Dang Khoa Do

INST414

**Final Project**

1. **Topic:**

For the final project topic, because COVID-19 is not a topic that people are concerned a lot about in this period, I want to analyze the COVID-19 vaccine distribution in the USA from 2021 to 2022 to see the trend of the vaccination distribution and the real situation of the vaccination based on the vaccination rate in the U.S.

1. **Dataset:**

* Link for the dataset: <https://ourworldindata.org/us-states-vaccinations>.
* The dataset is from kaggle.com which represents the vaccine for COVID-19 that is available in the USA. Also, the dataset also provides state-by-state data on COVID-19 vaccinations in the U.S. The dataset has 16 columns and 42981 rows.
* The dataset is updated daily, so it is up-to-date and the data is recent. The source of data is reliable and transparent.

1. **Questions:**

These are some questions that the result of the data analysis will attempt to answer to have insight into the vaccination campaign in the United States, so we can see the people’s concern about the COVID-19 pandemic.

* How many people are fully vaccinated per the vaccination distributed?
* How is the different vaccination distribution between states?
* Which state has the highest and lowest vaccination rate?
* How will people attempt to get the vaccine in the future?
* What is the vaccination attempt different between 2021 and 2022?

1. **Data Analysis Process:**

* Clean the dataset
* Analyze data using tools such as python and R
* Visualization of the dataset using tools such as Python, Power BI
* Prediction future trends of vaccination using Python and Power BI