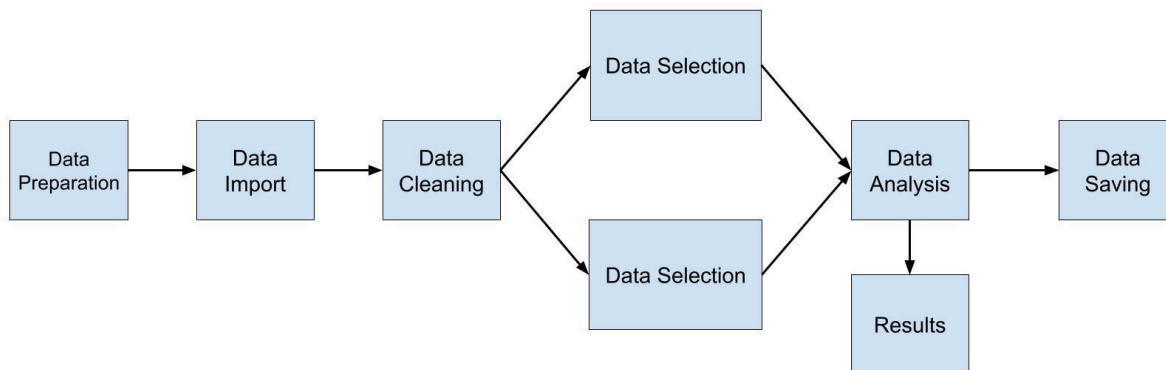1. A listing or drawing of the parts of the problem. What key operations and substeps need to happen in the program?

   Below are the key operations and substeps needed to happen in the program.
    1. Data preparation: Extract the data from the internet and save as a .csv file, (https://www.worldometers.info/gdp/gdp-by-country/)
    2. Data import: Read data from .csv file using 'pandas' library
    3. Data cleaning: Remove records with missing data
    4. Data selection: Extract top and bottom 5 countries with their respective GDP and GDP growth.
    5. Data analysis: Comparison of metrics for the top and bottom countries, including potential comparison plots
    6. Data analysis results: printout that may include plot(s) of GDP data analysis
    7. Data saving: Storing selected data of countries for further use as .csv files


2. A listing or drawing of the key input, intermediate, and output data that needs to pass through your program.

    Below are the key steps of data flow in the program.
    1. The data source, (https://www.worldometers.info/gdp/gdp-by-country/)
    2. Input dataset in the form of a .csv file
    3. 'pandas' dataframe with entire input data
    4. Cleaned dataset
    5. Datasets with selected data of top 5 countries
    6. Datasets with selected data of bottom 5 countries
    7. Combined datasets for top and bottom 5 countries
    8. .csv files with combined datasets
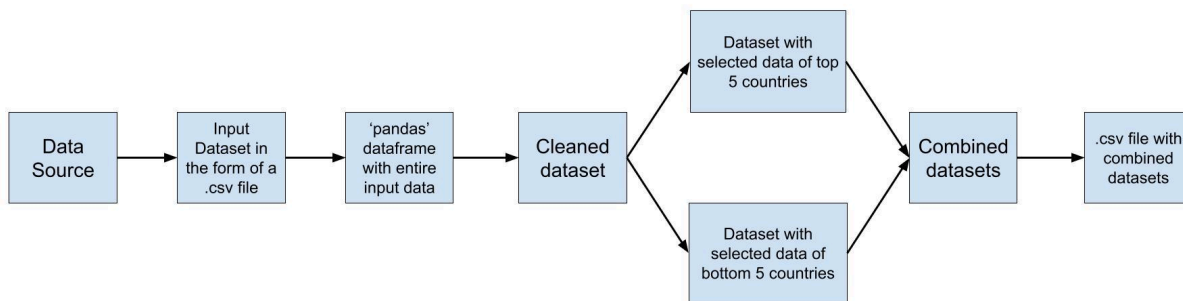
3. A drawing of how the parts of the problem relate logically, and what data is going in and out of each step.
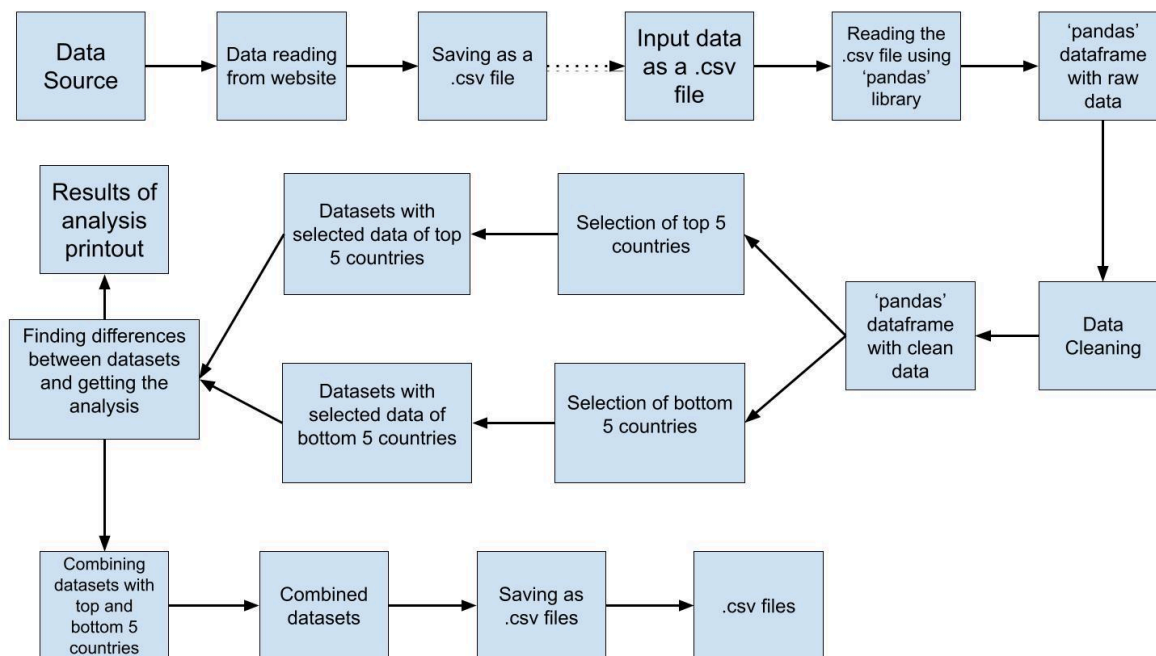
This project general concept is similar to problem idea #2. We have a dataset of countries (separated by their respective GDP's) (⊞ gdp_data). We would like to know how the GDP between countries varies across the world. More specifically, we want to know, **"Which 5**

countries have the largest GDP, and which 5 countries have the smallest GDP? And what are the gaps for these top and bottom 5 countries? What is the spread of GDP in those two groups? In addition, we want to find the 5 countries that have the largest GDP growth, and which 5 countries have the smallest GDP growth? And how do they compare? The analysis may also include plot(s) comparing the top and bottom 5 countries. GDP is defined in terms of total monetary value of all goods and services produced and sold by a country.

NOTE: the GDP values only represent one year (the year it represents is 2022). Also, the dataset, like many real-world datasets, *includes missing data*! For this analysis, we want to exclude countries that have missing data for their respective GDP's.

Your output should include:

1. *An answer to the final question, which can be shown in the outputs of one or more cells of your analysis notebook*
2. *A post-processed dataset in a new .csv file (named, e.g., `gdp_data.csv`) that we can use to directly explore different angles on the original question (e.g., top 10 countries with the largest GDP).*

Example analysis output that answers the question (what's important is the contents, formatting can be a bit different):

```
     #       Country GDP  (nominal, 2022)     GDP  (abbrev.) GDP growth  \
0  1  United States  $25,462,700,000,000  $25.463 trillion      2.06%
1  2          China  $17,963,200,000,000  $17.963 trillion      2.99%
2  3          Japan   $4,231,140,000,000   $4.231 trillion      1.03%
3  4        Germany   $4,072,190,000,000   $4.072 trillion      1.79%
4  5          India   $3,385,090,000,000   $3.385 trillion      7.00%

   Population  (2022) GDP per capita Share of  World GDP
0         338289857         $75,269            25.32%
1        1425887337         $12,598            17.86%
2         123951692         $34,135             4.21%
3          83369843         $48,845             4.05%
4        1417173173          $2,389             3.37%
```