Cleaning data is the process of detecting and correcting or removing corrupt or in accurate records from a record database.

The main steps are as follows:
1. First read excel spreadsheet,
2. Read different sheet_name using different sheet_name
3. Since the data read is not standard, slices are made according to the row where the normal value is located, and features are obtained again, and then associated with the acquisition of extraction
4. Perform melt operation according to the associated data to show the pre-processed data

| | Country code | International migrant stock at mid-year (both sexes) 1990 | International migrant stock at mid-year (both sexes) 1995 | International migrant stock at mid-year (both sexes) 2000 | International migrant stock at mid-year (both sexes) 2005 | International migrant stock at mid-year (both sexes) 2010.0 | International migrant stock at mid-year (both sexes) 2015.0 | International migrant stock at mid-year (male) 1990 | International migrant stock at mid-year (male) 1995 | International migrant stock at mid-year (male) 2000 | International migrant stock at mid-year (male) 2005 | Interna m st mi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 900 | 152563212 | 160801752 | 172703309 | 191269100 | 221714243.0 | 243700236.0 | 77747510 | 81737477 | 87884839 | 97866674 | 11461 |
| 17 | 901 | 82378628 | 92306854 | 103375363 | 117181109 | 132560325.0 | 140481955.0 | 40263397 | 45092799 | 50536796 | 57217777 | 6408 |
| 18 | 902 | 70184584 | 68494898 | 69327946 | 74087991 | 89153918.0 | 103218281.0 | 37484113 | 36644678 | 37348043 | 40648897 | 5053 |
| 19 | 941 | 11075966 | 11711703 | 10077824 | 9809634 | 10018128.0 | 11951316.0 | 5843107 | 6142712 | 5361902 | 5383009 | 546 |
| 20 | 934 | 59105261 | 56778501 | 59244124 | 64272611 | 79130668.0 | 91262036.0 | 31641006 | 30501966 | 31986141 | 35265888 | 4506 |

| ıational migrant stock at ıid-year le) 1990 | International migrant stock at mid-year (male) 1995 | International migrant stock at mid-year (male) 2000 | International migrant stock at mid-year (male) 2005 | International migrant stock at mid-year (male) 2010.0 | International migrant stock at mid-year (male) 2015.0 | International migrant stock at mid-year (female) 1990 | International migrant stock at mid-year (female) 1995 | International migrant stock at mid-year (female) 2000 | International migrant stock at mid-year (female) 2005 | International migrant stock at mid-year (female) 2010.0 | International migrant stock at mid-year (female) 2015.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 7747510 | 81737477 | 87884839 | 97866674 | 114613714.0 | 126115435.0 | 74815702 | 79064275 | 84818470 | 93402426 | 107100529.0 | 117584801.0 |
| )263397 | 45092799 | 50536796 | 57217777 | 64081077.0 | 67618619.0 | 42115231 | 47214055 | 52838567 | 59963332 | 68479248.0 | 72863336.0 |
| 7484113 | 36644678 | 37348043 | 40648897 | 50532637.0 | 58496816.0 | 32700471 | 31850220 | 31979903 | 33439094 | 38621281.0 | 44721465.0 |
| 5843107 | 6142712 | 5361902 | 5383009 | 5462714.0 | 6463217.0 | 5236216 | 5573685 | 4721920 | 4432371 | 4560536.0 | 5493028.0 |
| 1641006 | 30501966 | 31986141 | 35265888 | 45069923.0 | 52033599.0 | 27464255 | 26276535 | 27257983 | 29006723 | 34060745.0 | 39228437.0 |

The first table I ran came out after processing the data, adding feature names, removing fields and printing the data. The main purpose of this is to make table1 clearer and more concise to express the number of immigrants corresponding to each sort and each main area. This is because there are many branches (countries) in the table in each main area.

| | Country code | stock year | value |
|---|---|---|---|
| 0 | 900 | International migrant stock at mid-year (both ... | 152563212 |
| 1 | 901 | International migrant stock at mid-year (both ... | 82378628 |
| 2 | 902 | International migrant stock at mid-year (both ... | 70184584 |
| 3 | 941 | International migrant stock at mid-year (both ... | 11075966 |
| 4 | 934 | International migrant stock at mid-year (both ... | 59105261 |
| ... | ... | ... | ... |
| 4765 | 882 | International migrant stock at mid-year (femal... | 2460 |
| 4766 | 772 | International migrant stock at mid-year (femal... | 254 |
| 4767 | 776 | International migrant stock at mid-year (femal... | 2604 |
| 4768 | 798 | International migrant stock at mid-year (femal... | 63 |
| 4769 | 876 | International migrant stock at mid-year (femal... | 1411 |

The second table I ran is the immigration value of all countries (regions) in the

corresponding year are listed, the country is used TABLE1 in each country representative corresponding to the country code to represent.

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 3 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 4 | NaN | NaN | NaN | NaN | United Nations | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 276 | 261 | Samoa | NaN | 882 | 162.865 | 170.158 | 174.614 | 179.928 | 186.029 | 193.228 | 85.009 | 88.488 | 90.932 | 93.185 | 95.95 | 99.644 | 77.856 | 81.67 | 83.682 |
| 277 | 262 | Tokelau | NaN | 772 | 1.609 | 1.520 | 1.552 | 1.210 | 1.135 | 1.250 | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| 278 | 263 | Tonga | NaN | 776 | 95.152 | 95.889 | 97.898 | 100.858 | 103.947 | 106.170 | 48.247 | 48.614 | 49.788 | 50.574 | 52.055 | 53.239 | 46.905 | 47.275 | 48.11 |
| 279 | 264 | Tuvalu | NaN | 798 | 9.004 | 9.227 | 9.419 | 9.694 | 9.827 | 9.916 | .. | .. | .. | .. | .. | .. | .. | .. | .. |
| 280 | 265 | Wallis and Futuna Islands | NaN | 876 | 13.88 | 14.143 | 14.497 | 14.246 | 13.565 | 13.151 | .. | .. | .. | .. | .. | .. | .. | .. | .. |

This table is about TABLE2, the NaN stand for “not a number”, it might be cut off. There are two ways to deal with “NaN”, one way is just remove it. Another way is creating a new value for these ones where there might be a missing value.

| | Country code | pupulation year | value |
|---|---|---|---|
| 0 | 900 | Total population of both sexes at mid-year (th... | 5.30967e+06 |
| 1 | 901 | Total population of both sexes at mid-year (th... | 1.14446e+06 |
| 2 | 902 | Total population of both sexes at mid-year (th... | 4.1652e+06 |
| 3 | 941 | Total population of both sexes at mid-year (th... | 510058 |
| 4 | 934 | Total population of both sexes at mid-year (th... | 3.65515e+06 |
| ... | ... | ... | ... |
| 4765 | 882 | Total female population at mid-year (thousands... | 93.584 |
| 4766 | 772 | Total female population at mid-year (thousands... | .. |
| 4767 | 776 | Total female population at mid-year (thousands... | 52.931 |
| 4768 | 798 | Total female population at mid-year (thousands... | .. |
| 4769 | 876 | Total female population at mid-year (thousands... | .. |

And this table is about total value of total population, female population and male population during these mid-year and by region. All of regions have displayed in this table.

| | Country code | stock percentage year | value |
|---|---|---|---|
| 0 | 900 | International migrant stock as a percentage of... | 2.87331 |
| 1 | 901 | International migrant stock as a percentage of... | 7.19802 |
| 2 | 902 | International migrant stock as a percentage of... | 1.68502 |
| 3 | 941 | International migrant stock as a percentage of... | 2.17151 |
| 4 | 934 | International migrant stock as a percentage of... | 1.61704 |
| ... | ... | ... | ... |
| 4765 | 882 | International migrant stock as a percentage of... | 2.62865 |
| 4766 | 772 | International migrant stock as a percentage of... | .. |
| 4767 | 776 | International migrant stock as a percentage of... | 4.91961 |
| 4768 | 798 | International migrant stock as a percentage of... | .. |
| 4769 | 876 | International migrant stock as a percentage of... | .. |

This table is cleaned for TABLE3, which is about the value of each country code’s

international migrant stock percentage year.

| | Country code | percentage stock year | value |
|---|---|---|---|
| **0** | 900 | Female migrants as a percentage of the interna... | 49.0391 |
| **1** | 901 | Female migrants as a percentage of the interna... | 51.124 |
| **2** | 902 | Female migrants as a percentage of the interna... | 46.5921 |
| **3** | 941 | Female migrants as a percentage of the interna... | 47.2612 |
| **4** | 934 | Female migrants as a percentage of the interna... | 46.4667 |
| **...** | ... | ... | ... |
| **1585** | 882 | Female migrants as a percentage of the interna... | 49.9087 |
| **1586** | 772 | Female migrants as a percentage of the interna... | 52.1561 |
| **1587** | 776 | Female migrants as a percentage of the interna... | 45.4371 |
| **1588** | 798 | Female migrants as a percentage of the interna... | 44.6809 |
| **1589** | 876 | Female migrants as a percentage of the interna... | 49.5261 |

The table is about TABLE4, which is about female migrants as a percentage of the international migrant stock of each area. And this table displays the value for each year.

| | Country code | migrant stock year | value |
|---|---|---|---|
| **0** | 900 | Annual rate of change of the migrant stock (bo... | 1.05186 |
| **1** | 901 | Annual rate of change of the migrant stock (bo... | 2.27585 |
| **2** | 902 | Annual rate of change of the migrant stock (bo... | -0.487389 |
| **3** | 941 | Annual rate of change of the migrant stock (bo... | 1.11817 |
| **4** | 934 | Annual rate of change of the migrant stock (bo... | -0.803244 |
| **...** | ... | ... | ... |
| **3970** | 882 | Annual rate of change of the migrant stock (fe... | -0.545343 |
| **3971** | 772 | Annual rate of change of the migrant stock (fe... | 2.60325 |
| **3972** | 776 | Annual rate of change of the migrant stock (fe... | 2.52632 |
| **3973** | 798 | Annual rate of change of the migrant stock (fe... | -1.81944 |
| **3974** | 876 | Annual rate of change of the migrant stock (fe... | 0.516899 |

For TABLE5, this table shows that annual rate of change of the migrant stock by sex and different areas.

| | Country code | international migrant stock year | value |
|---|---|---|---|
| **0** | 900 | Estimated refugee stock at mid-year (both sexe... | 18836571 |
| **1** | 901 | Estimated refugee stock at mid-year (both sexe... | 2014564 |
| **2** | 902 | Estimated refugee stock at mid-year (both sexe... | 16822007 |
| **3** | 941 | Estimated refugee stock at mid-year (both sexe... | 5048391 |
| **4** | 934 | Estimated refugee stock at mid-year (both sexe... | 11773616 |
| **...** | ... | ... | ... |
| **4500** | 882 | Annual rate of change of the refugee stock 201... | .. |
| **4501** | 772 | Annual rate of change of the refugee stock 201... | .. |
| **4502** | 776 | Annual rate of change of the refugee stock 201... | .. |
| **4503** | 798 | Annual rate of change of the refugee stock 201... | .. |
| **4504** | 876 | Annual rate of change of the refugee stock 201... | .. |

This tables displays about TABLE6, it shows estimate refugee stock at mid-year by sex and area. The variable is estimated refugee stock at each year during mid-year,

refugee as a percentage of the international migrant stock, and annual rate of change of the refugee stock.

The first three with higher correlation coefficients are distinguished according to the feature correlation coefficients as the features of the three models, and the target is prices, a regression model.