Compiled successes by parameter count across LLMs (baseline filtered)