

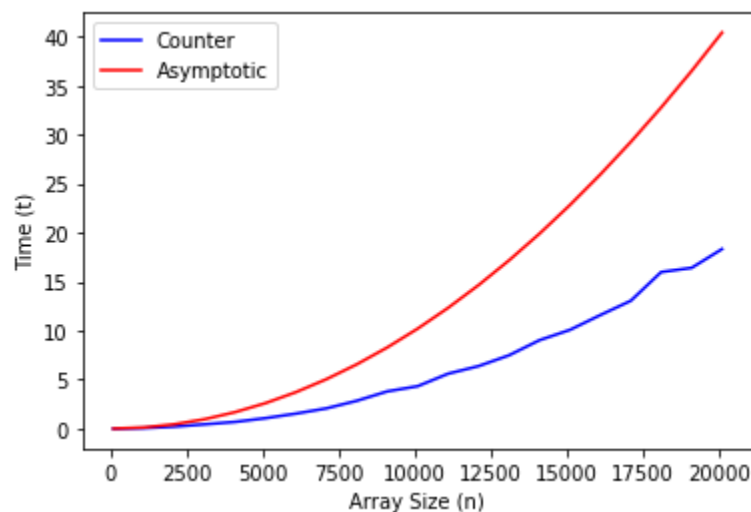
Problem Set 2

Q1

Write a short note on how you were able to achieve $O(N)$ complexity for the given problem.

=> $O(N)$ time complexity is obtained by increasing the space complexity. We iterate the loop only once, during iteration we add new values in the dictionary and increase the count during the repetition. The dictionary stores the num and its count. At the end of one iteration, we obtain a dictionary of all unique numbers with the count. We then run another loop around the dictionary to get the unique number having more than x counts.

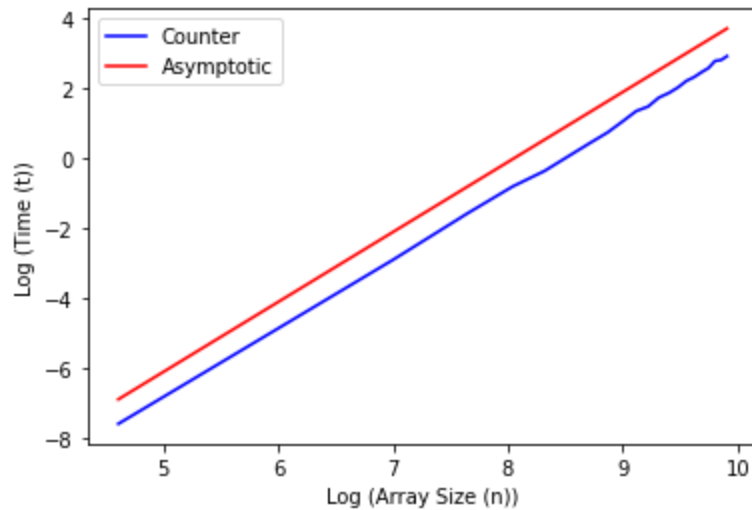
Plot the size of array (n) vs time Taken(t) for the counter() function and plot an asymptotical line which denotes the theoretical limit for $O(N^2)$.



Identify the C value that was used to construct the asymptotical line in the previous graph.

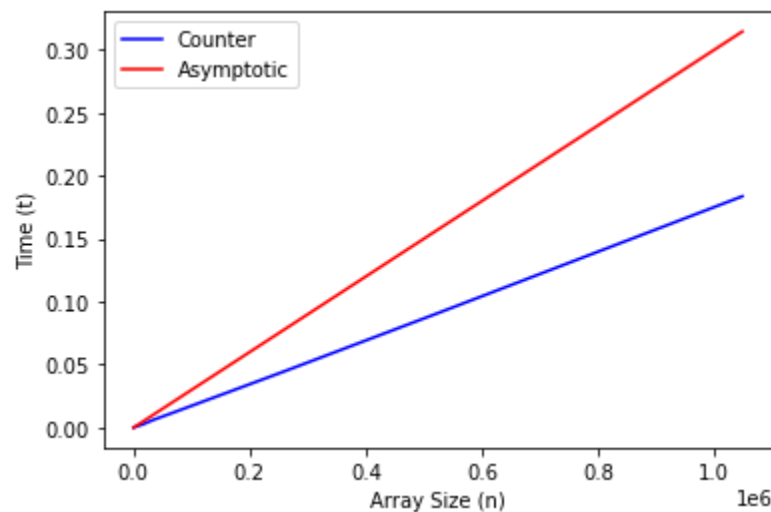
=> C value = $1e-7$

Plot the log-log plot of size vs time ($\log(\text{size})$ and $\log(\text{time})$) plot and identify the slope of the line. Plot also the asymptotical limit in the same plot in log scale. What is the value of slope of the line?



Slope of the line for both the curve is $1.982 \sim 2$.

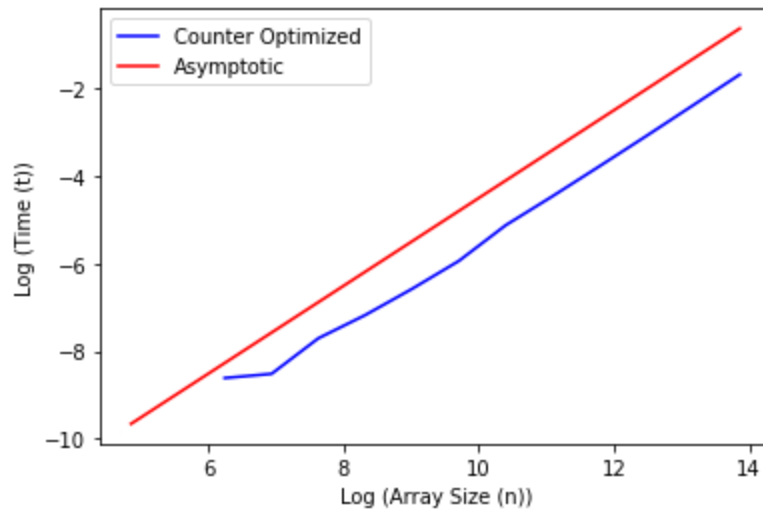
Plot the size of array (n) vs time Taken(t) for the counter optimized (t) function and plot an asymptotical line which denotes the theoretical limit for $O(N)$.



Identify the C value that was used to construct the asymptotical line in the previous graph

=> C value = $3e-7$

Plot the log-log plot of size vs time ($\log(\text{size})$ and $\log(\text{time})$) plot and identify the slope of the line. plot also the asymptotical limit in the same plot in log scale. What is the value of slope of the line?

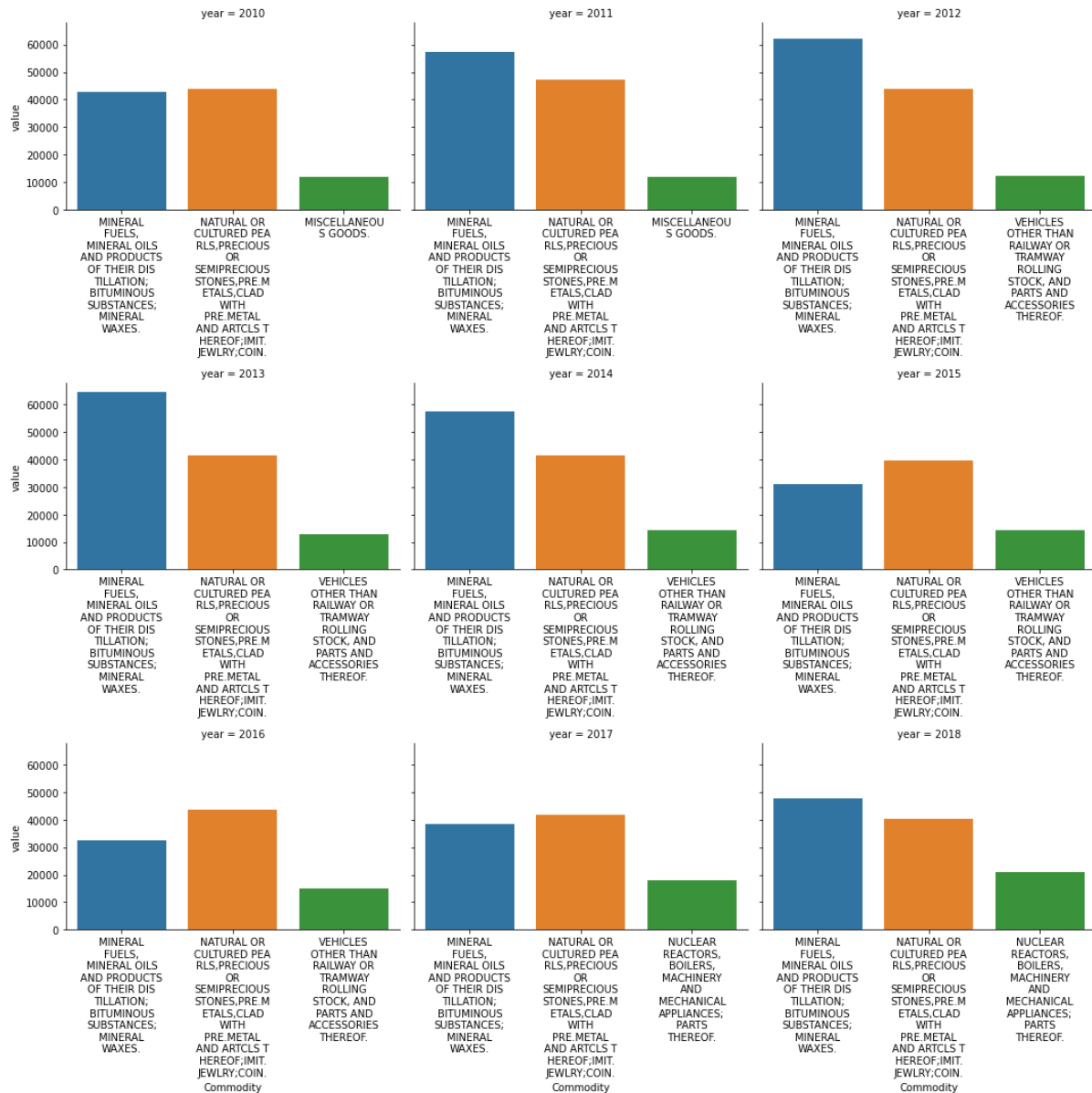


=> Slope of both line is $0.785 \sim 1$.

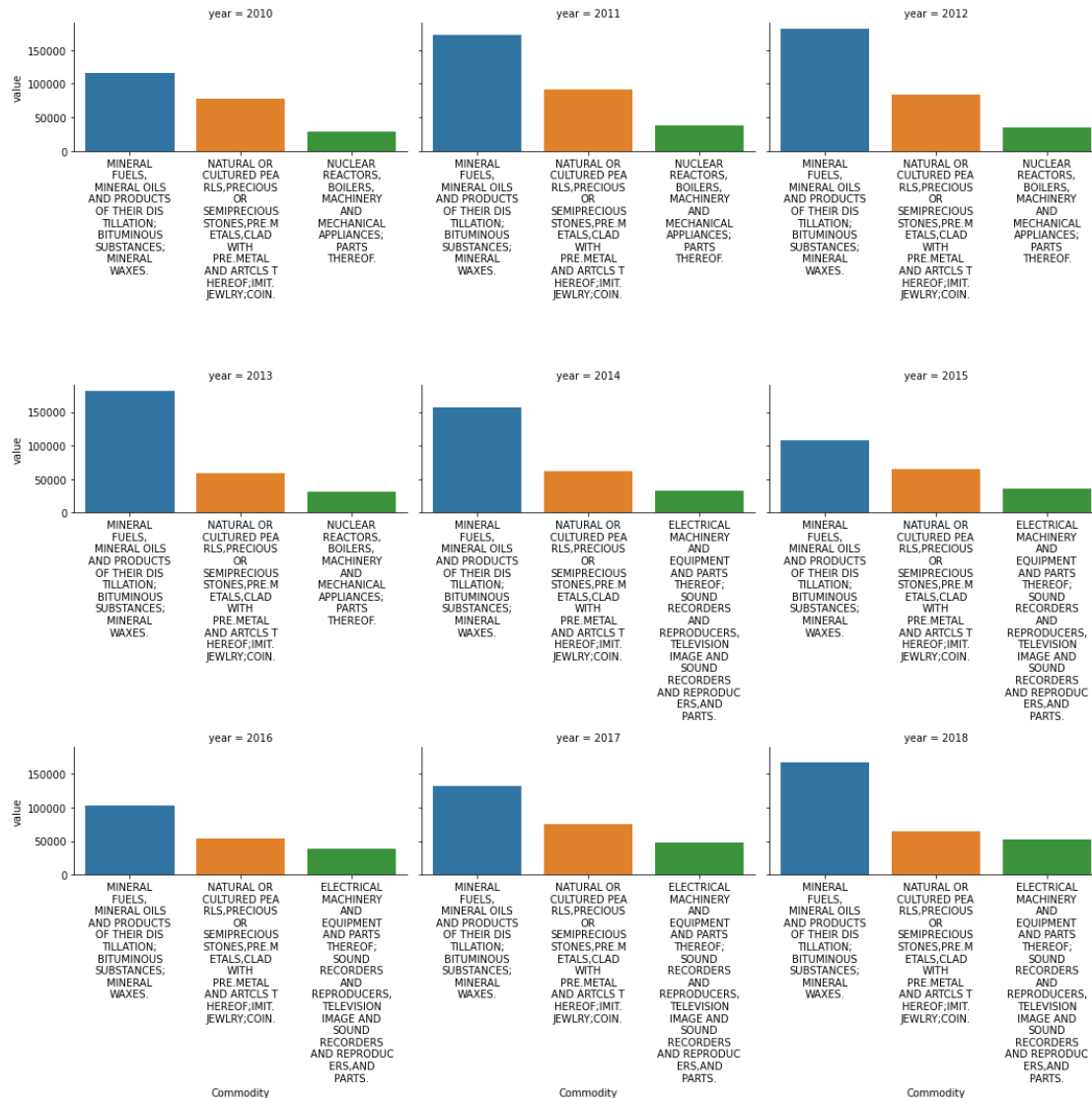
A multiple line chart with the total export/import value across years from 2010 to 2018. (x - years, y- total value of export and import)



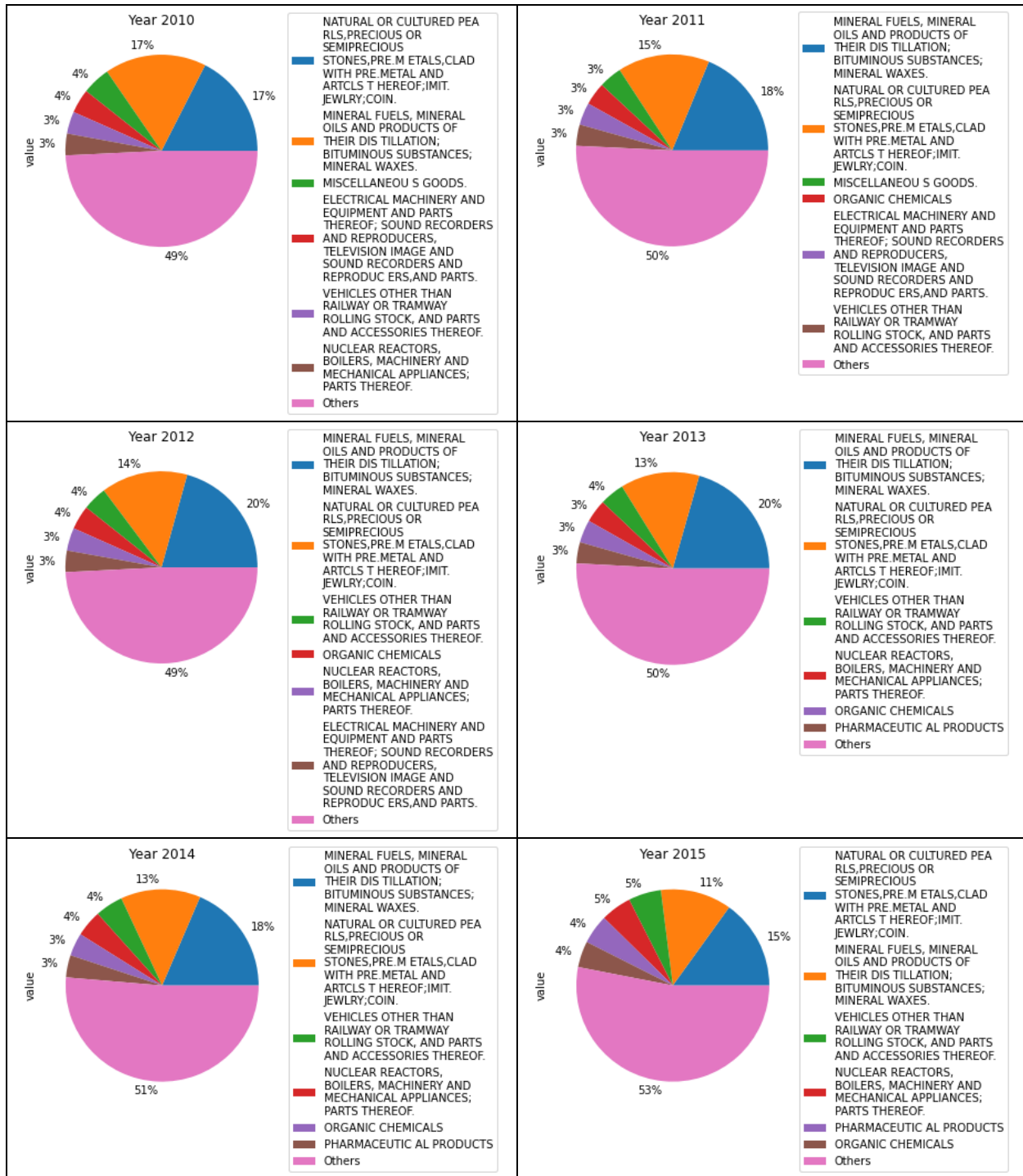
A multiple bar chart (stacked next to each other) with the top three export categories which were exported highest each year. [x - years, y - 3 bars with total export value of those categories]

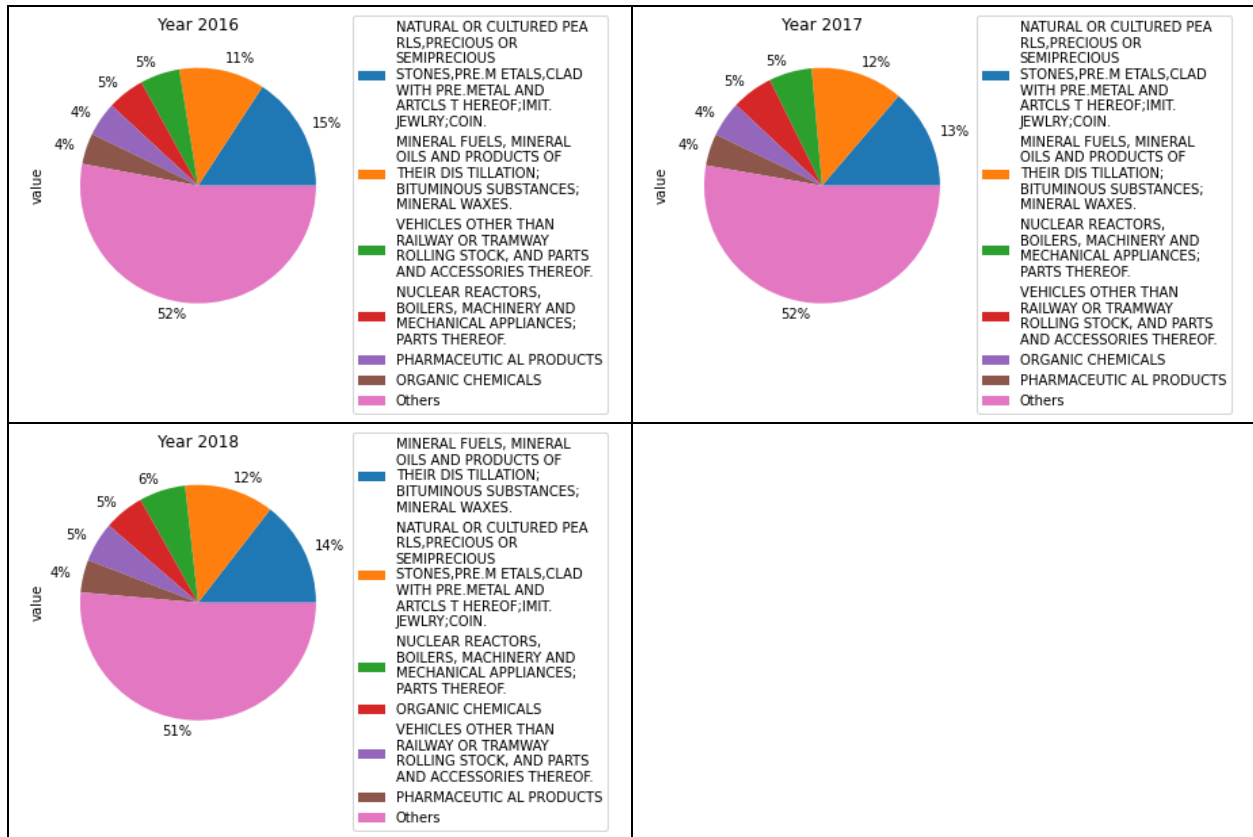


A multiple bar chart (stacked next to each other) with the top three import categories which we imported highest value from each year. [x - years, y - 3 bars with total import value of those categories]



A pie chart of average export value across all years per each category (only top 6, include all the remaining in “others” category)





A simple forecast of the total export and import value during the year 2019 and 2020

=> Used Autoregressive and moving average combination (ARIMA) for forecasting.

```

from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.ar_model import AutoReg
exports = pd.read_csv("ProblemSet_2/Questions/export.csv")
imports = pd.read_csv("ProblemSet_2/Questions/import.csv")

sd = exports.groupby("year").sum()["value"]
sd.index = pd.to_datetime(sd.index, format = "%Y")
# fit model
model = ARIMA(sd, order=(1, 1, 1))
model_fit = model.fit()
# make prediction
y_pred_export = model_fit.predict(start = '2010-01-01', end='2020-01-01')
plt.figure()
plt.plot(sd, label="Orig")
plt.plot(y_pred_export, label="Predicted")
plt.title("Export Forecast")

sd = imports.groupby("year").sum()["value"]
sd.index = pd.to_datetime(sd.index, format = "%Y")
# fit model
model = ARIMA(sd, order=(1, 1, 1))
model_fit = model.fit()
# make prediction
y_pred_import = model_fit.predict(start = '2010-01-01', end='2020-01-01')
plt.figure()
plt.plot(sd, label="Orig")
plt.plot(y_pred_import, label="Predicted")
plt.title("Import Forecast")

```

