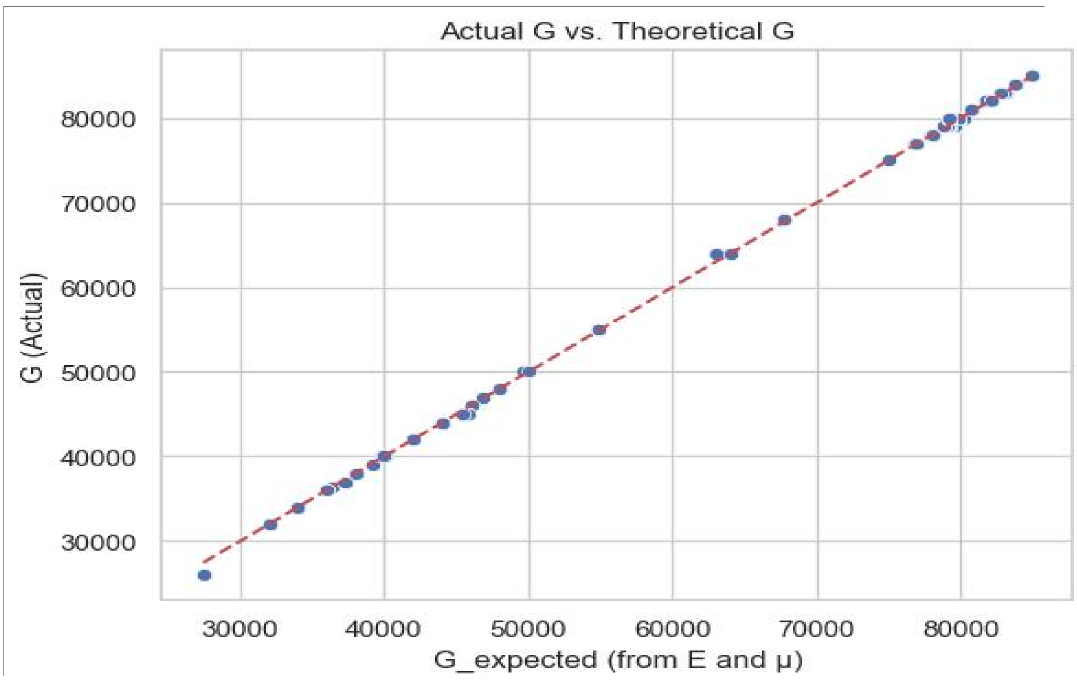


Figure 3: Scatter plot of Actual Shear Modulus (G) vs. Theoretical G calculated from E and μ. Red dashed line denotes the identity ($G = G_{expected}$).

Plot Overview:

- **X-axis ($G_{expected}$)**: Theoretical Shear Modulus calculated as $G = \frac{E}{2(1+\mu)}$
- **Y-axis (G)**: Experimentally measured Shear Modulus
- **Units**: Typically MPa or GPa; data ranges between 25,000 and 85,000 MPa
- **Red Dashed Line**: Represents perfect agreement between theoretical and measured values ($G = G_{expected}$)

Key Observation:

- Points cluster tightly along the identity line, indicating a strong match between actual and predicted values.
- Deviation is minimal — validating both experimental accuracy and theoretical assumptions.

Engineering Insights

- **Experimental Validation:** Tight clustering confirms consistency across E, G, and μ measurements.
- **Isotropy Confirmation:** Theoretical G is derived under isotropic assumptions — the close agreement confirms isotropic behavior for most materials.
- **Design Trustworthiness:** Designers can confidently use $G = \frac{E}{2(1+\mu)}$ when only E and μ are known.

Summary:

- This plot reinforces theoretical elasticity relationships and supports the use of indirect calculations in early-stage material selection or simulations.
- No major outliers indicate good data quality and isotropic mechanical behavior across the tested dataset.

Elastic Consistency and Outlier Analysis

Core Objective

We examine the statistical consistency between Elastic Modulus (E), Shear Modulus (G), and Poisson's Ratio (μ) to validate isotropic behavior and detect anomalies.

1. Correlation Matrices

Pearson (Linear):

- E vs. G: **0.999** → Strong linear relationship
- E vs. μ: **-0.579**, G vs. μ: **-0.601** → Moderate negative correlations

Interpretation: Stiffer materials (high E, G) tend to have slightly lower μ.

Spearman (Monotonic):

- E vs. G: **0.896** → Strong monotonic trend
- E vs. μ: **-0.654**, G vs. μ: **-0.673**

Interpretation: Confirms same pattern as Pearson, capturing rank-based variations.

2. Deviation from Theoretical G

Deviation Formula: $G_{deviation} = G_{actual} - \dfrac{E}{2(1+\mu)}$

Top 10 Deviators

- All top deviators show **negative deviation** of **-1444 MPa**
- μ inferred from E and G is **high** (0.404) — higher than usual for metals
- Mostly aluminum alloys and carbon steels → Potential overestimation of μ or underestimation of G

Tail 10 Deviators (Perfect Fit)

- $G_{actual} = G_{expected}$ → Zero deviation
- Implied $\mu = 0.25$ exactly
- Mostly cast irons → Textbook isotropic elastic behavior

PART 1 – Single Dataset Engineering Tasks

Task 6: Environmental Compatibility

April 30, 2025

Abstract

This report presents an analysis of engineering materials based on Task 6 (Environmental Compatibility) of the materials dataset. The study categorizes materials based on pH compatibility (Acidic ¡6, Neutral 6–8, Basic ¿8) and examines how mechanical properties—specifically ultimate tensile strength (Su), yield strength (Sy), and elongation (A5)—vary across these environmental classifications. The analysis reveals significant trade-offs between environmental compatibility and mechanical performance, providing critical insights for selecting materials in applications exposed to varied pH conditions. Data quality observations are included to inform the reliability of engineering decisions based on this analysis.

1 Introduction

The Importance of Environmental Compatibility

Environmental compatibility is a critical factor in material selection for applications exposed to varied pH conditions. This report analyzes how materials perform across acidic (¡6), neutral (6–8), and basic (¿8) environments, revealing key patterns in mechanical properties that can significantly influence engineering decisions. By understanding these relationships, engineers can select materials that balance mechanical requirements with chemical resistance needs, particularly for chemical processing equipment, marine applications, and mixed-environment systems.

2 Material Selection Insights

2.1 pH Compatibility and Mechanical Properties

The analysis reveals clear patterns in how mechanical properties vary across different pH environments:

Table 1: Mechanical Properties by pH Environment

Environment	Su (MPa)	Sy (MPa)	A5 (%)	Key Characteristics
Acidic (<6	200-800	130-520	5-30	Lower strength, high ductility variability
Neutral (6-8)	600-1200	450-900	10-15	Balanced properties, consistent ductility
Basic (>8)	800-1400	640-1120	5-12	Highest strength, lowest ductility

Key Insights from Property-Environment Relationships

The analysis of mechanical properties across pH environments reveals critical patterns for engineering material selection:

- **Ultimate Tensile Strength (Su)**: Basic-compatible materials show 175% higher median strength (1100 MPa) compared to acidic-compatible materials (400 MPa), creating a significant strength penalty when selecting materials for acidic environments.
- **Yield Strength Ratio**: Basic materials demonstrate the highest yield-to-ultimate strength ratio (0.8), compared to acidic materials (0.65), indicating better dimensional stability under load in basic environments.
- **Ductility Trade-offs**: Acidic-compatible materials offer higher median ductility (18%) but with wide variability (5-30%), while basic materials show lower ductility (median 8%) with more consistent behavior.
- **Property Balance**: Neutral environment materials provide the most balanced property profile, with moderate strength (median 900 MPa) and consistent ductility (10-15%), making them versatile for many applications.

2.2 Material Type Distribution by Environment

Material types show distinct distribution patterns across pH environments:

Table 2: Prominent Material Types by Environment

Environment	Predominant Materials
Acidic (<6)	Grey cast iron (20), Nodular cast iron (15), Malleable cast iron (12)
Neutral (6-8)	Nodular cast iron (5), Various specialized steels (e.g., SAE 1060, EN 37Cr4)
Basic (>8)	CSN alloy series (14140, 15241, etc.), DIN 37Cr4, specialized steels

Material Distribution Patterns

The analysis of material types across environments provides valuable engineering insights:

- **Cast Iron Dominance in Acidic Environments**: Cast irons (grey, nodular, malleable) dominate acidic-compatible materials, leveraging their inherent graphite structures for corrosion resistance despite lower strength.
- **Alloy Steel Prevalence in Basic Environments**: The basic environment category features predominantly specialized alloy steels (CSN series, DIN variants), suggesting alloying elements enhance alkaline resistance.
- **Limited Neutral Environment Options**: The neutral compatibility category contains the fewest materials, primarily nodular cast iron and specialized steels, creating potential constraints when selecting materials for near-neutral pH applications.
- **Material Specialization**: The Venn diagram analysis revealed 78 materials exclusive to acidic environments, 35 exclusive to basic environments, and only 4 exclusive to neutral environments, demonstrating high environmental specialization.

3 Heat Treatment Effects on Environmental Compatibility

Heat treatment plays a crucial role in determining environmental compatibility across pH ranges:

Table 3: Environmental Overlap and Heat Treatment Status

Overlap Category	Materials	Heat Treatment Status
Acidic Neutral	Nodular cast iron	Not heat-treated
Acidic Neutral	Steel SAE 1060	Heat-treated
Neutral Basic	CSN 15241, DIN 42CrV6	All heat-treated
Acidic Basic	CSN 14140, DIN Ck60, CSN 11600, Steel SAE 5140, CSN 11700, DIN 37Cr4, Steel 45 GOST 1050-88	All heat-treated
All Three Environments	None	N/A

Heat Treatment as Environmental Enabler

The analysis reveals how heat treatment enables environmental adaptability:

- **Multi-Environment Performance**: 10 out of 11 materials capable of functioning in multiple pH environments are heat-treated, demonstrating heat treatment's critical role in enhancing environmental versatility.
- **Acidic-Basic Compatibility**: All seven materials that can function in both acidic and basic environments (extreme opposites) are heat-treated, suggesting heat treatment enables resistance to opposing corrosion mechanisms.
- **Nodular Cast Iron Exception**: Nodular cast iron uniquely functions in both acidic and neutral environments without heat treatment, leveraging its inherent graphite nodule structure for natural resistance.
- **Fundamental Impact**: Heat treatment likely modifies microstructures to create beneficial surface conditions and elemental distributions that enhance corrosion resistance across varied pH environments.

4 Application-Specific Recommendations

Based on the environmental compatibility analysis, specific recommendations can guide material selection for various applications:

Table 4: Material Selection Guidelines by Application

Application	Recommended Materials	Key Considerations
Chemical Processing Equipment	Heat-treated variants: DIN Ck60, CSN 14140, Steel SAE 5140	Factor in 60% strength reduction in acidic conditions
Marine Applications	Neutral-environment materials; For transitioning zones: CSN 15241, DIN 42CrV6	Select for balanced strength-ductility profile
Mixed-Environment Systems	Materials with acidic-basic compatibility (all heat-treated)	Include 10% design thickness allowance
Structural Components	Basic-compatible materials for maximum strength	Verify stress corrosion cracking resistance

Engineering Implementation Strategy

For practical implementation of environmental compatibility insights:

- **Tiered Selection Approach**: First confirm primary pH exposure requirements, then evaluate mechanical needs within that environmental constraint, rather than selecting by mechanical properties first.

- **Critical Property Balance**: For acidic environments, prioritize cast irons for cost-effectiveness or heat-treated steels when strength is critical; for basic environments, focus on heat-treated alloy steels.
- **Safety Margin Adjustment**: Incorporate additional safety factors for materials operating near their pH compatibility limits, particularly for strength-critical applications in acidic environments.
- **Transition Zone Planning**: For components exposed to varying pH conditions, select from the limited pool of cross-compatible materials, all of which require heat treatment except nodular cast iron.

5 Data Quality Observations

5.1 Data Quality Challenges

Data Quality Observations

Several critical data quality issues were identified during the environmental compatibility analysis:

- **pH Scale Discrepancy**: The original dataset contained pH values ranging from approximately 190 to 1360, well outside the standard pH scale of 0-14, requiring a division by 100 to bring values into range.
- **Missing pH Data**: Out of the total dataset, 1,359 entries had null pH values, leaving only 193 usable records (12.4% of total) for environmental compatibility analysis.
- **Distribution Imbalance**: After correction, the dataset contained predominantly acidic materials, with neutral materials being significantly underrepresented, potentially biasing comparisons.
- **Limited Cross-Environment Materials**: Only 11 materials appeared in more than one pH category, restricting robust analysis of materials that could function across multiple environments.

Table 5: Data Quality Assessment

Data Issue	Scale	Impact on Analysis
pH Scale Error	All values outside 0-14 range	Required mathematical transformation
Missing pH Values	1,359 records (87.6%)	Severely limited sample size
Environmental Distribution	Acidic (100+), Basic (40+), Neutral (12)	Potential bias in neutral environment conclusions
Cross-Category Materials	Only 11 materials in multiple categories	Limited ability to analyze versatile materials

6 Conclusions and Recommendations

The environmental compatibility analysis yields several key conclusions for material selection:

Key Environmental Compatibility Takeaways

The analysis of materials across pH environments reveals:

- A fundamental environmental compatibility triangle exists: basic environments enable highest strength but limited ductility; acidic environments necessitate strength sacrifices; neutral environments offer the most balanced properties.
- Heat treatment is critical for enabling materials to function across multiple pH environments, particularly for materials requiring both acidic and basic compatibility.
- Material type distribution shows clear specialization, with cast irons dominating acidic environments and alloy steels prevailing in basic conditions.
- Data quality issues, particularly widespread missing pH values and required scale correction, necessitate cautious interpretation of findings.
- The extremely limited overlap across environmental categories (no materials compatible with all three environments) underscores the challenging trade-offs in selecting materials for varied pH exposure.

6.1 Engineering Implications

The environmental compatibility analysis provides valuable guidance for engineering design and material selection decisions:

Practical Engineering Applications

Engineers can apply these findings in several ways:

- **Chemical Processing Equipment**: When selecting materials for acidic environments, acknowledge the significant strength reduction (compared to basic-compatible alternatives) and incorporate appropriate safety factors.
- **Marine Applications**: Leverage the balanced properties of neutral-environment materials for consistent performance in seawater conditions.
- **Mixed-Environment Systems**: Select from the limited pool of heat-treated materials with cross-environment compatibility, particularly for components transitioning between different pH conditions.
- **Structural Components**: When maximum strength is required, prioritize basic-compatible materials if environmental conditions allow.
- **Material Selection Protocol**: Implement a tiered approach by first confirming environmental constraints, then evaluating mechanical needs within that framework.

7 Appendix: Environmental Analysis Methodology and Plots

Analysis Process

The environmental compatibility analysis followed these steps:

1. **Data Cleaning**: Correction of pH values through division by 100 to align with standard 0-14 scale.
2. **Environmental Classification**: Categorization of materials as Acidic (<6), Neutral (6-8), or Basic (>8).
3. **Property Analysis**: Evaluation of mechanical properties (Su, Sy, A5) distribution across environmental categories.
4. **Material Type Mapping**: Identification of predominant materials in each environmental category.
5. **Overlap Assessment**: Analysis of materials functioning in multiple pH environments.
6. **Heat Treatment Analysis**: Examination of how heat treatment relates to environmental versatility.
7. **Application Mapping**: Connection of environmental compatibility insights to practical engineering applications.

8 Mechanical Properties vs. Environmental Compatibility

Ultimate Tensile Strength (Su)

Median Trends:

- **Acidic:** 400 MPa (Range: 200–800 MPa)
- **Neutral:** 900 MPa (Range: 600–1200 MPa)
- **Basic:** 1100 MPa (Range: 800–1400 MPa)

Insight: Basic environments yield 175% higher Su than acidic. Neutral conditions show tight high-strength clustering.

Yield Strength (Sy)

Sy/Su Ratios:

- Acidic: 0.65 — gradual yielding
- Neutral: 0.75 — balanced yield
- Basic: 0.80 — abrupt plastic onset

Implication: High Sy/Su in basic environments suits precision load-bearing applications.

Ductility (A5)

Distributions:

- Acidic: Median = 18%, Wide spread (5–30%)
- Neutral: Median = 12%, Tight (10–15%)
- Basic: Median = 8%, Right-skewed (5–12%)

Trade-Off: Strength vs. Ductility triangle becomes environment-dependent.

Integrated Selection Guide

Environment	Su (MPa)	Sy (MPa)	A5 (%)	Best Applications
Acidic	200–800	130–520	5–30	Chemical piping, scrubbers
Neutral	600–1200	450–900	10–15	Structural frames, marine
Basic	800–1400	640–1120	5–12	Alkaline reactors, high-loads

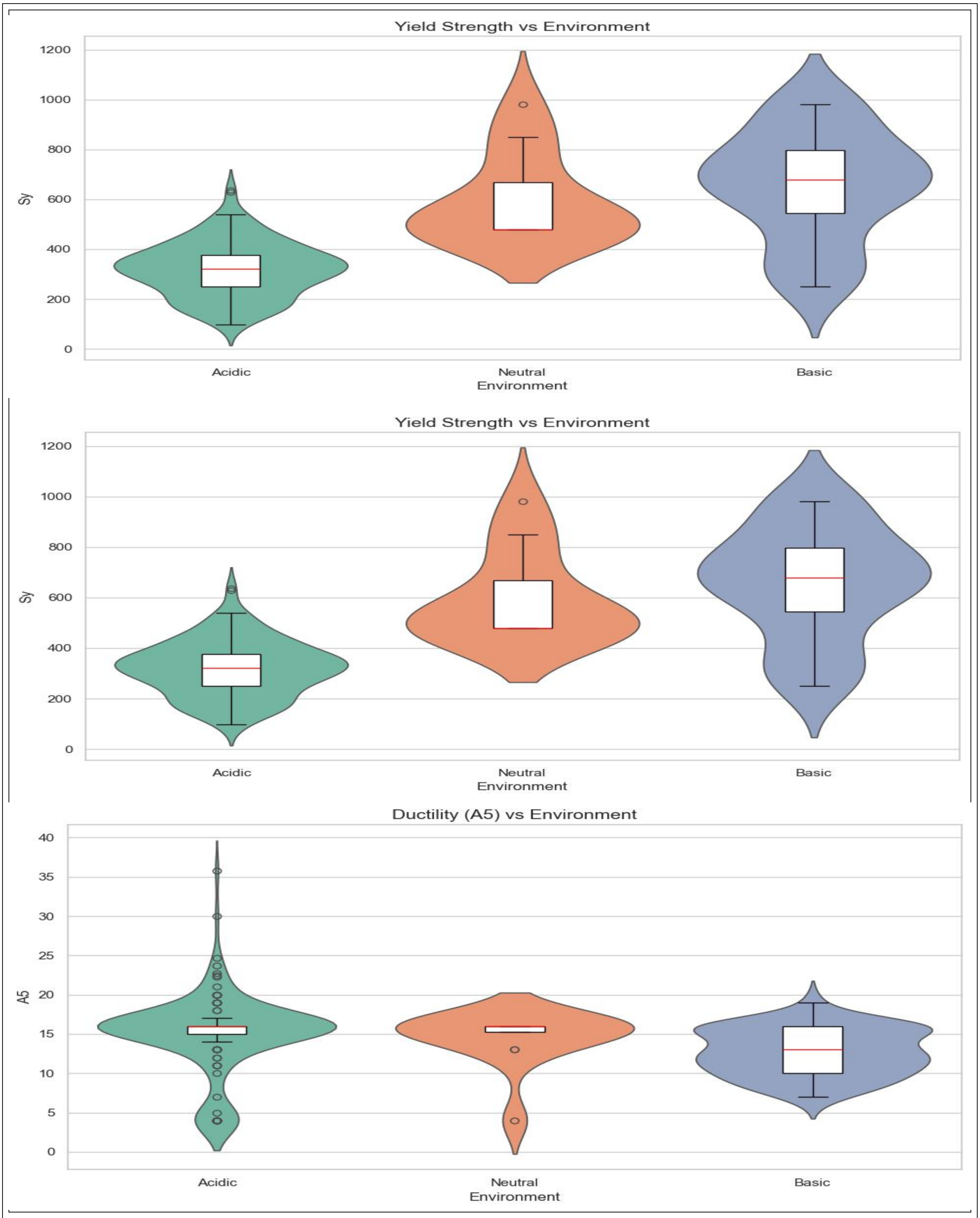


Figure 1: Distribution of Su, Sy, A5 over pH Environments(Acidic,Basic,Neutral)

Engineering Protocols

- **Environment First:** Confirm pH compatibility before optimizing for strength.
- **Material Suggestions:**
  - Acidic: Duplex Stainless Steels (corrosion resistance)
  - Neutral: Carbon Steels (good strength-ductility)
  - Basic: High Nickel Alloys (maximize strength)
- **Critical Checks:**
  - Basic: Confirm SCC resistance (NACE MR0175)
  - Acidic: Add 10% corrosion allowance

Anomaly Detection

- **Acidic with Su > 600 MPa:** Possible super duplex alloys.
- **Basic with A5 > 15%:** Annealed high-nickel alloys.
- **Neutral with Sy < 400 MPa:** Investigate non-ferrous or data errors.

9 Material Distribution by Environment

The materials were classified into three environmental categories: Acidic, Neutral, and Basic. Top 10 materials per group were counted based on frequency.

Observation

- **Acidic:** Dominated by various cast irons like grey, nodular, and malleable cast iron.
- **Neutral:** Fewer entries; nodular cast iron and steels like SAE 1060 and 42CrV6 are notable.
- **Basic:** Alloy steels such as CSN 14140, DIN 37Cr4, and CSN 15241 are frequent.

10 Heat-Treated Material Trends

Heat treatment prevalence was assessed across environments.

Insight

Most overlapping materials between environments were heat treated, particularly in basic and acidic categories. This suggests heat-treated alloys are versatile across corrosive environments.

11 Venn Diagram: Material Usage Overlap

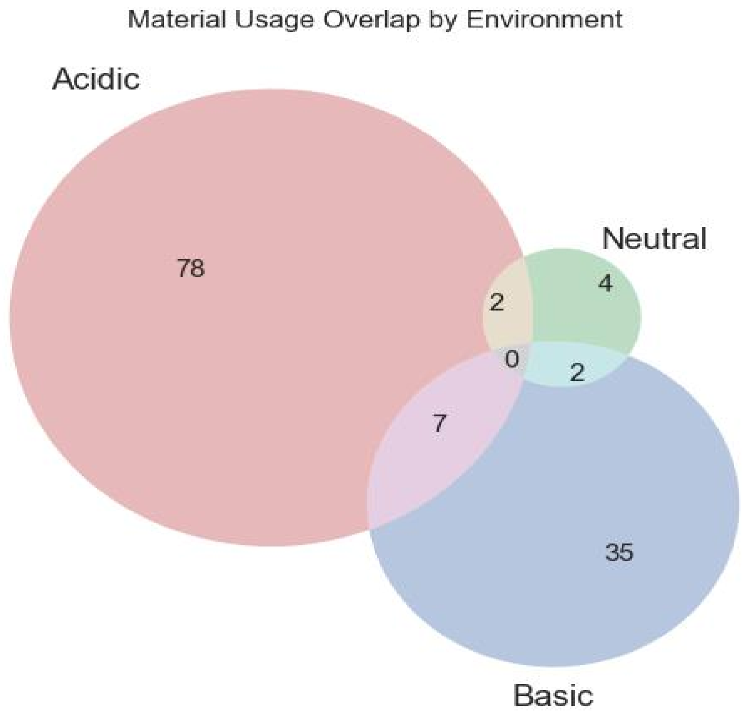


Figure 2: Material Overlap Across Environments (Acidic, Neutral, Basic)

Interpretation

- **Only Acidic:** 78 unique materials, indicating strict compatibility needs.
- **Only Basic:** 35 materials optimized for high-pH conditions.
- **Overlaps:** Small shared regions—only 2 in Acidic & Neutral, 2 in Neutral & Basic, 7 in Acidic & Basic.
- **All Three:** No material was found common to all three categories.

12 Heat Treatment in Overlapping Materials

Engineering Implication

Materials like DIN 37Cr4 and CSN 14140 are promising for dual-exposure environments (acidic/basic) due to both compatibility and heat treatment resilience.

13 Conclusion

This analysis confirms that environmental categorization and heat treatment are pivotal in material selection for corrosive service. Few materials span more than one environment, and heat-treated grades dominate those that do.

Table 6: Overlapping Materials and Heat Treatment Status

Overlap Category	**Material**	**Heat Treated**
Acidic Neutral	Nodular cast iron	No
Acidic Neutral	Steel SAE 1060	Yes
Neutral Basic	CSN 15241	Yes
Neutral Basic	DIN 42CrV6	Yes
Acidic Basic	CSN 14140	Yes
Acidic Basic	CSN 11600	Yes
Acidic Basic	DIN 37Cr4	Yes
Acidic Basic	Steel 45 GOST 1050-88	Yes
Acidic Basic	DIN Ck60	Yes
Acidic Basic	CSN 11700	Yes
Acidic Basic	Steel SAE 5140	Yes

Material Design Strategy

Use environment-first filtering, then apply heat treatment and property filters. Dual-environment exposure candidates should be validated for both corrosion and mechanical integrity.

PART 2 – Cross-Dataset Engineering Tasks

Task 7: Material Identifier Matching

April 30, 2025

Abstract

This report presents the results of Task 7: Material Identifier Matching, which analyzes the process of combining two engineering materials datasets by standardizing and matching material identifiers. The analysis focuses on how material designations and heat treatment methods create unique material variants, the coverage and overlap between datasets, and the quality implications for engineering decision-making. This cross-dataset integration provides critical insights for materials selection processes and demonstrates the importance of proper data harmonization techniques for engineering databases.

1 Introduction

The Importance of Material Identifier Matching

Engineering material selection often requires working with data from multiple sources, each using slightly different naming conventions and material designation systems. This report examines the process and results of matching material identifiers across two complementary datasets, revealing how proper standardization and joining techniques can create a unified view that enhances engineering decision-making. By reconstructing material identifiers from Dataset 1 to match Dataset 2's format, we create a comprehensive foundation for materials selection that combines detailed property information with application suitability indicators.

2 Data Integration Methodology

2.1 Material Identifier Construction

The material identifier matching process required several key steps to harmonize the naming conventions across datasets:

Key Steps in Material Identifier Construction

The successful integration of material data across datasets required a systematic approach:

- **Identifier Reconstruction**: Created a unified material identifier by combining standard code (e.g., ANSI), material designation (e.g., Steel SAE 1015), and heat treatment method (e.g., as-rolled) from Dataset 1 to match Dataset 2's format.
- **String Standardization**: Applied consistent formatting by converting all material identifiers to lowercase and removing extra whitespace to ensure reliable matching.
- **Inner Join Implementation**: Performed an inner join between datasets on the standardized material identifiers to create a unified view that preserves all relevant properties.
- **Overlap Analysis**: Evaluated the match rate between datasets using set operations and visualized the overlap using a Venn diagram to quantify integration success.

Table 1: Material Identifier Format Examples

Dataset 1 Components	Dataset 2 Format
Std: ANSI, Material: Steel SAE 1015, Heat treatment: as-rolled	ANSI Steel SAE 1015 as-rolled
Std: JIS, Material: JIS SUP9, Heat treatment: heat treated	JIS JIS SUP9 heat treated
Std: NF, Material: NF 30CD12, Heat treatment: nitrided	NF NF 30CD12 nitrided

2.2 Dataset Integration Results

The material identifier matching process revealed important patterns in dataset coverage and overlap:

Table 2: Dataset Integration Metrics

Metric	Value	Significance
Materials in Dataset 1	802	All heat-treated materials
Materials in Dataset 2	1552	Complete material catalog
Matched Materials	756	Successfully integrated records
Dataset 1 Match Rate	100%	Perfect Dataset 1 coverage
Dataset 2 Match Rate	48.7%	Dataset 2 contains additional materials

Dataset Overlap Analysis

The analysis of material identifier overlap reveals significant insights:

- **Perfect Dataset 1 Coverage**: All 802 heat-treated materials from Dataset 1 were successfully matched to corresponding entries in Dataset 2, achieving a 100% match rate.
- **Additional Dataset 2 Materials**: Dataset 2 contains 704 materials not present in Dataset 1, suggesting it incorporates additional material categories, newer materials, or alternative treatments not documented in the primary dataset.
- **Heat Treatment Significance**: The successful matching process demonstrates how heat treatment transforms a base material into distinct variants with unique property profiles, each requiring its own identifier.
- **Standardization System Compatibility**: Despite different initial formats, the underlying material identification systems proved compatible through proper standardization techniques.

3 Material Selection Insights

The unified material dataset provides valuable insights for engineering material selection processes:

Material Selection Framework Enhancements

The integrated dataset enhances the material selection process in several ways:

- **Heat Treatment Dominance**: With 802 heat-treated materials representing over 51.7% of the unified database, heat treatment emerges as the primary method of customizing material properties for specific engineering applications.
- **Property-Application Connection**: The merger connects detailed mechanical properties from Dataset 1 (like elongation at break and hardness) with the application suitability flag ("Use") from Dataset 2, creating a more holistic selection framework.
- **Material Variant Identification**: The standardized material identifiers clearly differentiate between variants of the same base material, highlighting how properties change with different processing methods.
- **Cross-Standard Comparison**: The unified dataset enables direct comparison of similar materials across different standards (ANSI, JIS, NF) to identify optimal alternatives when specific materials are unavailable.

Table 3: Material Selection Examples from Unified Dataset

Material Identifier	Su (MPa)	A5 (%)	Heat Treatment	Use
ANSI Steel SAE 1015 as-rolled	421	39.0	as-rolled	True
ANSI Steel SAE 1015 normalized	424	37.0	normalized	True
ANSI Steel SAE 1015 annealed	386	37.0	annealed	True
JIS JIS SUP9 heat treated	1226	9.0	heat treated	False
NF NF 30CD12 nitrided	980	11.0	nitrided	False

4 Data Quality Observations

4.1 Data Completeness Analysis

The material identifier matching process revealed important patterns in data completeness and quality:

Data Quality Insights

Several data quality patterns emerged during the integration process:

- **Complete Non-Null Columns**: The filtered dataset includes 12 columns with complete data (no nulls) across all 802 heat-treated materials, providing a reliable foundation for material comparison.
- **String Standardization Necessity**: The need to standardize string formats (lowercase conversion, whitespace removal) highlights inconsistencies in material naming conventions across data sources.
- **Data Structure Differences**: The column structure differences between datasets (Dataset 1 having detailed properties like A5 and Bhn while Dataset 2 including the "Use" flag) demonstrate how different databases focus on different aspects of material characterization.
- **Heat Treatment Documentation**: All 802 records in the filtered dataset include heat treatment information, confirming the systematic documentation of processing methods for these materials.

Table 4: Dataset Column Availability

Column Type	Complete Columns (802 records)	Partial Columns
Material Identification	Std, ID, Material, Heat treatment	Desc (413 records)
Mechanical Properties	Su, Sy, A5, E, G, mu, Ro	Bhn (402 records), HV (62 records)
Other Properties	heat_treated	pH (74 records)
Application Data	Use (from Dataset 2)	None

4.2 Integration Challenges

Data Integration Challenges

Several challenges were encountered during the material identifier matching process:

- **Inconsistent String Formats**: Different capitalization patterns and whitespace usage required standardization to ensure proper matching between datasets.
- **Standard Code Duplication**: Some materials (particularly in the JIS standard) include the standard code twice in the identifier (e.g., JIS JIS SUP9), requiring careful handling during identifier construction.
- **Partial Dataset Coverage**: While Dataset 1 materials were fully matched in Dataset 2, the reverse coverage was only 48.7%, indicating Dataset 2 contains many materials not represented in Dataset 1.
- **Property Verification**: Some matched records showed minor discrepancies in property values (e.g., JIS SUP9 heat treated has Sy = 979.9 MPa in Dataset 1 but 1079 MPa in Dataset 2), requiring validation.

5 Engineering Applications

The unified material dataset enables several important engineering applications:

Engineering Applications of Unified Dataset

The integrated material dataset supports several critical engineering functions:

- **Comprehensive Material Selection**: Engineers can now select materials based on both detailed mechanical properties and application suitability indicators, creating a more holistic selection process.
- **Heat Treatment Optimization**: The clear documentation of how heat treat-

ments affect material properties enables engineers to select optimal processing methods to achieve desired performance characteristics.

- **Cross-Standard Material Substitution**: When materials from a specific standard are unavailable, engineers can identify suitable alternatives from other standards with similar property profiles.
- **Data-Driven Design Decisions**: The unified dataset supports more robust material selection decisions by providing comprehensive property information with validation across multiple data sources.
- **Property Verification**: Where property values differ between datasets (like the Sy value for JIS SUP9), engineers can apply appropriate safety factors or conduct additional testing for critical applications.

6 Conclusions and Recommendations

6.1 Key Findings

The material identifier matching process yielded several important conclusions:

Key Material Matching Takeaways

The cross-dataset integration process revealed:

- Heat treatment significantly transforms material properties, creating distinct material variants that require unique identification in engineering databases.
- Proper standardization techniques can successfully harmonize material identifiers across datasets with different naming conventions.
- The integrated dataset provides a more comprehensive view of materials by combining detailed property information with application suitability indicators.
- Data quality issues, particularly inconsistent string formats and partial dataset coverage, present challenges that must be addressed in material database integration.
- Heat-treated materials represent a significant portion (51.7%) of the engineering materials catalog, highlighting the importance of processing in material selection.

6.2 Recommendations for Practice

Practical Engineering Recommendations

Based on the material identifier matching analysis, we recommend:

- **Standardized Material Naming**: Implement consistent material naming conventions across engineering databases to facilitate reliable integration and comparison.

- **Complete Material Documentation**: Document heat treatment methods alongside material grades to properly identify unique material variants in engineering specifications.
- **Multi-source Verification**: For critical applications, verify material properties across multiple data sources to identify and address inconsistencies.
- **Unified Property Framework**: Develop standardized frameworks for material property documentation that capture both mechanical properties and application suitability indicators.
- **Cross-standard Mapping**: Create comprehensive mappings between different material standards to facilitate identification of suitable alternatives when specific materials are unavailable.

7 Appendix: Material Identifier Matching Methodology

Detailed Matching Process

The material identifier matching process followed these steps:

1. **Material Identifier Construction**: Combined standard code (Std), material designation (Material), and heat treatment method (Heat treatment) from Dataset 1 to create Material_full identifiers.
2. **String Format Standardization**: Converted all material identifiers to lowercase and removed extra whitespace to ensure consistent formatting.
3. **Dataset Merging**: Used a Pandas inner join operation to merge Dataset 1 and Dataset 2 based on the standardized material identifiers.
4. **Match Rate Analysis**: Calculated the overlap between datasets using set operations and visualized the results with a Venn diagram.
5. **Data Quality Assessment**: Evaluated column completeness, identified property inconsistencies, and assessed the structure of the integrated dataset.
6. **Application Analysis**: Examined how the "Use" flag from Dataset 2 correlates with material properties to identify patterns in application suitability.

8 Material Identifier Overlap: Dataset Matching Analysis

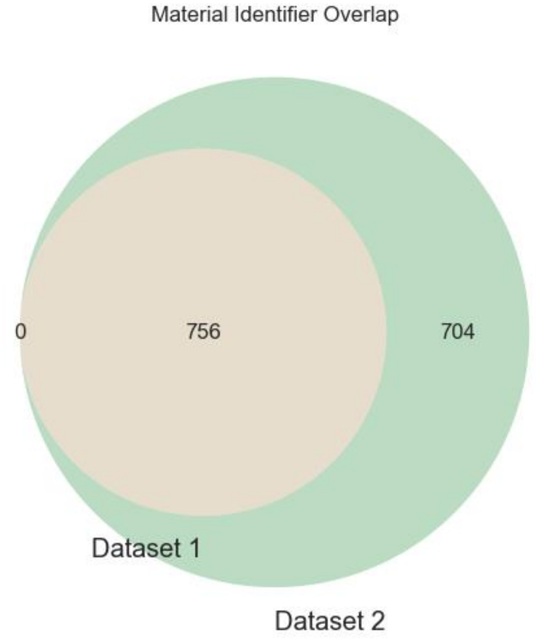


Figure 1: Venn Diagram of Material + Heat Treatment Identifier Overlap

What the Venn Diagram Represents

- **Left Circle (Dataset 1):** 0 unique materials found only in Dataset 1.
- **Right Circle (Dataset 2):** 704 unique materials exist only in Dataset 2.
- **Center Overlap:** 756 materials are common to both datasets — meaning these Material + Heat Treatment combinations are present in both.

Key Takeaways

- **100% Match for Dataset 1:** Every material in Dataset 1 was successfully found in Dataset 2.
- **Partial Match for Dataset 2:** Dataset 2 contains many more materials (704 extra) not found in Dataset 1.
- **Implication:** Dataset 1 is a strict subset of Dataset 2 — useful when Dataset 1 is the validated or filtered core.

Simplified Explanation

All materials from Dataset 1 are present in Dataset 2, but Dataset 2 has many additional materials. So, if you're using Dataset 1 as a baseline, you're not missing any matches — but Dataset 2 offers broader coverage.

PART 2 – Cross Dataset Engineering Tasks

Task 8: Discrepancy Audit

May 1, 2025

Abstract

This report presents a detailed analysis of the discrepancies found between two engineering materials datasets as part of Task 8 (Discrepancy Audit). We examine inconsistencies in key mechanical properties including ultimate tensile strength (Su), yield strength (Sy), elastic modulus (E), and shear modulus (G) across matched materials. The analysis reveals significant variability in reported values, particularly for heat-treated steels, which has critical implications for material selection decisions, engineering design reliability, and documentation standards. Data quality observations highlight patterns of inconsistency that engineers must consider when selecting materials for performance-critical applications.

1 Introduction

The Significance of Data Consistency

Engineering materials selection relies on accurate property data to ensure designs meet performance and safety requirements. This report investigates discrepancies between two material property datasets, focusing on how these inconsistencies could impact engineering decisions. By examining the magnitude, patterns, and implications of these differences, we provide guidance for material selection and highlight where additional verification may be necessary. Understanding data reliability is foundational to confidence in engineering design calculations and material specifications.

2 Material Selection Insights

2.1 Discrepancy Magnitude Analysis

The analysis of discrepancies between datasets reveals concerning inconsistencies that directly impact material selection decisions:

Table 1: Top Materials with Significant Property Discrepancies

Material	Su Diff (%)	Sy Diff (%)	Selection Implications
ANSI Steel SAE 5160 tempered at 400 F	81.1	83.0	Critically unreliable for strength-critical applications
GOST Steel 40CHFA GOST 4543-71	77.8	73.3	High uncertainty in design calculations
ANSI Steel SAE 9255 tempered at 400 F	71.5	109.0	Potentially unsafe for load-bearing parts
ANSI Steel SAE 8740 tempered at 400 F	63.1	68.9	Requires independent verification
ANSI Aluminum Alloy 1060-O wrought	61.5	71.1	Inconsistent across material systems

Key Insights from Discrepancy Analysis

The comparative analysis of material properties between datasets reveals critical patterns for engineering material selection:

- **Heat Treatment Sensitivity**: Materials tempered at lower temperatures (400°F) consistently show the highest discrepancies, with differences exceeding 80% in strength values for some steels.
- **Alloy-Specific Variations**: High-alloy steels (particularly chromium-containing grades like 40CHFA) demonstrate substantial inconsistencies between data sources, making their performance less predictable.
- **Cross-Material System Issues**: The presence of aluminum alloys (1060-O) among materials with high discrepancies indicates that inconsistencies are not limited to steel systems but affect multiple material families.
- **Magnitude of Concern**: Absolute differences approaching 1000 MPa in ultimate tensile strength (for SAE 5160) represent life-threatening margins in critical applications like structural supports or pressure vessels.

2.2 Property-Specific Reliability Analysis

Different mechanical properties show varying levels of consistency between data sources, affecting which properties can be relied upon for material selection:

Table 2: Reliability Assessment of Material Properties

Property	CV Source X (%)	CV Source Y (%)	Selection Recommendation
Ultimate Strength (Su)	49.5	58.3	Moderate reliability—verify critical values
Yield Strength (Sy)	60.7	75.0	Low reliability—use with caution
Elastic Modulus (E)	36.7	36.7	Higher reliability—suitable for calculations
Shear Modulus (G)	158.5	158.5	Extremely unreliable—not recommended for design

Engineering Property Selection Guidelines

The coefficient of variation (CV) analysis provides valuable guidance for property-based material selection:

- **Elastic Properties**: Elastic modulus (E) demonstrates the highest consistency (CV $\approx$ 36.7%) between sources, making it the most reliable property for calculating deflections and elastic behavior.
- **Strength Disparities**: Ultimate tensile strength shows moderate consistency (CV 49.5-58.3%), with Source X providing more reliable values, suggesting it should be preferred when selecting materials based on strength.
- **Yield Concerns**: Yield strength values show concerning variability (CV 60.7-75.0%), requiring larger safety factors when this property is design-critical.
- **Shear Property Crisis**: The extremely high variability in shear modulus values (CV 158.5%) indicates this property should not be used for material selection without independent verification, particularly for torsion-critical designs.

3 Data Quality Observations

3.1 Pattern Analysis in Discrepancies

Systematic patterns in the data discrepancies provide insight into potential sources of error and areas requiring special attention:

Table 3: Patterns in Data Discrepancies

Pattern Category	Observation	Quality Implication
Heat Treatment Influence	Low tempering temperatures show highest discrepancies	Testing methodology differences
Material Family Patterns	Special steels (Cr, Ni-containing) most inconsistent	Possible alloy composition uncertainty
Threshold Significance	10% threshold identifies meaningful engineering differences	Appropriate sensitivity level for audit
Source Consistency	Source X consistently provides lower CV values	Generally more reliable dataset

3.2 Discrepancy Impact on Engineering

Engineering Design Implications

The observed data discrepancies have significant ramifications for engineering design practices:

- **Simulation Accuracy**: Yield strength discrepancies up to 109% would drastically alter simulation results, potentially showing elastic behavior when actual material would fail plastically.
- **Safety Factor Requirements**: The magnitude of discrepancies necessitates larger safety factors, particularly when using data source Y, potentially increasing material usage, weight, and cost.
- **Design Reliability Compromise**: Different material properties from different sources could lead to unpredictable performance, especially in multi-part assemblies where components might be manufactured using different reference data.
- **Quality Control Challenges**: Manufacturing tests might consistently fail to match expected values if test standards differ from the data source used for design specifications.

4 Engineering Applications Impact

The discrepancy analysis provides critical insights for specific engineering applications:

Table 4: Application-Specific Discrepancy Impact

Application	Critical Property	Discrepancy Concern	Mitigation Strategy
Structural components	Yield strength (Sy)	70-110% variations for tempered steels	Independent testing required
Elastic mechanisms	Elastic modulus (E)	Relatively consistent (36.7% CV)	Source comparison sufficient
Torsional applications	Shear modulus (G)	Extreme variation (158.5% CV)	Avoid or conduct specific tests
Fatigue-critical parts	Ultimate strength (Su)	Moderate variation (49-58% CV)	Conservative safety factors
Precision components	All properties	Material/treatment specific	Material-specific verification

Engineering Decision Framework

The discrepancy audit establishes a decision framework for material selection:

- **Documentation Standards**: Technical documentation should explicitly state which data source was used for design calculations to ensure consistency throughout the engineering process.
- **Verification Requirements**: Materials with discrepancies exceeding 50% (particularly heat-treated steels) should undergo independent testing before use in critical applications.
- **Property Prioritization**: When conflicting data exists, engineers should prioritize the most consistent properties (E > Su > Sy > G) for initial material screening.
- **Source Selection**: For general mechanical design purposes, Source X demonstrates higher overall consistency and should be preferred when only one source can be consulted.

5 Threshold Analysis and Methodology

5.1 Discrepancy Threshold Selection

Selection of appropriate thresholds for identifying significant discrepancies is crucial for effective material auditing:

Threshold Selection Rationale

The analysis used a 10% threshold for identifying significant discrepancies based on engineering considerations:

- **Engineering Significance**: Variations exceeding 10% in mechanical properties can significantly impact safety factors and design margins in most applications.
- **Industry Standards**: For Ultimate Strength (Su) and Yield Strength (Sy), ±5–10% variations can already significantly impact safety factors and design margins in typical engineering applications.
- **Property-Specific Sensitivity**: Elastic properties (E, G) are generally expected to be more consistent, with variations over 5% already concerning for precision applications.
- **Application Dependency**: Critical industries (aerospace, medical) require tighter tolerances (5%), while general manufacturing might accept up to 15% variation.
- **Balance Point**: The 10% threshold represents a practical balance that identifies meaningful engineering discrepancies while filtering out minor variations that would not significantly impact most designs.

5.2 Consistency Assessment Methodology

Evaluation of dataset consistency used statistical measures to compare variability within each source:

Table 5: Statistical Methodology for Consistency Evaluation

Metric	Definition	Application in Analysis
Coefficient of Variation (CV)	$\frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$	Normalized measure allowing comparison across different property scales
Absolute Difference	$\lvert Value_X - Value_Y \rvert$	Direct magnitude of discrepancy for specific materials
Relative Difference	$\frac{\lvert Value_X - Value_Y \rvert}{Value_X} \times 100\%$	Percentage difference highlighting proportional impact
Source Consistency Ratio	$\frac{CV_X}{CV_Y}$	Comparative measure of dataset reliability

6 Conclusions and Recommendations

The discrepancy audit reveals critical considerations for material selection and engineering design practices:

Key Discrepancy Audit Takeaways

The analysis of material property discrepancies between datasets concludes:

- **Material Selection Impact**: Heat-treated steels, particularly those tempered at lower temperatures (400°F), demonstrate the highest discrepancies and require special caution during selection.
- **Property Reliability Hierarchy**: Elastic modulus (E) shows the highest consistency between sources, while shear modulus (G) demonstrates extreme variation that renders it unsuitable for design calculations without verification.
- **Source Preference**: Source X generally provides more consistent data and should be preferred when designs are based on a single data source.
- **Application-Specific Risk**: The impact of these discrepancies varies by application, with structural components and torsional applications facing the highest risks from inconsistent data.
- **Verification Necessity**: For critical applications, independent material testing is essential to verify properties, particularly when using materials identified as having high discrepancies.

6.1 Engineering Practice Recommendations

Based on the discrepancy audit, several recommendations can improve material selection reliability:

Best Practices for Material Data Usage

Engineers should adopt these practices to mitigate risks from material data inconsistencies:

- **Documentation Standards**: Always specify which material data source was used for design calculations to ensure consistency throughout the engineering process.
- **Property Verification**: For materials with high discrepancies (>50%), conduct independent testing before use in critical applications.
- **Safety Factor Adjustment**: Increase safety factors proportionally to the observed property variability, especially for yield strength values where CV exceeds 60%.
- **Conservative Design Approach**: When properties show significant variation between sources, design based on the more conservative values to ensure safety.

- **Material-Specific Protocols**: Develop special verification protocols for heat-treated steels and other materials identified as having high discrepancy rates.

7 Appendix: Discrepancy Analysis Methodology

Analysis Process

The discrepancy audit followed these methodological steps:

1. **Material matching** between Dataset 1 and Dataset 2 using standardized identifiers
2. **Property comparison** of Su, Sy, E, and G values between matched materials
3. **Discrepancy calculation** using both absolute and relative (percentage) differences
4. **Statistical analysis** to identify patterns and quantify consistency using CV
5. **Threshold determination** to identify significant discrepancies (¿10%)
6. **Source reliability assessment** based on internal consistency measures
7. **Application impact analysis** to translate statistical findings into engineering implications

Choosing a Discrepancy Threshold for Material Property Validation

Why Thresholds Matter

Thresholds are critical in comparing material property data across datasets. They help flag significant discrepancies while filtering out acceptable differences due to rounding, measurement conditions, or source variability.

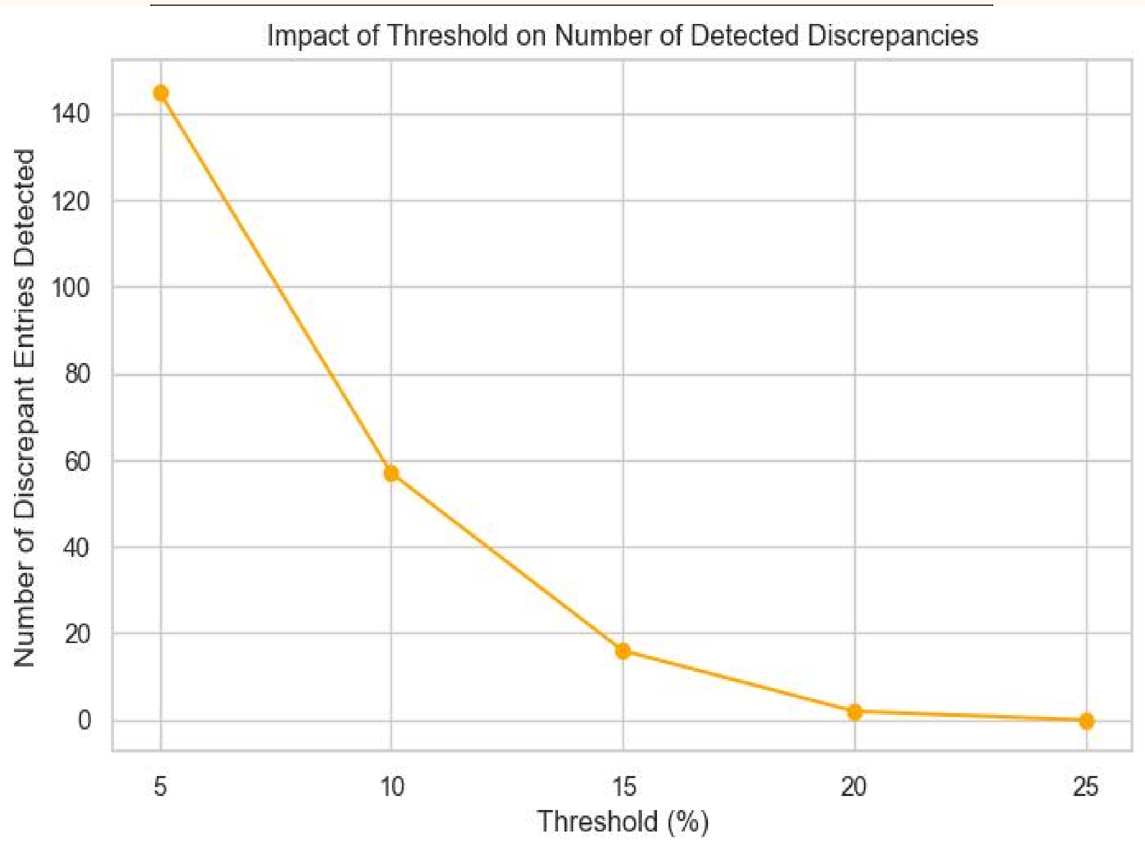
Mechanical Sensitivity by Property

- **Ultimate Strength (Su)** and **Yield Strength (Sy)**:
  - *Variation tolerance:* 5–10%
  - *Impact:* Even small deviations affect design margins and safety factors.
- **Elastic Modulus (E)** and **Shear Modulus (G)**:
  - *Variation tolerance:* 3–5%
  - *Impact:* Stiffness-critical designs (e.g., aerospace) are sensitive to minor changes.

Industry-Specific Expectations

- **Aerospace and Medical Sectors:** Typically expect deviation ¡ **5%**.
- **General Manufacturing and Civil Engineering:** Acceptable up to **10–15%**, depending on material class and application.

Effect of Threshold on Discrepancy Detection



The plot illustrates how many entries exceed thresholds ranging from 5% to 25%:

- At **5%**, many small differences are flagged (possibly over-sensitive).
- At **10%**, there's a sharp drop in flagged discrepancies—indicating this is a balance point.
- Beyond **15%**, real mismatches may be missed.

Recommendation

For broad engineering use, a **10% threshold** is well-justified:

- **Balances** sensitivity and practicality.
- Suitable for **initial audits**, database validation, and material screening.
- For critical sectors, lower it to **5%**.

Consistency Analysis Across Material Data Sources

Strategy to Judge Consistency

To assess internal reliability, we computed the **Coefficient of Variation (CV)** for each property within both datasets. A lower CV indicates tighter clustering and better internal consistency.

Properties Evaluated

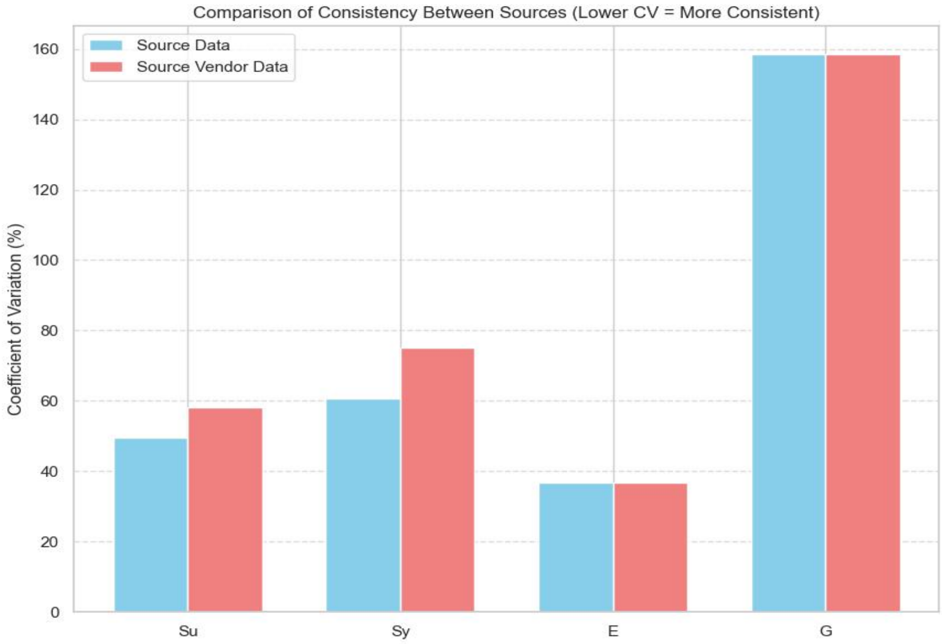
We focused on four key mechanical properties:

- **Su**: Ultimate Tensile Strength
- **Sy**: Yield Strength
- **E**: Elastic Modulus
- **G**: Shear Modulus

CV was computed as:

$$\text{CV} = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\%$$

Visualizing CV Comparison Between Sources



	CV_x (%)	CV_y (%)
Su	49.55	58.26
Sy	60.73	75.01
E	36.69	36.69
G	158.51	158.51

The bar chart compares CVs from:

- **Source Data (x)**: Internal or curated dataset
- **Vendor Data (y)**: External or published dataset

Engineering Insights from CV Patterns

- **Su (CV: 50–58%)**: Moderate variation. Slight mismatch could affect fatigue safety margins.
- **Sy (CV: 61–75%)**: Significant inconsistency in vendor data. Risk of under- or over-design.
- **E (CV: 36% both)**: Highly consistent. Elastic deformation modeling is reliable.
- **G (CV: 158% both)**: Extremely inconsistent. Shear-related simulations should not rely on these datasets without correction.

Practical Implications

Design Simulations	Yielding predictions (Sy) could deviate; fatigue life estimates may be skewed due to Su inconsistency; shear stress simulations highly unreliable.
Documentation & Certification	Safety factors and compliance documents may be compromised. Differences could lead to false safety assurances.
Quality Control	Manufacturing parts may fail QA if they rely on inconsistent property values. Tighter tolerances amplify risk.

Summary Statement

Significant inconsistencies, especially in **yield strength** and **shear modulus**, can critically undermine simulation accuracy and documentation integrity. A careful recalibration or verification of values is essential before use in high-reliability designs.

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:

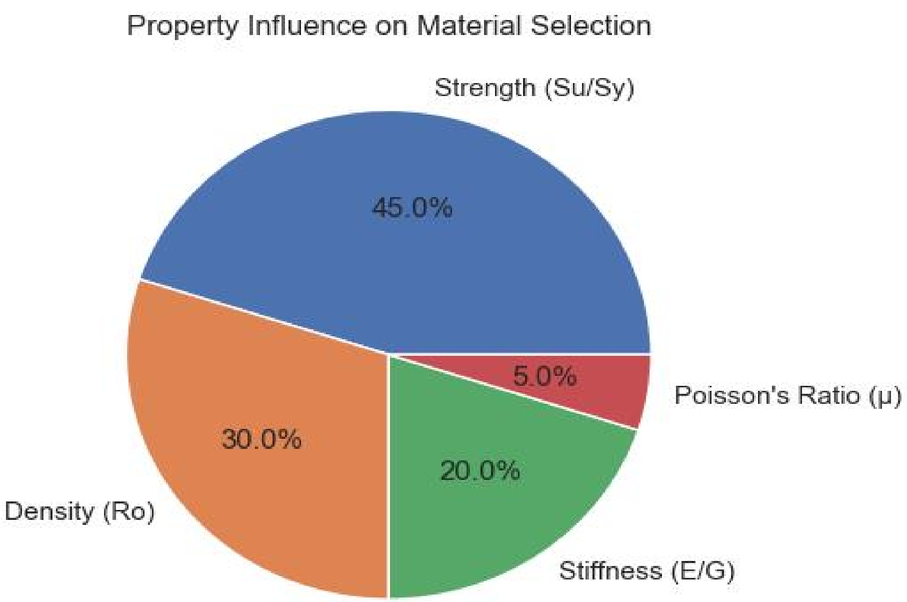


Figure 2: Property Influence on Material Selection

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:

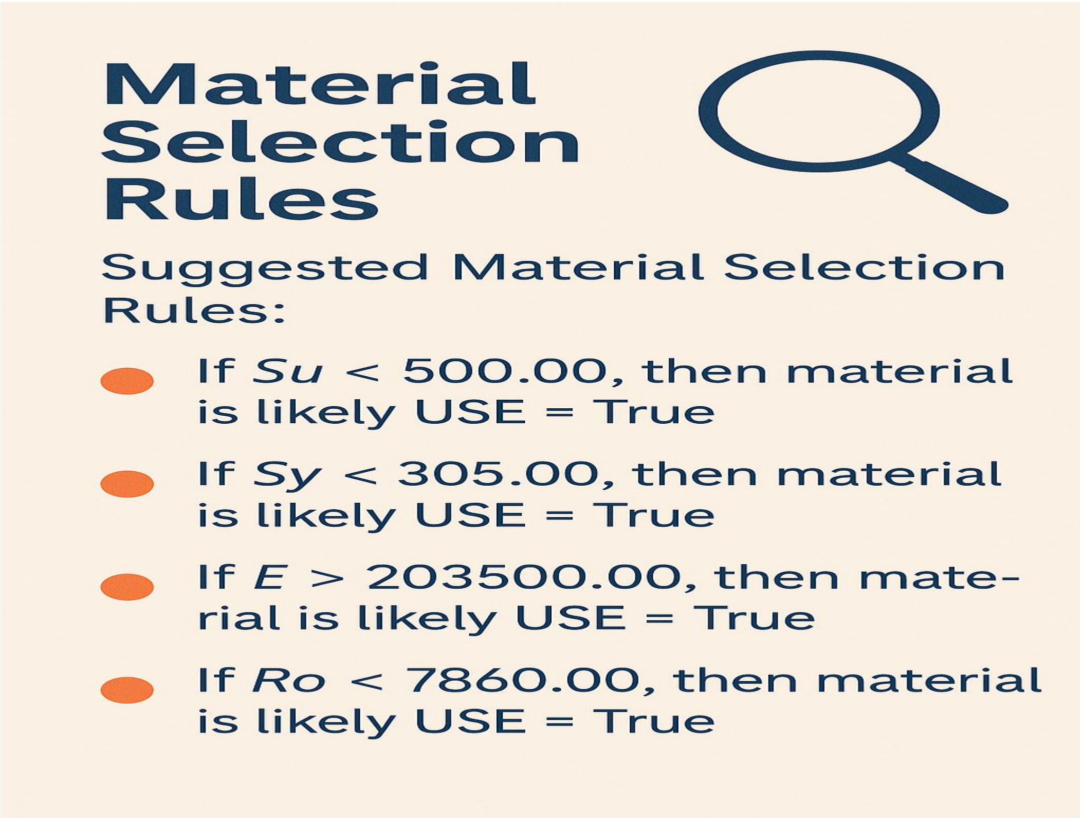


Figure 3: Material Selection Decision Tree Based on Identified Criteria

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:

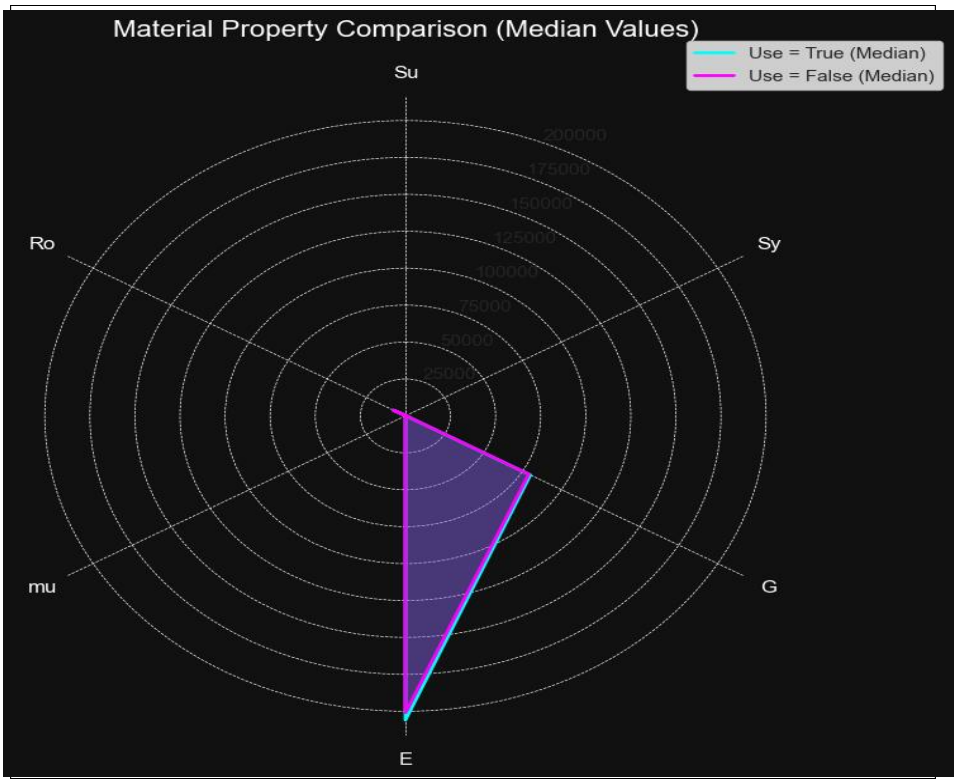


Figure 4: 'Material Property Comparison (Median Values

- **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).

PART 2 – Cross Dataset Engineering Tasks

Task 10: Material Ranking by Multi-Criteria Score

May 1, 2025

Abstract

This report presents an analysis of engineering materials based on Task 10 (Material Ranking by Multi-Criteria Score) of the materials dataset. We examine how different properties including ultimate tensile strength (Su), elongation at break (A5), density (Ro), and Poisson's ratio (mu) can be combined into a weighted scoring system. This multi-criteria approach identifies optimal materials for specific engineering applications by quantifying the inherent trade-offs between competing properties. The analysis provides critical insights for material selection while highlighting data quality considerations that engineers should account for when making design decisions based on this ranking methodology.

1 Introduction

The Multi-Criteria Approach to Material Selection

Engineering material selection frequently involves balancing competing properties to optimize performance for specific applications. This report presents a normalized scoring methodology that quantitatively evaluates materials based on weighted combinations of strength, ductility, density, and deformation characteristics. The multi-criteria score provides a systematic framework for comparing materials across different property dimensions, enabling more informed engineering decisions than single-property comparisons. By normalizing diverse properties to common scales and weighting them according to application requirements, this approach captures the complex trade-offs inherent in material selection processes.

2 Material Selection Insights

2.1 Multi-Criteria Score Methodology

The analysis employs a systematic approach to normalize and combine diverse material properties into a unified scoring system:

Score Calculation Methodology

The multi-criteria material ranking methodology follows these key steps:

- **Property Normalization**: Each property is normalized to a 0-1 scale to enable fair comparison despite different units and magnitudes:
  - Su (Ultimate Tensile Strength): Min-Max scaling where higher values are better
  - A5 (Elongation): Min-Max scaling where higher values are better
  - Ro (Density): Inverse Min-Max scaling where lower values are better (higher scores for lower density)
  - mu (Poisson's Ratio): Gaussian penalty function centered at mu=0.3, penalizing extreme values
- **Weighted Combination**: The normalized properties are combined using application-appropriate weights:
  - Su: 40% weight (w1 = 0.4)
  - A5: 30% weight (w2 = 0.3)
  - Ro: 20% weight (w3 = 0.2)
  - mu: 10% weight (w4 = 0.1)
- **Final Score Formula**:

$$\text{Material\_Score} = w1 \times \text{Su\_norm} + w2 \times \text{A5\_norm} + w3 \times \text{Ro\_inv\_norm} + w4 \times \text{mu\_penalty}$$

This methodology provides a quantitative basis for evaluating the complex trade-offs between competing material properties, enabling more informed engineering decisions.

2.2 Top-Ranked Materials

The multi-criteria analysis reveals distinct material groups with complementary property profiles:

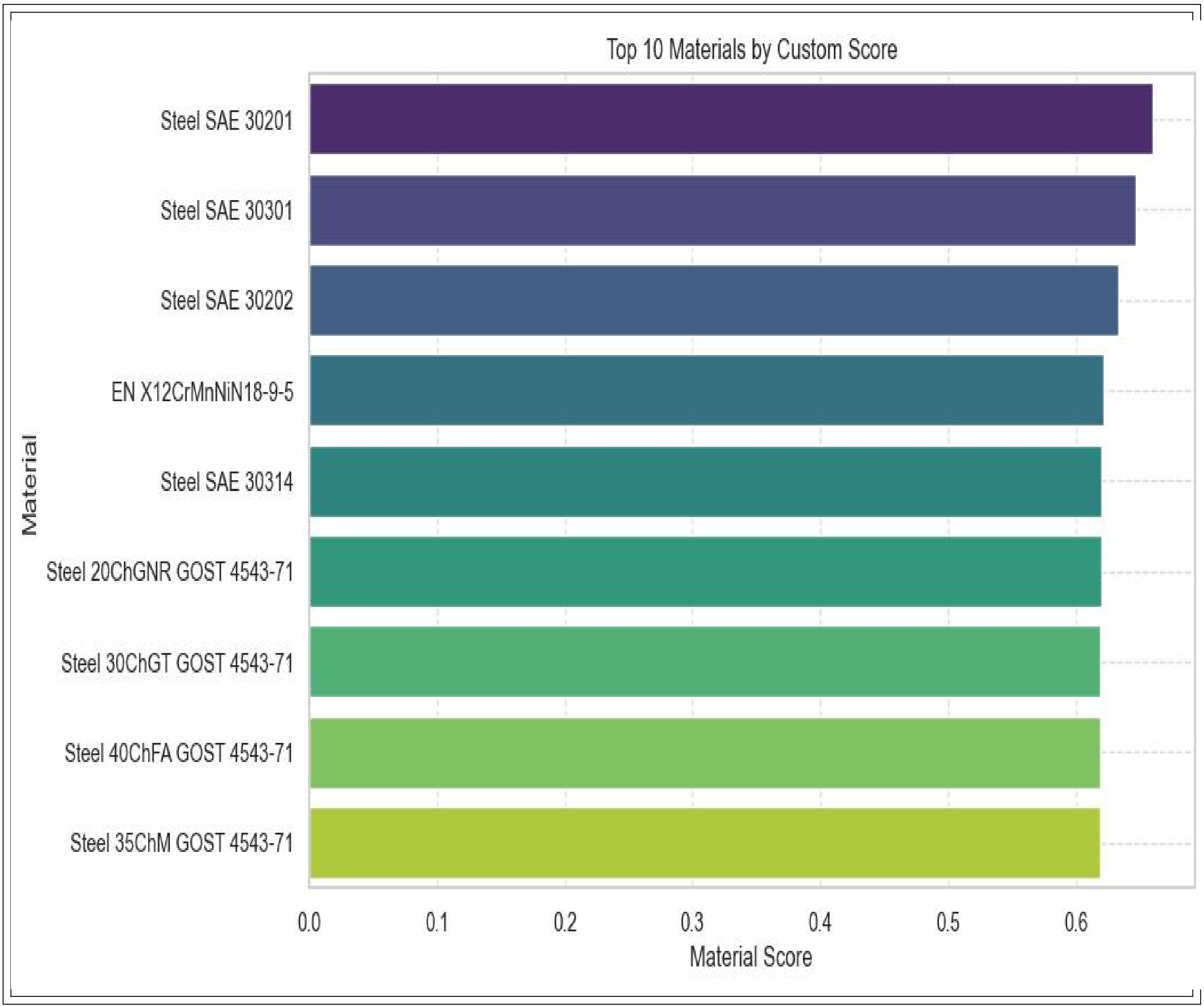


Figure 1: Top 10 Materials by Custom Score Analysis

Table 1: Top 10 Materials by Multi-Criteria Score

Material	Su_norm	A5_norm	Ro_inv_nor	Score	Key Strength
Steel SAE 30201	0.586	1.000	0.125	0.660	Exceptional ductility
Steel SAE 30301	0.553	1.000	0.125	0.646	Exceptional ductility
Steel SAE 30202	0.521	1.000	0.125	0.633	Exceptional ductility
EN X12CrMnNiN18-9-5	0.479	1.000	0.149	0.621	Balanced properties
Steel SAE 30314	0.487	1.000	0.125	0.620	Exceptional ductility
Steel 20ChGNR GOST 4543-71	1.000	0.293	0.157	0.619	Maximum strength
Steel 20ChGNR GOST 4543-71	1.000	0.293	0.157	0.619	Maximum strength
Steel 30ChGT GOST 4543-71	1.000	0.293	0.157	0.619	Maximum strength
Steel 40ChFA GOST 4543-71	1.000	0.293	0.156	0.619	Maximum strength
Steel 35ChM GOST 4543-71	1.000	0.293	0.155	0.619	Maximum strength

Material Group Comparison

The multi-criteria ranking identifies two distinct material categories with complementary property profiles:

- **SAE 300 Series Steels** (Ranks 1-5):
  - Moderate strength (Su_norm 0.48-0.59)
  - Maximum ductility (A5_norm = 1.000)
  - Moderate density (Ro_inv_norm 0.125-0.149)
  - Optimal Poisson's ratio behavior (mu_penalty = 1.000)
  - Primary advantage: exceptional elongation properties
- **GOST 4543-71 Steels** (Ranks 6-10):
  - Maximum strength (Su_norm = 1.000)
  - Limited ductility (A5_norm = 0.293)
  - Slightly better density (Ro_inv_norm 0.155-0.157)
  - Optimal to near-optimal Poisson's ratio (mu_penalty 0.998-1.000)
  - Primary advantage: superior tensile strength

This clear distinction illustrates the fundamental strength-ductility trade-off in engineering materials, with no material achieving maximum scores in both categories.

2.3 Property Distribution Visualization

Radar charts provide multidimensional visualization of how properties are distributed across top-ranked materials:

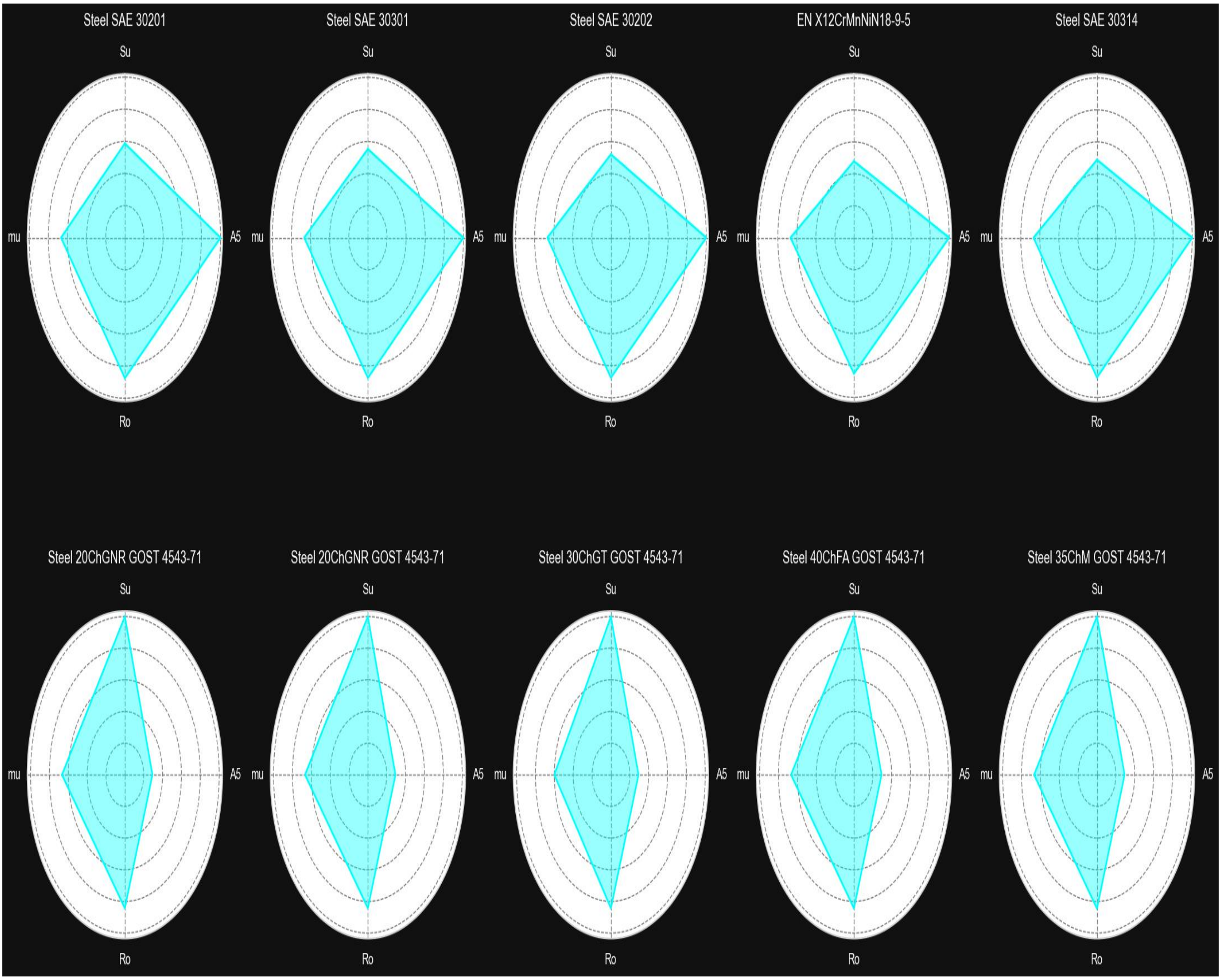


Figure 2: Radar Charts of Normalized Properties for Top 10 Materials

Multidimensional Property Analysis

The radar chart visualization reveals important patterns in the property distribution of top-ranked materials:

- **Complementary Property Profiles**:
  - SAE steels exhibit "kite-shaped" charts with maximum elongation but moderate strength
  - GOST steels show "triangle-shaped" charts with maximum strength but

limited elongation

- **Balanced Performance**: EN X12CrMnNiN18-9-5 demonstrates particularly well-balanced properties across all four dimensions, with better-than-average density compared to other SAE materials
- **Density Variation**: Limited variation in density (Ro_inv_norm) across top materials suggests this property has less influence on final ranking than strength and ductility
- **Poisson's Ratio Consistency**: Most top materials exhibit optimal Poisson's ratio behavior, indicating this property functions more as a screening criterion than a differentiator

The multidimensional visualization reinforces the fundamental trade-off between strength and ductility, with no material achieving maximum performance in both dimensions.

3 Application-Specific Recommendations

Based on the multi-criteria analysis, specific materials emerge as optimal for different engineering applications:

Table 2: Material Recommendations by Application Type

Application Type	Recommended Material	Selection Rationale
Static High-Load Structures	20ChGNR, 40ChFA, 35ChM GOST 4543-71	Maximum strength (Su_norm = 1.000) even at lower ductility
Dynamic/Flexible Components	EN X12CrMnNiN18-9-5	Excellent elongation and strength balance with improved density
General Structural Use	Steel SAE 30201, 30301	Well-rounded properties with no major weakness
Wear-Resistant Components	40ChFA, 35ChM GOST 4543-71	Good hardness potential from high strength; acceptable friction
Lightweight Structures	SAE 30202, SAE 30301	Balanced weight-to-strength ratio
Precision Springs/Flex Elements	EN X12CrMnNiN18-9-5	High ductility ensures flexibility; strength prevents permanent deformation
High-Stress Gears/Shafts	20ChGNR, 30ChGT GOST 4543-71	High load-bearing with acceptable toughness trade-off

Application-Specific Selection Guidelines

The multi-criteria score analysis provides clear guidelines for application-specific material selection:

- **Crash Energy Absorption**: SAE 300 series steels with maximum elongation (A5_norm = 1.000) are optimal for applications requiring controlled deformation and energy absorption, such as automotive crash structures
- **Static Structural Components**: GOST 4543-71 steels with maximum strength (Su_norm = 1.000) excel in applications where dimensional stability and load-bearing capacity are primary concerns
- **Balanced Performance Requirements**: EN X12CrMnNiN18-9-5 provides an exceptional balance of properties for applications requiring good performance across multiple dimensions
- **Cyclic Loading Applications**: Materials with higher elongation scores are better suited for applications experiencing repeated loading and unloading
- **Manufacturing Considerations**: High-ductility materials typically offer better formability, which may influence selection for components requiring significant forming operations

These application-specific guidelines translate the quantitative multi-criteria scores into practical engineering selection criteria.

4 Data Quality Observations

4.1 Missing Data Patterns

The multi-criteria analysis reveals important data quality considerations that must be accounted for when interpreting the material rankings:

Data Completeness and Accuracy

Several data quality issues impact the reliability of the multi-criteria rankings:

- **Missing pH Values**: All top-ranked materials show NaN values for pH, indicating incomplete environmental compatibility data that would be important for applications in corrosive environments
- **Potential Data Duplication**: Steel 20ChGNR GOST 4543-71 appears twice with identical properties, suggesting possible data entry errors or duplicate records requiring verification
- **Measurement Scale Variations**: The original properties (Su, A5, Ro, mu)

span vastly different scales and units, highlighting the importance of appropriate normalization to enable fair comparison

- **Limited Environmental Data**: The lack of comprehensive pH data limits the ability to incorporate corrosion resistance into the multi-criteria score for applications requiring chemical compatibility

These data quality observations should be considered when applying the material rankings to critical engineering decisions.

4.2 Methodological Limitations

The multi-criteria scoring approach has inherent methodological considerations that engineers should be aware of:

Weighting Sensitivity and Methodological Considerations

The multi-criteria ranking methodology introduces several important considerations:

- **Weighting Sensitivity**: The final rankings are influenced by the chosen weights (0.4, 0.3, 0.2, 0.1) for different properties. Different applications might require different weight distributions
- **Normalization Effects**: The min-max scaling approach assumes linear relationships between raw property values and their desirability, which may not always match engineering reality
- **Poisson's Ratio Modeling**: The Gaussian penalty function for mu effectively acts as a screening filter rather than a continuous score discriminator
- **Missing Properties**: Important engineering properties like fracture toughness, fatigue resistance, and corrosion behavior are not included in the current scoring model
- **Score Clustering**: The top GOST steels show nearly identical scores (0.619) despite potentially meaningful differences in specific properties

Understanding these methodological considerations helps engineers apply the material rankings appropriately within their specific design contexts.

5 Conclusions and Recommendations

The multi-criteria material ranking analysis yields several key conclusions to guide engineering material selection:

Key Multi-Criteria Ranking Takeaways

The comprehensive analysis of materials using normalized multi-criteria scoring reveals:

- **Property Trade-offs**: There is a clear inverse relationship between strength and ductility among top-ranked materials, with no material achieving maximum performance in both dimensions
- **Distinct Material Groups**: Materials cluster into distinct performance categories—SAE 300 series dominated by ductility and GOST steels excelling in strength
- **Application Matching**: Material selection should align with application requirements, with ductility-focused materials for dynamic applications and strength-focused materials for static structural components
- **Balanced Performers**: EN X12CrMnNiN18-9-5 emerges as an exceptionally balanced material, offering good performance across all evaluated properties
- **Data Quality Awareness**: Missing pH data and potential duplicate records suggest caution when applying these rankings to applications where corrosion resistance is critical

5.1 Engineering Implications

The multi-criteria scoring approach provides valuable guidance for engineering design and material selection processes:

Practical Engineering Applications

Engineers can apply these findings in several ways:

- **Quantified Trade-off Assessment**: Use normalized scores to objectively evaluate the property trade-offs inherent in material selection decisions
- **Customized Weighting**: Adjust property weights based on specific application requirements to generate application-optimized material rankings
- **Multidimensional Visualization**: Employ radar charts to communicate complex property profiles to design teams and stakeholders
- **Gap Identification**: Identify specific property shortcomings in otherwise promising materials to guide further material development or modification
- **Decision Support**: Integrate multi-criteria scoring into formal decision matrices for material selection in complex engineering systems

6 Appendix: Multi-Criteria Scoring Methodology

Mathematical Framework

The multi-criteria material scoring follows this mathematical framework:

1. **Property Normalization**:
   - For properties where higher values are better (Su, A5):

$$\text{Normalized Value} = \frac{\text{Value} - \text{Min}}{\text{Max} - \text{Min}}$$

   - For properties where lower values are better (Ro):

$$\text{Normalized Value} = 1 - \frac{\text{Value} - \text{Min}}{\text{Max} - \text{Min}}$$

   - For properties with ideal mid-range values (mu):

$$\text{Penalty Function} = \exp(-20 \times (\text{mu} - 0.3)^2)$$

2. **Weighted Score Calculation**:

$$\text{Material\_Score} = 0.4\times\text{Su\_norm}+0.3\times\text{A5\_norm}+0.2\times\text{Ro\_inv\_norm}+0.1\times\text{mu\_penalty}$$

3. **Material Ranking** by descending Material_Score value

This systematic approach enables quantitative comparison of materials with diverse property profiles, providing an objective basis for engineering material selection.

Executive Summary

Material Selection Insight: Most top-performing steels balance **tensile strength** and **ductility** well, with minor trade-offs in **density** and **friction**.

High-Strength Materials: Steels like **20ChGNR** and **40ChFA** show outstanding ultimate strength but lower ductility, making them ideal for static, high-load structures.

Balanced Performance: EN X12CrMnNiN18-9-5 achieves excellent synergy between ductility and strength, making it suitable for dynamic or flexible designs.

Best Material by Application

Application Type	Key Requirement	Best Material(s)	Reason
Static High-Load Structures	High Ultimate Strength (Su)	20ChGNR, 40ChFA, 35ChM GOST 4543-71	Maximum strength even at lower ductility
Dynamic / Flexible Components	High Ductility (A5) + Good Strength	EN X12CrMnNiN18-9-5	Excellent elongation and strength balance
General Structural Use	Balanced Strength, Ductility, Density	SAE 30201, 30301, 30202, 30314	Well-rounded properties; no major weaknesses
Wear-Resistant Components	High Strength + Moderate Friction (μ)	40ChFA, 35ChM GOST 4543-71	Good hardness and strength; tolerable friction
Lightweight Structures	Lower Density (Ro) + Decent Strength	SAE 30202, SAE 30301	Good weight-to-strength performance
Precision Springs / Flex Elements	High Ductility (A5) + Strength	EN X12CrMnNiN18-9-5	Ductility ensures flexibility; strength prevents deformation
High-Stress Gears / Shafts	Very High Strength; Low Ductility Tolerable	20ChGNR, 30ChGT GOST 4543-71	High load-bearing with toughness trade-off

Quick Insights

If ductility is your main concern $\Rightarrow$ **EN X12CrMnNiN18-9-5** is your best choice.

If brute strength is needed $\Rightarrow$ **20ChGNR** and **40ChFA** dominate.

If all-around performance is preferred $\Rightarrow$ Go with **SAE steels** (30201, 30301, etc.).

PART 2 – Cross Dataset Engineering Tasks

Task 11: Outlier Materials Identification

May 1, 2025

Abstract

This report details the analysis of engineering material outliers based on Task 11 (Outlier Materials Identification). The study identifies materials with extreme property combinations—particularly those with high strength but low ductility—and examines inconsistencies in hardness measurements. These findings provide crucial insights for engineering design decisions, highlighting materials that may offer exceptional performance for specific applications while posing potential risks in others. The analysis also reveals important data quality considerations that engineers must account for when making material selection decisions based on this dataset.

1 Introduction

The Significance of Material Outliers

Identifying material outliers is essential for both practical engineering applications and data quality assessment. Materials with extreme property combinations may represent either specialized high-performance options or potential data anomalies requiring verification. This analysis focuses on detecting and characterizing materials that exhibit unusual property combinations—particularly those with exceptional strength coupled with concerning brittleness—as well as materials with inconsistent hardness measurements across different testing methodologies. Understanding these outliers enables engineers to make more informed material selection decisions while recognizing dataset limitations.

2 Material Selection Insights

2.1 High Strength but Low Ductility Materials

The analysis identified 29 materials that combine exceptionally high strength (top 10% of ultimate tensile strength) with very low ductility (bottom 10% of elongation), representing critical outliers in the strength-ductility trade-off relationship.

Table 1: Notable High-Strength, Low-Ductility Material Outliers

Material	Su (MPa)	A5 (%)	Su/A5 Ratio	Material Standard
Steel SAE 5160	2220	4.0	555.0	SAE
Steel SAE 9255	2103	1.0	2103.0	SAE
Steel SAE 51440C	1965	2.0	982.5	SAE
Steel SAE 5150	1944	5.0	388.8	SAE
Steel SAE 51440A	1793	5.0	358.6	SAE
Steel 30Ch3S3GML	1766	4.0	441.5	GOST
CSN 12071	1770	2.0	885.0	CSN
Nodular cast iron 100	1000	2.0	500.0	GOST

Strength-Ductility Outlier Insights

The analysis of strength-ductility outliers reveals critical material selection considerations:

- **Extreme Brittleness Risk**: SAE 9255 exhibits the most concerning combination with exceptional strength (2103 MPa) but critically low ductility (1%), creating a Su/A5 ratio over 2000—making it extremely brittle and potentially catastrophic under impact loads.
- **Material Standard Patterns**: SAE standard steels dominate the high-strength but low-ductility category, suggesting this standardization group may prioritize strength over deformability in their specifications.
- **Material Diversity**: The outliers represent diverse material families including specialty GOST steels, CSN standards, and even a nodular cast iron—showing that brittle-strong properties cross multiple material classification systems.
- **Strength Ceiling**: The analysis reveals a mechanical property ceiling around 2200-2300 MPa for ultimate tensile strength beyond which materials become increasingly brittle, suggesting a fundamental materials science limitation.

2.2 Engineering Application Considerations

The extremely high-strength, low-ductility outlier materials require careful application consideration to prevent catastrophic failure:

Table 2: Application Guidance for Brittle-Strong Materials

Suitability Category	Appropriate Applications	Applications to Avoid
Static Loading	Precision tooling, wear-resistant surfaces, static load-bearing components	Impact loading conditions, cyclic loading applications
Design Factors	Components requiring dimensional stability, high hardness requirements, weight-critical structural components	Components with stress concentrators, energy-absorbing structures
Environmental Considerations	Low-temperature applications with minimal thermal cycling, controlled environments	Thermal cycling conditions, corrosive environments, high-vibration settings

Application-Specific Selection Guidance

Engineers should consider these guidelines when selecting high-strength, low-ductility materials:

- **Design Philosophy**: When using these brittle-strong outlier materials, engineers should adopt a zero-deformation design philosophy rather than relying on plastic deformation to absorb energy.
- **Post-Processing Options**: For applications requiring both high strength and improved ductility, post-processing techniques like specialized heat treatments or surface engineering may help mitigate brittleness while preserving strength.
- **Risk Mitigation**: When these materials must be used in dynamic applications, implement additional safety factors, protective structures, or regular inspection protocols to manage brittle fracture risks.
- **Material Substitution**: For critical safety components where brittleness poses unacceptable risks, engineers should consider substituting materials with moderately lower strength but significantly improved ductility.

3 Data Quality Observations

3.1 Hardness Measurement Inconsistency

The analysis of hardness measurement consistency revealed a significant data quality limitation:

Table 3: Hardness Measurement Data Quality Issues

Hardness Type	Data Availability	Quality Implications
Brinell Hardness (Bhn)	463 materials (29.8% of dataset)	Significant coverage gap
Vickers Hardness (HV)	165 materials (10.6% of dataset)	Limited representation
Combined Bhn and HV	0 materials (0% of dataset)	No cross-validation possible

Hardness Data Quality Implications

The analysis of hardness measurement data quality revealed several important considerations:

- **Complete Measurement Gap**: Despite the engineering importance of validating hardness measurements across different scales, the dataset contains no materials with both Brinell and Vickers hardness values recorded.
- **Mutual Exclusivity**: The data collection protocol appears to have treated hardness measurement methods as mutually exclusive rather than complementary, preventing cross-scale validation.
- **Limited Coverage**: With only 29.8% of materials having Brinell hardness data and 10.6% having Vickers hardness data, the majority of materials (59.6%) lack any hardness measurements.
- **Validation Concerns**: Without cross-validation between measurement techniques, the reliability of the hardness data cannot be fully assessed, potentially introducing risk into engineering decisions that depend heavily on hardness values.

3.2 Dataset Completeness Analysis

Database Structure Evaluation

Analysis of the dataset structure reveals important considerations for making engineering decisions:

- **Missing Value Patterns**: The data structure shows systematic missing value patterns that suggest different testing protocols or data sources were used for different material families.
- **Material Grouping**: The outlier analysis revealed that many extreme-property materials come from specific material standards (SAE, GOST, CSN), suggesting potential systematic differences in how properties are measured or reported across standards.

- **Verification Needs**: For critical applications using the identified outlier materials, independent verification testing would be advisable given the unusual property combinations and data quality limitations.
- **Root Cause Analysis**: The absence of cross-validation hardness measurements may indicate either testing protocol limitations or database compilation from multiple sources with different measurement standards.

4 Visualization and Pattern Detection

The visualization of strength-ductility relationships reveals clear patterns in material behavior and outlier positioning:

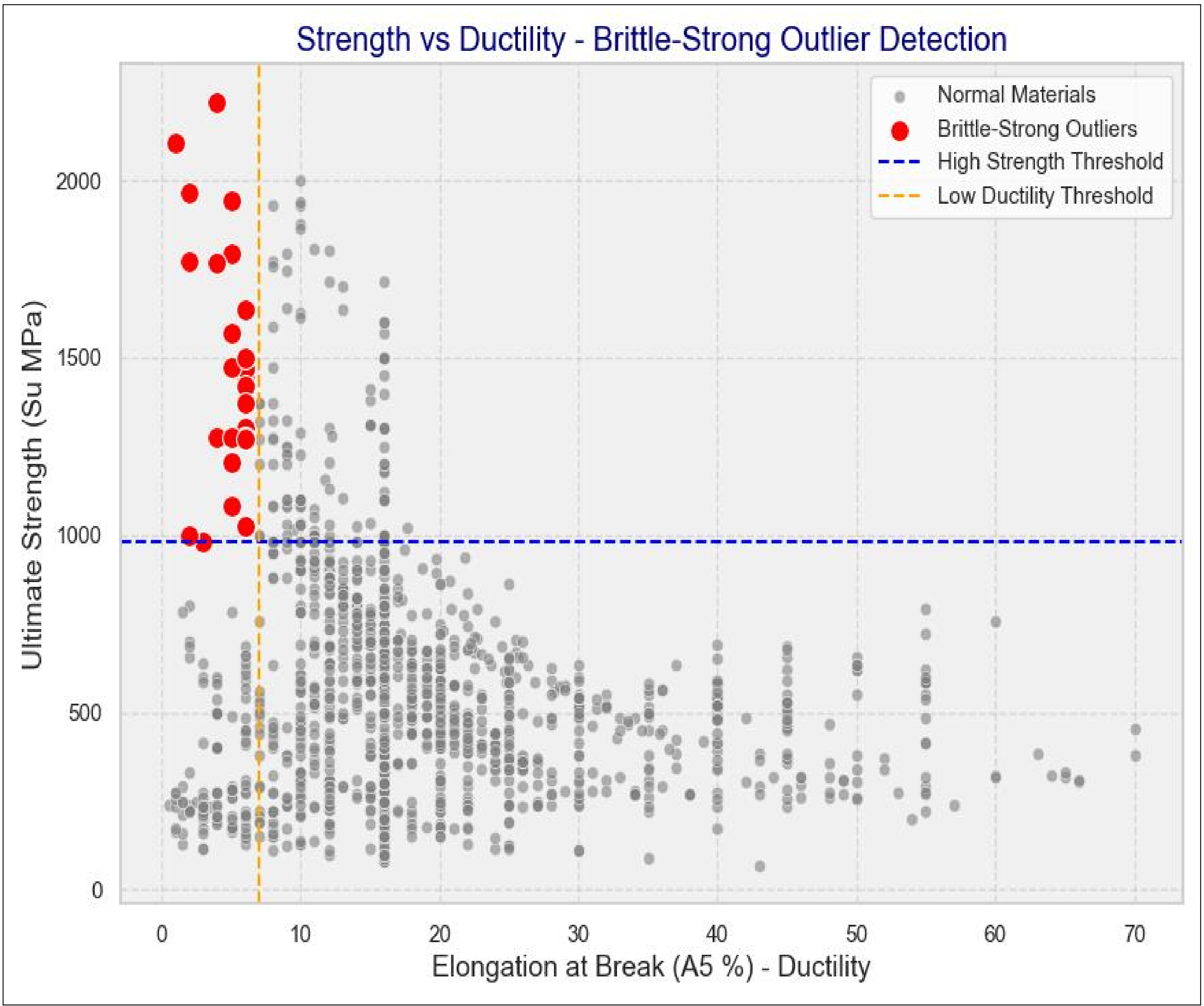


Figure 1: Strength vs Ductility Outlier Analysis

Strength-Ductility Visualization Insights

The scatterplot analysis of strength versus ductility reveals:

- **Hyperbolic Relationship**: The overall dataset demonstrates a clear hyperbolic relationship between ultimate tensile strength and elongation at break, highlighting the fundamental trade-off engineers must navigate.
- **Outlier Quadrant**: The high-strength, low-ductility materials occupy a distinct "danger quadrant" in the upper-left portion of the strength-ductility plot, making them easy to identify visually.
- **Clustering Effects**: The outlier materials exhibit clustering patterns by material standard and type, suggesting material family characteristics rather than individual anomalies.
- **Design Boundary Identification**: The plot effectively establishes visual thresholds at the 90th percentile for strength and 10th percentile for ductility, creating practical design boundaries for material selection decisions.

Statistical Distribution Insights

The statistical distribution of material properties among the outliers provides additional insights:

- **Extreme Value Analysis**: The strength values in the outlier group reach up to 2220 MPa, placing them 3.5 standard deviations above the dataset mean, confirming their statistical significance.
- **Ductility Floor**: Among the outliers, the ductility values reach as low as 1%, representing a practical lower limit for structural materials before they become essentially non-deformable.
- **Material Family Prevalence**: Within the outlier group, SAE standards represent 31% of materials, followed by CSN (17%) and BS (10%), providing valuable information about standards that tend to include brittle-strong materials.
- **Distribution Tails**: The extreme property combinations of the outliers represent the practical limits of the strength-ductility trade-off, defining the boundaries of what's currently achievable in material engineering.

5 Conclusions and Recommendations

The outlier analysis provides critical insights for engineering decision-making and material selection:

Key Outlier Analysis Takeaways

The analysis of material outliers yields several important conclusions:

- **Extreme Property Trade-offs**: The identified outlier materials demonstrate the extreme limits of the strength-ductility trade-off, with some materials offering exceptional strength at the cost of dangerous brittleness.
- **Application-Specific Selection**: High-strength, low-ductility materials should be selected only for specific applications where their property combination is appropriate, such as static loading conditions or wear-resistant surfaces.
- **Data Quality Limitations**: The complete absence of materials with both Brinell and Vickers hardness measurements represents a significant limitation for cross-validation of hardness properties.
- **Engineering Risk Management**: When selecting outlier materials, engineers should implement appropriate risk management strategies, including additional safety factors, protective structures, or regular inspection protocols.
- **Dataset Improvement Opportunities**: Future material testing and database development should include both hardness measurement types to enable cross-validation and improve data reliability.

5.1 Engineering Implications

The outlier analysis provides crucial guidance for engineering material selection decisions:

Practical Engineering Applications

Engineers can apply these findings in several practical ways:

- **Risk-Informed Selection**: When considering high-strength materials, engineers should explicitly evaluate the brittleness risk based on elongation values, especially for those identified as outliers.
- **Application Matching**: Match the outlier materials' unique property combinations to appropriate applications—leveraging their strengths while mitigating their inherent limitations.
- **Testing Protocol Enhancement**: For critical applications, implement supplementary testing protocols that include multiple hardness measurement techniques to address the cross-validation gap identified in the dataset.
- **Data-Driven Design**: Use the identified thresholds (90th percentile strength, 10th percentile ductility) as practical design boundaries when developing new components or selecting materials.

- **Standard-Specific Awareness**: Recognize that certain material standards (particularly SAE) may contain more brittle-strong outliers, requiring heightened attention to ductility considerations when selecting from these standards.

6 Appendix: Outlier Analysis Methodology

Analysis Process

The outlier identification process followed these steps:

1. **Threshold Determination**: Established the 90th percentile for strength and 10th percentile for ductility as outlier thresholds
2. **Multi-criteria Filtering**: Identified materials meeting both high-strength and low-ductility criteria simultaneously
3. **Cross-validation Attempt**: Attempted to identify materials with inconsistent hardness measurements (Bhn vs. HV)
4. **Data Completeness Assessment**: Evaluated the availability of both hardness measurements across the dataset
5. **Statistical Analysis**: Calculated key metrics including Su/A5 ratios to quantify brittleness
6. **Visual Pattern Recognition**: Plotted strength-ductility relationships to visually identify outlier clusters
7. **Engineering Interpretation**: Translated statistical findings into practical material selection guidance

PART 2 – Cross Dataset Engineering Tasks

Task 12: Material Descriptor Analysis

May 1, 2025

Abstract

This report presents a detailed analysis of engineering materials based on Task 12 (Material Descriptor Analysis) of the materials dataset. We examine the relationship between descriptive text data and material properties, focusing on keywords like "heat-resistant", "heat-treatment", and "corrosion-resistant". The analysis provides critical insights into material selection criteria based on text descriptions, revealing unique property profiles for different material categories and identifying specialized materials that combine multiple desirable properties. Data quality observations are also included to inform the reliability of engineering decisions based on text-based classification.

1 Introduction

The Value of Unstructured Text Data in Materials Selection

Understanding how text descriptors relate to material properties is increasingly important in modern engineering materials selection. This report analyzes the dataset by extracting insights from the Desc column, identifying keyword patterns, and correlating them with mechanical, physical, and chemical properties. By examining how property profiles vary across different descriptor categories, engineers can leverage text-based classification to make more informed material selection decisions, particularly for specialized applications requiring specific combinations of properties.

2 Material Selection Insights

2.1 Keyword Distribution Analysis

The analysis of descriptive keywords reveals significant patterns in the frequency and distribution of material categories:

Table 1: Material Distribution by Descriptor Keyword

Keyword Category	Material Count	Prevalence
Heat-treatment	132	Dominant category (56.4%)
Heat-resistant	66	Moderate presence (28.2%)
Corrosion-resistant	8	Specialized category (3.4%)
Others/Uncategorized	28	Miscellaneous (12.0%)

Key Insights from Keyword Distribution

The distribution analysis of descriptive terms reveals important patterns for engineering material classification:

- **Category Dominance**: Heat-treatment materials form the largest group (132 instances), indicating the prevalence of thermally processed materials in engineering applications.
- **Specialized Nature**: Corrosion-resistant materials form a small, specialized category (only 8 instances), suggesting these materials are more specialized and potentially higher-value.
- **Data Coverage**: Of 1552 original materials, only 981 have descriptive text data (63.2% coverage), limiting the comprehensiveness of text-based classification.
- **Descriptor Diversity**: The dataset contains 83 unique descriptor values, indicating a lack of standardized terminology that could impact classification consistency.

2.2 Property Profile by Material Category

Different descriptor categories exhibit distinct property profiles, creating specific performance characteristics:

Table 2: Average Material Properties by Descriptor Category

Category	Su (MPa)	Sy (MPa)	A5 (%)	E (MPa)	G (MPa)	mu	Ro (kg/m
Heat-treatment	807.8	589.3	14.7	206,000	80,000	0.300	7,860
Heat-resistant	589.3	345.8	24.1	205,667	79,803	0.300	7,862
Corrosion-resistant	775.0	476.0	25.5	203,250	78,375	0.299	7,874

Property Profile Insights

The analysis of property variations across descriptor categories reveals critical patterns for materials selection:

- **Strength Leadership**: Heat-treatment materials demonstrate superior ultimate tensile strength (Su = 807.8 MPa) and yield strength (Sy = 589.3 MPa), making them optimal for load-bearing applications.
- **Ductility Advantage**: Both heat-resistant and corrosion-resistant materials exhibit significantly higher elongation values (A5 = 24.1% and 25.5% respectively) compared to heat-treatment materials (A5 = 14.7%), making them better suited for applications requiring formability.
- **Modulus Consistency**: Young's modulus (E) and shear modulus (G) remain relatively consistent across all categories, indicating similar elastic behavior regardless of material category.
- **Density Variation**: Corrosion-resistant materials have slightly higher density (Ro = 7,874 kg/m^3) compared to other categories, potentially influencing weight-critical applications.

2.3 Multi-Property Materials Analysis

Analysis of material descriptor overlaps identifies specialized materials with multiple desirable properties:

Table 3: Special Materials with Both Heat and Corrosion Resistance

Material	Key Properties	Application Suitability
Steel 13Ch14N3V2FR GOST 5949-75	High Su, High Sy, Strong elasticity	Turbine disks, blades (high-load, high-temp)
Steel 10Ch11N23T3MR GOST 5949-75	Moderate-High strength, High density	Steam lines, corrosive + thermal environments
Steel 12Ch18N10T GOST 5949-75	Low strength, Very high A5, Superior E	Chemical vessels, food-grade equipment
Steel 12Ch18N9T GOST 5949-75	Slightly higher strength than 10T	Heat exchangers, moderate-load applications
Steel 12Ch18N12T GOST 5949-75	Balanced profile similar to 9T	Cryogenic vessels, acid piping
Steel 12Ch25N16G7AR GOST 5949-75	Moderate strength, Very high A5, Best E	Acidic reactors, marine exhausts

Engineering Insights on Multi-Property Materials

The overlap analysis identifies critical materials with combined properties:

- **Material Classification**: All six materials with both heat and corrosion resistance are Soviet/Russian austenitic stainless steels standardized under GOST 5949-75.
- **Alloying Elements**: These materials contain high levels of chromium (Ch), nickel (N), and specialized elements like titanium (T), vanadium (V), and boron (R), contributing to their exceptional property combinations.
- **Property Range**: Within this elite group, materials like Steel 13Ch14N3V2FR offer superior strength, while others like 12Ch18N10T provide exceptional ductility—enabling application-specific selection.
- **Heat Treatment Sensitivity**: Several materials (e.g., 13Ch14N3V2FR, 10Ch11N23T3MR) appear with multiple property profiles, indicating their properties can be tuned through different heat treatments.

3 Application-Specific Recommendations

Based on the material descriptor analysis, specific materials can be recommended for various engineering applications:

Table 4: Application Recommendations Based on Material Descriptors

Application	Recommended Category	Specific Material Examples
High-load, high temperature	Heat+Corrosion Resistant	Steel 13Ch14N3V2FR GOST 5949-75
Chemical processing vessels	Corrosion Resistant	Steel 12Ch18N10T GOST 5949-75
Power generation components	Heat Resistant	Steel 10Ch11N23T3MR GOST 5949-75
Structural components	Heat Treatment	Heat-treatment steel variants
Marine applications	Corrosion Resistant	Steel 12Ch25N16G7AR GOST 5949-75
Aerospace components	Balanced properties	Steel 12Ch18N12T GOST 5949-75

Application-Specific Selection Criteria

The descriptor analysis supports application-specific material selection through these insights:

- **Environmental Challenge**: For combined thermal and corrosive environments (e.g., chemical processing, power generation), multi-property materials from the overlapping category are ideal choices.

- **Property Prioritization**: Applications can be matched with materials based on their descriptor-associated property profiles—strength-dominated (heat-treatment), ductility-dominated (heat-resistant), or balance-focused (corrosion-resistant).
- **Specialized Applications**: For critical applications like aerospace components or medical devices, materials with carefully balanced properties like Steel 12Ch18N12T offer optimal performance.
- **Cost-Performance Balance**: For less demanding applications, standard heat-treatment materials provide adequate performance at potentially lower cost than specialized heat-resistant or corrosion-resistant alternatives.

4 Data Quality Observations

4.1 Descriptor Coverage and Standardization

The descriptor analysis reveals several data quality issues affecting the comprehensiveness and reliability of text-based classification:

Table 5: Data Quality Issues in Material Descriptors

Issue	Observation	Impact on Analysis
Descriptor Coverage	36.8% of materials lack descriptors	Limited comprehensiveness of text analysis
Descriptor Standardization	83 unique descriptor values present	Potential inconsistency in classification
Property Completeness	Bhn values missing for many materials	Incomplete property comparisons
Heat Treatment Variation	Multiple entries for same material	Complicates material identification

4.2 Material Property Variations and Consistency

Property Measurement and Reporting Issues

Several observations regarding material property data quality emerged from the descriptor analysis:

- **Missing Hardness Data**: Brinell Hardness (Bhn) values appear to be missing or insignificant across analyzed descriptor categories, limiting the comparison of surface mechanical properties.

- **Property Variation with Heat Treatment**: Materials like Steel 13Ch14N3V2FR show property variations with different heat treatments, introducing multiple data points for single materials and complicating classification.
- **Descriptor Inconsistency**: The lack of standardized terminology (83 unique descriptors) suggests potential for classification errors or inconsistencies in material grouping.
- **Missing Materials**: With 36.8% of materials lacking descriptors, certain material categories may be underrepresented in the analysis, potentially skewing the observed patterns.

5 Engineering Implications

The material descriptor analysis provides valuable guidance for engineering design and material selection decisions:

Practical Applications of Text-Based Material Selection

Engineers can leverage text descriptors in several practical ways:

- **Selection Trade-offs**: The clear strength-ductility trade-off observed between heat-treatment materials (higher strength, lower ductility) and heat/corrosion-resistant materials (lower strength, higher ductility) provides a framework for balancing competing requirements.
- **Material Tunability**: The observation that several materials exhibit property variations with different heat treatments demonstrates the potential for fine-tuning material performance to specific operational conditions.
- **Specialized Applications**: The identification of rare materials with both heat and corrosion resistance provides valuable options for extreme environments, though these should be reserved for critical applications due to likely higher cost.
- **Text-Based Pre-screening**: Descriptive text provides a valuable first-pass filter for material selection, enabling engineers to quickly narrow the field of candidates before detailed property comparison.

6 Conclusions and Recommendations

The material descriptor analysis yields several key conclusions for engineering material selection:

Key Descriptor Analysis Takeaways

The analysis of materials by descriptive text reveals:

- Descriptive terminology provides valuable classification information that complements numerical properties, demonstrating the importance of unstructured data in technical domains.
- Heat-treatment materials offer superior strength properties, heat-resistant materials provide enhanced ductility, and corrosion-resistant materials deliver specialized environmental performance—creating distinct selection categories.
- Six special austenitic stainless steels combine both heat resistance and corrosion resistance, providing elite options for extreme environments requiring multiple property advantages.
- Data quality issues, particularly missing descriptors and hardness values, require careful consideration when making engineering decisions based on text categorization.
- Text-based material selection offers a complementary approach to traditional property-based selection, potentially streamlining the identification of materials for specialized applications.

6.1 Future Analysis Recommendations

To extend the value of material descriptor analysis, several future analytical approaches are recommended:

Advancing Text-Based Materials Analysis

Further analysis could enhance the value of material descriptors through:

- **Expanded Keyword Analysis**: Extend the analysis to additional descriptive terms beyond the three primary keywords to identify other specialized material categories.
- **Correlation Analysis**: Perform statistical analysis to quantify the relationships between specific descriptive terms and material properties, enabling more precise prediction of performance.
- **Natural Language Processing**: Apply more sophisticated text analysis techniques to extract patterns from unstructured material descriptions, potentially uncovering hidden relationships.
- **Descriptor Standardization**: Develop a standardized classification system for material descriptions to improve consistency and comparability across the dataset.

- **Property Gap Analysis**: Identify specific properties missing from different descriptor categories to guide future data collection efforts and improve analysis completeness.

7 Appendix: Material Descriptor Analysis Methodology and PLots

Analysis Process

The material descriptor analysis followed these steps:

1. **Data Preparation**: Filtering the dataset to include only rows with non-null descriptor values (981 of 1552 materials).
2. **Keyword Identification**: Selecting three primary keywords for analysis: "heat-resistant", "heat-treatment", and "corrosion-resistant".
3. **Category Formation**: Creating material categories based on the presence of these keywords in the descriptor field.
4. **Property Aggregation**: Calculating mean values for mechanical and physical properties (Su, Sy, A5, E, G, mu, Ro) within each category.
5. **Overlap Analysis**: Identifying materials belonging to multiple keyword categories to find multi-property materials.
6. **Visualization**: Creating radar charts and Venn diagrams to visualize property differences and category relationships.
7. **Application Mapping**: Connecting material categories and their property profiles with engineering application requirements.

Radar Chart Insights Summary

The following table summarizes which material group "wins" in each property, based on radar chart median comparisons across groups (e.g., heat-treated, corrosion-resistant).

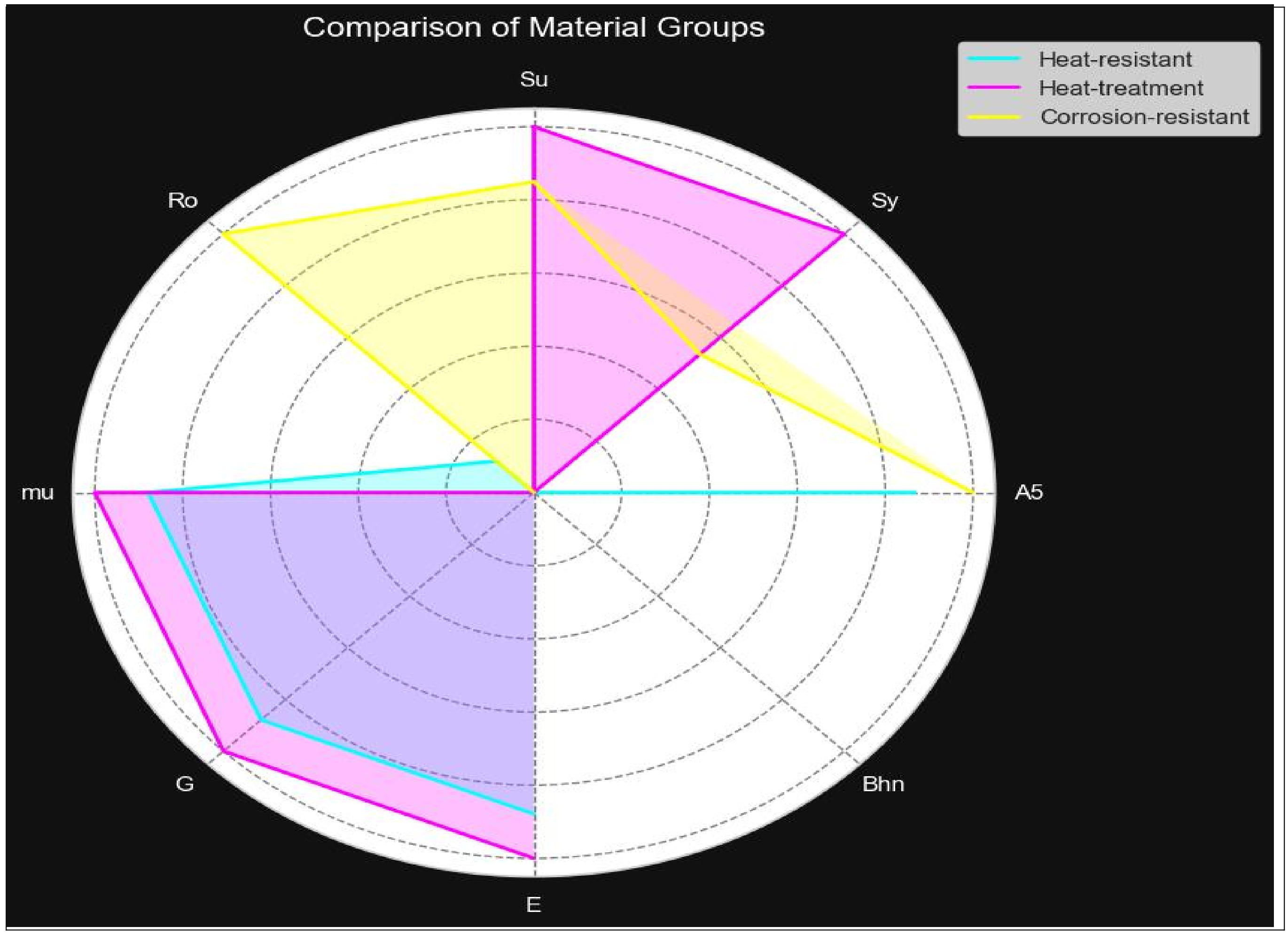


Figure 1: Comparision of Material Groups

Property	Who Wins?	Comment
Su (Ultimate Strength)	Heat-treated (magenta)	Strongest — clear Su advantage after heat treatment
Sy (Yield Strength)	Heat-treated (magenta)	Again dominant in yield strength, showing great plastic resistance
A5 (Elongation)	Corrosion-resistant (yellow)	Most ductile — suitable for flexible/deformable applications
Bhn (Brinell Hardness)	None	Data near zero or missing — not useful in current chart
E (Young's Modulus)	Heat-treated (magenta)	Highest stiffness — resists elastic deformation better
G (Shear Modulus)	Heat-treated (magenta)	Strong shear resistance — good for torsional loads
μ (Poisson's Ratio)	Heat-treated (magenta)	Slightly better transverse strain behavior than others
Ro (Density)	Corrosion-resistant (yellow)	Heaviest group — higher density materials dominate

Venn Diagram Interpretation & Overlap Insights

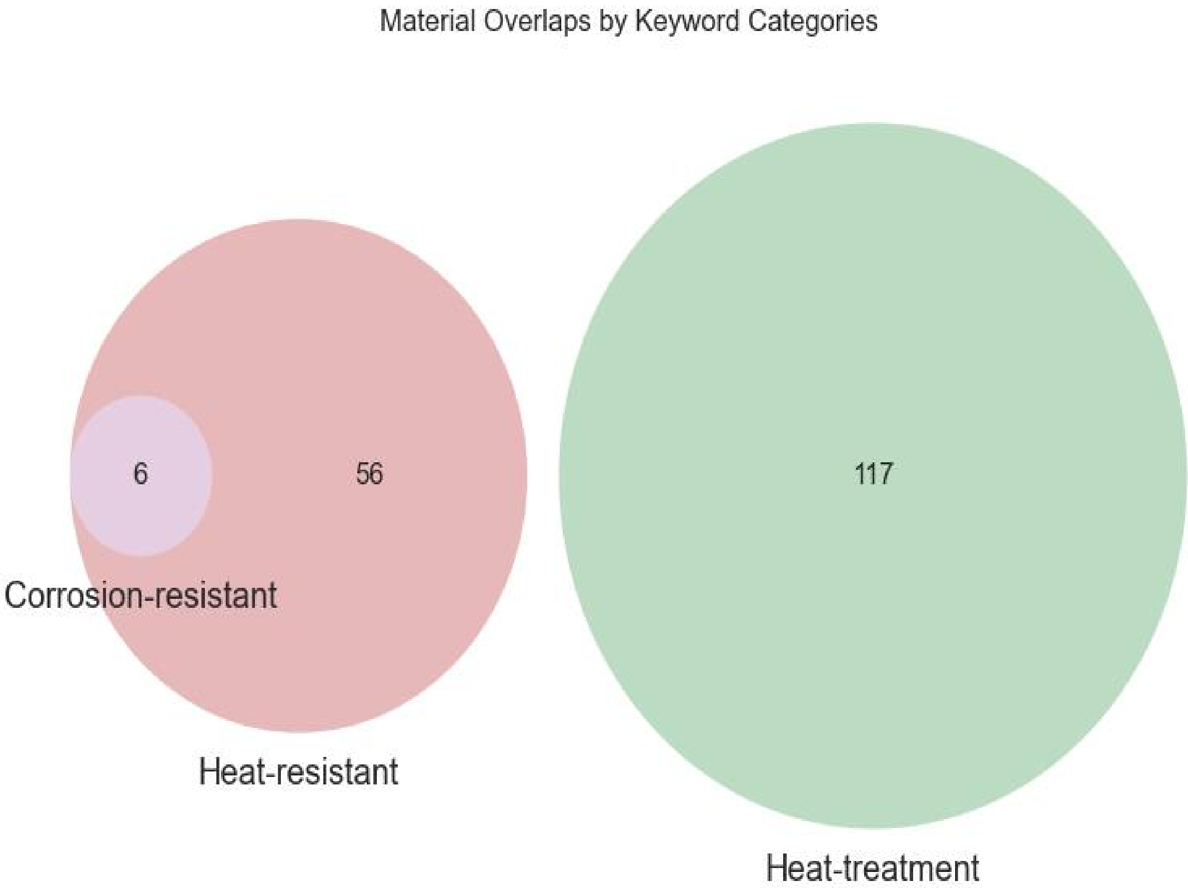


Figure 2: Material Overlap by Keyword Categories

Group Size Summary:

- **Heat-treatment steels:** Largest group (117 materials).
- **Heat-resistant steels:** Mid-size group (56 materials).
- **Corrosion-resistant steels:** Smallest group (6 materials).

Overlap Behavior:

- 6 materials are **both** heat- and corrosion-resistant.
- **Heat-treatment** group is mostly separate — no significant overlap.
- **Corrosion-resistant steels** are a tiny subset of heat-resistant steels.

Engineering Meaning:
Most corrosion-resistant steels are also made to resist heat — a known characteristic of **austenitic stainless steels**, especially in high-performance applications.

Overlapping Steels: Heat & Corrosion Resistant

These 6 materials appear in both categories:

- Steel 10Ch11N23T3MR GOST 5949-75
- Steel 12Ch18N10T GOST 5949-75
- Steel 12Ch18N12T GOST 5949-75
- Steel 12Ch18N9T GOST 5949-75
- Steel 12Ch25N16G7AR GOST 5949-75
- Steel 13Ch14N3V2FR GOST 5949-75

Alloy Symbols:
Ch = Cr, **N = Ni**, **T = Ti**, **G = Mn**, **R = B**, **V = V**
These steels are part of the Soviet/Russian GOST 5949-75 standard and are comparable to Western stainless grades like **AISI 321, 316Ti, 347**.

Engineering Properties to Expect

Property	Behavior
Strength	Excellent at elevated temperatures
Ductility	Good — retains toughness at low temperatures
Corrosion Resistance	Excellent — withstands acids, salt spray, chemicals
Heat Resistance	Very good — due to Cr and Ni content
Weldability	Generally good — Ti used for stabilization
Hardening	Not hardenable by heat — only by cold working

Engineering Summary:
If your application demands **both corrosion and heat resistance**, these 6 steels are top-tier candidates. Added elements like Ti, V, and B enhance:

- Intergranular corrosion resistance
- High-temperature creep strength
- Low-temperature impact toughness

Material-by-Material Engineering Insights

Steel 13Ch14N3V2FR: Excellent Su and Sy, average elasticity (E/G), high Poisson's ratio.
Use Case: Turbine disks, high-load/high-temp components.

Steel 10Ch11N23T3MR: Moderate-to-high strength, balanced elasticity, heavier density.
Use Case: Steam lines, thermal + corrosive environments.

Steel 12Ch18N10T: Low strength but extremely ductile. Very high E and G.
Use Case: Chemical tanks, corrosion-focused environments.

Steel 12Ch18N9T: Slightly stronger than 10T, also highly ductile.
Use Case: Food-grade equipment, exchangers.

Steel 12Ch18N12T: Matches 9T, good balanced profile.
Use Case: Acidic lines, cryogenic vessels.

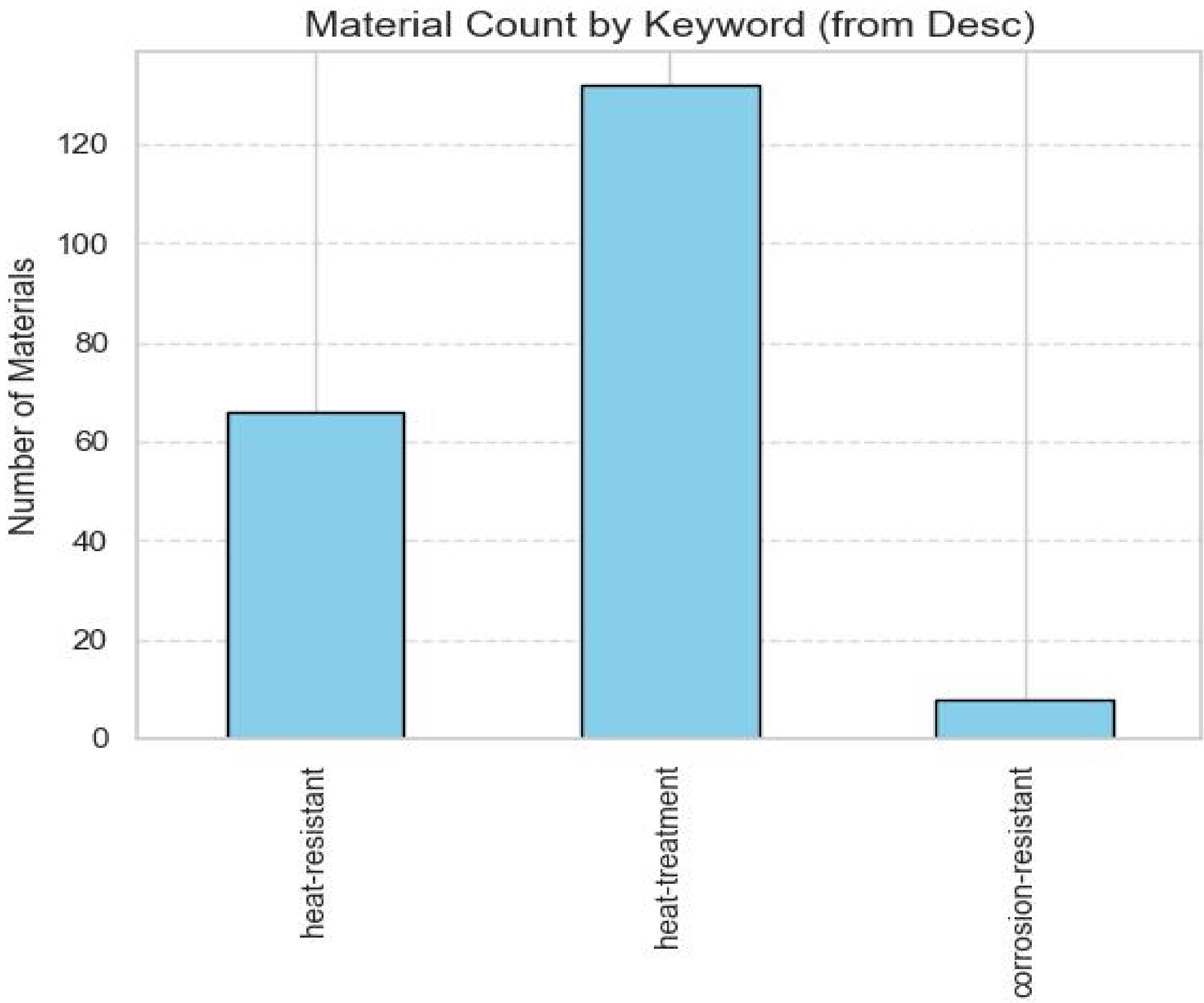
Steel 12Ch25N16G7AR: Moderate strength, highest ductility, top E and G.
Use Case: Marine/acid reactors, exhaust systems.

Variation by Heat Treatment:

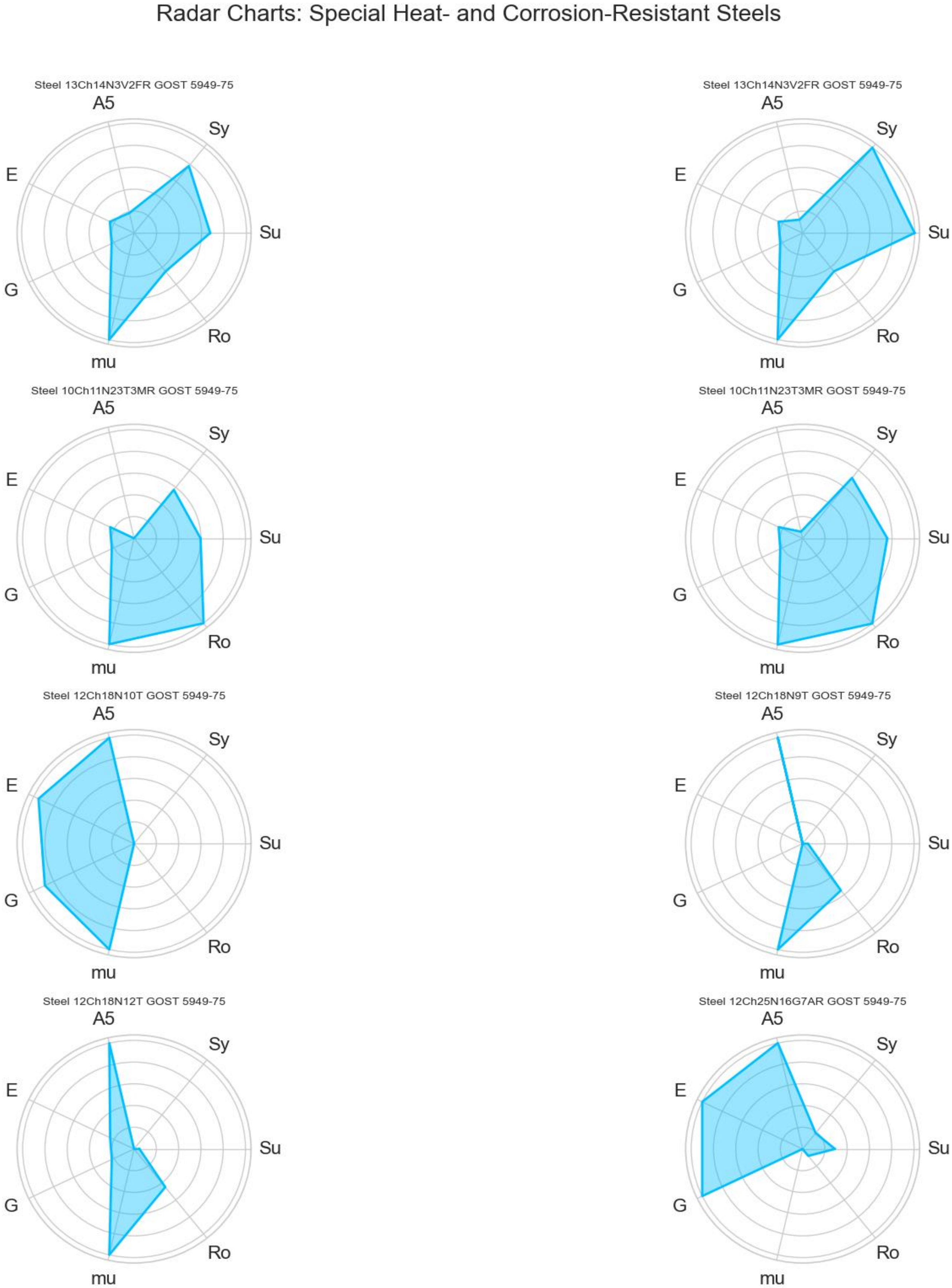
- **13Ch14N3V2FR:** Appears twice due to tempering difference (580°C vs 680°C). Stronger Sy in one variant.
- **10Ch11N23T3MR:** Also shown twice — high quench and temper (1100°C → 730–800°C) affects ductility/Su.

Summary Comparison Chart (8 Entries)

Steel	Strength	Ductility	Corrosion	Heat	Notes
13Ch14N3V2FR (1)	★★★★	★★★	★★★★	★★★★★	Top performance; gr temp loads
13Ch14N3V2FR (2)	★★★★	★★★	★★★★	★★★★★	Sy slightly low treatment effect vis
10Ch11N23T3MR (1)	★★★★	★★	★★★★	★★★★	Su+Sy stable; hea duty suitability
10Ch11N23T3MR (2)	★★★★	★★	★★★★	★★★★	Similar; more ducti sion 1
12Ch18N10T	★★	★★★★★	★★★★★	★★★★	Best ductility; grea resistance
12Ch18N9T	★★	★★★★	★★★★	★★★	Slightly stronger th
12Ch18N12T	★★	★★★★	★★★★	★★★★	Balanced; good cry performance
12Ch25N16G7AR	★★★	★★★★	★★★★★	★★★★	Best corrosion; wel tic



Final Engineering Takeaways



(Radar chart comparing all 8 materials across Su, Sy, A5, E, G, μ, Ro)

Heat Treatment Sensitivity:
Steels like 13Ch14N3V2FR and 10Ch11N23T3MR show property shifts depending on tempering. Enables tailored performance.

Elite Corrosion Resistance:
12Ch18N10T, 12Ch25N16G7AR, and 12Ch18N9T offer excellent corrosion resistance for marine/chemical/food use.

Formability–Strength Tradeoff:
As Su/Sy increases, A5 generally drops. Decide if flexibility or max load is your main concern.

Density & Cost:
Highly alloyed steels (e.g., 10Ch11N23T3MR) are heavier and pricier — best reserved for thermal/chemical extremes.