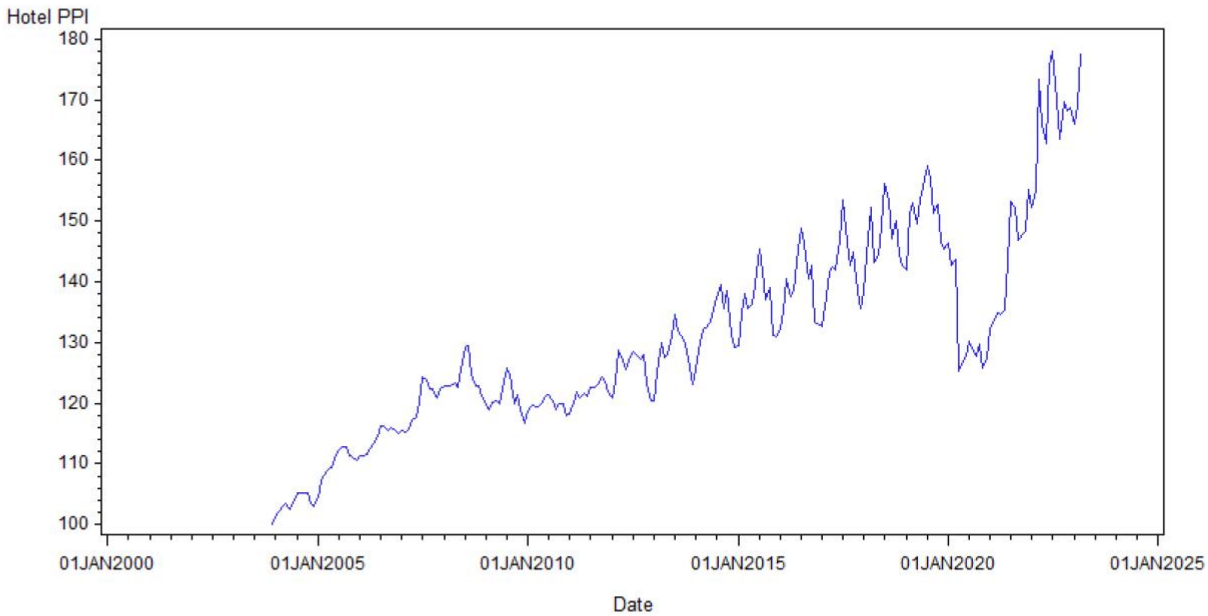
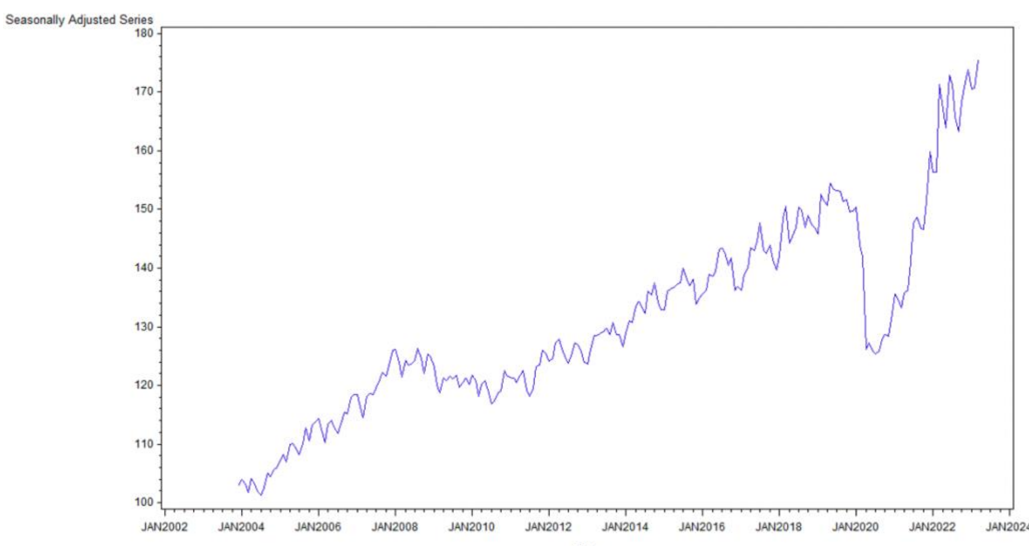
Introduction

This report will look at a dataset of Hotel Producer Price Index (PPI) from December 2003 to March 2023. This dataset was of particular interest to me as my family is in the hotel business. Below is a graph of the dataset.



Seasonal Decomposition

The first analysis done on this data was a decomposition model. As you can see in the above graph of the data a seasonal trend is seen in the middle starting before January 2015 and ending around January 2020. This seasonal trend appears to be increasing over time therefore multiplicative seasonal decomposition was performed.



Above is the deseasonalized data. The trend appears to be a mostly linear trend with hotel PPI seeming to increase over time minus the dip in 2020 for Covid. If you were to draw a line through this graph it would be line going up from the bottom left to the top right indicating a positive linear trend. This is confirmed below with the slope being 0.21753. The deseasonalized model is 105.60246 + 0.21753t.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	49241	49241	974.60	<.0001
Error	230	11621	50.52433		
Corrected Total	231	60862			

Root MSE	7.10805	R-Square	0.8091
Dependent Mean	130.94515	Adj R-Sq	0.8082
Coeff Var	5.42826		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	105.60247	0.93636	112.78	<.0001
t		1	0.21753	0.00697	31.22	<.0001

0.97395
0.99113
1.01160
0.99254
0.99330
1.01529
1.03843
1.02492
1.00095
1.00715
0.97975
0.97101

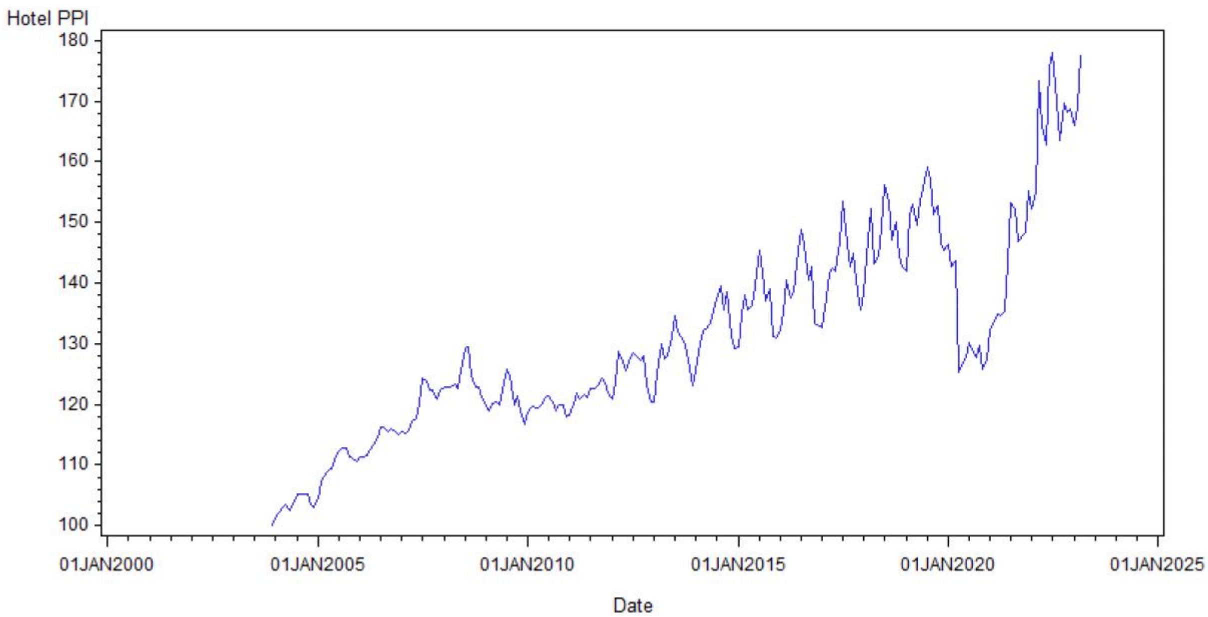
The seasonality factor for each month are shown above on the right. July is the largest at 1.03843 and December is the lowest at 0.97101. This intuitively makes sense as the summer months are much more known for travel with kids having off from school so hotels are able to charge more for their rooms as there is more demand while in winter months people travel less and therefore hotels charge less.

233	.	156.2878	0.9364	142.1616	170.4140	.
234	.	156.5054	0.9424	142.3776	170.6331	.
235	.	156.7229	0.9485	142.5936	170.8522	.

The predictions for the next three months after March 2023 are shown above. The deseasonalized prediction for month 233 is 156.2878. The reseasonalized predication is 156.2878 * (.99254) = 155.1219. The margin of error is (170.4140 – 142.1616)/2= 14.1262.
The 95% confidence interval for month 233 is 155.1219 +- 14.1262 = (140.9957, 169.2481). The deseasonalized prediction for month 234 is 156.5054. The reseasonalized predication is 156.5054 * (.99330) = 155.4568. The margin of error is (170.6331– 142.3776)/2= 14.1278.
The 95% confidence interval for month 234 is 155.4568 +- 14.1278 = (141.329, 169.5846).

Finally, the deseasonalized prediction for month 235 is 156.7229. The reseasonalized predication is 156.7229* (1.01529) = 159.1700. The margin of error is (170.8522– 142.5936)/2= 14.1293. The 95% confidence interval for month 235 is 159.1700 +- 14.1293 = (145.0407, 173.2993).

Exponential Smoothing

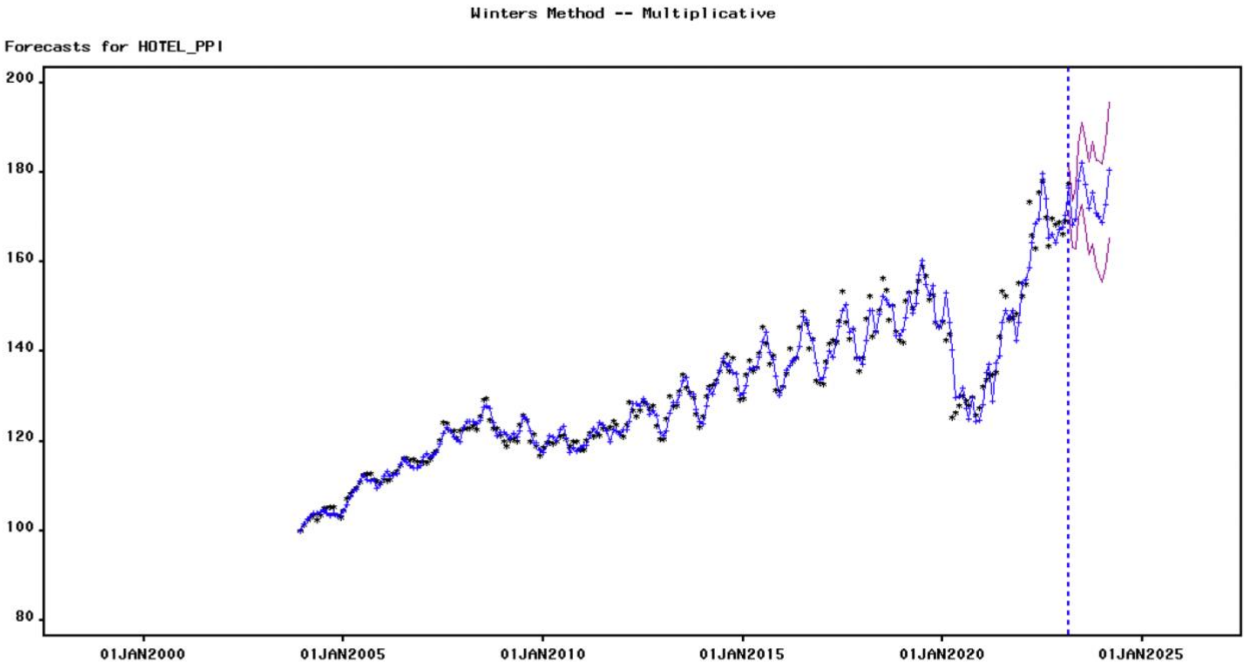


Again, above is the graph of our hotel dataset from January 2003 to March 2023. There appears to be an overall linear and in the middle of the dataset an increasing seasonal trend. For these reasons the Holts winter multiplicative model was chosen to be the most appropriate type of exponential smoothing analysis to use on this data.

Winters Method -- Multiplicative

Model Parameter	Estimate	Std. Error	T	Prob>|T|
LEVEL Smoothing Weight	0.79327	0.0416	19.0678	<.0001
TREND Smoothing Weight	0.00100	0.0062	0.1617	0.8717
SEASONAL Smoothing Weight	0.99900	0.2364	4.2251	<.0001
Residual Variance (sigma squared)	7.01698	.	.	.
Smoothed Level	171.94043	.	.	.
Smoothed Trend	0.25008	.	.	.
Smoothed Seasonal Factor 1	0.96706	.	.	.
Smoothed Seasonal Factor 2	0.98918	.	.	.
Smoothed Seasonal Factor 3	1.03182	.	.	.
Smoothed Seasonal Factor 4	0.97762	.	.	.
Smoothed Seasonal Factor 5	0.98432	.	.	.
Smoothed Seasonal Factor 6	1.03095	.	.	.
Smoothed Seasonal Factor 7	1.05245	.	.	.
Smoothed Seasonal Factor 8	1.02323	.	.	.
Smoothed Seasonal Factor 9	0.99197	.	.	.
Smoothed Seasonal Factor 10	1.01061	.	.	.
Smoothed Seasonal Factor 11	0.98197	.	.	.
Smoothed Seasonal Factor 12	0.97635	.	.	.

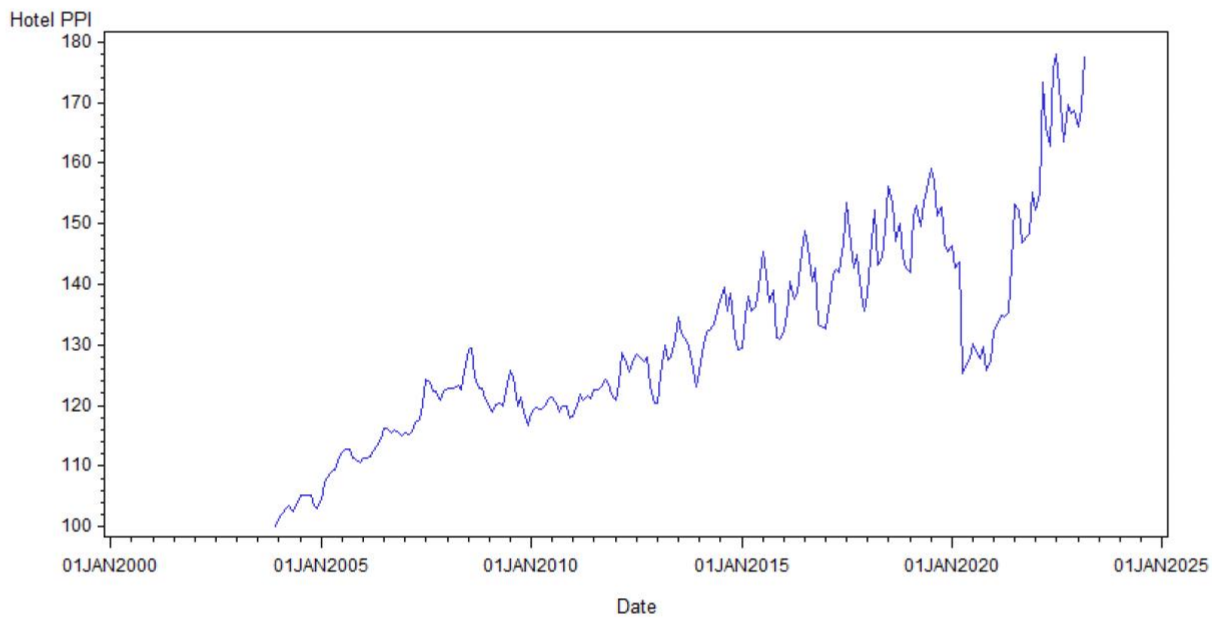
As shown above, the alpha constant for our model is 0.79327 and is significant with a p-value of <.0001. The gamma smoothing constant is 0.00100 which is not significant with a p-value of 0.8717. The estimate for the seasonality in the model is 0.99900 which is also significant with a p-value of <.0001. The smoothed level is 171.94042 and the smoothed trend is 0.25008. July again had the highest seasonality factor at 1.05245 and January was the lowest at 0.96706.



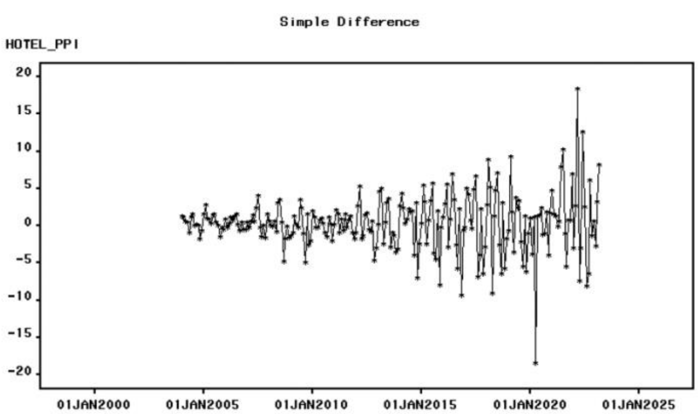
01APR2023	.	168.3370	173.5289	163.1452	.	.	172.1905	0.2501	0.9776
01MAY2023	.	169.7374	176.3847	163.0902	.	.	172.4406	0.2501	0.9843
01JUN2023	.	178.0346	186.0682	170.0010	.	.	172.6907	0.2501	1.0309

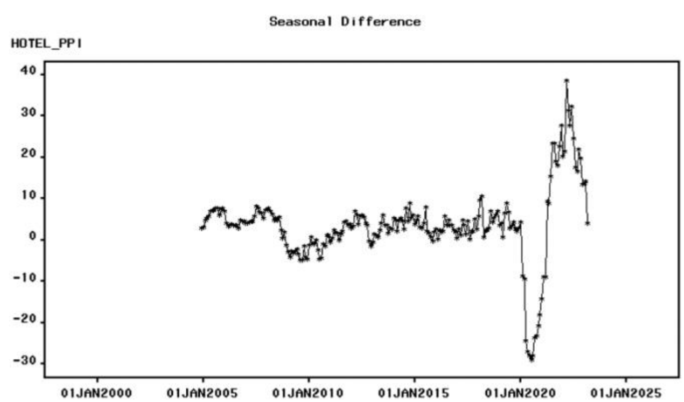
The above graph shows the forecast for the next year. It appeared to me to do a good job at sticking with the new trend showed after Covid and continued off the previous years' trends of being lower in winter months and higher in the summer months. Also shown above is the predictions and prediction intervals for the next three months after March 2023. The predicted value for April 2023 is 168.3370 with a 95% prediction interval between 163.1452 and 173.5289. The predicted value for May 2023 is 169.7374 with a 95% prediction interval between 163.0902 and 176.3847. The predicted value for June 2023 is 178.0346 with a 95% prediction interval between 170.0010 and 186.0682.

Box and Jenkins Model

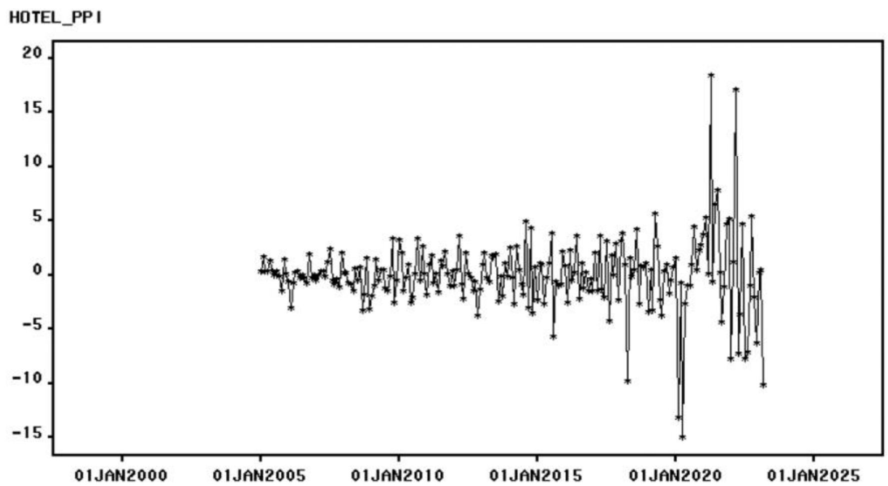


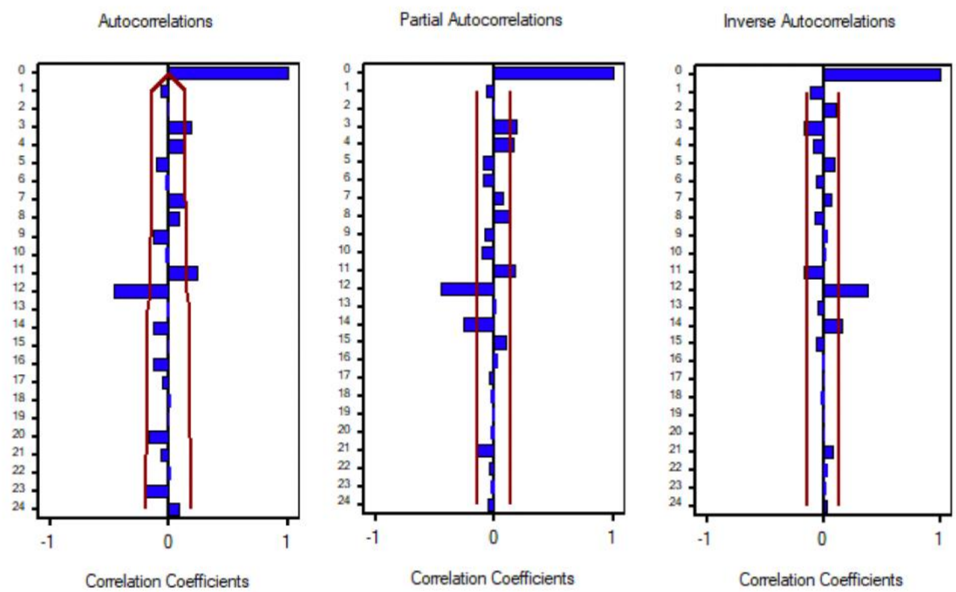
Again, above is the graph of our hotel dataset from January 2003 to March 2023. There appears to be an overall linear and in the middle of the dataset an increasing seasonal trend. Before we can proceed with the ARIMA model we needed to achieve stationarity in our data.





The above graph on the left is the simple difference transformation of our data. This did not appear to be sufficient as there was still what looked to be a violation of the constant variation. The above graph on the right is the seasonal difference transformation of our data. This also did not appear to be sufficient as there was still problems with the violation of constant variation.

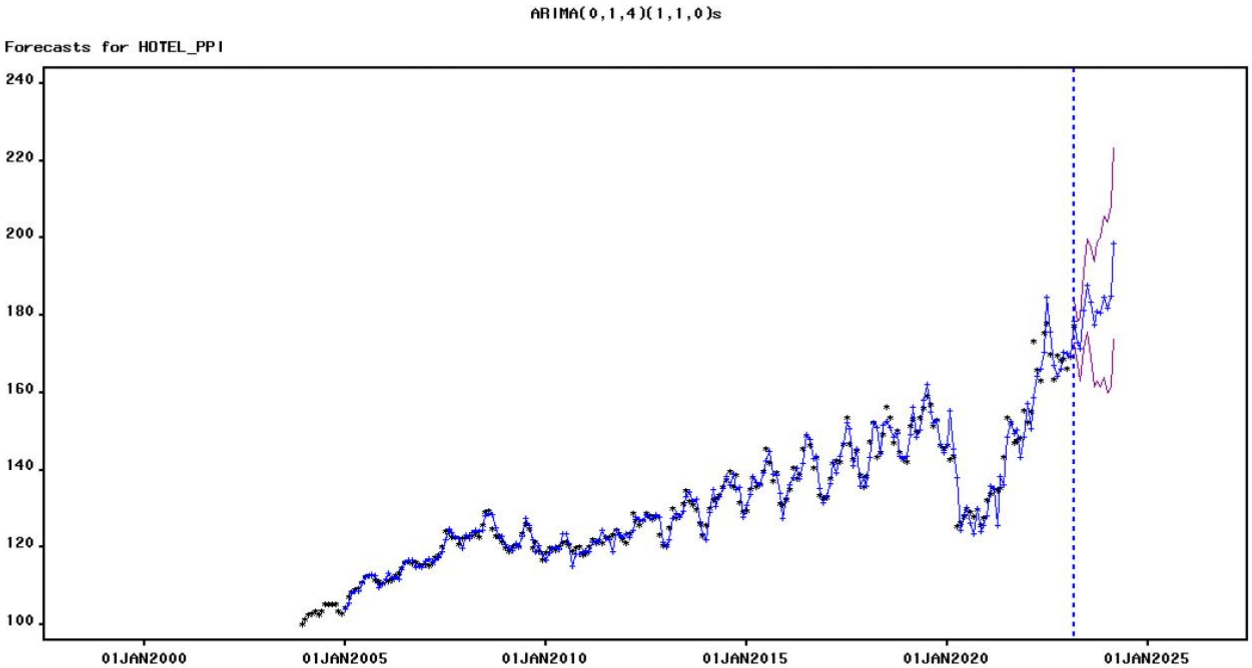




Above is the simple and seasonal difference transformation graph of our data. This did appear to achieve stationarity. A few spike points above and below zero are observed but overall it appears mostly stationary. Above is the SAC and SPAC for our data. Because we were looking for a seasonal ARIMA model here the time points 12 and 24 should be looked at for the lags. At time points 12 for both the SAC and SPAC we notice a spike that is significant and outside of the redlines. This confirms that we are seeing a seasonal trend. It appears to me that the SAC is dying down while the SPAC is cutting off. Both are pretty similar at lag 1 but at lag 2 (time point 24) SPAC seems to be a lot closer to that middle line where as the SAC is nearly half way to the redline. This means we should do an autoregressive seasonal ARIMA model of order 1. For the non-seasonal level the SAC appears to cut off more dramatically than the SPAC here at order 4 indicating a moving average model of order 4.

Model Parameter	Estimate	Std. Error	T	Prob>¦T¦
Intercept	0.04796	0.1724	0.2782	0.7811
Moving Average, Lag 1	-0.07371	0.0688	-1.0712	0.2853
Moving Average, Lag 2	0.05519	0.0661	0.8355	0.4044
Moving Average, Lag 3	-0.19186	0.0665	-2.8842	0.0043
Moving Average, Lag 4	-0.16219	0.0686	-2.3642	0.0190
Seasonal Autoregressive, Lag 12	-0.53757	0.0642	-8.3788	<.0001
Model Variance (sigma squared)	8.06658	.	.	.

Above shows the parameter estimates for our model. You can see only the seasonal autoregressive, lag 12 parameter is significant. These are type III p-values so that parameter is the only term that adds to the model above and beyond all the other parameters.



01APR2023	.	172.8821	178.4487	167.3155
01MAY2023	.	171.2919	179.4597	163.1242
01JUN2023	.	181.2277	191.1704	171.2850

The above graph shows the forecast for the next year. Also shown above is the predictions and prediction intervals for the next three months after March 2023. The predicted value for April 2023 is 172.8821 with a 95% prediction interval between 167.3155 and 178.4487. The predicted value for May 2023 is 171.2919 with a 95% prediction interval between 163.1242 and 179.4597. The predicted value for June 2023 is 181.2277 with a 95% prediction interval between 171.2850 and 191.1704.

Comparison

The first analysis done on this hotel PPI dataset was seasonal decomposition. This analysis seemed to be the weakest of the three analyses done. The dataset had a seasonal component to it in the middle years as discussed previously which seasonal decomposition was able to smooth out however it also had a decent amount of non-seasonal data with the crash from covid and a lesser seasonal effect in the early part of the dataset. Because of this the seasonal decomposition model predicted much lower hotel PPI values than the other two analyses which I believe may be under forecasting future hotel PPI.

The second analysis done on this hotel PPI dataset was exponential smoothing. This analysis appeared to be the strongest of the three analyses. Exponential smoothing was able to forecast and predict values for future hotel PPI that looked accurate and when looking at the smoothing model both the level and seasonal components had p-values that were significant. One downside to this model was that the trend or gamma constant was insignificant however this data as previously mentioned, did have a large dip due to covid and this was likely the cause of this trend being insignificant.

The third analysis done on this hotel PPI dataset was the ARIMA Box and Jenkins model. This model gave a forecast that looked accurate based off the model provided previously but was much higher in predicted hotel PPI than the other two models. As mentioned previously the dataset needs to have stationarity to be able to run this type of model. Stationarity was difficult to achieve with some terms still having larger variances than others. This was caused mostly by the huge dip and recover from the covid pandemic and I believe this led the model to have mostly insignificant parameters and thus I believe this model was not as useful as the exponential smoothing model.

Forecast Model	Model Title	Schwarz Bayesian Information Criterion
☑	Winters Method -- Multiplicative	465.33391
☐	ARIMA(0,1,4)(1,1,0)s	483.78765

Forecast Model	Model Title	Akaike Information Criterion
☑	Winters Method -- Multiplicative	454.99370
☐	ARIMA(0,1,4)(1,1,0)s	463.45322

Overall all three of these analyses had pros and cons when it came to model this dataset. Many of the problems with these models came from the huge dip and recovery from covid. For future work on this dataset it may be best to try and omit these dates out of the dataset. As mentioned before the seasonal decomposition model severely underestimated the forecasted hotel PPI and was the worst of the three models. The next best was the ARIMA Box and Jenkins model. This was confirmed above as both the AIC and the BIC for the exponential smoothing model was lower than the ARIMA model. The exponential smoothing model should be used to forecast and predict future Hotel PPI.