**DATA SCIENCE**

**PROJECT ( STAGE - 1 )**

**By**

***Praneeth Aluru***

**ENRICHMENT DATA**

**Employment Dataset**

**DATA DESCRIPTION :**

The enrichment data which I’ve chosen here is “Employment Dataset”. This data is the “Quarterly Census of Employment and Wages” data which is uploaded by “U.S. BUREAU OF LABOR STATISTICS”. This website shows about the employment in a year for every month with respect to every state where every state is divided into multiple counties. Apart from that, the dataset contains the information about the “Total Quaterly wages” and “Average weekly wage” for every state.

The data is available up to June, 2021. As, covid started from the year 2020, we will take the data from the year 2020 and is available according to quarterly basis of the year. That is, every file has the data of every 3 months. For the year 2021, the data is available up to the month of June. So, we take the data from January 2020 to June 2021 and show the analysis. Not only the columns above, but also other columns are present and the description of every field is in the below table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr No** | **Field Name** | **Field Description** | **Data**  **Type** |
| 1) | Year | 4-digit year | Int 64 |
| 2) | Qtr | 1-character quarter (always A for annual) | Int 64 |
| 3) | Area type | Category of the given area. (State, County,Nation, etc...) | String |
| 4) | St Name | Multi-character State name | String |
| 5) | Area | Area title associated with the area's FIPS code. | String |
| 6) | Ownership | Ownership title associated with the ownership code. | String |
| 7) | Industry | Industry title associated with the industry code. | String |
| 8) | Status Code | Status code, or disclosure code ('N' for not disclosed) | Object |
| 9) | Establishment  Count | Quarterly establishment counts for a given quarter | Int 64 |
| 10) | Employement  (In every Month) | Employment level for all the months. | Int 64 |
| 11) | Total  Quaterly  Wages | Total quarterly wage level for the given quarter. | Int 64 |
| 12) | Average weekly  wage | Average weekly wage based on the 12-monthly employment levels and total annual wage levels | Int64 |
| 13) | Employment Location Quotient Relative to U.S. | Employment Location Quotient Relative to U.S | Float 64 |
| 14) | Total Wage Location Quotient Relative to U.S. | Total Wage Location Quotient Relative to U.S. | Float 64 |

**HOW CAN YOU MERGE THE DATA WITH THE PRIMARY COVID-19 DATASET. IDENTIFY THE INDIVIDUAL VARIABLE WHICH MAP BETWEEN THE DATASETS.**

There are three files for Covid-19 dataset and those files data have been combined into one data frame where we named it as “Super Covid” data. We combine that data with enrichment data for the analysis of affect of covid on employment. For that, first we have to make the enrichment data accordingly with super covid data.

For that, I added extra columns for super covid data. They are year and month, as these columns are present in the enrichment dataset too. I transformed the months employment columns of enrichment data into rows by which we get employment numbers month wise for every state and their counties. After that, I merged both the data with Year, Month and State columns. I considered only Employment numbers to see the effect on it by covid data. Finally, I will get the year, month, state, cases, death and employment fields in a final Dataframe.

**DESCRIBE HOW YOUR ENRICHMENT DATA CAN HELP IN THE ANALYSIS OF COVID-19 SPREAD.**

The employment data contains the employment numbers pertaining to every state and their counties. We can analyse how the employment is affected by rise and fall of covid cases for every state. Also vice-versa, we can analyse the spread of covid, based on the rise and fall of employment numbers.

**HYPOTHESIS QUESTIONS**

How varying number of covid cases affects the employment ?

How varying number of covid deaths affects the employment ?

Does more the number of covid cases means there are several deaths ?

Are all the states employments affected by covid ?