

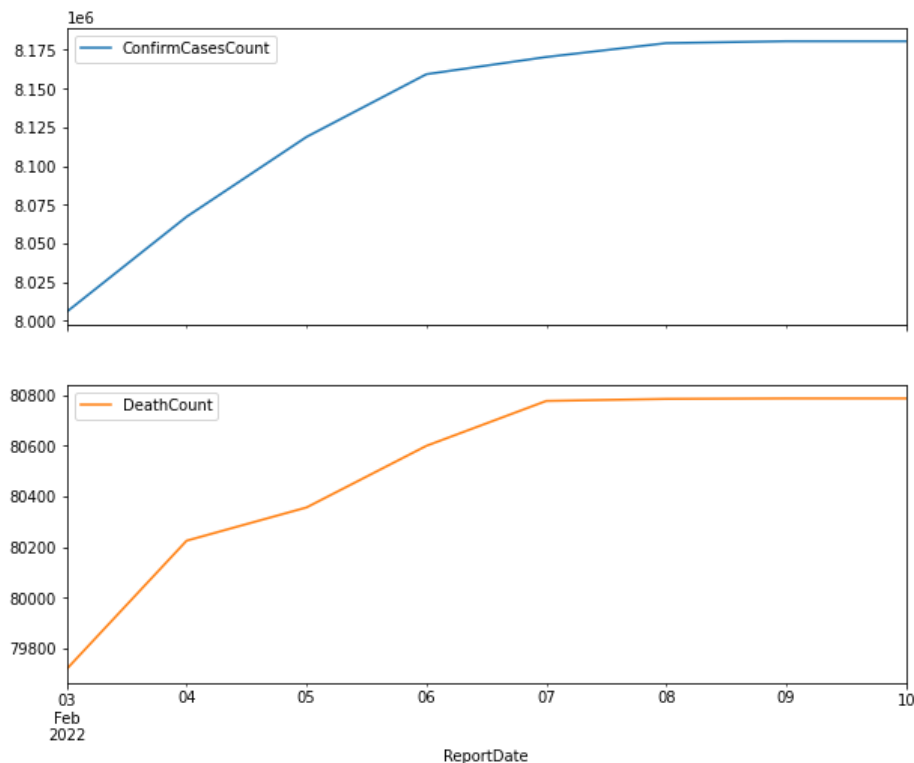
California

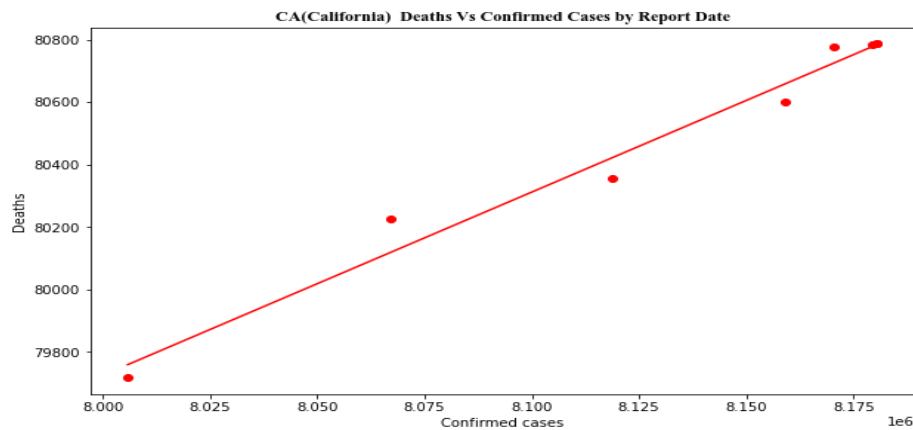
The confirmed COVID-19 cases and the death counts were assessed for California for the week of February 3rd– February 10th, 2022. From the analysis, it was found that the confirmed covid cases and deaths due to COVID-19 were increasing steadily in California. Additionally, the death cases rose from 79800 to 80800 which is the drastic change. The greatest number of confirmed covid cases of 8180551 people and the greatest number of deaths of 80786 people due to covid in this period. As we could see from the above graph, the cases rose from very drastically in the mid-week of February. Also, the death count seems to be increased in the mid of the week. The scatter plot for Death Vs Confirmed cases looks linear.

```
CovidCADF = covid_last_week.copy()
CovidCADF = CovidCADF[['ConfirmCasesCount', 'DeathCount']]
CovidCADF = CovidCADF.reset_index()
```

```
CovidCAAggDF = CovidCADF[['ConfirmCasesCount', 'DeathCount']].groupby(CovidCADF['ReportDate']).sum() #sum of cases grouped by date
CovidCAAggDF
```

ReportDate	ConfirmCasesCount	DeathCount
2022-02-03	8005751	79719
2022-02-04	8067176	80225
2022-02-05	8118735	80356
2022-02-06	8159238	80599
2022-02-07	8170356	80776
2022-02-08	8179312	80784
2022-02-09	8180551	80786
2022-02-10	8180551	80786





Enrichment Data:

Socio-Economic ACS data: Employment and Housing

The Social Economic dataset is related to the employment of 2020 by county. This dataset contains 20 columns. It has the state, county, and few employment parameters and housing parameters which decides the employment and housing analytics.

For merging this dataset with the super COVID-19 dataset, first, a mapping was created on the state column of the ACS socio-economic County dataset that mapped the state name and the county names with its abbreviation. Then, the current state values were overridden with the correlating abbreviation. This was done so that the state abbreviations would match that of the super COVID-19 dataset. Next, the state and county column name in the dataset was renamed to State and County Name so that it would match with the same name in the super COVID-19 dataset. Secondly, the string values were trimmed to ensure that the matches for the merge were not missed by leading or trailing spaces. Finally, the merge was performed using the State and Column Name columns by using left joins of both datasets. There were many controversies on the mismanagement of the COVID-19 pandemic from the

COVID-19 analysis

The enrichment data can help confirm two hypotheses in the analysis of COVID-19 spread.

- Hypothesis I: ACS socio-economic inequality between regions (states or counties), and industry sectors is highly correlated to the level of spread of COVID-19 over time. In particular, the COVID-19 spread has affected a severe drop in employment of the contact business industry, compared to the non-contact business industry over time.
- Hypothesis II: Wage inequality between regions (states or counties), and industry sectors is highly associated with the level of spread of COVID-19 and its spread contributes to the wage inequality.