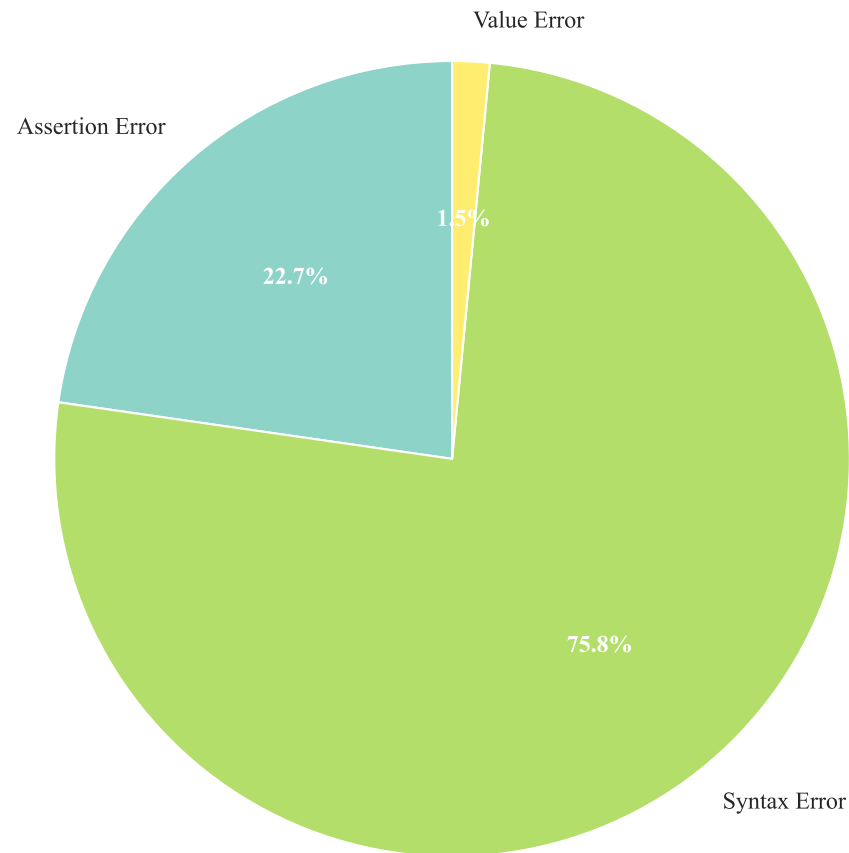
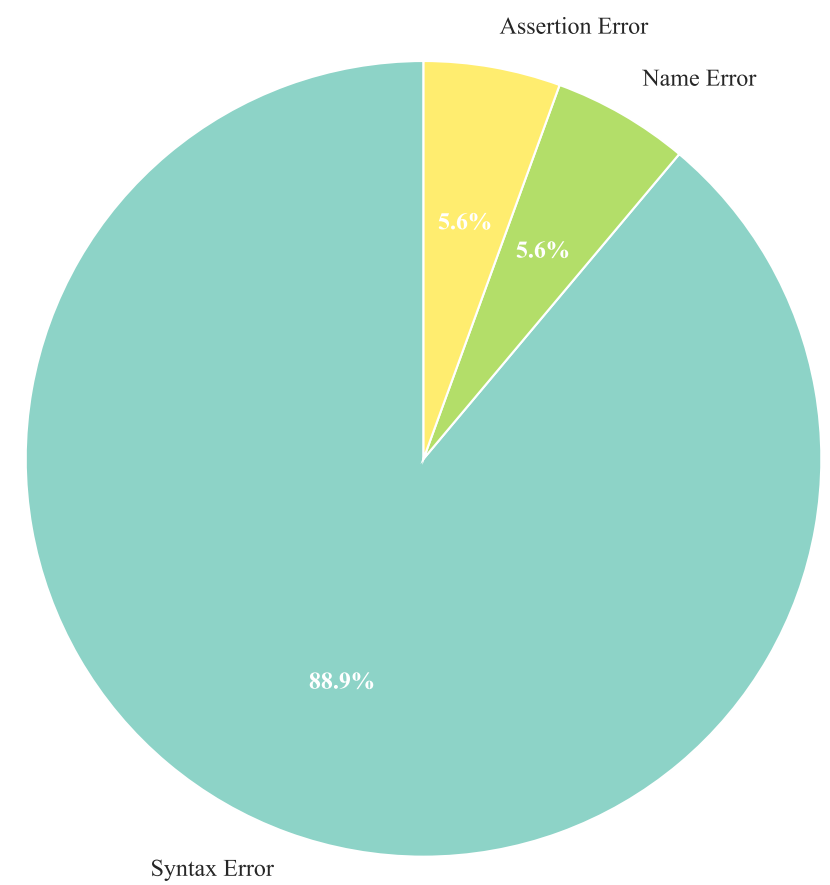


# Failure Analysis: Error Types Distribution

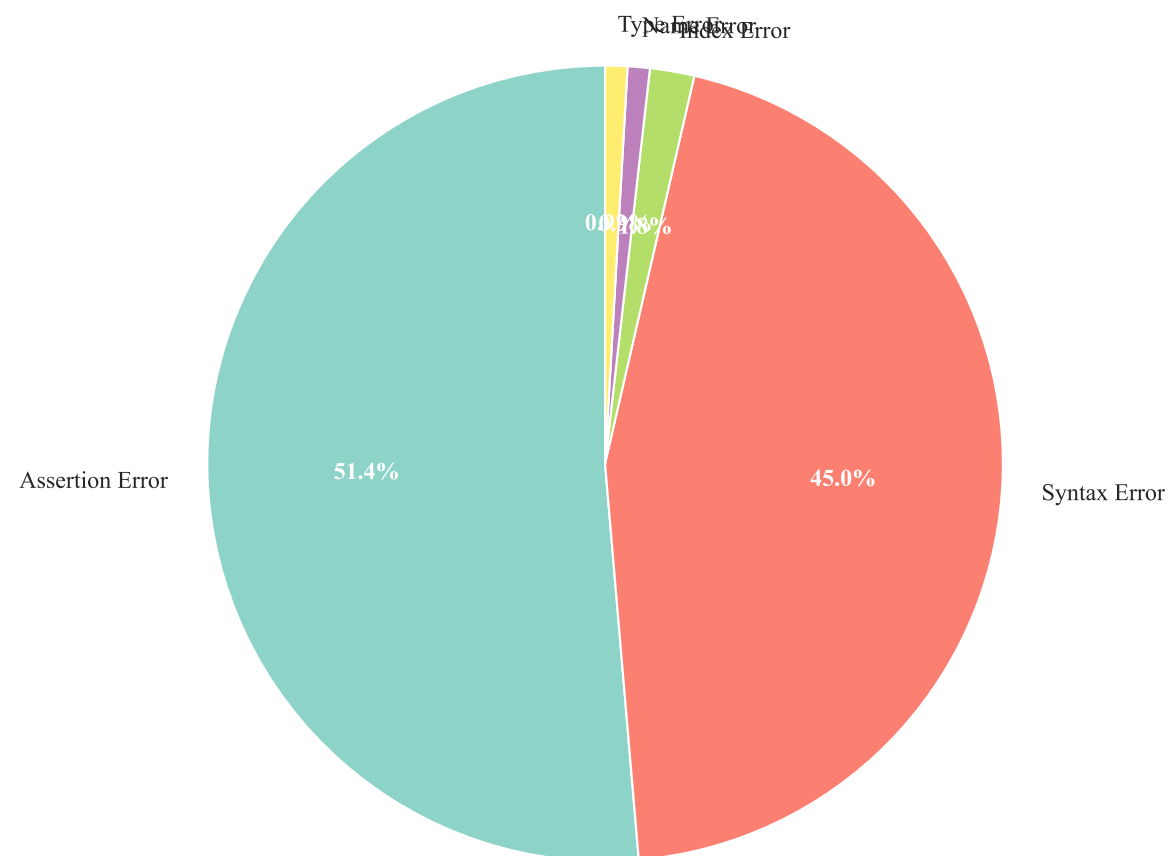
**GPT-4o-mini**  
(66 failures)



**Gemini 2.5 Flash-Lite**  
(90 failures)



**Llama 3.1 8B**  
(111 failures)



**Qwen 2.5 Coder 32B**  
(261 failures)

