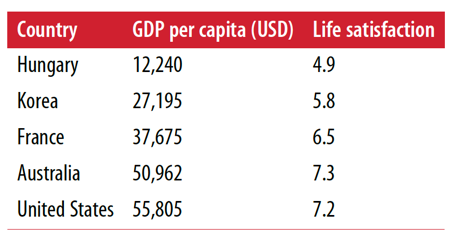
Your goal for this assignment is to explore whether there is a relationship between a country’s GDP (gross domestic product) and their reported life satisfaction index. The LSI is a number that represents how happy the average person reports feeling living in that country. You will use two files for this analysis: WEO\_Data.csv and Life Metrics.csv.

Your first step will be to analyze and clean up the data. At a minimum, you will need to join the two files into one file and begin exploring. The table below shows an example of what you might get after combining the two files and sorting by GDP per capita.



Next, create a plot that shows the (possible) relationship between the GDP and the LSI. A sample plot is show below (yours may look different). It suggests a definite relation.

A screenshot of a cell phone

Description automatically generated

Use the data analysis commands you have learned (*head*, *info*, etc.) to do some summary analysis. Are there any outliers? Are there any missing values? Develop a strategy to handle any anomalies.

For this exercise, you will not need to split your data into a training set and a test set. We will simply fit the model to all of our data and then use the resulting model to make a prediction about a country whose LSI is not known. For example, Cypriot has a per capita GDP around $27,000. What does your model predict its LSI will be? (You should get an answer somewhere around 5.9). Can you find two other countries whose GDP and LSI support this answer?