Details

1. Reading Data (10 pts)

Unlike Assignment 3, you should write code to read in the data from the `persons_of_concern.csv` file directly into a pandas DataFrame. Do not use the code from Assignment 3, but instead use pandas to read in the data. If you see a `DtypeWarning`, you may safely ignore it; we will fix this later.

Hints:

- Don't forget to import pandas and numpy (remember, if you want to use the shorthand (`pd` and `np`), you need to use the `import ... as ...` syntax).
- Check that the DataFrame looks correct by displaying it. If the variable `df` references the DataFrame, just typing `df` in a cell will display the table.

2. Rename Columns (10 pts)

Again, the column names are too long to do interactive exploration. Let's fix them via the same remapping as in Assignment 3:

```
{'Asylum-seekers (pending cases)': 'asylum',
 'Country / territory of asylum/residence': 'country',
 'Internally displaced persons (IDPs)': 'idps',
 'Origin': 'origin',
 'Others of concern': 'others',
 'Refugees (incl. refugee-like situations)': 'refugees',
 'Returned IDPs': 'ret_idps',
 'Returned refugees': 'ret_refugees',
 'Stateless persons': 'stateless',
 'Total Population': 'total',
 'Year': 'year'
}
```

You should use the `rename` method to accomplish this. You may find its documentation either by typing `?df.rename` (if your DataFrame is named `df`) or on the pandas website.

Hints:

- Remember, you want to convert the **column** names, not the row labels.
- Like most pandas operations, `rename` produces a **new** DataFrame by default rather than modifying the current DataFrame in place.

3. Update Values and Convert Column Types

a. Replace Asterisks and Missing Values. (10 pts)

Recall that the documentation for the dataset states:

> In the 2016 data, figures between 1 and 4 have been replaced with an asterisk (*). These represent situations where the figures are being kept confidential to protect the anonymity of individuals.

so we will overestimate by setting the value for any asterisk (`*`) to four (4). Similarly, we will replace any missing value (pandas has helpfully set all of these to `NaN`) to 0.

Hints:

- You can fill missing values using the `fillna` method.
- The `replace` method will work over the entire data frame.
- Remember to either update the DataFrame in place or create/update a reference to the updated DataFrame.

b. Convert Column Types (10 pts)

Now, we want to make sure that the columns match the data they contain. Specifically, we wish to update the columns

```
['asylum', 'idps', 'others','refugees', 'ret_idps','ret_refugees','stateless','total']
```