Assignment #1 Linked data

Issue: *there is a missing link between files "tbd-1" and "tbd-2"*

Task: is it possible to create a primary key column in file "raw_data.csv" that would link records between "processed_data.csv" and "raw_data.csv". If this link can be established, please provide a technological solution for it in a form of Python or SQL script (e.g. a script that creates a primary key field in "raw_data.csv" and foreign key field in "processed_data.csv" or a new table or other).

Hint: *Field "id" in "processed_data" has been created based on this pattern "'{latitude}'_'{longitude}'_'{rent in EUR}'_'{floor_area}'_'{date_created}'"*

Resources: *link to data*

Answer:

The table called "raw data" has three columns and its primary key column is or should be the one called "_RowNumber". The "processed_data" tables' primary key is most likely the "row_id" column but that table has also the same column called "_RowNumber" that is in the "raw data" table. That column could be set as a primary key in "raw_data" table and as a foreign key in "processed_data" table to link the records between the two.

Solution in SQL:

The table "raw_data" technological solution:

```
CREATE TABLE raw_data (
    _RowNumber int NOT NULL PRIMARY KEY,
    Value varchar,
    created varchar NOT NULL,
);
```

or if the table is already created:

```
ALTER TABLE raw_data
ADD PRIMARY KEY (_RowNumber);
```

The table "processed_data" technological solution:

```
CREATE TABLE processed_data (
    row_id int NOT NULL PRIMARY KEY,
    (here would be listed all the other 30 columns)
    _RowNumber int FOREIGN KEY REFERENCES raw_data (_RowNumber)
);
```