ANA-522-OL1 Spring 2022
Mod04 Week07 Lab: Data Plot and Visualization
Due: Friday February 25th at midnight

COVID-19 Vaccinations in the United States

Overall US COVID-19 Vaccine deliveries and administration data at national and jurisdiction level. Data represents all vaccine partners including jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities.

COVID-19 Vaccinations in the United States,Jurisdiction used in this lab exercise.

New weekly allocations of doses are posted every Tuesday. Beginning the following Thursday, states can begin ordering doses from that week's new allocation of 1st doses. Beginning two weeks (Pfizer) or three weeks (Moderna) from the following Sunday, states can begin ordering doses from that week's new allocation of 2nd doses. After doses are ordered by states, shipments begin the following Monday. The entire order may not arrive in one shipment or on one day, but over the course of the week.

Second doses are opened up for orders on Sundays, at the appropriate interval two or three weeks later according to the manufacturer's label, with shipments occurring after jurisdictions place orders. Shipments of an FDA-authorized safe and effective COVID-19 vaccine continue to arrive at sites across America. Vaccinations began on December 14, 2020.

Reference : **About COVID-19 Vaccine Delivered and Administration Data**

This lab exercise is to zoom in on the subset of state of Maryland in January 2022.

df_Jan_2022_MD_diff dataframe, derived in the following, is resulted from extracting the subset.

```
import numpy as np
import pandas as pd
df= pd.read_csv('/home/ANA522/mod04/
 ↪COVID-19_Vaccinations_in_the_United_States_Jurisdiction.csv', sep=',')
```

```
from datetime import datetime
# convert Data column to be of type datetime.date
obj_to_date = lambda x: datetime.strptime(x,'%m/%d/%Y').date()
```

```python
df['Date'] = df['Date'].apply(obj_to_date)
#df['Date']=pd.to_datetime(df['Date']).date()
```

```python
### Retrieve State of Maryland data
# sorting by Date
df_date=df.sort_values(by='Date')

# reset index with order of values in Date
df_date.reset_index(inplace=True)

# drop original index column
df_date= df_date.drop(['index'],axis=1)

# use Date of datetime.date as index
df_date.index = df_date['Date']

# focus on Maryland (MD) Location
df_date_MD = df_date[df_date['Location']=='MD']
df_date_MD
```

```python
### Zoom in on January 2022 data
# Transform index with DatetimeIndex from original date string index
dates_index_MD = pd.DatetimeIndex(df_date_MD.index)

df_date_MD.index = dates_index_MD

# select January 2022 (including last day of Dec 2021) from index with␣
 ↪Distributed and Administered columns
df_Jan_2022_MD = df_date_MD['2021/12/31':'2022/01/
 ↪31'][['Distributed','Administered','Series_Complete_Yes',

                                       ␣
 ↪'Administered_Janssen','Administered_Moderna',

                                       ␣
 ↪'Administered_Pfizer','Administered_Unk_Manuf']]
df_Jan_2022_MD
```
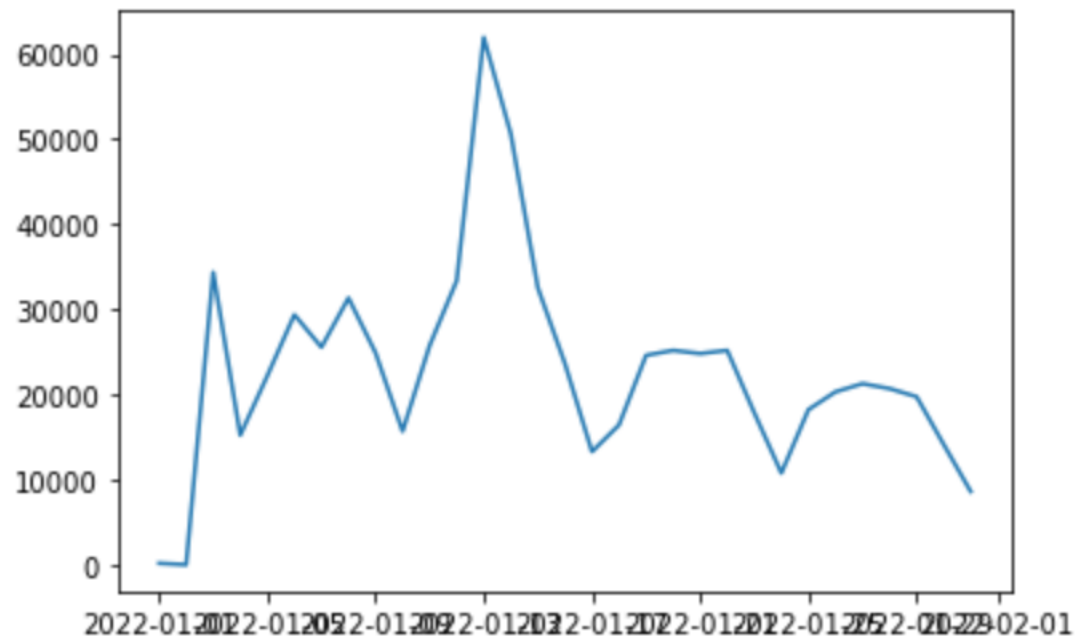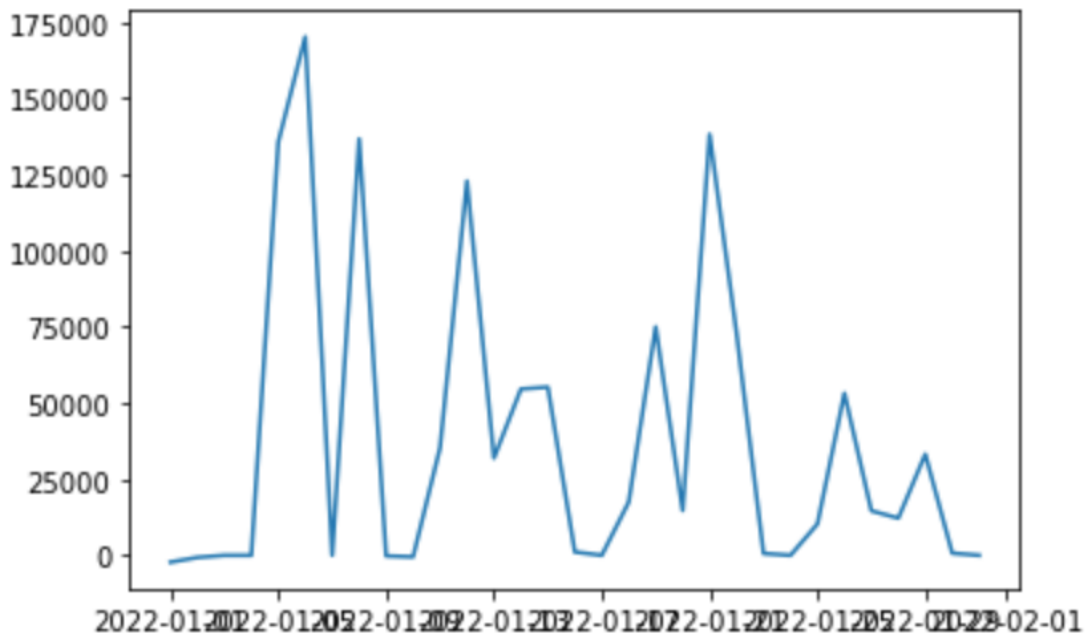
```python
# calculate the differences of dose increments from the day before
# the new table of doeses added daily

df_Jan_2022_MD_diff = df_Jan_2022_MD.diff().loc['2022/01']
df_Jan_2022_MD_diff
```
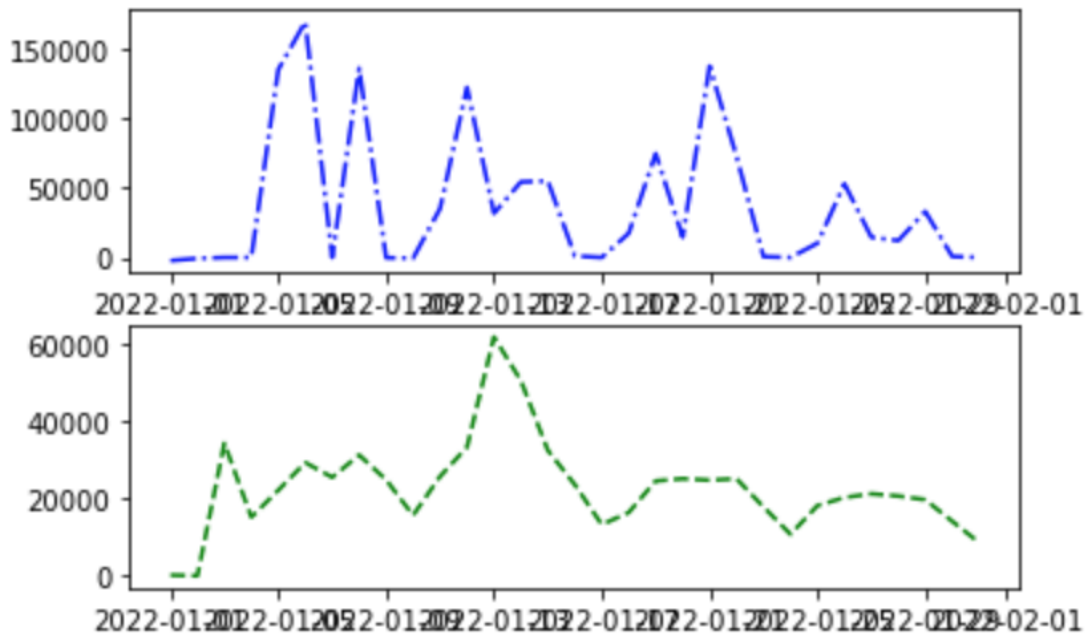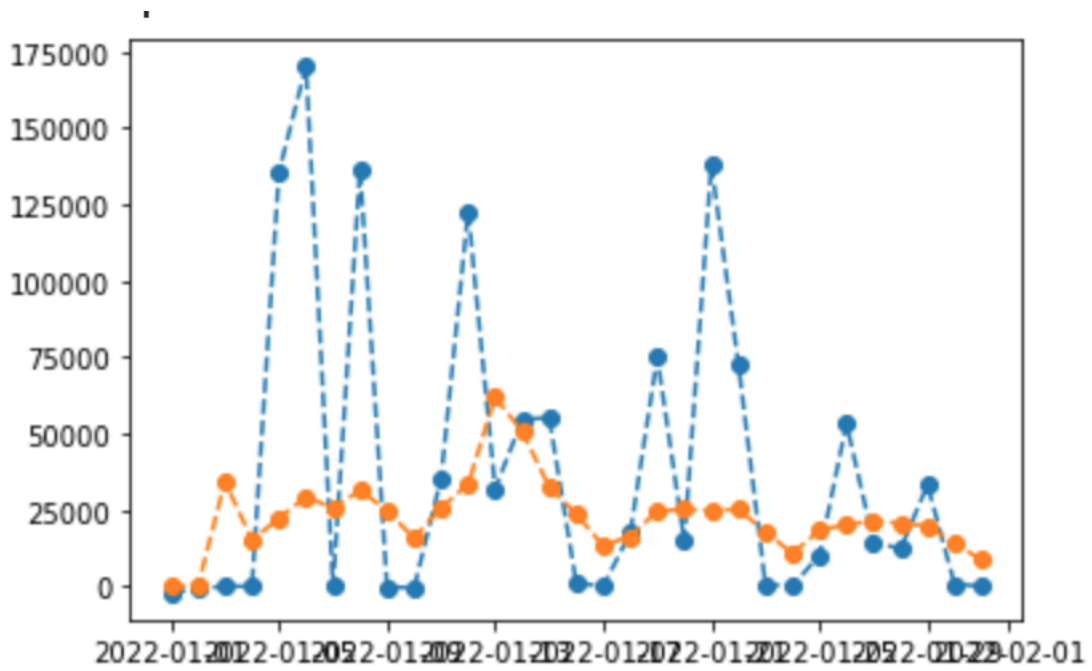
Q01: Draw a Line plot for Administered doses

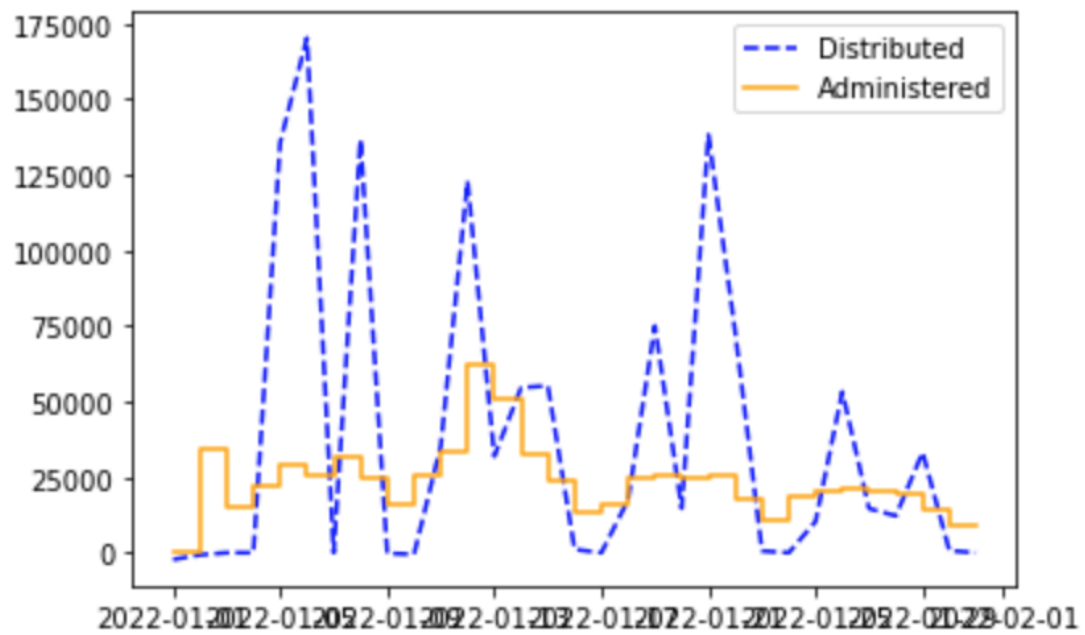Q02: Draw a Line plot for Distributed doses

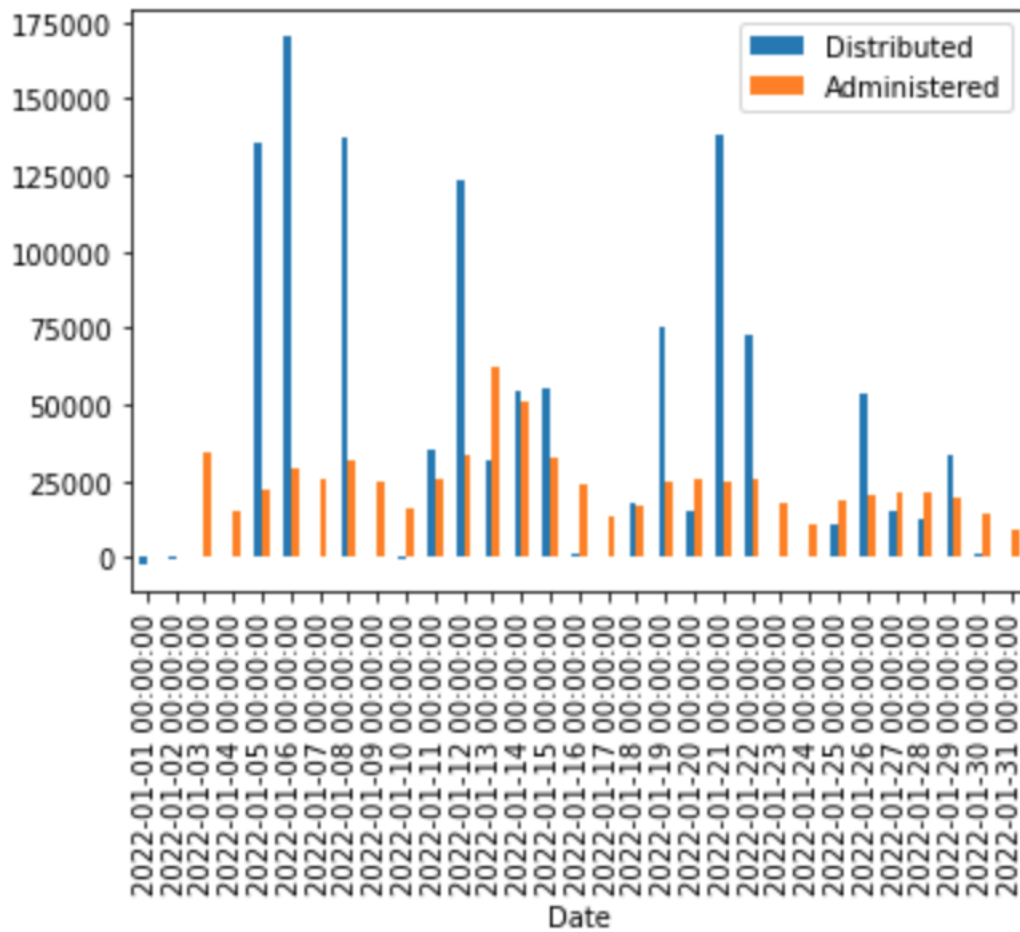Q03: Draw a Subplots for Distributed and Administered doses

Q04: Draw a Line plot for both Distributed and Administered doses with markers and using different colors.

Q05: Draw a Line plot for Distributed does with dashed line and Administered doses with steps with legends and different colors

Q06: Draw a Bar chat for both Distributed and Administered doses with legends and using different colors

Q07: Horizontal bar chat for both Distributed and Administered doses with legends and using different colors
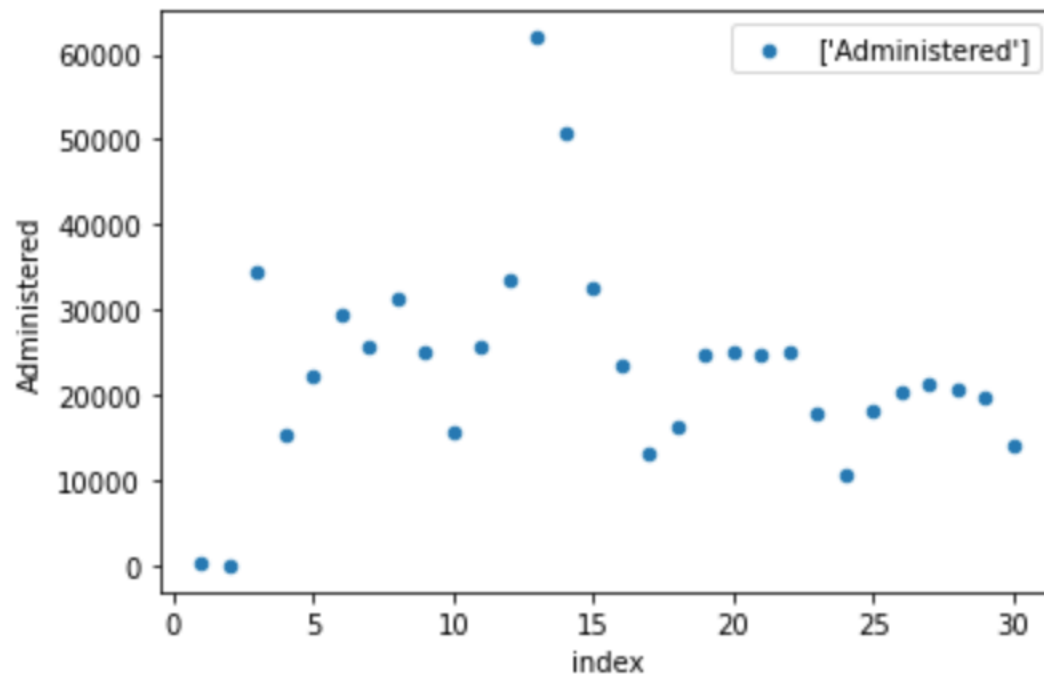
```
[ ]: # Prepare for Q08.
     # Prepare df_Jan_2022_MD_diff_idx dataframe to be used for Q08.
     # First Reset index to have just counting numbers to be added to columns

     df_Jan_2022_MD_diff_idx = df_Jan_2022_MD_diff.reset_index()
     df_Jan_2022_MD_diff_idx['index']=df_Jan_2022_MD_diff_idx.index.to_series().
       ↪shift(-1)

     df_Jan_2022_MD_diff_idx
```
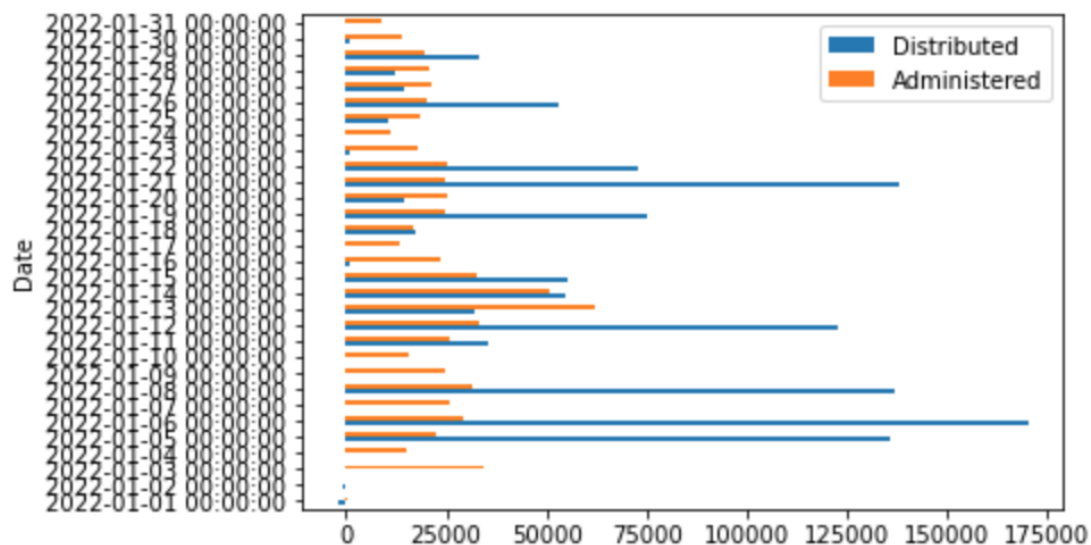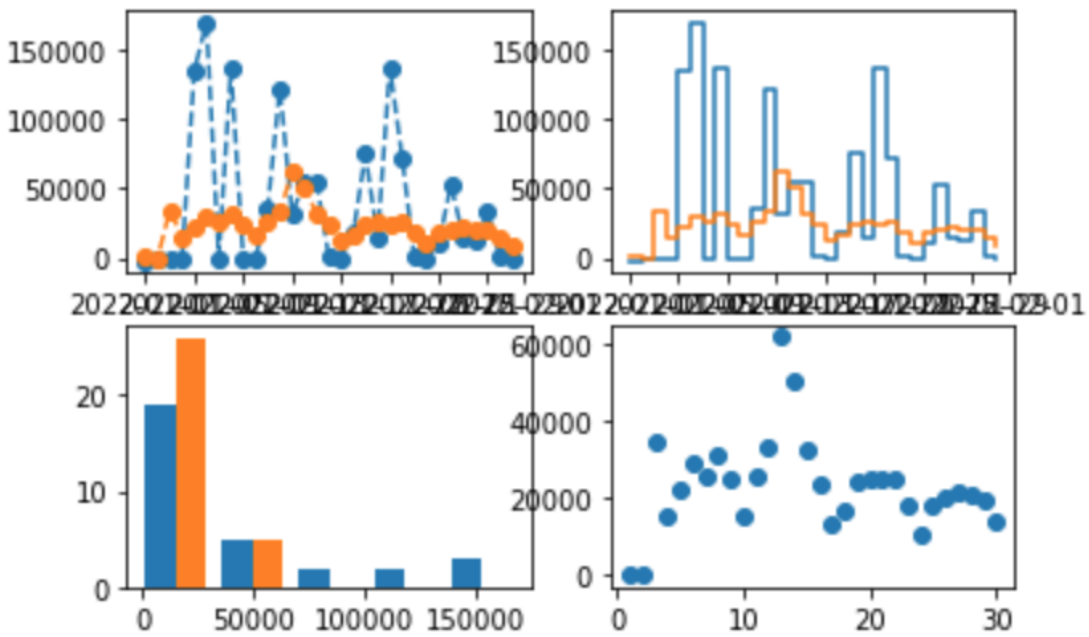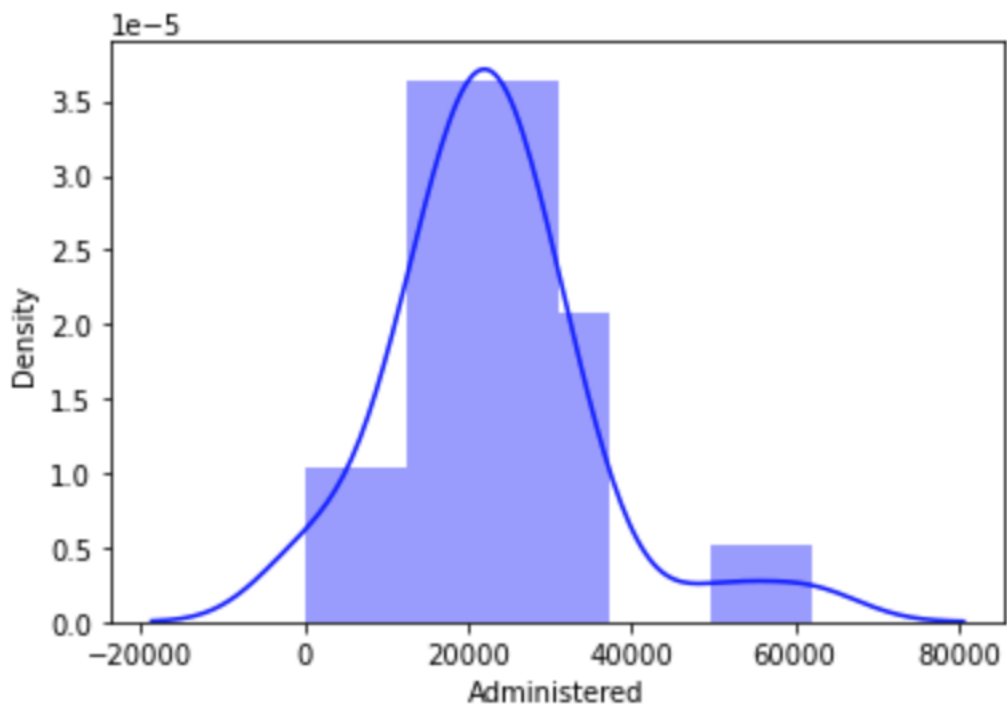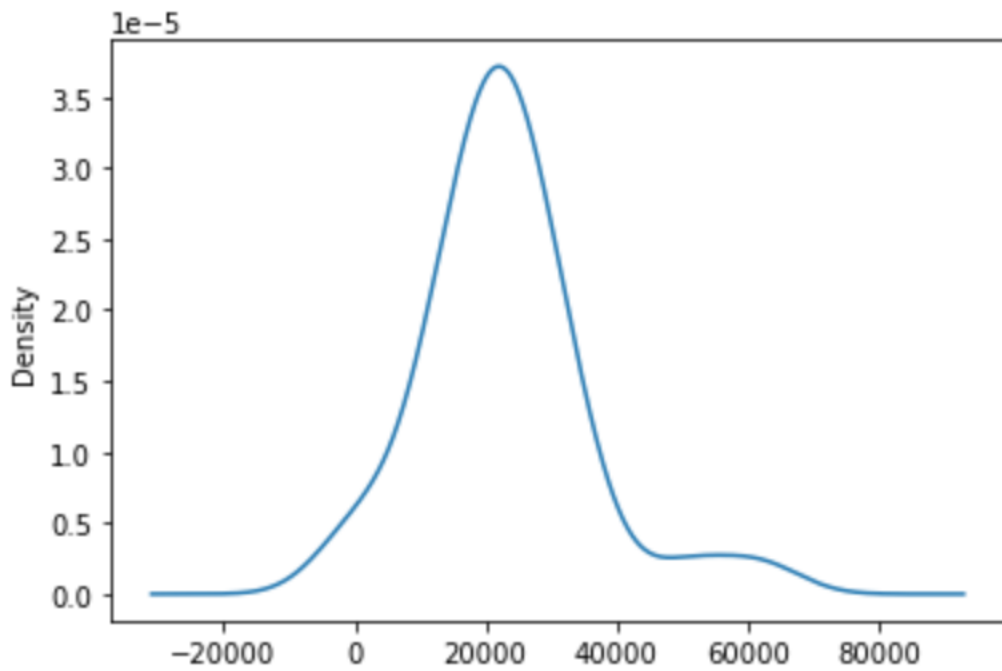
Q08: Draw a Scatter plot for Administered doses with legends.

Q09: Create a Matplot figure to accommodate line charts, histogram, and scatter plot as subplots.

Q10: Create either a density distribution plot from Administered column (or with histogram bins set to 10 using seaborn)

```
[ ]:  # Preparation for Q11
      # Preparation for the normalized 3-day dataframe,␣
      ↪df_Jan_2022_MD_diff_Vaccin_normal_3days, for plotting
      # Create subset DataFrame of various vaccines administered
      df_Jan_2022_MD_diff_Vaccin = df_Jan_2022_MD_diff.loc[:,'Administered_Janssen':]
      df_Jan_2022_MD_diff_Vaccin
```

```
[ ]:  # Normalize the numbers of doses from different vaccines
      df_Jan_2022_MD_diff_Vaccin_normal = df_Jan_2022_MD_diff_Vaccin.
      ↪div(df_Jan_2022_MD_diff_Vaccin.sum(1),axis=0)
      df_Jan_2022_MD_diff_Vaccin_normal
```

```
[ ]:  # Get data entries of every three days
      df_Jan_2022_MD_diff_Vaccin_normal_3days = df_Jan_2022_MD_diff_Vaccin_normal[::3]
      df_Jan_2022_MD_diff_Vaccin_normal_3days
```

Q11: Draw a Bar plot of normalized 3-day vaccines,of Administered_Janssen, Administered_Moderna, Administered_Pfizer, Administered_Unk_Manuf columns
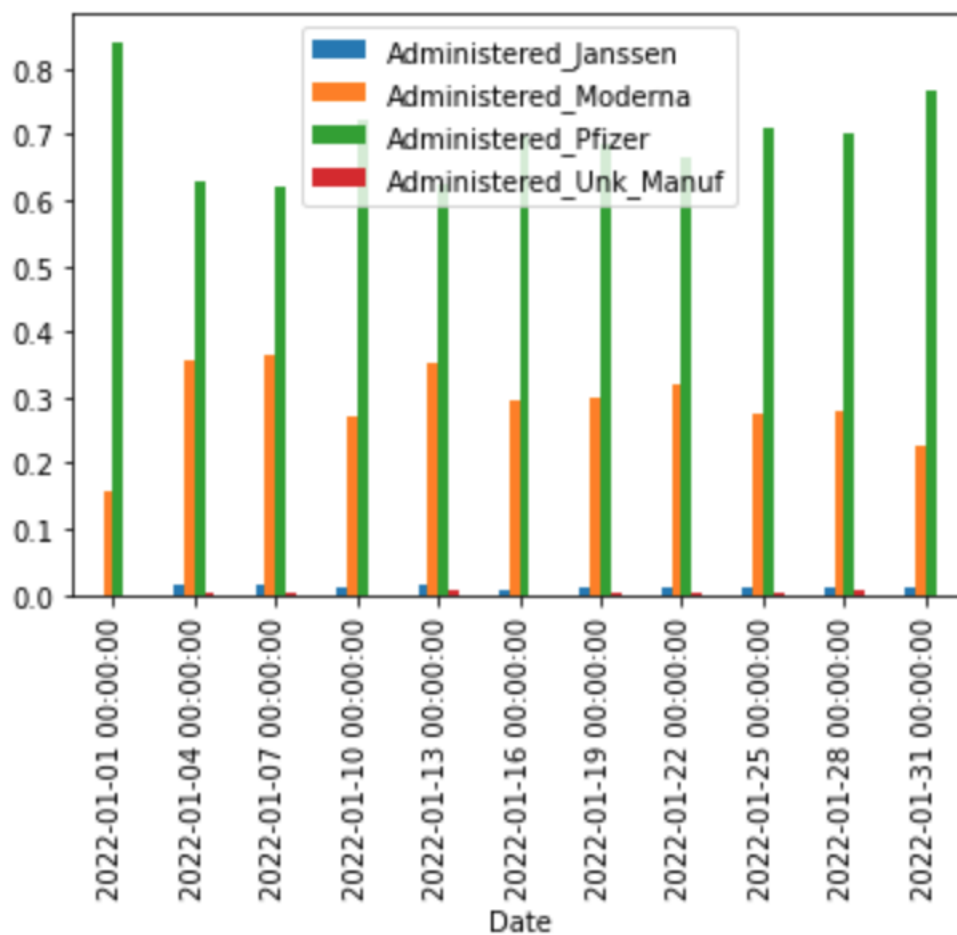
Q12: Create a pairs plot or scatter plot matrix in between Administered_Janssen, Administered_Moderna, Administered_Pfizer, Administered_Unk_Manuf columns using seaborn. (either one)