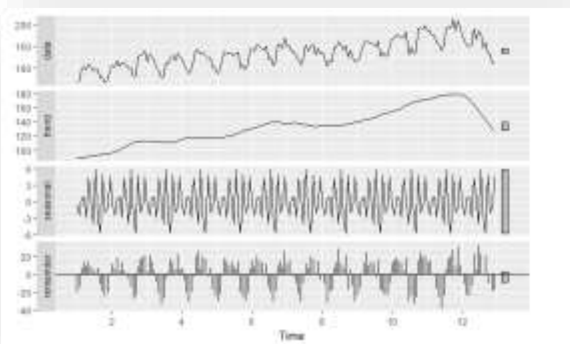
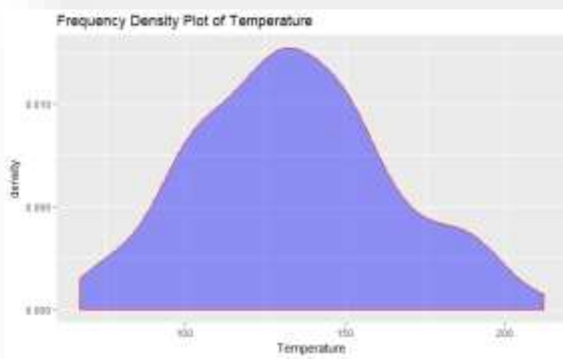
US Temperature Index Time Series Analysis – Jai Katariya

The purpose of this report is to analyze the US Temperature index time series. Perform exploratory analysis to explore the time series components. Perform transformations if there is skewness in the time series distribution. Perform differencing, double differencing, differencing on log and seasonal differencing to see if all the time series components are captured. Build SARIMA models, select the best model and use that model to forecast.

Initial Analysis

As we explore the data we see that there is a gradual upward trend and the trend decreases sharply towards the end. There seems to be high amount of seasonality in the time series. From the density plot we can see that there is slight right skewness, so we log transform our raw data. Also from the ACF and PACF we can see that there is strong correlation between the lags.





After transforming the data, we now difference the log(temp) to make it stationary. We try double differencing on the raw data as well as the log(temp). We also try seasonal differencing on the raw data as well as log(temp) data. Finally based on the ACF and PACF we select seasonally differenced raw data and seasonally differenced log(temp). When we compare both we see no much difference between them so we go ahead seasonally differenced log(temp) with period 12.

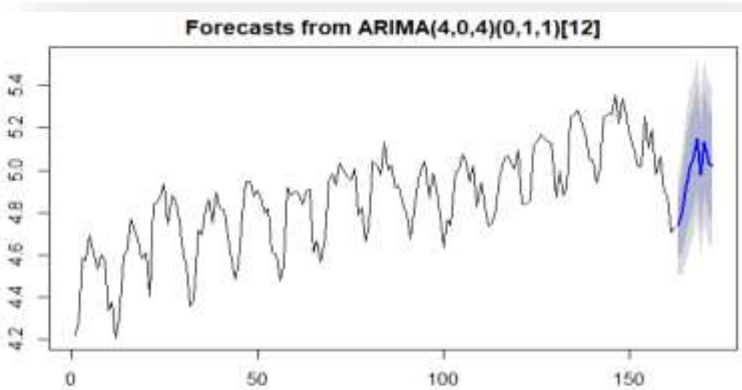
Model Building, Selection and Forecasting

We build close to 10 SARIMA models and do model selection based on good AIC score, Box-test, tsdiag, significant coefficients, ADF test and ACF of residuals. From the below table of models we can that the our selected model has ARMA order (4,0,4) and seasonal order (0,1,1)

```
arima(logtemp,order=c(4,0,4),seasonal=list(order=c(0,1,1),period=12))
```

Model	AIC	Box Test(p-value)	tsdiag	Significant Co-efficient > 2	ADF Test(p value)	ACF of residuals
ARMA (4,0,0) Seasonal (0,1,2)	-461.37	0.372	Looks Good	Yes	<0.01	Lag crossing blue line
ARMA (4,0,6) Seasonal (0,1,1)	-476.67	0.7492	Looks Good	NaN produced	<0.01	Looks Good
ARMA (4,0,4) Seasonal (0,1,1)	-475.28	0.3621	Looks Good	Yes	<0.01	Lag crossing blue line
ARMA (4,0,0) Seasonal (0,1,1)	-454.1	0.4326	Looks Good	Yes	<0.01	Lag crossing blue line
ARMA (4,0,6) Seasonal (0,1,2)	-474.49	0.5663	Looks Good	NaN produced	<0.01	Looks Good

We then use this model to forecast. Below is how the forecast looks like. To perform additional testing we split our data into train and test. Fit our selected model to the training set and predict the results. Upon prediction we compare our accuracy between the predicted values and the actual values which is the testing set. We get an accuracy close to 77% which is not bad. We feel that there is room for improvement and better models can be built using some advanced modelling techniques.




```
accuracy(pred$pred, test)

                  ME      RMSE       MAE       MPE     MAPE
Test set -0.05807026 0.2434127 0.2126423 -1.374867 4.338111
```