team_stage_1

September 13, 2024

0.1 COVID-19 Dataset (Cases Confirmed)

Name	Description	Data Type	Possible Values
countyFIPS	Federal Information Processing Standards for the county	Integer	1001 - 56045
County Name	Name of the county in the state	String	"[name] County"
State	State abbreviation	String	"GA", "NC" for the 50 states
StateFIPS	Federal Information Processing Standards for the state	Integer	1 – 56
Date	Date in mm-dd-yyyy	Date	1/22/2020 - 7/23/2023
Confirmed Cases	Number of confirmed cases	Integer	0 - 3534088

0.2 COVID-19 Dataset (County Populations)

Name	Description	Data Type	Possible Values
countyFIPS	Federal Information Processing Standards for the county	Integer	1001 – 56045 or 0
County Name	Name of the county	String	"[name] County" or "Statewide Unallocated"
State	State abbreviation	String	"GA", "NC" for the 50 states
Population	Number of people in county	Integer	0-10039107

0.3 COVID-19 Dataset (Deaths)

Name	Description	Data Type	Possible Values
countyFIPS	Federal Information Processing Standards for the county	Integer	1001 - 56045

Name	Description	Data Type	Possible Values
County Name	Name of the county in the state	String	"[name] County"
State	State abbreviation	String	"GA", "NC" for the 50 states
StateFIPS	Federal Information Processing Standards for the state	Integer	1 - 56
Date	Date in mm-dd-yyyy	Date	1/22/2020 - 7/23/2023
Deaths	Number of deaths	Integer	0 - 52102

[3]:
```python
import pandas as pd
df_county = pd.read_csv('covid_county_population_usafacts.csv')
df_deaths = pd.read_csv('covid_deaths_usafacts.csv')
df_confirmed = pd.read_csv('covid_confirmed_usafacts.csv')
```

[7]:
```python
df_super = pd.merge(df_confirmed, df_deaths, on = ['countyFIPS', 'County Name',
                                                    'State', 'StateFIPS'])
```

Merging Covid confirmed cases with Covid deaths data on the columns of countyFIPS, County Name, State, StateFIPS

[9]:
```python
df_super = pd.merge(df_county, df_super, on = ['countyFIPS', 'State'])
```

Merging the Covid cases plus deaths data with county population database

[11]:
```python
df_super.to_csv('covid_super.csv', index=False)
```

Save the merged super dataset to csv file, without an index column

[13]:
```python
print(pd.read_csv('covid_super.csv')) # Output the result
```

```
      countyFIPS          County Name_x 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
...          ...                    ...   ...          ...
3188       56037      Sweetwater County    WY        42343
3189       56039           Teton County    WY        23464
3190       56041           Uinta County    WY        20226
3191       56043        Washakie County    WY         7805
3192       56045          Weston County    WY         6927

              County Name_y  StateFIPS  2020-01-22_x  2020-01-23_x  \
0     Statewide Unallocated          1             0             0
1            Autauga County          1             0             0
```

```
2          Baldwin County       1           0           0
3          Barbour County       1           0           0
4             Bibb County       1           0           0
...                    ...    ...         ...         ...
3188    Sweetwater County      56           0           0
3189         Teton County      56           0           0
3190         Uinta County      56           0           0
3191      Washakie County      56           0           0
3192        Weston County      56           0           0

      2020-01-24_x  2020-01-25_x  …  2023-07-14_y  2023-07-15_y  \
0                0             0  …             0             0
1                0             0  …           235           235
2                0             0  …           731           731
3                0             0  …           104           104
4                0             0  …           111           111
...            ...           ... ..  …           ...           ...
3188             0             0  …           142           142
3189             0             0  …            16            16
3190             0             0  …            43            43
3191             0             0  …            51            51
3192             0             0  …            24            24

      2023-07-16_y  2023-07-17_y  2023-07-18_y  2023-07-19_y  2023-07-20_y  \
0                0             0             0             0             0
1              235           235           235           235           235
2              731           731           731           731           731
3              104           104           104           104           104
4              111           111           111           111           111
...            ...           ...           ...           ...           ...
3188           142           142           142           142           142
3189            16            16            16            16            16
3190            43            43            43            43            43
3191            51            51            51            51            51
3192            24            24            24            24            24

      2023-07-21_y  2023-07-22_y  2023-07-23_y
0                0             0             0
1              235           235           235
2              731           731           731
3              104           104           104
4              111           111           111
...            ...           ...           ...
3188           142           142           142
3189            16            16            16
3190            43            43            43
3191            51            51            51
3192            24            24            24
```

[3193 rows x 2536 columns]