

# Data Description on Employment Data

This dataset seems to contain detailed economic and employment data, broken down by industry, geography (state and county), and time (quarterly). It can be used for various types of economic analysis, including assessing employment trends, wage levels, and industry performance in different regions of the United States.

## Dataset Details:

In [102]:

employment

Out[102]:

tr	Area Type	St Name	Area	...	Industry	Status Code	Establishment Count	October Employment	November Employment	December Employment	Total Quarterly Wages	Average Weekly Wage	Employment Location Quotient Relative to U.S.	Total Wage Location Quotient Relative to U.S.
4	County	Alabama	Autauga County, Alabama	...	10 Total, all industries	NaN	936	10870	11008	10958	126028952	886	1.00	1.00
4	County	Alabama	Autauga County, Alabama	...	10 Total, all industries	NaN	11	92	86	85	1449696	1272	0.38	0.41
4	County	Alabama	Autauga County, Alabama	...	10 Total, all industries	NaN	19	562	563	563	8035294	1099	1.58	2.02
4	County	Alabama	Autauga County, Alabama	...	10 Total, all industries	NaN	32	1876	1923	1883	20537266	834	1.76	1.95
4	County	Alabama	Autauga County, Alabama	...	10 Total, all industries	NaN	874	8340	8436	8427	96006696	879	0.91	0.89
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4	County	Wyoming	Unknown Or Undefined, Wyoming	...	1024 Professional and business services	NaN	182	927	963	1076	13284245	1034	2.38	1.28
4	County	Wyoming	Unknown Or Undefined, Wyoming	...	1025 Education and health services	NaN	33	133	132	129	1620357	949	0.26	0.22
4	County	Wyoming	Unknown Or Undefined, Wyoming	...	1026 Leisure and hospitality	N	2	0	0	0	0	0	0.00	0.00
4	County	Wyoming	Unknown Or Undefined, Wyoming	...	1027 Other services	N	26	0	0	0	0	0	0.00	0.00
4	County	Wyoming	Unknown Or Undefined, Wyoming	...	1029 Unclassified	N	1	0	0	0	0	0	0.00	0.00

## Dtypes of the Data

In [171]:

print(employment.dtypes)

Area\nCode	int64
St	object
Cnty	float64
Own	int64
NAICS	int64
Year	int64
Qtr	int64
Area Type	object
St Name	object
Area	object
Ownership	object
Industry	object
Status Code	object
Establishment Count	int64
October Employment	int64
November Employment	int64
December Employment	int64
Total Quarterly Wages	int64
Average Weekly Wage	int64
Employment Location Quotient Relative to U.S.	float64
Total Wage Location Quotient Relative to U.S.	float64
dtype:	object

## Details about Column names:

**Area Code:** This column contains unique codes or identifiers for different geographical areas. Each code represents a specific region or area of interest. For example, "US000" likely

represents the entire United States as a whole, while other codes may represent individual states, counties, or metropolitan areas.

**St:** The "St" column contains state abbreviations, which are standardized two-letter codes used to represent U.S. states. "US" typically represents the entire United States as a single entity.

**Cnty:** The "Cnty" column contains county codes, which are numerical identifiers for individual counties within a state. These codes help to uniquely identify and categorize different counties.

**Own:** This column likely represents ownership types of business establishments. Ownership types can include private, public, or other classifications based on the legal structure and ownership of businesses.

**NAICS:** The "NAICS" code stands for the North American Industry Classification System. It's a standardized coding system used to categorize businesses and industries based on their primary economic activities. Each code corresponds to a specific industry or sector. In this dataset, "10" represents the total of all industries, indicating a broad overview.

**Year:** This column indicates the year to which the data pertains. It's a critical dimension for tracking economic changes and trends over time. In this dataset, the year is 2020.

**Qtr:** The "Qtr" column represents the quarter of the year during which the data was collected. It divides the year into four distinct time periods (quarters). In this case, "4" represents the fourth quarter of the year.

**Area Type:** This column specifies the type of geographical area being referred to in the dataset. In this case, "Nation" indicates that the data represents the entire United States as a nation.

**St Name:** The "St Name" column typically contains the full names of U.S. states corresponding to the state abbreviations in the "St" column. However, in this dataset, it seems to contain missing or undefined values ("NaN").

**Area:** The "Area" column contains the name of the geographical area being analyzed. In this dataset, it is "U.S. TOTAL," which signifies the entire United States as a whole.

**Industry:** This column provides a textual description of the industry or sector being analyzed. In this dataset, "10 Total, all industries" describes the total of all industries, indicating a comprehensive overview.

**Status Code:** The "Status Code" column may contain additional information or codes related to the data. In this dataset, it appears to contain missing or undefined values ("NaN").

**Establishment Count:** This column represents the count of business establishments within the specified area. It helps quantify the number of businesses operating in a given region.

**October Employment, November Employment, December Employment:** These columns contain employment data, specifically the number of employed individuals in the specified area for the respective months (October, November, and December).

**Total Quarterly Wages:** This column provides the total wages paid in the specified area for the given quarter. It reflects the overall income generated within the area during that time frame.

**Average Weekly Wage:** The "Average Weekly Wage" column represents the average weekly wage earned by workers in the specified area. It serves as a measure of income for employees within the region.

**Employment Location Quotient Relative to U.S.:** This column offers a relative measure of employment concentration within the specified area compared to the entire United States. A value of 1.0 indicates that employment in the area is in line with the national average. Values greater than 1.0 suggest over-employment relative to the national average, while values less than 1.0 indicate under-employment.

**Total Wage Location Quotient Relative to U.S.:** Similar to the previous column, this measure assesses how total wages in the specified area compare to the national average. A value of 1.0 signifies that the area's total wages are in line with the national average.

### Merging Covid super data with Employment:

#### Step1: Reading the Files

```
import pandas as pd

# Read the Excel file into a DataFrame
excel_df = pd.read_excel('allh1cn204.xlsx')

# Save the DataFrame as a CSV file
excel_df.to_csv('emp.csv', index=False)
```

My dataset contains many Area Type like nation, county to map with covid, I chosen only county type.

```
employment = employment.loc[employment['Area Type'] ==
'County']
```

Here two column names were different, but the values are same, so I used left\_on and right\_on, what I observed means countyFIPS has same values with the Area code

```
emp_merge = pd.merge(Supercovid_df, employment, left_on =
'countyFIPS', right_on = 'Area\nCode')
```

