In [1]:

```python
# Names:- Smit Modi, Smit Patel
import pandas as pd
#pd.to_datetime('2020-12-15 3:45pm')
import numpy as np
import matplotlib.pyplot as plt
```

In [ ]:

```python
# DATA CLEANING/GATHERING
```

In [2]:

```python
df =  pd.read_csv("CA_Vaccine_Data.csv")
df
```

Out[2]:

	county	administered_date	total_doses	cumulative_total_doses	pfizer_doses	cumulative_pfizer_doses	moderna_doses	cumulative_moderna_doses	jj_doses	cumulative_jj_doses	partially_vaccinated	total_partially_vaccinated	fully_vaccinated	cun
0	All CA Counties	2020-12-15	1238	1238	1238	1238	0	0	0	0	1238	1238	0	
1	All CA Counties	2020-12-16	9021	10259	9021	10259	0	0	0	0	9021	10259	0	
2	All CA Counties	2020-12-17	27137	37396	27137	37396	0	0	0	0	27139	37398	0	
3	All CA Counties	2020-12-18	46006	83402	45929	83325	77	77	0	0	46006	83403	1	
4	All CA Counties	2020-12-19	25773	109175	25740	109065	33	110	0	0	25772	109174	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
15309	Yuba	2021-08-14	138	53901	90	20333	47	30134	1	3434	75	4725	61	
15310	Yuba	2021-08-15	85	53986	57	20390	28	30162	0	3434	61	4764	22	
15311	Yuba	2021-08-16	216	54202	100	20490	96	30258	20	3454	136	4844	76	
15312	Yuba	2021-08-17	189	54391	96	20586	88	30346	5	3459	113	4891	71	
15313	Yuba	2021-08-18	130	54521	59	20645	59	30405	12	3471	77	4928	52	

15314 rows × 17 columns

In [3]:

```python
# DATA CLEANING:-
# RENAME ALL THE COLUMNS TO PROPERLY VISUALIZE!
```

In [4]:

```python
df.rename(columns={'county':'California Counties','administered_date':'Dose Date YYY-DD-MM', 'total_doses':'Total Doses', 'cumulative_total_doses':'Cumulative Total',
                   'pfizer_doses': 'PFIZER', 'cumulative_pfizer_doses':'PFIZER TOTAL','moderna_doses':'MODERNA','cumulative_moderna_doses':'MODERNA TOTAL',
                     'jj_doses':'JANSSEN','cumulative_jj_doses':'JANSSEN TOTAL','partially_vaccinated':'Partially Vaccinated',
                       'total_partially_vaccinated':'TOTAL Partially Vaccinated','fully_vaccinated':'Fully Vaccinated','cumulative_fully_vaccinated':'TOTAL Fully Vaccinated',
                         'at_least_one_dose':'One Dose Taken','cumulative_at_least_one_dose':'TOTAL One Doses Taken'}, inplace = True)
df
```

Out[4]:

	California Counties	Dose Date YYY-DD-MM	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken	california_flag

	California Counties	Dose Date YYY-DD-MM	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken	california_flag
0	All CA Counties	2020-12-15	1238	1238	1238	1238	0	0	0	0	1238	1238	0	0	1238	1238	NaN
1	All CA Counties	2020-12-16	9021	10259	9021	10259	0	0	0	0	9021	10259	0	0	9021	10259	NaN
2	All CA Counties	2020-12-17	27137	37396	27137	37396	0	0	0	0	27139	37398	0	0	27139	37398	NaN
3	All CA Counties	2020-12-18	46006	83402	45929	83325	77	77	0	0	46006	83403	1	1	46006	83404	NaN
4	All CA Counties	2020-12-19	25773	109175	25740	109065	33	110	0	0	25772	109174	1	2	25772	109176	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
15309	Yuba	2021-08-14	138	53901	90	20333	47	30134	1	3434	75	4725	61	26202	76	30927	California
15310	Yuba	2021-08-15	85	53986	57	20390	28	30162	0	3434	61	4764	22	26224	61	30988	California
15311	Yuba	2021-08-16	216	54202	100	20490	96	30258	20	3454	136	4844	76	26300	156	31144	California
15312	Yuba	2021-08-17	189	54391	96	20586	88	30346	5	3459	113	4891	71	26371	118	31262	California
15313	Yuba	2021-08-18	130	54521	59	20645	59	30405	12	3471	77	4928	52	26423	89	31351	California

15314 rows × 17 columns

In [5]:

```
# Dropping flag column becuase it doens't provide any use
# Also taking out any rows that aren't a part of just California

df.drop(["california_flag"],axis=1, inplace = True)
df.set_index('California Counties', inplace = True)
df.drop(labels = ["Unknown", "All CA and Non-CA Counties", "Outside California"], axis=0, inplace = True)
df.reset_index(inplace = True)
df
```

Out[5]:

	California Counties	Dose Date YYY-DD-MM	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken
0	All CA Counties	2020-12-15	1238	1238	1238	1238	0	0	0	0	1238	1238	0	0	1238	1238
1	All CA Counties	2020-12-16	9021	10259	9021	10259	0	0	0	0	9021	10259	0	0	9021	10259
2	All CA Counties	2020-12-17	27137	37396	27137	37396	0	0	0	0	27139	37398	0	0	27139	37398
3	All CA Counties	2020-12-18	46006	83402	45929	83325	77	77	0	0	46006	83403	1	1	46006	83404
4	All CA Counties	2020-12-19	25773	109175	25740	109065	33	110	0	0	25772	109174	1	2	25772	109176
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14568	Yuba	2021-08-14	138	53901	90	20333	47	30134	1	3434	75	4725	61	26202	76	30927
14569	Yuba	2021-08-15	85	53986	57	20390	28	30162	0	3434	61	4764	22	26224	61	30988
14570	Yuba	2021-08-16	216	54202	100	20490	96	30258	20	3454	136	4844	76	26300	156	31144
14571	Yuba	2021-08-17	189	54391	96	20586	88	30346	5	3459	113	4891	71	26371	118	31262
14572	Yuba	2021-08-18	130	54521	59	20645	59	30405	12	3471	77	4928	52	26423	89	31351

14573 rows × 16 columns

In [6]:

```
# Printing all the unique values of the dataframe to gain a better understanding of the data
for col in df:
    print(df[col].unique())
```

```
['All CA Counties' 'Alpine' 'Amador' 'Butte' 'Calaveras' 'Colusa'
 'Del Norte' 'Glenn' 'Humboldt' 'Imperial' 'Inyo' 'Kings' 'Lake' 'Lassen'
 'Madera' 'Marin' 'Mariposa' 'Mendocino' 'Merced' 'Modoc' 'Mono'
 'Monterey' 'Napa' 'Nevada' 'Plumas' 'San Benito' 'Santa Barbara' 'Shasta'
 'Sierra' 'Siskiyou' 'Stanislaus' 'Sutter' 'Tehama' 'Trinity' 'Tulare'
 'Tuolumne' 'Yuba' 'Alameda' 'Contra Costa' 'El Dorado' 'Fresno' 'Kern'
 'Los Angeles' 'Orange' 'Placer' 'Riverside' 'Sacramento' 'San Bernardino'
 'San Diego' 'San Francisco' 'San Joaquin' 'San Luis Obispo' 'San Mateo'
 'Santa Clara' 'Santa Cruz' 'Solano' 'Sonoma' 'Ventura' 'Yolo']
['2020-12-15' '2020-12-16' '2020-12-17' '2020-12-18' '2020-12-19'
 '2020-12-20' '2020-12-21' '2020-12-22' '2020-12-23' '2020-12-24'
 '2020-12-25' '2020-12-26' '2020-12-27' '2020-12-28' '2020-12-29'
 '2020-12-30' '2020-12-31' '2021-01-01' '2021-01-02' '2021-01-03'
 '2021-01-04' '2021-01-05' '2021-01-06' '2021-01-07' '2021-01-08'
 '2021-01-09' '2021-01-10' '2021-01-11' '2021-01-12' '2021-01-13'
 '2021-01-14' '2021-01-15' '2021-01-16' '2021-01-17' '2021-01-18'
 '2021-01-19' '2021-01-20' '2021-01-21' '2021-01-22' '2021-01-23'
 '2021-01-24' '2021-01-25' '2021-01-26' '2021-01-27' '2021-01-28'
 '2021-01-29' '2021-01-30' '2021-01-31' '2021-02-01' '2021-02-02'
 '2021-02-03' '2021-02-04' '2021-02-05' '2021-02-06' '2021-02-07'
 '2021-02-08' '2021-02-09' '2021-02-10' '2021-02-11' '2021-02-12'
 '2021-02-13' '2021-02-14' '2021-02-15' '2021-02-16' '2021-02-17'
 '2021-02-18' '2021-02-19' '2021-02-20' '2021-02-21' '2021-02-22'
 '2021-02-23' '2021-02-24' '2021-02-25' '2021-02-26' '2021-02-27'
 '2021-02-28' '2021-03-01' '2021-03-02' '2021-03-03' '2021-03-04'
 '2021-03-05' '2021-03-06' '2021-03-07' '2021-03-08' '2021-03-09'
 '2021-03-10' '2021-03-11' '2021-03-12' '2021-03-13' '2021-03-14'
 '2021-03-15' '2021-03-16' '2021-03-17' '2021-03-18' '2021-03-19'
 '2021-03-20' '2021-03-21' '2021-03-22' '2021-03-23' '2021-03-24'
 '2021-03-25' '2021-03-26' '2021-03-27' '2021-03-28' '2021-03-29'
 '2021-03-30' '2021-03-31' '2021-04-01' '2021-04-02' '2021-04-03'
 '2021-04-04' '2021-04-05' '2021-04-06' '2021-04-07' '2021-04-08'
 '2021-04-09' '2021-04-10' '2021-04-11' '2021-04-12' '2021-04-13'
 '2021-04-14' '2021-04-15' '2021-04-16' '2021-04-17' '2021-04-18'
 '2021-04-19' '2021-04-20' '2021-04-21' '2021-04-22' '2021-04-23'
 '2021-04-24' '2021-04-25' '2021-04-26' '2021-04-27' '2021-04-28'
 '2021-04-29' '2021-04-30' '2021-05-01' '2021-05-02' '2021-05-03'
 '2021-05-04' '2021-05-05' '2021-05-06' '2021-05-07' '2021-05-08'
 '2021-05-09' '2021-05-10' '2021-05-11' '2021-05-12' '2021-05-13'
 '2021-05-14' '2021-05-15' '2021-05-16' '2021-05-17' '2021-05-18'
 '2021-05-19' '2021-05-20' '2021-05-21' '2021-05-22' '2021-05-23'
 '2021-05-24' '2021-05-25' '2021-05-26' '2021-05-27' '2021-05-28'
 '2021-05-29' '2021-05-30' '2021-05-31' '2021-06-01' '2021-06-02'
 '2021-06-03' '2021-06-04' '2021-06-05' '2021-06-06' '2021-06-07'
 '2021-06-08' '2021-06-09' '2021-06-10' '2021-06-11' '2021-06-12'
 '2021-06-13' '2021-06-14' '2021-06-15' '2021-06-16' '2021-06-17'
 '2021-06-18' '2021-06-19' '2021-06-20' '2021-06-21' '2021-06-22'
 '2021-06-23' '2021-06-24' '2021-06-25' '2021-06-26' '2021-06-27'
 '2021-06-28' '2021-06-29' '2021-06-30' '2021-07-01' '2021-07-02'
 '2021-07-03' '2021-07-04' '2021-07-05' '2021-07-06' '2021-07-07'
 '2021-07-08' '2021-07-09' '2021-07-10' '2021-07-11' '2021-07-12'
 '2021-07-13' '2021-07-14' '2021-07-15' '2021-07-16' '2021-07-17'
 '2021-07-18' '2021-07-19' '2021-07-20' '2021-07-21' '2021-07-22'
 '2021-07-23' '2021-07-24' '2021-07-25' '2021-07-26' '2021-07-27'
 '2021-07-28' '2021-07-29' '2021-07-30' '2021-07-31' '2021-08-01'
 '2021-08-02' '2021-08-03' '2021-08-04' '2021-08-05' '2021-08-06'
 '2021-08-07' '2021-08-08' '2021-08-09' '2021-08-10' '2021-08-11'
 '2021-08-12' '2021-08-13' '2021-08-14' '2021-08-15' '2021-08-16'
 '2021-08-17' '2021-08-18']
```

```
[ 1238  9021 27137 ...   489   386   744]
[ 1238 10259 37396 ... 54202 54391 54521]
[ 1238  9021 27137 ...  1000  1224   302]
[ 1238 10259 37396 ... 20333 20390 20645]
[   0   77   33 ...  871 1733  669]
[    0    77   110 ... 30162 30258 30346]
[  0  33  15 ... 291 292 355]
[   0   33   48 ... 3454 3459 3471]
[ 1238  9021 27139 ...   811  1121  1031]
[ 1238 10259 37398 ...  4725  4764  4928]
[   0    1    2 ... 1502 1863 1977]
[    0     1     2 ... 26300 26371 26423]
[ 1238  9021 27139 ...  2194  1117  1038]
[ 1238 10259 37398 ... 31144 31262 31351]
```

In [7]:

```
import matplotlib
%matplotlib inline
```

In [ ]:

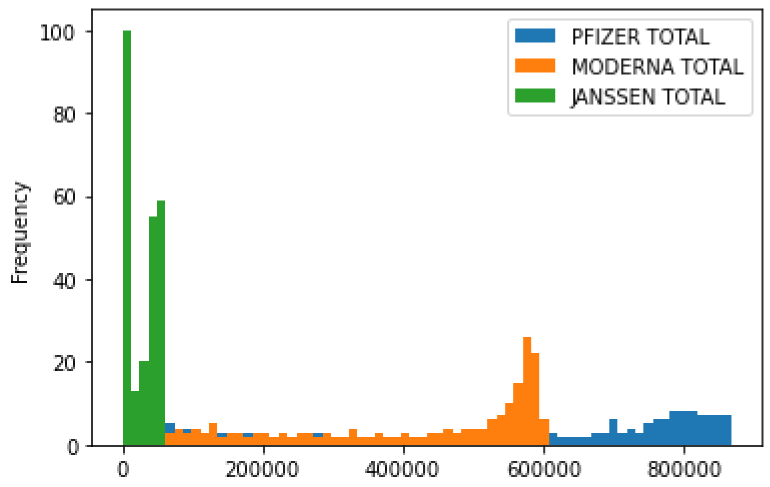
```
# DATA VISULIZATION
```

In [8]:

```
df.groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.hist(bins=70)
```

Out[8]:

```
<AxesSubplot:ylabel='Frequency'>
```



In [9]:

```
# Checking the data type
df.dtypes
```

Out[9]:

```
California Counties           object
Dose Date YYY-DD-MM           object
Total Doses                    int64
Cumulative Total               int64
PFIZER                         int64
PFIZER TOTAL                   int64
MODERNA                        int64
MODERNA TOTAL                  int64
JANSSEN                        int64
JANSSEN TOTAL                  int64
Partially Vaccinated           int64
TOTAL Partially Vaccinated     int64
Fully Vaccinated               int64
TOTAL Fully Vaccinated         int64
One Dose Taken                 int64
TOTAL One Doses Taken          int64
dtype: object
```

In [10]:

```
# Setting 'Dose Date' as an index
df = df.set_index('Dose Date YYY-DD-MM')
df.index
```

Out[10]:

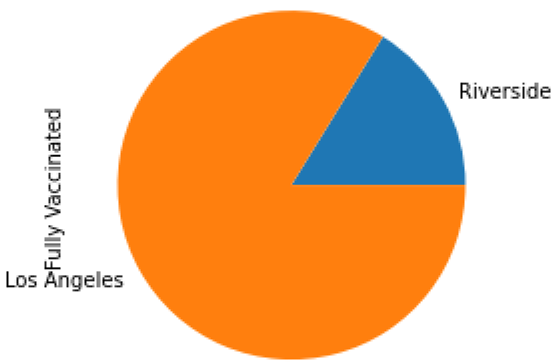
```
Index(['2020-12-15', '2020-12-16', '2020-12-17', '2020-12-18', '2020-12-19',
       '2020-12-20', '2020-12-21', '2020-12-22', '2020-12-23', '2020-12-24',
       ...
       '2021-08-09', '2021-08-10', '2021-08-11', '2021-08-12', '2021-08-13',
       '2021-08-14', '2021-08-15', '2021-08-16', '2021-08-17', '2021-08-18'],
      dtype='object', name='Dose Date YYY-DD-MM', length=14573)
```

In [12]:

```
# Comparing how many people are fully vaccinated in Riverside and Los Angeleous counties.
df.groupby(["California Counties"])["Fully Vaccinated"].mean().loc[["Riverside", "Los Angeles"]].plot.pie()
```

Out[12]:

```
<AxesSubplot:ylabel='Fully Vaccinated'>
```

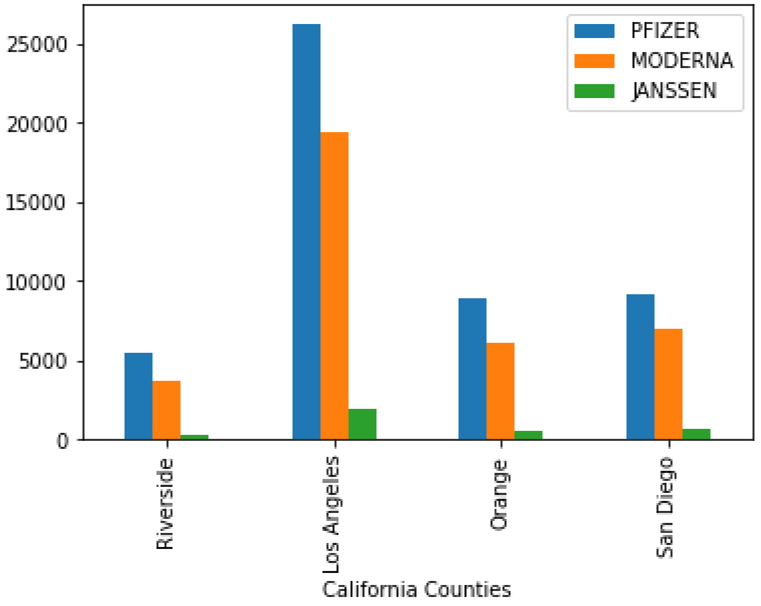


In [13]:

```
# Here we are comparing four CA counties to see which vaccine people have taken the most regardless of dose count.
df.groupby(["California Counties"])[["PFIZER","MODERNA","JANSSEN"]].mean().loc[["Riverside", "Los Angeles", "Orange", "San Diego"]].plot.bar()
```

Out[13]:

```
<AxesSubplot:xlabel='California Counties'>
```

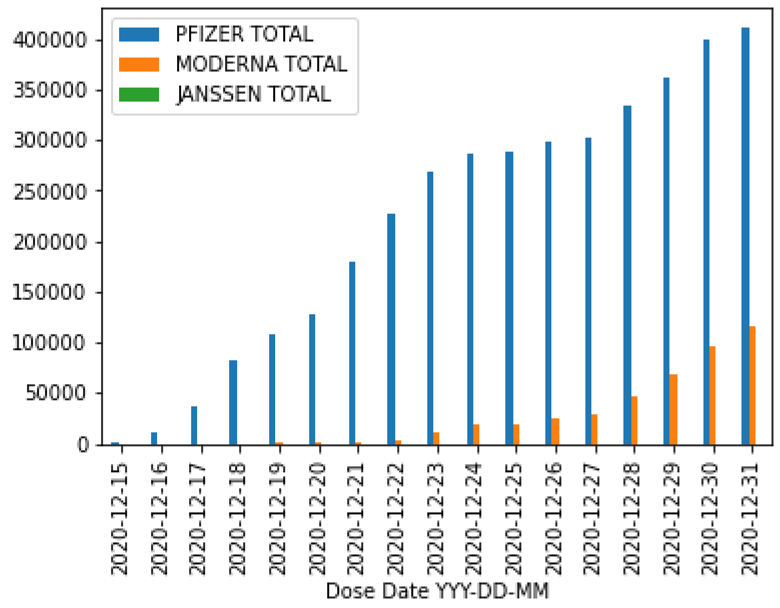


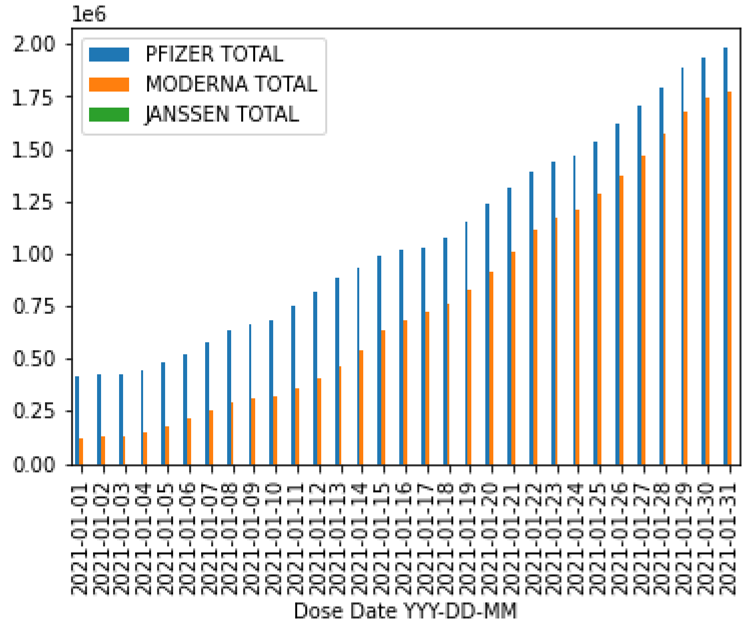
In [14]:

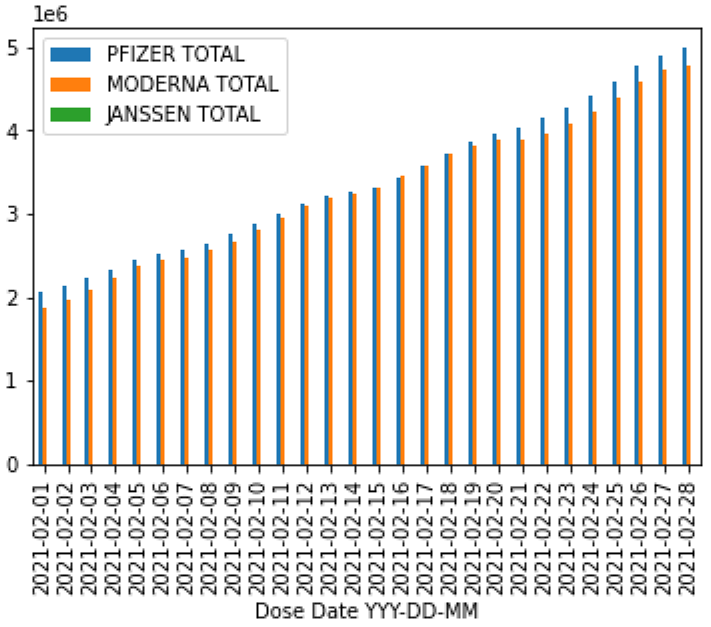
```
# These charts show how many of each type of vaccine was administered every
# day August 18, 2021
```

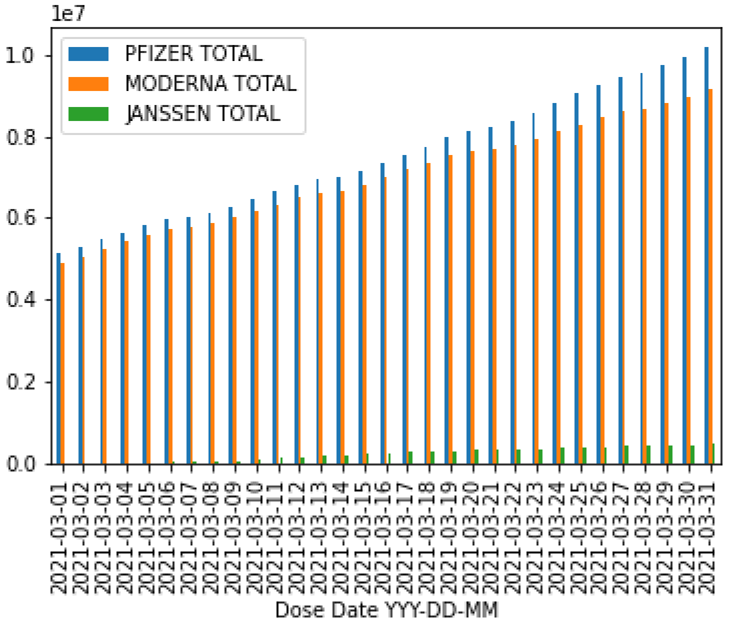
```
contain_values = df[df['California Counties'].str.contains('All CA Counties')]
contain_values.reset_index(inplace=True)

contain_values2 = contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2020-12')]
contain_values2.groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-01')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-02')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-03')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-04')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-05')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-06')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-07')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-08')].groupby('Dose Date YYY-DD-MM')[['PFIZER TOTAL', 'MODERNA TOTAL', 'JANSSEN TOTAL']].mean().plot.bar()
```

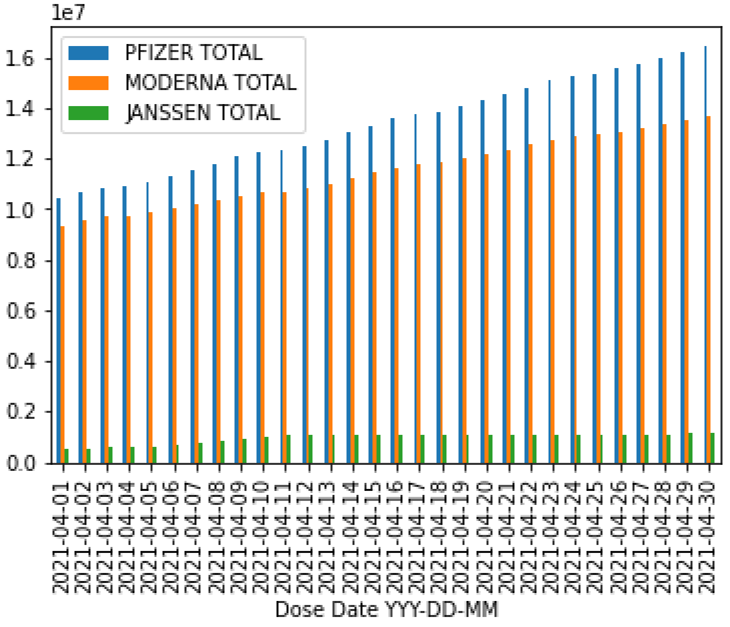
```
Out[14]: <AxesSubplot:xlabel='Dose Date YYY-DD-MM'>
```

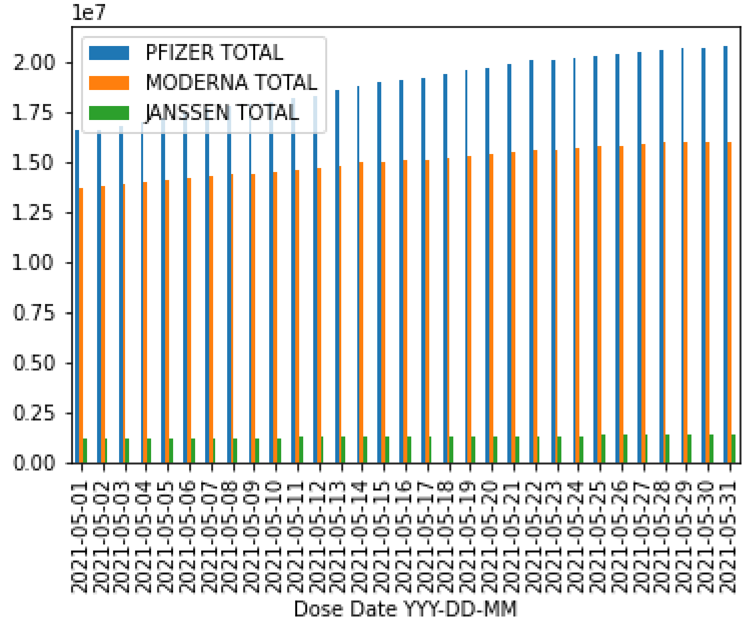


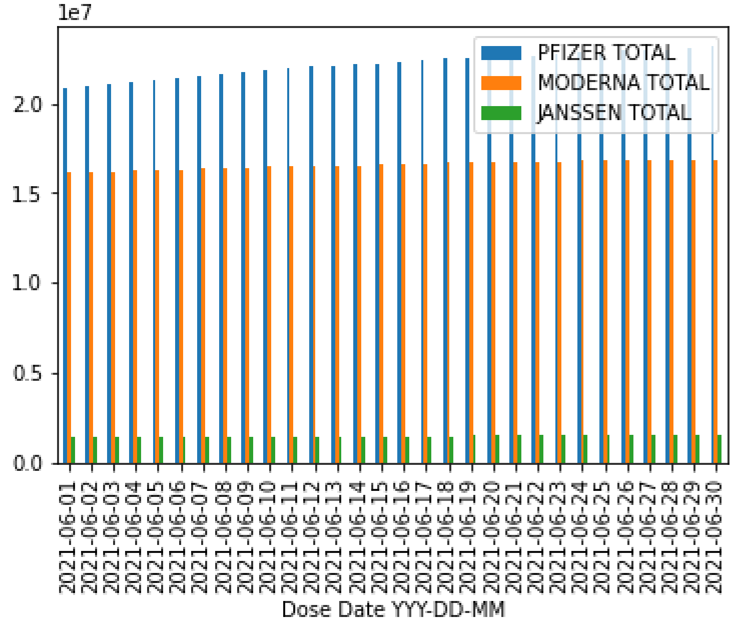


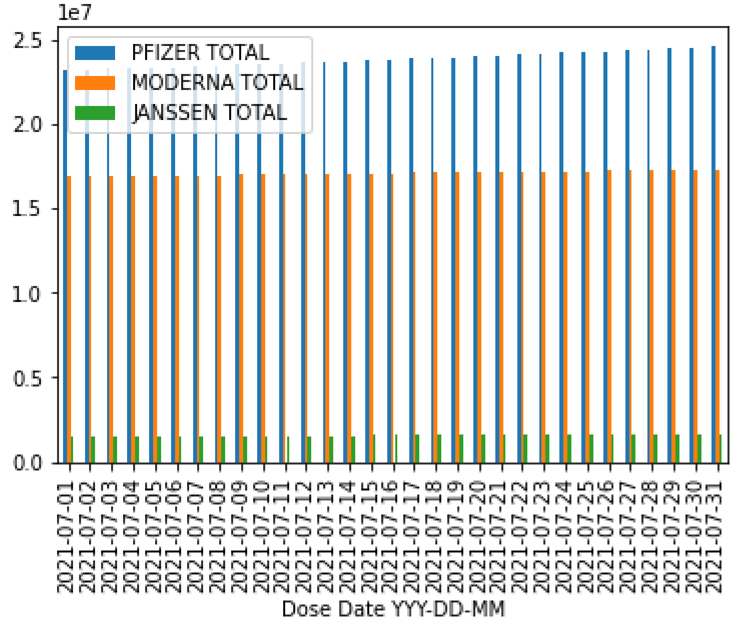

1e6
5
4
3
2
1
0
PFIZER TOTAL
MODERNA TOTAL
JANSSEN TOTAL
2021-02-01
2021-02-02
2021-02-03
2021-02-04
2021-02-05
2021-02-06
2021-02-07
2021-02-08
2021-02-09
2021-02-10
2021-02-11
2021-02-12
2021-02-13
2021-02-14
2021-02-15
2021-02-16
2021-02-17
2021-02-18
2021-02-19
2021-02-20
2021-02-21
2021-02-22
2021-02-23
2021-02-24
2021-02-25
2021-02-26
2021-02-27
2021-02-28
Dose Date YYY-DD-MM

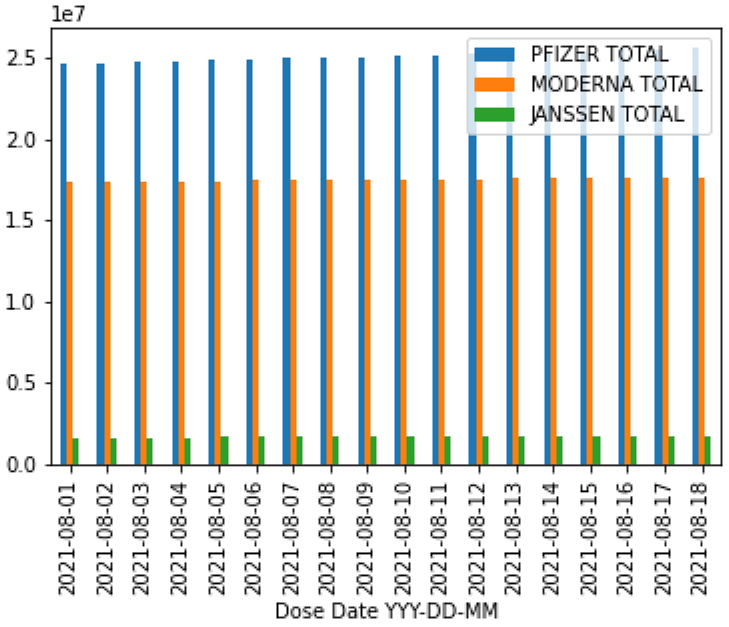

1e7
1.0
0.8
0.6
0.4
0.2
0.0
PFIZER TOTAL
MODERNA TOTAL
JANSSEN TOTAL
2021-03-01
2021-03-02
2021-03-03
2021-03-04
2021-03-05
2021-03-06
2021-03-07
2021-03-08
2021-03-09
2021-03-10
2021-03-11
2021-03-12
2021-03-13
2021-03-14
2021-03-15
2021-03-16
2021-03-17
2021-03-18
2021-03-19
2021-03-20
2021-03-21
2021-03-22
2021-03-23
2021-03-24
2021-03-25
2021-03-26
2021-03-27
2021-03-28
2021-03-29
2021-03-30
2021-03-31
Dose Date YYY-DD-MM


1e7
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
PFIZER TOTAL
MODERNA TOTAL
JANSSEN TOTAL
2021-04-01
2021-04-02
2021-04-03
2021-04-04
2021-04-05
2021-04-06
2021-04-07
2021-04-08
2021-04-09
2021-04-10
2021-04-11
2021-04-12
2021-04-13
2021-04-14
2021-04-15
2021-04-16
2021-04-17
2021-04-18
2021-04-19
2021-04-20
2021-04-21
2021-04-22
2021-04-23
2021-04-24
2021-04-25
2021-04-26
2021-04-27
2021-04-28
2021-04-29
2021-04-30
Dose Date YYY-DD-MM


1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
PFIZER TOTAL
MODERNA TOTAL
JANSSEN TOTAL
2021-05-01
2021-05-02
2021-05-03
2021-05-04
2021-05-05
2021-05-06
2021-05-07
2021-05-08
2021-05-09
2021-05-10
2021-05-11
2021-05-12
2021-05-13
2021-05-14
2021-05-15
2021-05-16
2021-05-17
2021-05-18
2021-05-19
2021-05-20
2021-05-21
2021-05-22
2021-05-23
2021-05-24
2021-05-25
2021-05-26
2021-05-27
2021-05-28
2021-05-29
2021-05-30
2021-05-31
Dose Date YYY-DD-MM


1e7
2.0
1.5
1.0
0.5
0.0
PFIZER TOTAL
MODERNA TOTAL
JANSSEN TOTAL
2021-06-01
2021-06-02
2021-06-03
2021-06-04
2021-06-05
2021-06-06
2021-06-07
2021-06-08
2021-06-09
2021-06-10
2021-06-11
2021-06-12
2021-06-13
2021-06-14
2021-06-15
2021-06-16
2021-06-17
2021-06-18
2021-06-19
2021-06-20
2021-06-21
2021-06-22
2021-06-23
2021-06-24
2021-06-25
2021-06-26
2021-06-27
2021-06-28
2021-06-29
2021-06-30
Dose Date YYY-DD-MM


1e7
2.5
2.0
1.5
1.0
0.5
0.0
PFIZER TOTAL
MODERNA TOTAL
JANSSEN TOTAL
2021-07-01
2021-07-02
2021-07-03
2021-07-04
2021-07-05
2021-07-06
2021-07-07
2021-07-08
2021-07-09
2021-07-10
2021-07-11
2021-07-12
2021-07-13
2021-07-14
2021-07-15
2021-07-16
2021-07-17
2021-07-18
2021-07-19
2021-07-20
2021-07-21
2021-07-22
2021-07-23
2021-07-24
2021-07-25
2021-07-26
2021-07-27
2021-07-28
2021-07-29
2021-07-30
2021-07-31
Dose Date YYY-DD-MM

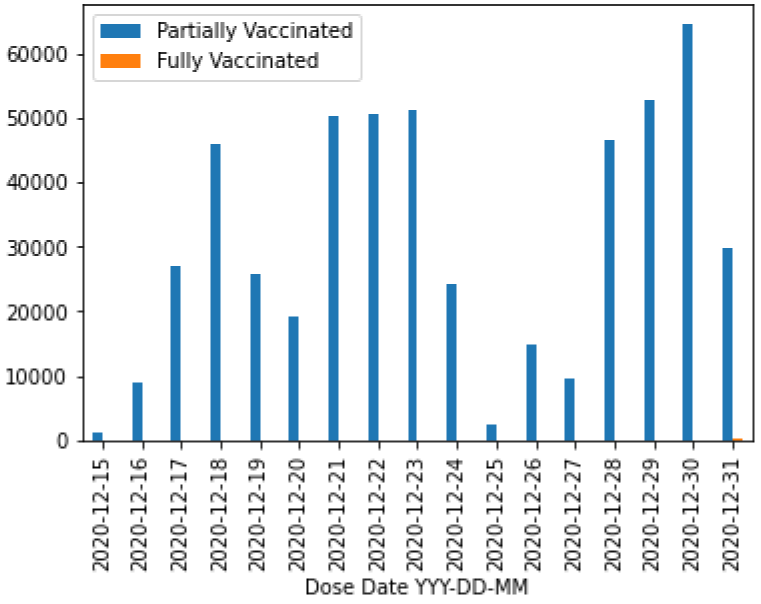


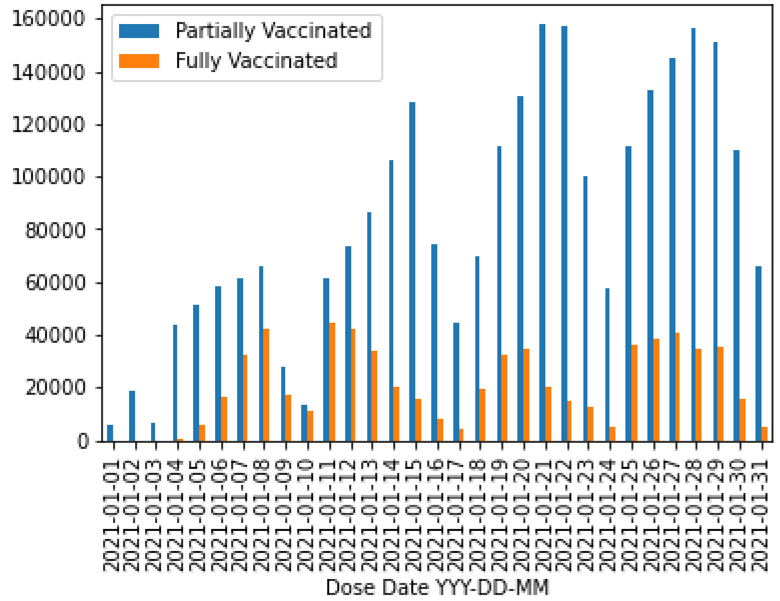
In [15]:

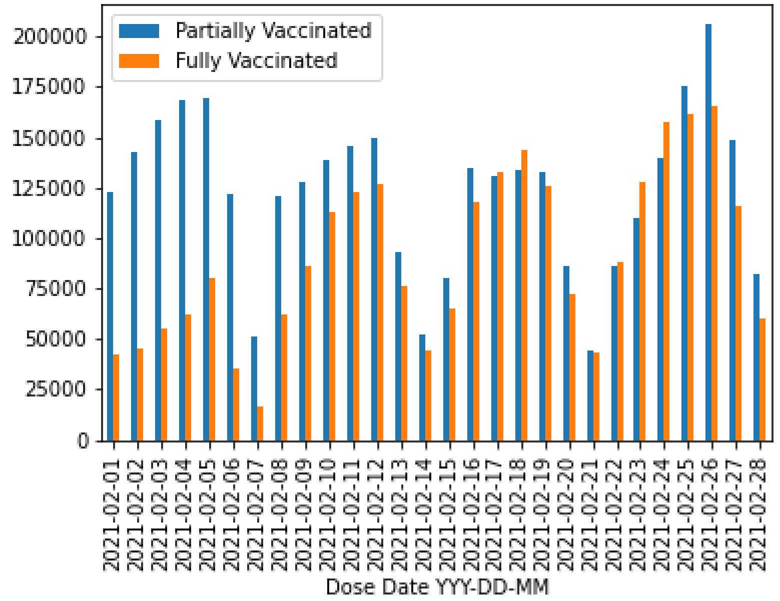
```
# This data focuses on showing the change of how many people are
# partially vaccinated and fully vaccinated over time

contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2020-12')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-01')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-02')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-03')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-04')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-05')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-06')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-07')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
contain_values[contain_values['Dose Date YYY-DD-MM'].str.contains('2021-08')].groupby('Dose Date YYY-DD-MM')[['Partially Vaccinated', 'Fully Vaccinated']].mean().plot.bar()
```

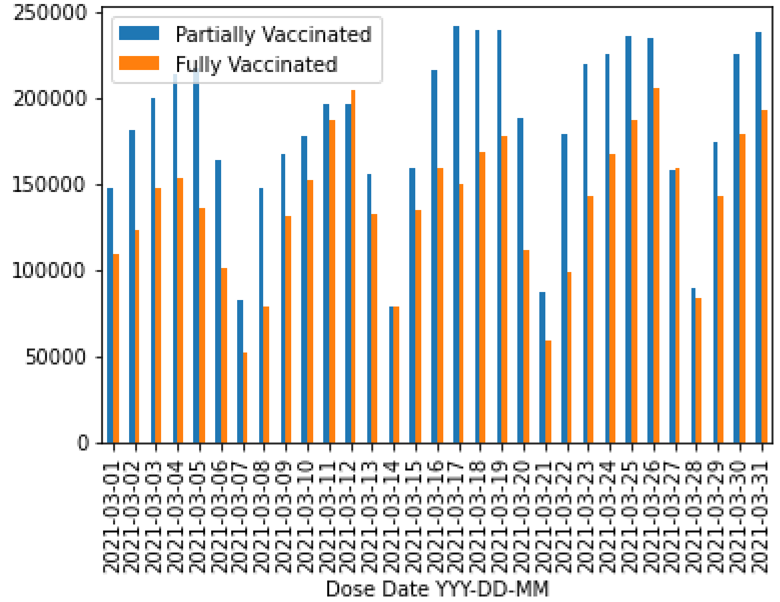
Out[15]:

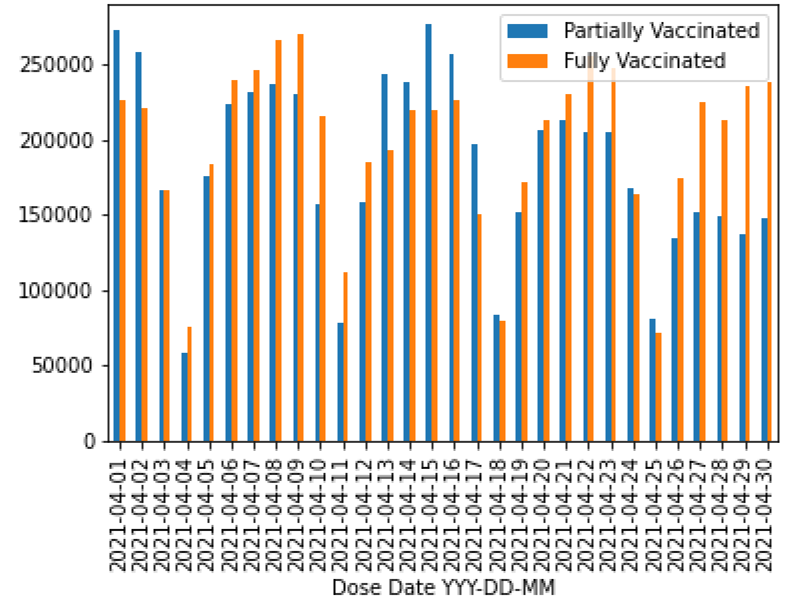
```
<AxesSubplot:xlabel='Dose Date YYY-DD-MM'>
```

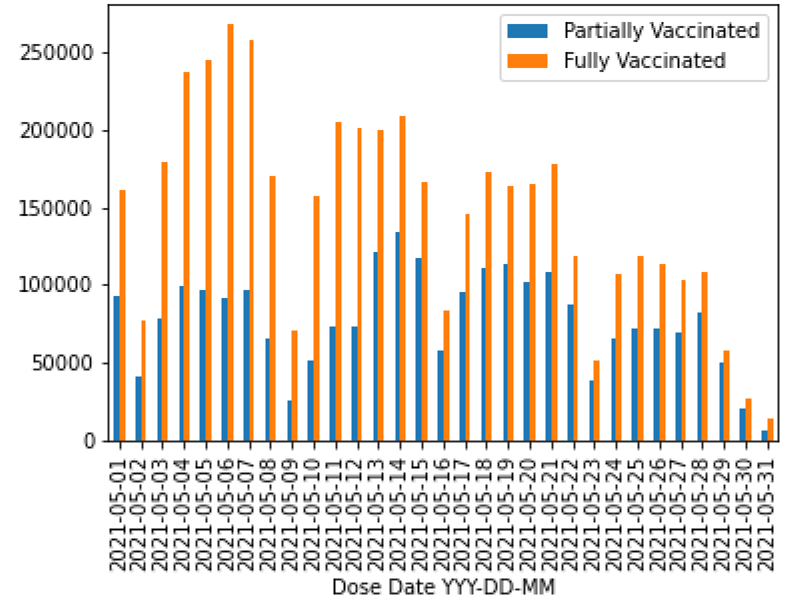


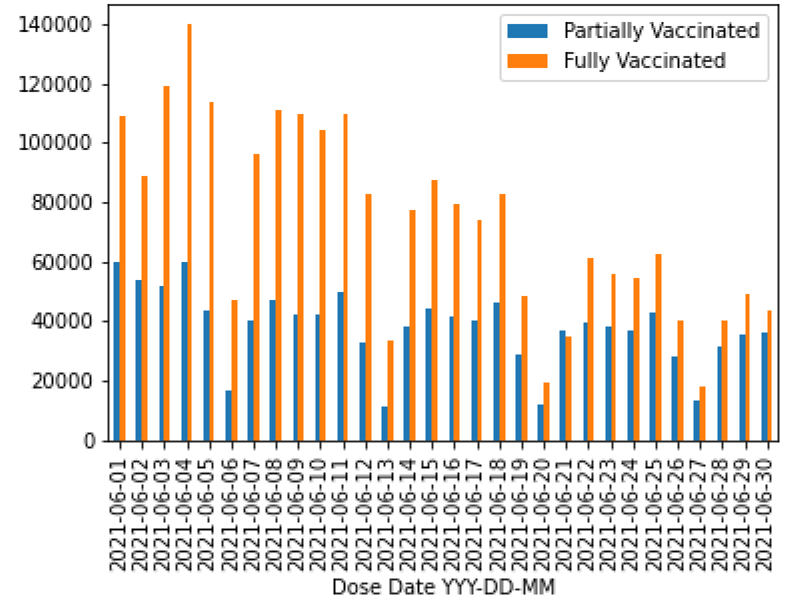

Partially Vaccinated
Fully Vaccinated
160000
140000
120000
100000
80000
60000
40000
20000
0
2021-01-01
2021-01-02
2021-01-03
2021-01-04
2021-01-05
2021-01-06
2021-01-07
2021-01-08
2021-01-09
2021-01-10
2021-01-11
2021-01-12
2021-01-13
2021-01-14
2021-01-15
2021-01-16
2021-01-17
2021-01-18
2021-01-19
2021-01-20
2021-01-21
2021-01-22
2021-01-23
2021-01-24
2021-01-25
2021-01-26
2021-01-27
2021-01-28
2021-01-29
2021-01-30
2021-01-31
Dose Date YYY-DD-MM

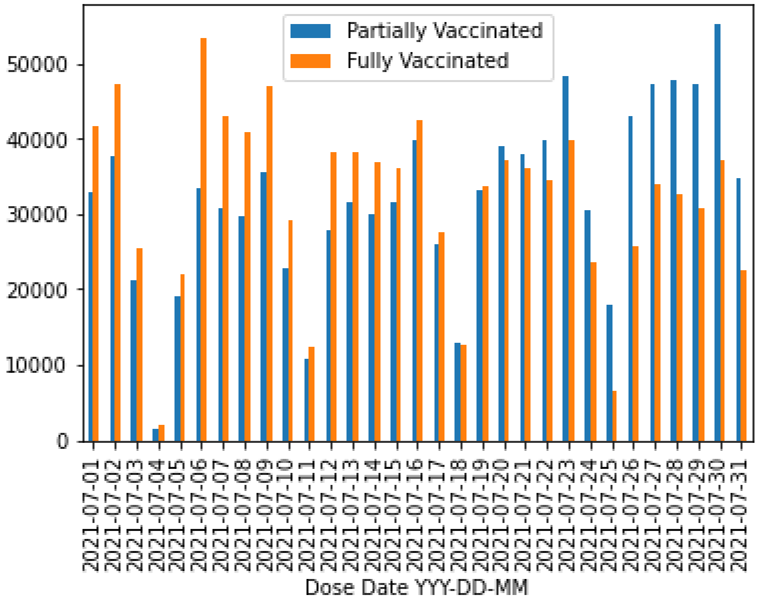

Partially Vaccinated
Fully Vaccinated
200000
175000
150000
125000
100000
75000
50000
25000
0
2021-02-01
2021-02-02
2021-02-03
2021-02-04
2021-02-05
2021-02-06
2021-02-07
2021-02-08
2021-02-09
2021-02-10
2021-02-11
2021-02-12
2021-02-13
2021-02-14
2021-02-15
2021-02-16
2021-02-17
2021-02-18
2021-02-19
2021-02-20
2021-02-21
2021-02-22
2021-02-23
2021-02-24
2021-02-25
2021-02-26
2021-02-27
2021-02-28
Dose Date YYY-DD-MM

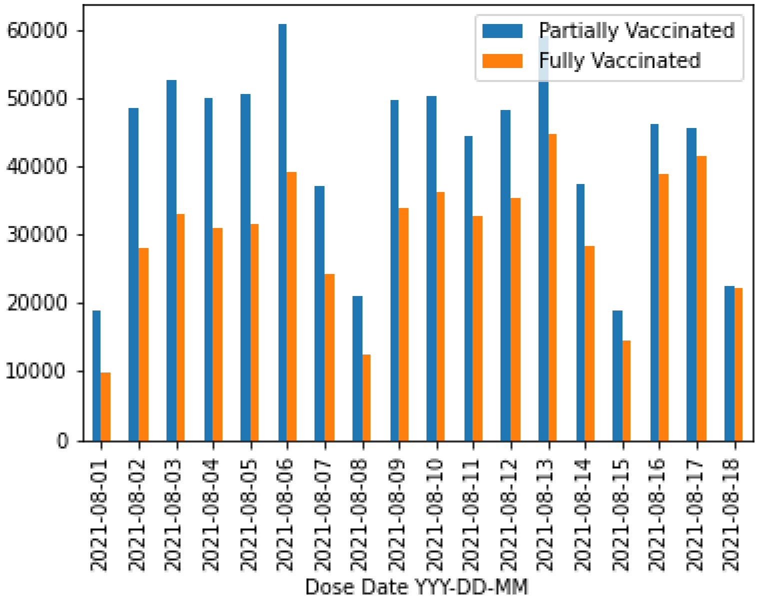

Partially Vaccinated
Fully Vaccinated
250000
200000
150000
100000
50000
0
2021-03-01
2021-03-02
2021-03-03
2021-03-04
2021-03-05
2021-03-06
2021-03-07
2021-03-08
2021-03-09
2021-03-10
2021-03-11
2021-03-12
2021-03-13
2021-03-14
2021-03-15
2021-03-16
2021-03-17
2021-03-18
2021-03-19
2021-03-20
2021-03-21
2021-03-22
2021-03-23
2021-03-24
2021-03-25
2021-03-26
2021-03-27
2021-03-28
2021-03-29
2021-03-30
2021-03-31
Dose Date YYY-DD-MM


Partially Vaccinated
Fully Vaccinated
0
50000
100000
150000
200000
250000
2021-04-01
2021-04-02
2021-04-03
2021-04-04
2021-04-05
2021-04-06
2021-04-07
2021-04-08
2021-04-09
2021-04-10
2021-04-11
2021-04-12
2021-04-13
2021-04-14
2021-04-15
2021-04-16
2021-04-17
2021-04-18
2021-04-19
2021-04-20
2021-04-21
2021-04-22
2021-04-23
2021-04-24
2021-04-25
2021-04-26
2021-04-27
2021-04-28
2021-04-29
2021-04-30
Dose Date YYY-DD-MM


Partially Vaccinated
Fully Vaccinated
0
50000
100000
150000
200000
250000
2021-05-01
2021-05-02
2021-05-03
2021-05-04
2021-05-05
2021-05-06
2021-05-07
2021-05-08
2021-05-09
2021-05-10
2021-05-11
2021-05-12
2021-05-13
2021-05-14
2021-05-15
2021-05-16
2021-05-17
2021-05-18
2021-05-19
2021-05-20
2021-05-21
2021-05-22
2021-05-23
2021-05-24
2021-05-25
2021-05-26
2021-05-27
2021-05-28
2021-05-29
2021-05-30
2021-05-31
Dose Date YYY-DD-MM


Partially Vaccinated
Fully Vaccinated
0
20000
40000
60000
80000
100000
120000
140000
2021-06-01
2021-06-02
2021-06-03
2021-06-04
2021-06-05
2021-06-06
2021-06-07
2021-06-08
2021-06-09
2021-06-10
2021-06-11
2021-06-12
2021-06-13
2021-06-14
2021-06-15
2021-06-16
2021-06-17
2021-06-18
2021-06-19
2021-06-20
2021-06-21
2021-06-22
2021-06-23
2021-06-24
2021-06-25
2021-06-26
2021-06-27
2021-06-28
2021-06-29
2021-06-30
Dose Date YYY-DD-MM





In [16]:

```
df.reset_index(inplace=True)
df
```

Out[16]:

	Dose Date YYY-DD-MM	California Counties	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken
0	2020-12-15	All CA Counties	1238	1238	1238	1238	0	0	0	0	1238	1238	0	0	1238	1238
1	2020-12-16	All CA Counties	9021	10259	9021	10259	0	0	0	0	9021	10259	0	0	9021	10259
2	2020-12-17	All CA Counties	27137	37396	27137	37396	0	0	0	0	27139	37398	0	0	27139	37398
3	2020-12-18	All CA Counties	46006	83402	45929	83325	77	77	0	0	46006	83403	1	1	46006	83404
4	2020-12-19	All CA Counties	25773	109175	25740	109065	33	110	0	0	25772	109174	1	2	25772	109176
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
14568	2021-08-14	Yuba	138	53901	90	20333	47	30134	1	3434	75	4725	61	26202	76	30927
14569	2021-08-15	Yuba	85	53986	57	20390	28	30162	0	3434	61	4764	22	26224	61	30988
14570	2021-08-16	Yuba	216	54202	100	20490	96	30258	20	3454	136	4844	76	26300	156	31144

	Dose Date YYY-DD-MM	California Counties	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken
14571	2021-08-17	Yuba	189	54391	96	20586	88	30346	5	3459	113	4891	71	26371	118	31262
14572	2021-08-18	Yuba	130	54521	59	20645	59	30405	12	3471	77	4928	52	26423	89	31351

14573 rows × 16 columns

In [17]:

```
# going to limit the data set to only the month of august 2021
# Then visualize how the Total Fully Vaccinated for each county

contain_values3 = df[df['Dose Date YYY-DD-MM'].str.contains('2021-08-18')]
#df.drop(labels = ["Unknown", "All CA and Non-CA Counties", "Outside California"], axis=0, inplace = True)
contain_values3.set_index("California Counties", inplace = True)
contain_values3.drop(labels = ["All CA Counties"], axis=0, inplace = True)
contain_values3.reset_index(inplace = True)
```

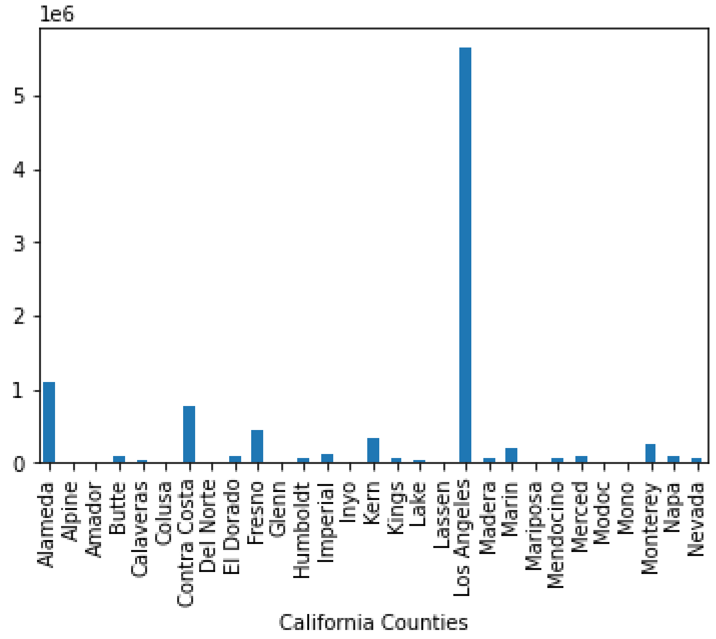
```
/opt/anaconda/lib/python3.8/site-packages/pandas/core/frame.py:4308: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  return super().drop(
```

In [18]:

```
# This chart shows how many people are fully vaccinated by county

contain_values3[:29].groupby('California Counties')['TOTAL Fully Vaccinated'].mean().plot.bar()
```

Out[18]:

```
<AxesSubplot:xlabel='California Counties'>
```

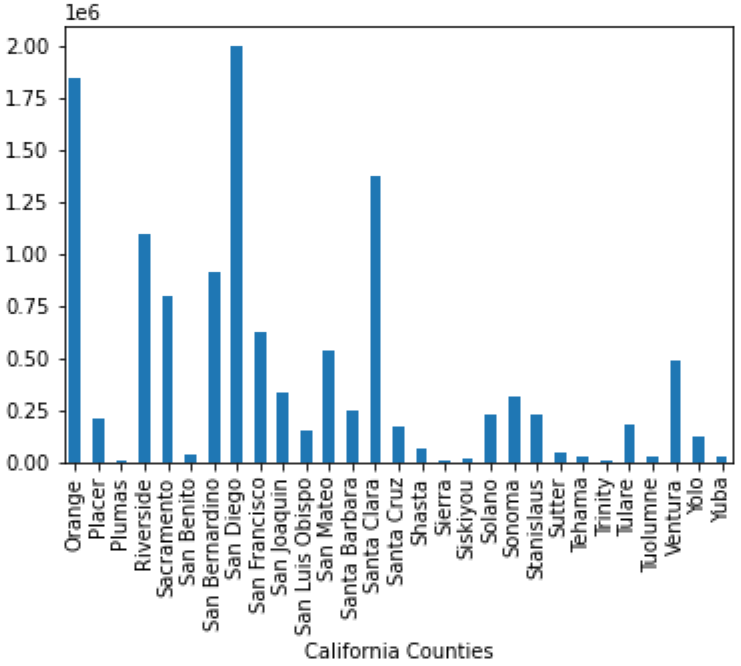


In [19]:

```
# This chart shows how many people are fully vaccinated by county

contain_values3[29:].groupby('California Counties')['TOTAL Fully Vaccinated'].mean().plot.bar()
```

Out[19]:

```
<AxesSubplot:xlabel='California Counties'>
```

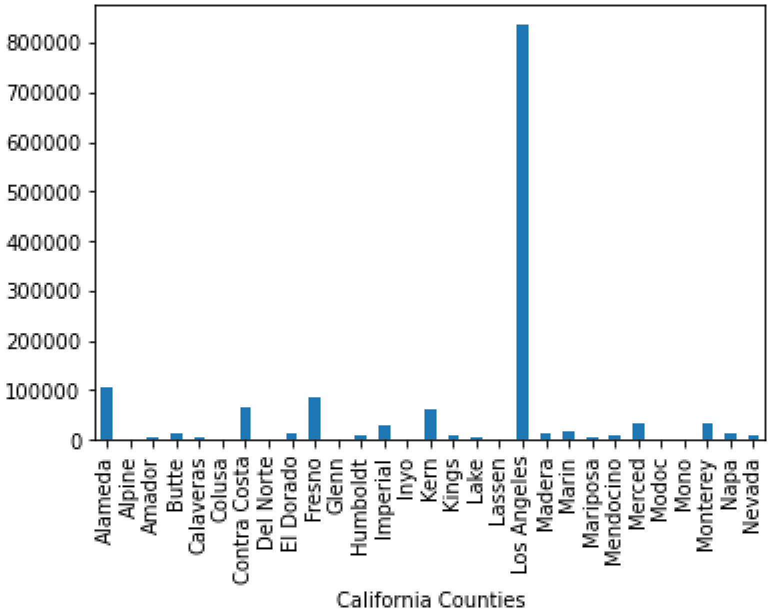


In [20]:

```
# This chart shows how many people are partially vaccinated by county

contain_values3[:29].groupby('California Counties')['TOTAL Partially Vaccinated'].mean().plot.bar()
```

Out[20]:

```
<AxesSubplot:xlabel='California Counties'>
```

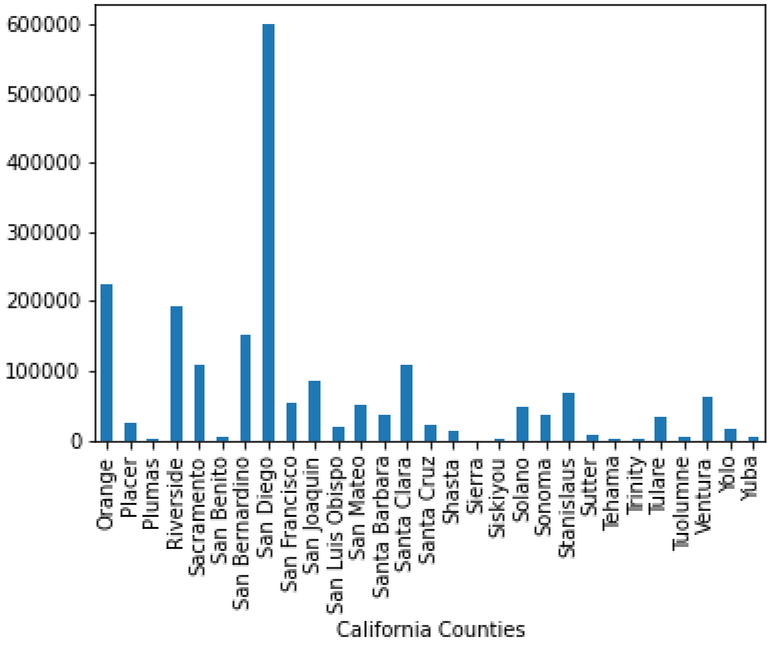


In [21]:

```
# This chart shows how many people are partially vaccinated by county

contain_values3[29:].groupby('California Counties')['TOTAL Partially Vaccinated'].mean().plot.bar()
```

Out[21]:

```
<AxesSubplot:xlabel='California Counties'>
```



In [22]:

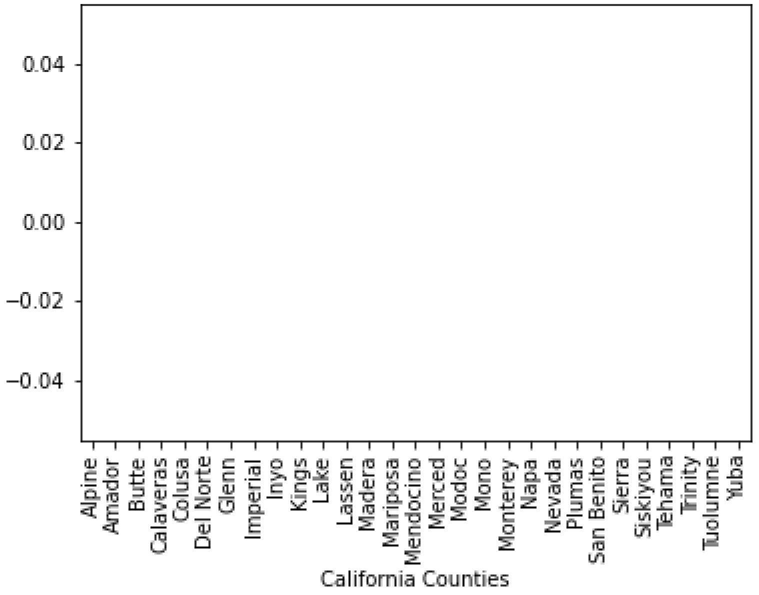
```
# Limiting our databse to only check for December 2020. We will not see any graph.
contain_values4 = df[df['Dose Date YYY-DD-MM'].str.contains('2020-12-15')]
#df.drop(labels = ["Unknown", "All CA and Non-CA Counties", "Outside California"], axis=0, inplace = True)
contain_values4.set_index("California Counties", inplace = True)
contain_values4.drop(labels = ["All CA Counties"], axis=0, inplace = True)
contain_values4.reset_index(inplace = True)
```

In [23]:

```
contain_values4[:29].groupby('California Counties')['TOTAL Fully Vaccinated'].mean().plot.bar()
```

Out[23]:

```
<AxesSubplot:xlabel='California Counties'>
```



In [24]:

```
# As expected we don't see any visualization because vaccination was not really started for most people.
```

In [26]:

```
# This is more data collection
# We thought it would be useful to get the population of each county to see how many people
# are vaccinated compared to how many people are actually in the county
```

```
import requests
response = requests.get("https://www.california-demographics.com/counties_by_population")

if response.status_code == 200:
    print('all good')
else:
    print('error')
```

```
all good
```

In [27]:

```
from bs4 import BeautifulSoup
soup = BeautifulSoup(response.content, "html.parser")
```

In [28]:

```
headings = soup.find_all(['tr'])
```

In [29]:

```
rows = []


for name in headings[1:-1]:
    county = name.find('a').text
    population = name.find('td').find_next('td').find_next('td').text

    population = population.replace('\n', '')
    population = population.replace(' ', '')
    population = population.replace(',', '')




    rows.append({
        "California Counties": county,
        "Population": int(population)
    })
```

In [30]:

```
import pandas as pd
import warnings
warnings.filterwarnings("ignore")

df_pop = pd.DataFrame(rows)
#df_pop
```

In [31]:

```
df_pop.sort_values('California Counties', ascending=True, inplace = True)
#df_pop
```

In [32]:

```
contain_values3.sort_values('California Counties', ascending=True, inplace = True)
```

In [33]:

```
contain_values3["Population"]  = df_pop["Population"].values
contain_values3
```

Out[33]:

	California Counties	Dose Date YYY-DD-MM	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken	Population
1	Alameda	2021-08-18	1877	2209568	1451	1403820	319	696168	107	109580	810	104950	798	1102418	915	1207368	1656754

	California Counties	Dose Date YYY-DD-MM	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken	Population
0	Alpine	2021-08-18	0	1509	0	28	0	1474	0	7	0	139	0	687	0	826	1039
2	Amador	2021-08-18	24	36299	7	9663	14	25333	3	1303	14	6078	10	15734	17	21812	38429
3	Butte	2021-08-18	280	186443	196	90252	65	88653	19	7538	108	12288	164	90403	127	102691	225817
4	Calaveras	2021-08-18	41	40877	16	17293	16	22359	9	1225	18	4399	21	18748	28	23147	45514
5	Colusa	2021-08-18	14	18921	11	4980	3	13178	0	763	9	1627	5	9006	9	10633	21454
6	Contra Costa	2021-08-18	1816	1545626	1455	1039515	262	455691	99	50420	784	63395	816	763750	881	827145	1142251
7	Del Norte	2021-08-18	59	19952	20	7152	33	11682	6	1118	32	1730	13	9610	38	11340	27495
8	El Dorado	2021-08-18	209	197376	140	94496	50	95269	19	7611	94	12627	95	95706	113	108333	188563
9	Fresno	2021-08-18	1466	928182	1019	496933	359	397639	88	33610	745	85225	669	437812	833	523037	984521
10	Glenn	2021-08-18	67	23518	59	9763	5	12514	3	1241	41	1296	26	11690	44	12986	27976
11	Humboldt	2021-08-18	167	144108	108	73003	34	63713	25	7392	70	9289	79	70665	95	79954	135940
12	Imperial	2021-08-18	461	231559	342	118147	78	98552	41	14860	208	27206	248	109413	250	136619	180701
13	Inyo	2021-08-18	24	17635	14	6731	10	10360	0	544	16	981	6	8579	16	9560	17977
14	Kern	2021-08-18	1229	700611	727	362884	438	306285	64	31442	662	62156	500	333294	726	395450	887641
15	Kings	2021-08-18	199	97495	127	45402	60	47128	12	4965	103	10041	94	46131	114	56172	150691
16	Lake	2021-08-18	64	56865	38	17499	26	36685	0	2681	37	4469	23	27437	37	31906	64195
17	Lassen	2021-08-18	10	12331	4	1586	6	9674	0	1071	5	786	4	6264	5	7050	30818
18	Los Angeles	2021-08-18	12413	11742429	8448	6482717	3253	4798727	712	460985	5305	835810	6056	5652474	6007	6488284	10081570
19	Madera	2021-08-18	353	128986	251	66777	91	56934	11	5275	227	12044	110	61266	238	73310	155433
20	Marin	2021-08-18	190	391246	113	248268	63	130615	14	12363	58	15745	62	192868	72	208613	259943
21	Mariposa	2021-08-18	2	13438	0	2835	2	10050	0	553	1	2958	0	5504	1	8462	17420
22	Mendocino	2021-08-18	87	98088	49	53369	32	41277	6	3442	49	7798	24	46549	55	54347	87224
23	Merced	2021-08-18	320	218264	192	128352	114	80474	14	9438	204	33963	101	95704	218	129667	271382
24	Modoc	2021-08-18	3	6018	0	908	3	4714	0	396	2	308	1	3047	2	3355	8907
25	Mono	2021-08-18	17	16978	11	9607	6	7044	0	327	12	874	3	8189	12	9063	14310
26	Monterey	2021-08-18	899	494814	666	238043	221	235414	12	21357	438	34958	418	239461	450	274419	433410
27	Napa	2021-08-18	136	185310	98	97358	35	82468	3	5484	71	12776	50	88636	74	101412	139623
28	Nevada	2021-08-18	116	107168	63	57435	39	46734	14	2999	55	7441	52	50936	69	58377	99244
29	Orange	2021-08-18	3184	3809385	2090	2181852	896	1507674	198	119859	1396	225432	1416	1842509	1593	2067941	3168044
30	Placer	2021-08-18	530	433930	353	249811	128	172523	49	11596	260	24746	220	209602	308	234348	385512
31	Plumas	2021-08-18	7	17183	3	2182	4	13913	0	1088	6	900	0	8634	6	9534	18660
32	Riverside	2021-08-18	3064	2313509	2067	1337398	851	908020	146	68091	1481	191239	1288	1092705	1626	1283944	2411439
33	Sacramento	2021-08-18	2460	1655374	1646	934339	624	673840	190	47195	1159	107920	1164	793900	1347	901820	1524553
34	San Benito	2021-08-18	90	68749	48	38629	31	26976	11	3144	35	5765	47	32815	46	38580	60376

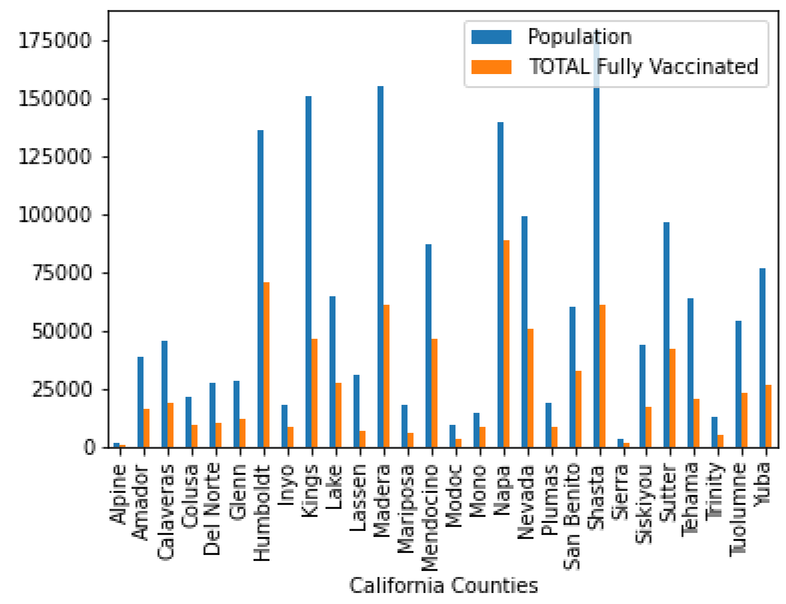
	California Counties	Dose Date YYY-DD-MM	Total Doses	Cumulative Total	PFIZER	PFIZER TOTAL	MODERNA	MODERNA TOTAL	JANSSEN	JANSSEN TOTAL	Partially Vaccinated	TOTAL Partially Vaccinated	Fully Vaccinated	TOTAL Fully Vaccinated	One Dose Taken	TOTAL One Doses Taken	Population
35	San Bernardino	2021-08-18	2988	1925865	2268	1147680	555	717777	165	60408	1423	150560	1467	914184	1590	1064744	2149031
36	San Diego	2021-08-18	3395	4115586	2227	2240656	803	1718387	365	156543	1629	598449	1538	1994991	2297	2593440	3316073
37	San Francisco	2021-08-18	897	1260800	582	812904	245	395536	70	52360	329	54597	432	625161	396	679758	874961
38	San Joaquin	2021-08-18	1009	730069	698	416103	274	289437	37	24529	552	85483	430	332554	613	418037	742603
39	San Luis Obispo	2021-08-18	252	305860	162	157878	67	136446	23	11536	115	19849	98	147963	137	167812	282165
40	San Mateo	2021-08-18	810	1079026	620	667996	104	372056	86	38974	354	51966	358	530437	440	582403	767423
41	Santa Barbara	2021-08-18	557	506310	385	296286	121	187982	51	22042	272	35351	216	245337	323	280688	444829
42	Santa Clara	2021-08-18	1920	2766793	1286	1790593	500	874608	134	101592	826	107501	841	1373913	961	1481414	1927470
43	Santa Cruz	2021-08-18	192	345030	134	169347	53	166837	5	8846	74	21225	88	165605	79	186830	273962
44	Shasta	2021-08-18	128	128974	64	57862	53	64478	11	6634	60	13013	53	61012	70	74025	179212
45	Sierra	2021-08-18	1	2896	0	234	1	2616	0	46	1	79	0	1422	1	1501	3040
46	Siskiyou	2021-08-18	33	35148	13	14546	19	19558	1	1044	20	2802	11	16634	21	19436	43468
47	Solano	2021-08-18	881	482485	555	264352	299	204783	27	13350	429	47022	409	223567	456	270589	441829
48	Sonoma	2021-08-18	700	648382	515	360898	125	266726	60	20758	295	36549	306	314591	355	351140	499772
49	Stanislaus	2021-08-18	774	510466	497	295146	241	197677	36	17643	422	67065	309	228847	455	295912	543194
50	Sutter	2021-08-18	112	88950	59	37016	47	47815	6	4119	71	7714	41	42368	76	50082	96109
51	Tehama	2021-08-18	97	41895	73	15447	20	24859	4	1589	55	2905	39	20226	59	23131	63912
52	Trinity	2021-08-18	3	9611	3	1905	0	7063	0	643	2	747	1	4738	2	5485	12700
53	Tulare	2021-08-18	456	373475	330	220206	106	142385	20	10884	245	33072	198	174857	265	207929	461898
54	Tuolumne	2021-08-18	42	48561	12	13225	30	34292	0	1044	30	4116	11	22602	31	26718	54045
55	Ventura	2021-08-18	1174	1002338	832	537102	291	429920	51	35316	590	60930	486	486113	640	547043	847263
56	Yolo	2021-08-18	281	254992	188	124336	79	123018	14	7638	140	16227	126	122765	155	138992	217352
57	Yuba	2021-08-18	130	54521	59	20645	59	30405	12	3471	77	4928	52	26423	89	31351	76360

In [34]:

```
contain_values3.sort_values('Population', ascending=True, inplace = True)
contain_values3[:28].groupby('California Counties')[['Population', 'TOTAL Fully Vaccinated']].mean().plot.bar()
# Now, comparing vaccinated people with total population
```

Out[34]:

```
<AxesSubplot:xlabel='California Counties'>
```

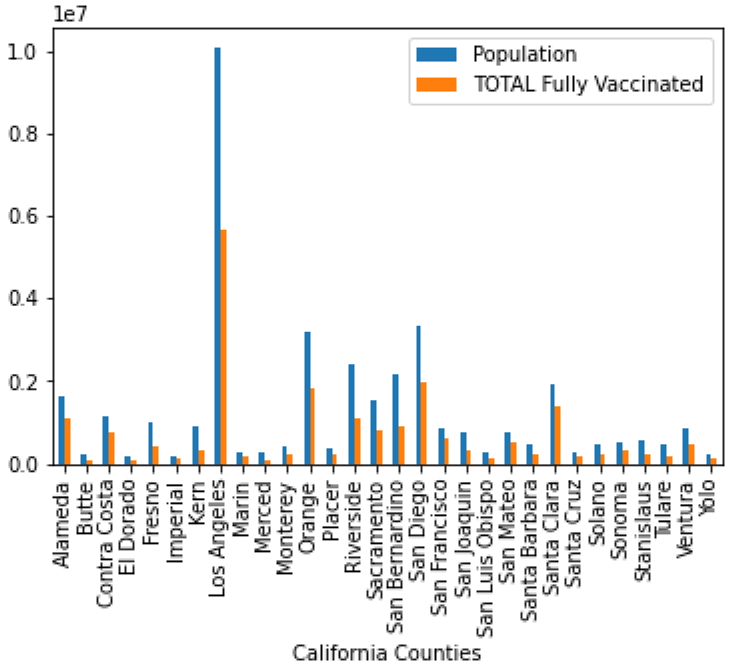


In [35]:

```
contain_values3[28:].groupby('California Counties')[['Population', 'TOTAL Fully Vaccinated']].mean().plot.bar()
```

Out[35]:

```
<AxesSubplot:xlabel='California Counties'>
```

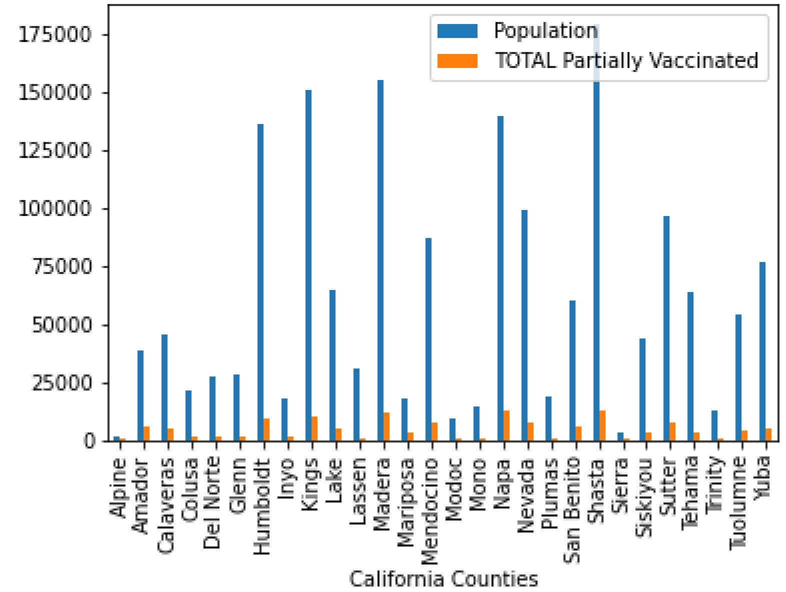


In [36]:

```
contain_values3[:28].groupby('California Counties')[['Population', 'TOTAL Partially Vaccinated']].mean().plot.bar()
# Checking to see how many people are partially vaccinated
```

Out[36]:

```
<AxesSubplot:xlabel='California Counties'>
```

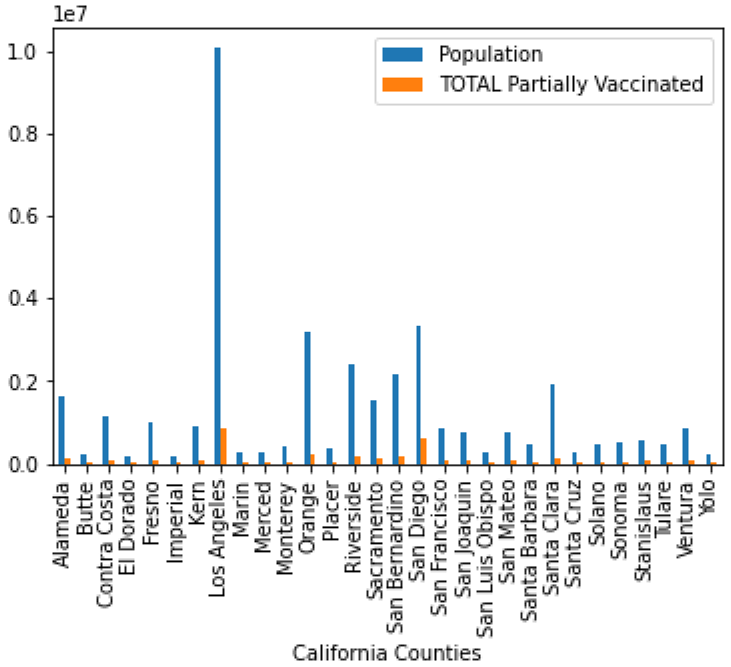


In [37]:

```
contain_values3[28:].groupby('California Counties')[['Population', 'TOTAL Partially Vaccinated']].mean().plot.bar()
```

Out[37]:

```
<AxesSubplot:xlabel='California Counties'>
```

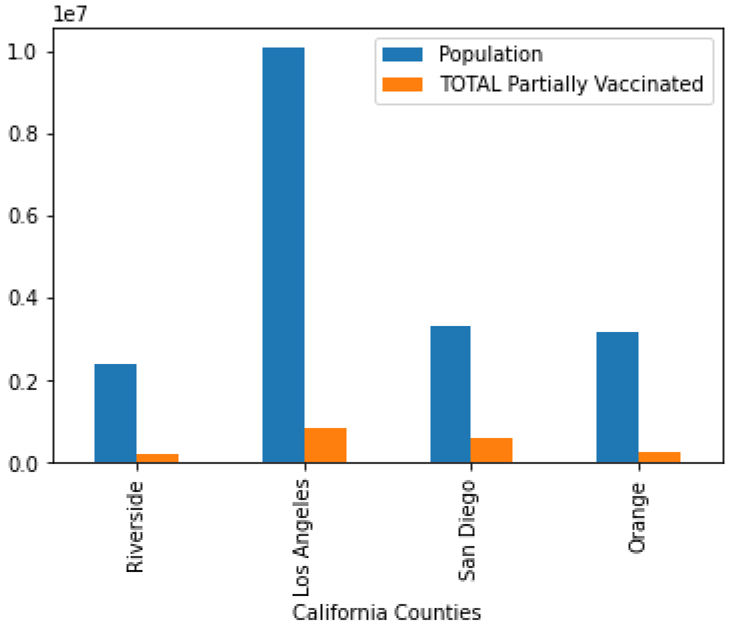


In [38]:

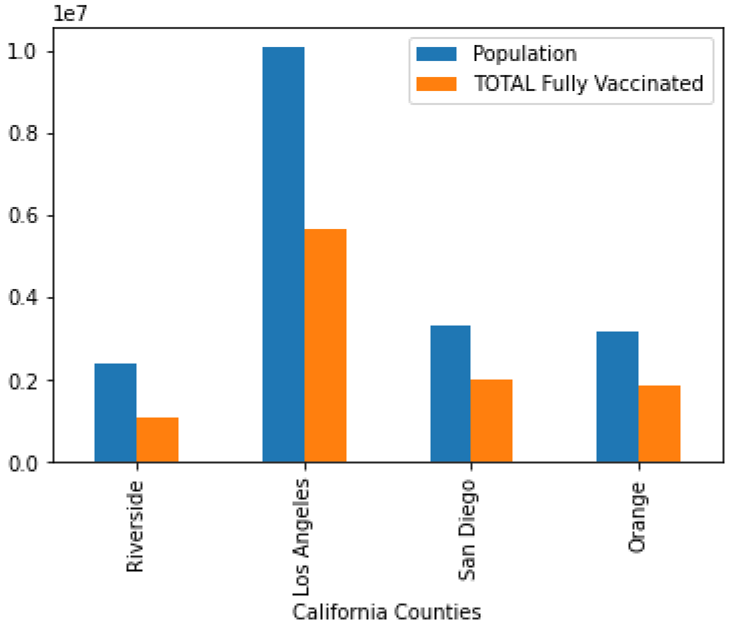
```
# Here we check only top 4 populated counties of california to see how many people are partially vaccinated
contain_values3.groupby("California Counties")[["Population", "TOTAL Partially Vaccinated"]].mean().loc[["Riverside", "Los Angeles", "San Diego", "Orange"]].plot.bar()
```

Out[38]:

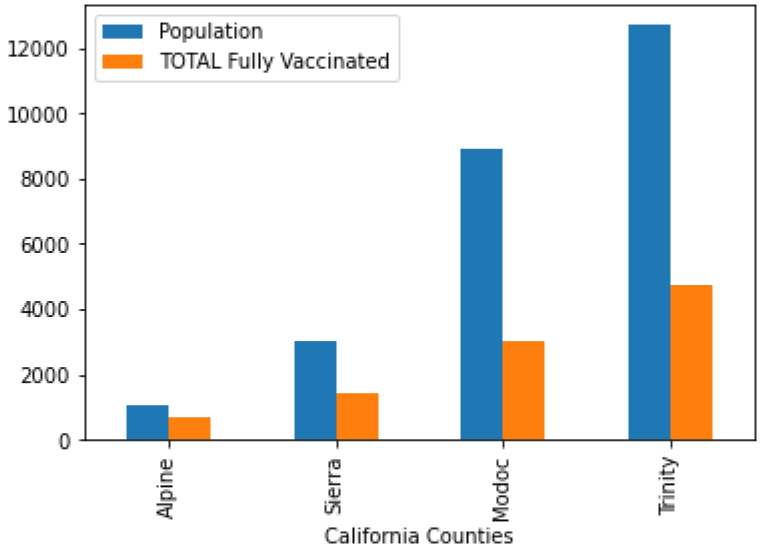
```
<AxesSubplot:xlabel='California Counties'>
```



```
In [39]: # Comparing how many people are fully vaccinated in populated counties.
         contain_values3.groupby("California Counties")[["Population", "TOTAL Fully Vaccinated"]].mean().loc[["Riverside", "Los Angeles", "San Diego", "Orange"]].plot.bar()

Out[39]: <AxesSubplot:xlabel='California Counties'>
```



```
In [40]: # Now, doing the same comparasion but with least populated counties.
         contain_values3.groupby("California Counties")[["Population", "TOTAL Fully Vaccinated"]].mean().loc[["Alpine", "Sierra", "Modoc", "Trinity"]].plot.bar()

Out[40]: <AxesSubplot:xlabel='California Counties'>
```


Population
TOTAL Fully Vaccinated
12000
10000
8000
6000
4000
2000
0
Alpine
Sierra
Modoc
Trinity
California Counties

```
In [ ]:
```