**1. Project Description:** *“Covid-19 Vaccines Administered By Demographics By County By Age”*

The COVID-19 vaccinations are separated among a range of age segments whose vaccination status provides insight about population trends in the State of California. The data is primarily sorted through age segments and the completeness of the vaccination. For example, the age segments are 5 years and older, and 12 years and older. The categories of completeness of the vaccination are partially vaccinated, fully vaccinated, booster, at least one dose, and unvaccinated.

With this in mind, our group will consider 2 counties (Alameda and San Diego) to exhibit the key trends, progress and differences in each county. The demographics of age and race/ethnicity suggests the distinct patterns in their vaccination status. This can predict potential avenues to improve the percent of vaccinated population in CA, particularly Alameda and San Diego counties, and decrease the spread of the COVID-19 disease.

**2.Objective: Why are you visualizing this?**

The visualization of the COVID-19 vaccination status and its relation to age depicts groups that need to be targeted, or in other words, age groups that are lacking in a particular vaccination category. This can be noted by visually seeing the key differences among the two counties. With the help of these visualizations, we can take a deeper look into the driving forces for these numbers. This can be crucial for decision making when encountered with a similar contagious disease. Moreover, the research of data and visual representations convey the motive of the COVID-19 vaccination to the public and other health care professionals.

**3.Hypothesis: At least 5 hypotheses explaining the phenomenon of data that is considered in your proposed DV project.**

1. Comparing covid vaccine boosters' willingness to administer rates in two regions and whether there is a significant effect on all age groups (vaccinators)
2. Whether the situation of different races under the vaccination has a significant effect of suppressing the epidemic
3. Compared with the northern cities, the epidemic control in the southern counties is more unstable.
4. Finding out the progress in vaccinating groups and communities with the most urgent need which can be viewed on basis of age, race and ethnicity, or Vaccine Equity Metric (VEM).
5. Comparison of Whites and Latinos with Primary vaccination and boosted for Ages Under 5 and Ages 65+.

**4.Data Sources: Provide the data link(s) for primary and secondary (if any)**

Primary Data Source Link:

“Covid-19 Vaccines Administered By Demographics By County By Age”

<https://data.ca.gov/dataset/covid-19-vaccine-progress-dashboard-data>

**5.Number of Records: How many records in the data set?**

7,99,517 records

**6.Visualization Tool: What visualization tool will you use? Why?**

we will be using Tableau tool because most of the members are familiar with the use of the interface and it is readily available for students. The Tableau visualization tools has many features like the simple drag-and-drop features where the analysis the data and insights is much easier than Power BI or other tools. Further, Tableau understands and recognizes reports and makes suggestions or adjustments to enhance it.

For the purpose of this project, we will be using interactive maps and charts to depict the two chosen counties. Tableau charts are customizable which aids our group to make changes to the colors and text. The map will be used to put in perspective where the counties are in the state of CA. This helps the audience get situated and sets the stage for further conversations about the findings of COVID-19.

**7. Data Cleansing Tool: What data cleansing tool will you use? Why?**

We will be using Excel to clean the data as it is a collaborative tool where in I had cleaned the data. It is also helpful to use Excel with a large number of records and store it efficiently with no loss of data. In addition, it is easier to perform quick calculations with help of cell formulas, and use the features of “find and replace” when encountered with duplication of data.