ADS-507 – Team Project Status Update Form

Fill out this form and submit it by the end of Module 4 in Blackboard.

Team Number: 6

Team Leader/Representative: Dingyi Duan

Full names of team members:

1. Dingyi Duan

2. Emina Belekanich

3. Ryan Dunn

Title of your Production-Ready Data Pipeline project: Covid-19 Post-Vaccination Monitoring System

Short description of your project and objectives:

This pipeline will start with data from California Open Data Portal and store the data in Amazon S3 (Simple Storage Service) bucket, then uses Amazon Glue to extract, transform and load (ETL) the data using AWS Glue's GUI; Then we can either store the structured data back in AWS S3 or locally; Lastly we will use MySQL or MS SQL to query the data for the end users (California Department of Public Health (CDPH)) to be able to identify vaccination status of COVID-19 cases, hospitalizations, and deaths.

Name of your selected dataset: **COVID-19 Post Vaccination Statewide Stats**

Description of your selected dataset (data source, format, size of dataset, etc.):

The dataset is taken from California Department of Public Health in CSV format; Contains 350 rows and 25 columns; The type of data involved are timestamp, text and integer. The columns cover mainly the date, cases, deaths, hospitalizations and population of three types of patients of age 12 and above that are unvaccinated, vaccinated and boosted.

Please provide the link for your GitHub repository here:
https://github.com/dingyiduan7/ADS507-Final_Project

How many times have your members met in the last two weeks?   2

What was the agreed upon method of communication? Are you using any teamwork project management software, such as Trello or Asana?   Email, Slack, Zoom and Google Docs.

Comments/Roadblocks:

- Need to figure out if AWS Redshift is needed for building the pipeline since our dataset is small;
- All three of us are full-time working professionals and first-timers for data engineering, so we need to make sure we allocate adequate enough time for the project;
- It can be challenging to get the data lively updated.