1340 Midterm Project

Bohao Cui
1009314467

November 2022

1 Main Problem

In the entire project data, we can find that the same problem "Column headers are values, not variable names" exists in Tables 1 to 5. These five tables all use 1990, 1995, 2000, 2005, 2010 and 2015 as headers, but these headers should be the value of the variable name "Year". Also, the values of these headers should belong to another variable name. To solve this problem, we need to melt the data. The first step is to select the column that does not need to be changed as the identifier. In this table, the selection is "Major area, region, country or area of destination, Notes, Country code, Type of data(a)". The second step is to move the column names 1990, 1995, 2000, 2005, 2010, 2015 to the variable year. The third step, move the values of these 6 columns into a new column, the column name will be renamed according to the data type.

The second problem is that the second column of all tables contains three values that Major area, region and country or area of destination, which will lead to multiple variables stored in one column" problem. We'll get the values for the Major area, region and country or area of destination according to the table "ANNEX" and remove duplicate elements in the sequence. The next step is to extract the corresponding data from tables 1 to 5 based on these names and form a new table.

The third problem is that except for the fourth table, the other tables have three observations. For example, table 1 has three observations for male, female and both sexes. First of all, we need to ensure that an observation should be in a table, rather than a table with multiple observations. So use the "iloc" command to extract the data of male, female and both sexes respectively. At the same time, the column names male and female also violate the rule that the column name should be a variable name and not a variable value, so these two columns should belong to the value of the variable name Gender, not as two separate columns of data, and the same two column values should also be attributed to the same column.

The fourth problem is that the table contains many null values, which are represented by ".." in the Excel document. For null values we will use the command notation "$na_values = '..'$ "toreplaceall"" with NAN, which is the notation for null values in the python language, which also makes it possible to use commands to process null values.

The first 14 lines of all tables are some introduction and description about the table. And because the column names are distributed in rows 15 and 16, this will cause some column names to be unreadable. So, we will create a new list of column names to rename the table.

2 Define Function

The main problems that need to be modified in the data have been introduced in the above article. Since the six tables contain the same problem, in order to avoid repeating the steps, we will create multiple functions to clean the data,

so that when cleaning other tables, only Just need to call the function.

(1) The "*get_table*" function is to reads the table from the UN data according to the table name and selects the number of rows to discard according to "*skip_rows*".

```
def get_table(table_name, skip_rows):
    return pd.read_excel('/content/drive/MyDrive/
        INF1340_Midterm/UN_MigrantStockTotal_2015.xlsx',
        sheet_name=table_name,skiprows=skip_rows, na_values =
         '..')
```

(2) The "*split_table*" function is to split the left part of the table according to the index number and insert the column name. This will solve the problem of not being able to read the column names.

```
def split_table(table, index, col_name):
    temp = table.iloc[:,1:index]
    temp.columns = col_name
    return temp
```

(3) The "*melt_function*" is to convert the data frame from wide format to long format, this is the method to solve the problem that the column name is not the variable name but the variable value.

```
def melt_func(table, ID_Vars,Var_Name,Value_Name):
      return table.melt(id_vars=ID_Vars, var_name=Var_Name,
          value_name=Value_Name)
```

(4) The "*select_row*" function extracts the specified number of rows from the table according to the parameters and combines them into a new table. Then rename the table and use the "*rest_index*" function to reassign new index numbers to the rows of the table.

```
def select_row(table,col_name,row_name):
    temp = table[table[col_name].isin(row_name)]
    if row_name[1] == country[1]:
        temp = temp.rename(columns={'Major area, region,
            country or area of destination':'Country or area
            of destination'})
    elif row_name[1] == major_area[1]:
        temp = temp.rename(columns={'Major area, region,
            country or area of destination':'Major area'})
    elif row_name[1] == region[1]:
        temp = temp.rename(columns={'Major area, region,
            country or area of destination':'Region'})
    temp = reset_index(temp)
    return temp
```

(5) The row number in the table starts from 0, this function will reset the row number to start from 1.

```
def reset_index(table):
    table = table.reset_index()
    table.index = table.index+1
    table = table.drop(columns=['index'])
    return table
```

(6) This function is to extract both sexes observations from the table and create a new table to store the data. Then delete the empty values in the table.

```
def creat_both_sexes_table(table,name1,name2):
    temp = table
    temp = temp.drop(columns=[name1,name2])
    #Delete a row based on missing values in a column
    col_name = temp.columns[temp.columns.size-1]
    temp = temp.dropna(subset=[col_name])
    return temp
```

(7) This function is to extract the male and female observations in the table and use the melt function to merge the two columns of data. Then create a new table to store the data to store the data. Finally delete the empty values in the table.

```
def creat_gender_table(table,name1,name2,name3,ID_Vars,
    var_name,value_name):
    Gender_table = table
    both_sexes_table = table
    Gender_table = Gender_table.rename(columns={name1:'male'
        })
    Gender_table = Gender_table.rename(columns={name2:'
        female'})
    Gender_table = Gender_table.drop(columns=[name3])
    Gender_table = melt_func(Gender_table,ID_Vars,var_name,
        value_name)
    Gender_table = reset_index(Gender_table)
    #Delete a row based on missing values in a column
    Gender_table = Gender_table.dropna(subset=[value_name])
    return Gender_table
```

3 Steps

Table 1 has 23 columns and contains three distinct observations, with six years of data under each observation. So, the first step is to divide the table into

three tables by observations, then use the function "$melt_function$" to melt the six years together to get a new table. This step solves the problem of using multiple year values as column names. Table 1 shows the data for both sexes of observations.

	Major area, region, country or area of destination	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)
0	WORLD	NaN	900	NaN	1990	152563212.0
1	Developed regions	(b)	901	NaN	1990	82378628.0
2	Developing regions	(c)	902	NaN	1990	70184584.0
3	Least developed countries	(d)	941	NaN	1990	11075966.0
4	Less developed regions excluding least develop...	NaN	934	NaN	1990	59105261.0

Figure 1: International migrant stock at mid-year (both sexes)

The second step is to combine the three observations into a single table, which is convenient for us to separate the values for Major area, region, and country. The data for these three variables has been obtained according to the table "ANNEX", so use the function "$select_row$" to extract the data from table 1. The four images below show the original table and the table with extracted country, major area and region. Note the change in the second column.

	Major area, region, country or area of destination	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)
0	WORLD	NaN	900	NaN	1990	152563212.0
1	Developed regions	(b)	901	NaN	1990	82378628.0
2	Developing regions	(c)	902	NaN	1990	70184584.0
3	Least developed countries	(d)	941	NaN	1990	11075966.0
4	Less developed regions excluding least develop...	NaN	934	NaN	1990	59105261.0

Figure 2: Original table

	Country or area of destination	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)	International migrant stock at mid-year (male)	International migrant stock at mid-year (female)
1	Burundi	NaN	108	B R	1990	333110.0	163267.0	169843.0
2	Comoros	NaN	174	B	1990	14079.0	6717.0	7362.0
3	Djibouti	NaN	262	B R	1990	122221.0	64242.0	57979.0
4	Eritrea	NaN	232	I	1990	11848.0	6228.0	5620.0
5	Ethiopia	NaN	231	B R	1990	1155390.0	607284.0	548106.0

Figure 3: Include data for all countries only

	Major area	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)	International migrant stock at mid-year (male)	International migrant stock at mid-year (female)
1	Africa	NaN	903	NaN	1990	15690623.0	8279564.0	7411059.0
2	Asia	NaN	935	NaN	1990	48142261.0	26011875.0	22130386.0
3	Europe	NaN	908	NaN	1990	49219200.0	23946673.0	25272527.0
4	Latin America and the Caribbean	NaN	904	NaN	1990	7169728.0	3597037.0	3572691.0
5	Northern America	NaN	905	NaN	1990	27610542.0	13497319.0	14113223.0

Figure 4: Include data for all major area only

	Region	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)	International migrant stock at mid-year (male)	International migrant stock at mid-year (female)
1	Eastern Africa	NaN	910	NaN	1990	5964031.0	3071189.0	2892842.0
2	Middle Africa	NaN	911	NaN	1990	1460530.0	744494.0	716036.0
3	Northern Africa	NaN	912	NaN	1990	2403200.0	1230643.0	1172557.0
4	Southern Africa	NaN	913	NaN	1990	1392359.0	840899.0	551460.0
5	Western Africa	NaN	914	NaN	1990	4470503.0	2392339.0	2078164.0

Figure 5: Include data for all region only

The last step is to keep one observation in one table, so we'll continue to split the table above. Each table contains three observations male, female, and both sexes. Since male and female are variable values belonging to the variable name gender, we use the functions "*creat_both_sexes_table*" and "*creat_gender_table*" to generate two tables respectively. Since this is the last step, we will deal with missing values. There is only one observation in the table, so if the observation is missing, then this row is invalid data and needs to be deleted. When initially reading the table, we have used the function "*get_table*" to mark missing values with "NAN". So, the two functions used above already contain instructions to remove missing values. So according to the rules of cleaning data, we divided the first table 1 into 6 tables.

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Gender	International migrant stock at mid-year
1	Burundi	NaN	108	B R	1990	male	163267.0
2	Comoros	NaN	174	B	1990	male	6717.0
3	Djibouti	NaN	262	B R	1990	male	64242.0
4	Eritrea	NaN	232	I	1990	male	6228.0
5	Ethiopia	NaN	231	B R	1990	male	607284.0

Figure 6: International migrant stocks mid-year for all countries by gender

	Country or area of destination	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)
1	Burundi	NaN	108	B R	1990	333110.0
2	Comoros	NaN	174	B	1990	14079.0
3	Djibouti	NaN	262	B R	1990	122221.0
4	Eritrea	NaN	232	I	1990	11848.0
5	Ethiopia	NaN	231	B R	1990	1155390.0

Figure 7: International migrant stocks mid-year (both sexes) for all countries

	Major area	Notes	Country code	Type of data(a)	Year	Gender	International migrant stock at mid-year
1	Africa	NaN	903	NaN	1990	male	8279564.0
2	Asia	NaN	935	NaN	1990	male	26011875.0
3	Europe	NaN	908	NaN	1990	male	23946673.0
4	Latin America and the Caribbean	NaN	904	NaN	1990	male	3597037.0
5	Northern America	NaN	905	NaN	1990	male	13497319.0

Figure 8: International migrant stocks mid-year for all major area by gender

	Major area	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)
1	Africa	NaN	903	NaN	1990	15690623.0
2	Asia	NaN	935	NaN	1990	48142261.0
3	Europe	NaN	908	NaN	1990	49219200.0
4	Latin America and the Caribbean	NaN	904	NaN	1990	7169728.0
5	Northern America	NaN	905	NaN	1990	27610542.0

Figure 9: International migrant stocks mid-year (both sexes) for all major area

	Region	Notes	Country code	Type of data(a)	Year	Gender	International migrant stock at mid-year
1	Eastern Africa	NaN	910	NaN	1990	male	3071189.0
2	Middle Africa	NaN	911	NaN	1990	male	744494.0
3	Northern Africa	NaN	912	NaN	1990	male	1230643.0
4	Southern Africa	NaN	913	NaN	1990	male	840899.0
5	Western Africa	NaN	914	NaN	1990	male	2392339.0

Figure 10: International migrant stocks mid-year for all region by gender

	Region	Notes	Country code	Type of data(a)	Year	International migrant stock at mid-year (both sexes)
1	Eastern Africa	NaN	910	NaN	1990	5964031.0
2	Middle Africa	NaN	911	NaN	1990	1460530.0
3	Northern Africa	NaN	912	NaN	1990	2403200.0
4	Southern Africa	NaN	913	NaN	1990	1392359.0
5	Western Africa	NaN	914	NaN	1990	4470503.0

Figure 11: International migrant stocks mid-year (both sexes) for all region

Table 2: The only difference between Table 2 and Table 1 is the direction of the statistics. Table1 counts the immigrant population, while Table2 counts the total population. Since their data is presented in the same way, the data cleaning method is also the same. So, I will take the same approach. The following six figures show the final results of Table 2.

	Country or area of destination	Notes	Country code	Year	Gender	Total population at mid-year (thousands)
1	Burundi	NaN	108	1990	male	2755.028
2	Comoros	NaN	174	1990	male	208.212
3	Djibouti	NaN	262	1990	male	295.933
4	Eritrea	NaN	232	1990	male	1558.486
5	Ethiopia	NaN	231	1990	male	23965.647

Figure 12: Total population at mid-year for all countries by gender

	Country or area of destination	Notes	Country code	Year	Total population of both sexes at mid-year (thousands)
1	Burundi	NaN	108	1990	5613.141
2	Comoros	NaN	174	1990	415.144
3	Djibouti	NaN	262	1990	588.356
4	Eritrea	NaN	232	1990	3139.083
5	Ethiopia	NaN	231	1990	48057.094

Figure 13: Total population of both sexes at mid-year for all countries

	Major area	Notes	Country code	Year	Gender	Total population at mid-year (thousands)
1	Africa	NaN	903	1990	male	315071.378
2	Asia	NaN	935	1990	male	1634734.677
3	Europe	NaN	908	1990	male	347356.281
4	Latin America and the Caribbean	NaN	904	1990	male	221989.776
5	Northern America	NaN	905	1990	male	137757.875

Figure 14: Total population at mid-year for all major area by gender

	Major area	Notes	Country code	Year	Total population of both sexes at mid-year (thousands)
1	Africa	NaN	903	1990	631614.304
2	Asia	NaN	935	1990	3202474.692
3	Europe	NaN	908	1990	721086.311
4	Latin America and the Caribbean	NaN	904	1990	446888.767
5	Northern America	NaN	905	1990	280633.063

Figure 15: Total population of both sexes at mid-year for all major area

	Region	Notes	Country code	Year	Gender	Total population at mid-year (thousands)
1	Eastern Africa	NaN	910	1990	male	98208.646
2	Middle Africa	NaN	911	1990	male	35035.128
3	Northern Africa	NaN	912	1990	male	70480.841
4	Southern Africa	NaN	913	1990	male	20760.400
5	Western Africa	NaN	914	1990	male	90586.363

Figure 16: Total population at mid-year for all region area by gender

	Region	Notes	Country code	Year	Total population of both sexes at mid-year (thousands)
1	Eastern Africa	NaN	910	1990	198231.687
2	Middle Africa	NaN	911	1990	70886.433
3	Northern Africa	NaN	912	1990	140116.613
4	Southern Africa	NaN	913	1990	42049.013
5	Western Africa	NaN	914	1990	180330.558

Figure 17: Total population of both sexes at mid-year for all region area

Table3: The arrangement of Table 3 is the same as that of Tables 1 and 2, so the same method is used for data cleaning. Table 3 is also split into six new tables. But we noticed that the data in table 3 is related to both tables 1 and 2. The logic for table 3 is to divide the data in table 1 by the data in table 2. If the user wants to use these detailed data in a table, just use the command "table3 name [column name] = table1 or table 2 [column name]" and the column data of the table 1 or 2 will be inserted into table 3.

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Gender	International migrant stock as a percentage of the total population
1	Burundi	NaN	108	B R	1990	male	5.926147
2	Comoros	NaN	174	B	1990	male	3.226039
3	Djibouti	NaN	262	B R	1990	male	21.708292
4	Eritrea	NaN	232	I	1990	male	0.399619
5	Ethiopia	NaN	231	B R	1990	male	2.533977

Figure 18: International migrant stock as a percentage of the total population for all countries by gender

	Country or area of destination	Notes	Country code	Type of data(a)	Year	International migrant stock as a percentage of the total population (both sexes)
1	Burundi	NaN	108	B R	1990	5.934467
2	Comoros	NaN	174	B	1990	3.391353
3	Djibouti	NaN	262	B R	1990	20.773307
4	Eritrea	NaN	232	I	1990	0.377435
5	Ethiopia	NaN	231	B R	1990	2.404203

Figure 19: International migrant stock as a percentage of the total population (both sexes) for all countries

	Major area	Notes	Country code	Type of data(a)	Year	Gender	International migrant stock as a percentage of the total population
1	Africa	NaN	903	NaN	1990	male	2.627838
2	Asia	NaN	935	NaN	1990	male	1.591199
3	Europe	NaN	908	NaN	1990	male	6.893980
4	Latin America and the Caribbean	NaN	904	NaN	1990	male	1.620362
5	Northern America	NaN	905	NaN	1990	male	9.797857

Figure 20: International migrant stock as a percentage of the total population for all major area by gender

	Major area	Notes	Country code	Type of data(a)	Year	International migrant stock as a percentage of the total population (both sexes)
1	Africa	NaN	903	NaN	1990	2.484210
2	Asia	NaN	935	NaN	1990	1.503283
3	Europe	NaN	908	NaN	1990	6.825702
4	Latin America and the Caribbean	NaN	904	NaN	1990	1.604365
5	Northern America	NaN	905	NaN	1990	9.838663

Figure 21: International migrant stock as a percentage of the total population (both sexes) for all major area

	Region	Notes	Country code	Type of data(a)	Year	Gender	International migrant stock as a percentage of the total population
1	Eastern Africa	NaN	910	NaN	1990	male	3.127208
2	Middle Africa	NaN	911	NaN	1990	male	2.124993
3	Northern Africa	NaN	912	NaN	1990	male	1.746067
4	Southern Africa	NaN	913	NaN	1990	male	4.050495
5	Western Africa	NaN	914	NaN	1990	male	2.640948

Figure 22: International migrant stock as a percentage of the total population for all region area by gender

	Region	Notes	Country code	Type of data(a)	Year	International migrant stock as a percentage of the total population (both sexes)
1	Eastern Africa	NaN	910	NaN	1990	3.008616
2	Middle Africa	NaN	911	NaN	1990	2.060380
3	Northern Africa	NaN	912	NaN	1990	1.715143
4	Southern Africa	NaN	913	NaN	1990	3.311276
5	Western Africa	NaN	914	NaN	1990	2.479060

Figure 23: International migrant stock as a percentage of the total population (both sexes) for all region area

Table 4: Table 4 has only one observation, so we just need to follow the previous steps. No need to split the table by columns

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Female migrants as a percentage of the international migrant stock
1	Burundi	NaN	108	B R	1990	50.987061
2	Comoros	NaN	174	B	1990	52.290646
3	Djibouti	NaN	262	B R	1990	47.437838
4	Eritrea	NaN	232	I	1990	47.434166
5	Ethiopia	NaN	231	B R	1990	47.439047

Figure 24: Female migrants as a percentage of the international migrant stock for all countries

	Major area	Notes	Country code	Type of data(a)	Year	Female migrants as a percentage of the international migrant stock
1	Africa	NaN	903	NaN	1990	47.232408
2	Asia	NaN	935	NaN	1990	45.968730
3	Europe	NaN	908	NaN	1990	51.346887
4	Latin America and the Caribbean	NaN	904	NaN	1990	49.830217
5	Northern America	NaN	905	NaN	1990	51.115342

Figure 25: Female migrants as a percentage of the international migrant stock for all major area

	Region	Notes	Country code	Type of data(a)	Year	Female migrants as a percentage of the international migrant stock
1	Eastern Africa	NaN	910	NaN	1990	48.504812
2	Middle Africa	NaN	911	NaN	1990	49.025765
3	Northern Africa	NaN	912	NaN	1990	48.791486
4	Southern Africa	NaN	913	NaN	1990	39.606165
5	Western Africa	NaN	914	NaN	1990	46.486134

Figure 26: Female migrants as a percentage of the international migrant stock for all region

Table 5: Table 5 still uses the same cleaning method and is divided into 6 tables in total.

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Gender	Annual rate of change of the migrant stock
1	Burundi	NaN	108	B R	1990-1995	male	-5.475511
2	Comoros	NaN	174	B	1990-1995	male	-0.309060
3	Djibouti	NaN	262	B R	1990-1995	male	-4.046026
4	Eritrea	NaN	232	I	1990-1995	male	0.983754
5	Ethiopia	NaN	231	B R	1990-1995	male	-7.179744

Figure 27: Annual rate of change of the refugee stock for all countries by gender

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Annual rate of change of the migrant stock (both sexes)
1	Burundi	NaN	108	B R	1990-1995	-5.355717
2	Comoros	NaN	174	B	1990-1995	-0.199873
3	Djibouti	NaN	262	B R	1990-1995	-4.058465
4	Eritrea	NaN	232	I	1990-1995	0.910748
5	Ethiopia	NaN	231	B R	1990-1995	-7.179771

Figure 28: Annual rate of change of the refugee stock (both sexes) for all countries

	Major area	Notes	Country code	Type of data(a)	Year	Gender	Annual rate of change of the migrant stock
1	Africa	NaN	903	NaN	1990-1995	male	0.798774
2	Asia	NaN	935	NaN	1990-1995	male	-0.636150
3	Europe	NaN	908	NaN	1990-1995	male	1.363213
4	Latin America and the Caribbean	NaN	904	NaN	1990-1995	male	-1.424381
5	Northern America	NaN	905	NaN	1990-1995	male	3.898245

Figure 29: Annual rate of change of the refugee stock for all major area by gender

	Major area	Notes	Country code	Type of data(a)	Year	Annual rate of change of the migrant stock (both sexes)
1	Africa	NaN	903	NaN	1990-1995	0.826734
2	Asia	NaN	935	NaN	1990-1995	-0.673431
3	Europe	NaN	908	NaN	1990-1995	1.420702
4	Latin America and the Caribbean	NaN	904	NaN	1990-1995	-1.371210
5	Northern America	NaN	905	NaN	1990-1995	3.771892

Figure 30: Annual rate of change of the refugee stock (both sexes) for all major area

	Region	Notes	Country code	Type of data(a)	Year	Gender	Annual rate of change of the migrant stock
1	Eastern Africa	NaN	910	NaN	1990-1995	male	-3.446375
2	Middle Africa	NaN	911	NaN	1990-1995	male	11.845106
3	Northern Africa	NaN	912	NaN	1990-1995	male	-2.139210
4	Southern Africa	NaN	913	NaN	1990-1995	male	-3.266338
5	Western Africa	NaN	914	NaN	1990-1995	male	3.611415

Figure 31: Annual rate of change of the refugee stock for all region area by gender

	Region	Notes	Country code	Type of data(a)	Year	Annual rate of change of the migrant stock (both sexes)
1	Eastern Africa	NaN	910	NaN	1990-1995	-3.435412
2	Middle Africa	NaN	911	NaN	1990-1995	11.885810
3	Northern Africa	NaN	912	NaN	1990-1995	-2.872903
4	Southern Africa	NaN	913	NaN	1990-1995	-3.114352
5	Western Africa	NaN	914	NaN	1990-1995	3.817706

Figure 32: Annual rate of change of the refugee stock (both sexes) for all region area

Table6: The difference between Table 6 and the previous table is that there are 3 independent observations that do not need to be combined. So according to the cleaning rules, it is divided into 9 tables.

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Estimated refugee stock at mid-year (both sexes)
1	Burundi	NaN	108	B R	1990	267929.0
2	Comoros	NaN	174	B	1990	0.0
3	Djibouti	NaN	262	B R	1990	54508.0
4	Eritrea	NaN	232	I	1990	0.0
5	Ethiopia	NaN	231	B R	1990	741965.0

Figure 33: Estimated refugee stock at mid-year (both sexes) for all countries

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Refugees as a percentage of the international migrant stock
1	Burundi	NaN	108	B R	1990	80.432590
2	Comoros	NaN	174	B	1990	0.000000
3	Djibouti	NaN	262	B R	1990	44.597901
4	Eritrea	NaN	232	I	1990	0.000000
5	Ethiopia	NaN	231	B R	1990	64.217710

Figure 34: Refugees as a percentage of the international migrant stock for all countries

	Country or area of destination	Notes	Country code	Type of data(a)	Year	Annual rate of change of the refugee stock
1	Burundi	NaN	108	B R	1990	-3.390926
2	Comoros	NaN	174	B	1990	NaN
3	Djibouti	NaN	262	B R	1990	-9.763426
4	Eritrea	NaN	232	I	1990	NaN
5	Ethiopia	NaN	231	B R	1990	-5.505717

Figure 35: Annual rate of change of the refugee stock for all countries

	Major area	Notes	Country code	Type of data(a)	Year	Estimated refugee stock at mid-year (both sexes)
1	Africa	NaN	903	NaN	1990	5687352.0
2	Asia	NaN	935	NaN	1990	9937007.0
3	Europe	NaN	908	NaN	1990	1321884.0
4	Latin America and the Caribbean	NaN	904	NaN	1990	1197198.0
5	Northern America	NaN	905	NaN	1990	583450.0

Figure 36: Estimated refugee stock at mid-year (both sexes) for all major area

	Major area	Notes	Country code	Type of data(a)	Year	Refugees as a percentage of the international migrant stock
1	Africa	NaN	903	NaN	1990	36.246821
2	Asia	NaN	935	NaN	1990	20.640923
3	Europe	NaN	908	NaN	1990	2.685708
4	Latin America and the Caribbean	NaN	904	NaN	1990	16.697956
5	Northern America	NaN	905	NaN	1990	2.113142

Figure 37: Refugees as a percentage of the international migrant stock for all major area

	Major area	Notes	Country code	Type of data(a)	Year	Annual rate of change of the refugee stock
1	Africa	NaN	903	NaN	1990	0.076037
2	Asia	NaN	935	NaN	1990	-3.819461
3	Europe	NaN	908	NaN	1990	13.201700
4	Latin America and the Caribbean	NaN	904	NaN	1990	-23.096408
5	Northern America	NaN	905	NaN	1990	1.917003

Figure 38: Annual rate of change of the refugee stock for all major area

	Region	Notes	Country code	Type of data(a)	Year	Estimated refugee stock at mid-year (both sexes)
1	Eastern Africa	NaN	910	NaN	1990	3168001.0
2	Middle Africa	NaN	911	NaN	1990	446609.0
3	Northern Africa	NaN	912	NaN	1990	1202360.0
4	Southern Africa	NaN	913	NaN	1990	135525.0
5	Western Africa	NaN	914	NaN	1990	734857.0

Figure 39: Estimated refugee stock at mid-year (both sexes) for all region

	Region	Notes	Country code	Type of data(a)	Year	Refugees as a percentage of the international migrant stock
1	Eastern Africa	NaN	910	NaN	1990	53.118453
2	Middle Africa	NaN	911	NaN	1990	30.578557
3	Northern Africa	NaN	912	NaN	1990	50.031625
4	Southern Africa	NaN	913	NaN	1990	9.733481
5	Western Africa	NaN	914	NaN	1990	16.437904

Figure 40: Refugees as a percentage of the international migrant stock for all region

	Region	Notes	Country code	Type of data(a)	Year	Annual rate of change of the refugee stock
1	Eastern Africa	NaN	910	NaN	1990	-5.308010
2	Middle Africa	NaN	911	NaN	1990	12.964162
3	Northern Africa	NaN	912	NaN	1990	-3.456178
4	Southern Africa	NaN	913	NaN	1990	-1.954547
5	Western Africa	NaN	914	NaN	1990	8.717581

Figure 41: Refugees as a percentage of the international migrant stock for all region

4 SUMMARY

Compared with the data before and after cleaning, the new tabular data is more standardized and easier to understand. Cleaning the data can also improve efficiency, because the new data is error-free and consistent, so future developers can focus more on algorithm development and analysis. The new data also removes invalid values and retains valid data, thus improving accuracy. Unfortunately, there is only one data source, if other data sources can be used for auxiliary verification, then the accuracy will be greatly improved. Because only better data can make better analysis and decision-making, whether it is the development direction of the enterprise or the investment plan, it needs huge data support.