This is the teamwork section of our project stage 1

Tasks to do

(Project Stage I)

Understanding the COVID-19 Dataset: Look at the COVID-19 data and make a list of the key variables (like number of cases, deaths, and population).

Create a data dictionary for each of these variables.

Merge COVID-19 Data: In Jupyter Notebook, load the COVID-19 data (cases, deaths, population). and Combine all this data into one big dataset.

Save the combined dataset as a CSV file.

make sure to have reports on each task/step etc

turn it in

TASK ONE

step one

Load the COVID-19 Data from the Teamwork Project Stage 1 Folder

In [37]:

```
import pandas as pd
import os

# path to the teamwork project stage 1 folder on saras desktop?
desktop_path = os.path.join(os.path.expanduser("~"), "Desktop")
teamwork_project_folder = os.path.join(desktop_path, "teamwork project stage

# loading confirmed cases, deaths, and population data
confirmed_cases_path = os.path.join(teamwork_project_folder, 'covid_confirme
deaths_data_path = os.path.join(teamwork_project_folder, 'covid_deaths_usafa
population_data_path = os.path.join(teamwork_project_folder, 'covid_county_p

# reading the data into pandas
confirmed_data = pd.read_csv(confirmed_cases_path)
deaths_data = pd.read_csv(deaths_data_path)
population_data = pd.read_csv(population_data_path)

# printing the first few rows of each dataset to understand and get comforta
print("confirmed cases data:")
print(confirmed_data.head())
print("\ndeaths data:")
print(deaths_data.head())

print("\npopulation data:")
print(population_data.head())
```

```
confirmed cases data:
   countyFIPS            County Name State  StateFIPS  2020-01-22  2020-01-2
3  \
0           0  Statewide Unallocated    AL          1           0
0
1        1001         Autauga County    AL          1           0
0
2        1003         Baldwin County    AL          1           0
0
3        1005         Barbour County    AL          1           0
0
4        1007            Bibb County    AL          1           0
0

   2020-01-24  2020-01-25  2020-01-26  2020-01-27  ...  2023-07-14  \
0           0           0           0           0  ...           0
1           0           0           0           0  ...       19913
2           0           0           0           0  ...       70521
3           0           0           0           0  ...        7582
4           0           0           0           0  ...        8149

   2023-07-15  2023-07-16  2023-07-17  2023-07-18  2023-07-19  2023-07-20  \
0           0           0           0           0           0           0
1       19913       19913       19913       19913       19913       19913
2       70521       70521       70521       70521       70521       70521
3        7582        7582        7582        7582        7582        7582
4        8149        8149        8149        8149        8149        8149

   2023-07-21  2023-07-22  2023-07-23
0           0           0           0
1       19913       19913       19913
2       70521       70521       70521
3        7582        7582        7582
4        8149        8149        8149

[5 rows x 1269 columns]

deaths data:
   countyFIPS            County Name State  StateFIPS  2020-01-22  2020-01-2
3  \
0           0  Statewide Unallocated    AL          1           0
0
1        1001         Autauga County    AL          1           0
0
2        1003         Baldwin County    AL          1           0
0
3        1005         Barbour County    AL          1           0
0
4        1007            Bibb County    AL          1           0
0

   2020-01-24  2020-01-25  2020-01-26  2020-01-27  ...  2023-07-14  \
0           0           0           0           0  ...           0
1           0           0           0           0  ...         235
2           0           0           0           0  ...         731
3           0           0           0           0  ...         104
```

```
4            0            0            0            0  ...          111

   2023-07-15  2023-07-16  2023-07-17  2023-07-18  2023-07-19  2023-07-20  \
0           0           0           0           0           0           0   
1         235         235         235         235         235         235   
2         731         731         731         731         731         731   
3         104         104         104         104         104         104   
4         111         111         111         111         111         111   

   2023-07-21  2023-07-22  2023-07-23  
0           0           0           0  
1         235         235         235  
2         731         731         731  
3         104         104         104  
4         111         111         111  

[5 rows x 1269 columns]

population data:
   countyFIPS            County Name State  population
0           0  Statewide Unallocated    AL           0
1        1001         Autauga County    AL       55869
2        1003         Baldwin County    AL      223234
3        1005         Barbour County    AL       24686
4        1007            Bibb County    AL       22394
```

step 2

Create a Data Dictionary for key variables

In [7]:

```
# data dictionary for confirmed cases, deaths, and population datasets
data_dictionary = {
    "countyFIPS": {
        "Data Type": "int64",
        "Description": "A unique identifier for each county."
    },
    "County Name": {
        "Data Type": "object (string)",
        "Description": "The name of the county."
    },
    "State": {
        "Data Type": "object (string)",
        "Description": "The state where the county is located."
    },
    "Population": {
        "Data Type": "int64",
        "Description": "The total population of the county (in the populatic
    },
    "Date Columns": {
        "Data Type": "int64 (for confirmed cases and deaths)",
        "Description": "Daily data for the confirmed cases or deaths startir
    }
}
```

```
# displaying the data dictionary
for key, value in data_dictionary.items():
    print(f"Variable: {key}")
    print(f"  Data Type: {value['Data Type']}")
    print(f"  Description: {value['Description']}\n")
```

```
Variable: countyFIPS
  Data Type: int64
  Description: A unique identifier for each county.

Variable: County Name
  Data Type: object (string)
  Description: The name of the county.

Variable: State
  Data Type: object (string)
  Description: The state where the county is located.

Variable: Population
  Data Type: int64
  Description: The total population of the county (in the population datase
t).

Variable: Date Columns
  Data Type: int64 (for confirmed cases and deaths)
  Description: Daily data for the confirmed cases or deaths starting from th
e earliest date in the dataset.
```

For Task 1: Step 1 & 2, I began by inspecting the confirmed cases, deaths, and population datasets thats located in the teamwork project folder. These datasets were loaded into pandas and the key columns were reviewed. The datasets contained information for various counties across the United States, by daily covid 19 case counts, deaths, and population

Key Columns that were found countyFIPS: A unique identifier for each county County Name: The name of the county in each state. State: The abbreviation of the U.S. state where the county is located. StateFIPS: A numerical code representing each state. Population: The total population of each county Date Columns: Each dataset contains daily covid 19 records (for confirmed cases and deaths) across date columns, starting from early 2020.

TASK 2 load confirmed cases, deaths, and population datasets

In [62]:

```
import pandas as pd
import os

# path to the teamwork project stage 1 folder on sara's desktop
desktop_path = os.path.join(os.path.expanduser("~"), "Desktop")
teamwork_project_folder = os.path.join(desktop_path, "teamwork project stage
# loading confirmed cases, deaths, and population data again
confirmed_cases_path = os.path.join(teamwork_project_folder, 'covid_confirme
deaths_data_path = os.path.join(teamwork_project_folder, 'covid_deaths_usafa
population_data_path = os.path.join(teamwork_project_folder, 'covid_county_p

# reading
confirmed_data = pd.read_csv(confirmed_cases_path)
deaths_data = pd.read_csv(deaths_data_path)
population_data = pd.read_csv(population_data_path)

# printing the first few rows of each dataset to confirm they are correct ar
print("\nconfirmed cases data:")
print(confirmed_data.head())

print("\ndeaths data:")
print(deaths_data.head())

print("\npopulation data:")
print(population_data.head())

# merging confirmed cases and deaths data on countyFIPS, County Name, and St
merged_data = pd.merge(confirmed_data, deaths_data, on=['countyFIPS', 'Count

# merging with population data on countyFIPS
final_merged_data = pd.merge(merged_data, population_data[['countyFIPS', 'po

# filtering out rows where countyFIPS is 0 (Statewide Unallocated data) beca
filtered_data = final_merged_data[final_merged_data['countyFIPS'] != 0]

# printing the first few rows of the filtered data to confirm its there
print("\nFiltered Merged Data (without Statewide Unallocated):")
print(filtered_data.head())

# saving the filtered data to a CSV file that we named final_merged_data.csv
filtered_output_path = os.path.join(teamwork_project_folder, 'final_merged_c
filtered_data.to_csv(filtered_output_path, index=False)

print(f"The filtered merged dataset is saved as {filtered_output_path}")
```

```
confirmed cases data:
   countyFIPS            County Name State  StateFIPS  2020-01-22  2020-01-2
3  \
0           0  Statewide Unallocated    AL          1           0
0
1        1001         Autauga County    AL          1           0
0
2        1003         Baldwin County    AL          1           0
0
3        1005         Barbour County    AL          1           0
0
4        1007            Bibb County    AL          1           0
0

   2020-01-24  2020-01-25  2020-01-26  2020-01-27  ...  2023-07-14  \
0           0           0           0           0  ...           0
1           0           0           0           0  ...       19913
2           0           0           0           0  ...       70521
3           0           0           0           0  ...        7582
4           0           0           0           0  ...        8149

   2023-07-15  2023-07-16  2023-07-17  2023-07-18  2023-07-19  2023-07-20  \
0           0           0           0           0           0           0
1       19913       19913       19913       19913       19913       19913
2       70521       70521       70521       70521       70521       70521
3        7582        7582        7582        7582        7582        7582
4        8149        8149        8149        8149        8149        8149

   2023-07-21  2023-07-22  2023-07-23
0           0           0           0
1       19913       19913       19913
2       70521       70521       70521
3        7582        7582        7582
4        8149        8149        8149

[5 rows x 1269 columns]

deaths data:
   countyFIPS            County Name State  StateFIPS  2020-01-22  2020-01-2
3  \
0           0  Statewide Unallocated    AL          1           0
0
1        1001         Autauga County    AL          1           0
0
2        1003         Baldwin County    AL          1           0
0
3        1005         Barbour County    AL          1           0
0
4        1007            Bibb County    AL          1           0
0

   2020-01-24  2020-01-25  2020-01-26  2020-01-27  ...  2023-07-14  \
0           0           0           0           0  ...           0
1           0           0           0           0  ...         235
2           0           0           0           0  ...         731
3           0           0           0           0  ...         104
```

```
4           0           0           0           0  ...         111

   2023-07-15  2023-07-16  2023-07-17  2023-07-18  2023-07-19  2023-07-20  \
0           0           0           0           0           0           0
1         235         235         235         235         235         235
2         731         731         731         731         731         731
3         104         104         104         104         104         104
4         111         111         111         111         111         111

   2023-07-21  2023-07-22  2023-07-23
0           0           0           0
1         235         235         235
2         731         731         731
3         104         104         104
4         111         111         111

[5 rows x 1269 columns]

population data:
   countyFIPS            County Name State  population
0           0  Statewide Unallocated    AL           0
1        1001         Autauga County    AL       55869
2        1003         Baldwin County    AL      223234
3        1005         Barbour County    AL       24686
4        1007            Bibb County    AL       22394

Filtered Merged Data (without Statewide Unallocated):
    countyFIPS      County Name State  StateFIPS_cases  2020-01-22_cases  \
51        1001  Autauga County     AL                1                0
52        1003  Baldwin County     AL                1                0
53        1005  Barbour County     AL                1                0
54        1007     Bibb County     AL                1                0
55        1009   Blount County     AL                1                0

    2020-01-23_cases  2020-01-24_cases  2020-01-25_cases  2020-01-26_cases
\
51                 0                 0                 0                 0
52                 0                 0                 0                 0
53                 0                 0                 0                 0
54                 0                 0                 0                 0
55                 0                 0                 0                 0

    2020-01-27_cases  ...  2023-07-15_deaths  2023-07-16_deaths  \
51                 0  ...                235                235
52                 0  ...                731                731
53                 0  ...                104                104
54                 0  ...                111                111
55                 0  ...                261                261

    2023-07-17_deaths  2023-07-18_deaths  2023-07-19_deaths  \
51                235                235                235
52                731                731                731
53                104                104                104
54                111                111                111
55                261                261                261
```

```
    2023-07-20_deaths  2023-07-21_deaths  2023-07-22_deaths  \
51                235                235                235   
52                731                731                731   
53                104                104                104   
54                111                111                111   
55                261                261                261   

    2023-07-23_deaths  population  
51                235       55869  
52                731      223234  
53                104       24686  
54                111       22394  
55                261       57826  

[5 rows x 2536 columns]
The filtered merged dataset is saved as /Users/saraabukhalaf/Desktop/teamwor
k project stage 1/covid 19 data/final_merged_data.csv
```

TASK 2

The next and final steps will involve merging the three datasets (confirmed cases, deaths, and population) using the countyFIPS column. this will create one unified covid dataset, so we can analyze how the virus spread in different counties, and consider each county's population.

For merging we used countyFIPS as the key to combine the data from the confirmed cases, deaths, and population datasets. this will help us get a view of the impact of covid.After merging we saved it as a csv file

Findings: We noticed that many counties had zero cases and deaths in the early months of 2020. This makes sense because covid 19 hadn’t spread widely in the U.S. during that time, especially in rural counties. From January to early March 2020, there were very few reported cases. It wasn’t until April 2020 that we started seeing a noticeable increase in confirmed cases especially in larger counties.

```
In [ ]:
```