# User Guide for Neo4j Clinical Trials Project

## Overview

This project processes clinical trials data and generates recommendations based on similarities.

Follow the steps below to set up and execute the workflow.

## Prerequisites

1. Python: Ensure Python is installed (3.8 or higher).
2. Dependencies: Install required Python libraries by running:

   pip install -r requirements.txt

3. Neo4j: Install and configure a Neo4j database with Graph Data Science (GDS) enabled.

## Execution Steps

Follow the steps below in sequence:

1. Create Relationships

   Run the `CreateRelationship.py` script to process the input data and extract relationships.

   python CreateRelationship.py

   Output: Generates `relationships.csv` containing extracted relationships.

2. Find Similar Entities

   Run the `SimilarEntities.py` script to identify and merge similar entities.

   python SimilarEntities.py

   Output: Generates `filtered_results_with_similars.csv` containing merged entities.

3. Ingest Data into Neo4j

Run the `ingest.py` script to load the relationships into the Neo4j database.

python ingest.py

4. Generate Recommendations

Run the `model.py` script to query the Neo4j database and view the top 10 results for a given trial ID.

python model.py

Input: Provide an NCT ID when prompted.

Output: Displays the top 10 similar trials based on the input ID.

**Notes**

- Ensure the Neo4j database is running before executing `ingest.py` and `model.py`.
- All outputs are saved in the same directory as the scripts.

**File Overview**

- CreateRelationship.py: Extracts relationships from clinical trials data.
- SimilarEntities.py: Finds and merges similar entities.
- ingest.py: Loads relationships into Neo4j.
- model.py: Generates recommendations from Neo4j.
- data_200.csv: Input clinical trials data.
- requirements.txt: Lists required Python libraries.