AI Agent Analyst – Home Assignment

Overview

In this assignment, you will build a chatbot application that can answer analytical questions about two key datasets: **reservations** and **reviews** (see table descriptions below). These will be provided as CSV files.

The goal of this chatbot is to make Guesty data accessible to internal users, enabling even non-technical users to quickly retrieve the information they need and work more efficiently.

Focus first on the specific questions listed below, ensuring your chatbot can answer them correctly. If time permits, feel free to extend support for additional types of questions.

You may use any modern AI agent framework, such as **LangGraph, LangChain, LlamaIndex, PydanticAI**, or any equivalent framework of your choice.

Minimum Requirements

- Answer the provided analytical questions based on the data.
- Support **follow-up queries** (maintain conversational context).
- Restrict answers to the provided CSV datasets (don't hallucinate unrelated responses).
- Implement **session management** to keep each user's session isolated.

Be creative — additional functionality is welcome.

Data Provided

1. `reservations.csv`

In Guesty, a reservation represents a booking for a listing (property) made through a distribution channel (Airbnb, Booking.com, direct booking site, etc.). Each row represents a single reservation.

Columns:

- **`reservation_id`**: Unique identifier of the reservation.
- **`account_id`**: ID of the account that owns the reservation
- **`listing_id`**: Unique identifier of the listing booked.
- **`status`**: Status of reservation (e.g., confirmed, canceled).
- **`created_at`**: Reservation creation timestamp
- **`check_in`**: Check-in date.
- **`check_out`**: Check-out date.
- **`payout`**: Amount paid to the host.
- **`guest_count`**: Number of guests in the reservation.
- **`nights_count`**: Number of nights stayed.
- **`booking_window`**: Days between booking date and check-in.
- **`reservation_channel`**: Channel used for booking (e.g., Airbnb, Booking.com).

2. `reviews.csv`

Reviews submitted by guests after their stay, rating the property and their overall experience.

Columns:

- **`review_id`**: Unique review identifier
- **`listing_id`**: The listing reviewed.
- **`reservation_id`**: Reservation associated with this review.

- **`overall_rating`**: Overall rating given by the guest.
- **`cleaniness_rating`**: Cleanliness rating.
- **`location_rating`**: Location rating.

Analytical Questions

Your agent should be able to answer the following questions:

1. How many reservations did listing XXX have last week?
2. From which channel did listing YYY receive the most reservations?
3. Show me the distribution of reservations by creation time for account XXX.
4. What is the average number of nights per reservation for listing XXX / account YYY?
5. Which reservation channel generated the highest total payout last month?
6. How many reservations were created in the last 30 days for account YYY?
7. What percentage of reservations were canceled?
8. How many reviews does listing XXX have?
9. Which are the top 10 listings with the highest overall score?

Deliverables

- You have **3 days** to complete the assignment.
- Create a **Git repository** with your code and share the link.

Evaluation Criteria

We will evaluate based on:

- **Correctness** – Does the agent return accurate answers to the analytical questions?
- **Code Quality** – Code should be clear and modular.
- **Architecture & Design –** The solution should demonstrate good architectural practices.
- **Creativity** – Innovative use of tools, extra features