G-Eval Score Distribution by Model and Prompt Type (soundness)