**Data Science Program Final Project**

**Covid Vaccine Data: for the State of California**

**Executive Summary**

On March 11, 2022 represents the three-year anniversary since the World Health Organization declared Covid-19 a global pandemic, and shortly after, former President Donald Trump had declared a national emergency, with states beginning to issue stay-at-home orders, mask mandates, and capacity limits at businesses like restaurants and bars.[[1]](#footnote-1)

If 2020 was dominated by the news of how Covid-19 spread across the globe, putting normal life on hold for most people and overwhelming health care providers, then 2021 was by far focused on ending the pandemic through vaccine distribution through 2022.[[2]](#footnote-2)

With research into hundreds of vaccine candidates, there are now 3 vaccines approved by the FDA for use in the United States. As of May 13, [the CDC was reporting](https://covid.cdc.gov/covid-data-tracker/#vaccinations) that more than 580.9 million vaccine doses had been administered.[[3]](#footnote-3) Moreover, California counties like Imperial, Merin, Santa Clara, San Mateo, and San Francisco have administered the most doses per 100,000 of the total population. And with close to 75% of the California population fully vaccinated, updated guidance has changed the State.[[4]](#footnote-4)

This data project will explain the data information of vaccine rollout in the State of California as it progressed from 2021 into 2022 for the counties and demographics of community members. This data project will provide a point of view and ask questions to understand the impact of the Covid-19 vaccination administrated data sets.

**Business Objectives**

To showcase the skills that Jasmin, Lacey, and Arkajia have acquired through the Data Science program. They will be using R, Python, SQL, and other programs to process data wrangling, analyze, and visualization of the “Covid Vaccine Data: for the State of and California,” dataset made available by the California Open Data and CA Health and Human Services on Google Data Sets.

This data is from the same source as the Vaccine Progress Dashboard at <https://covid19.ca.gov/vaccination-progress-data/> which 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 statewide total vaccinations due to out-of-state residents vaccinated in California.

At the end of the data project, Jasmin, Lacey, and Arkajia should be able to explain their work in layman’s term, and present their findings to the students, faculty, staff, and potential employers, along with other interested parties via Zoom.

**Background**

Jasmin, Lacey, and Arkajia have selected the “Covid Vaccine Data: for the State of California” dataset because the topic carries interest in healthcare and affected the lives of millions of people in the United States, in general. They plan to collect understanding from this document for vaccine rollout in the State of California as it progressed from 2021 into 2022 for the counties and demographics of community members.

On December 11, 2020, The Food and Drug Administration issues an Emergency Use Authorization (EUA) for the first COVID-19 vaccine – the Pfizer-BioNTech COVID-19 vaccine.[[5]](#footnote-5)

Thereafter, December 18, 2020, The U.S. Food and Drug Administration issues an Emergency Use Authorization for the second COVID-19 vaccine – the Moderna COVID-19 vaccine. [[6]](#footnote-6)

Lastly, on February 27, 2021, FDA approves emergency use authorization for Johnson and Johnson one shot COVID-19 vaccine.[[7]](#footnote-7)

**Scope**

Jasmin, Lacey, and Arkajia will be using all resources and software we have learned throughout the course to complete the project in a timely matter. We will use the tool of our interest and/or tools that may help us finish our task at hand. We may use other software and tool if need be but that would be in each one’s interest.

**Functional requirements**

**Data Wrangling**: The downloaded dataset will be cleaned up for analyzing. Columns and unusable columns will be removed. As the dataset has a lot of important data, Jasmin, Lacey, and Arkajia will sub-stetted and prepared for proper analysis.

**Git-Hub:** Will be one of our connections for passing and reviewing the dataset. This will also help us work simultaneously on the dataset and upload data.

**Trello:** Will be the other program that will help us organize our “To Do” work and keeps us notified on due dates. Here we can communicate with each other and leave notes for each team member to view and correct or look over anything that needs it be reviewed. It also lets us know what team member is work on what data at a glance. Also lets us know who has completed what data that way there not multiple team members working on the same job.

**Data Analysis**: Jasmin, Lacey, and Arkajia will go thoroughly through the dataset to understand everything the dataset is stating. We will go through each column means, test values and make sure they are measured correctly. We will brainstorm all our questions to ask, also ask what we can gather from the dataset. We will then, compute our information with the proper functions to create models, predictions, correct any errors that may/or may not occur.

**Data Visualization:** Jasmin, Lacey, and Arkajia will run dataset through tableau to produce beautiful detail graphs. To use in our final presentation. We will create different graph and choose the best one that explains and showcases our dataset with ease understanding. Once chosen we will compile all the information and showcase it on a Power Point presentation.

**Presentation:** for our final presentation we will work with our mentor and professor to schedule a date and time to present via Zoom. We will make sure all communication is ready and clear for everyone attending can hear and understand us. We will all be very professional for this important event. We will keep the presentation about 20 mins.

**Personnel Requirements**

Our team consist of varies background and careers. We will bring all of our knowledge learned from life and other institutes to make sure the final project is a success. We will meet once a week through Zoom, Slack, Trello, and other means of communication to have everyone on the same page. On our weekly meeting we review what we did the past week and what we need to do accomplish the following week. We will take turns being scrum manager to gain experience in delegating and reporting to mentors and professor.

**Delivery schedule**

Week 1: Import and begin to analyze which dataset to use to perform each analysis for the questions that are proposed. Begin to review the data chosen for the proposed questions. Educate ourselves on the topic at hand. Set Up Github.

Week 2: Study the dataset and begin to ask questions. Break up the questions proposed and identify the methods used to analyze the data and see if there were any inconsistencies. Are there any correlations? What possible methods of presentation would best showcase the analysis? Whether to use R Studio or Python etc.

Conduct weekly meeting to discuss the proposed working going forward.

Week 3: Using the different components of the program from what we learned and demonstrate how to use; Data Wrangling and Visualization: Data Transformations. Intermediate Statistics- Single Sample T-Tests, Independent T-Test. Tableau to make visual presentation of the data.

Conduct weekly meeting to discuss the proposed working going forward.

Week 4: Review and validate findings from the previous week, and draw insights/conclusions.

Conduct weekly meeting to discuss the proposed working going forward.

Week 5: Begin to piece together the presentation. Reviewing each component to ensure it will be presented clear and concise. Work out any visual media to best present the project.

Conduct weekly meeting to discuss the proposed working going forward.

Week 6: Make final touches to the Power Point presentation. Jasmin, Lacey, and Ar’kajia will not attempt to come up with a brand-new analysis. There will not be enough time to verify their findings. We will practice presenting and determine suitable placement. Present the presentation at least once with their instructor.

**Other requirements**

All programs used should be free of charge.

**Assumptions**

All of the programs that will be used will be readily available to Jasmin, Lacey, and Ar’kajia.

**Limitations**

Limitations collectively will be scheduling. Respecting the schedules of Jasmin, Lacey, and Ar’kajia may be a limitation, but we are eager to take ownership and maintain the weekly scheduled deadlines. Availability and not being able to maintain the weekly schedule. We expect to have a timely delivery of the project.

**Risks**

There may be many risks associated with completing a group project. One of them being, failing to maintain a schedule and meeting deadlines. Jasmin, Lacey, and Ar’kajia are determined to be efficient in communication and conveying all discrepancies. Collectively we will be working diligently to complete the project in the proposed timeline.

1. [↑](#footnote-ref-1)
2. https://www.ajmc.com/view/a-timeline-of-covid-19-vaccine-developments-in-2021 [↑](#footnote-ref-2)
3. https://covid.cdc.gov/covid-data-tracker/#vaccinations\_vacc-total-admin-rate-total [↑](#footnote-ref-3)
4. https://covid19.ca.gov/vaccination-progress-data/ [↑](#footnote-ref-4)
5. [↑](#footnote-ref-5)
6. [↑](#footnote-ref-6)
7. https://www.cdc.gov/museum/timeline/covid19.html#Late-2020 [↑](#footnote-ref-7)