**Note:** Initially planned to perform this transformation using Databricks notebook activity. But due to exhausted quotas, I cannot perform the activity.

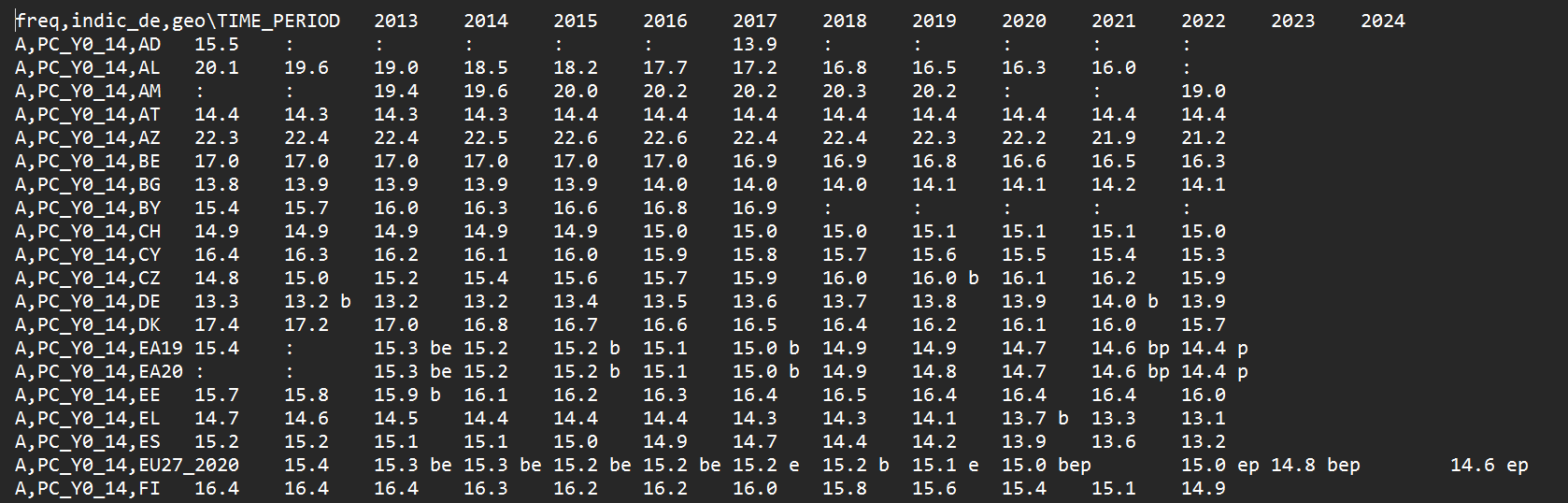
 

**Transform EU population data from official European Union website**

Data:

A screenshot of a computer

AI-generated content may be incorrect.



The file has been directly downloaded from official EU website containing latest population data till 2024.

**Requirement to transform the data:**

A screenshot of a computer

AI-generated content may be incorrect.

1. The column ‘freq,indic\_de,geo\TIME\_PERIOD’ is a complex type containing values by comma separated. So, we used derived column transformation to derive it as array format.
2. Then, we used Country LookUp table to add country code 2 digit and 3-digit values using LookUp transformation.
3. The above image may only show year 2019 to be pivoted, but we performed for all other years including 2020 to 2024.
4. The array ‘indic\_de’ then used to unpivot the years and later used conditional split data based on years.
5. The we pivoted the year data again to load each year separately based on the Year.

**LookUp data – Dim Country:**

A screenshot of a computer

AI-generated content may be incorrect.

**Data flow transformation:**

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

**Result Dataset:**

The transformed data are loaded in the Azure DataLake processed folder after successful pipeline on data flow activity.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.