**DSC 540 – Project Milestone 1.**

Part of the DSC540 Project, I would like to access data, clean and merge data sets to related to Covid-19 data containing Infections information by County/State, Summary of infection information and vaccination. I am identifying these data sets part of the current milestone 1 (identified below).

Once the data is cleansed from CSV (Milestone2), Website(Milestone3) and API sources(Milestone4), I would like to merge this data part of Milestone 5 and produce visualizations that would give insight on serious infection outcomes and correlation with general vaccination trends in particular states of US.

**CSV File:** This is a comprehensive COVID infection details data collected by local city, county and state officials on each infection and the serious outcome by case to CDC.

This data set contains detailed information on case occurrence month, location, age group, Demographic information (sex, race, ethnicity), exposure and the serious outcome of infection(hospitalization, ICU admittance and death) and if the patient already has underlying condition.

Data set is location: <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4>

Csv format to be downloaded ~28M records, part of Milestone 2 – this data will be cleansed and structured.

**Website**: contains current infection summary by each state and trend for that day, including the data of total infections, new infections, death summary, infection recovery summary, total test summary and active cases summary.

Data location: <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4> - Data contains one row per state, territories, and other special cases like US military.

Part of milestone 3, this data will cleansed and transformed.

**API**: This is an API that generates data in CSV or JSON formats. The data can be generated and downloaded at state or county level with an apiKey generated using an email id.

Data set contains information on population/positivity rate, case density, infection rate, icu capacity, risk levels, infection rates, hospital bed availability/usage, vaccination distribution and delivery information for covid at state/county level.

Location: <https://apidocs.covidactnow.org/#register>

API link : [https://api.covidactnow.org/v2/states.timeseries.csv?apiKey=**YOUR\_KEY\_HERE**](https://api.covidactnow.org/v2/states.timeseries.csv?apiKey=YOUR_KEY_HERE)

Part of milestone4, this data will be pulled from API , and I will cleanse and transform it.

At the end during milestone5 activity, I will merge all the cleansed data and provide visualizations on various factors that could help in ascertaining possible trends causing serious outcomes due to COVID infections, and as well potential efficacy of overall vaccination program in reducing the Coronavirus infections.