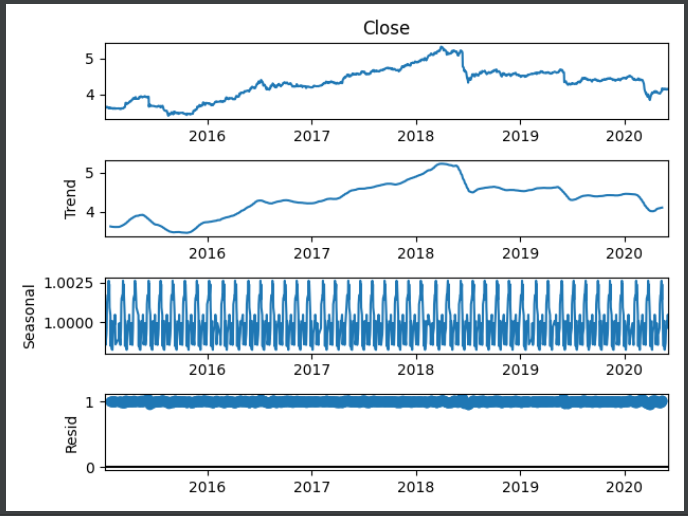
Câu 1:

1. Phân rã dữ liệu chuỗi thời gian này? Nhận xét?

decompose\_results=seasonal\_decompose(df\_train, period=30, model="multiplicative")

decompose\_results.plot()

plt.show()



Phân rã dữ liệu chuỗi thời gian thành 4 thành phần. Trực quan hóa dữ liệu đã phân rã bằng biểu đồ plot như hình bên trên. Ta có thể thấy riêng tính xu hướng, tính mùa vụ

b) Kiểm định tính dừng trên dữ liệu này.

def adf\_test(data):

indices = ["ADF: Test statistic", "p value","H of lags", "H of observatión"]

test = adfuller(data, autolag = "AIC")

results = pd.Series(test[:4], index=indices)

for key, value in test[4].items():

results[f"Critical Value ({key})"] = value

return results

def kpss\_test(data):

indices = ["KPSS: Test statistic","p value","H of lags" ]

test = kpss(data)

results = pd.Series(test[:3], index=indices)

for key, value in test[3].items():

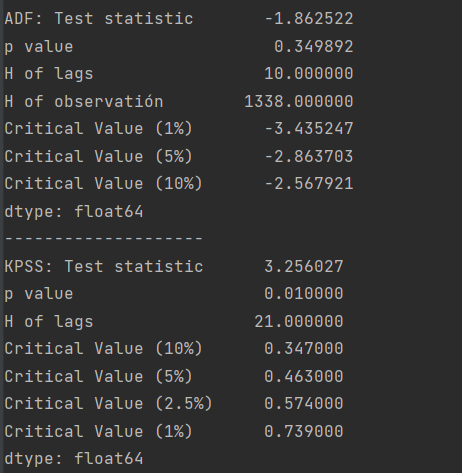
results[f"Critical Value ({key})"] = value

return results

print(adf\_test(df\_train))

print("----"\*5)

print(kpss\_test(df\_train))



Từ kết quả kiểm định qua ADF và KPSS có thể thấy p-value>0.05, và giá trị kiểm định bé hơn giá trị tới hạn trong trong kiểm định ADF vì vậy dữ liệu không có tính dừng

c) Kiểm định tự tương quan trên dữ liệu này

#%%-kiem dinh tự tương quan (Auto correlation)

pd.plotting.lag\_plot(df\_train)

plt.show()

#%%

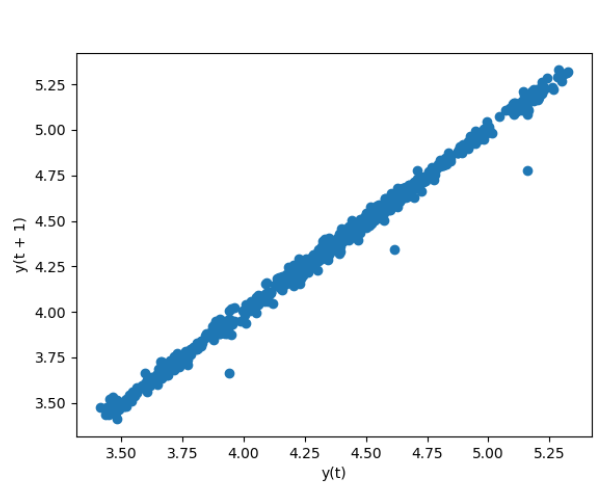
plot\_pacf(df\_train)

plt.show()

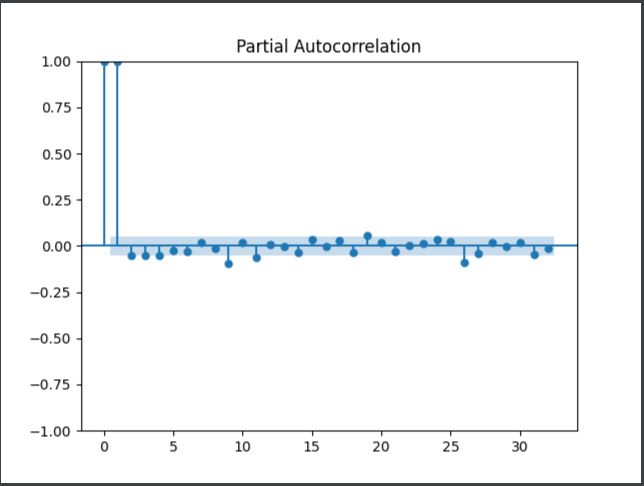
#%%

plot\_acf(df\_train)

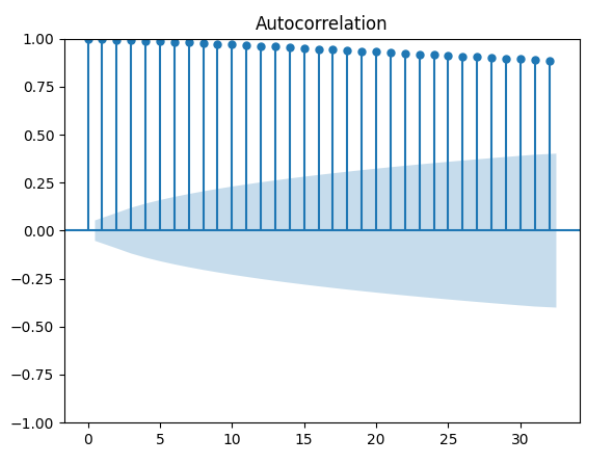
plt.show()



Từ biểu đồ có thể thấy dữ liệu ngày hôm nay và dữ liệu ngày hôm sau có sự tương quan mạnh



Trước khi chuyển đổi về chuỗi dừng, đồ thị tự tương quan PACF cho ta thấy được đồ thị có tương quan bậc 1



Còn với đồ thị ACF cho thấy được sự tương quan mạnh giữa dữ liệu giữa các kỳ trong quá khứ và có khuynh hướng giảm dần đều

d) Áp dụng mô hình ARIMA trên dữ liệu này, sử dụng mô hình để dự báo giá cổ phiếu trong tương lai.

#%% - Xac dinh tham so p, d, q cho mo hinh ARIMA

stepwise\_fit=auto\_arima(df\_train, trace=True, suppress\_warnings=True)

print(stepwise\_fit.summary())

stepwise\_fit.plot\_diagnostics(figsize=(15,8))

plt.show()

#%%-create model

model=sm.tsa.arima.ARIMA(df\_train, order=(2,1,1))

fitted=model.fit()

print(fitted.summary())

#Dự báo giá trị cổ phiếu trong tương lai

preds=fitted.forecast(len(df\_test), alpha=0.05)

preds.index = df\_test.index

plt.figure(figsize=(16,10),dpi=150)

plt.plot(df\_train,color="r",label="Training data")

plt.plot(df\_test, color="orange", label="Actual stock price")

plt.plot(preds, color="b", label="Predicted stock price")

# plt.fill\_between(lower\_series.index, lower\_series, upper\_series, color="b", alpha=.10)

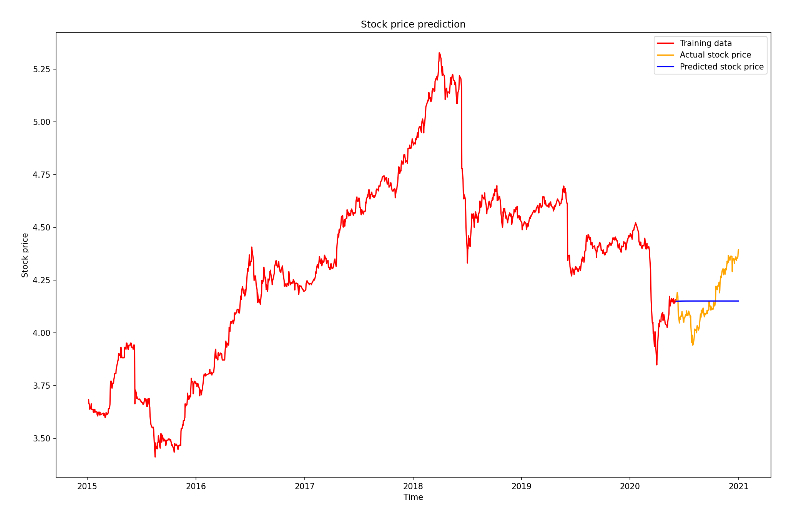
plt.title("Stock price prediction")

plt.xlabel("Time")

plt.ylabel("Stock price")

plt.legend()

plt.show()



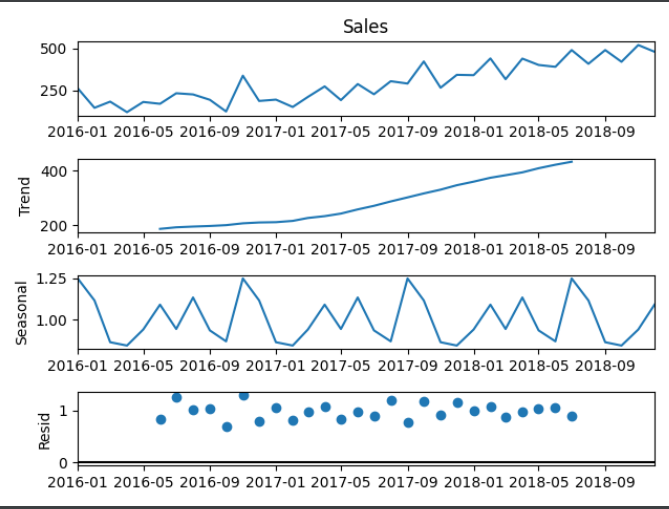
Câu 2:

a) Phân rã dữ liệu chuỗi thời gian này? Nhận xét?

decompose\_result=seasonal\_decompose(df\_sales, period=10, model="multiplicative")

decompose\_result.plot()

plt.show()



Phân rã dữ liệu chuỗi thời gian thành 4 thành phần. Trực quan hóa dữ liệu đã phân rã bằng biểu đồ plot như hình bên trên. Ta có thể thấy dữ liệu có xu hướng tăng và có tính xu hướng, tính mùa vụ, tĩnh ngẫu nhiên.

b) Kiểm định tính dừng trên dữ liệu này.

#%%- Kiểm định tính dừng của dữ liệu

def adf\_test(data):

indices = ["ADF: Test statistic", "p value","H of lags", "H of observatión"]

test = adfuller(data, autolag = "AIC")

results = pd.Series(test[:4], index=indices)

for key, value in test[4].items():

results[f"Critical Value ({key})"] = value

return results

def kpss\_test(data):

indices = ["KPSS: Test statistic","p value","H of lags" ]

test = kpss(data)

results = pd.Series(test[:3], index=indices)

for key, value in test[3].items():

results[f"Critical Value ({key})"] = value

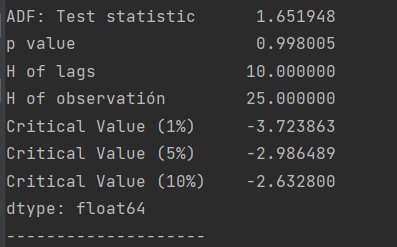
return results

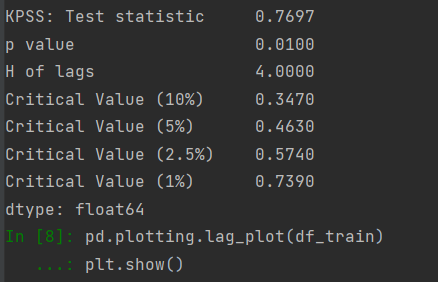
print(adf\_test(df\_sales))

print("----"\*5)

print(kpss\_test(df\_sales))

Từ bảng kết quả ở bên dưới bằng phương pháp kiểm định ADF và KPSS ta có thể thấy dữ liệu không có tính dừng vì với ADF test,trị tuyệt đối giá trị kiểm định =1,65 bé hơn trị tuyệt đối của các giá trị tới hạn





c) Kiểm định tự tương quan trên dữ liệu này.

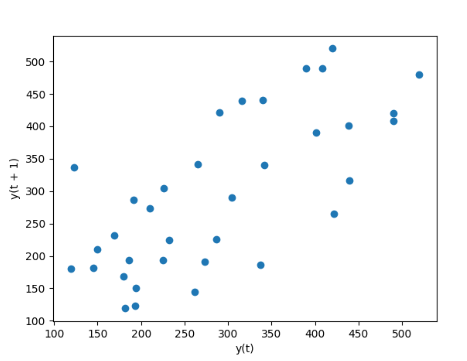
pd.plotting.lag\_plot(df\_train)

plt.show()

#%%

plot\_acf(df\_train)

plt.show()



Từ biểu đồ trên có thể thấy dữ liệu doanh thu ngày hôm nay không có sự tương quan so với ngày hôm sau.

Vì là chuỗi nhiễu trắng nên không thể tiến hành dự báo nên mình dừng lại ở bước này