Table of contents

Cleaning of table 1	2
Cleaning of table 2	7
Cleaning of table 3	12
Cleaning of table 4	16
Cleaning of table 5	19
Cleaning of table 6	28
Cleaning of Annex	34

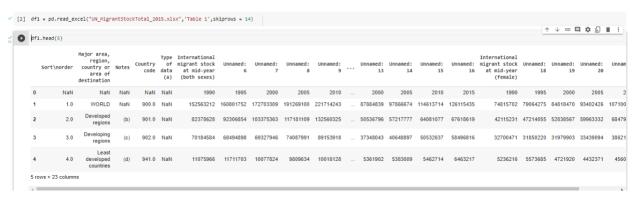
Cleaning of table 1

Step 1: Importing essential libraries

```
import numpy as np import pandas as pd
```

These libraries help to build the data frame that support interchange protocol. It has also been imported to get high performance data structure and data analysis tools.

Step 2: Importing and reading the datasheet as a dataframe



According to the tidy data principle, data cleaning requires importing of the dataset as a dataframe. In this regard, the 14 rows have been skipped as they contain headings of the dataset which are not included in the data. After importing the data, df1.head(5) function is used to read the data.

Step 3: Checking the information of the dataframe

```
// [4] df1.info()
             <class 'pandas.core.frame.DataFrame'>
             RangeIndex: 266 entries, 0 to 265
Data columns (total 23 columns):
                                                                                                            Non-Null Count Dtype
                   Column
                   r 265 non-r
Major area, region, country or area of destination
Notes
Country code
Type of data (a)
International migrant stock at mid-year (both sexes)
                                                                                                            266 non-null
                                                                                                            266 non-null
                    Unnamed:
                                                                                                            266 non-null
                                                                                                            266 non-null
                                                                                                                                     int64
                    International migrant stock at mid-year (male)
                                                                                                            266 non-null
                    Unnamed: 13
Unnamed: 14
Unnamed: 15
Unnamed: 16
Internationa
Unnamed: 18
Unnamed: 19
Unnamed: 20
                                      .
nal migrant stock at mid-year (female)
                                                                                                            266 non-null
             dtypes: float64(2), int64(6), object(15)
memory usage: 47.9+ KB
```

df1.info() function is used to check the information of the data type, non-null count and columns.

Step 4: Renaming the columns' name

```
[5] df1.rename(columns = {'International migrant stock at mid-year (both sexes)':'International migrant stock at mid-year (both sexes)(1990)'}, inplace = True)
df1.rename(columns = {'Unnamed: 6':'International migrant stock at mid-year (both sexes)(1995)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 7':'International migrant stock at mid-year (both sexes)(2000)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 8':'International migrant stock at mid-year (both sexes)(2005)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 9':'International migrant stock at mid-year (both sexes)(2010)'}, inplace = True
        df1.rename(columns = {'Unnamed: 10':'International migrant stock at mid-year (both sexes)(2015)'}, inplace = True)
        df1.rename(columns = {'International migrant stock at mid-year (male)':'International migrant stock at mid-year (male)(1990)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 12':'International migrant stock at mid-year (male)(1995)'}, inplace = True
        df1.rename(columns = {'Unnamed: 13':'International migrant stock at mid-year (male)(2000)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 14':'International migrant stock at mid-year (male)(2005)
        df1.rename(columns = {'Unnamed: 15':'International migrant stock at mid-year (male)(2010)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 16':'International migrant stock at mid-year (male)(2015)'}, inplace = True)
        dfi.rename(columns = {'International migrant stock at mid-year (female)':'International migrant stock at mid-year (female)(1990)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 18':'International migrant stock at mid-year (female)(1995)'}, inplace = True'
        df1.rename(columns = {'Unnamed: 19':'International migrant stock at mid-year (female)(2000)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 20':'International migrant stock at mid-year (female)(2005)'}, inplace = True
        df1.rename(columns = {'Unnamed: 21':'International migrant stock at mid-year (female)(2010)'}, inplace = True)
        df1.rename(columns = {'Unnamed: 22':'International migrant stock at mid-year (female)(2015)'}, inplace = True
```

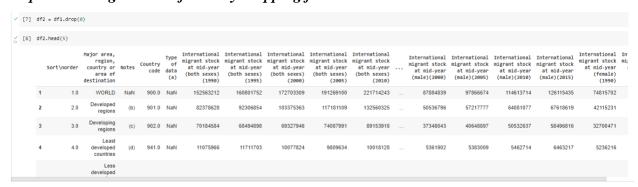
The names of some columns were unnamed 6, 7, 8 and etc. that do not signify the column appropriately due to which, it has been renamed with the help of giving an appropriate name to the columns.

Step 5: Reading the dataframe to check changed columns' name



After renaming the columns' name, the information of the data frame has been checked.

Step 6: Creating new dataframe by dropping first row



It is checked that the renamed columns itself contain the details of year, due to which the first row after the heading was not required. Hence, it is removed to get accuracy as some null values were also there. The df2.head(5) function is also used to check the data frame structure.

Step 7: Removing null values

newdf newdf	2 = df2.drops	na()													
	Sort\norder	Major area, region, country or area of destination	Notes	Country code	Type of data (a)		International migrant stock at mid-year (both sexes) (1995)						International migrant stock at mid-year (male)(2010)		Internat migrant at mid (fe
17	17.0	Mauritius	(1)	480.0	С	3613	7493	15543	19647	24836	5705	8943	13188	15832	
26	26.0	United Republic of Tanzania	(2)	834.0	BR	574025	1106043	928180	770846	308600	470962	486774	153984	130404	29
44	44.0	Sudan	(3)	729.0	BR	1402896	1053396	801883	541994	578363	403048	275915	294782	254530	7
67	67.0	Saint Helena	(4)	654.0	В	383	394	405	487	569	215	263	312	335	
79	79.0	China	(5)	156.0	С	376361	442198	508034	678947	849861	254082	379920	505758	600136	1
80	80.0	China, Hong Kong Special Administrative Region	(6)	344.0	В	2218473	2443798	2669122	2721235	2779950	1225629	1185121	1147539	1119957	10
81	81.0	China, Macao Special Administrative Region	(7)	446.0	В	205047	224929	240791	279308	318506	109391	130387	146407	155692	1
91	91.0	Malaysia	(8)	458.0	CR	695920	937368	1277223	1722344	2406011	712912	995607	1437206	1529630	2
110	110.0	Azerbaijan	(9)	31.0	BR	360600	344070	327540	302220	276901	141161	135336	129512	126600	2
112	112.0	Cyprus	(10)	196.0	В	43805	61941	80076	117165	187923	34745	50421	82284	86980	
					-										

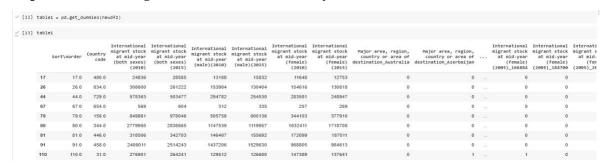
This step consists of dropping all the null values from the dataframe that can improve the accuracy of the result.

Step8: ".." in the datasheet has been removed with 0 using regex function



There are some fields in the dataframe that contain ".." values which do not signify anything and are not required for the data analytical part. Due to this reason, it is replaced with 0 value with the help of using regex function followed by reading the dataframe.

Step 9: Convert categorical variable into dummy/indicator variables



Get_dummies is used to convert categorical variables into dummy/indicator variables.

Step 10: Creating a series using pd.series

Pd.series is used to create a series of country code up to 10 in the list and reading the data. Pd.to_numeric is used to convert arguments to numeric types so that the data give accurate results.

Step 11: Hashing the value of a column



Hash function is used to return the hash value of the chosen column that returns an array of deterministic integers. The iloc function is used to select specific columns from the dataset.

Step 12: Renaming column name



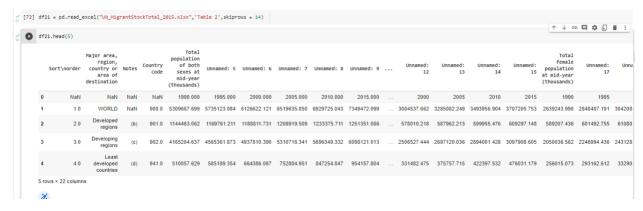
Columns' names are renamed to signify appropriate meaning.

Step 13: .eval function usage

Newdf3.eval function is used to evaluate the python expression as a string using various backbends.

Cleaning of table 2

Step 1: Importing the datasheet table 2



Import the datasheet as a dataframe df21 by skipping 14 rows of the above part that contains the heading of the sheet and reading the data frame.

Step 2: Getting the information of the dataframe

```
df21.info()
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 266 entries, 0 to 265
    Data columns (total 22 columns):
                                                                    Non-Null Count Dtype
     # Column
     0 Sort
    order
                                                         265 non-null
                                                                         float64
         Major area, region, country or area of destination
                                                                   265 non-null
                                                                                    object
         Notes
                                                                    26 non-null
                                                                                    object
         Country code
                                                                   265 non-null
                                                                                    float64
         Total population of both sexes at mid-year (thousands)
                                                                   266 non-null
                                                                                    float64
         Unnamed: 5
                                                                    266 non-null
                                                                                    float64
         Unnamed:
                                                                   266 non-null
                                                                                    float64
         Unnamed:
                                                                   266 non-null
                                                                                    float64
                                                                    266 non-null
                                                                    266 non-null
                                                                                    float64
     10
         Total male population at mid-year (thousands)
                                                                   266 non-null
                                                                                    object
     11 Unnamed: 11
                                                                   266 non-null
                                                                                    object
     12 Unnamed: 12
                                                                   266 non-null
                                                                                    object
     13
         Unnamed: 13
                                                                   266 non-null
                                                                                    object
         Unnamed: 14
                                                                   266 non-null
     14
                                                                                    object
         Unnamed: 15
                                                                   266 non-null
     15
                                                                                    object
         Total female population at mid-year (thousands)
                                                                   266 non-null
     16
                                                                                    object
         Unnamed: 17
                                                                   266 non-null
                                                                                    object
     18 Unnamed: 18
                                                                   266 non-null
                                                                                    object
         Unnamed: 19
                                                                    266 non-null
                                                                                    object
     20 Unnamed: 20
                                                                   266 non-null
     21 Unnamed: 21
                                                                   266 non-null
    dtypes: float64(8), object(14)
memory usage: 45.8+ KB
```

The detailed information of the data frame has been identified using df21.info() function.

Step 3: Renaming all the columns

```
df21.rename(columns = {'Total population of both sexes at mid-year (thousands)': 'Total population of both sexes at mid-year (thousands)(1995)'}, inplace = True)

df21.rename(columns = {'Unnamed: 6': 'Total population of both sexes at mid-year (thousands)(2009)'}, inplace = True)

df21.rename(columns = {'Unnamed: 7': 'Total population of both sexes at mid-year (thousands)(2005)'}, inplace = True)

df21.rename(columns = {'Unnamed: 8': 'Total population of both sexes at mid-year (thousands)(2005)'}, inplace = True)

df21.rename(columns = {'Unnamed: 8': 'Total population of both sexes at mid-year (thousands)(2010)'}, inplace = True)

df21.rename(columns = {'Unnamed: 9': 'Total population at mid-year (thousands)(2010)'}, inplace = True)

df21.rename(columns = {'Unnamed: 12': 'Total male population at mid-year (thousands)(2005)'}, inplace = True)

df21.rename(columns = {'Unnamed: 12': 'Total male population at mid-year (thousands)(2005)'}, inplace = True)

df21.rename(columns = {'Unnamed: 12': 'Total male population at mid-year (thousands)(2005)'}, inplace = True)

df21.rename(columns = {'Unnamed: 13': 'Total male population at mid-year (thousands)(2005)'}, inplace = True)

df21.rename(columns = {'Unnamed: 15': 'Total male population at mid-year (thousands)(2010)'}, inplace = True)

df21.rename(columns = {'Unnamed: 15': 'Total male population at mid-year (thousands)(2010)'}, inplace = True)

df21.rename(columns = {'Unnamed: 15': 'Total female population at mid-year (thousands)(2005)'}, inplace = True)

df21.rename(columns = {'Unnamed: 15': 'Total female population at mid-year (thousands)(2006)'}, inplace = True)

df21.rename(columns = {'Unnamed: 20': 'Total female population at mid-year (thousands)(2006)'}, inplace = True)

df21.rename(columns = {'Unnamed: 20': 'Total female population at mid-year (thousands)(2006)'}, inplace = True)

df21.rename(columns = {'Unnamed: 20': 'Total female population at mid-year (thousands)(2006)'}, inplace = True)

df21.rename(columns = {'Unnamed: 20': 'Total female population at mid-y
```

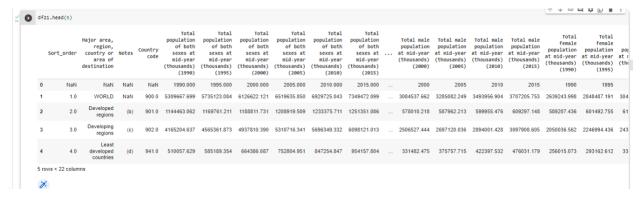
The renaming of columns has been done to get accurate insights of the column's name.

Step 4: Checking the data frame information

```
df21.info()
       <class 'pandas.core.frame.DataFrame'>
       RangeIndex: 266 entries, 0 to 265
       Data columns (total 22 columns):
       # Column
                                                                        Non-Null Count Dtype
       0 Sort order
                                                                         265 non-null
                                                                                        float64
           Major area, region, country or area of destination
                                                                         265 non-null
        1
                                                                                        object
                                                                        26 non-null
                                                                                        object
           Country code
                                                                        265 non-null
                                                                                        float64
           Total population of both sexes at mid-year (thousands)(1990) 266 non-null
        4
                                                                                        float64
            Total population of both sexes at mid-year (thousands)(1995) 266 non-null
                                                                                        float64
           Total population of both sexes at mid-year (thousands)(2000) 266 non-null
           Total population of both sexes at mid-year (thousands)(2005) 266 non-null
                                                                                        float64
           Total population of both sexes at mid-year (thousands)(2010) 266 non-null
                                                                                        float64
           Total population of both sexes at mid-year (thousands)(2015) 266 non-null
                                                                                        float64
        10 Total male population at mid-year (thousands)(1990)
                                                                        266 non-null
        11 Total male population at mid-year (thousands)(1995)
                                                                        266 non-null
                                                                                        object
        12 Total male population at mid-year (thousands)(2000)
                                                                        266 non-null
                                                                                        object
        13 Total male population at mid-year (thousands)(2005)
                                                                        266 non-null
                                                                                        object
        14 Total male population at mid-year (thousands)(2010)
                                                                       266 non-null
                                                                                        object
        15 Total male population at mid-year (thousands)(2015)
                                                                        266 non-null
                                                                                        object
        16 Total female population at mid-year (thousands)(1990)
                                                                        266 non-null
                                                                                        object
        17 Total female population at mid-year (thousands)(1995)
                                                                        266 non-null
                                                                                        object
        18 Total female population at mid-year (thousands)(2000)
                                                                        266 non-null
                                                                                        object
        19 Total female population at mid-year (thousands)(2005)
                                                                        266 non-null
                                                                                        obiect
        20 Total female population at mid-year (thousands)(2010)
                                                                        266 non-null
                                                                                        object
        21 Total female population at mid-year (thousands)(2015)
                                                                        266 non-null
                                                                                        object
       dtypes: float64(8), object(14)
       memory usage: 45.8+ KB
```

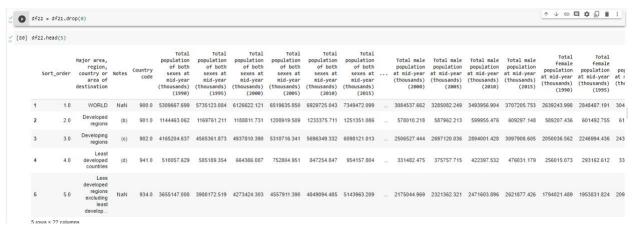
The data frame information has been obtained after the renaming of the data frame columns to check its functionality.

Step 5: Reading the dataframe



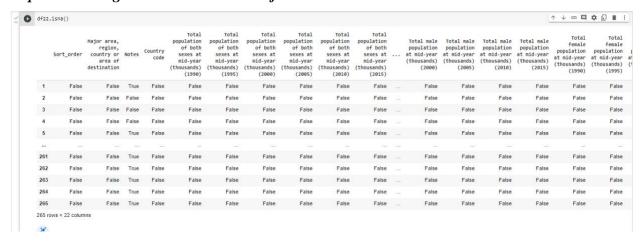
The data is read to visualize it appropriately.

Step 6: Dropping extra row and reading the data frame



Extra row of year is dropped from the dataframe as it has no utilization.

Step 7: Checking null values in the dataframe



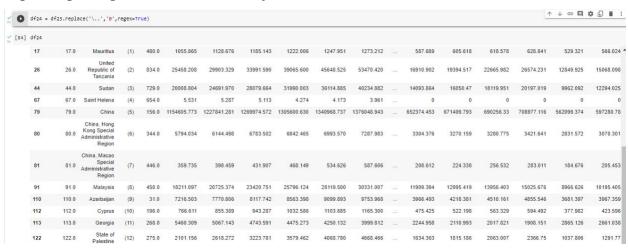
.isna() function is used to detect missing values of the dataframe.

Step 8: Dropping Null values from the dataframe and reading it



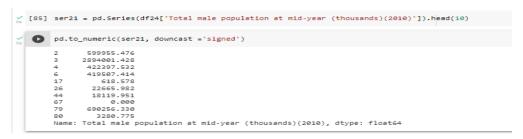
.dropna() is used to drop null values of the data frame followed by reading the data frame.

Step 9: Replacing ".." with 0 in the dataframe and read



".." values were there in the dataset that have been replaced with 0 using regex function so that accuracy can be improved followed by reading the dataframe.

Step 10: Converting numeric value to get accurate results



pd.Series function is used to take a series of 10 values of the total male population at mid-year (thousands) (2010) followed by converting the value to numeric to increase the accuracy of the data

Step 11: Hashing the dataframe value



Hash function is used to get the hash value of the column in the dataframe.

Step 12: Column rename

```
[88] df24.rename(columns = {'Country_code'; 'Country_code'}, inplace = True)
df24.rename(columns = {'Total_population of both sexes at mid-year (thousands)(2005)': 'Total_population_of_both_sexes_at_mid_year_(thousands)_(2005)'}, inplace = True)
```

Column names have been renamed to get the correct understanding.

Step 13: Used melt function

```
[89] df25 = pd.melt(df24, id_vars=['sort_order'], value_vars=list(df24.columns)[1:], var_name='Country_code', value_name='Total_population_of_both_sexes_at_mid_year_(thousands)_(2005)')

/usr/local/lib/python3.7/dist-packages/IPython/core/interactiveshell.py:3326: FutureWarning: This dataframe has a column name that matches the 'value_name' column name of the resulting Dataframe. In the fut exec(code_obj, self.user_global_ns, self.user_ns)
```

Melt function is used to manipulate the data frame in which Unpivot of the DataFrame has been done from wide to long format, optionally leaving identifiers set.

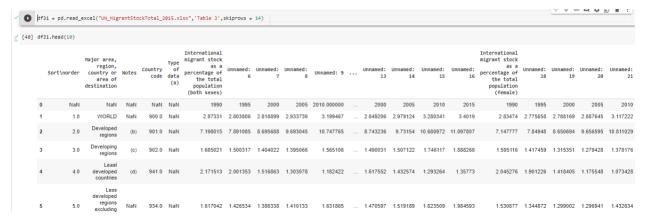
Step 14: Reading first 10 row data



The data frame is read to get the accurate value of the data.

Cleaning of table 3

Step 1: reading the table three



The table 3 has been imported by skipping 14 rows and it has read.

Step 2: Information of the data is checked

```
(41] df31.info()
       <class 'pandas.core.frame.DataFrame'>
       RangeIndex: 266 entries, 0 to 265
       Data columns (total 23 columns):
        # Column
                                                                                              Non-Null Count Dtype
           -----
        0
           Sort
       order
                                                                                    265 non-null
                                                                                                   float64
        1 Major area, region, country or area of destination
                                                                                             265 non-null
                                                                                                             object
        2
                                                                                              26 non-null
                                                                                                             object
           Notes
        3
            Country code
                                                                                             265 non-null
                                                                                                             float64
            Type of data (a)
            International migrant stock as a percentage of the total population (both sexes) 266 non-null
                                                                                                             obiect
           Unnamed: 6
                                                                                              266 non-null
                                                                                                             object
            Unnamed: 7
                                                                                              266 non-null
                                                                                                             object
        8
           Unnamed: 8
                                                                                              266 non-null
                                                                                                             object
           Unnamed: 9
                                                                                             266 non-null
                                                                                                             float64
        10 Unnamed: 10
                                                                                             266 non-null
                                                                                                             float64
        11 International migrant stock as a percentage of the total population (male)
                                                                                             266 non-null
                                                                                                             object
                                                                                             266 non-null
        12 Unnamed: 12
        13 Unnamed: 13
                                                                                              266 non-null
                                                                                                             object
        14 Unnamed: 14
                                                                                             266 non-null
                                                                                                             obiect
        15 Unnamed: 15
                                                                                              266 non-null
                                                                                                             object
                                                                                              266 non-null
        17 International migrant stock as a percentage of the total population (female)
                                                                                             266 non-null
                                                                                                             object
                                                                                              266 non-null
        18 Unnamed: 18
                                                                                                             obiect
        19 Unnamed: 19
                                                                                             266 non-null
                                                                                                             object
        20 Unnamed: 20
                                                                                              266 non-null
                                                                                                             object
        21 Unnamed: 21
                                                                                              266 non-null
                                                                                                             object
                                                                                              266 non-null
        22 Unnamed: 22
                                                                                                             object
       dtypes: float64(4), object(19)
       memory usage: 47.9+ KB
```

The information of the data is checked in the system to get proper information about data type non-null count and columns.

Step 3: columns are renamed

```
| Variance (columns = {'International migrant stock as a percentage of the total population (both sexes)': 'International migrant stock as a percentage of the total population (both sexes)(1990)'}, inplace = True)

df31.rename(columns = {'Unnamed: 6': 'International migrant stock as a percentage of the total population (both sexes)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 8': 'International migrant stock as a percentage of the total population (both sexes)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 8': 'International migrant stock as a percentage of the total population (both sexes)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 18': 'International migrant stock as a percentage of the total population (both sexes)(2016)'}, inplace = True)

df31.rename(columns = {'Unnamed: 18': 'International migrant stock as a percentage of the total population (male)(1995)'}, inplace = True)

df31.rename(columns = {'Unnamed: 12': 'International migrant stock as a percentage of the total population (male)(1995)'}, inplace = True)

df31.rename(columns = {'Unnamed: 12': 'International migrant stock as a percentage of the total population (male)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 13': 'International migrant stock as a percentage of the total population (male)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 13': 'International migrant stock as a percentage of the total population (male)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 15': 'International migrant stock as a percentage of the total population (male)(2005)'}, inplace = True)

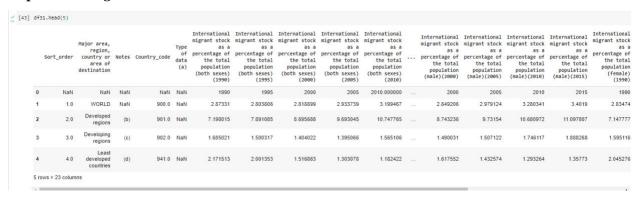
df31.rename(columns = {'Unnamed: 15': 'International migrant stock as a percentage of the total population (male)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 15': 'International migrant stock as a percentage of the total population (female)(2005)'}, inplace = True)

df31.rename(columns = {'Unnamed: 15': 'International migrant stock as a percentage of the total population (female)(2005)'}, inp
```

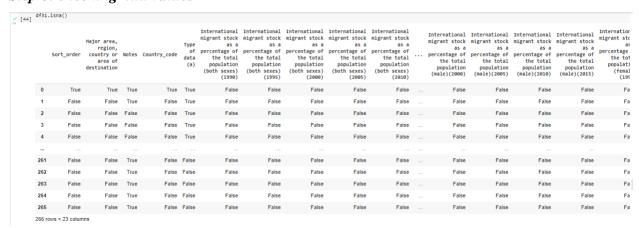
This step consists of renaming the columns to get accurate insights of the data.

Step 4: Reading the data



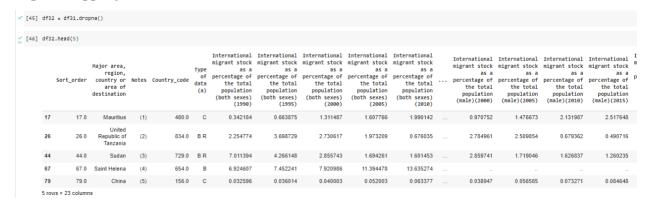
The data frame has been read to analyze the change columns name

Step 6: checking null values



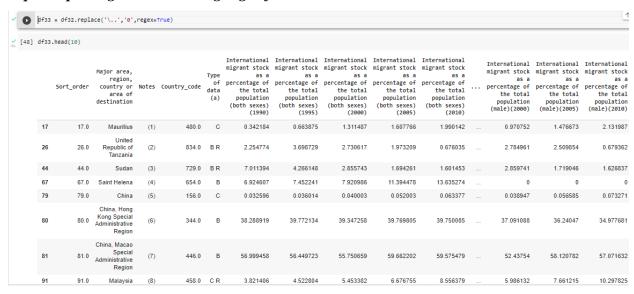
This step consist of checking the null values in the data frame

Step 7: dropping null values



This is consist of dropping the null values from the data frame

Step 8: replacing '..' with 0 using regex function



[&]quot;.." is replaced with zero using regex function since the row has no functionality in the data frame.

Step 9: hashing the value

```
hash(df33['International migrant stock as a percentage of the total population (male)(2015)'].values.tobytes())
       5140964633953251780
/ [50] df33.info()
       <class 'pandas.core.frame.DataFrame'>
       Int64Index: 22 entries, 17 to 240
       Data columns (total 23 columns):
        # Column
                                                                                                    Non-Null Count Dtype
        0 Sort_order
                                                                                                    22 non-null
                                                                                                                    float64
           Major area, region, country or area of destination
                                                                                                    22 non-null
                                                                                                                    object
        1
        2
           Notes
                                                                                                    22 non-null
                                                                                                                    object
            Country code
                                                                                                    22 non-null
                                                                                                                    float64
            Type of data (a)
                                                                                                    22 non-null
                                                                                                                    object
            International migrant stock as a percentage of the total population (both sexes)(1990)
                                                                                                    22 non-null
                                                                                                                    object
            International migrant stock as a percentage of the total population (both sexes)(1995)
                                                                                                    22 non-null
                                                                                                                    object
            International migrant stock as a percentage of the total population (both sexes)(2000)
                                                                                                    22 non-null
                                                                                                                    object
        8 International migrant stock as a percentage of the total population (both sexes)(2005) 22 non-null
                                                                                                                    object
            International migrant stock as a percentage of the total population (both sexes)(2010)
                                                                                                    22 non-null
                                                                                                                    float64
        10 International migrant stock as a percentage of the total population (both sexes)(2015) 22 non-null
                                                                                                                    float64
        11 International migrant stock as a percentage of the total population (male)(1990)
                                                                                                    22 non-null
                                                                                                                    object
        12 International migrant stock as a percentage of the total population (male)(1995)
                                                                                                    22 non-null
                                                                                                                    object
        13 International migrant stock as a percentage of the total population (male)(2000)
                                                                                                    22 non-null
                                                                                                                    object
        14 International migrant stock as a percentage of the total population (male)(2005)
                                                                                                    22 non-null
                                                                                                                    object
        15 International migrant stock as a percentage of the total population (male)(2010)
                                                                                                    22 non-null
                                                                                                                    object
        16 International migrant stock as a percentage of the total population (male)(2015)
                                                                                                    22 non-null
                                                                                                                    object
        17 International migrant stock as a percentage of the total population (female)(1990)
                                                                                                    22 non-null
                                                                                                                    object
        18 International migrant stock as a percentage of the total population (female)(1995)
                                                                                                    22 non-null
                                                                                                                    object
        19 International migrant stock as a percentage of the total population (female)(2000)
                                                                                                    22 non-null
                                                                                                                    object
        20 International migrant stock as a percentage of the total population (female)(2005)
                                                                                                    22 non-null
                                                                                                                    object
        21 International migrant stock as a percentage of the total population (female)(2010)
                                                                                                    22 non-null
                                                                                                                    object
        22 International migrant stock as a percentage of the total population (female)(2015)
                                                                                                    22 non-null
                                                                                                                    object
       dtvpes: float64(4), object(19)
```

This is step consist of hashing the value to buy and getting information of the data frame

Step 10: Renaming the columns

This step consists of renaming the columns 9f the data frame to get an accurate understanding.

Step 11: melt function usage

Melt function is used in the data frame to manipulate the data with the help of unpivot a data frame from white to long format.

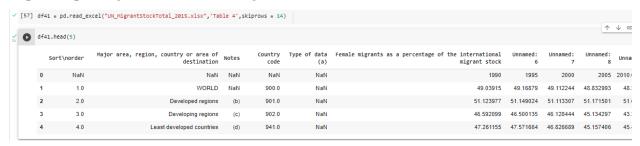
Step 12: reading the data frame



Data manipulation is done using the .melt() function in which Unpivot a DataFrame from wide to long format, optionally leaving identifiers set.

Cleaning of table 4

Step 1: Importing the data set in Google colab



This stage consists of importing and reading the data set to the Google colab using pd.read and df.head() functions.

Step 2: Checking the data information

```
 [59] df41.info()
         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 266 entries, 0 to 265
         Data columns (total 11 columns):
# Column
                                                                                              Non-Null Count Dtype
                                                                                  265 non-null
                                                                                                    float64
         order
              Major area, region, country or area of destination
          2 Notes
                                                                                              26 non-null
                                                                                                                 object
              Country code
Type of data (a)
                                                                                              232 non-null
                                                                                                                 obiect
              Female migrants as a percentage of the international migrant stock 266 non-null Unnamed: 6 266 non-null
                                                                                                                 object
                                                                                                                object
object
              Unnamed: 7
                                                                                              266 non-null
              Unnamed: 9
                                                                                              266 non-null
                                                                                                                 float64
        dtypes: float64(4), object(7) memory usage: 23.0+ KB
```

In this stage df41.info() is used to check the data information from the dataframe.

Step 3: renaming the columns and reading the data frame



Renaming of the columns has been done along with reading the data frame.

Step 4: Dropping null values



.dropna() is used to drop the null values.

Step 5: Checking the data frame information

```
df42.info()
      <class 'pandas.core.frame.DataFrame'>
     Int64Index: 22 entries, 17 to 240 Data columns (total 11 columns):
           Column
                                                                                                          Non-Null Count
                                                                                                                              Dtype
                                                                                                          22 non-null
                                                                                                          22 non-null
            Major area, region, country or area of destination
                                                                                                                               object
            Notes
                                                                                                          22 non-null
                                                                                                                              object
           Country_code
Type of data (a)
                                                                                                          22 non-null
                                                                                                                               float64
           Female_migrants_as_a_percentage_of_the_international_migrant_stock_1990
Female_migrants_as_a_percentage_of_the_international_migrant_stock_1995
                                                                                                          22 non-null
                                                                                                                              object
                                                                                                          22 non-null
                                                                                                                              obiect
            Female_migrants_as_a_percentage_of_the_international_migrant_stock_2000
                                                                                                                               object
           Female_migrants_as_a_percentage_of_the_international_migrant_stock_2005
Female_migrants_as_a_percentage_of_the_international_migrant_stock_2010
                                                                                                          22 non-null
                                                                                                                              object
                                                                                                         22 non-null
                                                                                                                               float64
       10 Female_migrants_as_a_percentage_of_the_international_migrant_stock_2015
      dtypes: float64(4), object(7)
      memory usage: 2.1+ KB
```

Data frame information is checked using .info() function

Step 6: Replacing unnecessary value with zero

df4	43 = df42.rep	olace('\','0	',regex	=True)			
df4	43.head(5)						↑ ↓ © □ ‡ [] ■
	Sort_order	Major area, region, country or area of destination	Notes	Country_code	Type of data (a)	Female_migrants_as_a_percentage_of_the_international_migrant_stock_1990	Female_migrants_as_a_percentage_of_the_international_migrant_stock_1995
17	17.0	Mauritius	(1)	480.0	С	51.203986	56.919792
26	26.0	United Republic of Tanzania	(2)	834.0	BR	50.63107	50.536372
44	44.0	Sudan	(3)	729.0	BR	50.360896	50.094646
67		Saint Helena	(4)	654.0	В	42.036554	44.670051
79	79.0	China	(5)	156.0	C	48.997372	49.566032
4							

".." value is replaced with 0 using the regex function.

Step 7: Creating a series of 10 values in the concerned column of data frame and its numeric conversation

```
/ [68] ser43 = pd.Series(df43['Female_migrants_as_a_percentage_of_the_international_migrant_stock_2015']).head(10)
/ [69] pd.to_numeric(ser43, downcast ='signed')
       17
              44.614308
       26
             50.079243
       44
             49.445556
       67
             44.536424
       79
              38.639287
       80
              60.546348
       81
             54.569408
       91
             39.161410
       110
             52.089191
       112
              55.660228
       Name: Female_migrants_as_a_percentage_of_the_international_migrant_stock_2015, dtype: float64
```

A series of 10 values have been created and it is converted to numeric values.

Step 8: Returning unique values based on hash table and reading the data frame

```
7 [70] df44 = df43['Female_migrants_as_a_percentage_of_the_international_migrant_stock_2015'].unique()
7 [72] print(df44)

[44.6143082 50.07924294 49.44555561 44.53642384 38.6392869 60.54634837 54.5694085 39.16140962 52.08919131 55.66022827 56.83641189 55.68144904 64.63010133 52.20576622 49.14224027 47.81730706 53.25 56.01164667 51.21814578 51.60744501 55.47727573 50.65425347]
```

Unique values of the column are returned using the .unique() function and the value is printed to check the values.

Cleaning of table 5

Step 1: Reading the table five by importing it as a data frame

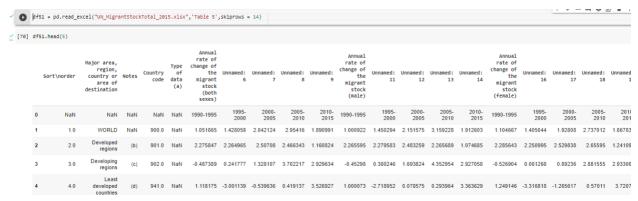


Table 5 is imported to google colab as a dataframe by skipping 14 rows and it has read.

Step 2: Checking the data frame information

```
// [71] df51.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 266 entries, 0 to 265
       Data columns (total 20 columns):
           Column
                                                                      Non-Null Count Dtype
        0
            Sort
        order
                                                            265 non-null
            Major area, region, country or area of destination
        2
            Notes
                                                                      26 non-null
                                                                                      object
            Country code
                                                                      265 non-null
                                                                                      float64
            Type of data (a)
                                                                      232 non-null
            Annual rate of change of the migrant stock (both sexes)
        5
                                                                      266 non-null
                                                                                      object
            Unnamed: 6
                                                                      266 non-null
                                                                                      object
            Unnamed: 7
                                                                      266 non-null
                                                                                      object
        8
            Unnamed: 8
                                                                      266 non-null
                                                                                      object
            Unnamed: 9
                                                                      266 non-null
                                                                                      object
           Annual rate of change of the migrant stock (male)
                                                                      266 non-null
                                                                                      object
        11 Unnamed: 11
                                                                      266 non-null
                                                                                      object
        12 Unnamed: 12
                                                                      266 non-null
                                                                                      object
            Unnamed: 13
                                                                      266 non-null
        14
            Unnamed: 14
                                                                      266 non-null
                                                                                      object
            Annual rate of change of the migrant stock (female)
        15
                                                                      266 non-null
                                                                                      object
        16 Unnamed: 16
                                                                      266 non-null
                                                                                      object
            Unnamed: 17
                                                                      266 non-null
        17
                                                                                      object
        18
            Unnamed: 18
                                                                      266 non-null
                                                                                      object
        19 Unnamed: 19
                                                                      266 non-null
                                                                                      object
       dtypes: float64(2), object(18)
       memory usage: 41.7+ KB
```

The information of the dataframe is checked using .info() function

Step 3: Renaming the columns name

```
/*Total fish.rename(columns = {'Annual rate of change of the migrant stock (both sexes)': 'Annual rate of change of the migrant stock (both sexes)(1995-2080)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 7': 'Annual rate of change of the migrant stock (both sexes)(2096-2080)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 8': 'Annual rate of change of the migrant stock (both sexes)(2096-2080)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 9': 'Annual rate of change of the migrant stock (both sexes)(2096-2080)'}, inplace = True)

dfSi.rename(columns = {'Annual rate of change of the migrant stock (male)': 'Annual rate of change of the migrant stock (male)(1996-1995)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 11': 'Annual rate of change of the migrant stock (male)(2008-2005)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 12': 'Annual rate of change of the migrant stock (male)(2008-2005)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 12': 'Annual rate of change of the migrant stock (male)(2008-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 14': 'Annual rate of change of the migrant stock (male)(2008-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 16': 'Annual rate of change of the migrant stock (male)(2008-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 16': 'Annual rate of change of the migrant stock (female)(2008-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 18': 'Annual rate of change of the migrant stock (female)(2008-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 18': 'Annual rate of change of the migrant stock (female)(2008-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 18': 'Annual rate of change of the migrant stock (female)(2008-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 18': 'Annual rate of change of the migrant stock (female)(2018-2010)'}, inplace = True)

dfSi.rename(columns = {'Unnamed: 18': 'Annual rate of change of the migrant stock (female)(2018-2010)'}, inplace = True
```

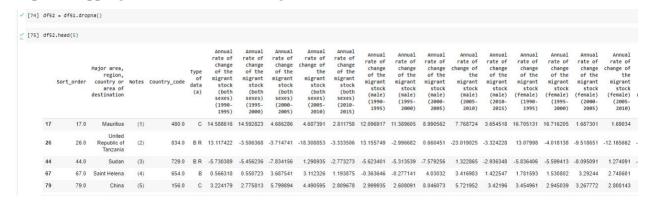
Column names are renamed to get accurate insights of the data.

Step 4: Reading the data frame after the renaming of columns

	Sort_order	Major area, region, country or area of destination	Notes	Country_code	Type of data (a)	Annual rate of change of the migrant stock (both sexes) (1990- 1995)	Annual rate of change of the migrant stock (both sexes) (1995- 2000)	Annual rate of change of the migrant stock (both sexes) (2000- 2005)	Annual rate of change of the migrant stock (both sexes) (2005- 2010)	Annual rate of change of the migrant stock (both sexes) (2010- 2015)	Annual rate of change of the migrant stock (male) (1990- 1995)	Annual rate of change of the migrant stock (male) (1995- 2000)	Annual rate of change of the migrant stock (male) (2000- 2005)	Annual rate of change of the migrant stock (male) (2005- 2010)	Annual rate of change of the migrant stock (male) (2010- 2015)	Annual rate of change of the migrant stock (female) (1990- 1995)	Annual rate of change of the migrant stock (female) (1995- 2000)	Annual rate of change of the migrant stock (female) (2000- 2005)	Annirate character character structure (fema. (20) 20)
0	NaN	NaN	NaN	NaN	NaN	1990- 1995	1995- 2000	2000- 2005	2005- 2010	2010- 2015	1990- 1995	1995- 2000	2000- 2005	2005- 2010	2010- 2015	1990- 1995	1995- 2000	2000- 2005	2
1	1.0	WORLD	NaN	900.0	NaN	1.051865	1.428058	2.042124	2.95416	1.890991	1.000922	1.450294	2.151575	3.159228	1.912603	1.104667	1.405044	1.92808	2.737
2	2.0	Developed regions	(b)	901.0	NaN	2.275847	2.264965	2.50708	2.466343	1.160824	2.265595	2.279583	2.483259	2.265689	1.074685	2.285643	2.250995	2.529838	2.65
3	3.0	Developing regions	(c)	902.0	NaN	-0.487389	0.241777	1.328107	3.702217	2.929634	-0.45298	0.380246	1.693824	4.352954	2.927058	-0.526904	0.081268	0.89236	2.88
4	4.0	Least developed countries	(d)	941.0	NaN	1.118175	-3.001139	-0.539636	0.419137	3.526927	1.000073	-2.718952	0.078575	0.293964	3.363629	1.249146	-3.316818	-1.265617	0.5

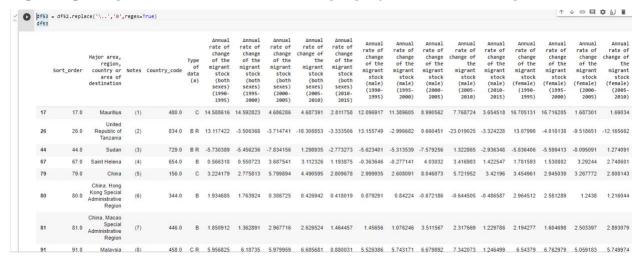
.head() function is used to read the data and check if the column names are renamed.

Step 5: Dropping null values and reading the data



Null values of dropped from the dataframe using .dropna() and it has read using .head()

Step 6: Replacing unwanted value with zero using regex function and reading the data



".." value from the dataframe is replaced with 0 using the regex function and the data is read using the .head() function.

Step 7: Checking data information

```
// [77] df53.info()
       <class 'pandas.core.frame.DataFrame'>
       Int64Index: 22 entries, 17 to 240
       Data columns (total 20 columns):
        # Column
                                                                                Non-Null Count Dtype
                                                                                22 non-null
        0
            Sort order
                                                                                                float64
            Major area, region, country or area of destination
                                                                                22 non-null
                                                                                                object
        2
            Notes
                                                                                22 non-null
                                                                                                object
            Country_code
                                                                                                float64
        3
                                                                                22 non-null
            Type of data (a)
                                                                                22 non-null
                                                                                                object
                                                                                22 non-null
            Annual rate of change of the migrant stock (both sexes)(1990-1995)
                                                                                                object
            Annual rate of change of the migrant stock (both sexes)(1995-2000)
                                                                                                object
            Annual rate of change of the migrant stock (both sexes)(2000-2005)
                                                                                22 non-null
                                                                                                object
            Annual rate of change of the migrant stock (both sexes)(2005-2010) 22 non-null
                                                                                                object
            Annual rate of change of the migrant stock (both sexes)(2010-2015) 22 non-null
                                                                                                object
        10
            Annual rate of change of the migrant stock (male)(1990-1995)
                                                                                22 non-null
                                                                                                object
           Annual rate of change of the migrant stock (male)(1995-2000)
                                                                                22 non-null
                                                                                                object
            Annual rate of change of the migrant stock (male)(2000-2005)
                                                                                22 non-null
        12
                                                                                                object
        13 Annual rate of change of the migrant stock (male)(2005-2010)
                                                                                22 non-null
                                                                                                object
            Annual rate of change of the migrant stock (male)(2010-2015)
                                                                                22 non-null
                                                                                                object
            Annual rate of change of the migrant stock (female)(1990-1995)
                                                                                22 non-null
                                                                                                object
            Annual rate of change of the migrant stock (female)(1995-2000)
                                                                                22 non-null
            Annual rate of change of the migrant stock (female)(2000-2005)
                                                                                22 non-null
                                                                                                object
        18 Annual rate of change of the migrant stock (female)(2005-2010)
                                                                                22 non-null
                                                                                                object
        19 Annual rate of change of the migrant stock (female)(2010-2015)
                                                                                22 non-null
                                                                                                object
        dtypes: float64(2), object(18)
        memory usage: 3.6+ KB
```

Data information is obtained using .info()

Step 8: Evaluate python expression as string using various back ends for top level evaluation

```
( [78] print(df53.eval("Country_code + Sort_order"))
        17
              497.0
        26
              860.0
        44
              773.0
        67
              721.0
        79
              235.0
        80
              424.0
              527.0
        81
        91
              549.0
        110
              141.0
        112
              308.0
        113
              381.0
        122
              397.0
        134
              632.0
        140
              970.0
        144
              390.0
        150
              728.0
        160
              496.0
        166
               854.0
        168
        187
              722.0
        195
               507.0
        240
               276.0
        dtype: float64
```

.eval() is used to get top level evaluation of the dataframe.

Step 9: Returning unique values based on hash table

```
/ [79] df54 = df53['Country_code'].unique()

/ [80] print(df54)

[480. 834. 729. 654. 156. 344. 446. 458. 31. 196. 268. 275. 498. 830. 246. 578. 336. 688. 724. 535. 312. 36.]
```

Unique values of the column are returned based on the hash table and the value is printed using print function to check the values.

Cleaning of table 6

Step 1: Importing the table 6 in the form of data frame by skipping 14 rows and reading it

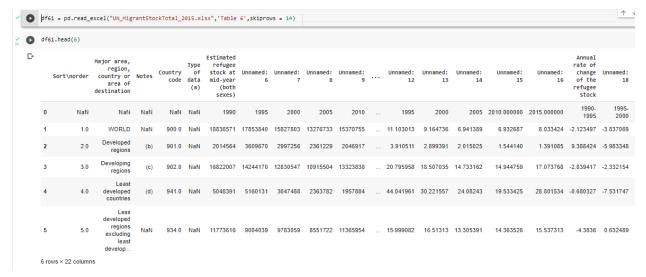


Table 6 is imported to google colab and its data is read using .head() function.

Step 2: Getting the information of the data frame

```
[83] df61.info()
       <class 'pandas.core.frame.DataFrame'>
       RangeIndex: 266 entries, 0 to 265
       Data columns (total 22 columns):
           Column
                                                                         Non-Null Count Dtype
        0
           Sort
       order
                                                               265 non-null
                                                                               float64
        1 Major area, region, country or area of destination
                                                                         265 non-null
                                                                                         object
        2
            Notes
                                                                         26 non-null
                                                                                         object
            Country code
                                                                         265 non-null
                                                                                         float64
            Type of data (a)
                                                                         232 non-null
                                                                                         object
            Estimated refugee stock at mid-year (both sexes)
                                                                         266 non-null
                                                                                         object
                                                                         266 non-null
            Unnamed: 6
            Unnamed: 7
                                                                         266 non-null
                                                                                         object
            Unnamed: 8
                                                                         266 non-null
                                                                                         object
            Unnamed: 9
                                                                         266 non-null
                                                                         266 non-null
                                                                                         int64
        10 Unnamed: 10
        11 Refugees as a percentage of the international migrant stock 266 non-null
                                                                                         object
                                                                         266 non-null
                                                                                         object
        13 Unnamed: 13
                                                                         266 non-null
                                                                                         object
        14 Unnamed: 14
                                                                         266 non-null
                                                                                         object
            Unnamed: 15
                                                                         266 non-null
                                                                                         float64
                                                                                         float64
        16 Unnamed: 16
                                                                         266 non-null
        17 Annual rate of change of the refugee stock
                                                                         266 non-null
                                                                                         object
        18
            Unnamed: 18
                                                                         266 non-null
                                                                                         object
        19 Unnamed: 19
                                                                         266 non-null
                                                                                         object
        20 Unnamed: 20
                                                                         266 non-null
                                                                                         object
        21 Unnamed: 21
                                                                         266 non-null
       dtypes: float64(4), int64(2), object(16)
       memory usage: 45.8+ KB
```

Dataframe information is obtained using .info()

Step 3: Renaming the column name

```
### df61.rename(columns = {'Estimated refugee stock at mid-year (both sexes)': 'Estimated refugee stock at mid-year (both sexes)(1999)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 6': 'Estimated refugee stock at mid-year (both sexes)(1999)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 8': 'Estimated refugee stock at mid-year (both sexes)(2005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 8': 'Estimated refugee stock at mid-year (both sexes)(2005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 10': 'Estimated refugee stock at mid-year (both sexes)(2010)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 10': 'Estimated refugee stock at mid-year (both sexes)(2015)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 12': 'Refugees as a percentage of the international migrant stock(1005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 12': 'Refugees as a percentage of the international migrant stock(1005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 12': 'Refugees as a percentage of the international migrant stock(2005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 15': 'Refugees as a percentage of the international migrant stock(2005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 15': 'Refugees as a percentage of the international migrant stock(2005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 15': 'Refugees as a percentage of the international migrant stock(2015)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 15': 'Refugees as a percentage of the international migrant stock(2015)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 15': 'Refugees stock': 'Annual rate of change of the refugee stock(2008-2005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 15': 'Annual rate of change of the refugee stock(2008-2005)'}, inplace = True)

### df61.rename(columns = ('Unnamed: 20': 'Annual rate of change of the refugee stock(2010-2015)'}, inplace = True)

### df61.rename(columns = ('Unna
```

Column names are renamed to get accurate insights of the column values.

Step 4: Reading the data frame

df6	1.head(5)														
	Sort_order	Major area, region, country or area of destination	Notes	Country_code	Type of data (a)	Estimated refugee stock at mid-year (both sexes) (1990)	Estimated refugee stock at mid-year (both sexes) (1995)	Estimated refugee stock at mid-year (both sexes) (2000)	Estimated refugee stock at mid-year (both sexes) (2005)	Estimated refugee stock at mid-year (both sexes) (2010)	 percentage of the	percentage of the	Refugees as a percentage of the international migrant stock(2005)	percentage of the	percentage o
0	NaN	NaN	NaN	NaN	NaN	1990	1995	2000	2005	2010	 1995	2000	2005	2010.000000	2015.00000
1	1.0	WORLD	NaN	900.0	NaN	18836571	17853840	15827803	13276733	15370755	 11.103013	9.164736	6.941389	6.932687	8.03342
2	2.0	Developed regions	(b)	901.0	NaN	2014564	3609670	2997256	2361229	2046917	 3.910511	2.899391	2.015025	1.544140	1.39108
3	3.0	Developing regions	(c)	902.0	NaN	16822007	14244170	12830547	10915504	13323838	 20.795958	18.507035	14.733162	14.944759	17.073768
4	4.0	Least developed countries	(d)	941.0	NaN	5048391	5160131	3047488	2363782	1957884	44.041961	30.221557	24.08243	19.533425	28.80153

The data frame is read using .head() function

Step 5: Dropping null values

```
✓ [86] df62 = df61.dropna()
```

Null values of the dataframe are dropped using .dropna()

Step 6: Replacing unwanted value with zero using regex and checking data

		Sort_order	Major area, region, country or area of destination	Notes	Country_code	Type of data (a)	Estimated refugee stock at mid-year (both sexes) (1990)	Estimated refugee stock at mid-year (both sexes) (1995)	Estimated refugee stock at mid-year (both sexes) (2000)	Estimated refugee stock at mid-year (both sexes) (2005)
	17	17.0	Mauritius	(1)	480.0	С	0	0	0	0
	26	26.0	United Republic of Tanzania	(2)	834.0	BR	265184	829671	680862	548824
	44	44.0	Sudan	(3)	729.0	BR	1031050	674071	414928	147256
	67	67.0	Saint Helena	(4)	654.0	В	0	0	0	0
	79	79.0	China	(5)	156.0	С	285788	289747	293705	297346
5	5 rov	vs × 22 column	IS							

[&]quot;.." value is replaced with 0 using the regex function.

Step 7: Creating dummies data to convert categorical variable into dummy

table table		_dummies(df63)								
	Sort_order	Country_code	Estimated refugee stock at mid-year (both sexes) (2010)	Estimated refugee stock at mid-year (both sexes) (2015)	percentage of the	Refugees as a percentage of the international migrant stock(2015)	Major area, region, country or area of destination_Australia	Major area, region, country or area of destination_Azerbaijan	Major area, region, country or area of destination_Bonaire, Sint Eustatius and Saba	Major area, region, country or area of destination_channel Islands
17	17.0	480.0	0	0	0.000000	0.000000	0	0	0	0
26	26.0	834.0	109286	90650	35.413480	34.702284	0	0	0	0
44	44.0	729.0	144008	205174	24.899241	40.751415	0	0	0	0
67	67.0	654.0	0	0	0.000000	0.000000	0	0	0	0
79	79.0	156.0	300986	301052	35.415909	30.780965	0	0	0	0
80	80.0	344.0	154	184	0.005540	0.006482	0	0	0	0
81	81.0	446.0	0	0	0.000000	0.000000	0	0	0	0
91	91.0	458.0	81516	98207	3.388014	3.906027	0	0	0	0
110	110.0	31.0	1908	1380	0.689055	0.522251	0	1	0	0
112	112.0	196.0	3394	4281	1.806059	2.182324	0	0	0	0
113	113.0	268.0	639	857	0.350710	0.507695	0	0	0	0
122	122.0	275.0	1955469	2051096	757.839725	802.755306	0	0	0	0
134	134.0	498.0	148	283	0.093868	0.198035	0	0	0	0

Dummy data is created to convert categorical variables into dummies using .get_dummies().

Step 8: Checking the data frame information

```
/ [89] df63.info()
        <class 'pandas.core.frame.DataFrame'>
       Int64Index: 22 entries, 17 to 240
       Data columns (total 22 columns):
        # Column
                                                                              Non-Null Count Dtype
        0 Sort_order
                                                                              22 non-null
                                                                                              float64
        1
            Major area, region, country or area of destination
                                                                               22 non-null
                                                                                              object
        2 Notes
                                                                              22 non-null
                                                                                              obiect
        3
           Country_code
                                                                              22 non-null
                                                                                              float64
            Type of data (a)
                                                                              22 non-null
                                                                                              object
           Estimated refugee stock at mid-year (both sexes)(1990)
                                                                              22 non-null
                                                                                              object
        6 Estimated refugee stock at mid-year (both sexes)(1995)
                                                                             22 non-null
            Estimated refugee stock at mid-year (both sexes)(2000)
                                                                              22 non-null
                                                                                              object
        8 Estimated refugee stock at mid-year (both sexes)(2005)
                                                                              22 non-null
                                                                                              object
        9 Estimated refugee stock at mid-year (both sexes)(2010)
                                                                              22 non-null
                                                                                              int64
        10 Estimated refugee stock at mid-year (both sexes)(2015)
                                                                              22 non-null
                                                                                              int64
        11 Refugees as a percentage of the international migrant stock(1990) 22 non-null
                                                                                              object
        12 Refugees as a percentage of the international migrant stock(1995) 22 non-null
                                                                                              object
        13 Refugees as a percentage of the international migrant stock(2000) 22 non-null
                                                                                              object
            Refugees as a percentage of the international migrant stock(2005)
                                                                              22 non-null
                                                                                              object
        15 Refugees as a percentage of the international migrant stock(2010) 22 non-null
                                                                                              float64
        16 Refugees as a percentage of the international migrant stock(2015) 22 non-null
                                                                                              float64
        17 Annual rate of change of the refugee stock(1990-1995)
                                                                              22 non-null
                                                                                              object
        18 Annual rate of change of the refugee stock(1995-2000)
                                                                              22 non-null
                                                                                              object
        19 Annual rate of change of the refugee stock(2000-2005)
                                                                              22 non-null
                                                                                              object
        20 Annual rate of change of the refugee stock(2005-2010)
                                                                              22 non-null
                                                                                              object
        21 Annual rate of change of the refugee stock(2010-2015)
                                                                              22 non-null
                                                                                              object
        dtypes: float64(4), int64(2), object(16)
       memory usage: 4.0+ KB
```

The dataframe information is checked using .info()

Step 9: Rename in the column and checking the data information

```
bf63.rename(columns = {'Estimated refugee stock at mid-year (both sexes)(2010)': Estimated_refugee_stock_at_mid_year_both_sexes_2010'}, inplace = True)
[91] df63.info()
           <class 'pandas.core.frame.DataErame's</pre>
          Data columns (total 22 columns):
                                                                                                              Non-Null Count Dtype
                Column
                 Sort order
                                                                                                               22 non-null
                                                                                                                                     float64
                 Major area, region, country or area of destination
                                                                                                                                     object
                 Notes
                                                                                                               22 non-null
                                                                                                                                     object
                                                                                                                                     float64
                 Type of data (a)
                                                                                                               22 non-null
                                                                                                                                     object
                 Estimated refugee stock at mid-year (both sexes)(1990)
                                                                                                               22 non-null
                                                                                                                                     object
                 Estimated refugee stock at mid-year (both sexes)(1995)
                                                                                                                                     object
                 Estimated refugee stock at mid-year (both sexes)(2000)
                                                                                                               22 non-null
                                                                                                                                     object
                 Estimated refugee stock at mid-year (both sexes)(2005)
           9 Estimated_refugee_stock_at_mid_year_both_sexes_2010 22 non-null
10 Estimated refugee stock at mid-year (both sexes)(2015) 22 non-null
11 Refugees as a percentage of the international migrant stock(1990) 22 non-null
                                                                                                                                     int64
                                                                                                                                     int64
                                                                                                                                     object
           12 Refugees as a percentage of the international migrant stock(1995)
                                                                                                              22 non-null
                                                                                                                                     object
                 Refugees as a percentage of the international migrant stock(2009)
Refugees as a percentage of the international migrant stock(2005)
                                                                                                                                     object
                                                                                                              22 non-null
                                                                                                                                     object
           15 Refugees as a percentage of the international migrant stock(2010) 22 non-null
16 Refugees as a percentage of the international migrant stock(2015) 22 non-null
17 Annual rate of change of the refugee stock(1990-1995) 22 non-null
                                                                                                                                     float64
                                                                                                                                     float64
                                                                                                                                     object
           18 Annual rate of change of the refugee stock(1995-2000)
19 Annual rate of change of the refugee stock(2000-2005)
                                                                                                                                     object
                                                                                                              22 non-null
                                                                                                                                     object
           20 Annual rate of change of the refugee stock(2005-2010)
21 Annual rate of change of the refugee stock(2010-2015)
                                                                                                              22 non-null
                                                                                                                                    object
          dtypes: float64(4), int64(2), object(16) memory usage: 4.0+ KB
```

Renaming the column and checking the data information has been done in this stage using .rename() and .info() functions

Step 10: Creating a series of a column and converting the data to numeric

A series of 5 values of the column has been created and its values are converted to numeric values.

Step 11: Top level evaluation

```
// [94] print(df63.eval("Sort_order + Estimated_refugee_stock_at_mid_year_both_sexes_2010"))
        17
                    17.0
        26
                109312.0
        44
                144052.0
        67
                    67.0
        79
                301065.0
        80
                   234.0
        81
                    81.0
        91
                81607.0
                 2018.0
        110
                  3506.0
        112
        113
                  752.0
        122
              1955591.0
        134
                   282.0
        140
                   140.0
        144
                 8868.0
        150
                40410.0
                   160.0
        160
        166
                 73774.0
        168
                  3988.0
        187
                   187.0
        195
                   195.0
                 25745.0
        240
        dtype: float64
```

.eval () is used to get top level evaluation and its value is printed.

Step 12: Manipulating the data



Manipulation of the data is done using .melt() and the data is read.

Cleaning of table 6

Step 1: Importing the table 6 in the form of a data frame and reading it.

df	61.head(6)																		
	Sort\norder	Major area, region, country or area of destination	Notes	Country code	Type of data (a)	Estimated refugee stock at mid-year (both sexes)	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	 Unnamed:	Unnamed:	Unnamed: 14	Unnamed: 15	Unnamed: 16	Annual rate of change of the refugee stock	Unnamed: 18	Unnamed:	Unnam
0	NaN	NaN	NaN	NaN	NaN	1990	1995	2000	2005	2010	1995	2000	2005	2010.000000	2015.000000	1990- 1995	1995- 2000	2000- 2005	20 2
1	1.0	WORLD	NaN	900.0	NaN	18836571	17853840	15827803	13276733	15370755	11.103013	9.164736	6.941389	6.932687	8.033424	-2.123497	-3.837069	-5.557223	-0.025
2	2.0	Developed regions	(b)	901.0	NaN	2014564	3609670	2997256	2361229	2046917	3.910511	2.899391	2.015025	1.544140	1.391085	9.388424	-5.983348	-7.277379	-5.323
3	3.0	Developing regions	(c)	902.0	NaN	16822007	14244170	12830547	10915504	13323838	20.795958	18.507035	14.733162	14.944759	17.073768	-2.839417	-2.332154	-4.561	0.285
4	4.0	Least developed countries	(d)	941.0	NaN	5048391	5160131	3047488	2363782	1957884	44.041961	30.221557	24.08243	19.533425	28.801534	-0.680327	-7.531747	-4.541459	-4.187
5	5.0	Less developed regions excluding least develop	NaN	934.0	NaN	11773616	9084039	9783059	8551722	11365954	 15.999082	16.51313	13.305391	14.363526	15.537313	-4.3836	0.632489	-4.319731	1.530

Table 6 is imported to google colab by skipping 14 rows and the data frame is read.

Step 2: Checking the data frame information

```
colass 'pandas.core.frame.DataFrame'>
RangeIndex: 266 entries, 0 to 265
Data columns (total 22 columns):

# Column

Column

Bort

Bort

Column

Conder

1 Major area, region, country or area of destination

265 non-null floats4
28 non-null object
28 non-null object
29 non-null object
20 non-null object
20 non-null object
20 non-null object
20 non-null object
21 non-null object
22 non-null object
23 non-null object
24 Type of data (a)
25 Index of the series of t
```

Dataframe info is checked using .info()

Step 3: Renaming the column name

```
[] df61.rename(columns = {'Estimated refugee stock at mid-year (both sexes)':'Estimated refugee stock at mid-year (both sexes)(1990)'}, inplace = True)
df61.rename(columns = {'Unnamed: 6':'Estimated refugee stock at mid-year (both sexes)(2000)'}, inplace = True)
df61.rename(columns = {'Unnamed: 8':'Estimated refugee stock at mid-year (both sexes)(2000)'}, inplace = True)
df61.rename(columns = {'Unnamed: 8':'Estimated refugee stock at mid-year (both sexes)(2005)'}, inplace = True)
df61.rename(columns = {'Unnamed: 9':'Estimated refugee stock at mid-year (both sexes)(2010)'}, inplace = True)
df61.rename(columns = {'Unnamed: 10':'Estimated refugee stock at mid-year (both sexes)(2015)'}, inplace = True)

df61.rename(columns = {'Nefugees as a percentage of the international migrant stock(1995)'}, inplace = True)
df61.rename(columns = {'Unnamed: 12':'Refugees as a percentage of the international migrant stock(2000)'}, inplace = True)
df61.rename(columns = {'Unnamed: 13':'Refugees as a percentage of the international migrant stock(2000)'}, inplace = True)
df61.rename(columns = {'Unnamed: 14':'Refugees as a percentage of the international migrant stock(2000)'}, inplace = True)
df61.rename(columns = {'Unnamed: 15':'Refugees as a percentage of the international migrant stock(2010)'}, inplace = True)
df61.rename(columns = {'Unnamed: 16':'Refugees as a percentage of the international migrant stock(2010)'}, inplace = True)

df61.rename(columns = {'Unnamed: 16':'Refugees as a percentage of the international migrant stock(2010)'}, inplace = True)

df61.rename(columns = {'Unnamed: 18':'Annual rate of change of the refugee stock(1990-1995)'}, inplace = True)

df61.rename(columns = {'Unnamed: 199':'Annual rate of change of the refugee stock(1990-2005)'}, inplace = True)

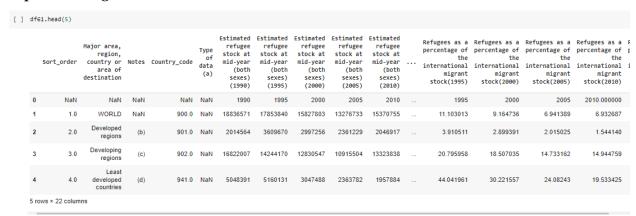
df61.rename(columns = {'Unnamed: 199':'Annual rate of change of the refugee stock(2010-2015)'}, inplace = True)

df61.rename(columns = {'Unnamed: 29':'Annual rate of change of the refugee stock(2010-2015)'}, inplace = True)

df61.rename(columns = {'Unnamed: 29':'
```

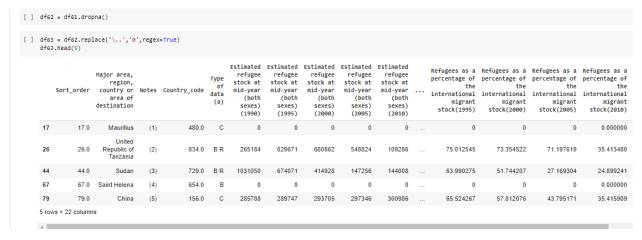
Renaming of the columns has been done

Step 4: Reading the data



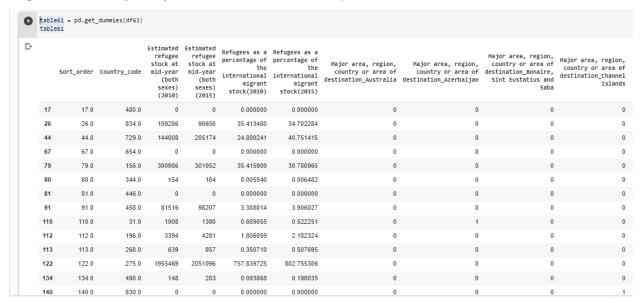
The data are read using .head()

Step 5: Dropping null values and unwanted value



Null values are dropped using .dropna() and ".." value are replaced with 0 using regex() function

Step 6: Converting categorical variable into dummy



Dummy data is created using .get_dummies() to converting categorical variable into dummy

Step 7: Checking the data frame information

```
[ ] df63.info()
     <class 'pandas.core.frame.DataFrame'>
     Int64Index: 22 entries, 17 to 240
     Data columns (total 22 columns):
                                                                             Non-Null Count Dtype
      # Column
     Ø Sort order
                                                                             22 non-null
                                                                                              float64
         Major area, region, country or area of destination
                                                                             22 non-null
                                                                                             object
         Notes
                                                                             22 non-null
                                                                                             object
         Country_code
                                                                                              float64
         Type of data (a)
                                                                             22 non-null
         Estimated refugee stock at mid-year (both sexes)(1990)
                                                                             22 non-null
                                                                                             object
         Estimated refugee stock at mid-year (both sexes)(1995)
                                                                             22 non-null
                                                                                             object
         Estimated refugee stock at mid-year (both sexes)(2000)
                                                                             22 non-null
                                                                                             object
         Estimated refugee stock at mid-year (both sexes)(2005)
                                                                                              object
         Estimated refugee stock at mid-year (both sexes)(2010)
                                                                             22 non-null
                                                                                              int64
      10 Estimated refugee stock at mid-year (both sexes)(2015)
                                                                             22 non-null
                                                                                              int64
      11 Refugees as a percentage of the international migrant stock(1990) 22 non-null
                                                                                             object
      12 Refugees as a percentage of the international migrant stock(1995) 22 non-null
                                                                                             object
         Refugees as a percentage of the international migrant stock(2000)
                                                                             22 non-null
                                                                                             object
      14
         Refugees as a percentage of the international migrant stock(2005)
                                                                             22 non-null
      15 Refugees as a percentage of the international migrant stock(2010)
                                                                             22 non-null
                                                                                              float64
         Refugees as a percentage of the international migrant stock(2015) 22 non-null
                                                                                              float64
      16
      17 Annual rate of change of the refugee stock(1990-1995)
                                                                             22 non-null
                                                                                             object
      18 Annual rate of change of the refugee stock(1995-2000)
                                                                                             object
      19 Annual rate of change of the refugee stock(2000-2005)
                                                                             22 non-null
                                                                                             object
      20 Annual rate of change of the refugee stock(2005-2010)
                                                                             22 non-null
                                                                                             object
     21 Annual rate of change of the refugee stock(2010-2015)
                                                                             22 non-null
                                                                                             object
     dtypes: float64(4), int64(2), object(16)
memory usage: 4.0+ KB
```

The dataframe information is checked to understand the renaming of the columns

Step 8: Renaming the column name

```
bf63.rename(columns = {'Estimated refugee stock at mid-year (both sexes)(2010)':|'Estimated_refugee_stock_at_mid_year_both_sexes_2010'}, inplace = True)
[ ] df63.info()
      <class 'pandas.core.frame.DataFrame'>
      Int64Index: 22 entries, 17 to 240
      Data columns (total 22 columns):
       # Column
                                                                                                  Non-Null Count Dtype
       0 Sort order
                                                                                                  22 non-null
                                                                                                                        float64
            Major area, region, country or area of destination
                                                                                                   22 non-null
                                                                                                                        object
            Notes
                                                                                                   22 non-null
                                                                                                                        object
            Country_code
Type of data (a)
                                                                                                   22 non-null
                                                                                                                       object
            Estimated refugee stock at mid-year (both sexes)(1990)
Estimated refugee stock at mid-year (both sexes)(1995)
Estimated refugee stock at mid-year (both sexes)(2000)
                                                                                                   22 non-null
                                                                                                                        object
                                                                                                   22 non-null
            Estimated refugee stock at mid-year (both sexes)(2005)
                                                                                                   22 non-null
                                                                                                                        object
            Estimated_refugee_stock_at_mid_year_both_sexes_2010
Estimated_refugee_stock_at_mid_year_(both_sexes)(2015)
                                                                                                                       int64
int64
                                                                                                   22 non-null
            Refugees as a percentage of the international migrant stock(1990)
Refugees as a percentage of the international migrant stock(1995)
                                                                                                  22 non-null
                                                                                                                        object
       13
            Refugees as a percentage of the international migrant stock(2000)
                                                                                                  22 non-null
                                                                                                                        object
            Refugees as a percentage of the international migrant stock(2005)
            Refugees as a percentage of the international migrant stock(2010)
       15
                                                                                                  22 non-null
                                                                                                                        float64
            Refugees as a percentage of the international migrant stock(2015)
Annual rate of change of the refugee stock(1990-1995)
                                                                                                  22 non-null
                                                                                                                        object
            Annual rate of change of the refugee stock(1995-2000)
            Annual rate of change of the refugee stock(2000-2005)
                                                                                                  22 non-null
                                                                                                                        object
            Annual rate of change of the refugee stock(2005-2010)
Annual rate of change of the refugee stock(2010-2015)
                                                                                                  22 non-null
22 non-null
      dtypes: float64(4), int64(2), object(16)
memory usage: 4.0+ KB
```

The column name is renamed and its information is checked.

Step 9: Creating a series of 10 value of the column and converting it to numeric

A series has been created using 10 values of the columns and its values are converted to numeric to get accurate results with the data frame.

Step 10: Top level evaluation

```
[ ] print(df63.eval("Sort_order + Estimated_refugee_stock_at_mid_year_both_sexes_2010"))
               17.0
           109312.0
    26
    44
           144052.0
               67.0
    67
    79
           301065.0
    80
              234.0
    81
               81.0
    91
            81607.0
    110
             2018.0
    112
             3506.0
              752.0
    122
         1955591.0
    134
              282.0
    140
              140.0
    144
              8868.0
    150
            40410.0
    160
              160.0
    166
            73774.0
    168
             3988.0
    187
              187.0
    195
              195.0
             25745.0
    240
    dtype: float64
```

.eval() is used to get the top level evaluation of the dataset.

Step 11: Manipulating the data



The data is manipulated using the .melt() function in which Unpivot a DataFrame from wide to long format, optionally leaving identifiers are set along with reading the dataframe.

Cleaning of Annex

Step 1: Importing Annex table by skipping 14 rows and reading.



Annex table is imported by skipping 14 rows and it is read using the .head() function.

Step 2: Checking the data information

```
df71.info()
r. <class 'pandas.core.frame.DataFrame'>
   RangeIndex: 232 entries, 0 to 231
   Data columns (total 12 columns):
    # Column
                           Non-Null Count Dtype
                             -----
    0 Country code
                            232 non-null int64
    1 Country or area
                            232 non-null object
    2 Sort order
                            232 non-null int64
    3 Major area
                            232 non-null object
                            232 non-null int64
    4 Code
    5 Sort order.1
                           232 non-null int64
    6 Region
                            232 non-null object
    7 Code.1
                            232 non-null int64
    8 Sort order.2
    8 Sort order.2 232 non-null int64
9 Developed region 232 non-null object
    10 Least developed country 232 non-null object
    11 Sub-Saharan Africa 232 non-null object
   dtypes: int64(6), object(6)
   memory usage: 21.9+ KB
```

Dataframe information is checked to understand the columns to be renamed

Step 3: Renaming the column name

```
df71.rename(columns = {'Country code':'Country_code'}, inplace = True)
df71.rename(columns = {'Sort order':'Sort_order'}, inplace = True)
df71.rename(columns = {'Sort order.1':'Sort_order_1'}, inplace = True)
df71.rename(columns = {'Code.1':'Code_1'}, inplace = True)
df71.rename(columns = {'Sort order.2':'Sort_order_2'}, inplace = True)
```

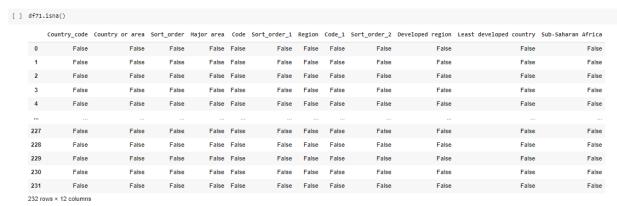
Renaming of the columns has been done to get accurate insights of the columns.

Step 4: checking the data information

```
[ ] df71.info()
    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 232 entries, 0 to 231
    Data columns (total 12 columns):
     # Column
                               Non-Null Count Dtype
     0
        Country_code
                                232 non-null
                                             int64
                                            object
     1
        Country or area
                               232 non-null
                                             int64
         Sort order
     2
                               232 non-null
                                             object
         Major area
                                232 non-null
     4
        Code
                                232 non-null
                                              int64
     5 Sort_order_1
                               232 non-null
                                              int64
     6 Region
                                232 non-null
                                              object
        Code_1
                               232 non-null
                                              int64
     8 Sort_order_2
                               232 non-null
                                              int64
       Developed region
     9
                                232 non-null
                                              object
     10 Least developed country 232 non-null
                                              object
     11 Sub-Saharan Africa
                                232 non-null
                                              object
    dtypes: int64(6), object(6)
    memory usage: 21.9+ KB
```

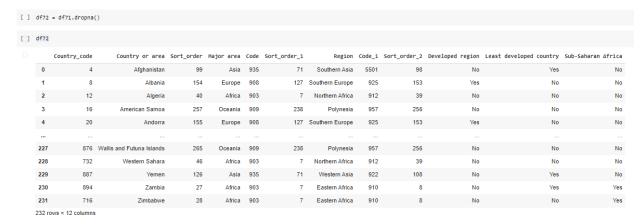
Information of the data frame has been checked in this stage.

Step 5: Checking null values in the data frame



Null value in the dataset has been checked so that it can be dropped.

Step 6: Dropping null values from the data frame and reading the data



Null value in the data frame has been dropped so that the accuracy of the data can be increased.

Step 7: Replacing unwanted value with zero using regex function



The ".." value has been replaced with 0 using the regex function to get accurate data in the dataframe.

Step 8: To get top level evaluation

```
[ ] print(df73.eval("Country_code + Code_1"))
     1
             933
     2
             924
     3
             973
            1833
     228
            1644
     229
            1809
     230
            1804
     231
            1626
     Length: 232, dtype: int64
```

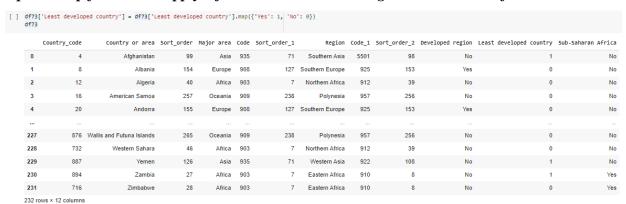
Top level evaluation is done in the system using .eval() function

Step 9: Data manipulation



Data manipulation is done using the melt function.

Step 10: Map function to apply a function to each item to get iterable value of the data



Map function is used to input a list and to get iterable value of the data

Step 11: Getting unique value based on hash table and print it

This step consists of getting the union value based on the hash table and getting the value of it.