**Project Report**

* Purpose: By joining the vaccination and mobility dataset, we are able to view the relationship between vaccination and the active level of certain places, such as parks or retail stores, in each States.
* **E**xtract:
  + Vaccinations.csv: data source is from Kaggle, formatted as csv file.
  + 2021\_US\_Region\_Mobility\_Report.csv: data source is from google formatted as csv file.
* **T**ransform:
  + Vaccinations.csv: cleaned and specify data to only the date, location, people vaccinated, and people fully vaccinated columns, then dropped the null values, and renamed location into states to merge with other datasets.
  + 2021\_US\_Region\_Mobility\_Report.csv: data cleaned by keeping data that only has the state data by dropping all the nulls in sub\_region\_1(states), and keeping all the datas that are null in sub\_region\_2 (county). Then reset indexes with sub\_region\_1, date, and all the percentage changes in different industry to merge with other datasets.
* **L**oad: we loaded both cleaned datasets into pgAdmin as a relational database for people to easily select categories they want to compare with