**Adwait Kalsekar**

**Business Forecasting Assignment 04**

1. **Explain your Model output**

The dataset chosen for this assignment consists of the House Price Index for the State of New Jersey over the last 50 years calculated quarterly.

The time series of the data set shows seasonality in the House Price Index every 10 years. The HPI grows exponentially over the decade and drops towards the start of a new decade. This seasonality holds true for the entire time series data.

Looking at the time series data it can also be inferred that there has been a visible trend in the growth of HPI over the last 50 years. The HPI has increased exponentially since the year 1975 to the year 2023.

Looking at the time series data, certain forecasting models applied have predicted similar future data points that account for the existing trend and seasonality.

Models used:

* Average Methods
  + Mean Forecast: it produces visually inaccurate results as it calculates the mean of all the data points in the time series data.
  + Moving Averages : this model adds weights to the data points with a higher weight to the recent data points and hence can provide a better prediction.
* Naive Methods
  + Naive Forecast: it accounts only for the most recent data point and forecasts it for the next data points. It can be used for a small data set without any trend or seasonality forecasted for a small time period. In the time series data used in this assignment, the naive method provides inaccurate results.
  + Random walk with a drift: this model considers the most recent data point with a drift term and hence can provide good results for a short time period. In this case, it shows good results as it accounts for the growing trend.
  + Seasonal Naive: this model is perfect for a highly seasonal data without accounting for the trend and gives inaccurate results in this case.
* Exponential smoothing
  + ETS: this method applies exponential smoothing to the time series data and predicts the data points. THe drawback is that it does not account for trend or seasonality.
  + Holt Winters: this model is also known as triple exponential smoothing, as it applies exponential smoothing and accounts for trend as well as seasonality and turns out to be the most accurate model for this time series data.

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

-> Accuracy Measures Calculated:

* 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 absolute percentage error (MAPE) measures the average of forecast errors in percentages. It's a helpful accuracy metric to use because many people can understand forecast accuracy in terms of percentages.

For example, a MAPE of 3% means there was a 3% difference between the actual and projected data. Typically, a lower MAPE indicates a higher forecast accuracy.

MAPE for different Models:

* Mean Forecast: 88.20295
* Naive Forecast: 1.869083
* Random Walk Forecast: 1.731568
* Simple Naive Forecast: 6.396315
* Exponential smoothing Forecast: 1.148609
* Holt-Winter Forecast: 1.218338
* SSE Simple Forecast: 1.86915

-> Looking at the MAPE for all the models the best models to use for forecasting are:

* Exponential Smoothing Forecast
* Holt-Winters Forecast

But, considering that the Exponential Smoothing Forecast does not account for trend or seasonality, it may provide inaccurate forecasts for a larger time period.

Hence, the best model for this time series dataset is the Holt-Winters Forecast.