

INF1340 Midterm Project
UN Dataset Cleaning Write-Up

Mingli Wang

1003100716

INF1340
Professor Shion Guha
University of Toronto
November 16, 2022

Table of Contents

Introduction.....	3
Preparation: import libraries and upload the raw dataset.....	3
Defining some frequently used functions and variables.....	3
Cleaning Table 1, 2, and 3	5
Cleaning Table 4.....	6
Cleaning Table 5.....	9
Cleaning Table 6.....	10
Result	11
Screenshots of Cleaned Data Frames	13

Introduction

The dataset: *Trends in International Migrant Stock: The 2015 Revision* is a dataset from the UN's Department of Economic and Social Affairs. The raw dataset is in excel format and its contents are as follows:

TABLE	TITLE
Table 1	International migrant stock at mid-year by sex and by major area, region, country or area, 1990-2015
Table 2	Total population at mid-year by sex and by major area, region, country or area, 1990-2015 (thousands)
Table 3	International migrant stock as a percentage of the total population, 1990-2015
Table 4	Female migrants as a percentage of the international migrant stock by major area, region, country or area, 1990-2015
Table 5	Annual rate of change of the migrant stock by sex and by major area, region, country or area, 1990-2015 (percentage)
Table 6	Estimated refugee stock at mid-year by major area, region, country or area, 1990-2015
ANNEX	Classification of countries and areas by major area and region
NOTES	NOTES

The goal is to clean and tidy up data in Table 1 to 6 following tidy data principles, other sheets in the raw dataset will not be altered by the script. The cleaning and tidying process eventually generates 24 cleaned data frames as result.

Preparation: import libraries and upload the raw dataset

This notebook is written on Google Colab. Pandas is the primary library used in this script, so along with other useful libraries, they are imported to this notebook. The notebook will then ask you to upload the raw UN dataset from your local drives.

Defining some frequently used functions and variables

Throughout the data cleaning process, there are some repeated actions performed on different tables from the raw dataset, thus I have defined 7 customized functions for these repeated cleaning actions so that I can keep the code concise and the variables easy to manage.

```
def prep(table, tb_name, col_name):  
    table = pd.read_excel('UN_MigrantStockTotal_2015.xlsx', tb_name)  
    table.columns = col_name  
    table = table.drop(table.index[0 : 21])  
    return table
```

The first function, `prep`, reads a targeted table from the raw data, renames the columns, and then drops the table's first 22 rows which include a descriptive header, the original column names, and the first 6 rows of data. Although the header contains good contextual information, it will impede the data cleaning, thus it is temporarily dropped. The original column names are

replaced with a new shorter set of names for easier coding. Once the data cleaning process is complete, these column names will be replaced by a complete and descriptive set of names. The first 6 rows of data in every table (*WORLD, Developed regions, Developing regions, Least developed countries, Less developed regions excluding least developed countries, Sub-Saharan Africa*) are dropped because they are summative data calculated based on values below, which makes them not unique observations. If these 6 rows are kept, they will cause complications in future data science processes after the cleaning.

The `melt_func` is directly derived from pandas melt function for the purpose of easier parameter management. When melting another dataset, it allows me to enter the parameters all together inside the bracket without the need to go through the code line by line to alter the parameters. The same principle also applies to `split_year` and `split_period` functions designed to split a column called “year and gender” or “period and gender” into two columns with each containing one variable.

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

def split_year(table):
    return table.join(table['year and gender'].str.partition(' ')[[0, 2]].rename({0: 'year', 2: 'gender'}, axis=1)).drop("year and gender", axis = 1)

def split_period(table):
    return table.join(table['period and gender'].str.partition(' ')[[0, 2]].rename({0: 'period', 2: 'gender'}, axis=1)).drop("period and gender", axis = 1)
```

A set of frequently used variables are also defined in this step, they are parameters for the functions just defined. In the raw dataset, gender and year are stored separately in 2 rows (figure 1.1), so during the column renaming, year/period and gender information are combined into 1 row, for example, “male” and “1990” are now “1990 m”.

Total male population at mid-year (thousands)					
1990	1995	2000	2005	2010	2015

(Figure 1.1)

Hence, a list of column names for a dataset of males from 1990 to 2015 would be: ['1990 m', '1995 m', '2000 m', '2005 m', '2010 m', '2015 m']. Lists like this are then stored in variables for `melt_func` to use as parameters. (figure 1.2)

```
value_vars_b = ['1990 all', '1995 all', '2000 all', '2005 all', '2010 all', '2015 all']
value_vars_m = ['1990 m', '1995 m', '2000 m', '2005 m', '2010 m', '2015 m']
value_vars_f = ['1990 f', '1995 f', '2000 f', '2005 f', '2010 f', '2015 f']

value_vars_b_p = ['1990-1995 all', '1995-2000 all', '2000-2005 all', '2005-2010 all', '2010-2015 all']
value_vars_m_p = ['1990-1995 m', '1995-2000 m', '2000-2005 m', '2005-2010 m', '2010-2015 m']
value_vars_f_p = ['1990-1995 f', '1995-2000 f', '2000-2005 f', '2005-2010 f', '2010-2015 f']
```

(Figure 1.2 Variable Names)

In all 6 tables of the UN dataset, there is a column named “Major area, region, country or area of destination”. This column stores 3 types of variables, which is a violation of tidy data principles. Therefore, to separate major areas, regions, and countries or areas of destination, a list of major areas and a list of regions from the “ANNEX” table are extracted and then the two

lists are then used as the index to split the messy column from the raw dataset. Since all 6 tables require this separation procedure, the final 3 functions are defined for this procedure: `extract_country`, `extract_region`, and `extract_majorarea`.

```
#Define function for extracting country and area from column "area"
def extract_country(table):
    table = table[~table['area'].isin(region_list)]
    table = table[~table['area'].isin(major_area_list)]
    return table

#Define function for extracting regions from column "area"
def extract_region(table):
    table = table[table['area'].isin(region_list)]
    return table

#Define function for extracting major area from column "area"
def extract_majorarea(table):
    table = table[table['area'].isin(major_area_list)]
    return table
```

Cleaning Table 1, 2, and 3

Table 1, 2, and 3 share a similar structure and they have the same problems, hence the cleaning procedure taken on these 3 tables is identical.

Step 1: The procedure starts with importing the table into Notebook using the `prep` function. With this function, the first 22 rows in the excel are dropped (detailed reasoning was mentioned while defining the function) and a tidier set of column names is assigned to the data frame (figure 2.1).

order		area	notes	country_code	type	1990	all	1995	all	2000	all	2005	all	2010	all	...
21	7	Africa	NaN	903	NaN	15690623	16352814	14800306	15191146	16840014.0	...					
22	8	Eastern Africa	NaN	910	NaN	5964031	5022742	4844795	4745792	4657063.0	...					
23	9	Burundi	NaN	108	B R	333110	254853	125628	172874	235259.0	...					
24	10	Comoros	NaN	174	B	14079	13939	13799	13209	12618.0	...					
25	11	Djibouti	NaN	262	B R	122221	99774	100507	92091	101575.0	...					
...					
275	261	Samoa	NaN	882	B	3357	4694	5998	5746	5122.0	...					
276	262	Tokelau	NaN	772	B	270	266	262	258	429.0	...					
277	263	Tonga	NaN	776	B	2911	3274	3684	4301	5022.0	...					
278	264	Tuvalu	NaN	798	C	318	263	217	183	154.0	...					
279	265	Wallis and Futuna Islands	NaN	876	B	1402	1680	2015	2365	2776.0	...					

259 rows x 23 columns

(Figure 2.1 Table 1 after the `Prep` function)

Step 2: Each of Table 1 to 3 holds data for 3 categories based on gender - male, female, and both sexes. Under each category, there are 6 columns based on year. The primary problem of these 3 tables is that data from different years are stored horizontally instead of in the same column. Moreover, data of males, females, and both sexes are stored horizontally as well. Since year is a variable and should be stored in one column, the first step is to melt each of the 3

categories' 6 columns under years into one column using `melt_func` (figure 2.2), and then the “year and gender” column is split into two columns named “year” and “gender” using `split_year`.

order		area	notes	country_code	type	International migrant stock at mid-year (male)	year	gender
0	7	Africa	NaN	903	NaN	8279564	1990	m
1	8	Eastern Africa	NaN	910	NaN	3071189	1990	m
2	9	Burundi	NaN	108	B R	163267	1990	m
3	10	Comoros	NaN	174	B	6717	1990	m
4	11	Djibouti	NaN	262	B R	64242	1990	m
...
1549	261	Samoa	NaN	882	B	2469.0	2015	m
1550	262	Tokelau	NaN	772	B	233.0	2015	m
1551	263	Tonga	NaN	776	B	3127.0	2015	m
1552	264	Tuvalu	NaN	798	C	78.0	2015	m
1553	265	Wallis and Futuna Islands	NaN	876	B	1438.0	2015	m

1554 rows x 8 columns

(Figure 2.2 The 6 columns based on years under each gender is melted into one)

Step 3: The same procedure is conducted on all 3 categories of male, female, and both sexes. The resulting 3 new columns are then joined together (figure 2.3). Because the values under “both sexes” are summative results of the values of male and female combined, the “both sexes” column is not a unique observation, and the column is therefore dropped.

order		area	notes	country_code	type	year	both sexes	male	female
0	7	Africa	NaN	903	NaN	1990	15690623	8279564	7411059
1	8	Eastern Africa	NaN	910	NaN	1990	5964031	3071189	2892842
2	9	Burundi	NaN	108	B R	1990	333110	163267	169843
3	10	Comoros	NaN	174	B	1990	14079	6717	7362
4	11	Djibouti	NaN	262	B R	1990	122221	64242	57979
...
1549	261	Samoa	NaN	882	B	2015	4929.0	2469.0	2460.0
1550	262	Tokelau	NaN	772	B	2015	487.0	233.0	254.0
1551	263	Tonga	NaN	776	B	2015	5731.0	3127.0	2604.0
1552	264	Tuvalu	NaN	798	C	2015	141.0	78.0	63.0
1553	265	Wallis and Futuna Islands	NaN	876	B	2015	2849.0	1438.0	1411.0

1554 rows x 9 columns

(Figure 2.3 The 2 new columns are joined in one data frame)

Step 4: Since gender is also a variable and data of different genders should not be stored horizontally, the 2 new columns are again melted into 1 column (figure 2.4).

	Order		Region	Notes	Country Code	Type	Year	Gender	International migrant stock at mid-year
1	8		Eastern Africa	NaN	910	NaN	1990	male	3071189
22	29		Middle Africa	NaN	911	NaN	1990	male	744494
32	39		Northern Africa	NaN	912	NaN	1990	male	1230643
40	47		Southern Africa	NaN	913	NaN	1990	male	840899
46	53		Western Africa	NaN	914	NaN	1990	male	2392339
...
3074	232		Northern America	NaN	905	NaN	2015	female	27902348.0
3081	239		Australia and New Zealand	NaN	927	NaN	2015	female	3963032.0
3084	242		Melanesia	NaN	928	NaN	2015	female	47782.0
3090	248		Micronesia	NaN	954	NaN	2015	female	57159.0
3098	256		Polynesia	NaN	957	NaN	2015	female	33361.0

264 rows × 8 columns

(Figure 2.4 Data from 3 genders are melted into one column)

Step 5: The final step for Table 1, 2, and 3 is to separate major areas and regions from countries and areas, because they are 3 different types of variables and should be stored in 3 separate data frames. Using the `extract_country`, `extract_region`, and `extract_majorarea` functions previously defined in step 2, data for major areas and regions and separated from countries. Empty values in the raw data marked with “.” are replaced with “NaN” for better readability, the columns are also properly renamed, and these data frames are now cleaned (figure 2.5).

	Order		Region	Notes	Country Code	Type	Year	Gender	International migrant stock at mid-year
1	8		Eastern Africa	NaN	910	NaN	1990	male	3071189
22	29		Middle Africa	NaN	911	NaN	1990	male	744494
32	39		Northern Africa	NaN	912	NaN	1990	male	1230643
40	47		Southern Africa	NaN	913	NaN	1990	male	840899
46	53		Western Africa	NaN	914	NaN	1990	male	2392339
...
3074	232		Northern America	NaN	905	NaN	2015	female	27902348.0
3081	239		Australia and New Zealand	NaN	927	NaN	2015	female	3963032.0
3084	242		Melanesia	NaN	928	NaN	2015	female	47782.0
3090	248		Micronesia	NaN	954	NaN	2015	female	57159.0
3098	256		Polynesia	NaN	957	NaN	2015	female	33361.0

264 rows × 8 columns

(Figure 2.5: Example of separated data frame for regions)

Cleaning Table 4

Table 4 is different from the previous 3 tables because it only contains data regarding females, so it requires a slightly different approach, although the 6 columns of data based on year still need to be melted into one column.

Step 1: The cleaning procedure starts with the `prep` function, which again dropped the header, original column names, and the first 6 rows of redundant data (figure 3.1).

	order	area	notes	country_code	type	1990 f	1995 f	2000 f	2005 f	2010 f	2015 f
21	7	Africa	NaN	903	NaN	47.232408	47.306127	46.917598	45.814246	46.322408	46.132389
22	8	Eastern Africa	NaN	910	NaN	48.504812	48.533032	48.798989	46.700993	49.190810	49.272007
23	9	Burundi	NaN	108	B R	50.987061	51.279757	51.369122	50.94404	50.767877	50.730100
24	10	Comoros	NaN	174	B	52.290646	52.550398	52.815421	52.411235	51.973371	51.644763
25	11	Djibouti	NaN	262	B R	47.437838	47.405136	47.346951	44.277942	47.531381	47.413908
...
275	261	Samoa	NaN	882	B	47.244564	47.784406	48.299433	48.833971	49.355720	49.908704
276	262	Tokelau	NaN	772	B	44.444444	44.736842	45.038168	48.449612	51.981352	52.156057
277	263	Tonga	NaN	776	B	48.883545	47.525962	46.226927	45.873053	45.698925	45.437096
278	264	Tuvalu	NaN	798	C	43.396226	43.726236	44.239631	44.808743	44.805195	44.680851
279	265	Wallis and Futuna Islands	NaN	876	B	48.216833	48.869048	49.478908	49.513742	49.531700	49.526150

259 rows x 11 columns

(Figure 3.1 Table 4 after the `prep` function)

Step 2: Since “year” is a variable and according to tidy data principles, they should be stored vertically in one column, thus the last 6 rows of data based on years from 1990 – 2015 are melted using the `melt_func` (figure 3.2).

	order	area	notes	country_code	type	Female migrants as a percentage of the international migrant stock	year	gender
0	7	Africa	NaN	903	NaN	47.232408	1990	f
1	8	Eastern Africa	NaN	910	NaN	48.504812	1990	f
2	9	Burundi	NaN	108	B R	50.987061	1990	f
3	10	Comoros	NaN	174	B	52.290646	1990	f
4	11	Djibouti	NaN	262	B R	47.437838	1990	f
...
1549	261	Samoa	NaN	882	B	49.908704	2015	f
1550	262	Tokelau	NaN	772	B	52.156057	2015	f
1551	263	Tonga	NaN	776	B	45.437096	2015	f
1552	264	Tuvalu	NaN	798	C	44.680851	2015	f
1553	265	Wallis and Futuna Islands	NaN	876	B	49.52615	2015	f

1554 rows x 8 columns

(Figure 3.2 The new column resulting from the `melt_func`)

Step 3: Finally, major areas and regions are again separated from the “area” column and stored in a data frame respectively. The columns are properly reordered and renamed, and empty values are replaced with “NaN” (figure 3.3).

Order	Country and Area	Notes	Country Code	Type	Year	Gender	Female migrants as a percentage of the international migrant stock
2	Burundi	NaN	108	B R	1990	female	50.987061
3	Comoros	NaN	174	B	1990	female	52.290646
4	Djibouti	NaN	262	B R	1990	female	47.437838
5	Eritrea	NaN	232	I	1990	female	47.434166
6	Ethiopia	NaN	231	B R	1990	female	47.439047
...
1549	Samoa	NaN	882	B	2015	female	49.908704
1550	Tokelau	NaN	772	B	2015	female	52.156057
1551	Tonga	NaN	776	B	2015	female	45.437096
1552	Tuvalu	NaN	798	C	2015	female	44.680851
1553	Wallis and Futuna Islands	NaN	876	B	2015	female	49.52615

1392 rows x 8 columns

(Figure 3.3: Example of separated data frame for countries and areas)

Cleaning Table 5

Table 5 has a similar structure to Table 1, 2, and 3, data for male, female, and both sexes are listed horizontally in 3 categories. But unlike Table 1 – 3 which has 6 columns of data based on years under each gender category, Table 5 has 5 columns based on periods under each gender category, therefore it requires a slight adjustment in the codes. However, Table 5 violates the same principles of tidy data, namely the horizontally stored variables such as periods and genders, and the messy area column where major areas and countries are mixed in one column.

Step 1: Using the `prep` function, load the table from raw data, rename the columns, and drop the first 6 rows of redundant data (figure 4.1).

order	area	notes	country_code	type	1990-1995 all	1995-2000 all	2000-2005 all	2005-2010 all	2010-2015 all	1990-1995 m	1995-2000 m	2000-2005 m	2005-2010 m	2010-2015 m	1990-1995 f	1995-2000 f	2000-2005 f	2005-2010 f	2010-2015 f	
21	7	Africa	NaN	903	NaN	0.826734	-1.995043	0.521296	2.582902	4.078721	0.798774	-1.846117	0.932749	1.872454	4.149296	0.857926	-2.189983	0.045343	2.281516	3.99651
22	8	Eastern Africa	NaN	910	NaN	-3.435412	-0.721421	-0.412932	-0.377467	5.493301	-3.446375	-0.82504	0.390236	-1.334276	5.461314	-3.423779	-0.612122	-1.291815	0.661361	5.526287
23	9	Burundi	NaN	108	B R	-5.355717	-14.147235	6.384757	6.162481	3.962659	-5.475511	-14.183953	6.558817	6.234173	3.977999	-5.241234	-14.112412	6.218568	6.093201	3.947771
24	10	Comoros	NaN	174	B	-0.199873	-0.201891	-0.873954	-0.915481	-0.100107	-0.30906	-0.313911	-0.703362	-0.732303	0.036271	-0.10077	-0.10128	-1.027599	-1.08327	-0.226961
25	11	Djibouti	NaN	262	B R	-4.058465	0.146395	-1.749003	1.960404	2.016609	-4.048026	0.168508	-0.615965	0.757188	2.061337	-4.072257	0.121832	-3.089319	3.378474	1.967118
...
275	261	Samoa	NaN	882	B	6.704748	4.90282	-0.858442	-2.299179	-0.768177	6.499035	4.704571	-1.066301	-2.50417	-0.987758	6.931983	5.117229	-0.638315	-2.08663	-0.545343
276	262	Tokelau	NaN	772	B	-0.298513	-0.303036	-0.307698	10.169947	2.536144	-0.404054	-0.412386	-1.589283	8.750541	2.463246	-0.167365	-0.168777	1.152882	11.577161	2.60325
277	263	Tonga	NaN	776	B	2.350316	2.359733	3.096969	3.099614	2.641235	2.874558	2.648819	3.228155	3.163851	2.737439	1.787022	1.80546	2.943277	3.023552	2.526318
278	264	Tuvalu	NaN	798	C	-3.797947	-3.845134	-3.408224	-3.450671	-1.763854	-3.914892	-4.028435	-3.613401	-3.449385	-1.718849	-3.646431	-3.611679	-3.152079	-3.452255	-1.819436
279	265	Wallis and Futuna Islands	NaN	876	B	3.61788	3.636508	3.203177	3.20466	0.51914	3.364378	3.396526	3.189382	3.197545	0.52134	3.886601	3.884553	3.217252	3.211913	0.516899

259 rows x 20 columns

(Figure 4.1: Columns are renamed for easier coding)

Step 2: Using the `melt_func`, the 5 columns of data based on periods under every gender category are melted into one column, generating 3 new columns in total (figure 4.2). Again, because the values under “both sexes” are the summative results of the values of male and female combined, the “both sexes” column is not a unique observation, and the column is therefore dropped.

order	area	notes	country_code	type	period	Annual rate of change of the migrant stock (both sexes)	Annual rate of change of the migrant stock (male)	Annual rate of change of the migrant stock (female)
0	7	Africa	NaN	903	1990-1995	0.826734	0.798774	0.857926
1	8	Eastern Africa	NaN	910	1990-1995	-3.435412	-3.446375	-3.423779
2	9	Burundi	NaN	108	B R 1990-1995	-5.355717	-5.475511	-5.241234
3	10	Comoros	NaN	174	B 1990-1995	-0.199873	-0.30906	-0.10077
4	11	Djibouti	NaN	262	B R 1990-1995	-4.058465	-4.048026	-4.072257
...
1290	261	Samoa	NaN	882	B 2010-2015	-0.768177	-0.987758	-0.545343
1291	262	Tokelau	NaN	772	B 2010-2015	2.536144	2.463246	2.60325
1292	263	Tonga	NaN	776	B 2010-2015	2.641235	2.737439	2.526318
1293	264	Tuvalu	NaN	798	C 2010-2015	-1.763854	-1.718849	-1.819436
1294	265	Wallis and Futuna Islands	NaN	876	B 2010-2015	0.51914	0.52134	0.516899

(Figure 4.2: 3 new columns from the melting are joined together)

Step 3: The 3 new columns with data for each gender are again melted into a single column because genders are variables and should be stored in cells (figure 4.3).

	Order	Country and Area	Notes	Country Code	Type	Year	Gender	Annual rate of change of the migrant stock
2	9	Burundi	NaN	108	B R	1990-1995	male	-5.475511
3	10	Comoros	NaN	174	B	1990-1995	male	-0.30906
4	11	Djibouti	NaN	262	B R	1990-1995	male	-4.046026
5	12	Eritrea	NaN	232	I	1990-1995	male	0.983754
6	13	Ethiopia	NaN	231	B R	1990-1995	male	-7.179744
...
2585	261	Samoa	NaN	882	B	2010-2015	female	-0.545343
2586	262	Tokelau	NaN	772	B	2010-2015	female	2.60325
2587	263	Tonga	NaN	776	B	2010-2015	female	2.526318
2588	264	Tuvalu	NaN	798	C	2010-2015	female	-1.819436
2589	265	Wallis and Futuna Islands	NaN	876	B	2010-2015	female	0.516899

2320 rows × 8 columns

(Figure 4.3: 3 columns for each gender are melted into one)

Step 4: Split major areas and regions from countries and areas and put the split data in their own data frames using the `extract_country`, `extract_region`, and `extract_majorarea` functions (figure 4.4).

	Order	Country and Area	Notes	Country Code	Type	Year	Gender	Annual rate of change of the migrant stock
2	9	Burundi	NaN	108	B R	1990-1995	male	-5.475511
3	10	Comoros	NaN	174	B	1990-1995	male	-0.30906
4	11	Djibouti	NaN	262	B R	1990-1995	male	-4.046026
5	12	Eritrea	NaN	232	I	1990-1995	male	0.983754
6	13	Ethiopia	NaN	231	B R	1990-1995	male	-7.179744
...
2585	261	Samoa	NaN	882	B	2010-2015	female	-0.545343
2586	262	Tokelau	NaN	772	B	2010-2015	female	2.60325
2587	263	Tonga	NaN	776	B	2010-2015	female	2.526318
2588	264	Tuvalu	NaN	798	C	2010-2015	female	-1.819436
2589	265	Wallis and Futuna Islands	NaN	876	B	2010-2015	female	0.516899

2320 rows × 8 columns

(Figure 4.4: Example of separated data frame for countries and areas)

Cleaning Table 6

Table 6 holds 3 sets of data – estimated refugee stock based on years, refugees as a percentage of the international migrant stock based on years, and annual rate of change of the refugee stock based on periods. These 3 sets of data are different in type and should be split into separate tables. In addition, each of the 3 sets has used variables as column headers, which also violates tidy data principles and needs to be addressed.

Step 1: Using the `prep` function, load the table from raw data, rename the columns, and drop the first 6 rows of redundant data (figure 5.1).

order	area	notes	country_code	type	1990	1995	2000	2005	2010	...	1995 b	2000 b	2005 b	2010 b	2015 b	1990-1995	1995-2000	2000-2005	2005-2010	2010-2015
21	7	Africa	NaN	903	NaN	5687352	5949953	3609138	2750644	2413917.0	36.384888	24.385563	18.106889	14.334412	19.479173	0.076037	-8.003244	-5.953976	-4.672588	6.133656
22	8	Eastern Africa	NaN	910	NaN	3168001	2046088	1641559	1419685	1006358.0	40.736474	33.88294	29.914506	21.652230	34.058990	-5.30801	-3.684244	-2.491296	-6.464768	9.059711
23	9	Burundi	NaN	108	B R	267929	173017	27136	20681	29365.0	67.88894	21.60028	11.963048	12.481988	16.667829	-3.390926	-22.903337	-11.817674	0.849281	5.783876
24	10	Comoros	NaN	174	B	0	5	10	5	0.0	0.035871	0.072469	0.037853	0.000000	0.000000	...	14.064834	-12.968989
25	11	Djibouti	NaN	262	B R	54508	27310	23257	10456	15104.0	27.37186	23.139682	11.353987	14.869801	18.273091	-9.763426	-3.35933	-14.239601	5.395268	4.121943
...
275	261	Samoa	NaN	882	B	0	0	0	0	0.0	0	0	0	0.000000	0.000000
276	262	Tokelau	NaN	772	B	0	0	0	0	0.0	0	0	0	0.000000	0.000000
277	263	Tonga	NaN	776	B	0	0	0	0	0.0	0	0	0	0.000000	0.000000
278	264	Tuvalu	NaN	798	C	0	0	0	0	0.0	0	0	0	0.000000	0.000000
279	265	Wallis and Futuna Islands	NaN	876	B	0	0	0	0	0.0	0	0	0	0.000000	0.000000

259 rows x 22 columns

(Figure 5.1: The 3 types of data)

Step 2: For each set of data in Table 6, using the `melt_func`, melt the data stored in multiple columns under years or periods into a single column, and then store the column in a separate data frame. This step splits Table 6 into 3 data frames in total (figure 5.2).

order		area	notes	country_code	type	period	Annual rate of change of the refugee stock
0	7	Africa	NaN	903	NaN	1990-1995	0.076037
1	8	Eastern Africa	NaN	910	NaN	1990-1995	-5.30801
2	9	Burundi	NaN	108	B R	1990-1995	-3.390926
3	10	Comoros	NaN	174	B	1990-1995	NaN
4	11	Djibouti	NaN	262	B R	1990-1995	-9.763426
...
1290	261	Samoa	NaN	882	B	2010-2015	NaN
1291	262	Tokelau	NaN	772	B	2010-2015	NaN
1292	263	Tonga	NaN	776	B	2010-2015	NaN
1293	264	Tuvalu	NaN	798	C	2010-2015	NaN
1294	265	Wallis and Futuna Islands	NaN	876	B	2010-2015	NaN

1295 rows x 7 columns

(Figure 5.2: Example of one type of data separated in one data frame)

Step 3: For each new data frame, separate major areas and regions for countries and areas under the “area” column and store them in separate data frames using the `extract_country`, `extract_region`, and `extract_majorarea` functions. This step splits the 3 data frames from Table 6 into 9 data frames in total. For each of the 9 data frames, column headers are then renamed, and empty values are replaced with “NaN” for better readability (figure 5.3).

Order	Country and Area	Notes	Country Code	Type	Year	Estimated refugee stock at mid-year (both sexes)	
2	9	Burundi	NaN	108	B R	1990	267929
3	10	Comoros	NaN	174	B	1990	0
4	11	Djibouti	NaN	262	B R	1990	54508
5	12	Eritrea	NaN	232	I	1990	0
6	13	Ethiopia	NaN	231	B R	1990	741965
...
1549	261	Samoa	NaN	882	B	2015	0.0
1550	262	Tokelau	NaN	772	B	2015	0.0
1551	263	Tonga	NaN	776	B	2015	0.0
1552	264	Tuvalu	NaN	798	C	2015	0.0
1553	265	Wallis and Futuna Islands	NaN	876	B	2015	0.0

(Figure 5.3: One of the 9 separated data frames for countries and areas)

Final Result

The original raw data has 6 tables in the excel sheet, after the cleaning, the notebook generates 24 cleaned data frames as outputs, and their names are as follows:

No.	Origin	Data Frame Name	Topic
1	Table 1	df_tb1_ca	International migrant stock at mid-year for countries and areas for countries and areas
2	Table 1	df_tb1_ma	International migrant stock at mid-year for countries and areas for major areas
3	Table 1	df_tb1_rg	International migrant stock at mid-year for countries and areas for regions
4	Table 2	df_tb2_ca	Total population of both sexes at mid-year (thousands) for countries and areas
5	Table 2	df_tb2_ma	Total population of both sexes at mid-year (thousands) for major areas
6	Table 2	df_tb2_rg	Total population of both sexes at mid-year (thousands) for regions
7	Table 3	df_tb3_ca	International migrant stock as a percentage of the total population for countries and areas
8	Table 3	df_tb3_ma	International migrant stock as a percentage of the total population for major areas
9	Table 3	df_tb3_rg	International migrant stock as a percentage of the total population for regions
10	Table 4	df_tb4_ca	Female migrants as a percentage of the international migrant stock for countries and areas
11	Table 4	df_tb4_ma	Female migrants as a percentage of the international migrant stock for major areas
12	Table 4	df_tb4_rg	Female migrants as a percentage of the international migrant stock for regions
13	Table 5	df_tb5_ca	Annual rate of change of the migrant stock for countries and areas
14	Table 5	df_tb5_ma	Annual rate of change of the migrant stock for major areas
15	Table 5	df_tb5_rg	Annual rate of change of the migrant stock for regions
16	Table 6	df_tb6_a_ca	Estimated refugee stock at mid-year (both sexes) for countries and areas
17	Table 6	df_tb6_a_ma	Estimated refugee stock at mid-year (both sexes) for major areas
18	Table 6	df_tb6_a_rg	Estimated refugee stock at mid-year (both sexes) for regions
19	Table 6	df_tb6_b_ca	Refugees as a percentage of the international migrant stock for countries and areas
20	Table 6	df_tb6_b_ma	Refugees as a percentage of the international migrant stock for major areas

21	Table 6	df_tb6_b_rg	Refugees as a percentage of the international migrant stock for regions
22	Table 6	df_tb6_c_ca	Annual rate of change of the refugee stock for countries and areas
23	Table 6	df_tb6_c_ma	Annual rate of change of the refugee stock for major areas
24	Table 6	df_tb6_c_rg	Annual rate of change of the refugee stock for regions

Screenshots of Cleaned Data Frames

df_tb1_ca

	Order	Country and Area	Notes	Country Code	Type	Year	Gender	International migrant stock at mid-year	
	2	9	Burundi	NaN	108	B R	1990	male	163267
	3	10	Comoros	NaN	174	B	1990	male	6717
	4	11	Djibouti	NaN	262	B R	1990	male	64242
	5	12	Eritrea	NaN	232	I	1990	male	6228
	6	13	Ethiopia	NaN	231	B R	1990	male	607284

3103	261	Samoa	NaN	882	B	2015	female	2460.0	
3104	262	Tokelau	NaN	772	B	2015	female	254.0	
3105	263	Tonga	NaN	776	B	2015	female	2604.0	
3106	264	Tuvalu	NaN	798	C	2015	female	63.0	
3107	265	Wallis and Futuna Islands	NaN	876	B	2015	female	1411.0	

2784 rows × 8 columns

df_tb1_ma

	Order	Major Area	Notes	Country Code	Type	Year	Gender	International migrant stock at mid-year
0	7	Africa	NaN	903	NaN	1990	male	8279564
64	71	Asia	NaN	935	NaN	1990	male	26011875
120	127	Europe	NaN	908	NaN	1990	male	23946673
173	180	Latin America and the Caribbean	NaN	904	NaN	1990	male	3597037
225	232	Northern America	NaN	905	NaN	1990	male	13497319
...
2913	71	Asia	NaN	935	NaN	2015	female	31530709.0
2969	127	Europe	NaN	908	NaN	2015	female	39873338.0
3022	180	Latin America and the Caribbean	NaN	904	NaN	2015	female	4650938.0
3074	232	Northern America	NaN	905	NaN	2015	female	27902348.0
3080	238	Oceania	NaN	909	NaN	2015	female	4101334.0

72 rows × 8 columns

df_tb1_rg

	Order	Region	Notes	Country Code	Type	Year	Gender	International migrant stock at mid-year
1	8	Eastern Africa	NaN	910	NaN	1990	male	3071189
22	29	Middle Africa	NaN	911	NaN	1990	male	744494
32	39	Northern Africa	NaN	912	NaN	1990	male	1230643
40	47	Southern Africa	NaN	913	NaN	1990	male	840899
46	53	Western Africa	NaN	914	NaN	1990	male	2392339
...
3074	232	Northern America	NaN	905	NaN	2015	female	27902348.0
3081	239	Australia and New Zealand	NaN	927	NaN	2015	female	3963032.0
3084	242	Melanesia	NaN	928	NaN	2015	female	47782.0
3090	248	Micronesia	NaN	954	NaN	2015	female	57159.0
3098	256	Polynesia	NaN	957	NaN	2015	female	33361.0

264 rows × 8 columns

df_tb2_ca

	Order	Country and Area	Notes	Country Code	Year	Gender	Total population at mid-year (thousands)
2	9	Burundi	NaN	108	1990	male	2755.028
3	10	Comoros	NaN	174	1990	male	208.212
4	11	Djibouti	NaN	262	1990	male	295.933
5	12	Eritrea	NaN	232	1990	male	1558.486
6	13	Ethiopia	NaN	231	1990	male	23965.647
...
3103	261	Samoa	NaN	882	2015	female	93.584
3104	262	Tokelau	NaN	772	2015	female	NaN
3105	263	Tonga	NaN	776	2015	female	52.931
3106	264	Tuvalu	NaN	798	2015	female	NaN
3107	265	Wallis and Futuna Islands	NaN	876	2015	female	NaN

2784 rows × 7 columns

df_tb2_ma

	Order	Major Area	Notes	Country Code	Year	Gender	Total population at mid-year (thousands)
0	7	Africa	NaN	903	1990	male	315071.378
64	71	Asia	NaN	935	1990	male	1634734.677
120	127	Europe	NaN	908	1990	male	347356.281
173	180	Latin America and the Caribbean	NaN	904	1990	male	221989.776
225	232	Northern America	NaN	905	1990	male	137757.875
...
2913	71	Asia	NaN	935	2015	female	2146310.075
2969	127	Europe	NaN	908	2015	female	382275.191
3022	180	Latin America and the Caribbean	NaN	904	2015	female	320877.844
3074	232	Northern America	NaN	905	2015	female	180455.403
3080	238	Oceania	NaN	909	2015	female	19624.181

72 rows × 7 columns

df_tb2_rg

	Order	Region	Notes	Country Code	Year	Gender	Total population at mid-year (thousands)
1	8	Eastern Africa	NaN	910	1990	male	98208.646
22	29	Middle Africa	NaN	911	1990	male	35035.128
32	39	Northern Africa	NaN	912	1990	male	70480.841
40	47	Southern Africa	NaN	913	1990	male	20760.4
46	53	Western Africa	NaN	914	1990	male	90586.363
...
3074	232	Northern America	NaN	905	2015	female	180455.403
3081	239	Australia and New Zealand	NaN	927	2015	female	14308.441
3084	242	Melanesia	NaN	928	2015	female	4719.309
3090	248	Micronesia	NaN	954	2015	female	260.316
3098	256	Polynesia	NaN	957	2015	female	336.115

264 rows × 7 columns

df_tb3_ca

	Order	Country and Area	Notes	Country Code	Type	Year	Gender	International migrant stock as a percentage of the total population	
	2	9	Burundi	NaN	108	B R	1990	male	5.926147
	3	10	Comoros	NaN	174	B	1990	male	3.226039
	4	11	Djibouti	NaN	262	B R	1990	male	21.708292
	5	12	Eritrea	NaN	232	I	1990	male	0.399619
	6	13	Ethiopia	NaN	231	B R	1990	male	2.533977
...
3103	261	Samoa	NaN	882	B	2015	female		2.628654
3104	262	Tokelau	NaN	772	B	2015	female		NaN
3105	263	Tonga	NaN	776	B	2015	female		4.919612
3106	264	Tuvalu	NaN	798	C	2015	female		NaN
3107	265	Wallis and Futuna Islands	NaN	876	B	2015	female		NaN

2784 rows × 8 columns

df_tb3_ma

	Order		Major Area	Notes	Country Code	Type	Year	Gender	International migrant stock as a percentage of the total population
	0	7	Africa	NaN	903	NaN	1990	male	2.627838
	64	71	Asia	NaN	935	NaN	1990	male	1.591199
	120	127	Europe	NaN	908	NaN	1990	male	6.89398
	173	180	Latin America and the Caribbean	NaN	904	NaN	1990	male	1.620362
	225	232	Northern America	NaN	905	NaN	1990	male	9.797857
...
	2913	71	Asia	NaN	935	NaN	2015	female	1.469066
	2969	127	Europe	NaN	908	NaN	2015	female	10.430532
	3022	180	Latin America and the Caribbean	NaN	904	NaN	2015	female	1.449442
	3074	232	Northern America	NaN	905	NaN	2015	female	15.462185
	3080	238	Oceania	NaN	909	NaN	2015	female	20.899389

72 rows × 8 columns

df_tb3_rg

Order		Region	Notes	Country Code	Type	Year	Gender	International migrant stock as a percentage of the total population
1	8	Eastern Africa	NaN	910	NaN	1990	male	3.127208
22	29	Middle Africa	NaN	911	NaN	1990	male	2.124993
32	39	Northern Africa	NaN	912	NaN	1990	male	1.746067
40	47	Southern Africa	NaN	913	NaN	1990	male	4.050495
46	53	Western Africa	NaN	914	NaN	1990	male	2.640948
...
3074	232	Northern America	NaN	905	NaN	2015	female	15.462185
3081	239	Australia and New Zealand	NaN	927	NaN	2015	female	27.697161
3084	242	Melanesia	NaN	928	NaN	2015	female	1.012479
3090	248	Micronesia	NaN	954	NaN	2015	female	21.957544
3098	256	Polynesia	NaN	957	NaN	2015	female	9.925472

264 rows × 8 columns

df_tb4_ca

	Order	Country and Area	Notes	Country Code	Type	Year	Gender	Female migrants as a percentage of the international migrant stock
2	9	Burundi	NaN	108	B R	1990	female	50.987061
3	10	Comoros	NaN	174	B	1990	female	52.290646
4	11	Djibouti	NaN	262	B R	1990	female	47.437838
5	12	Eritrea	NaN	232	I	1990	female	47.434166
6	13	Ethiopia	NaN	231	B R	1990	female	47.439047
...
1549	261	Samoa	NaN	882	B	2015	female	49.908704
1550	262	Tokelau	NaN	772	B	2015	female	52.156057
1551	263	Tonga	NaN	776	B	2015	female	45.437096
1552	264	Tuvalu	NaN	798	C	2015	female	44.680851
1553	265	Wallis and Futuna Islands	NaN	876	B	2015	female	49.52615

1392 rows × 8 columns

df_tb4_ma

	Order	Major Area	Notes	Country Code	Type	Year	Gender	Female migrants as a percentage of the international migrant stock
0	7	Africa	NaN	903	NaN	1990	female	47.232408
64	71	Asia	NaN	935	NaN	1990	female	45.96873
120	127	Europe	NaN	908	NaN	1990	female	51.346887
173	180	Latin America and the Caribbean	NaN	904	NaN	1990	female	49.830217
225	232	Northern America	NaN	905	NaN	1990	female	51.115342
231	238	Oceania	NaN	909	NaN	1990	female	48.95129
259	7	Africa	NaN	903	NaN	1995	female	47.306127
323	71	Asia	NaN	935	NaN	1995	female	45.867921
379	127	Europe	NaN	908	NaN	1995	female	51.486537
432	180	Latin America and the Caribbean	NaN	904	NaN	1995	female	49.963418
484	232	Northern America	NaN	905	NaN	1995	female	50.805529

df_tb4_rg

	Order	Region	Notes	Country Code	Type	Year	Gender	Female migrants as a percentage of the international migrant stock
1	8	Eastern Africa	NaN	910	NaN	1990	female	48.504812
22	29	Middle Africa	NaN	911	NaN	1990	female	49.025765
32	39	Northern Africa	NaN	912	NaN	1990	female	48.791486
40	47	Southern Africa	NaN	913	NaN	1990	female	39.606165
46	53	Western Africa	NaN	914	NaN	1990	female	46.486134
...
1520	232	Northern America	NaN	905	NaN	2015	female	51.207563
1527	239	Australia and New Zealand	NaN	927	NaN	2015	female	50.785972
1530	242	Melanesia	NaN	928	NaN	2015	female	43.598704
1536	248	Micronesia	NaN	954	NaN	2015	female	49.372042
1544	256	Polynesia	NaN	957	NaN	2015	female	46.257626

132 rows × 8 columns

df_tb5_ca

	Order	Country and Area	Notes	Country Code	Type	Year	Gender	Annual rate of change of the migrant stock
2	9	Burundi	NaN	108	B R	1990-1995	male	-5.475511
3	10	Comoros	NaN	174	B	1990-1995	male	-0.30906
4	11	Djibouti	NaN	262	B R	1990-1995	male	-4.046026
5	12	Eritrea	NaN	232	I	1990-1995	male	0.983754
6	13	Ethiopia	NaN	231	B R	1990-1995	male	-7.179744
...
2585	261	Samoa	NaN	882	B	2010-2015	female	-0.545343
2586	262	Tokelau	NaN	772	B	2010-2015	female	2.60325
2587	263	Tonga	NaN	776	B	2010-2015	female	2.526318
2588	264	Tuvalu	NaN	798	C	2010-2015	female	-1.819436
2589	265	Wallis and Futuna Islands	NaN	876	B	2010-2015	female	0.516899

2320 rows × 8 columns

df_tb5_ma

	Order	Major Area	Notes	Country Code	Type	Year	Gender	Annual rate of change of the migrant stock
0	7	Africa	NaN	903	NaN	1990-1995	male	0.798774
64	71	Asia	NaN	935	NaN	1990-1995	male	-0.63615
120	127	Europe	NaN	908	NaN	1990-1995	male	1.363213
173	180	Latin America and the Caribbean	NaN	904	NaN	1990-1995	male	-1.424381
225	232	Northern America	NaN	905	NaN	1990-1995	male	3.898245
231	238	Oceania	NaN	909	NaN	1990-1995	male	0.973073
259	7	Africa	NaN	903	NaN	1995-2000	male	-1.848117
323	71	Asia	NaN	935	NaN	1995-2000	male	1.269413
379	127	Europe	NaN	908	NaN	1995-2000	male	1.211037
432	180	Latin America and the Caribbean	NaN	904	NaN	1995-2000	male	-0.398563

df_tb5_rg

	Order	Region	Notes	Country Code	Type	Year	Gender	Annual rate of change of the migrant stock
1	8	Eastern Africa	NaN	910	NaN	1990-1995	male	-3.446375
22	29	Middle Africa	NaN	911	NaN	1990-1995	male	11.845106
32	39	Northern Africa	NaN	912	NaN	1990-1995	male	-2.13921
40	47	Southern Africa	NaN	913	NaN	1990-1995	male	-3.266338
46	53	Western Africa	NaN	914	NaN	1990-1995	male	3.611415
...
2556	232	Northern America	NaN	905	NaN	2010-2015	female	1.29213
2563	239	Australia and New Zealand	NaN	927	NaN	2010-2015	female	2.776495
2566	242	Melanesia	NaN	928	NaN	2010-2015	female	0.648292
2572	248	Micronesia	NaN	954	NaN	2010-2015	female	-0.191872
2580	256	Polynesia	NaN	957	NaN	2010-2015	female	-0.186769

220 rows × 8 columns

df_tb6_a_ca

	Order	Country and Area	Notes	Country Code	Type	Year	Estimated refugee stock at mid-year (both sexes)
2	9	Burundi	NaN	108	B R	1990	267929
3	10	Comoros	NaN	174	B	1990	0
4	11	Djibouti	NaN	262	B R	1990	54508
5	12	Eritrea	NaN	232	I	1990	0
6	13	Ethiopia	NaN	231	B R	1990	741965
...
1549	261	Samoa	NaN	882	B	2015	0.0
1550	262	Tokelau	NaN	772	B	2015	0.0
1551	263	Tonga	NaN	776	B	2015	0.0
1552	264	Tuvalu	NaN	798	C	2015	0.0
1553	265	Wallis and Futuna Islands	NaN	876	B	2015	0.0

1392 rows × 7 columns

df_tb6_a_ma

	Order	Major Area	Notes	Country Code	Type	Year	Estimated refugee stock at mid-year (both sexes)
0	7	Africa	NaN	903	NaN	1990	5687352
64	71	Asia	NaN	935	NaN	1990	9937007
120	127	Europe	NaN	908	NaN	1990	1321884
173	180	Latin America and the Caribbean	NaN	904	NaN	1990	1197198
225	232	Northern America	NaN	905	NaN	1990	583450
231	238	Oceania	NaN	909	NaN	1990	109680
259	7	Africa	NaN	903	NaN	1995	5949953
323	71	Asia	NaN	935	NaN	1995	7937682
379	127	Europe	NaN	908	NaN	1995	2746090
432	180	Latin America and the Caribbean	NaN	904	NaN	1995	352256
484	232	Northern America	NaN	905	NaN	1995	775419
490	238	Oceania	NaN	909	NaN	1995	92440

df_tb6_a_rg

	Order	Region	Notes	Country Code	Type	Year	Estimated refugee stock at mid-year (both sexes)
1	8	Eastern Africa	NaN	910	NaN	1990	3168001
22	29	Middle Africa	NaN	911	NaN	1990	446609
32	39	Northern Africa	NaN	912	NaN	1990	1202360
40	47	Southern Africa	NaN	913	NaN	1990	135525
46	53	Western Africa	NaN	914	NaN	1990	734857
...
1520	232	Northern America	NaN	905	NaN	2015	423941.0
1527	239	Australia and New Zealand	NaN	927	NaN	2015	49408.0
1530	242	Melanesia	NaN	928	NaN	2015	4812.0
1536	248	Micronesia	NaN	954	NaN	2015	390.0
1544	256	Polynesia	NaN	957	NaN	2015	0.0

132 rows × 7 columns

df_tb6_b_ca

	Order	Country and Area	Notes	Country Code	Type	Year	Refugees as a percentage of the international migrant stock
	2	9	Burundi	NaN	108	B R 1990	80.43259
	3	10	Comoros	NaN	174	B 1990	0
	4	11	Djibouti	NaN	262	B R 1990	44.597901
	5	12	Eritrea	NaN	232	I 1990	0
	6	13	Ethiopia	NaN	231	B R 1990	64.21771
...
1549	261	Samoa	NaN	882	B	2015	0.0
1550	262	Tokelau	NaN	772	B	2015	0.0
1551	263	Tonga	NaN	776	B	2015	0.0
1552	264	Tuvalu	NaN	798	C	2015	0.0
1553	265	Wallis and Futuna Islands	NaN	876	B	2015	0.0

1392 rows × 7 columns

df_tb6_b_ma

	Order	Major Area	Notes	Country Code	Type	Year	Refugees as a percentage of the international migrant stock
0	7	Africa	NaN	903	NaN	1990	36.246821
64	71	Asia	NaN	935	NaN	1990	20.640923
120	127	Europe	NaN	908	NaN	1990	2.685708
173	180	Latin America and the Caribbean	NaN	904	NaN	1990	16.697956
225	232	Northern America	NaN	905	NaN	1990	2.113142
231	238	Oceania	NaN	909	NaN	1990	2.318396
259	7	Africa	NaN	903	NaN	1995	36.384888
323	71	Asia	NaN	935	NaN	1995	17.052599
379	127	Europe	NaN	908	NaN	1995	5.196729
432	180	Latin America and the Caribbean	NaN	904	NaN	1995	5.261762
484	232	Northern America	NaN	905	NaN	1995	2.325712

df_tb6_b_rg

	Order	Region	Notes	Country Code	Type	Year	Refugees as a percentage of the international migrant stock
1	8	Eastern Africa	NaN	910	NaN	1990	53.118453
22	29	Middle Africa	NaN	911	NaN	1990	30.578557
32	39	Northern Africa	NaN	912	NaN	1990	50.031625
40	47	Southern Africa	NaN	913	NaN	1990	9.733481
46	53	Western Africa	NaN	914	NaN	1990	16.437904
...
1520	232	Northern America	NaN	905	NaN	2015	0.778034
1527	239	Australia and New Zealand	NaN	927	NaN	2015	0.63316
1530	242	Melanesia	NaN	928	NaN	2015	4.390711
1536	248	Micronesia	NaN	954	NaN	2015	0.336869
1544	256	Polynesia	NaN	957	NaN	2015	0.0

132 rows × 7 columns

df_tb6_c_ca

	Order	Country and Area	Notes	Country Code	Type	Period	Annual rate of change of the refugee stock
2	9	Burundi	NaN	108	B R	1990-1995	-3.390926
3	10	Comoros	NaN	174	B	1990-1995	NaN
4	11	Djibouti	NaN	262	B R	1990-1995	-9.763426
5	12	Eritrea	NaN	232	I	1990-1995	NaN
6	13	Ethiopia	NaN	231	B R	1990-1995	-5.505717
...
1290	261	Samoa	NaN	882	B	2010-2015	NaN
1291	262	Tokelau	NaN	772	B	2010-2015	NaN
1292	263	Tonga	NaN	776	B	2010-2015	NaN
1293	264	Tuvalu	NaN	798	C	2010-2015	NaN
1294	265	Wallis and Futuna Islands	NaN	876	B	2010-2015	NaN

1160 rows × 7 columns

df_tb6_c_ma

	Order	Major Area	Notes	Country Code	Type	Period	Annual rate of change of the refugee stock
0	7	Africa	NaN	903	NaN	1990-1995	0.076037
64	71	Asia	NaN	935	NaN	1990-1995	-3.819461
120	127	Europe	NaN	908	NaN	1990-1995	13.2017
173	180	Latin America and the Caribbean	NaN	904	NaN	1990-1995	-23.096408
225	232	Northern America	NaN	905	NaN	1990-1995	1.917003
231	238	Oceania	NaN	909	NaN	1990-1995	-4.615626
259	7	Africa	NaN	903	NaN	1995-2000	-8.003244
323	71	Asia	NaN	935	NaN	1995-2000	1.732425
379	127	Europe	NaN	908	NaN	1995-2000	-4.942878
432	180	Latin America and the Caribbean	NaN	904	NaN	1995-2000	-41.213183
484	232	Northern America	NaN	905	NaN	1995-2000	-7.863684

df_tb6_c_rg

	Order	Region	Notes	Country Code	Type	Period	Annual rate of change of the refugee stock
1	8	Eastern Africa	NaN	910	NaN	1990-1995	-5.30801
22	29	Middle Africa	NaN	911	NaN	1990-1995	12.964162
32	39	Northern Africa	NaN	912	NaN	1990-1995	-3.456178
40	47	Southern Africa	NaN	913	NaN	1990-1995	-1.954547
46	53	Western Africa	NaN	914	NaN	1990-1995	8.717581
...
1261	232	Northern America	NaN	905	NaN	2010-2015	-1.526424
1268	239	Australia and New Zealand	NaN	927	NaN	2010-2015	8.829439
1271	242	Melanesia	NaN	928	NaN	2010-2015	-0.268521
1277	248	Micronesia	NaN	954	NaN	2010-2015	NaN
1285	256	Polynesia	NaN	957	NaN	2010-2015	NaN

110 rows × 7 columns