**CSC405-605-Spring2023-Project:**

**Introduction:**

Coronavirus is a continuing worldwide pandemic, which has affected a lot of people including you. Goal of the project is to develop an analytical framework to study the data coming from United States to understand patterns of COVID-19 effect and spread. We have utilized the data from usfacts.org. The dataset contains daily county-level tracker of COVID-19 cases. This makes it easy to follow COVID-19 cases on a granular level, as does the ability to break down infections per 100,000 people (with the population data).

**Data Dictionary:**

|  |  |  |
| --- | --- | --- |
| Variable Name | Datatype | Description |
| Employment Status-Employed (Above 16 years) | Object | Employed people whose age are above 16 years |
| Employment Status-Employed(Female) | Object | Employed females |
| Commuting to work with Public transport | Object | People commuting to work with public transport |
| Working from home | Object | People who are working from home |
| Occupation-Sales and office occupations | Object | People whose occupation is related to sales and office |
| Industry-Educational services, and health care and social assistance | Object | People whose occupation is related to education, health care or social services |
| Class of worker-Government workers | Object | People who are government employee |
| Unpaid family workers | Object | Family workers who are not paid |
| Income-$200,000 or more | Object | People with income greater than $200,000 |
| With private health insurance | Object | People who own a private health care insurance |
| No health insurance coverage | Object | People who has no health insurance |
| 18 to 64 years (People whose income in past 12 years is below poverty level) | Object | People whose income are below poverty level in past 12 years |

**Enrichment Datasets:**

For enrichment dataset we have used data from data.census.gov. It covers 100,000+ different geographies like states, counties, places, tribal areas, etc., and for each topics like education, employment, health, and housing are covered.

Enrichment data DP03 is selected for Economic characteristics. In this dataset stats of all the counties in US States along with their geographic ID. Count of population is given based on their employment status, occupation, commuting to work, industry they work, class of worker, income health insurance coverage.

Enrichment dataset contains Geography ID whose last 5 digits are same as CountyFIPS in Covid-19 dataset. Also, the name of state and county are same as Covid-19 dataset which can be used for mapping.

Enrichment dataset must help to analyze below questions –

1. If number of covid cases and deaths hugely vary between the people who are employed or not employed?
2. What kind of commute for employed workers is more prone to covid like taking public transport?
3. What kinds of occupations has most impact of covid like may be sales occupation where people meet several other people can have more cases?
4. Are covid cases and deaths found more in some industry like education or health care?
5. If worker is associated with government job how the impact of covid is seen.
6. Does income of the house matters when covid related deaths and cases are considered?
7. To know the impact of health insurance and age on covid deaths and cases.