**TaskC.1(b) Data cleaning**

**Explore Data:**

A screenshot of a cell phone

Description automatically generatedA close up of text on a white background

Description automatically generated

The result below shows the data for each table.

A screenshot of a cell phone

Description automatically generatedAddress table:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Advertisement table:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Agent table:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generatedAgent\_office table:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Client table:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generatedClient\_wish table:

A screenshot of a cell phone

Description automatically generated

A close up of a logo

Description automatically generated

Feature table:

A screenshot of a cell phone

Description automatically generated

Office table: A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Person table: A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Postcode table: 

A screenshot of a cell phone

Description automatically generated

Property table:A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Property\_Advert table:A close up of a logo

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Property\_feature table:

A picture containing book

Description automatically generatedA close up of a logo

Description automatically generated

Rent table:A close up of a logo

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Sales table:A picture containing drawing

Description automatically generated

A screenshot of a cell phone

Description automatically generated

State table:A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Visit table:A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Find errors:**

After exploring the data as shown above, the following checking process for detecting errors are performed as shown below.

This query is to check whether there are any duplicated records in address table.

A picture containing knife

Description automatically generated

The result shows that there are 39 addresses have duplicated records.

A close up of text on a white background

Description automatically generated

This query below is to check whether there are any null and illegal values in the address table.

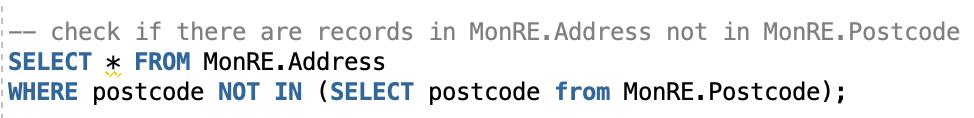
A picture containing bird

Description automatically generated

The result shows that there are no null or illegal values in address table.



This query below is to check whether all the postcodes recorded in the address table matches the records in the Postcode table.



The result is positive, which shows no finding.



This query below is to check whether there are any duplicated records in the advertisement table.

A picture containing knife

Description automatically generated

The result shows there are no duplicated records.

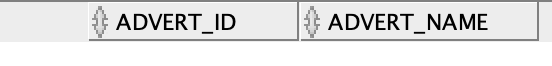


This query below is to check whether there are null and illegal values in the advertisement table.

A screenshot of a cell phone

Description automatically generated

The result shows no.



This query below is to check whether there are any null and illegal values in the agent table.

A screenshot of a cell phone

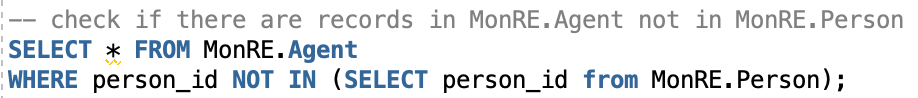
Description automatically generated

The result shows there are one 0 and two negative values of salary attribute in the agent table as shown below.

A screenshot of a cell phone

Description automatically generated

This query below is to check is there any records of person Id in agent table that are not in the person table.

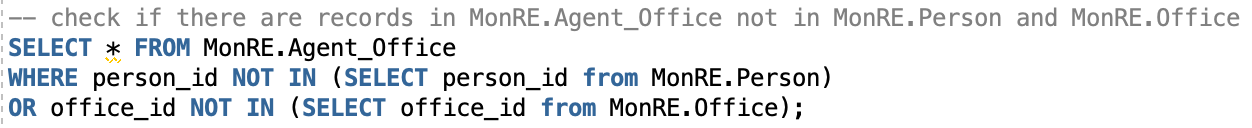


The result shows that person id 6997 is not recorded in the person table.

A screenshot of a cell phone

Description automatically generated

This query below is to check is there any records of person Id in the agent office table are not in the person table. Also, to check whether there are any records od office Id that are not recorded in the Office table.



The result shows there is one record as shown below.

A picture containing table

Description automatically generated

This query below is to check whether there are any null or illegal values in the client table.

A screenshot of a cell phone

Description automatically generated

The result shows that there is a record of maximum budget that is a negative value of -150. Besides, there are two records where the maximum budget is lower than the minimum budget.

A screenshot of a cell phone

Description automatically generated

This query below is to check whether there are any illegal values in the client wish table. A picture containing knife

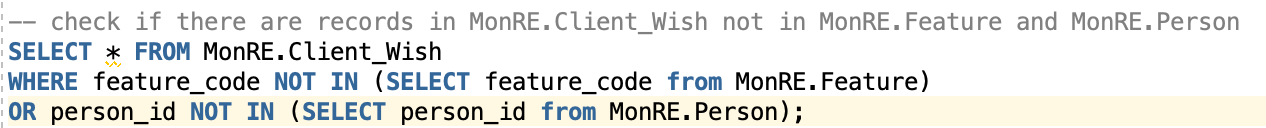
Description automatically generated

The result shows that there are no illegal values in the client wish table.

A picture containing clock

Description automatically generated

This query below is to check whether there are any records of feature code in the client wish table that are not in the feature table, and whether any records of person Id in the client wish table that are not in the person table.



The result shows no.

A close up of a logo

Description automatically generated

This query below is to check is there are any null and illegal values in feature table.

A picture containing knife

Description automatically generated

The result shows that there are no such values.



This query below is to check whether there are any duplicated records in the feature table.

A picture containing knife

Description automatically generated

The result shows that there are no records of duplicated values.

A close up of a logo

Description automatically generated

This query below is to check whether there are any illegal and null values in the office table.

A picture containing knife

Description automatically generated

The result shows that there are no null or illegal values in the table.

A picture containing table

Description automatically generated

This query below is to check whether there are any duplicated records in the office table.

A picture containing knife

Description automatically generated

The result shows that there are no duplicated records.

A close up of a logo

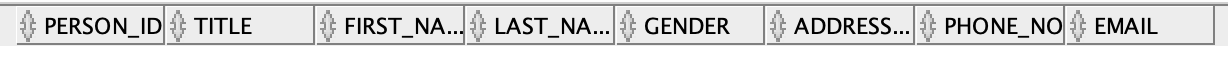
Description automatically generated

This query below is to check whether there are any illegal and null values in the person table.

A screenshot of a cell phone

Description automatically generated

The result shows no such record.



This query below is to check whether there are any duplicated records in the person table.

A picture containing knife

Description automatically generated

The result shows that there are four duplicated records for person with id 6995.

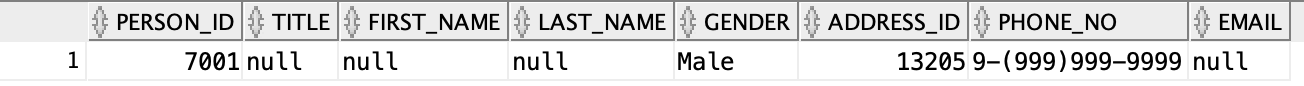
A screenshot of a cell phone

Description automatically generated

This query below is to check is there any records of address Id in the person table that are not in the address table.



The result shows that there is one record of address is not in the address table.



This query below is to check whether there are any illegal or null values in the postcode table. A picture containing knife, table

Description automatically generated

The result shows that there is no such record.

A picture containing drawing

Description automatically generated

This query below is to check whether there are any duplicated records in the postcode table.

A screenshot of a cell phone

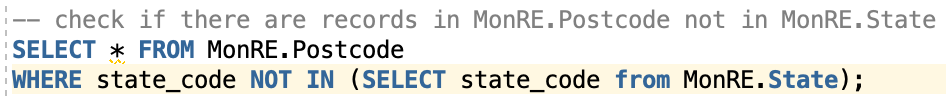
Description automatically generated

The result shows that there is no such record.

A picture containing drawing

Description automatically generated

This query below is to check whether there are any records of state code in the postcode table that are not in the state table.



The result shows that there is no such record.

A picture containing drawing

Description automatically generated

This query below is to check whether there are any illegal or null values in the property table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are 1550 rows of records contains illegal or null values.

A screenshot of a social media post

Description automatically generated

This query below is to check whether there are any duplicated records in property table.

A picture containing knife

Description automatically generated

The result shows that there are two property that have duplicated records.

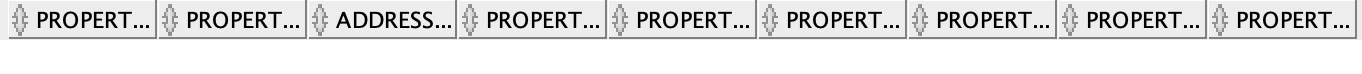
A screenshot of a cell phone

Description automatically generated

This query below is to check whether there are any records of address Id in the property table that are not in the address table.



The result shows that there is no such value.



The query below is to check whether there are any illegal or null values in the property advert table.

A picture containing knife

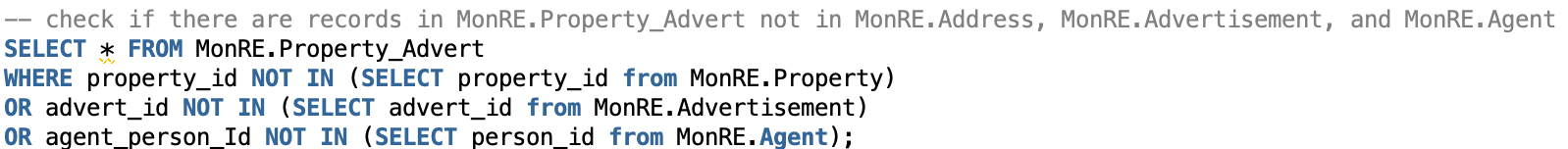
Description automatically generated

The result shows that there is no such value.

A close up of a logo

Description automatically generated

The query below is to check whether there are records of property\_id in the property advert table that are not from the property table. And to check whether there are any records of advert\_id in the property advert table that are not from the advertisement table. Also, to check whether there are any records of agent\_person\_id in the propery advert table that are not in the agent table.



The result shows no such values.



This query below is to check whether there are any illegal and null values in the property feature table.

A screenshot of a cell phone

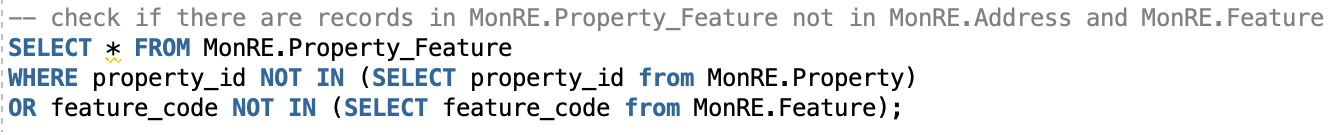
Description automatically generated

The result shows no such values.

A screenshot of a cell phone

Description automatically generated

This query below is to check whether records of property\_id in property feature table that are not in the property table. And to check whether the feature\_code in property feature table that are not in the feature table.



The result shows that there are no such values.

A picture containing drawing

Description automatically generated

This query below is to check whether there are any duplicated records in the rent table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are no such record.



This query below is to check whether there are any illegal and null values in rent table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are 50 rows of records that contains illegal and null values.

A screenshot of a cell phone

Description automatically generated

This query below is to check are there any records of agent\_person\_id and client\_person\_id in rent table that are not in person table. And are there any records of property\_id in rent table that are not in the property table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are no such records.



This query below is to check is there any duplicated records in sale table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are no such records.

A close up of a logo

Description automatically generated

This query below is to check whether there are any illegal or null values in the sales table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are 50 records of sales that contains illegal or null values.

A screenshot of a cell phone

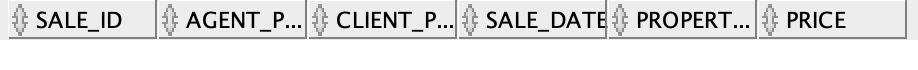
Description automatically generated

This query below is to check are there any records of agent\_person\_id and client\_person\_id in sale table that are not in person table. And are there any records of property\_id in sale table that are not in the property table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are no such value.



This query below is to check are there any null or illegal values in the state table.

A picture containing knife

Description automatically generated

The result shows that there is one record of null value.

A close up of a logo

Description automatically generated

This query below is to check are there any null or illegal values in the visit table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are no such records.



This query below is to check are there any records of agent\_person\_id and client\_person\_id in visit table that are not in person table. And are there any records of property\_id in visit table that are not in the property table.

A screenshot of a cell phone

Description automatically generated

The result shows that there are no such records.



**Clean data:**

After detecting errors in the previous process. The following detection strategy is applied to clean the error.

--------------------------------------------------------------------------------

-- Error 1: Duplicate record in ADDRESS table

--(39 out of 13204 records were redundant)

-- Detection Strategy:

A picture containing knife

Description automatically generated

A picture containing table

Description automatically generated

A picture containing knife

Description automatically generated

A screenshot of a cell phone

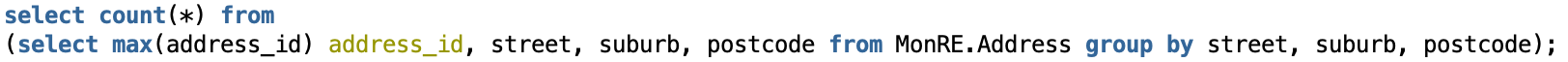
Description automatically generated



A close up of a logo

Description automatically generated

-- Solution:



A screenshot of a cell phone

Description automatically generated

-- Error 2: Dirty record in Agent table

--(There are 0 and negative values in MonRE.Agent)

-- Detection Strategy:



A picture containing table

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

-- Solution:

A close up of a logo

Description automatically generated

A close up of a logo

Description automatically generated

-- Error 3: Dirty record in Client table

--(There is negative value in MonRE.Client and there are maximum budget that are smaller than minimum budget)

-- Detection Strategy:



A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

-- Solution:

A picture containing table

Description automatically generated

A close up of a logo

Description automatically generated

-- Error 4: Dirty record in Person table

--(There are 4 duplicate values in MonRE.Person, we will just keep 1 record in the database)

-- Detection Strategy:



A screenshot of a cell phone

Description automatically generated

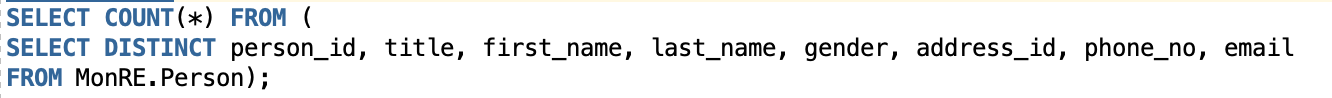
A picture containing knife

Description automatically generated

A close up of a logo

Description automatically generated

-- Solution:



A screenshot of a cell phone

Description automatically generated

-- Error 5: Dirty record in Person table

--(There is one person with no title, no name, no email, and meaningless phone number in Person table)

-- Detection Strategy:



A screenshot of a cell phone

Description automatically generated

A picture containing knife

Description automatically generated

A close up of a logo

Description automatically generated

-- Solution:



A screenshot of a cell phone

Description automatically generated

-- Error 6: Dirty record in Property table

--(There is 4403 records with no property size value in Property table, we will simply set it to 0 showing that it is not availabe)

-- Detection Strategy:



A screenshot of a cell phone

Description automatically generated

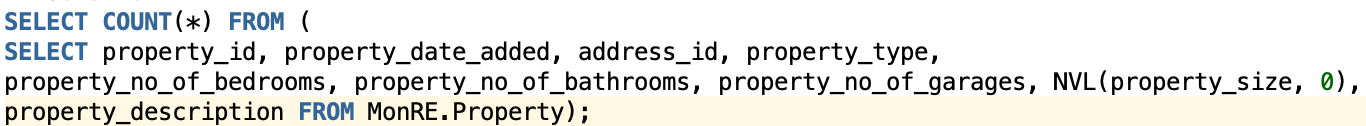
A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

-- Solution:



A screenshot of a cell phone

Description automatically generated

--Error 7: Dirty record in Property table

--(There are 20 redundant records in Property table, we will keep it distinct in the database)

-- Detection Strategy:



A screenshot of a cell phone

Description automatically generated

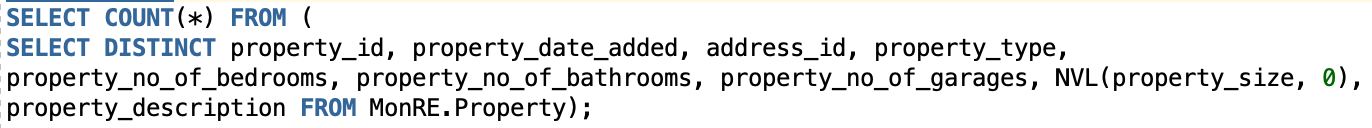
A picture containing knife

Description automatically generated

A picture containing table

Description automatically generated

-- Solution:



A screenshot of a cell phone

Description automatically generated

-- Error 8: Dirty record in Rent table

--(There are 1637 records without client\_person\_id, rent\_start\_date, and rent\_end\_date in Rent

-- table. In terms of integrity, we will delete it from the database)

-- Detection Strategy:



A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

-- Solution:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

-- Error 9: Dirty record in Sale table

--(There are 2009 records without client\_person\_id, sale\_date in Sale table. In terms of integrity, we will delete it from the database)

-- Detection Strategy:



A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

-- Solution:

A picture containing knife

Description automatically generated

A picture containing drawing

Description automatically generated

-- Error 10: Dirty record in State table

--(There is 1 record without state\_code in State table. In terms of integrity, we will delete it from the database)

-- Detection Strategy:



A close up of a logo

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

-- Solution:

A screenshot of a cell phone

Description automatically generated

A close up of a logo

Description automatically generated