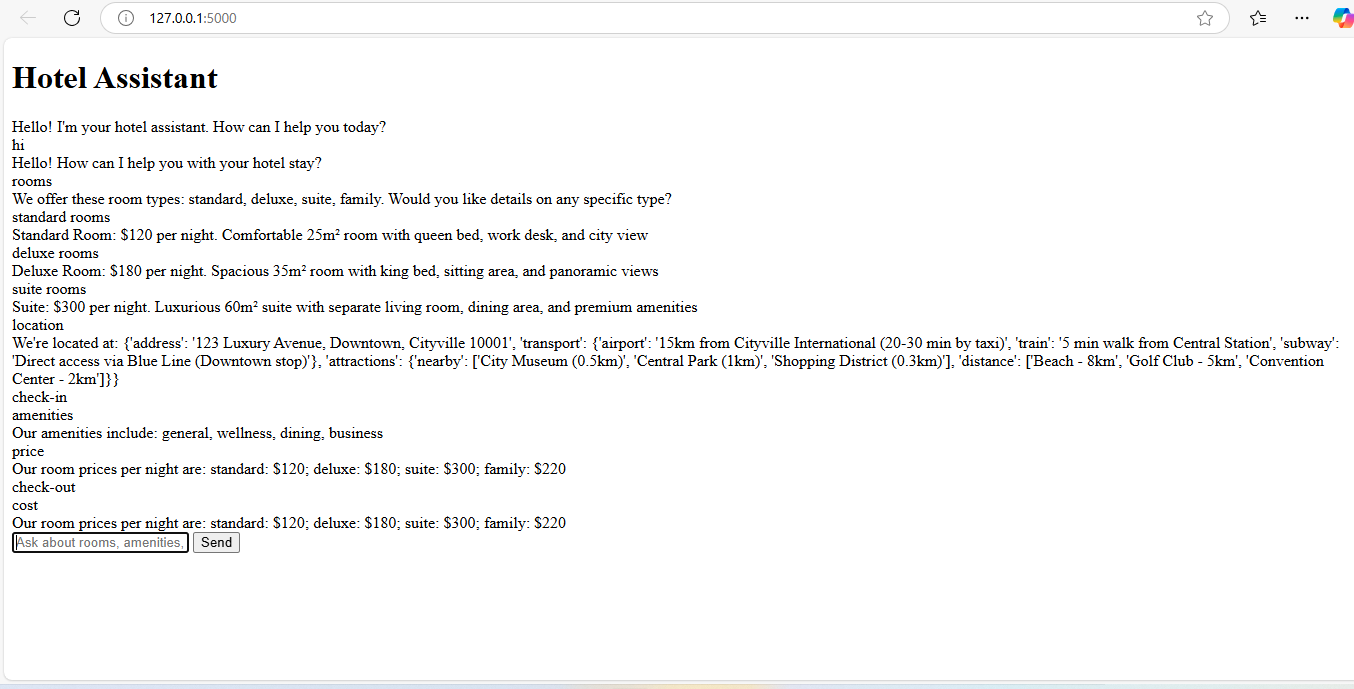
**HOTEL INFORMATION ASSISTANT**

**Hadith Chatbot Explanation**

**1. Data Collection**

* The code clones a GitHub repository (LK-Hadith-Corpus) containing Hadith texts organized by different Islamic scholars (Bukhari, Muslim, Abu Daud, etc.)
* The repository contains CSV files organized by chapters, with each file containing multiple Hadiths

**2. Data Processing Pipeline**

1. **File Reading**:
   * Uses glob to find all CSV files in the repository
   * Reads each CSV file into a pandas DataFrame with predefined column names
2. **Text Cleaning**:
   * Converts text to lowercase
   * Removes special characters using regex
   * Normalizes whitespace
3. **Data Consolidation**:
   * Combines Hadiths from all files into a single DataFrame
   * Keeps relevant columns: chapter info, section info, Hadith number, and the English text of the Hadith

**3. Key Components**

* **Columns Used**:
  + chapter\_Number, chapter\_English, chapter\_Arabic: Metadata about the chapter
  + Section\_number, Section\_English: Metadata about the section within the chapter
  + haidth\_number: The identifier for each Hadith
  + English\_Hadith: The actual text of the Hadith in English
* **Text Cleaning Function**:
  + Standardizes the text to make it easier to process and analyze later
  + Prepares the text for potential NLP tasks by removing noise

**4. Output**

The final output is a pandas DataFrame (hadith\_df) containing all Hadiths from all files with their associated metadata, ready for further analysis or to be used in a chatbot system