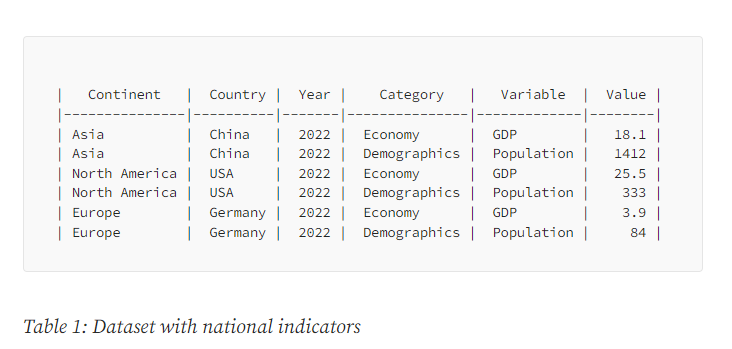
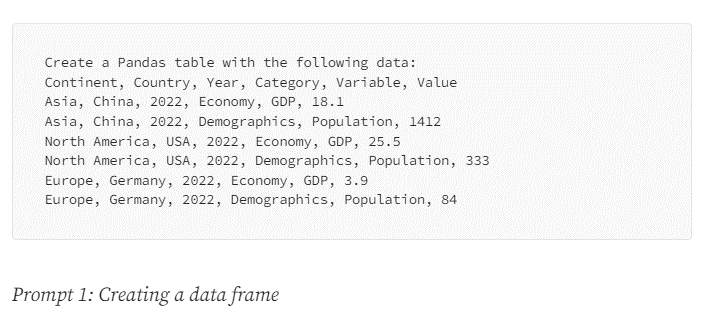
From Data Engineering to Prompt Engineering

1. **Creating a dataframe**

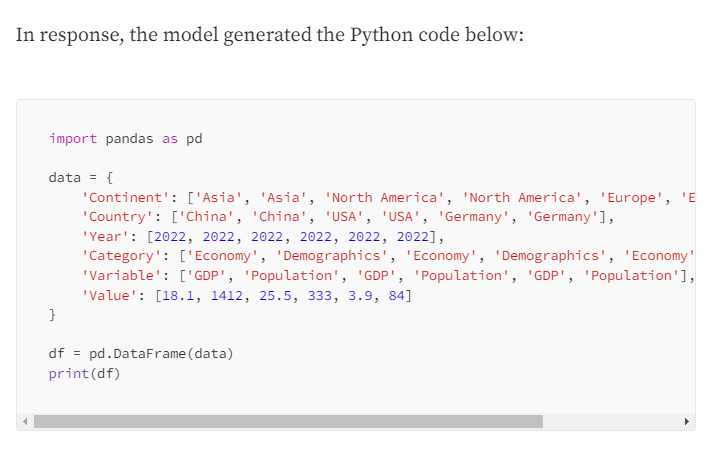


**Prompt for creating above dataframe in ChatGPT**

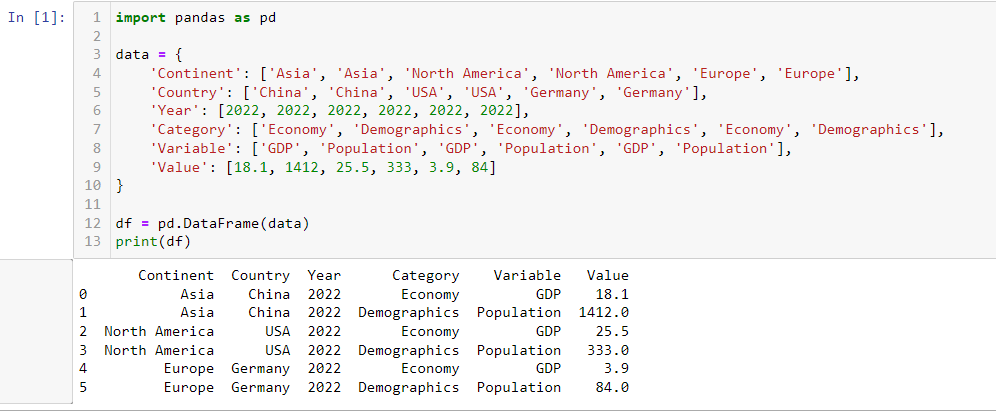


**Promt Code:**

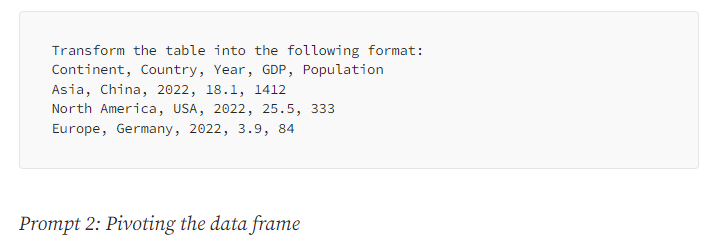
Create a Pandas table with the following data:  
Continent, Country, Year, Category, Variable, Value  
Asia, China, 2022, Economy, GDP, 18.1  
Asia, China, 2022, Demographics, Population, 1412  
North America, USA, 2022, Economy, GDP, 25.5  
North America, USA, 2022, Demographics, Population, 333  
Europe, Germany, 2022, Economy, GDP, 3.9  
Europe, Germany, 2022, Demographics, Population, 84



If we use the above code in jupyter notebook we get the required result:



1. **Pivoting the dataframe**



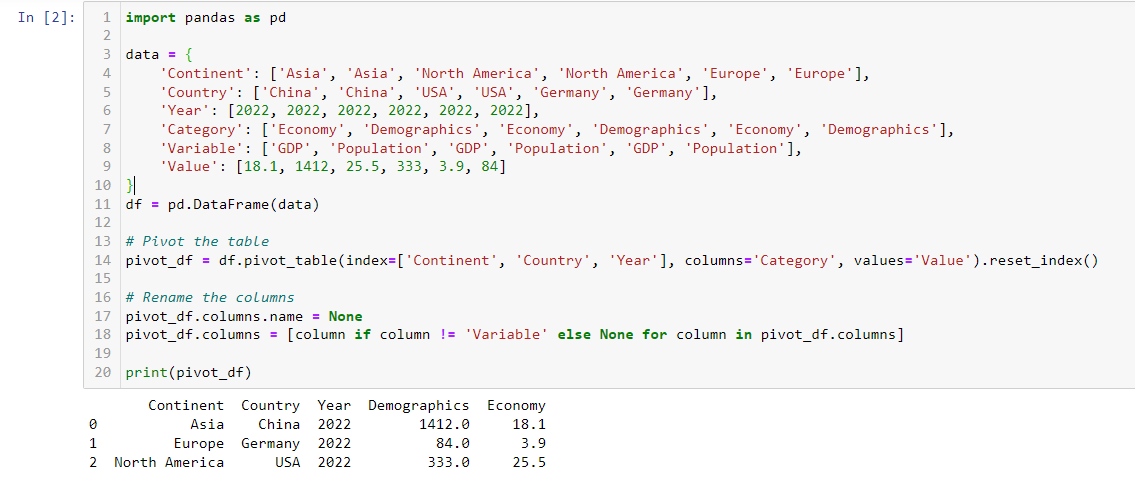
**Prompt code:**

Transform the table into the following format:  
Continent, Country, Year, GDP, Population  
Asia, China, 2022, 18.1, 1412  
North America, USA, 2022, 25.5, 333  
Europe, Germany, 2022, 3.9, 84

The code generated by this prompt:



The python output will look like:



But here we don’t get the desired output,

Instead of the expected columns “GDP” and “Population”, the table contains two columns named “Demographics” and “Economy”. To fix this problem, we posted the next prompt:

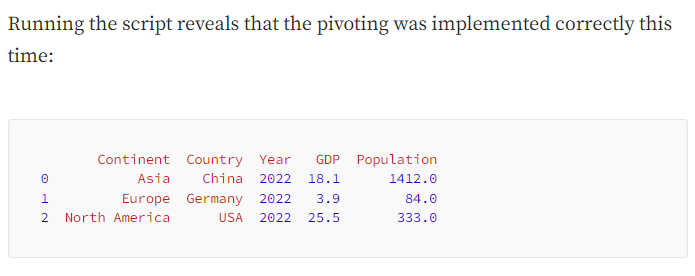
Prompt 3: to fix the pivoting code:

**Prompt code:**

The columns are not correct; you should have used

the variables GDP and population





1. **Enriching the dataframe**

To add country codes in the above dataset

**Prompt code:**

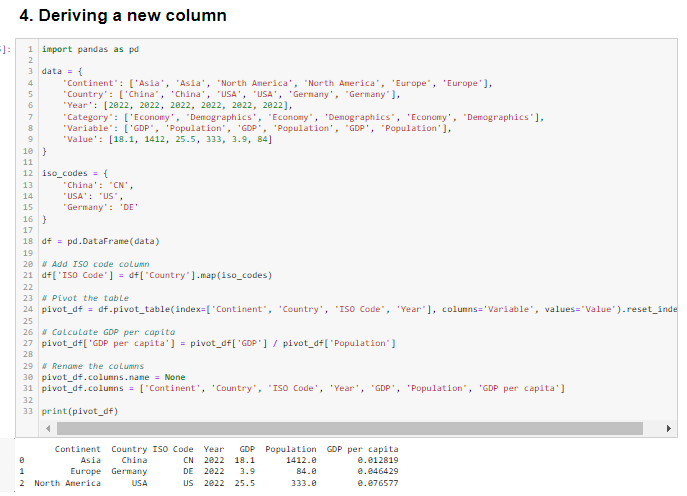
For each country, in the table above, add its ISO code



1. **Deriving a column**

**Prompt code:**

Derive a new column "GDP per capita" by dividing the GDP   
of a country by its population size



1. **Analysing the data**

Calculate the correlation coefficient between GDP & the population of the countries

Prompt:

Use the table above to calculate the correlation coefficient between   
the GDP and the population of the countries



Python code output for analysing the data:

