**Research proposal**

**Due date: Tues 8th February 2022**

**Project title: ETL project 2.**

**Team members:**

Jade, Yao, Antoinette, Stella

**Datasets:**

[**https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions/version/4?select=2021VAERSVAX.csv**](https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions/version/4?select=2021VAERSVAX.csv)

**2021 VaersData.csv**

**2021 Vaerssymptoms.csv**

**2021 Vaersvax.csv**

**Project description/proposal:**

For this project, we will be looking into the adverse reactions to Covid-19 vaccines reported to the vaccine adverse event reporting system (VAERS) in the United States in 2021.

We will be working with 3 csv files that hold different information and our aim is to clean, transform and merge all 3 files before loading the final product into a SQL database. A relational database structure is more suitable because of the already existing columns in the csv files and the information stored is already ordered.

Below is a breakdown on how we will be transforming the files:

**2021 VaersData.csv**

Drop columns:

CAGE\_YR, CAGE\_MO, RPT\_DATE, SYMPTOM\_TEXT, DIED, DATEDIED,

**2021 Vaerssymptoms.csv**

Drop columns:

SYMPTOM-VERSION 1-5

**2021 Vaersvax.csv**

Drop columns:

VAX\_LOT, VAX\_NAME,