**Data 602 – Final Project Proposal**

**By: Mubashira Qari**

**Project Title:** **Impact of Covid-19 on Student’s Learning Modalities (Year 2021-2022) & Prediction Using Supervised Machine Learning**

**Describing Dataset:** The first dataset is the weekly summary of school learning modalities. It has 923K row and 9 columns.

The second dataset is the US county level covid-19 cases and deaths counts. It has 2249807 rows and 6 columns.

These datasets require detailed data cleaning before performing the data manipulation steps.

The metrics for the columns in the dataset are as below:

Text

Description automatically generated with medium confidence

**Data Source Link:**

The datasets are acquired from the following sources,

* HealthData.gov
* NYTimes Covid-19 data – GitHub
* US Zip code to County State to FIPS Look Up – data.world
* States Names and Abbreviations - GitHub

and the links are provided below:

<https://healthdata.gov/National/School-Learning-Modalities/aitj-yx37>

<https://github.com/nytimes/covid-19-data>

<https://data.world/niccolley/us-zipcode-to-county-state>

<https://github.com/jasonong/List-of-US-States/blob/master/states.csv>

**Justification for Dataset Selection:**

The reason for choosing these datasets is that I am interested in finding out the impact of covid-19 on the learning modalities for the students.

My focus will be to investigate the state level trend for the hybrid, remote and in-person learning due to covid for the year 2021 and 2022.

**Research Questions & Objectives:**

The research questions for this project are the following:

1- Which state has the most covid cases for the year 2021 and 2022?

2- Which state has the most deaths due to covid for the year 2021 and 2022?

3- Which state has the highest average student count for hybrid, remote and in-person learning modality for the year 2021 and 2022?

7- Clean and merge the datasets.

8- Upload the dataset in the PostgreSQL database for further analysis.

9- Prediction of learning modalities using machine learning.

**Libraries, Visualization Apps & Database Used for Project Implementation:**

* Python Pandas
* Python NumPy
* Python Matplotlib
* PostgreSQL Database
* Python sklearn
* Python seaborn
* Plotly Dash App
* Tableau App
* SQLAlchemy python SQL Toolkit and Object Relational Mapper

**EDA and Summary Statistics:**

Below are the images of exploratory data analysis and summary statistics:

Text

Description automatically generated

Text

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

