**Applied Data Science Capstone ( Week# 4):- The-Battle-of Neighborhoods**

# Author: - ARUN KUMAR Date: - 12/21/2020

# Year-End is approaching. Year 2020 was very rough, stressful, and most difficult year for common public worldwide. 2020 has been heavily defined by the COVID 19 pandemic, which has led to global social and economic devastation, worldwide lockdowns, and the largest economic recession since Great Depression of the 1930s.

# Introduction: -

# Coronavirus ( COVID19 ) virus hits the United State of America (USA) extremely severely compare to rest of the world. The United States is struggling to keep the virus under control as new cases are reported each day. On December 17, 2020, USA surpassed 17.0 million official COVID 19 confirmed cases. COVID virus transmission is extremely fast and dangerously high in several big states such as California, Texas, Florida, Illinois, New York. I am resident of Texas state. Therefore, I decided to work on COVID 19 confirmed cases data set for Texas State. Texas is the second largest state in the United State by Area and Population. Nearly 29.9 million people live in Texas. Texas became the 28th state in the US in 1845. The largest cities in Texas are Houston, Dallas, Fort Worth, San Antonia, El Paso, and Austin. Austin is the Capital city of Texas.

# Project Objective:-

# COVID 19 analysis and visualization of confirmed cases for all the county (254) in Texas.

# Investigation of 5 most impacted county and Transmission prediction and cause.

# Characterize risk factor based on the confirmed cases

# Recommendation

# Dataset: - confirmed COVID 19 data set will be downloaded from [Coronavirus Disease 2019 (COVID-19) (texas.gov)](https://www.dshs.texas.gov/coronavirus/?gclid=CjwKCAiArIH_BRB2EiwALfbH1DTENV8kdWsdleFabix9f7swHSje2yLpNF1kfnV4hxDYzeq2fV1HgBoCAToQAvD_BwE) website.

**Methodology**

* Data will be downloaded from the website. If required, data will be Web Scraper with Python using the Beautiful Soup Library from  [Coronavirus Disease 2019 (COVID-19) (texas.gov)](https://www.dshs.texas.gov/coronavirus/?gclid=CjwKCAiArIH_BRB2EiwALfbH1DTENV8kdWsdleFabix9f7swHSje2yLpNF1kfnV4hxDYzeq2fV1HgBoCAToQAvD_BwE).
* Data cleaning, sorting, and preprocessing in Python.
* Supervised and unsupervised Machine Learning Algorithm application on clean data for prediction and clustering techniques.
* Finally, the data will be visually assessed using graphing from Python libraries.

**Problem Statement: -**

1. Identification of at least 5 most and least impacted counties ( Graphs and Maps\_ in Texas State.
2. Investigation of the factor associated with infection transmission ( population density, commercial/residential property etc ) in these counties.