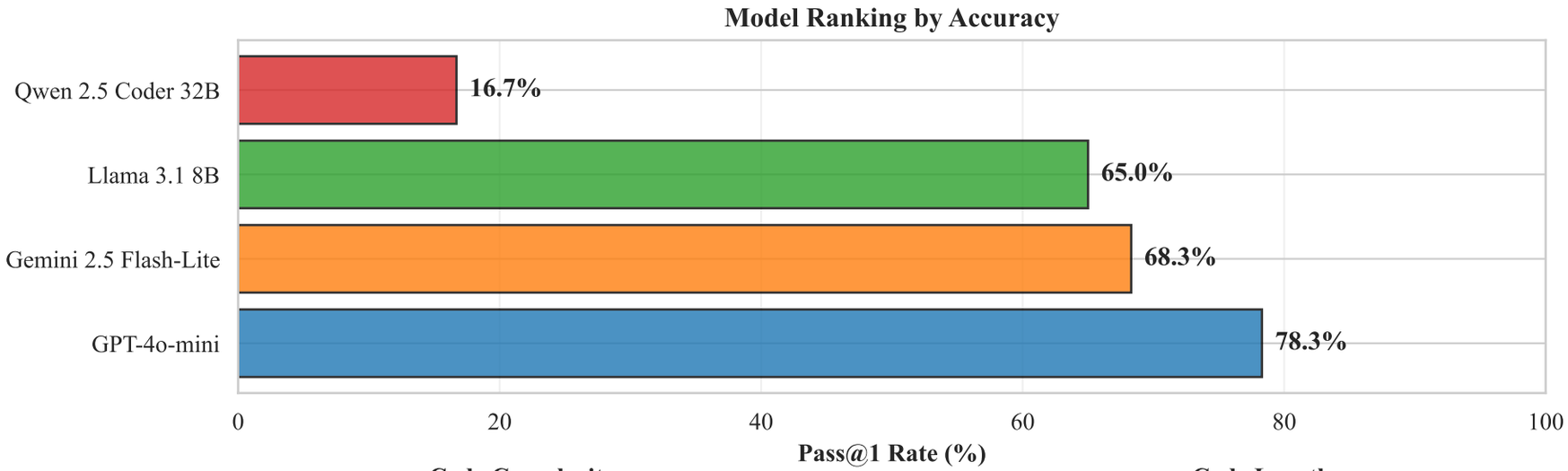
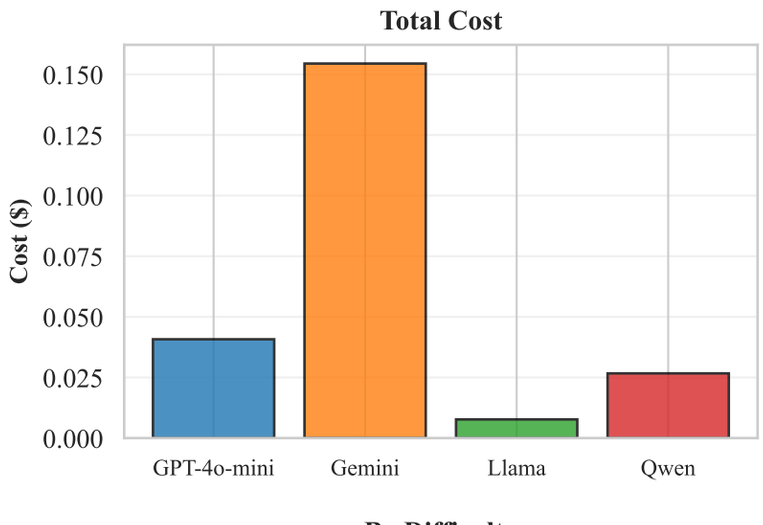
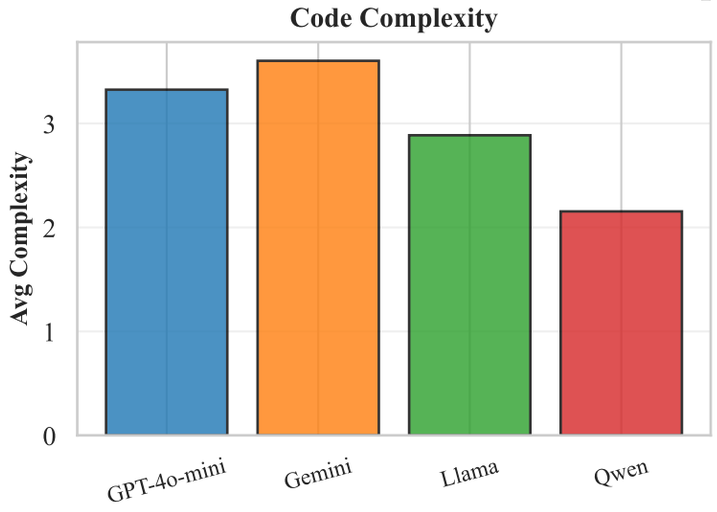
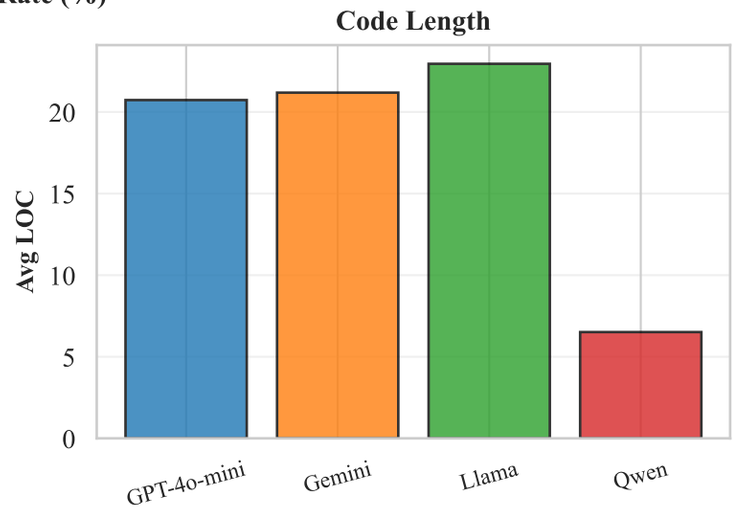
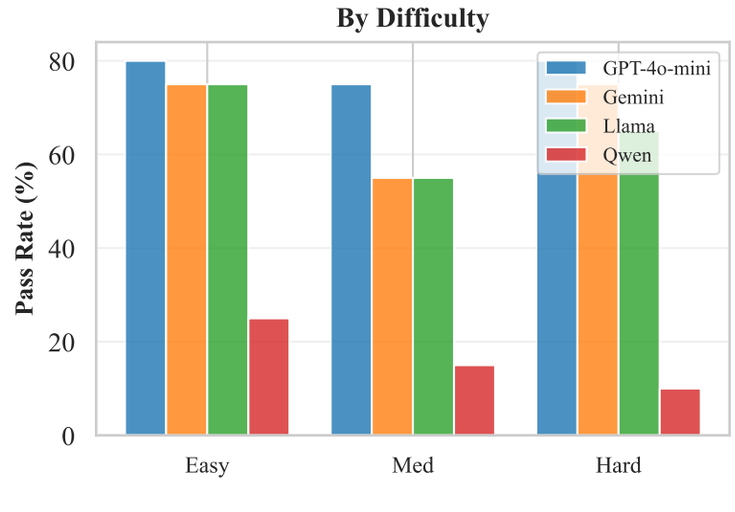
Comprehensive Analysis Dashboard












```
Statistical Significance (Mann-Whitney U with Bonferroni correction, α=0.05):

Significant differences found:
  • GPT vs QWEN: p=9.05e-14, Cohen's d=1.906 (grande)
  • GEMINI vs QWEN: p=2.93e-11, Cohen's d=1.612 (grande)
  • LLAMA vs QWEN: p=8.44e-10, Cohen's d=1.376 (grande)

No significant differences (3 pairs):
  • GPT vs GEMINI: p=0.131
  • GPT vs LLAMA: p=0.032
  • GEMINI vs LLAMA: p=0.463
```