Rouge Score Statistics by Model

- Median
- Mean

Claude 3.7 Sonnet — Mean: 0.498, Median: 0.508, Std: 0.197

Llama 3.1 8B — Mean: 0.177, Median: 0.161, Std: 0.083

Llama 3.3 70B — Mean: 0.321, Median: 0.327, Std: 0.144

Ministral 8B — Mean: 0.487, Median: 0.537, Std: 0.169

Qwen 2.5 72B — Mean: 0.490, Median: 0.490, Std: 0.185