The UN dataset has a content table and six tables with different contents. It also ends with a ANNEX table and a NOTES table. According to the tidy data principle, I chose to clean six main tables. The cleaning processes of six main tables are pretty similar.

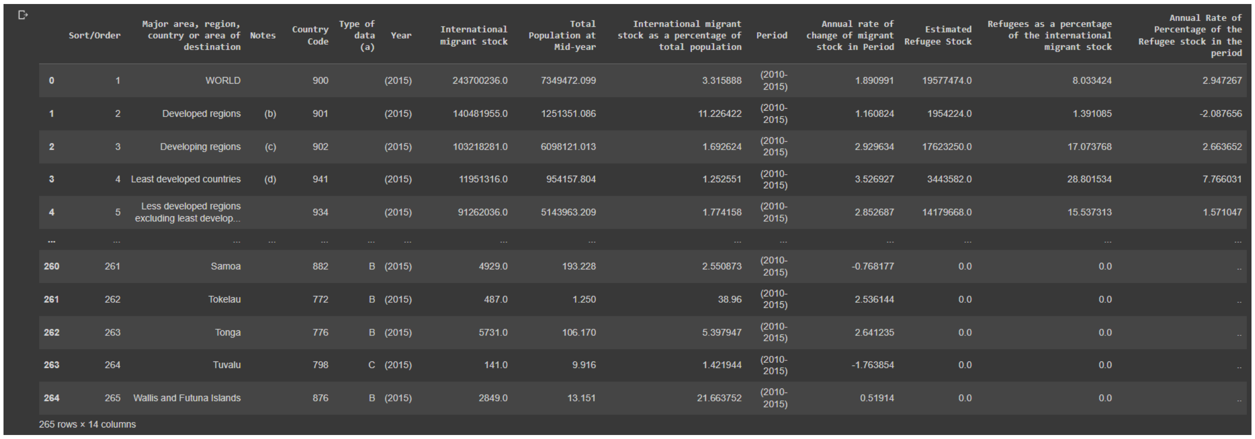
For the first table, the first step I chose is to drop the head part which is the green title area in the excel sheet so that we start from our first observation, the WORLD. After this step, I got a renaming process that renamed the column name in the Python dataframe since the original column name is stored in the row of the dataframe instead of the head. After the dropping and renaming process, we can deal with the data part. Here I first replace all NaN values with the empty strings to make it look better. I didn't drop these null values since they are also an important part of our dataset. And then, I used the melt function to separate the sheet to three different dataset that is male, female, and both sexes. After this step, I got two new columns named Year and International migrant stock. I did a little revision on the Year column since I got sex and year in this column at the same time, so I drop the sex part and only keep year number in this column. After these steps, we got 1590 observations in the table, but we actually only have 265 areas and countries which means each area and country name has repeated 6 times in one dataframes. The reason for this is that we have 6 different years from 1990 to 2015. In order to make dataframes much more tidy, I decided to separate these data frames to 18 different data frames by using the index number. Also I reset the index after the separating process.

After doing the same process for sheet 2, I got another 18 data frames of total population. Then I merge these total population datasets into those datasets about international migrant stock from sheet 1 so that I got 18 new different data frames:

1) UN_1990_both_sexes, 2) UN_1990_male, 3) UN_1990_female
4) UN_1995_both_sexes, 5) UN_1005_male, 6) UN_1995_female
7) UN_2000_both_sexes, 8) UN_2000_male, 9) UN_2005_female
10) UN_2005_both_sexes, 11) UN_2005_male, 12) UN_2005_female
13) UN_2010_both_sexes, 14) UN_2010_male, 15) UN_2010_female
14) UN_2015_both_sexes, 15) UN_2015_male, 16) UN_2015_female

After I get these data frames, I only need to merge new data into these data frames after I cleaned each sheet. As the result, I got 18 similar datasets as following:

This is all data of both sexes at 2015



	Sort/Order	Major area, region, country or area of destination	Notes	Country Code	Type of data (a)	Year	International migrant stock	Total Population at Mid-year	International migrant stock as a percentage of total population	Period	Annual rate of change of migrant stock in Period	Estimated Refugee Stock	Refugees as a percentage of the international migrant stock	Annual Rate of Percentage of the Refugee stock in the period
0	1	WORLD		900		(2015)	243700236.0	7349472.099	3.315888	(2010-2015)	1.890991	19577474.0	8.033424	2.947267
1	2	Developed regions	(b)	901		(2015)	140481955.0	1251351.086	11.226422	(2010-2015)	1.160824	1954224.0	1.391085	-2.087656
2	3	Developing regions	(c)	902		(2015)	103218281.0	6098121.013	1.692624	(2010-2015)	2.929634	17623250.0	17.073768	2.663652
3	4	Least developed countries	(d)	941		(2015)	11951316.0	954157.804	1.252551	(2010-2015)	3.526927	3443582.0	28.801534	7.766031
4	5	Less developed regions excluding least develop...		934		(2015)	91262036.0	5143963.209	1.774158	(2010-2015)	2.852687	14179668.0	15.537313	1.571047
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
260	261	Samoa		882	B	(2015)	4929.0	193.228	2.550873	(2010-2015)	-0.768177	0.0	0.0	..
261	262	Tokelau		772	B	(2015)	487.0	1.250	38.96	(2010-2015)	2.536144	0.0	0.0	..
262	263	Tonga		776	B	(2015)	5731.0	106.170	5.397947	(2010-2015)	2.641235	0.0	0.0	..
263	264	Tuvalu		798	C	(2015)	141.0	9.916	1.421944	(2010-2015)	-1.763854	0.0	0.0	..
264	265	Wallis and Futuna Islands		876	B	(2015)	2849.0	13.151	21.663752	(2010-2015)	0.51914	0.0	0.0	..

265 rows × 14 columns