Project Proposal Template:

Submitted by: Arjun Singh Pathania – axp220145

1. Project Description: This project aims to analyse the progress of COVID-19 vaccination in California by measuring the percentage of people who have been vaccinated, as well as identifying the remaining population that needs to be vaccinated. The analysis will also focus on the rate at which vaccination progress is occurring to determine if adjustments need to be made to ensure that all people are vaccinated in a timely manner..

2. Objective: Our objective is to understand the distribution of the COVID-19 vaccine doses across various counties and to study the impact of the vaccine on reducing the number of COVID-19 cases and fatalities. We aim to identify any disparities in vaccine distribution and to create a visual representation of the data to raise awareness and encourage further action.

3. Hypothesis:

* In the state of California during 2021, Pfizer vaccination has been preferred by 17.5 times than Johnson & Johnson while 1.5 times than Moderna because of availability.
* LA county of the state of California has the highest vaccination as compared to other counties of the state because of higher population density in LA county, this county has approximately 1/3rd of the state.
* In the state of California, in 2022, the mortality rate has gone down by 0.004% (0.007(2021 death)-0.003(2022 death)) because of the 72% higher vaccine administered in 2021.
* In the county of San Diego, the numbers of COVID-19 cases increased by 41% with the decrease in vaccination rate over the years 2021 and 2022 when compared to previous year 2020.
* In California, counties with higher Covid-19 cases are administering booster and bivalent dosage at a higher rate when compared to counties with lower case rates.

4. Data Sources: This Dataset summarizes vaccination data at the county level by county of residence. Where county of residence was not reported in a vaccination record, the county of provider that vaccinated the resident is included. This applies to less than 1% of vaccination records. The sum of county-level vaccinations does not equal state-wide total vaccinations due to out-of-state residents vaccinated in California. The dataset has the following attributes

* County- County of residence for vaccine recipient.
* Administered Date- Date on which the vaccine was administered.
* Total Doses- Total number of vaccines administered on a particular administration date. Includes both single and two-dose vaccines.
* Cumulative total doses- Cumulative number of vaccines administered up to a particular administration date. Includes both single and two-dose vaccines.
* Pfizer Doses- Pfizer doses administered on a particular administration date.
* Cumulative Pfizer Doses- Cumulative Pfizer doses administered up to that date.
* Moderna Doses- Moderna doses administered on a particular administration date.
* Cumulative Moderna Doses- Cumulative Moderna doses administered up to that date.
* Johnson & Johnson Doses- Johnson & Johnson doses administered on a particular administration date.
* Cumulative Johnson & Johnson Doses- Cumulative Johnson & Johnson doses administered up to that date.
* Partially Vaccinated- The number of people who became partially vaccinated on a particular administration date.
* Total Partially vaccinated- Total number of all partially vaccinated people for a particular administration date.
* Fully Vaccinated- Total number of people who became fully vaccinated.
* Cumulative Fully Vaccinated- Cumulative number of fully vaccinated people up to that date.
* At Least one dose- Total number of people who got their first dose on a particular administration date.
* Cumulative At Least one dose- Cumulative number of people with at least one dose on that date.
* Booster Recipient count- Total number of individuals who have completed a primary series of an approved or authorized COVID-19 vaccine and who have received another dose.
* Cumulative Booster Recipient count- Cumulative total number of individuals who have completed a primary series of an approved or authorized COVID-19 vaccine and who have received another dose.
* Booster Eligible Population- Total number of individuals eligible to receive a booster or additional dose based on CDC recommendations.
* Bivalent Booster Recipient Count- Total number of individuals who have completed a primary series of an approved or authorized COVID-19 vaccine and who have received a bivalent booster dose.
* Cumulative Bivalent Booster Recipient Count- Cumulative total number of individuals who have completed a primary series of an approved or authorized COVID-19 vaccine and who have received a bivalent booster dose.
* Bivalent Booster Eligible Population- Total number of individuals eligible to receive a bivalent booster based on CDC recommendations
* [https://data.chhs.ca.gov/dataset/e283ee5a-cf18-4f20-a92c-ee94a2866ccd/resource/d995e0fa-aa50-43e3-9753-94e9bf1df12f/download/covid-19-progress-dashboard-data-dictionary\_](https://data.chhs.ca.gov/dataset/e283ee5a-cf18-4f20-a92c-ee94a2866ccd/resource/d995e0fa-aa50-43e3-9753-94e9bf1df12f/download/covid-19-progress-dashboard-data-dictionary_jan2023.xlsx)

5. Number of Records: There are 54810 records in the dataset.

6. Data Cleansing Tool: We will use R and Excel as data cleansing tool because R has simple functions for examining and cleaning dirty data and Excel has many features to get data in precise format.

7. Visualization Tool: We will use Tableau as a visualization tool because this application helps make Big data small and, small data insightful and actionable. This software provides the most efficient way to change or transform the raw data into an easily understandable format.