Analyzing California Vulnerable Communities for COVID-19 Intervention Efforts

Correlation One: Data Science For All

Team 38: Marci Miller, Camilius Amevorku, Joseph Compaore, Symphony Hopkins, Taiwo Akinyemi

Table of Contents

1. Introduction

The pandemic has impacted the health of millions of individuals, disproportionately impacting people of color. Research has proven the positive effects of vaccinations in reducing the spread of COVID-19, ultimately decreasing negative symptoms and deaths from the virus. At the state level, government officials and the Public Health Department need to measure the impact of COVID 19 across the state, in addition to vaccination status in order to focus intervention efforts on the most vulnerable populations.

Through an analysis of COVID cases, hospitalizations, deaths, and vaccination status by county, zip code, and demographic, our goal is to help CA Public Health Officials target resources for intervention/vaccination efforts for those most in need. We plan to use our final model and data visualizations of our analysis to inform recommendations on COVID-19 Prevention efforts. Local, county, policymakers, and nonprofits in CA can use this data and analysis to target intervention efforts for populations most in need and decrease the spread of COVID-19. This analysis is also useful in measuring the impact of current immunization programs.

2. Data Analysis & Computation

2.1. Data Sources

Our data analysis focused on the following datasets from California Open Data: Statewide COVID-19 Cases Deaths Demographics, COVID-19 Vaccines Administered by Demographics, Statewide COVID-19 Vaccines Administered by Zip Code, COVID-19 Post Vaccination Statewide Stats, Statewide Covid-19 Hospital County Data, Statewide COVID-19 Vaccines Administered By County.

2.1.1. Statewide COVID-19 Cases Deaths Demographics

State COVID-19 Cases Deaths Demographics between report dates 4/13/2020 and 12/5/21.

10,644 rows & 8 columns. Size: 584 KB. Source: Statewide COVID-19 Cases Deaths Demographics

2.1.2. COVID-19 Vaccines Administered by Demographics

Statewide COVID-19 Vaccines Administered By Demographics between dates 7/27/2020 and 12/5/2021. Starting on November 10, 2021, columns added for age 5+ denominator for calculating vaccine coverage to reflect new vaccine eligibility criteria.

7,878 rows & 17 columns. Size: 820 KB. Source: COVID-19 Vaccines Administered by Demographics

2.1.3. Statewide COVID-19 Vaccines Administered by Zip Code

Data of the estimating Covid-19 vaccines administered by zip code in California for the period of 01/05/2021 to 11/30/2021.

84,672 rows & 14 columns. Size: ~10MB. Source: CA Statewide COVID-19 Vaccines Administered by zip code

2.1.4. COVID-19 Post Vaccination Statewide Stats

“The California Department of Public Health (CDPH) is identifying vaccination status of cases and deaths by analyzing the state immunization registry and registry of confirmed COVID-19 cases. Post-vaccination cases, also referred to as vaccine breakthrough cases, are individuals who have a positive SARS-Cov-2 molecular test at least 14 days after they have completed their full one-dose or two-dose vaccination series”(CA.gov).

300 rows and 17 columns. Size: 40 kb. Source: COVID-19 Post-Vaccination Infection Data - COVID-19 Post Vaccination Statewide Stats - California Open Data

2.1.5. Statewide COVID-19 Hospital County Data

Data is from the California COVID-19 State Dashboard at https://covid19.ca.gov/state-dashboard/. Cumulative totals are not available due to the fact that hospitals report the total number of patients each day (as opposed to new patients).

Source: COVID-19 Hospital Data - Statewide Covid-19 Hospital County Data - California Open Data

2.1.6. Statewide COVID-19 Vaccines Administered By County

This data is from the same source as the Vaccine Progress Dashboard at https://covid19.ca.gov/vaccines/ which summarizes vaccination data at the county level by county of residence. Where the county of residence was not reported in a vaccination record, the county of the 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.

Source: COVID-19 Vaccine Progress Dashboard Data - Statewide COVID-19 Vaccines Administered By County - California Open Data

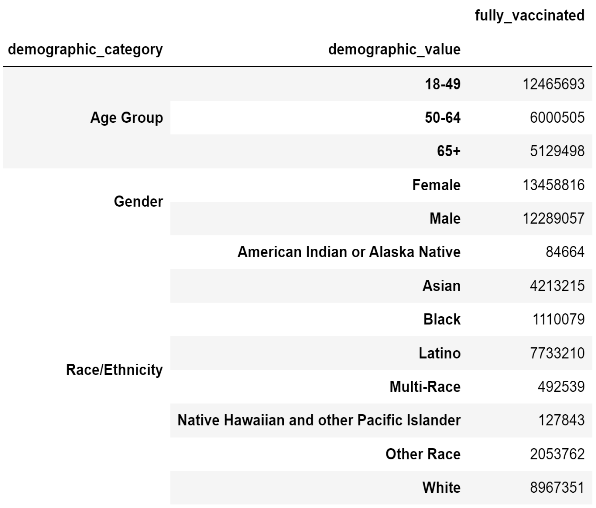
2.2. Datasets Cleaning & Exploratory Data Analysis

2.2.1. COVID-19 Cases & Vaccines Demographics Analysis

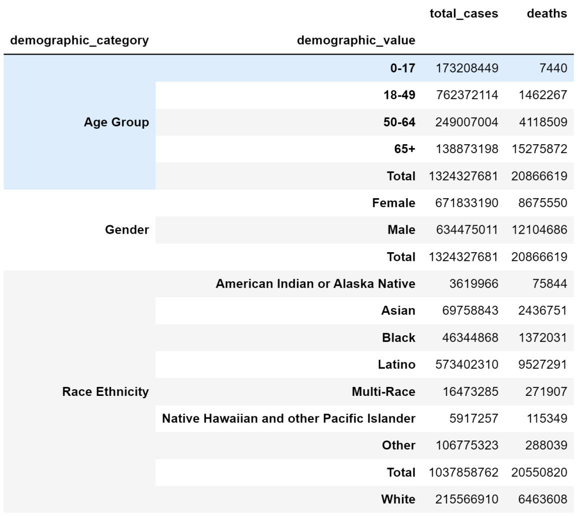
- **Data Sources**
  - Statewide COVID-19 Cases Deaths Demographics

    - COVID-19 Vaccines Administered by Demographics
- **Cleaning Process**
    - For the Statewide COVID-19 Cases Deaths Demographics dataset, we imported the csv file into Jupyter notebook, first analyzing the shape and head for any immediate data anomalies. We reviewed the data types to check if they are standard with format, in addition to reviewing unique values for demographic and value categories. We updated the report date to a date time. Initially, the demographic object variables were string types, but it was changed to object variables for a cleaner output. We added a month a year column, in order to look for any trends across months (min 1, max 12) and years (min 2020, max 2021). The initial data set included "missing", "Missing" and "Unknown" all to represent unknown demographic values; we updated those values to all be standardized as "Unknown." We used a for loop to review null values across columns, and found no pivotal columns with null values. We kept dropped columns for the initial exploratory process.
    - For the COVID-19 Vaccines Administered by Demographics dataset, a similar cleaning process was followed to address data anomalies. We updated incorrect values to "Unknown" (eg. May 11 as value for the age group updated). We replaced demographic racial/ethnicity values with the corresponding values in the demographic data in order to create standardized values that can be compared to each other. We used a for loop to review null values across columns, and found no columns with null values. We kept dropped columns for the initial exploratory process.

- **Exploratory Data Analysis**
    - Process Step 1: We selected columns of interest and target features.
        - Columns of Interest: Demographic Category and Value, Total Cases, Deaths, Year, and Month. Key pieces of information refer to the total cases and deaths for each demographic category: age group, gender, and race. Moreover, key information in partially vaccinated/fully vaccinated individuals.
            - Covid Cases DF: Demographic Columns: Category; Total cases and deaths: numeric, year and month: datetime
            - Vaccines DF: Demographic Columns: Category, Partially Vaccinated, Fully Vaccinated
        - Columns/Rows Removed in Analysis: We removed index/rows of demographic values with unknown demographic value variables. This was key in determining which known demographics public health interventions should target.
        - Preliminary Insights
            - Nulls values: No null values across columns kept for analysis.
            - Data from April 2020 to January 2021.
            - See figures below for more preliminary insights.

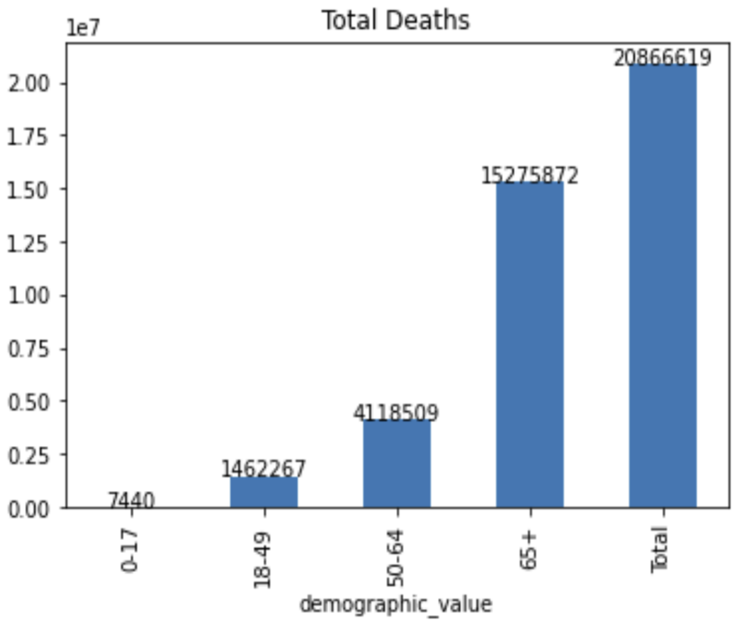


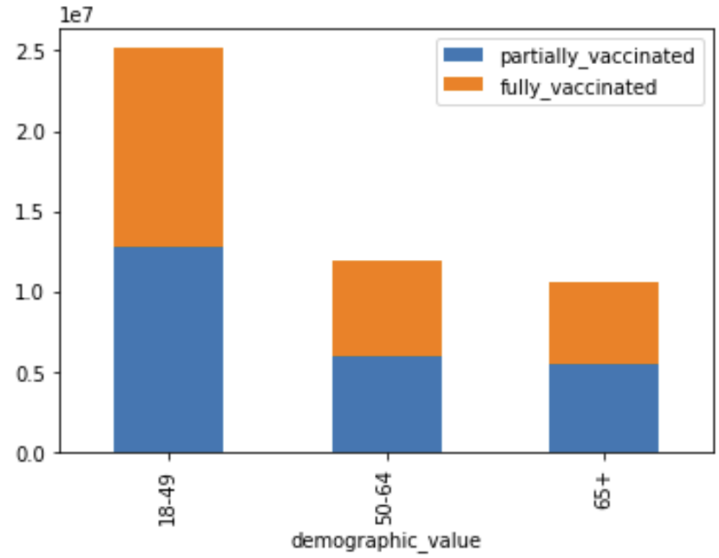
		fully_vaccinated
demographic_category	demographic_value	
Age Group	18-49	12465693
	50-64	6000505
	65+	5129498
Gender	Female	13458816
	Male	12289057
Race/Ethnicity	American Indian or Alaska Native	84664
	Asian	4213215
	Black	1110079
	Latino	7733210
	Multi-Race	492539
	Native Hawaiian and other Pacific Islander	127843
	Other Race	2053762
	White	8967351

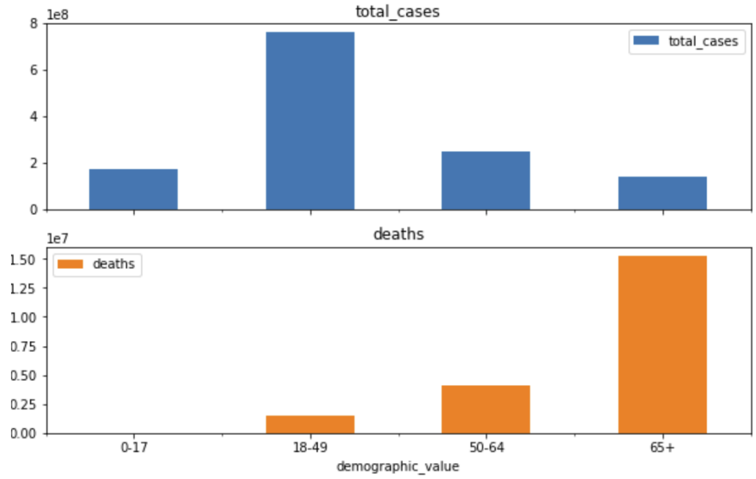


		total_cases	deaths
demographic_category	demographic_value		
Age Group	0-17	173208449	7440
	18-49	762372114	1462267
	50-64	249007004	4118509
	65+	138873198	15275872
	Total	1324327681	20866619
Gender	Female	671833190	8675550
	Male	634475011	12104686
	Total	1324327681	20866619
Race Ethnicity	American Indian or Alaska Native	3619966	75844
	Asian	69758843	2436751
	Black	46344868	1372031
	Latino	573402310	9527291
	Multi-Race	16473285	271907
	Native Hawaiian and other Pacific Islander	5917257	115349
	Other	106775323	288039
	Total	1037858762	20550820
	White	215566910	6463608

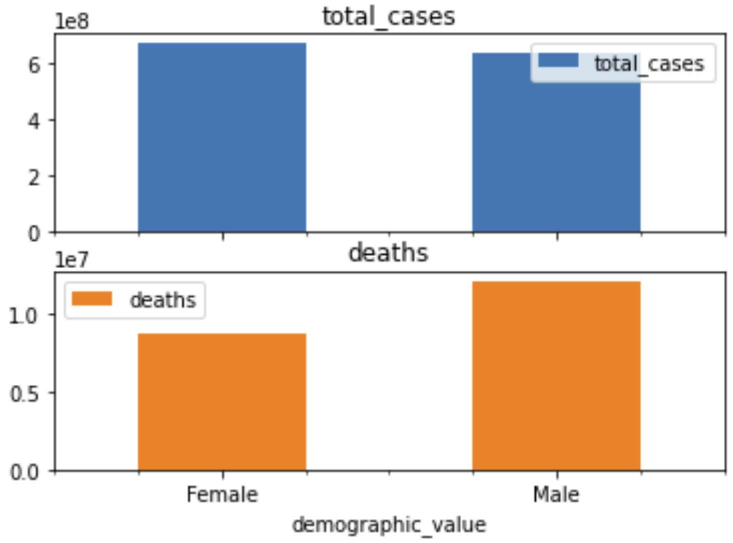
- Process Step 2: We created separate data frames for age group, gender, and race/ethnicity, to compare total cases and deaths across each demographic value. Additionally, separated data frames were created for partially vaccinated/fully vaccinated status.
  - Bar Graph Data Visualizations for Total Cases/Deaths and Vaccines Status (See Figures Below):
    - Age Group: 18-49 with higher cases, 65+ least cases but highest deaths. Younger individuals with a greater number of vaccinations.

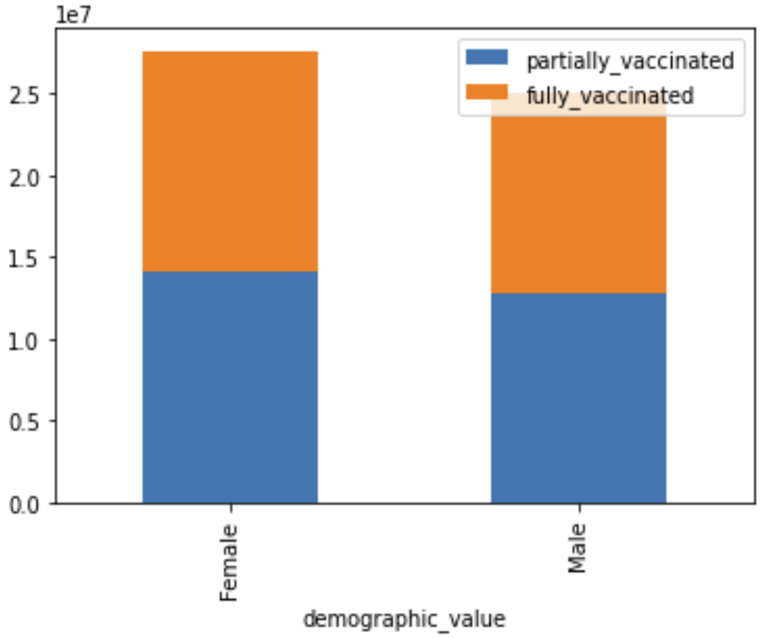




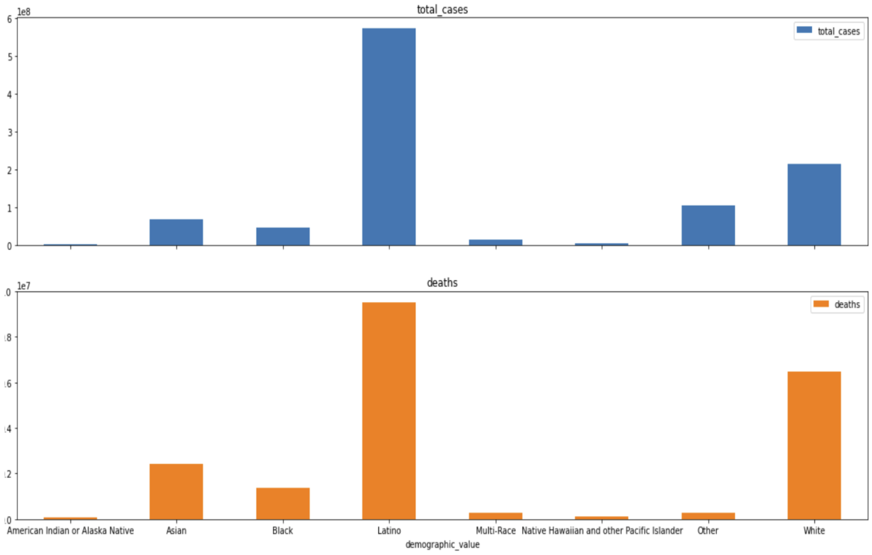


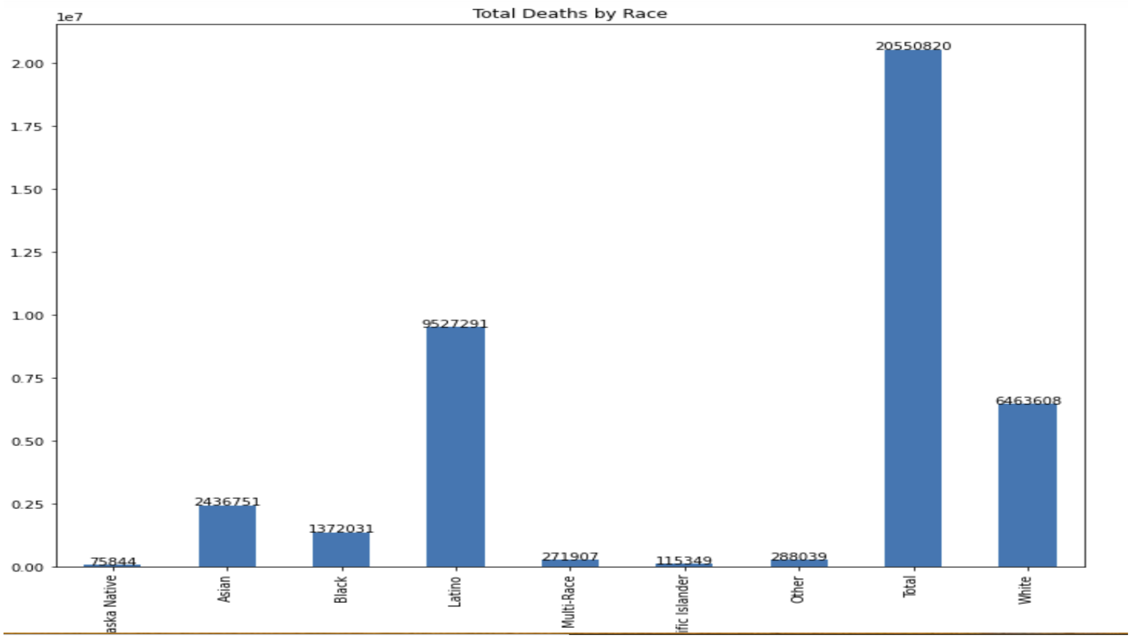
- Gender: Females and males with similar total cases, males with higher reported deaths. Females with a higher number of fully vaccinated.

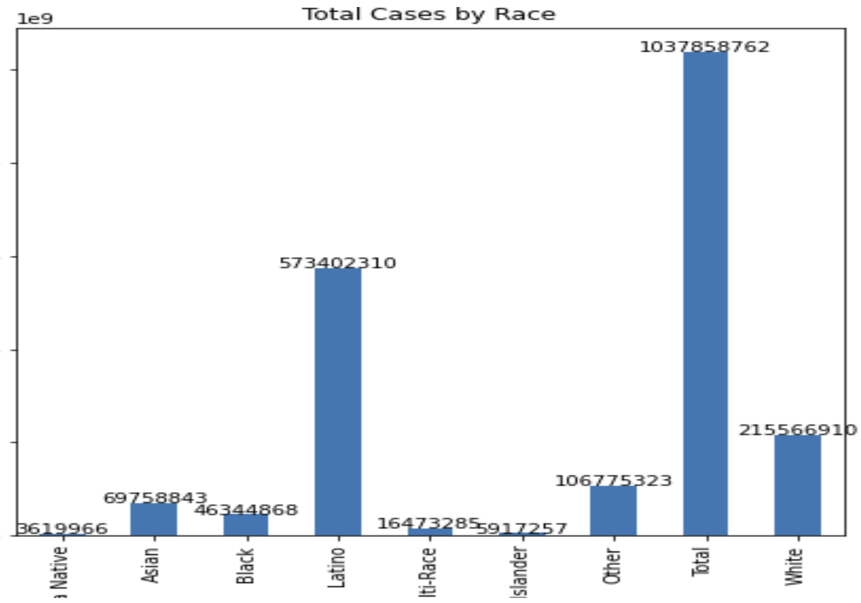


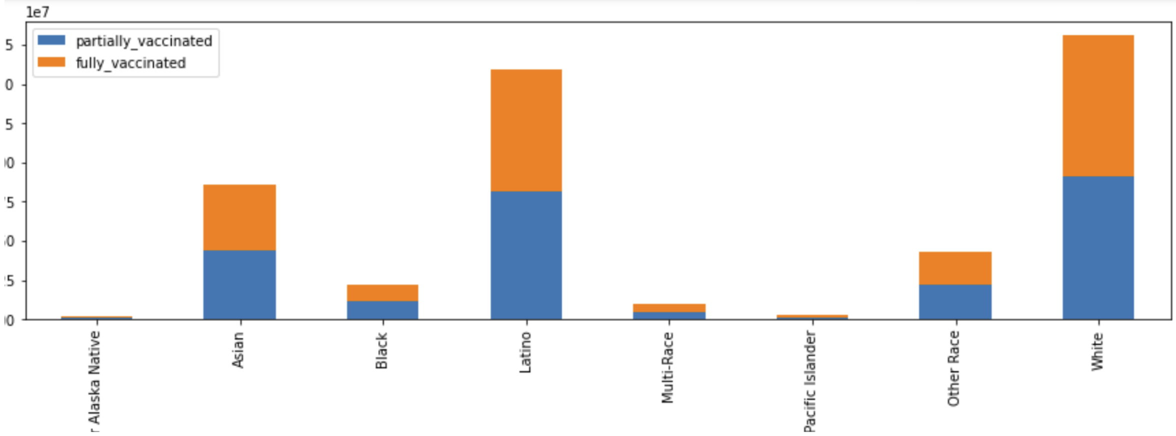


- Race/Ethnicity: Latino population with highest number of cases/deaths. White individuals have a higher number of fully vaccinations, followed by Latino, then Asian.



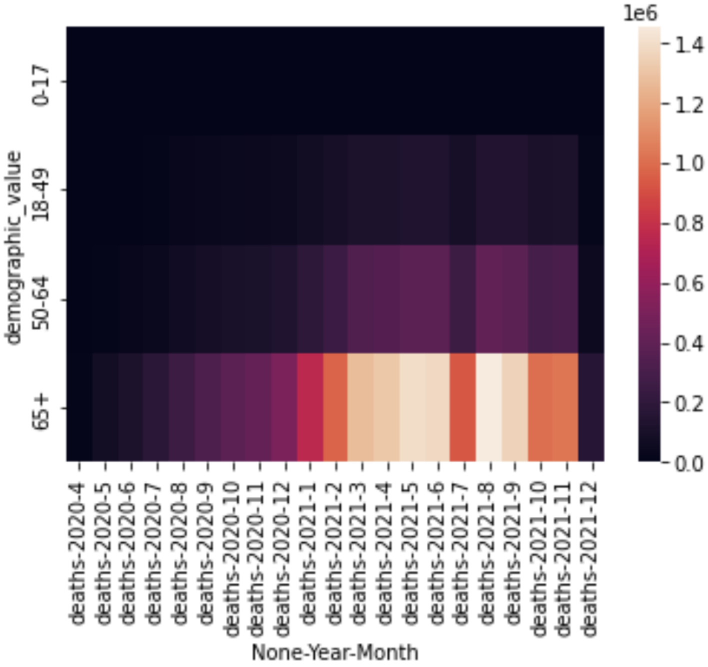


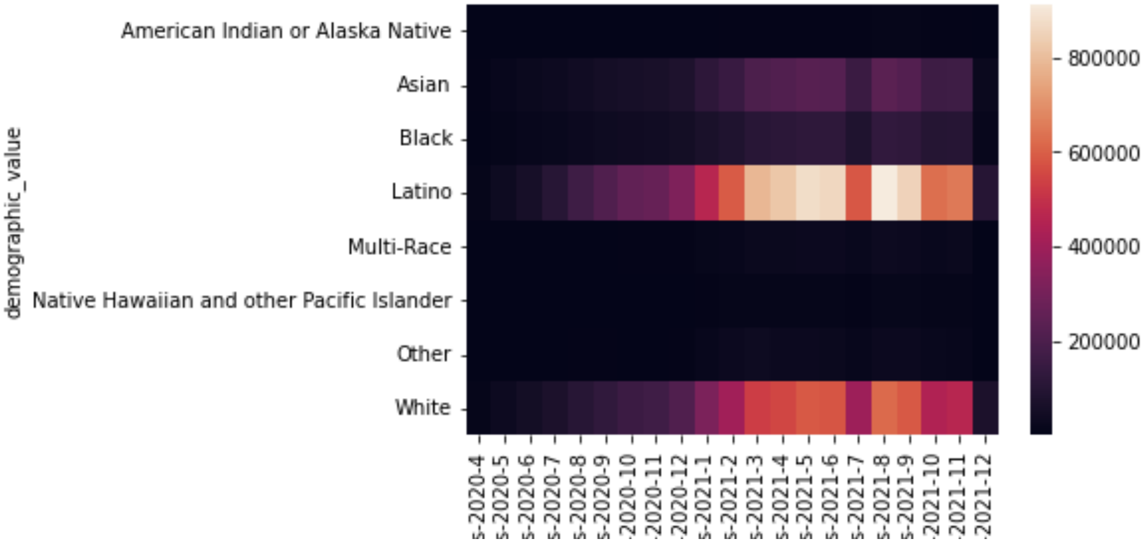




- Process Step 3: We created six pivot tables (two per demographic category representing death and total cases per month over the two years of the data time frame).
    - We used each pivot table to create a heat map across data to see trends in covid cases/deaths over time.
        - These heat maps showed similar trends to the bar graphs in which the most vulnerable populations were 65+ and Latino populations.
        - Additionally, it seems there was an uptick in cases/deaths springtime of 2021 up until summer.

- Concerning point is that 65+ had the least number of cases, but highest death rates.





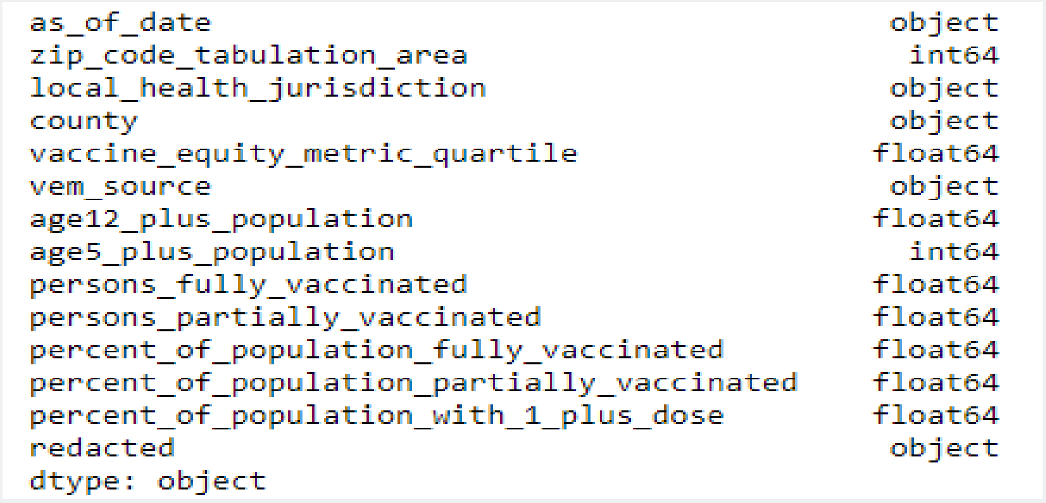
  - We found that correlations were evident between age, gender, and race on covid cases and deaths. We also found that the total cases/deaths were not spread out evenly across each demographic category.

- **Summary**
  - While younger individuals displayed the highest amount of cases, asides from the teens, they displayed the lowest amount of deaths. The older individuals displayed the lowest number of cases, and highest deaths making them a target population for intervention efforts. Lastly, across race, it was clear that the Latino

population is the most vulnerable to both cases and deaths making them a key target population. Vaccine data displayed high numbers among youth and white individuals, however, it should be noted that this data is limited to early 2021.

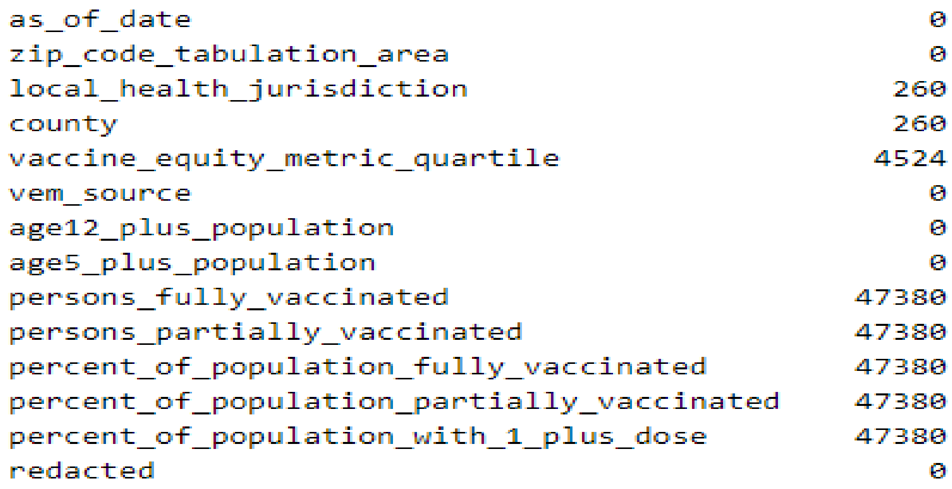
2.2.2. COVID-19 Vaccination by Zip Code Analysis

- **Data Sources**
  - Statewide COVID-19 Vaccines Administered by Zip Code
- **Cleaning Process**
  - All columns were renamed for clarity. Columns not relevant to the targeted goal were removed (e.g. local_health_jurisdiction, vem source, redacted). Columns containing numerical information, but not coded as such were converted to integers. It is quite a compulsory process to modify the data we have as the computer will show an error of invalid input as it is quite impossible to process the data having 'NaN' with it and it is not quite practically possible to manually change the 'NaN' to its mean. Therefore, we resolved this problem by processing the data and using various functions so that the 'NaN' was removed from our data and replaced with the particular mean so it could be processed by the system.
- **Exploratory Data Analysis**
  - Process Step 1: We identified columns in the dataset that would help us answer the question: What is the trend of the fully vaccinated (or partially vaccinated) by zip code?
    - We looked at the data types of each column:



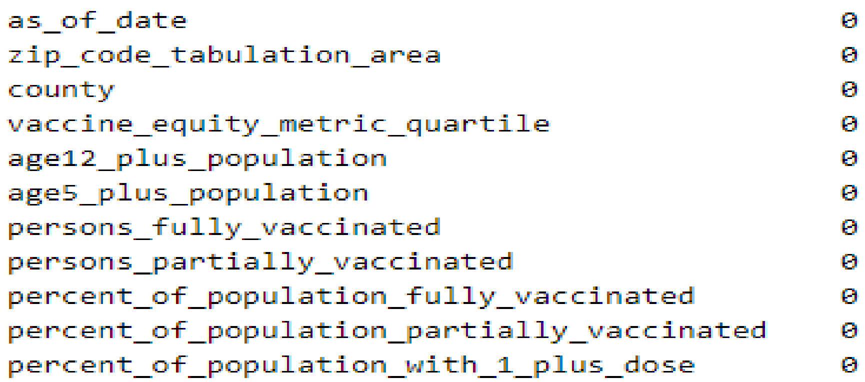
```
as_of_date                                  object
zip_code_tabulation_area                     int64
local_health_jurisdiction                   object
county                                      object
vaccine_equity_metric_quartile             float64
vem_source                                  object
age12_plus_population                      float64
age5_plus_population                         int64
persons_fully_vaccinated                   float64
persons_partially_vaccinated               float64
percent_of_population_fully_vaccinated     float64
percent_of_population_partially_vaccinated float64
percent_of_population_with_1_plus_dose     float64
redacted                                    object
dtype: object
```

  - Process Step 2: We explored the individual columns to identify if null values are present, and how to deal with them. Additionally, we examined each column to determine if we needed to drop them.



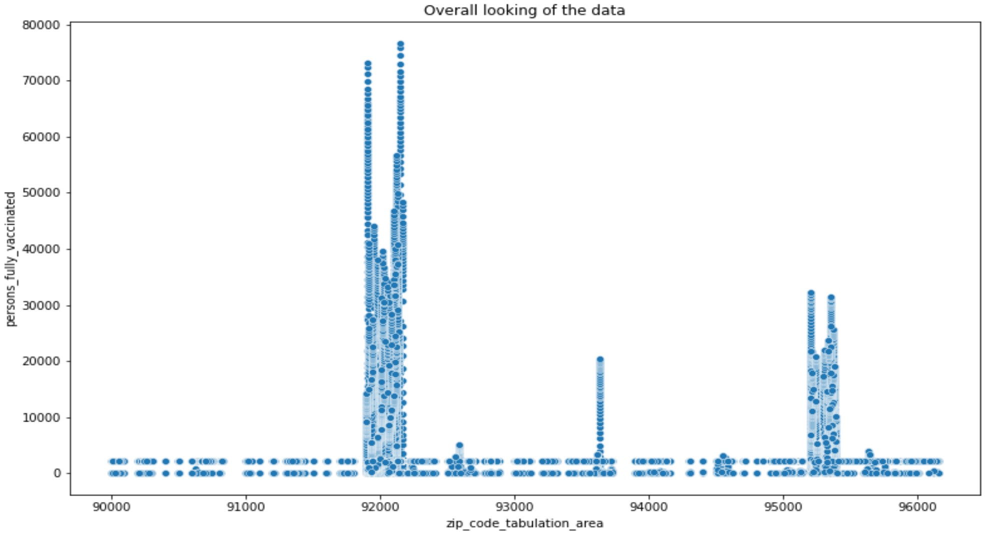
```
as_of_date                                       0
zip_code_tabulation_area                         0
local_health_jurisdiction                      260
county                                         260
vaccine_equity_metric_quartile                4524
vem_source                                       0
age12_plus_population                            0
age5_plus_population                             0
persons_fully_vaccinated                     47380
persons_partially_vaccinated                 47380
percent_of_population_fully_vaccinated       47380
percent_of_population_partially_vaccinated   47380
percent_of_population_with_1_plus_dose       47380
redacted                                         0
```

- From the exploration, we observed that…
    - There are some columns that are not useful for our analysis so we will drop them.
    - *Columns: local_health_jurisdiction, vem_source, redacted**
    - There are null values in some columns and we will use the mean() to replace numerical null values.
    - *Columns: vaccine_equity_metric_quartile, persons_fully_vaccinated, persons_partially_vaccinated, percent_of_population_fully_vaccinated, percent_of_population_partially_vaccinated, percent_of_population_with_1_plus_dose**
    - Some column data types need to be converted to the right data type for us to perform analysis on them.
    - *Columns: age12_plus_population, persons_fully_vaccinated, persons_partially_vaccinated**
- The below output gave us a simpler dataset and all null values removed for a better analysis:

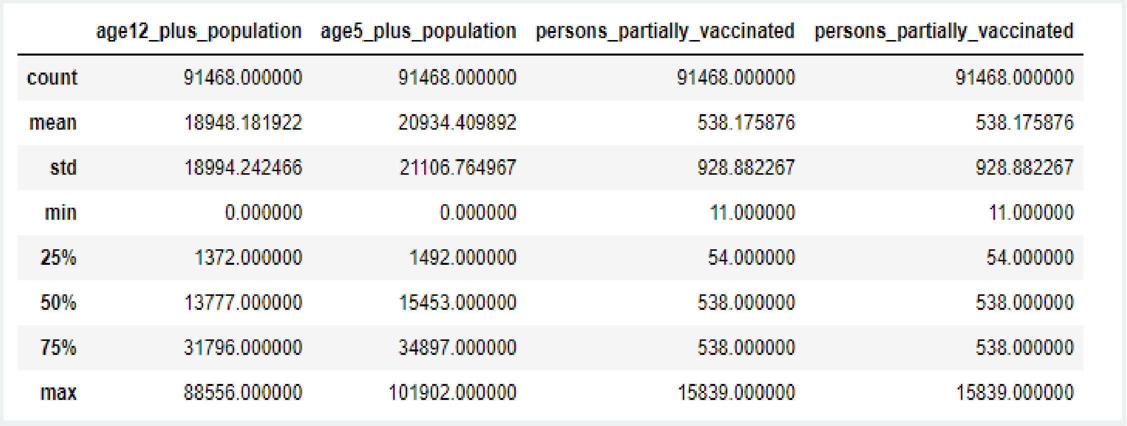


```
as_of_date                                    0
zip_code_tabulation_area                      0
county                                        0
vaccine_equity_metric_quartile                0
age12_plus_population                         0
age5_plus_population                          0
persons_fully_vaccinated                      0
persons_partially_vaccinated                  0
percent_of_population_fully_vaccinated        0
percent_of_population_partially_vaccinated    0
percent_of_population_with_1_plus_dose        0
```

- We plotted one dimensional distributions of the numerical columns to have a look of the trend of the data.



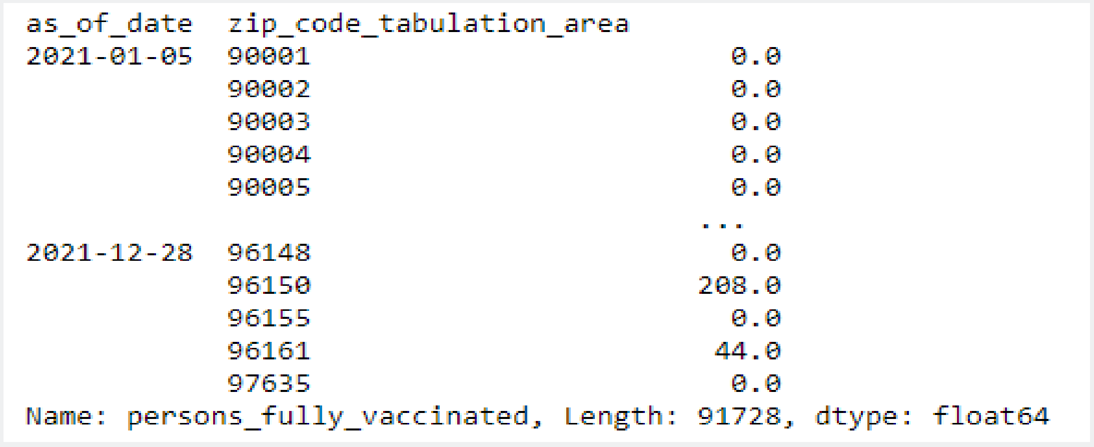
- The view of the overall plot of the data revealed a trend.
- To have a better understanding of all numerical columns, we proceeded to do a basic statistical analysis of some numerical columns:



	age12_plus_population	age5_plus_population	persons_partially_vaccinated	persons_partially_vaccinated
count	91468.000000	91468.000000	91468.000000	91468.000000
mean	18948.181922	20934.409892	538.175876	538.175876
std	18994.242466	21106.764967	928.882267	928.882267
min	0.000000	0.000000	11.000000	11.000000
25%	1372.000000	1492.000000	54.000000	54.000000
50%	13777.000000	15453.000000	538.000000	538.000000
75%	31796.000000	34897.000000	538.000000	538.000000
max	88556.000000	101902.000000	15839.000000	15839.000000

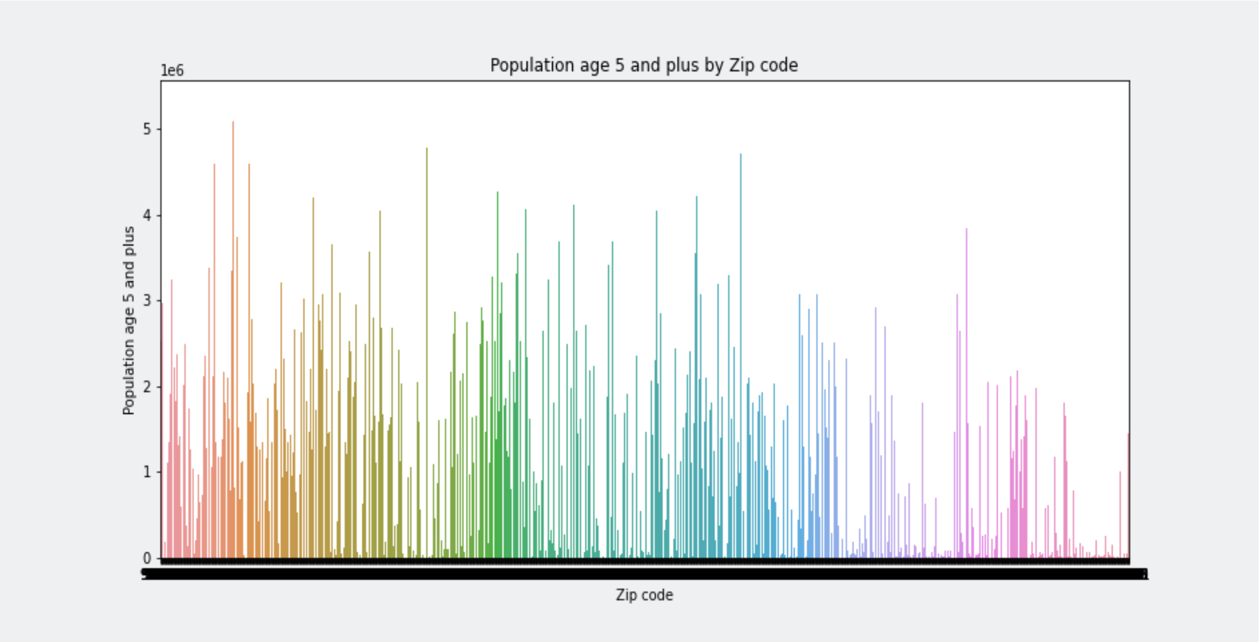
- We performed a subgroup calculation of categorical data (column) to see the distribution of these categorical columns.

        - We grouped the columns by time periods, and discovered that the data frame time period is from January 1, 2021 to December 28, 2021.

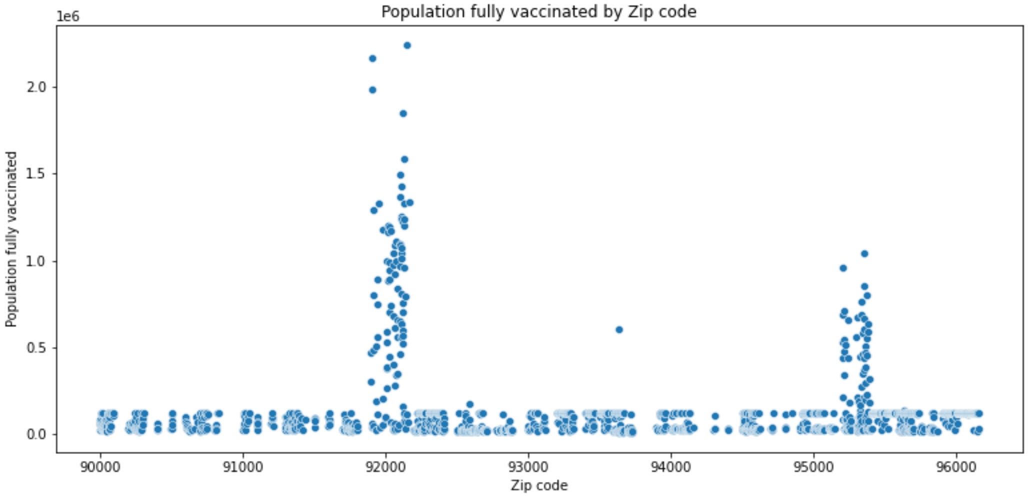


```
as_of_date  zip_code_tabulation_area
2021-01-05  90001                         0.0
            90002                         0.0
            90003                         0.0
            90004                         0.0
            90005                         0.0
                                         ...
2021-12-28  96148                         0.0
            96150                       208.0
            96155                         0.0
            96161                        44.0
            97635                         0.0
Name: persons_fully_vaccinated, Length: 91728, dtype: float64
```

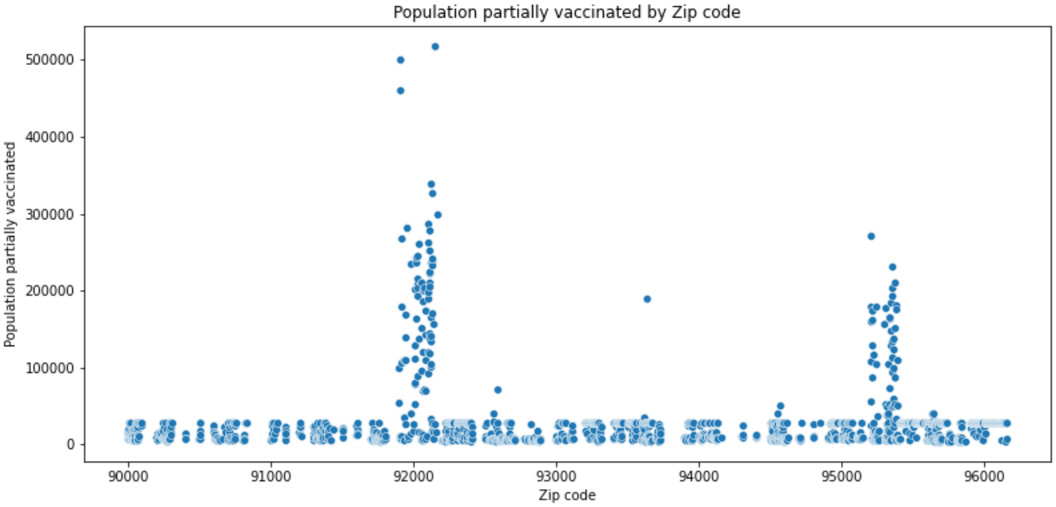
    - Process Step 3: To have a deep insight of the trend, we created data visualizations comparing one dimensional columns to another.
        - First, we plotted the total population age +5 by zip code. Looking at the graph below, we can see the distribution of the population across all zip codes (county).



        - Second, we plotted the total of people fully vaccinated by zip code to see the trend compared to the first graph.



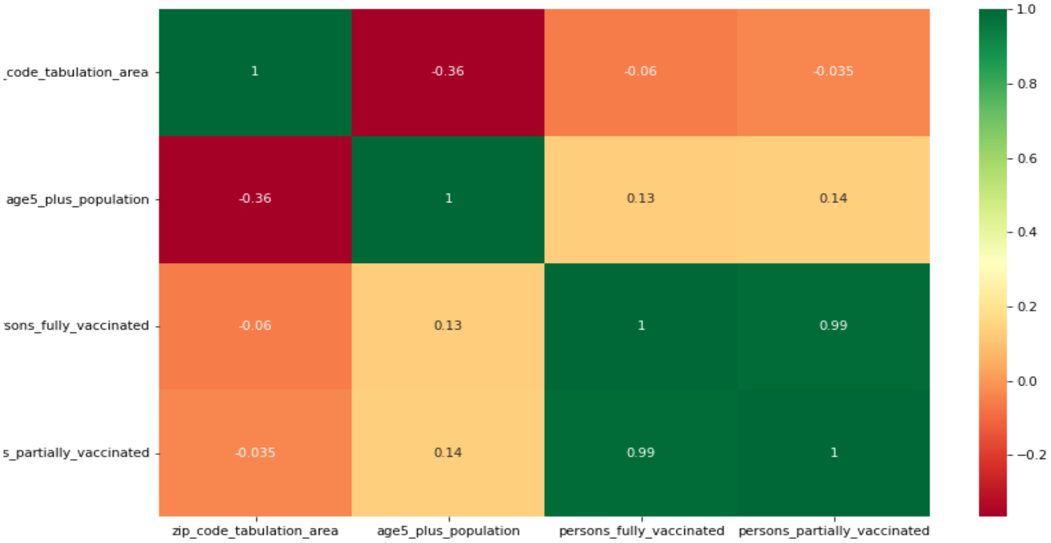
- Third, we plotted the total of the person partially vaccinated by zip code.



- Looking at the two last graphs, it seemed that they had the same trend over the zip code.

○ Process Step 4: To have a better look and to make the final analysis, we performed correlation analysis to answer the following questions: How does the zip code determine high vaccination or affect vaccination? What is the relation between zip code and persons fully vaccinated? What is the relation between population (by ages) and persons fully vaccinated?

  ■ From the heatmap below we can observe that there is not a strong correlation between the total number of persons fully vaccinated and the zip code.



- **Summary**
  - The scatter plot and bar plot gave us an overall visualization of the data. For our case, it gave us a look of the dispersion of the population overall zip code (county). Taking a deep look at the individual dimension, we could see that the rate of persons fully vaccinated and partially vaccinated is almost the same and there is not a strong correlation between the population and the population fully vaccinated (or partially) by zip code. So while a few zip codes see a very high rate of persons vaccinated, the overall zip code needs to do more to get at least the rate of the population vaccinated to the mean of the total population.

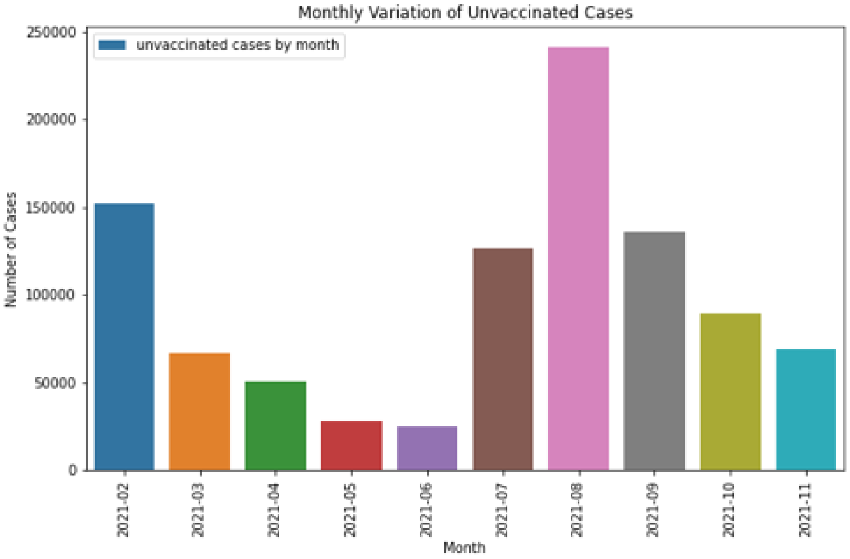
2.2.3. COVID-19 Post-Vaccination Infection Data Analysis

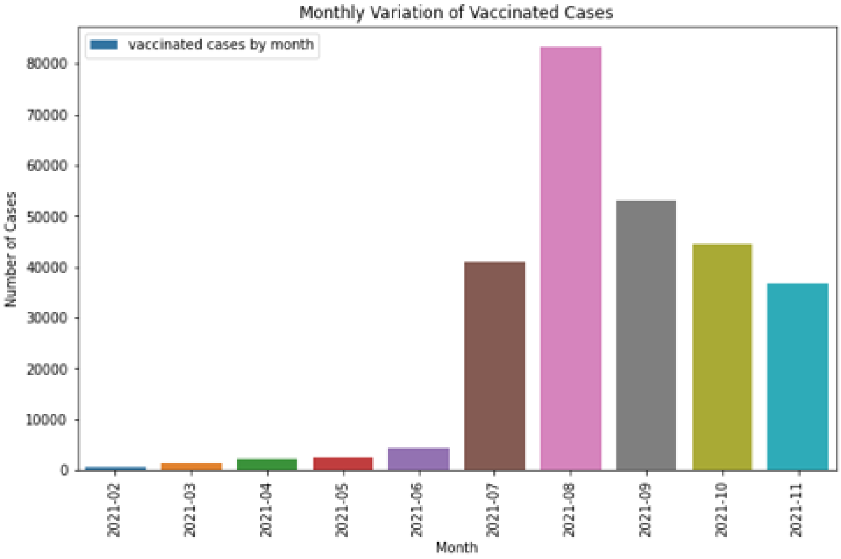
- **Data Sources**
  - COVID-19 Post Vaccination Statewide Stats
- **Cleaning Process**
  - The data was imported into a Jupyter notebook and inspected. The columns of the dataframe were checked for the existence of null values. This gave an output of all the columns and the total number of null values in the columns. From the results, there are no null values hence there was no need to deal with filing the null values. The elements of the columns “area” and “area_type” were checked for the number of unique values in them. It turned out that both of the columns have only one unique value each. That is, the unique value of “area” is “California” and that of “area_type” is “State”. Since those unique values are the same throughout, it makes it irrelevant to include them in the table. Hence the columns for “area” and “area_type” were dropped. A new column for “months” and “weekday” were derived from the “date” column and added to the dataframe

to enable investigation of the post-vaccination infection data on the basis of months and weekdays.

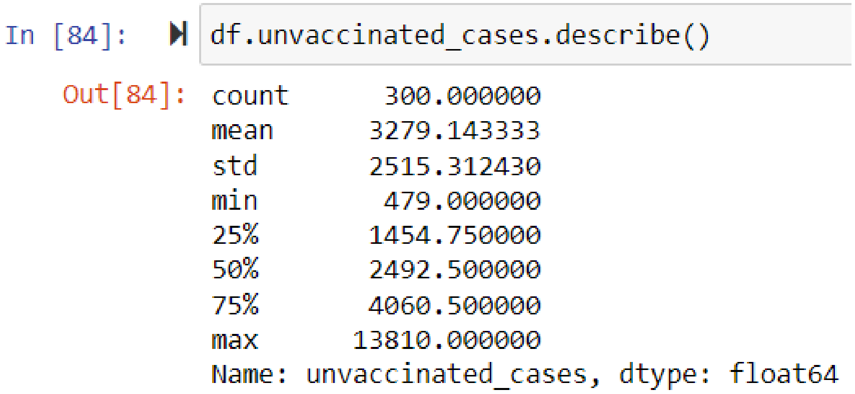
- **Exploratory Data Analysis**
    - Process Step 1: We selected columns of interest and target features.
        - The columns in the dataset that would help answer the questions in the problem statement are “date”, “unvaccinated_cases”, “vaccinated_cases”, “unvaccinated_hosp”, “vaccinated_hosp”, “unvaccinated_deaths”, “vaccinated_deaths“, “population_unvaccinated” and “population_vaccinated”
        - The columns that held key information that needed to be examined thoroughly were the “vaccinated_cases”, “vaccinated_hosp” and “vaccinated_deaths”. These columns were compared to their corresponding unvaccinated columns to highlight the significance of vaccination.
        - Of a total of 17 columns, three columns are of the datetime format and the remaining columns are of the numerical format.
    - Process Step 2: We explored the individual columns for preliminary insights
        - All the columns were inspected for the presence of null values but none of the columns have null values
        - The data was grouped by months and summed by unvaccinated and vaccinated cases respectively. Barplots were generated to visualize the month distributed of the cases.

- From the barplots, it can be seen that the month of August recorded the highest number of both unvaccinated and vaccinated cases.

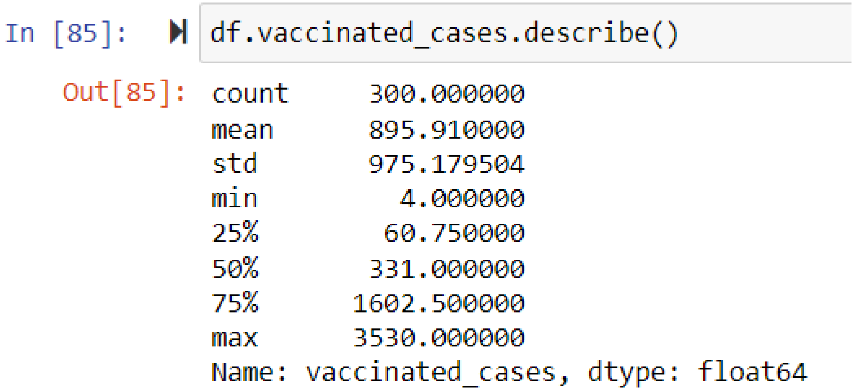




- The basic statistics of both unvaccinated and vaccinated are presented below:

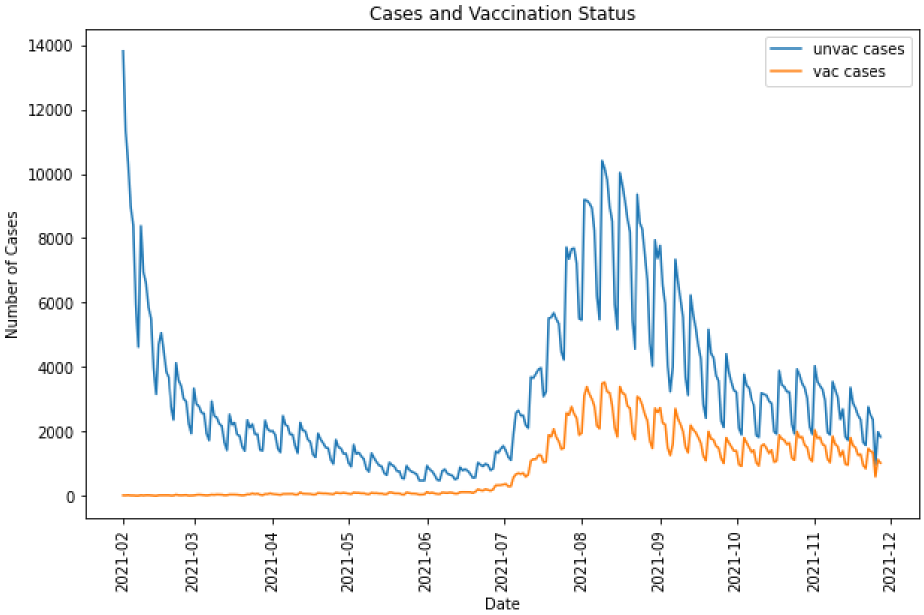


```
In [84]:  df.unvaccinated_cases.describe()

Out[84]:  count      300.000000
          mean      3279.143333
          std       2515.312430
          min        479.000000
          25%       1454.750000
          50%       2492.500000
          75%       4060.500000
          max      13810.000000
          Name: unvaccinated_cases, dtype: float64
```

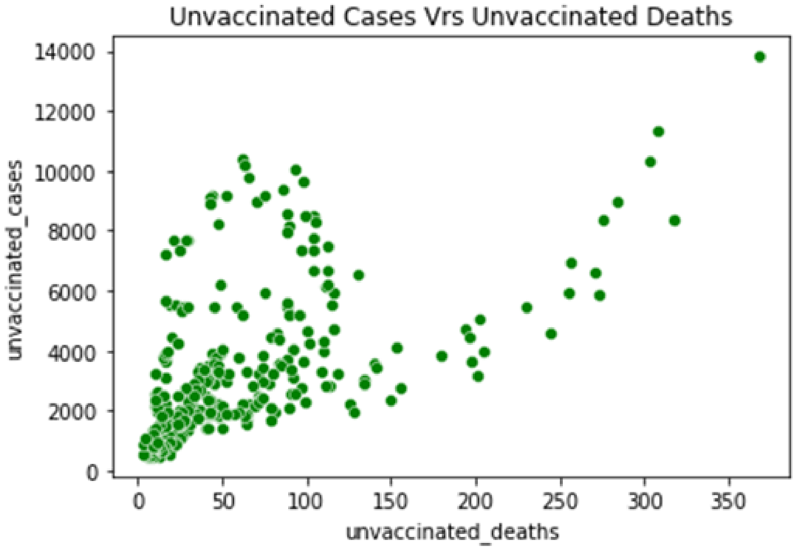


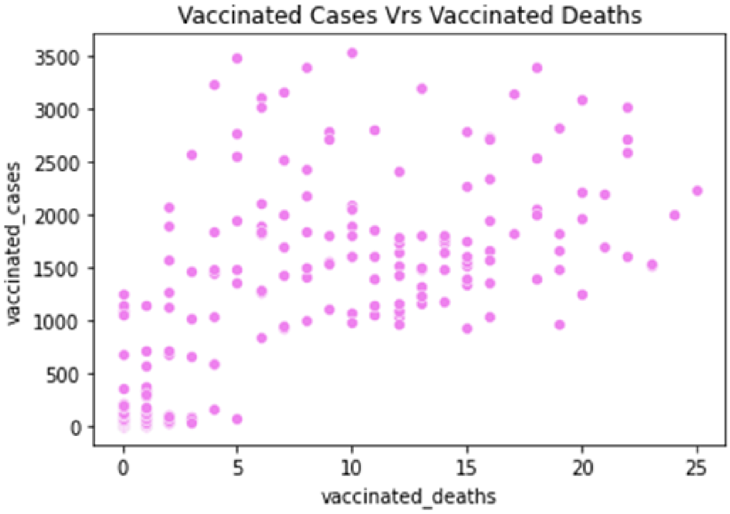
```
In [85]:  df.vaccinated_cases.describe()

Out[85]:  count     300.000000
          mean      895.910000
          std       975.179504
          min         4.000000
          25%        60.750000
          50%       331.000000
          75%      1602.500000
          max      3530.000000
          Name: vaccinated_cases, dtype: float64
```

- To explore the seasonality of the date/datetime columns for basic trends, line plots were generated for both unvaccinated and vaccinated cases to check for patterns in the data. From the line plot, it can be seen that though the number of unvaccinated cases are higher than the vaccinated cases, they both follow a similar pattern during the second half of the year.

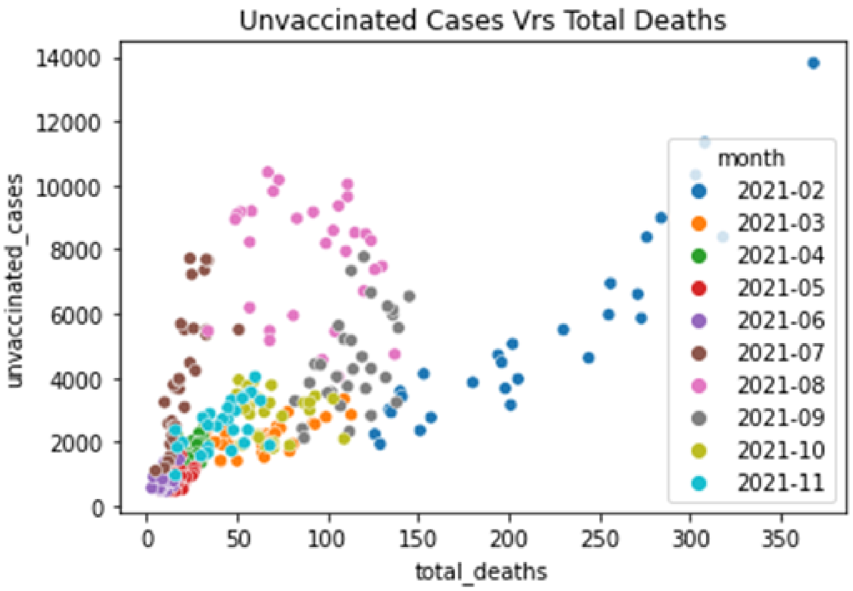


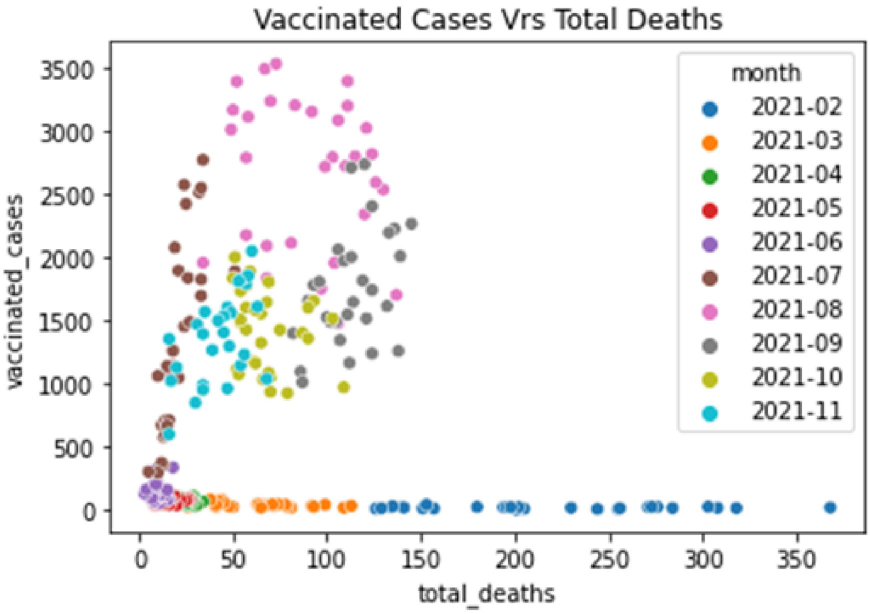
- Process Step 3: Plotting two-dimensional distributions of variables of interest against target variable(s).
    - The first set of scatter plots below show unvaccinated cases plotted against unvaccinated deaths (green); and vaccinated cases plotted against vaccinated deaths (violet). It can be seen that the points for the unvaccinated plot are more clustered than those of the vaccinated plot





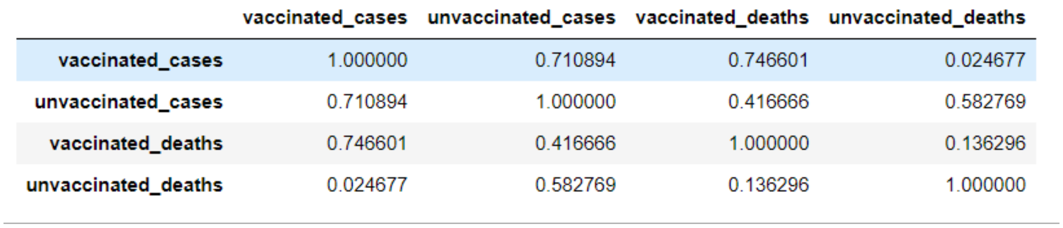
- The scatterplots below are the plots of unvaccinated cases against total deaths and vaccinated deaths against total deaths respectively. The markers on the plots are grouped/colored based on the month they are associated with. The total deaths were derived by adding the unvaccinated deaths to the vaccinated deaths. From both plots, it can be seen that the highest total number of deaths occured in the month of February (blue)





- Process Step 4: We analyzed any correlations between the independent and dependent variables.
    - The table below shows the correlation between the variables. The highest non-similar correlation is 0.746601 and this occurs between vaccinated cases and vaccinated deaths. There also seems to be a strong correlation between vaccinated cases and unvaccinated cases. These observations

will be further explored in the in-depth exploratory data analysis.



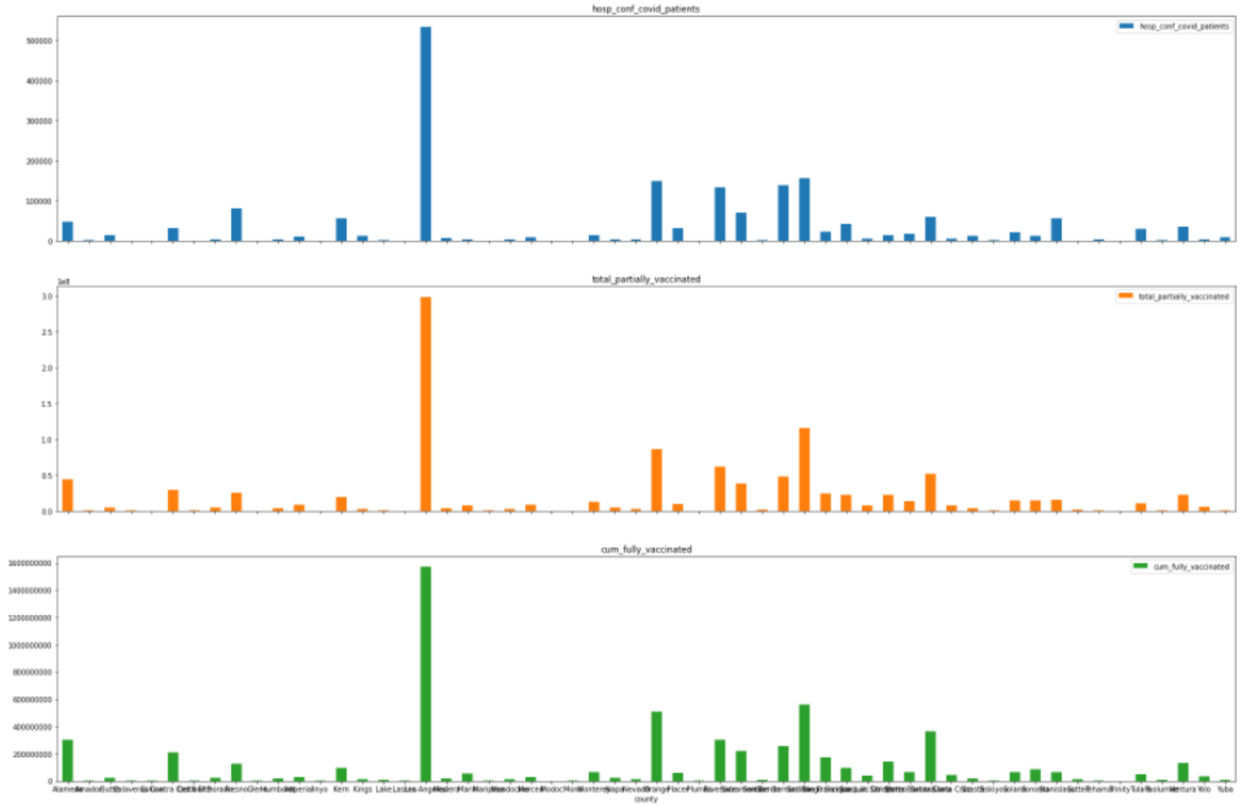
	vaccinated_cases	unvaccinated_cases	vaccinated_deaths	unvaccinated_deaths
vaccinated_cases	1.000000	0.710894	0.746601	0.024677
unvaccinated_cases	0.710894	1.000000	0.416666	0.582769
vaccinated_deaths	0.746601	0.416666	1.000000	0.136296
unvaccinated_deaths	0.024677	0.582769	0.136296	1.000000

- **Summary**
    - The exploratory data analysis revealed that both vaccinated and unvaccinated cases had their peaks in the month of August and followed a similar pattern at the latter part of the year. This makes it important to investigate the reason behind the seasonal pattern or trend in the two categories of cases. Also, the number of deaths for vaccinated and unvaccinated respectively show that vaccination reduced the death rate drastically. Additionally, the grouped deaths according to months show that the highest deaths occured in February and this is also worth investigating.

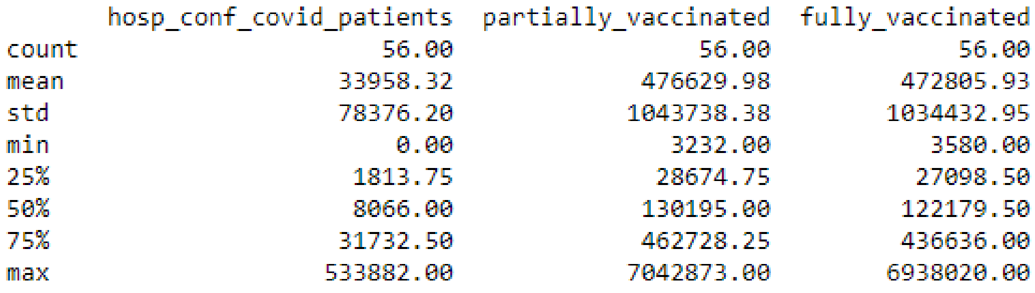
2.2.4. COVID-19 Vaccines and Hospitalized Cases by County Analysis

- **Data Sources**
    - Statewide Covid-19 Hospital County Data
    - Statewide COVID-19 Vaccines Administered By County
- **Cleaning Process**
    - We merged the two data frames COVID-19 hospital county data and COVID 19 vaccine progress by county on the county and the days because we may want to find the association between hospitalization and administered vaccines. We kept all the observations despite that vaccine data came later than hospital data. We renamed some of the columns for clarity. We kept all columns for the initial exploratory process. We created two new columns for month and weekday. We replace the missing values in all columns with the "Not Available" string. We can replace them with the mean for the columns where we can calculate mean, or interpolate the values, or drop them if necessary. For example vaccine doses earlier than when we have data for vaccines should all be zero.
- **Exploratory Data Analysis**
    - Process Step 1: We selected columns of interest and target feature.
        - Columns of Interest: County, hospitalized confirmed covid patients, partially vaccinated and fully vaccinated, month
        - Target Variables: Hospitalized confirmed covid patients, partially vaccinated and fully vaccinated
        - Preliminary Insights:
            - No null values in columns of interest.
            - Observations are from July 2020 to December 2021
    - Process Step 2: We plotted one-dimensional distributions of numerical columns and observed the overall shape of the data.

- The data were grouped by counties and summed by fully vaccinated, partially vaccinated, and hospitalized confirmed covid cases, respectively. Bar plots were generated to visualize the counties. There are similar trends in all the columns of interest. The county that had the highest hospitalized confirmed cases had the highest fully and partially vaccinated persons. For the next step, we created other visuals using California counties to check whether counties with high covid hospitalization cases are close to each other.

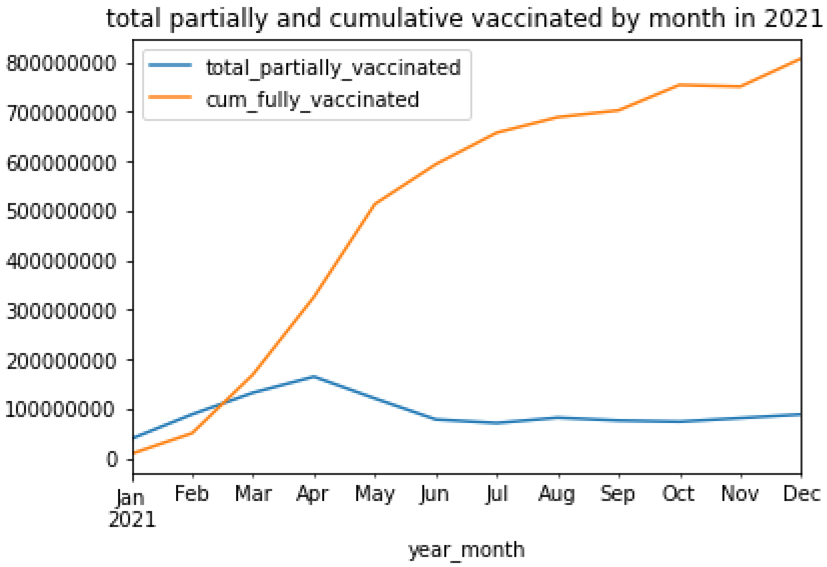


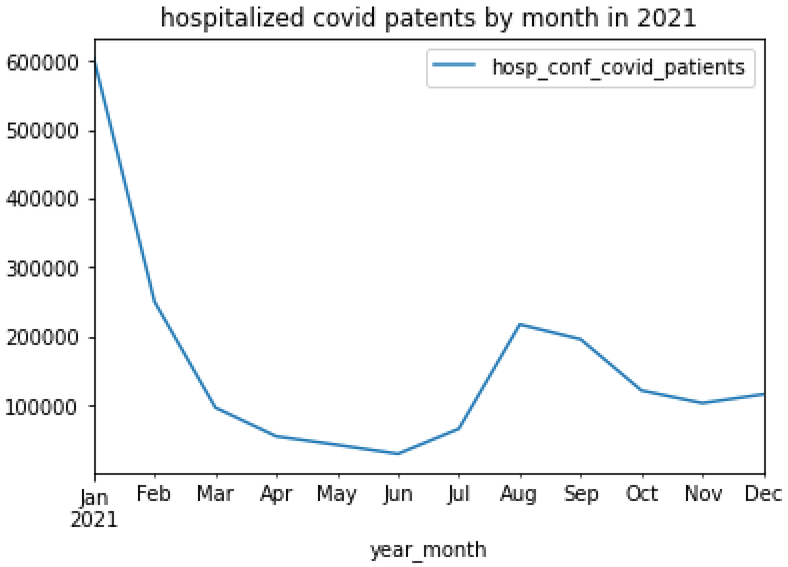
- We computed basic statistics for the numerical columns:



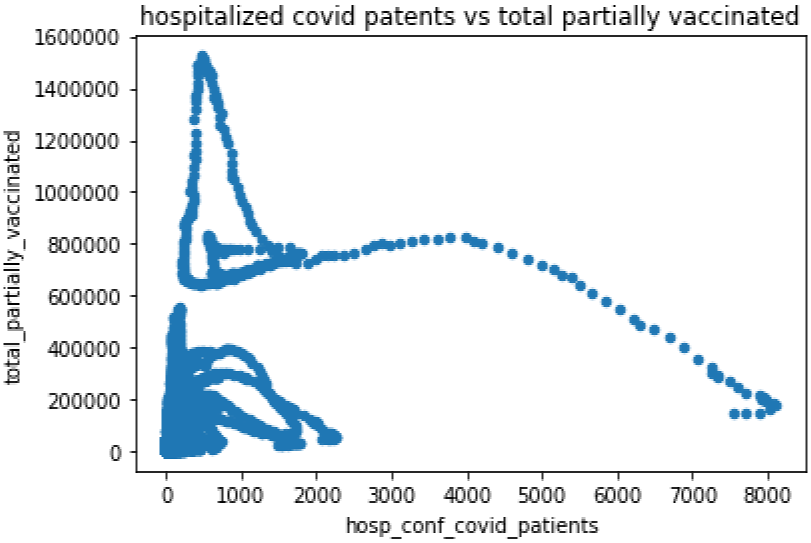
```
       hosp_conf_covid_patients  partially_vaccinated  fully_vaccinated
count                     56.00                 56.00             56.00
mean                   33958.32             476629.98         472805.93
std                    78376.20            1043738.38        1034432.95
min                        0.00               3232.00           3580.00
25%                     1813.75              28674.75          27098.50
50%                     8066.00             130195.00         122179.50
75%                    31732.50             462728.25         436636.00
max                   533882.00            7042873.00        6938020.00
```

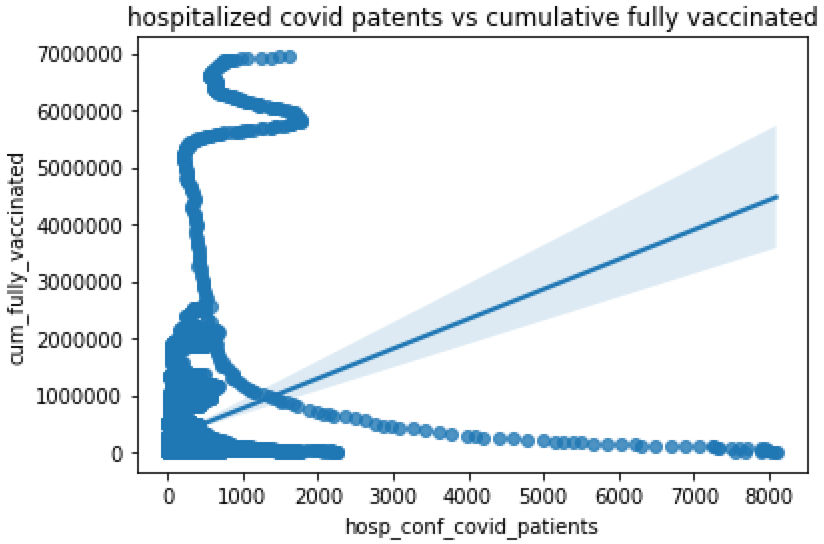
- Looking at seasonality trends, we discovered that the total partial vaccinations were highest at around march/April 2021 and after that dropped. Additionally, hospitalized covid cases were highest in Jan 2021.

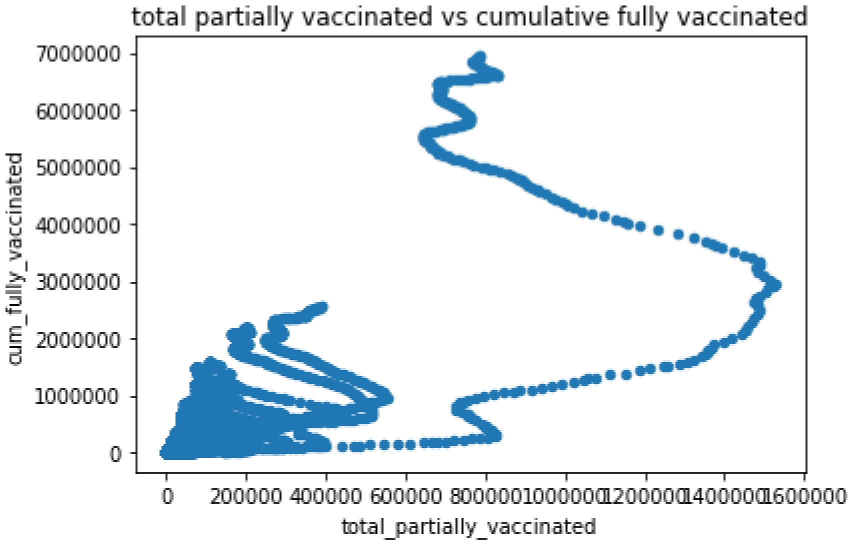




- Process Step 3: We plotted two-dimensional distributions of the variables of interest against the target variables.

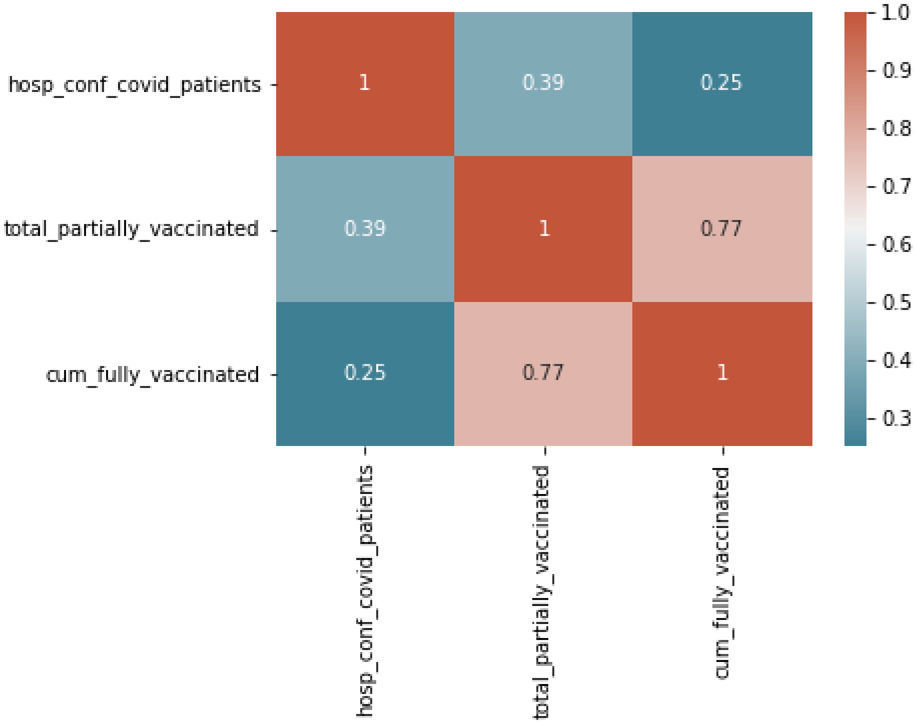

hospitalized covid patents vs total partially vaccinated
total_partially_vaccinated
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
0
1000
2000
3000
4000
5000
6000
7000
8000
hosp_conf_covid_patients

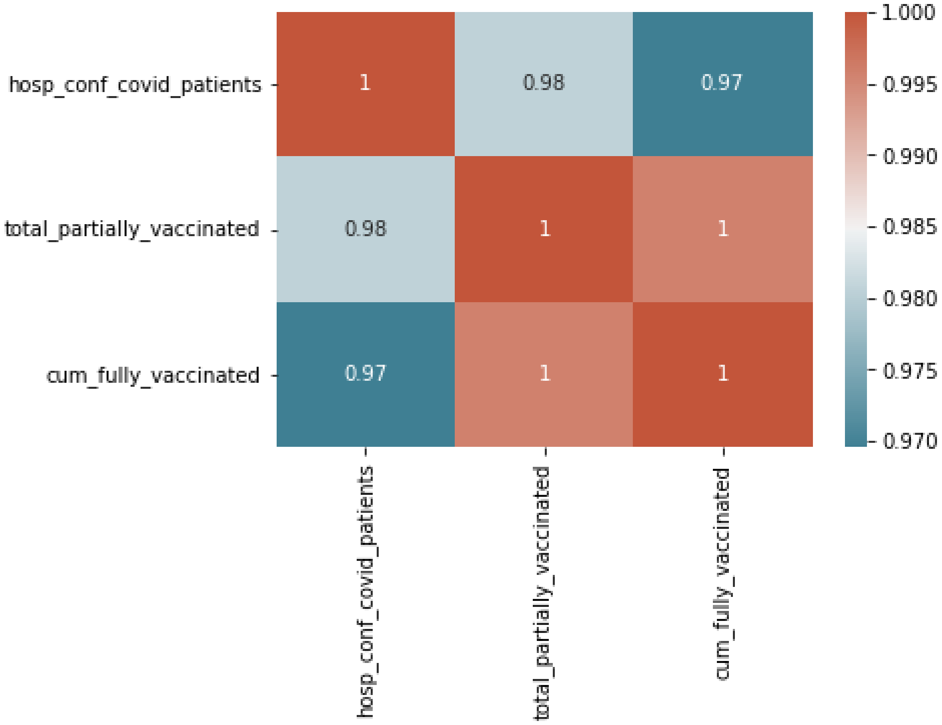




- As the number of fully/cumulative fully vaccinated increases, the hospitalized confirmed covid patients decrease. This was plausible, as this points to the fact that the vaccines could be helping to keep people out of the hospitals.

○ Process Step 4: We analyzed any correlations between the independent and dependent variables.

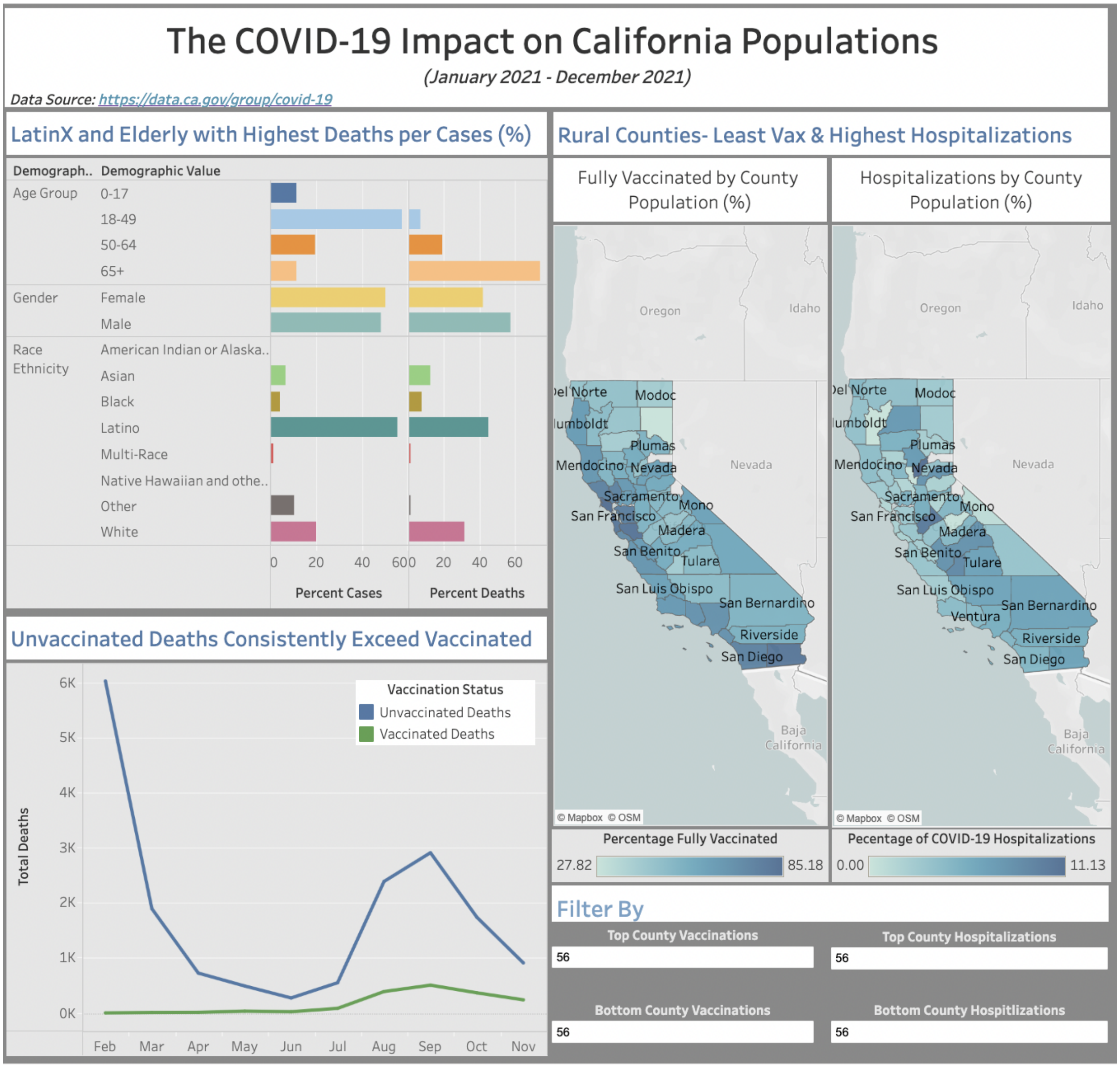




- All the variables of interest had a very high correlation with one another when grouped by county. We needed to check by county population since every variable has a positive relationship with other variables. It's most likely the visualization is being driven by counties with high populations. When considering the general data, cumulative fully vaccinated had the least correlation with hospitalized covid cases.

- **Summary**
  - Vaccines administered both fully and partially had a positive correlation with hospitalized covid cases in the raw data and when grouped by county, the correlation was almost perfect which is not intuitive for now. It is likely that the population of the various counties was also driving the results. Looking at the time trend analysis, hospitalized covid cases were highest at the beginning of the year (January) and lowest in the middle of the year (June). After that, it had another peak in August, but not up to half of what was experienced in January.

3. Dashboard



The COVID-19 Impact on California Populations

(January 2021 - December 2021)

Data Source: https://data.ca.gov/group/covid-19

LatinX and Elderly with Highest Deaths per Cases (%)

Rural Counties- Least Vax & Highest Hospitalizations

Unvaccinated Deaths Consistently Exceed Vaccinated

Filter By

Top County Vaccinations: 56

Top County Hospitalizations: 56

Bottom County Vaccinations: 56

Bottom County Hospitlizations: 56

Link to Dashboard:
https://public.tableau.com/shared/W6D6FCBKD?:display_count=n&:origin=viz_share_link

3.1. Use Case

The primary use case of our dashboard is to visualize the California populations impacted by COVID-19. The user can hover their mouse over sections of the visualizations to learn additional information about the affected populations. For the California counties visualization, the user can filter the counties by their ranking in vaccinations and hospitalization by entering a number ranging from 1 - 56 in the filter boxes below.

4. Conclusions & Future Work

From our analyses it was discovered that vaccines are effective towards reducing the number of Covid-19 cases and hospitalizations. This was observed by comparing the unvaccinated cases to the vaccinated cases across the year. Further analysis revealed that the people of age 65 and above have the highest mortality rate despite having a low acquisition rate of the Covid-19. This makes it important to pay attention to people of that age since they are more susceptible to death resulting from Covid-19. Additionally, the demographic data showed that the latinx group are more vulnerable to contracting the Covid-19 and have the highest mortality rate among other racial groups. It is therefore important to direct intervention efforts towards the latinx to ensure that the number of cases recorded by the latinx group is curtailed.

Since the data at county level had limited variables predictive modeling could not be carried out on data. Future work will entail gathering more data especially at the county level to create models that will be able to predict the likelihood of an individual contracting Covid-19 based on the features or variables of the individual.