**Engineering Materials – A Data-Driven Approach to Mechanical, Physical & Chemical Properties**

**Dataset 1 – Detailed Properties**

Mechanical, physical, and environmental properties with structured metadata

|  |  |  |  |
| --- | --- | --- | --- |
| **Abbreviation** | **What It Means** | **Simple Description** | **Real-Life Example** |
| Std | Standard | A rule or guideline everyone follows for parts or materials. | Using ANSI bolts to fit engine parts correctly. |
| ID | Identification Number | Identification Number | Identification Number |
| Material | Material | The stuff a part is made of. | Aluminium used for lightweight car wheels. |
| Heat treatment | Heat Treatment | Heating and cooling metal to make it stronger or tougher. | Gears that are hardened to last longer. |
| Su | Ultimate Strength | The most pulling force a material can handle before breaking. | Strong chassis frames made to survive crashes. |
| Sy | Yield Strength | When a material starts to bend and doesn't return to shape. | Control arms built to resist bending during bumps. |
| A5 | Elongation at Break | How much a material can stretch before snapping. | Car bumpers made from metals that can stretch in accidents. |
| Bhn | Brinell Hardness Number | How hard the surface of a material is when pressed? | Truck axles tested for surface hardness. |
| E | Young’s Modulus | How stiff a material is when you pull or push it. | Steel car frames that don't easily bend. |
| G | Shear Modulus | How stiff a material is when you twist it? | Driveshafts designed to resist twisting forces. |
| mu (μ) | Coefficient of Friction | How slippery or sticky two surfaces are. | Tires designed with a high grip for dry roads. |
| Ro (ρ) | Density | How heavy something is for its size. | Aluminium panels are light but strong, helping fuel economy. |
| pH | Acidity/Alkalinity Level | Shows if a fluid is more acidic or basic. | Coolant tested to keep engine parts from corroding. |
| Desc | Description | A short explanation about a part. | Labelling a part as "Forged steel crankshaft" in specs. |
| HV | Vickers Hardness | A way to measure hardness using a sharp diamond point. | Checking hardness of a camshaft after heat treatment. |

**Std, Material, Desc** → Tell you what a part is.

**Su, Sy, A5, Bhn, E, G, HV** → Tell you how strong or tough the material is.

**Ro, mu, pH** → Tell you physical properties important for real-world performance.

**Dataset 2 – Simplified View + Use Flag**

Flattened format with “Use = True/False” label indicating suitability for engineering application

|  |  |  |  |
| --- | --- | --- | --- |
| **Abbreviation** | **What It Means** | **Simple Description** | **Real-Life Automotive Example** |
| Material | Material | The type of stuff a part is made from. | Steel used for car body panels. |
| Su | Ultimate Strength | The maximum force a material can take before it breaks. | Crash bars built with high Su steel to survive impacts. |
| Sy | Yield Strength | The point where a material starts to bend and stays bent. | Suspension arms designed with high Sy to avoid bending on rough roads. |
| E | Young’s Modulus | How stiff a material is when stretched or squeezed? | Chassis frames made stiff to keep the car stable. |
| G | Shear Modulus | How stiff a material is when twisted. | Driveshafts designed to resist twisting during acceleration. |
| mu (μ) | Coefficient of Friction | How much two surfaces resist sliding against each other. | Tires made with high μ for better grip on dry pavement. |
| Ro (ρ) | Density | How heavy something is compared to its size. | Using aluminium instead of steel to make car parts lighter. |
| Use Flag | Use Flag | A marker showing if a material or part is allowed for use. | A material marked as “Approved” for making door beams. |

**PART 1 – Single Dataset Engineering Tasks (Dataset 1)**

**Task 1: Initial Exploration & Summary**

1. Identify the total number of materials and heat treatment types.
   1. Total number of materials in df1: **1552**
   2. Total number of unique materials in df1: **1225**
   3. Total number of heat treatments in df1: **1552**
   4. Total number of unique heat treatments in df1: **44**
2. Check for any missing or inconsistent data values.
3. **Missing Values per Column:** Heat treatment 750, A5 206, Bhn 1089, pH 1359, Desc 571, HV 1387
4. No duplicates.
5. **Columns containing outliers:** ['Su', 'Sy', 'A5', 'Bhn', 'G', 'mu', 'Ro', 'pH', 'HV']
6. **Number of outliers in each column:** Su: 73 outliers Sy: 96 outliers A5: 151 outliers Bhn: 463 outliers G: 51 outliers mu: 97 outliers Ro: 311 outliers pH: 193 outliers HV: 165 outliers
7. Summarize key statistics of mechanical properties like Su (ultimate tensile strength), Sy (yield strength), and A5 (elongation at break).
8. **Handle Missing Values:** Mean for numerical columns and ‘Unknown’ for descriptive column.
9. **Handle Outliers:** Using IQR – Capped lower bound to values below lower bound and similarly upper bound to values above upper bound.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **Before Cleaning** | **After Cleaning** | **What Changed? (Interpretation)** |
| **Mean** | 572.75 | 559.05 | Slight decrease — outliers pulled the average down. |
| **Standard Deviation** | 326.83 | 285.05 | Less spread — data is now more consistent. |
| **Minimum** | 69.00 | 69.00 | No change — weakest material remains. |
| **Maximum** | 2220.00 | 1252.50 | Capped — extremely high values adjusted using IQR cap. |

1. **Su (ultimate tensile strength) –** Before cleaning the data, the average material looked very strong but some values were very high like 2220. After handling the outliers, the average and standard deviation dropped providing a consistent and reliable value. On handling the extreme values, it helped to make a realistic decision.
2. **Sy (yield strength) -** Yield Strength is the stress level at which a material **starts to deform permanently.**

*The source data had few anomalies 280 max, 240 max, 210 max, 250 max, 210 max, 280 max, 240 max, 25 max and the datatype were also not integer. So first these were first cleaned and converted to integer before running the stats.*

**Before cleaning,** the data included a max yield strength of **2048** — extremely high, likely **exotic** or **erroneous.** This could mislead an engineer into thinking super-strong materials are standard. After cleaning, the max was capped at **867.5**, which is more realistic.

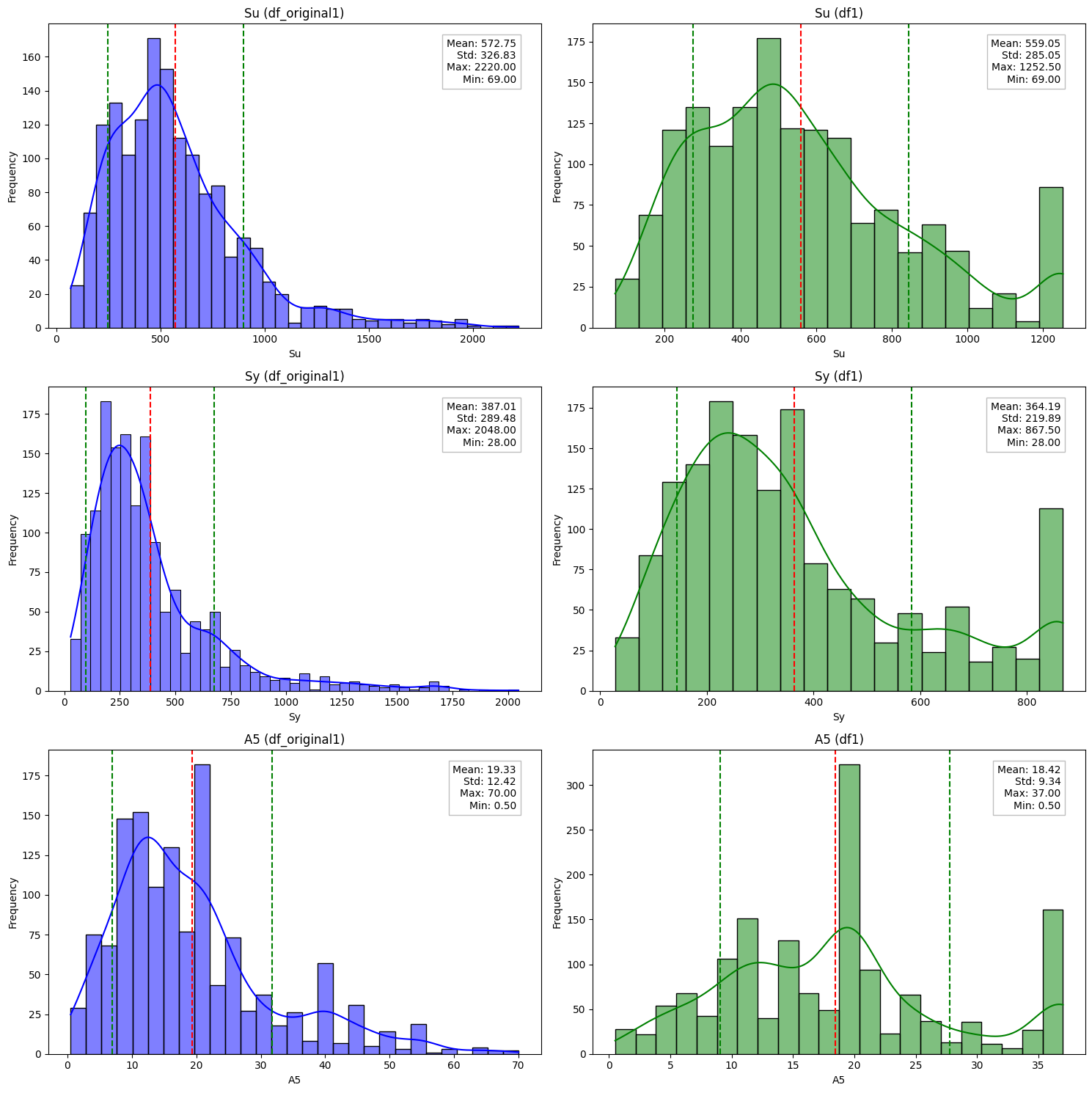
|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **Before Cleaning** | **After Cleaning** | **What Changed?** |
| **Mean** | 387.01 | 364.19 | Average lowered slightly after removing extreme highs. |
| **Standard Deviation** | 289.48 | 219.89 | Data spread reduced — results more consistent. |
| **Minimum** | 28.00 | 28.00 | Weakest material retained. |
| **Maximum** | 2048.00 | 867.50 | High outliers capped using IQR — unrealistic values removed. |

1. **A5 (elongation at break) -** It tells us **how much a material can stretch before it breaks,** expressed as a **percentage of its original length**. A higher A5 means the material is more **ductile** (can deform more before failing), while a lower value means it's **brittle**.

Before cleaning, we saw a max stretch of 70% — which is very high and might not reflect realistic metal behaviour. It could be a soft polymer or even an error. After capping outliers, the max value became 37%, which is more typical for high-ductility metals used in crash-absorbing components.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistic** | **Before Cleaning** | **After Cleaning** | **What Changed? (Simplified)** |
| **Mean** | 19.33 | 18.42 | Average slightly lowered — extreme values removed. |
| **Standard Deviation** | 12.42 | 9.34 | Variation reduced — values more consistent. |
| **Minimum** | 0.50 | 0.50 | No change — still includes very brittle materials. |
| **Maximum** | 70.00 | 37.00 | Unrealistically high stretch values capped using IQR. |

**Original data vs Cleaned data**

****

**Task 2: Groupwise Comparison**

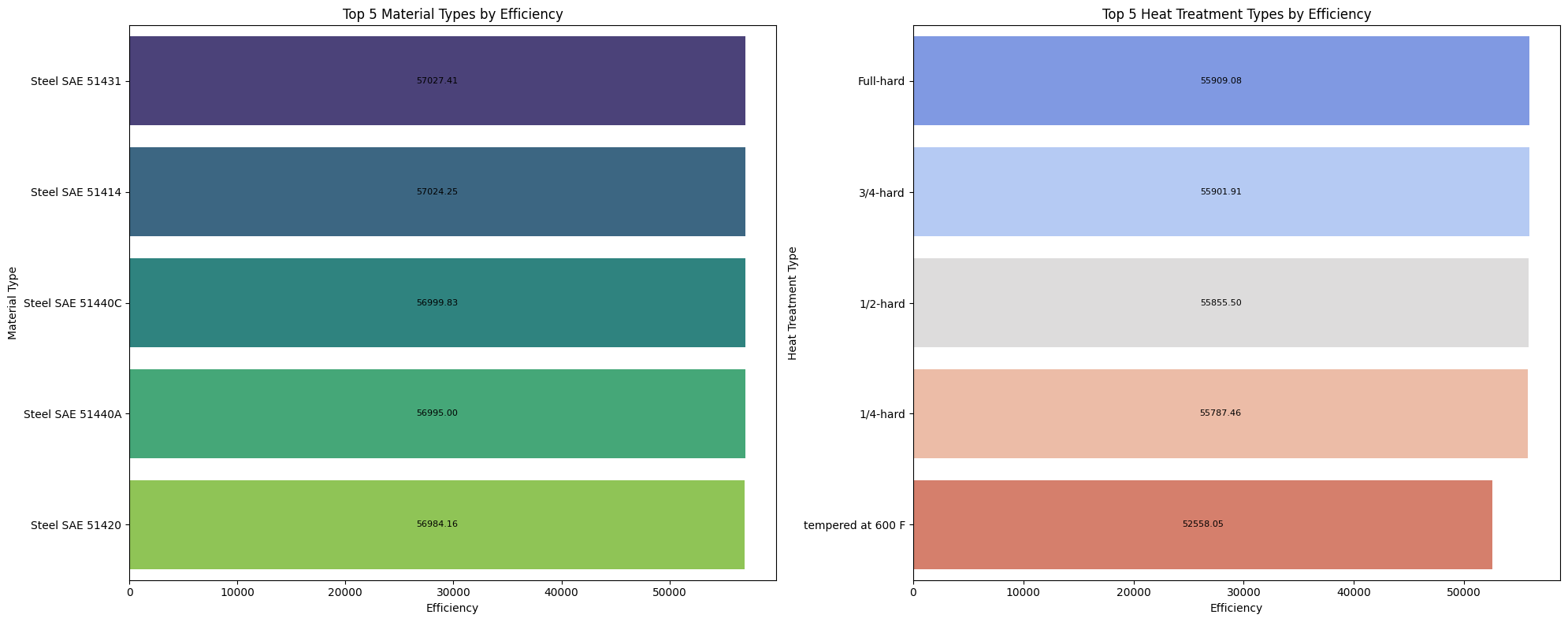
Compare average strength, ductility, and hardness values grouped by:

1. Material type
2. Heat treatment method

On grouping the Material Types and Heat Treatment Types **separately** with respect to Su, Sy, A5, E, G and HV,

**Material - Steel SAE 51431** has the highest efficiency of 57027.414141 and

**Heat treatment - Full-hard** has the highest efficiency of 55909.080808.



But when calculating the efficiency of combined grouping of Material Types and Heat Treatment Types with respect to Strength, Ductility and Hardness, below are the insights

Top Material and Heat Treatment Types by Strength Efficiency (Su, Sy, E, G):

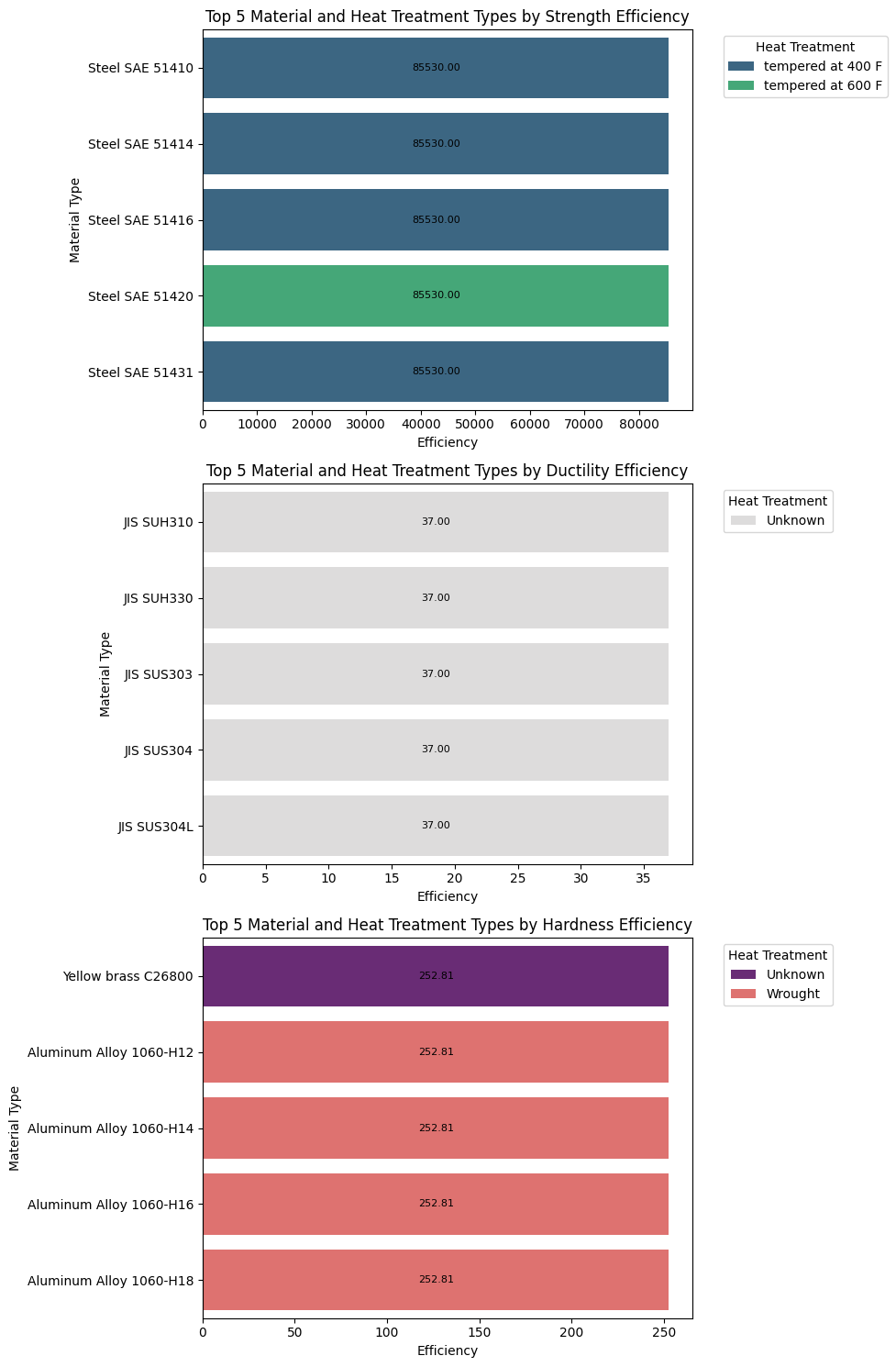
**Steel SAE 51410 - tempered at 400 has the highest efficiency of 85530.0**

Top Material and Heat Treatment Types by Ductility Efficiency (A5):

**JIS SUH310 – Unknown has the highest efficiency of 37.0**

Top Material and Heat Treatment Types by Hardness Efficiency (Bhn, HV):

**Yellow brass C26800 – Unknown has the highest efficiency of 252.811539**

****