Task List:

Task 1: Perform EDA on both CSV files and generate pandas profiling reports(10)

Task 2: In both CSV files, remove rows where column 'frm' and 'txt' are empty. Also, reset index once rows are deleted(20)

Task 3: In Records.CSV, create a new column by replacing "::" by "," from the string in column 'frm'. Remove additional spaces, commas if necessary (30)

Task 4: In Base.CSV, Create a new column by joining columns 'txt','addr','addr2','city','zip','nation'. Remove additional spaces, commas if necessary.(40)

Task 5: For both CSV files, create sentence embeddings for newly created strings (through tasks 3 & 4) using TF-IDF method(80)

Task 6: For both CSV files, create sentence embeddings for newly created strings (through tasks 3 & 4) using Word2Vec method.(80)

Task 7: For both CSV files, create sentence embeddings for newly created strings (through tasks 3 & 4) using GPT3.(80)

Task 8: For every record in Records.CSV, create a similarity score with every record in Base.CSV. Use cosine similarity option. Every record in Records.CSV is to be compared with every record in Base.CSV (80)

Task 9: Create a new CSV file having Maximum similarity scores of every record in Records.CSV(80)