Technical Test 1: Web Crawling & Data Extraction

Objective: Build a web crawler to extract and store medical professionals' details from Andalusia Health.

Task Details:

- Scrape all medical professionals practicing in **Andalusia, AL**.
- Extract the following details for each doctor:
  1. **Full Name**
  2. **Doctor Profile URL**
  3. **Address**
  4. **Phone Number**
  5. **Indicator if they have more than one location**
  6. **Whether they accept new patients**
  7. **Whether they are an employed provider**
- Store the extracted data in a **database** (MongoDB or PostgreSQL preferred).

Data Analysis & Reporting

After storing the data, analyze it to generate the following insights:

1. **Total number of doctors**
2. **Total number of doctors with ratings** (if ratings are available on the website)
3. **Doctors having the same phone number**
4. **Doctors with more than one location**

Expected Deliverables:

- **Codebase**:
  - A well-structured Python script to crawl the website.
  - A database schema .
  - Data analysis queries or scripts.
- **Database Dump**: Exported dataset of the extracted doctors' information.
- **Report**: JSON or CSV output of the analysis.

Technical Test 2: SQL

For Data please refer to **SQL_test.xlsx**

1. Find the total sales for each region.
2. List the top 5 customers with the highest total sales.
3. Find the average profit for each product category.
4. Identify the month with the highest total sales.
5. List all orders where the profit is negative (loss).
6. Find the total shipping cost for each ship mode.
7. List the top 3 products with the highest sales.
8. Find the total number of orders placed by each customer segment.
9. List all orders where the discount is greater than 0.05.
10. Find the total profit for each state.

Technical Test 3: Reverse Engineering RAG and Prompt

Reference File: `app.py`

Task 1: Code Analysis

Carefully review the `app.py` file and perform a detailed analysis by:

- Examining all imports and their roles in the application.
- Breaking down each function, explaining its logic and purpose.
- Providing a comprehensive explanation of how different components interact within the script.

Task 2: Structured Prompt Design

Create a well-structured prompt to extract diverse details about **Andalusia Health**. Consider the various types of data that could be retrieved and ensure the prompt is designed to elicit comprehensive and relevant responses.