Title: Database Relationship/Mapping Indication Graph Generation

Authors: Jinhao Lu, Tianbao Piao, Anish Rao

Summary:
The purpose of the package is to visualize the relationship between different tables in the database or the mapping relationship from the source data table to a standardized common data model. Export the visualization for use in project or client consulting presentations.
By creating the mapping visualization, loading statements for ETL processes can be created by the package. Create select/insert statement to query the database, copy/loading statement for data bucket.

External libraries needed:

Pandas: Read_csv/read_excel: This method needed for loading data from csv or excel file.
Dataframe: The data manipulation basically will be applied on the pandas dataframe.
SQLite3: We use sqlite3 as an instance of a relational database.
Create a cursor to fetch data, implement the SQL statement, test the output query.
Graphviz: Use the digraph class imported from graphviz package to create an object then export the object as a graph DOT file.

Implementation of external libraries:

To implement the package's intended functionality, we may need to use external libraries such as NumPy, Pandas, SQLite3, Graphviz and potentially other ones.The Graphviz package provides modules, classes, and functions for generating customizable graphs and diagrams in various formats. The 'graphviz' module includes the 'Digraph' and 'Graph' classes for creating directed and undirected graphs. The package also includes functions for rendering the graphs and displaying them in interactive viewers.Our team's work will involve designing and implementing functions to preprocess and format the input data, define the graph structure, customize its attributes, and provide a user-friendly interface. The libraries mentioned above are tools that our team will most likely need to use for the package development.

Functionality:

1. **Data Extraction Module:** This module will contain classes and functions that fetch table info from the database: Retrieve information about the tables in a database, such as their names, columns, data types, and relationships to other tables.

2. **Data Mapping Module:** This module will contain classes and functions that customize source data table graphic generation, allow the user to customize the appearance of the source data tables in the visualization. Indicate the mapping relation between tables. Specify the mapping relationship between the source data tables and the standardized common data model.

3. **Visualization Module:** Create visualization: Generate a visualization of the mapping relationship between the source data tables and the common data model using Graphviz.
Export the visualization: Allow the user to export the visualization as an image file or a Graphviz DOT file.

4. **ETL Module:** Create copy/loading script for use in ETL processes: Generate SQL statements or other scripts that can be used to copy data from the source tables into the common data model, and create SELECT statements or data export scripts that can be used to extract data from the source tables for analysis or reporting purposes.