**Project Proposal**

The outbreak of COVID-19 has become a global threat till now. Many researches and studies have been conducted where they have extracted humongous data on daily bases. In the upcoming project assignment we are going to extract the dataset from the New York Times ([dataset](https://github.com/nytimes/covid-19-data)). The dataset we are using consists of all the data which was published to provide a daily cumulative number of cases and deaths reported in each county and state across the U.S. We will clean the raw data by using pandas and then convert the cleaned/wanted data(sub data) into database tables by using COPY statement/query.

In our database we will create tables with attributes - number of cases, number of deaths, confirmed cases and confirmed deaths on daily bases. Also we will use the college dataset where it consists of the data related to state, city, county and college ID. The high level idea in using the above dataset is to create a sophisticated database which can be used to demonstrate most of the SQL features like joints, triggers, procedures, functions, aggregations etc,.

By -

[pendyala@pdx.edu](mailto:pendyala@pdx.edu)

[thoutam@pdx.edu](mailto:thoutam@pdx.edu)