Test Scenarios for Agent Evaluation

# 1. Normal Scenarios

• User asks a direct SQL-based question with clear intent.  
• User queries known tables with correct column references.  
• The user question is properly contextualized (e.g., "Show ICU readmission rates for March 2024").

# 2. Edge Case Scenarios

• User provides incomplete query input (e.g., "Show data for ICU").  
• SQL query references a non-existent table or column.  
• Ambiguous user questions with multiple possible interpretations.  
• Long and complex queries that test token limits.  
• Queries with special characters or malformed SQL syntax.  
• Questions outside the domain knowledge of the agent.

# 3. Error Handling Scenarios

• Backend service (like Athena) is down or returns a timeout.  
• SQL query runs successfully but returns no data.  
• Invalid data types or null values in response.

# 4. Contextual Understanding Scenarios

• Follow-up questions relying on context from previous queries.  
• Questions using pronouns or references like "that one", "previous month", etc.  
• Switching topics mid-conversation to test reset of context.