Florida Atlantic University
COLLEGE OF BUSINESS

ISM 6136 – Datamining/Predictive Analytics
Jacob Perrone
Class Assignment 10
Dr. Bharti Sharma

5 points

TASK: Time series forecasting – Data Mining Task using XLMiner

Perform time series forecasting on the Average Income of Tax payers in US dataset determine the best model for any 'two' of the states. Provide screen shots with your explanation below.

1. Perform Data Partitioning on the dataset (select Training set at least 60 % or higher)
A 70/30 and 80/20 splits were used for each model.

Variables		
# Selected Variables	25	
Time Variable	Year	
Value Variables	CA, CT, DC, DE, FL, IA, ID, IL, IN, KA	

Partitioning Parameters	
Partitioning type	SEQUENTIAL
Ratio - Training	0.7
Ratio - Validation	0.3

Partition Summary

Partition	# Records
Training	50
Validation	21

Partitioned Data

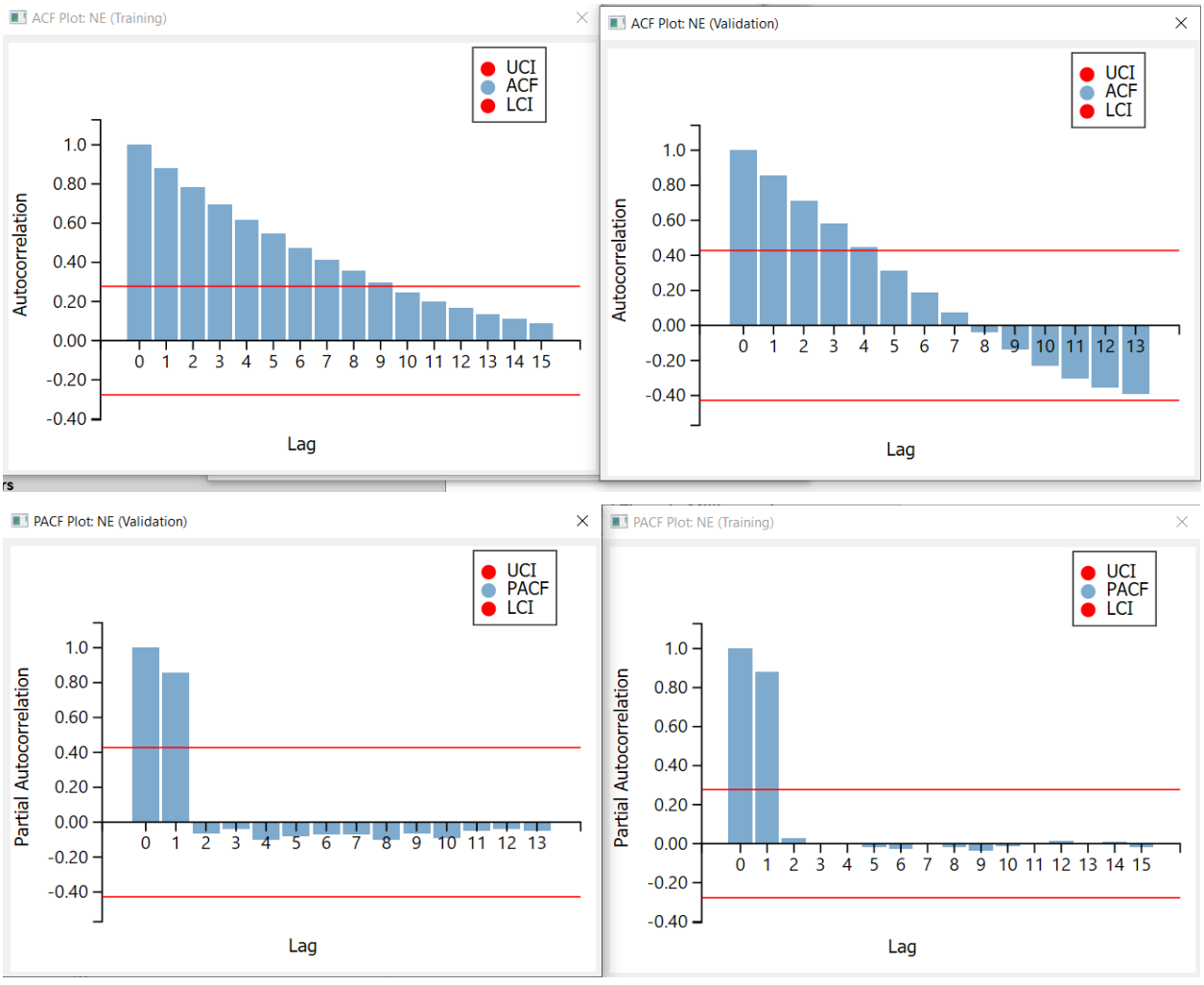
# Selected Variables	25	
Time Variable	Year	
Value Variables	CA	CT	DC	DE	FL

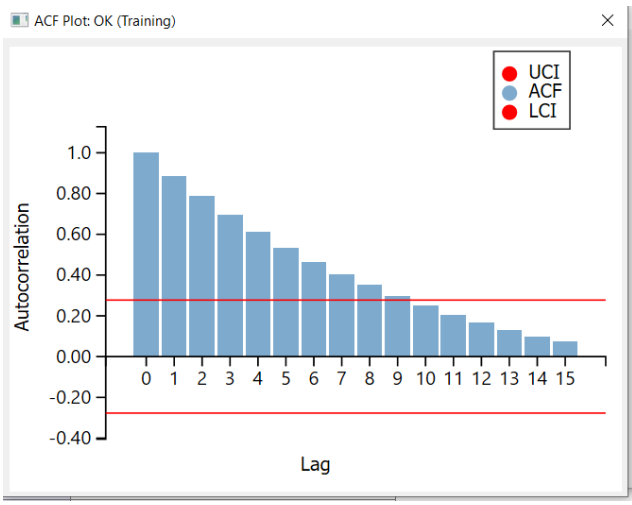
Partitioning Parameters	
Partitioning type	SEQUENTIAL
Ratio - Training	0.8
Ratio - Validation	0.2

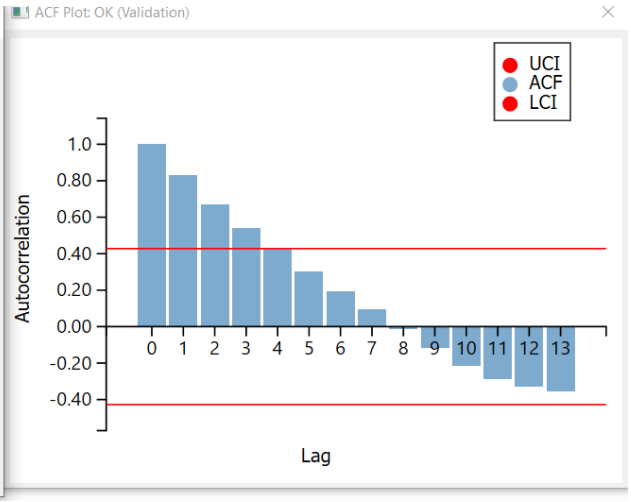
Partition Summary

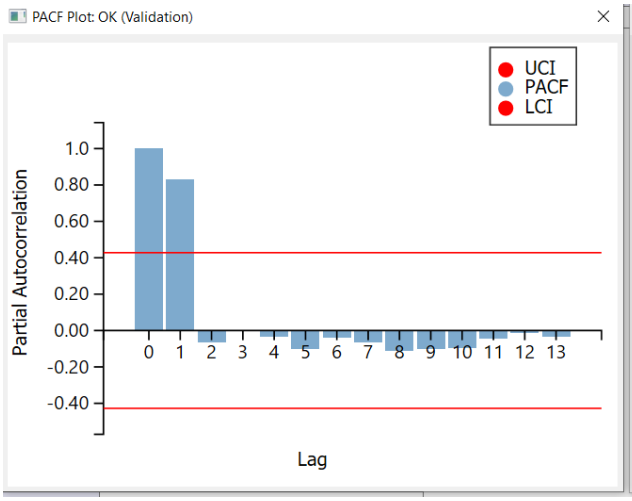
Partition	# Records
Training	57
Validation	14

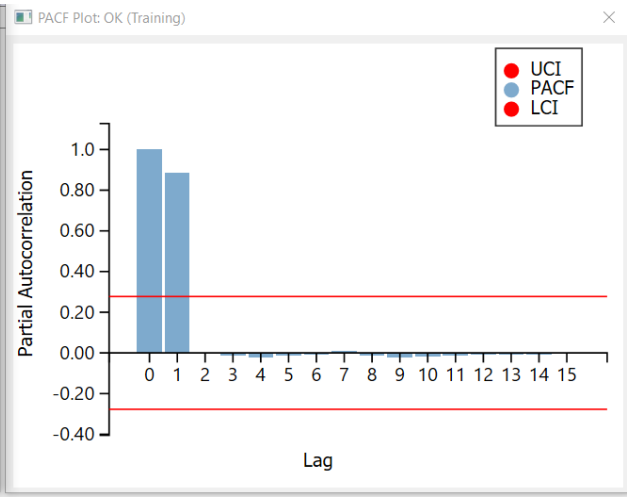
2. Perform Lag analysis and explain the ACF and PACF plots. Take Training lag of 15 and Validation lag < 14 as the max lag (lag analysis window).











Both chosen states ACF show a downward trend and exhibit a similar pattern to one another. Both chosen states PACF show a repeat in the first two lags which indicates seasonality. Both states show the same trend and seasonality so similar models can be used for both.

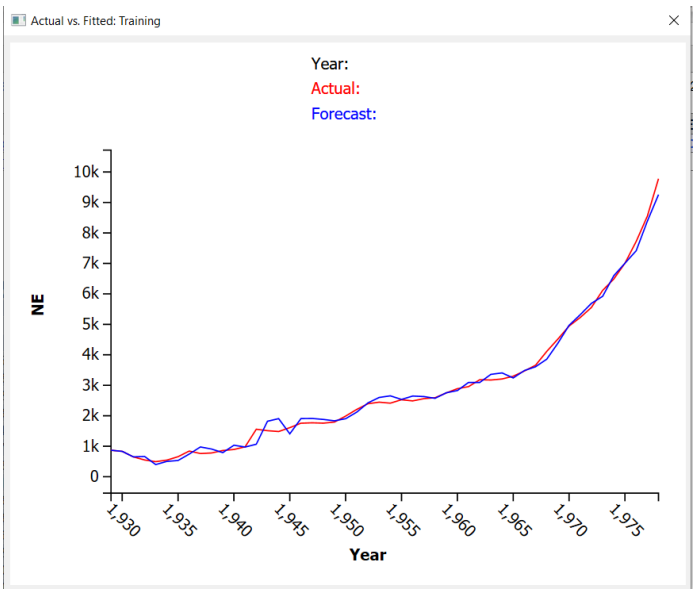
3. Develop 2 ARIMA models for each of the three states by changing the partitioning 'or' p,d,q parameters 'or' iterations. So total you should have four models (2 models for each State, choose the best model for each State). Create a table to show all the models.

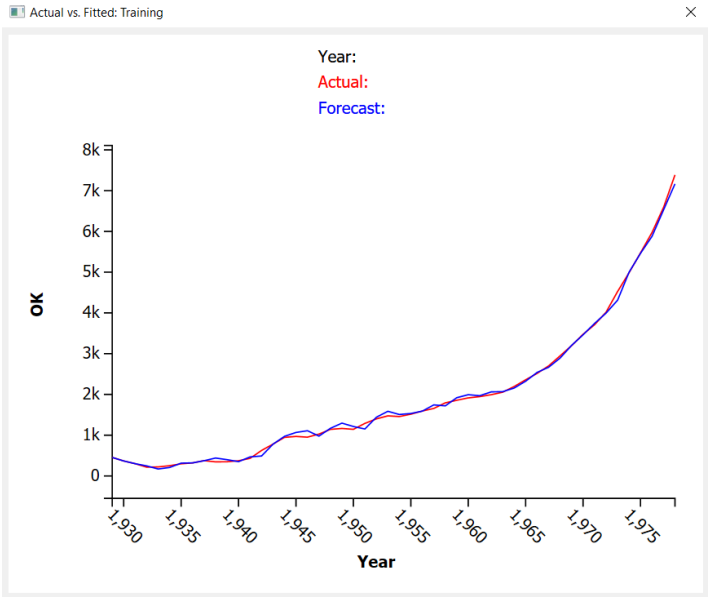
	State Chosen	Partitioning	Training P Value	Training MSE	Validation MSE
Model 1	NE	75/35	0.117749989	30921.25056	422431.0741
Model 2	OK	75/35	1.15282E-08	5568.22443	1070349.145
Model 3	NE	80/20	0.041765627	50136.41846	1004332.935
Model 4	OK	80/20	0	47801.00359	2929553.991

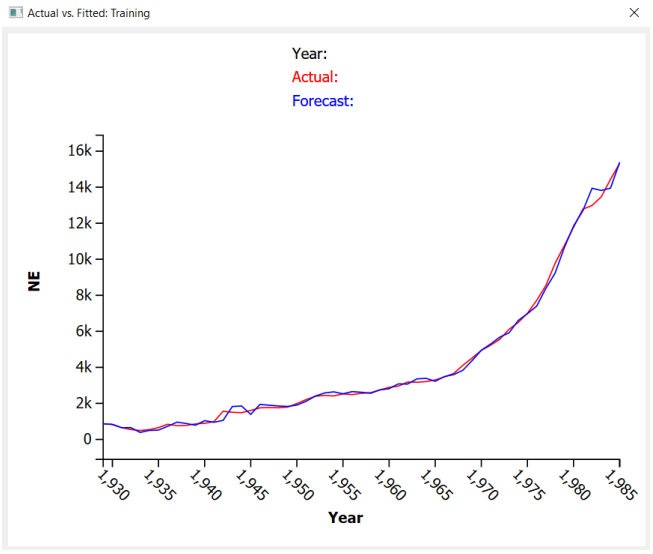
The best models per state would be model 3 for NE, and model 4 for OK. These models were chosen based on criteria that P value is lowest and the MSE values are low.

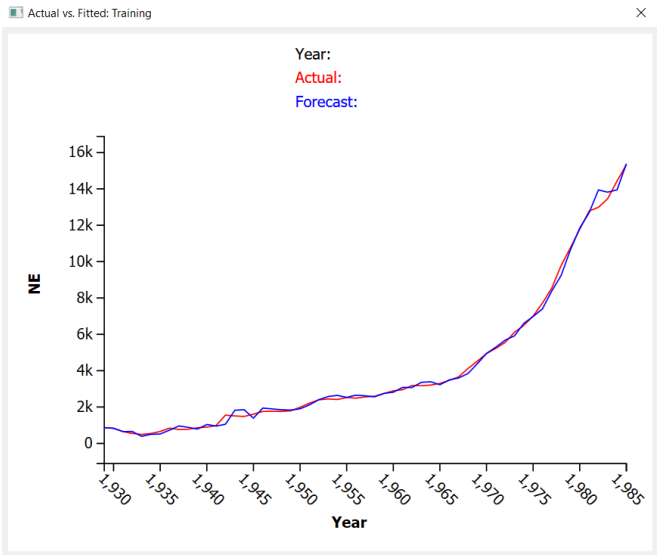
4. Compare the models and determine the best model for each of the two states - based on the following – provide screen shots and your comments for the following model selection criteria:

a) Forecast and Actual plot for Training
Model 1-4 respectively for each section

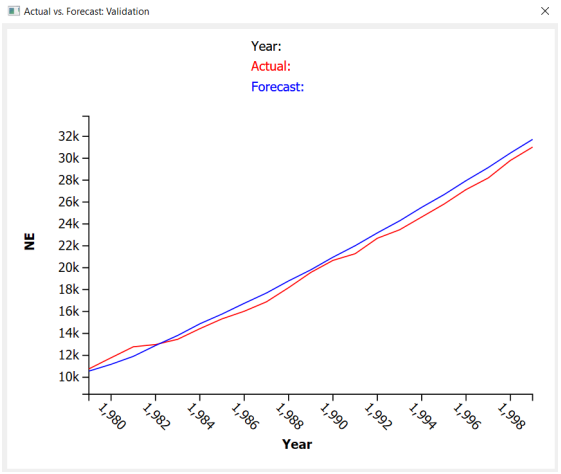




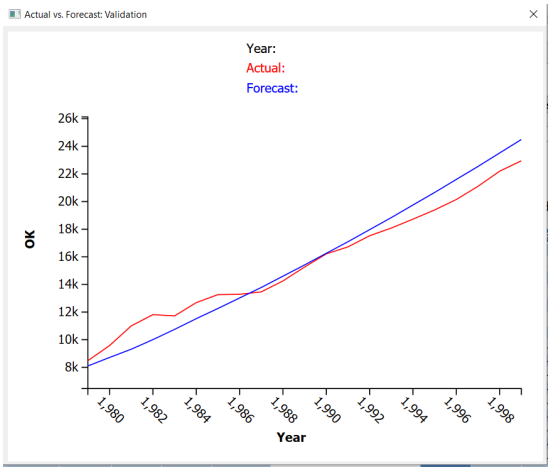




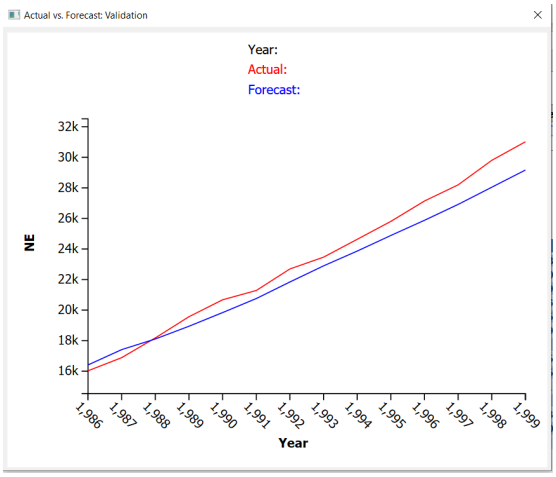
b) Forecast and Actual plot for Validation (select any year and compare the results with respect to that in all the models)



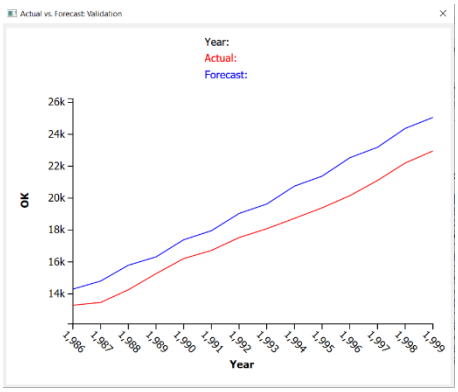
Model 1: year 1990 Act:20674 Forecast:20964



Model 2: year 1990 Act:16214 Forecast:16259



Model 3: year 1990 Act:20674 Forecast:198387



Model 4: year 1990 Act:16214 Forecast:17391

c) MSE training and validation error d) p-values of the coeff and AR1, AR2 etc

	State Chosen	Partitioning	Training P Value	Training MSE	Validation MSE
Model 1	NE	75/35	0.117749989	30921.25056	422431.0741
Model 2	OK	75/35	1.15282E-08	5568.22443	1070349.145
Model 3	NE	80/20	0.041765627	50136.41846	1004332.935
Model 4	OK	80/20	0	47801.00359	2929553.991

Record ID	Coeff	Std-Dev	p-value
Const	113.4889	65.49610038	0.083139
AR 1	0.224219	0.143336459	0.11775
AR 2	-0.3807	0.142884211	0.007713
SAR 1	-0.64384	1.414213562	0.648922
SAR 2	-0.38175	1.414213562	0.787209

Record ID	Coeff	Std-Dev	p-value
Const	52.71105	21.29177088	0.013299
AR 1	0.549409	0.096276656	1.15E-08
AR 2	-0.53377	0.096210174	2.89E-08
SAR 1	-0.44966	1.414213562	0.750516
SAR 2	-0.15645	1.414213562	0.911913

Record ID	Coeff	Std-Dev	p-value
Const	92.86883	82.58211889	0.260774
AR 1	0.252589	0.124070423	0.041766
AR 2	-0.4479	0.131134046	0.000636
SAR 1	-0.46526	1.414213562	0.742163
SAR 2	-0.5489	1.414213562	0.69792

Record ID	Coeff	Std-Dev	p-value
Const	51.68649	57.11363598	0.365478
AR 1	0.321427	0.000307598	0
AR 2	-0.4873	0.000153799	0
SAR 1	-0.30063	1.414213562	0.831658
SAR 2	-0.11259	1.414213562	0.936547