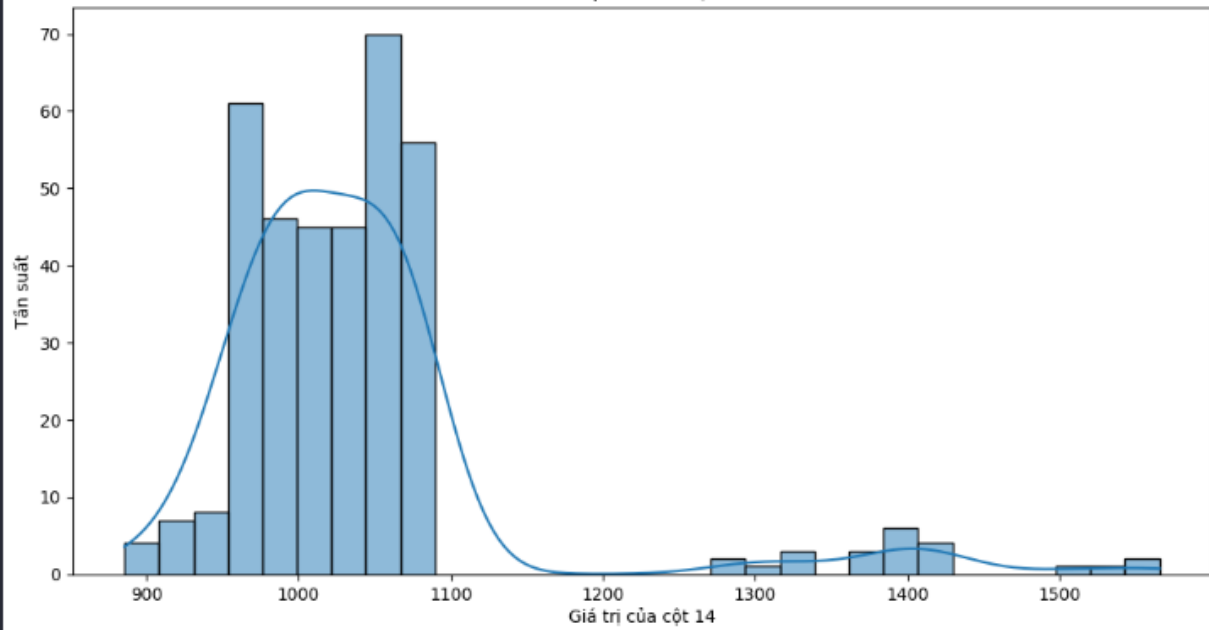
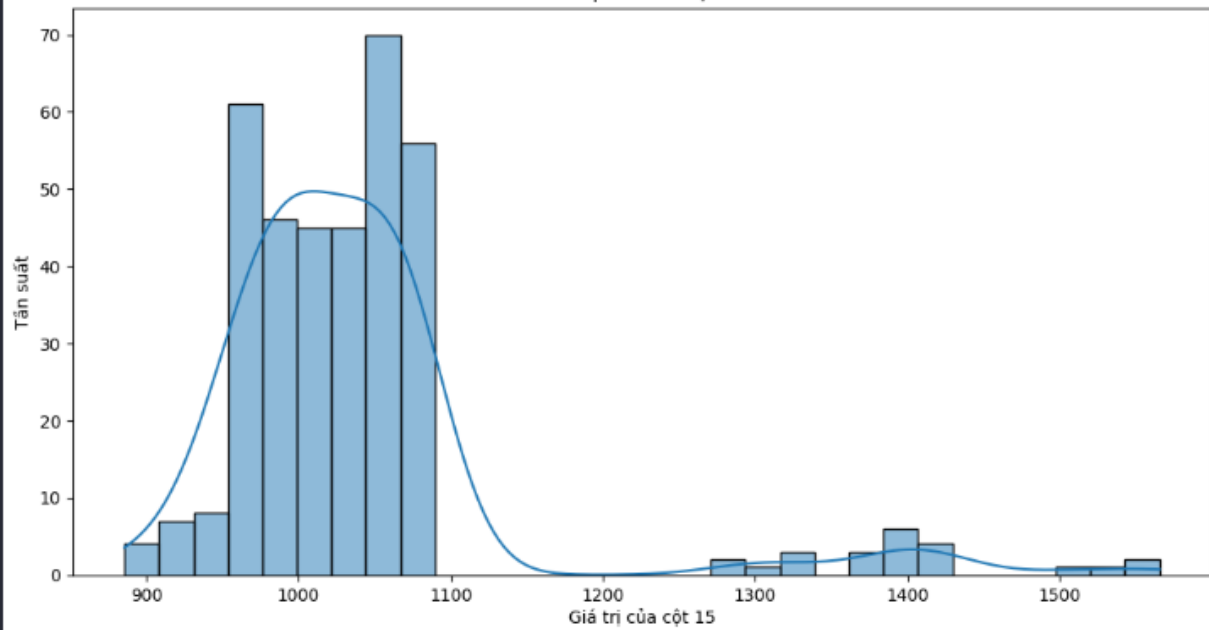


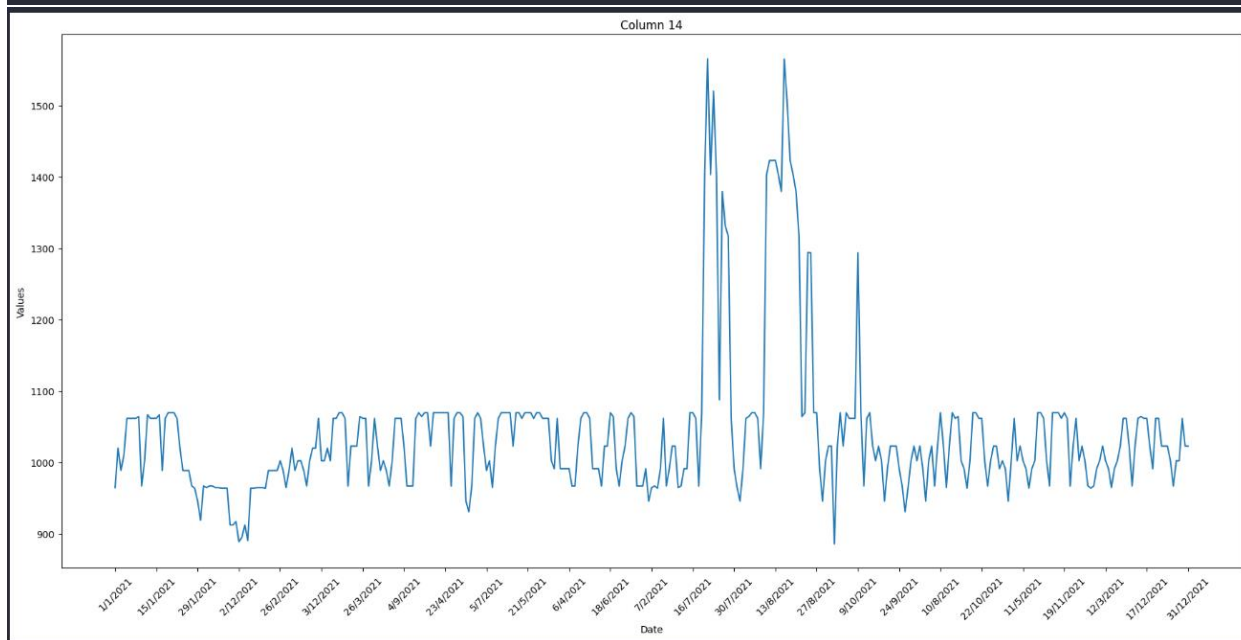
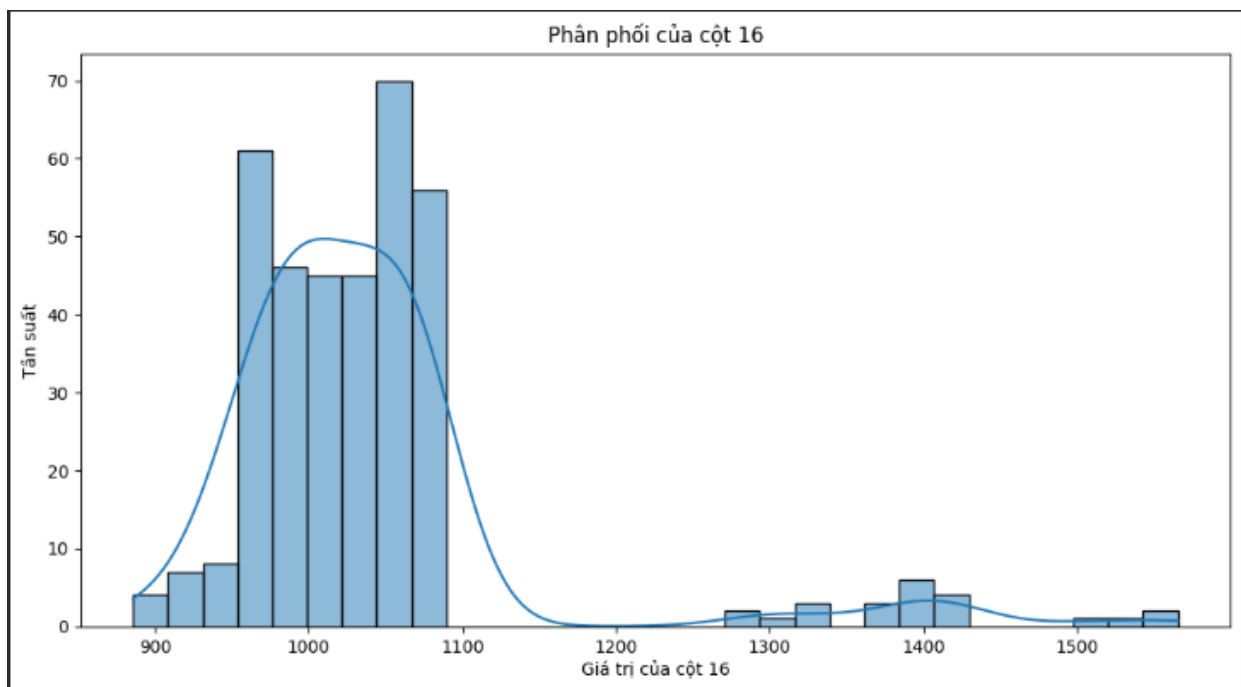
Biểu diễn EDA của các cột 14, 15, 16

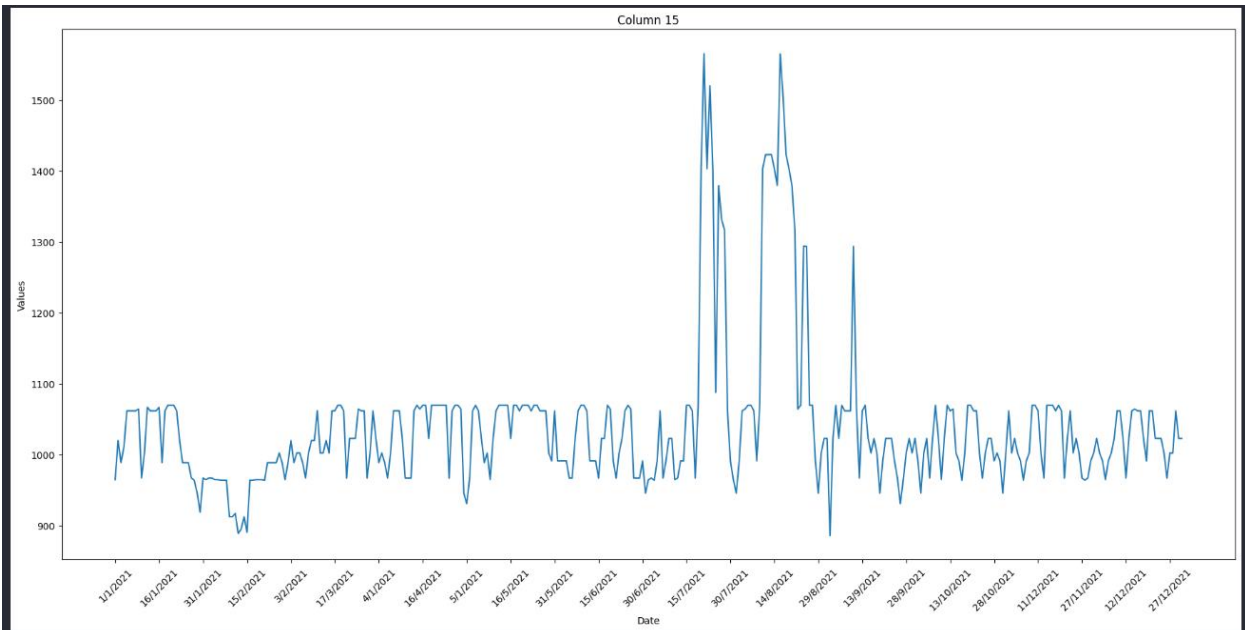
Phân phối của cột 14



Phân phối của cột 15

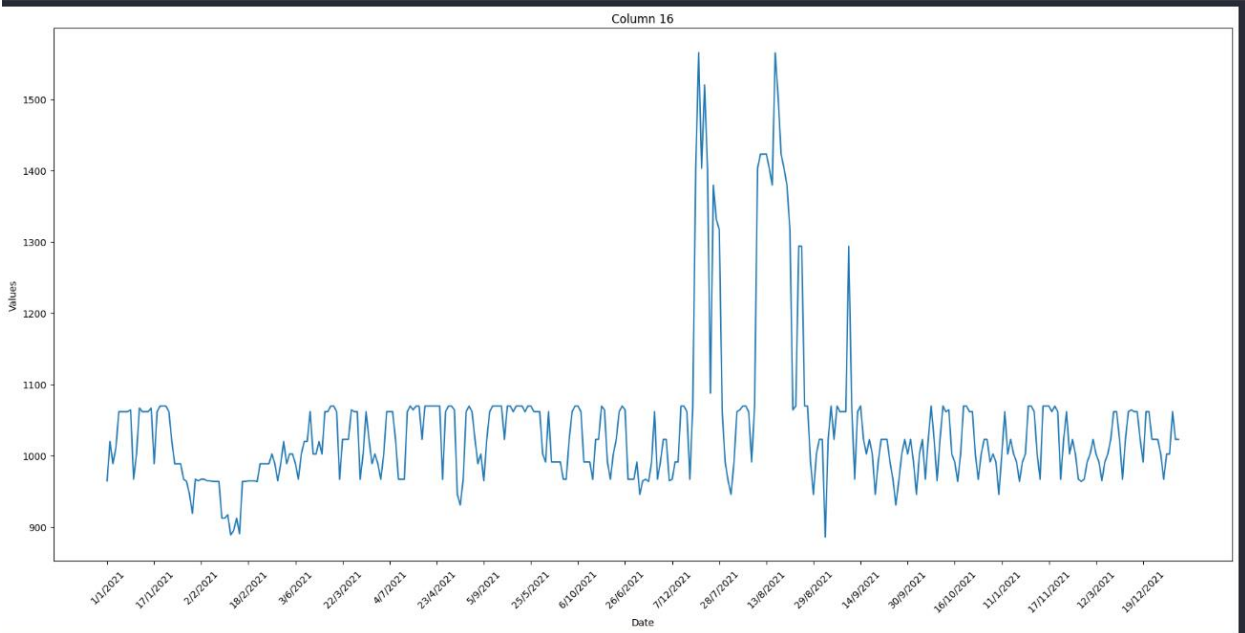




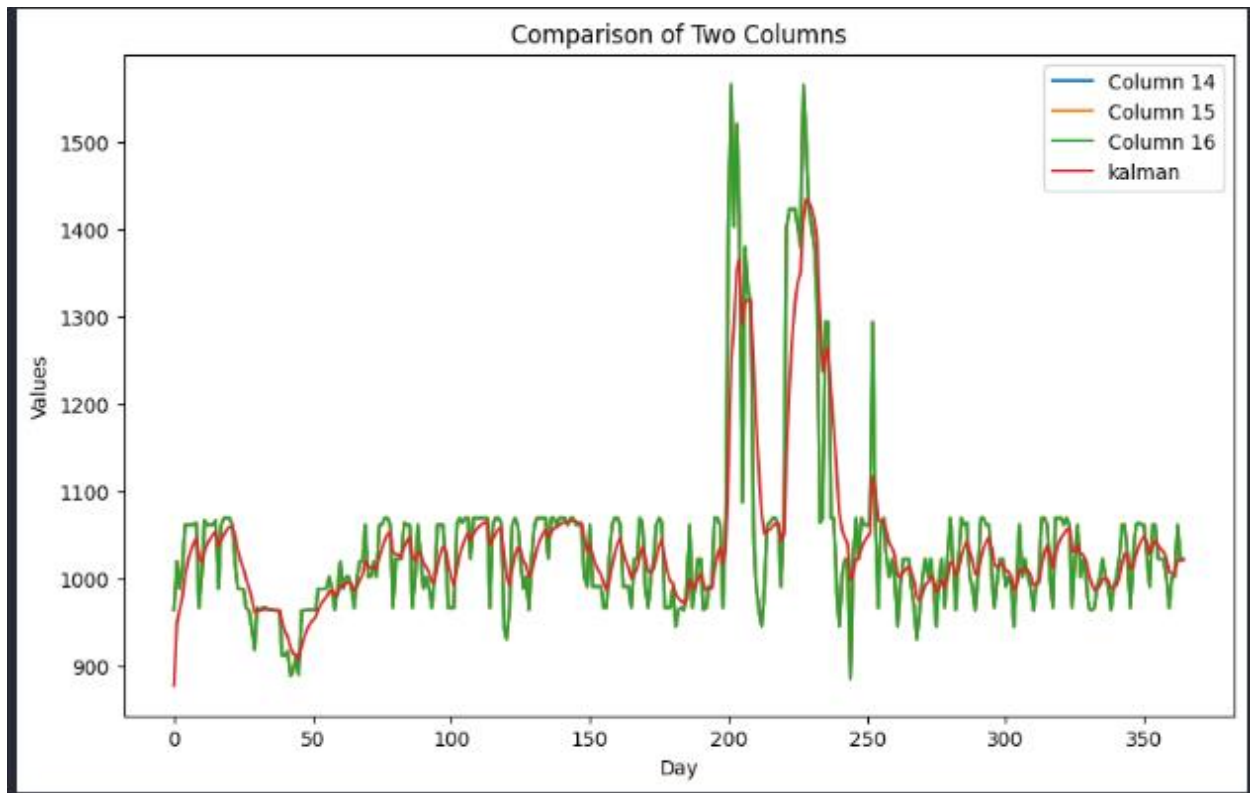


```
# biểu đồ biểu diễn giá trị của cột 16 theo thời gian
df_16 = df['16']
df16_bin = df['16']
plt.figure(figsize=(22,18))
plt.plot(df['Day'],df['16'])
plt.title("Column 16")
plt.xlabel("Date")
plt.ylabel("Values")
plt.xticks(df['Day'][::16], df['Day'][::16], rotation=45)
plt.show()
```

0.2s



Dùng mô hình Kalman để dự đoán



Dùng mô hình SARIMA để dự đoán và vẽ biểu đồ đánh giá:

SARIMAX Results						
=====						
Dep. Variable:	14	No. Observations:	365			
Model:	SARIMAX(1, 1, 1)x(1, 1, 1, 12)	Log Likelihood	-1966.358			
Date:	Tue, 21 May 2024	AIC	3942.716			
Time:	11:50:25	BIC	3962.035			
Sample:	0	HQIC	3950.404			
	- 365					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

ar.L1	0.9986	35.701	0.028	0.978	-68.974	70.971
ma.L1	-0.9993	71.658	-0.014	0.989	-141.446	139.448
ar.S.L12	-0.2391	0.038	-6.295	0.000	-0.314	-0.165
ma.S.L12	-0.9999	18.599	-0.054	0.957	-37.454	35.454
sigma2	3649.2068	1.87e+05	0.020	0.984	-3.62e+05	3.69e+05
=====						
Ljung-Box (L1) (Q):	1.29	Jarque-Bera (JB):	981.59			
Prob(Q):	0.26	Prob(JB):	0.00			
Heteroskedasticity (H):	1.62	Skew:	0.48			
Prob(H) (two-sided):	0.01	Kurtosis:	11.12			
=====						
Warnings:						
[1] Covariance matrix calculated using the outer product of gradients (complex-step).						

SARIMA - MSE: 7108.286061038034
SARIMA - MAE: 43.31316097086826
SARIMA - RMSE: 84.31065212082062

