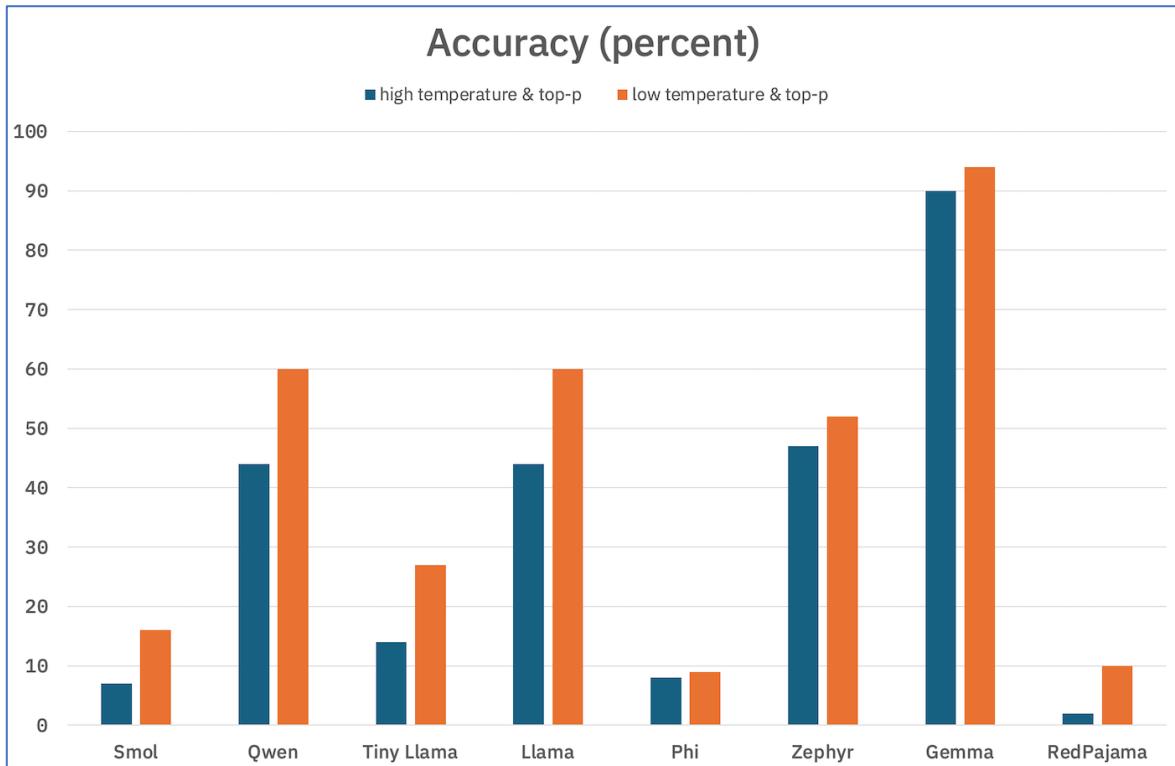


Teacher's Notes: "Benchmark"

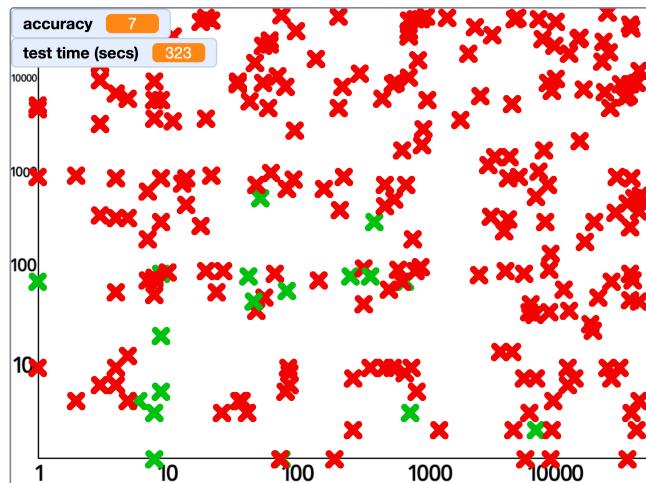
Your students can create graphs to understand the results of their testing.

Their testing will not necessarily produce identical results to these (especially for time, where computer hardware makes a huge difference), but here is an example of what it could look like:



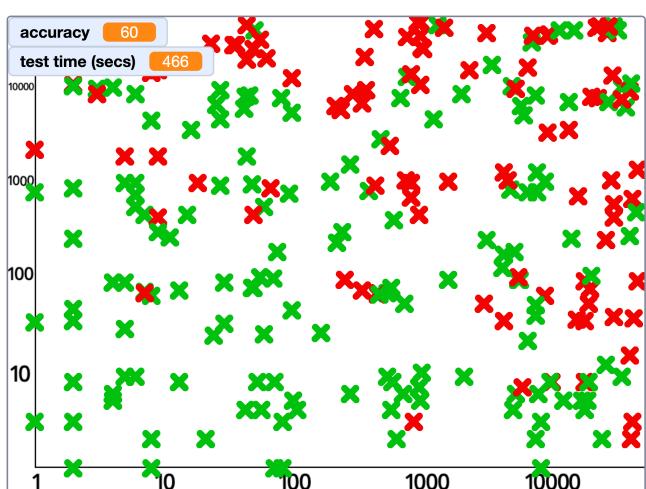
High temperature and Top-P

Smol

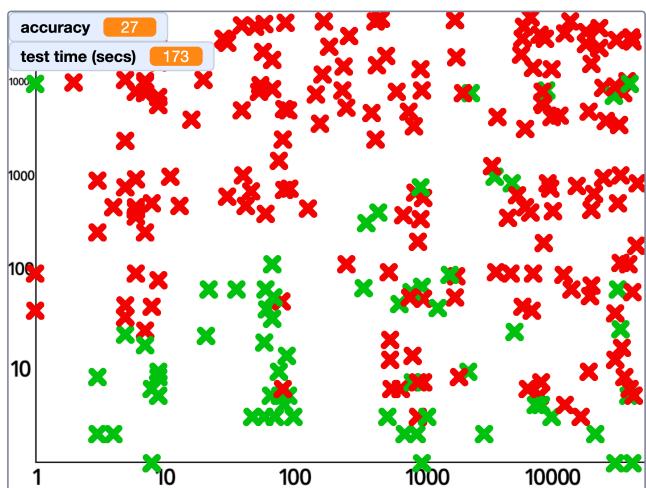
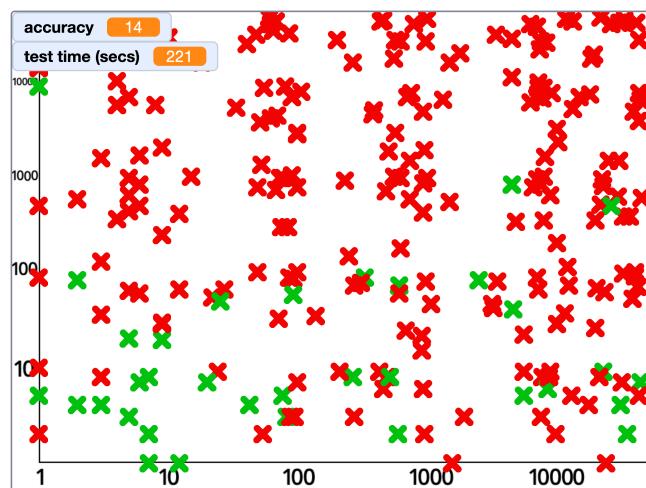


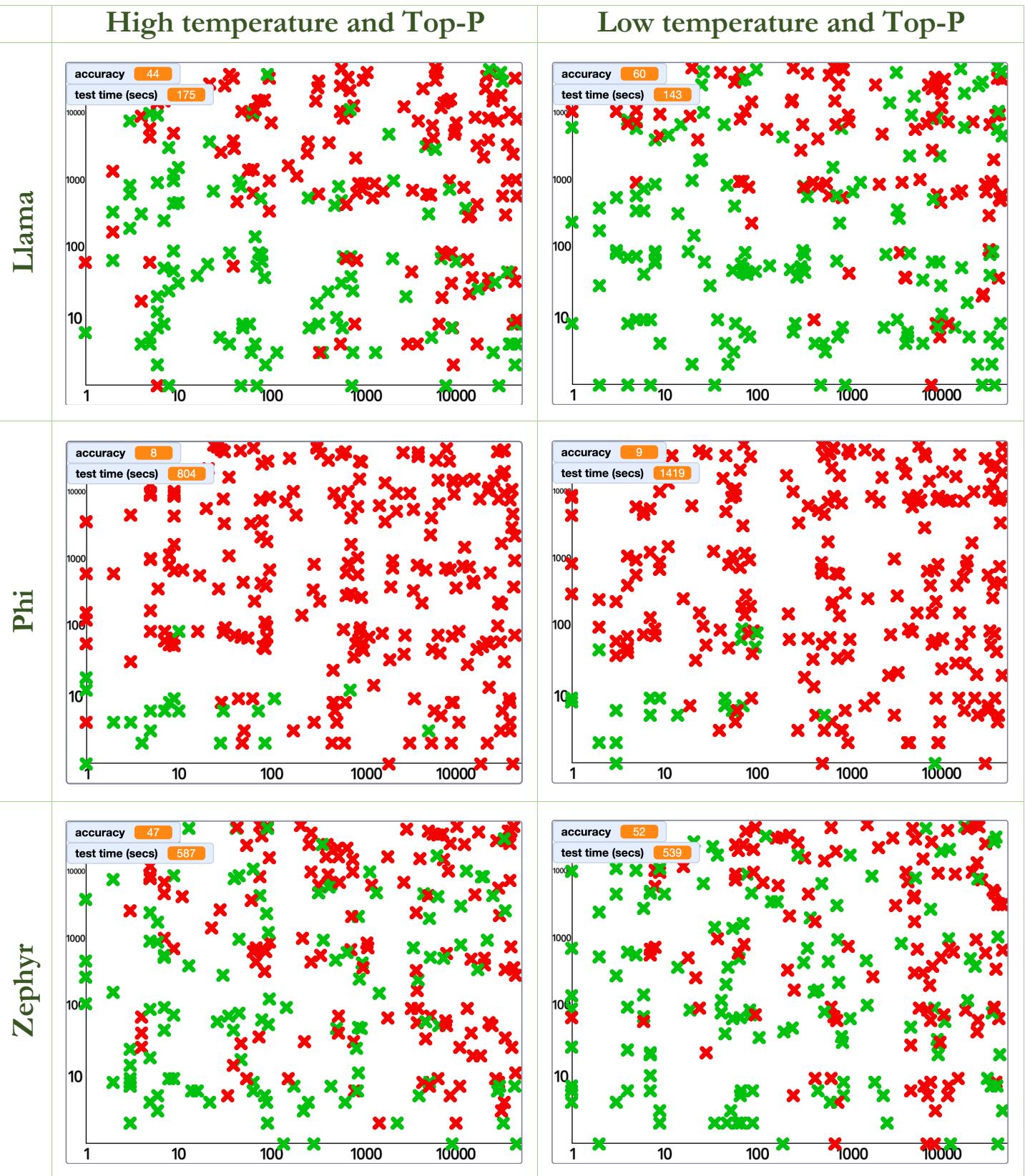
Low temperature and Top-P

Qwen



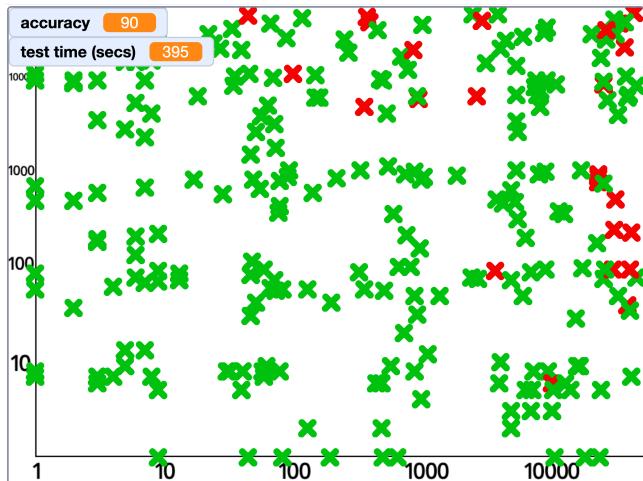
Tiny Llama





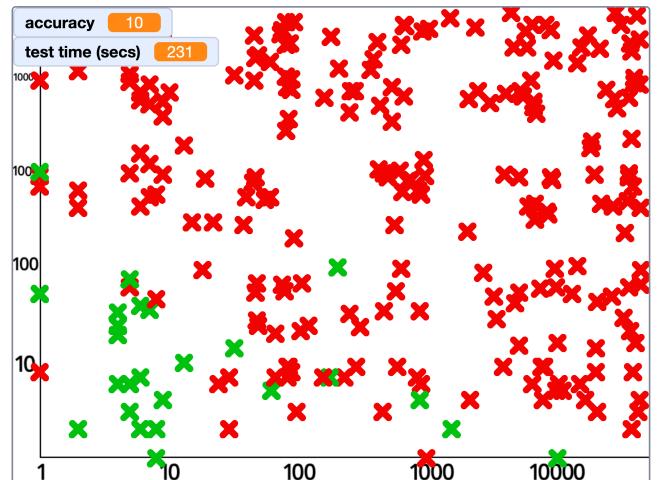
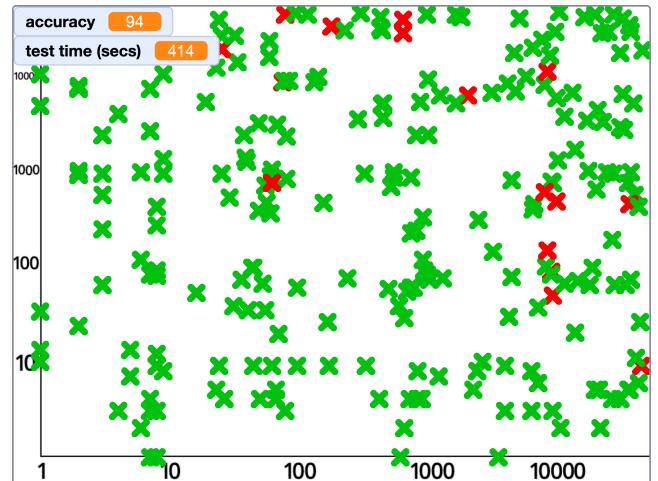
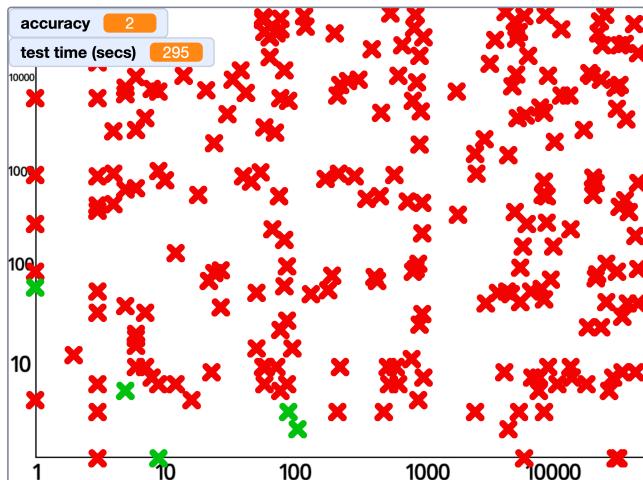
High temperature and Top-P

Gemma



Low temperature and Top-P

Red Pajama



If your students need help finding model cards for each of the models, these links could be helpful:

- [SmolLM2-135M-Instruct](#)
- [Qwen2.5-0.5B-Instruct](#)
- [TinyLlama-1.1B-Chat-v1.0](#)
- [Llama-3.2-1B-Instruct](#)
- [phi-1_5](#)
- [stablelm-2-zephyr-1_6b](#)
- [gemma-2-2b-it](#)
- [RedPajama-INCITE-Chat-3B-v1](#)

However, web searches for more generic terms (e.g. “gemma 2 model card maths”) can turn up simpler information.

For more info, see <https://dalelane.co.uk/blog/?p=5669>