**Assignment 4:**

**Name: Vishal Pawar**

**RUID: 227002379**

**Understand and Explain your model output**

Attributes (Columns): Date, Gold price

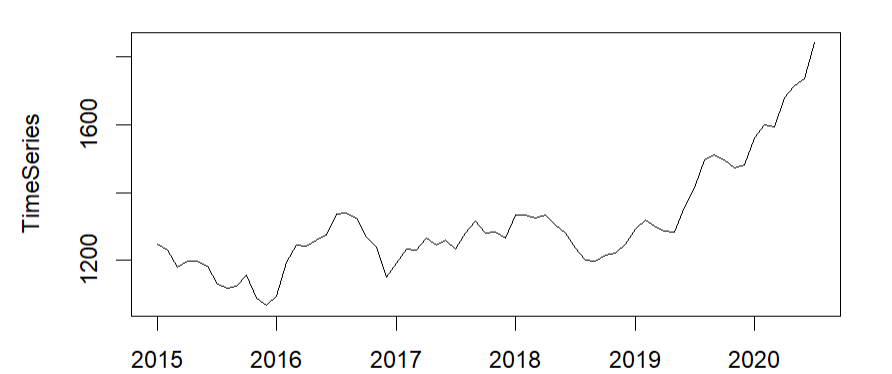
Operations performed:

* Data converted to timeseries using ts() function to map gold price values with time frame in order to compute different forecasting models. Ranging from (jan-2015 to july-2020)

syntax

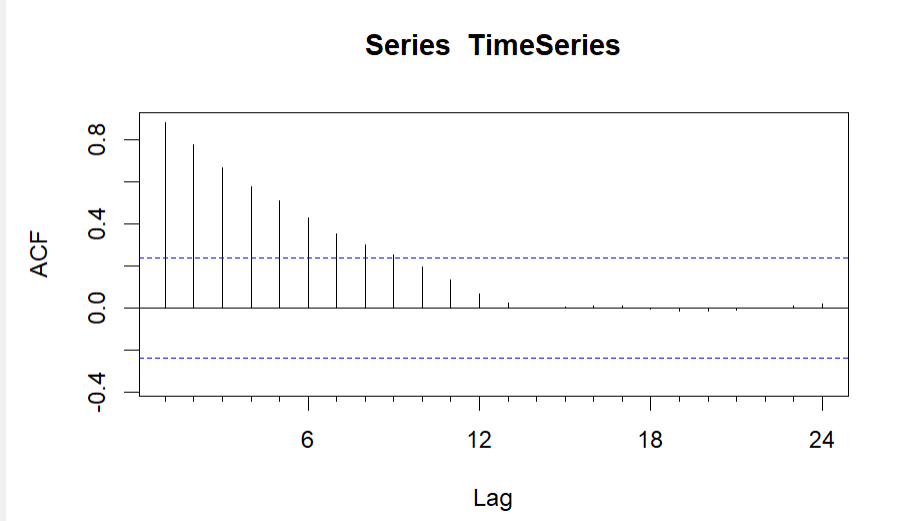
TimeSeries<-ts(DataFrame$gold\_price.Price,start=c(2015,01),end=c(2020,07),frequency=12) #convert to timeseries

* Time series plotted on line chart
* Analyzed based on obervations:
  + - Data is seasonal: As price of gold is observed to always drop in the late months of a year (sept - dec) and there is an increase by start of next year.
    - Also, the price of gold shows increasing curve (trend) as the years pass by



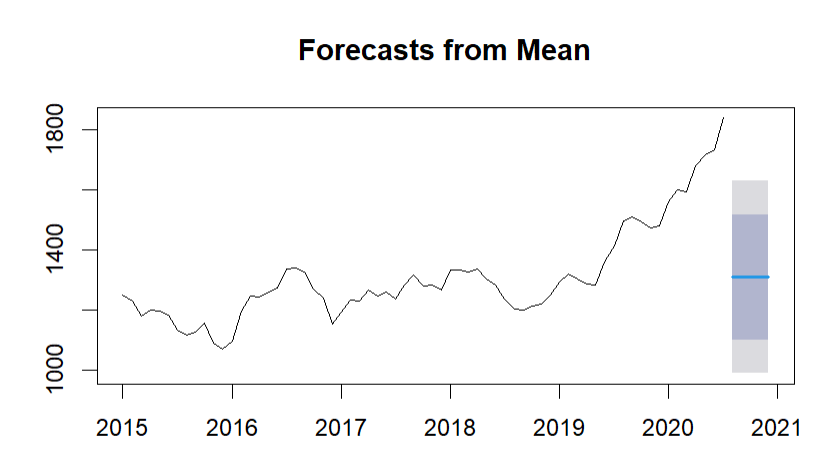
Syntax: plot(TimeSeries)

* **Acf (TimeSeries**): Displays correlation for all the lags in data. Indicating high correlation with the values in the beginning but the correlation decreases as years pass because there is an increasing trend which indicates increase the difference between initial price and price later in time series



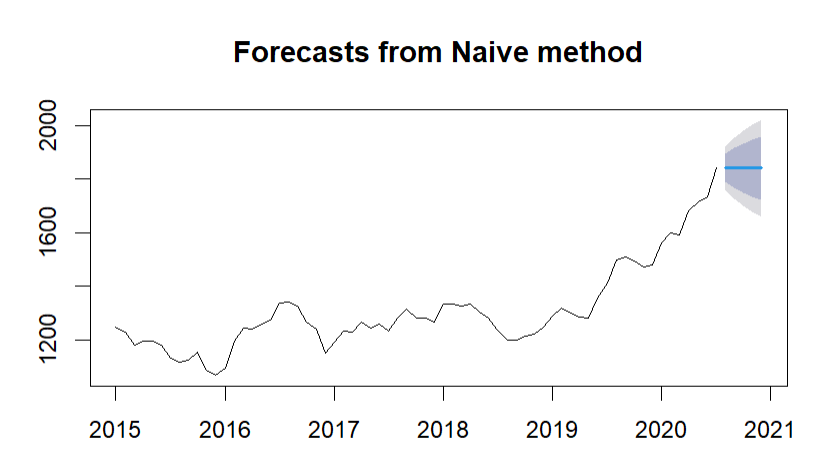
* Forecasting models:

1. **Mean Forecast:** simple forecasting method by assuming that future values will be equal to the historical mean. It's often used as a baseline or benchmark method for comparing with more sophisticated forecasting models.
   * + - Not suitable for the gold price data as it calculates mean of all data points due to which it starts forecasting from the mean of entire time series.

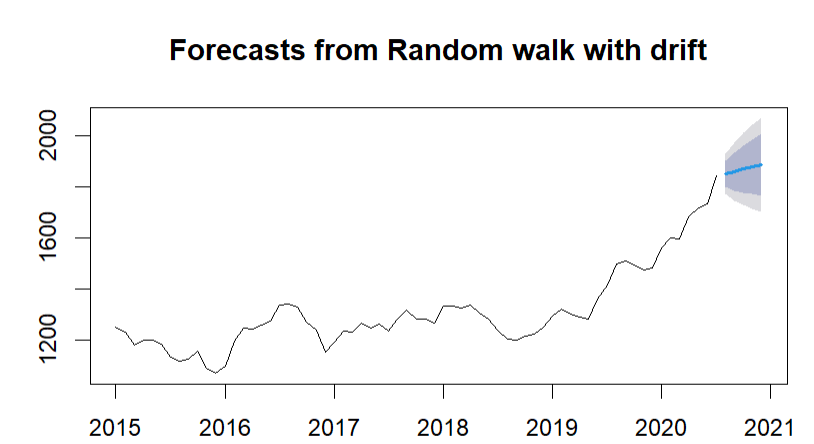


1. **Naïve Forecast:** The naive method is one of the simplest forecasting techniques and involves making predictions by assuming that the future values of a time series will be the same as the most recent observed value.

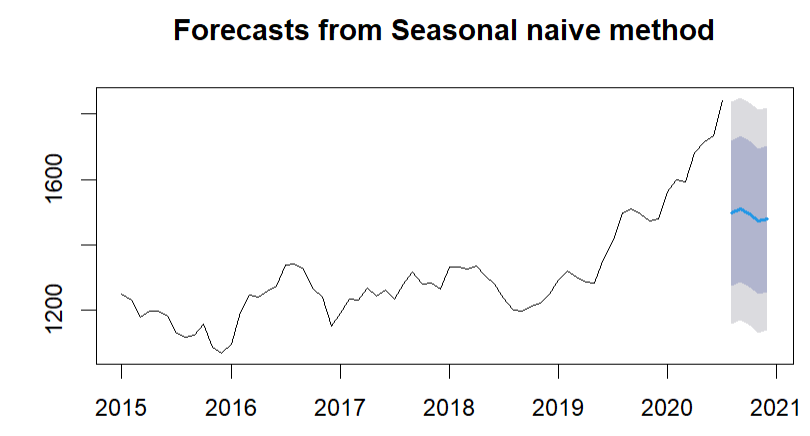
* It yields inaccurate results as it does not take into account seasonality and trend.



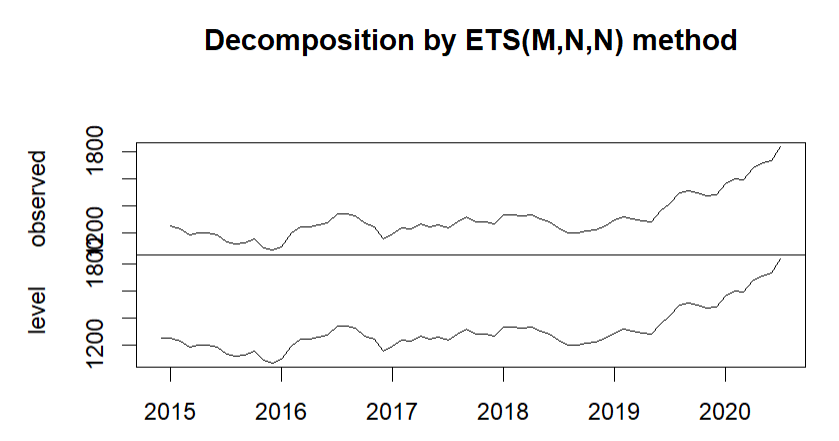
1. **RWF (Random walk with drift):** The function uses the most recent observed value in the time series as the forecast for all future periods. This is similar to the naive method, but it's called a "random walk" because it models the idea that future values are random (based on recent trend) and follows the same distribution as past values.

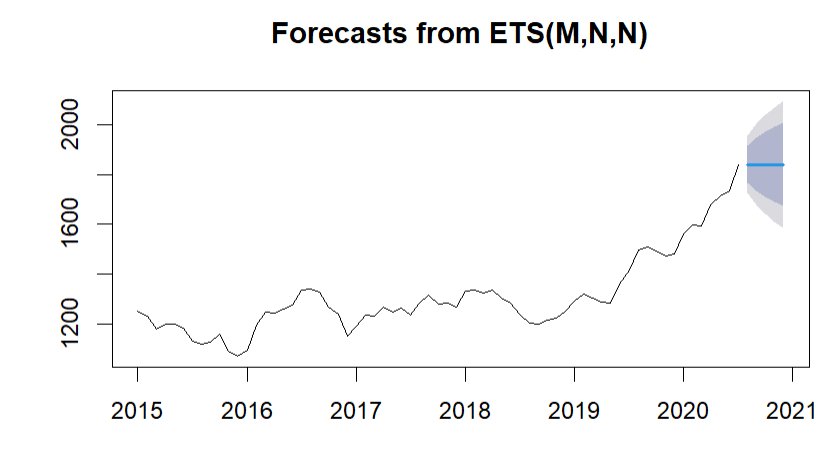


1. **Seasonal Naïve:** Takes into account seasonality but does not consider trend. Hence it yields inaccurate outcome.

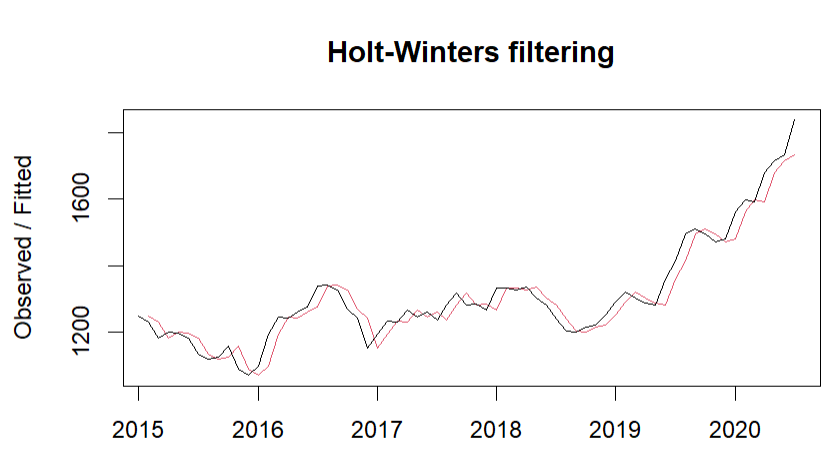


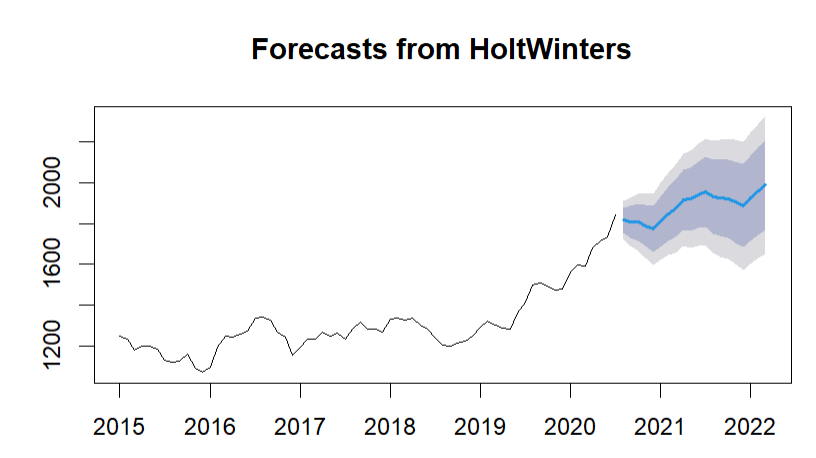
1. **ETS (Error, Trend, Seasonal):** it is a framework to estimate various exponential smoothing models. It automatically selects the best-fitting model based on the characteristics of your data and estimates a forecast model based on data points. It does not calculate trend and seasonality hence it yields inaccurate forecast in the following case.



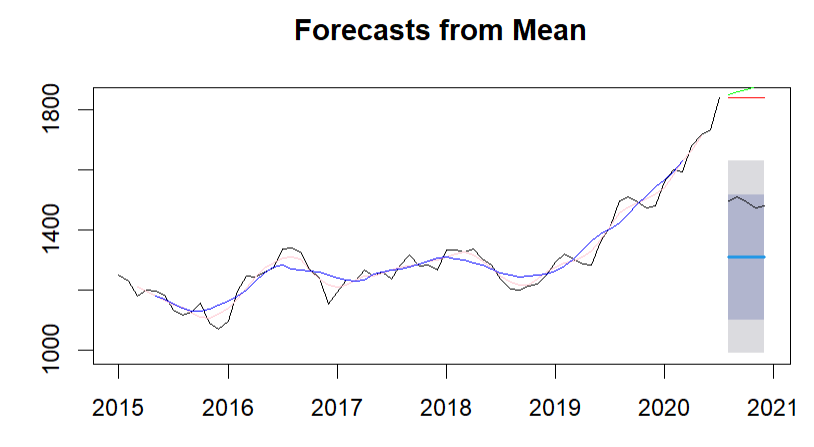


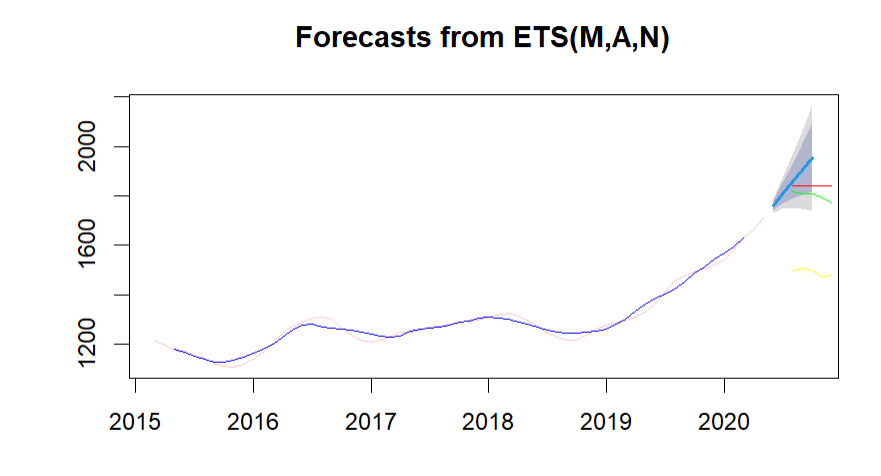
1. **Holt Winters**: It is an exponential smoothing model which takes into account all 3 components. Applies smoothing for both trend (gamma=true) and seasonality (Beta=true). Suitable for following data as it takes into consideration seasonality and trend.





**Plot the time series and different model forecasts in one chart**





**Pick an accuracy measure, compare your models, and state the best model based on the accuracy comparison**

Accuracy Measures:

* Mean Error (ME)
* Root Mean Squared Error (RMSE)
* Mean Absolute Error (MAE)
* Mean Percentage Error (MPE)
* Mean Absolute Percentage Error (MAPE)
* Mean Absolute Scaled Error (MASE)
* Autocorrelation of errors at lag 1 (ACF1)

-> Accuracy Measure Selected for Model Comparison:

* Mean Absolute Percentage Error (MAPE)

The Mean Percentage Error (MPE) measures the accuracy of forecasts with respect to actual observed values. It's a helpful accuracy metric to use because it calculates average percentage error indicating how well the forecast matches with actual values.

MPE= (1/N)\*∑[(Actual - Forecast)/Actual]\*100

N= No. of observations

MPE calculates the average percentage: by which the forecasts either overestimate or underestimate from the actual values.

Positive MPE illustrates that, on average, by what measure the forecasts are overestimating the actual values and a negative MPE indicates that the forecasts are underestimating the actual values.

MPE for different Models:

* Mean Forecast: -1.294273
* Naive Forecast: 0.538691
* Random Walk Forecast: -0.1546792
* Seasonal Naive Forecast 6.218464
* Moving Average order=5 : 0.06272455
* Holt-Winter Forecast: 0.1149414
* SSE Simple Forecast: 0.5387162

After comparing MPE for all the models, **Ḥolt-Winters Forecast** is best for the following case as it takes into account both seasonality and trend for calculation as observed in the data set