ANLT 207
Time Series Analysis
Assignment #3

For all problems below: Assume the 'robberies.csv' dataset from Brownlee Chapter 30
Show your work with a PY or an IPYNB file.

1	Perform a complete Box-Jenkins analysis to arrive at what you believe to be the optimum ARIMA(p,d,q) model		
	a.	Plot the series.	1
	b.	Perform a Dickey-Fuller test on the series. Is the series stationary?	2
	c.	Plot the ACF & PACF and observe the trends.	1
	d.	Perform differencing on the series.	5
	e.	Is the differenced series stationary?	1
	f.	Plot the ACF & PACF of the differenced data and select the p & q orders.	5
	g.	Run an ARIMA analysis with the original series and the (p,d,q) you've deduced from above.	5
	h.	Plot the residuals.	1
	i.	Plot the residual distribution (kind='kde').	1
	j.	Print the residuals summary.	1
	k.	Plot the ACF & PACF and observe for residual correlations.	1
	l.	Compare the model results series to the actuals:	
		i Convert the residuals to the series estimated by ARIMA (hint: resids = y - yhat).	5
		ii Add back the differencing to compare with the original series.	5
		iii Plot the actual and modeled results on the same graph.	1
			35 pts.

2 Create a AIC grid search and run 'robberies.csv' through it to find the optimum p,d,q order
What is the optimum p,d,q order? 10 pts.

3 Compare the results of your orginal (p,d,q) estimate to the results from the grid search (p,d,q). 10 pts.
Compare the residuals summaries.
Compare the graphs generated by the script made in 1.l above.

4 Forecast 'robberies.csv' based on an ARIMA(2,1,1) model: 20 pts.
Split the series into 75% train, 25% test
a. Perform a dynamic forecast for all points in 'test'
Plot the forecasted data and the prediction intervals on top of the actual data
b. Perform a rolling forecast for all points in 'test'
Plot the forecasted data and the prediction intervals on top of the actual data