Viveka Reddy Erram, Graduate Student

Employment Dataset-

Description on the Employment enrichment dataset:

This data set (Employment dataset) provides the information about the level of employment, ownership and establishment count by county level in every state in the United States for the month of January, February and March of the year 2021.By merging this dataset with COVID-19 dataset we can analyze and observe how the employment is affecting covid-19 cases based on the type of Industry, ownership, average weekly wages. Employment plays a key role while calculating the Covid-19 cases. Based on different categories of industrial employment also we see the increase and decrease in the covid-19 cases.

Variables and Data type table:

Name	Data type	Description
Area	object	Shows state name and area name
Area\nCode	int64	Unique ID for each county
Area Type	object	Has area in county
Year	int64	2021
Own	int64	Unique code for ownership
Ownership	object	This shows the employment details of Federal, state, local government or private sectors.
Industry	object	To see the employment details of each industry (Information, Manufacturing, trade)
State Name	object	Shows the name of the state
Establishment Count	int64	Number of people employed in each establishment.
January Employment	int64	Number of people who were employed in the month of January 2021.

February Employment	int64	Number of people who were employed in the month of February 2021.
March Employment	int64	Number of people who were employed in the month of March 2021.
Total Quarterly Wages	int64	Total quarterly wages in a county based on each industry.
Average Weekly Wage	int64	Average weekly wage of the county based on industry.
Employment Location Quotient Relative to U.S.	Float64	Employment Location quotient compared to the total country.
Total Wage Location Quotient to US	Float64	Wage location quotient compared to the total country.

Merging the employment data set with the COVID-19 dataset:

To merge the enrichment dataset with the COVID-19 data, first process the employment data set. Here we need to remove the inconsistencies in the Area Code column. Select the columns which are useful for the analysis of the data. Then I have merged the employment data set using 'Area Code' variable with the 'countyFIPS' variable in the large covid data set to get Employee covid dataset. I grabbed a sample random set of 500 rows from the data frame and converted it to a csv file.

Hypothesis Questions:

I have used the employment dataset and the population of each county to compute the employment rate in each county. This would assist me in determining the employment rate and the link between Covid cases and mortality. Then, based on the deaths/cases, I'd be able to determine which counties, industries, or ownerships were severely affected and which were not.

- Unemployed people in various counties may have a higher number of COVID reports?
- Due to federal reporting requirements, counties with government control may have more reports of COVID -19 cases than counties with private ownership?
- There might be higher number of COVID -19 cases in the Counties which have trade and transportation as employment?
- If there are more medical centers in the areas with high employment rates, that may lead to fewer deaths due to COVID-19?

- Since they can work from home, business management-related industries may have fewer COVID-19 cases?

Analysis on COVID-19 Data set:

Firstly, Calculate the COVID-19 cases confirmed trend for the state "Alabama" for last week which is 2021-08-10 to 2021-08-16. The below graph shows the data visualization of COVID-19 cases trend for the state Alabama.
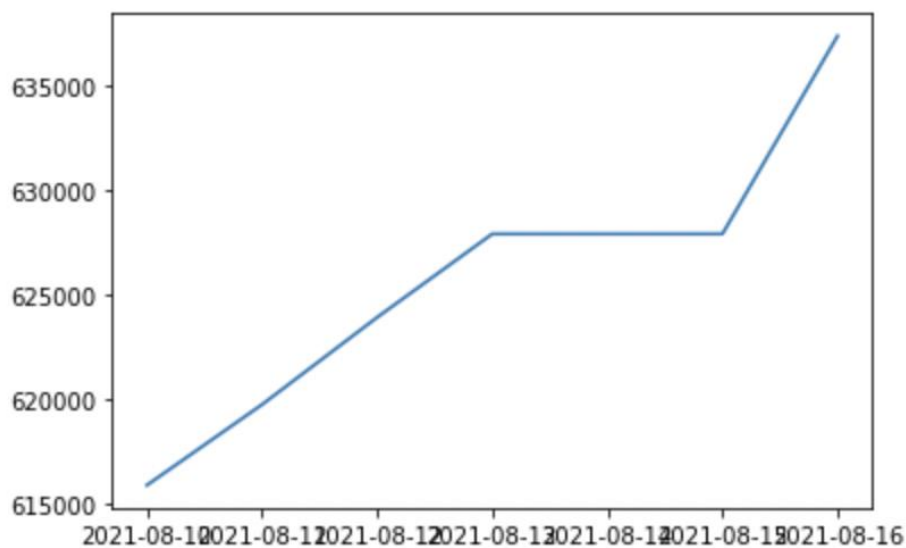
By analyzing the above graph, we can say that the deaths have constantly increased from the first day (2021-08-10) to 2021-08-13 later the number of deaths are stable from 2021-08-13 to 2021-08-15, from there we can see deaths has rocketed. So, we can say that the death cases have increased.