## Coca cola Sales Price Forecasting

## Work Flow

1. Import Libraries
2. Load The Data
3. EDA
4. Preprocessing
5. Split the data
6. Build The Model
7. Model Selection
8. Conclusion.

### Import Libraries

I had import the useful libraries that is required for project.

### Load the data

I loaded the airline passengers dataset using pandas library.

### EDA

I plotted the actual data and explored the dataset to find number of data points, shape of dataset, duplicates, null values etc. I plotted the different plot to find out is there any seasonality, trend are present or not. Unfortunately I found that there is seasonality present in the dataset and also data have upward trend, after that I checked for Stationary test because time series works well on stationary dataset for this I used Adfuller test to test the stationary nature of data.

### Preprocessing

I tested the data by Adfuller test and found data is non stationary. So, I used differencing method to convert non-stationary data to stationary. I did first differencing then I found slightly trend in the data then I did second differencing then I found now data is stationary and ready to fit the model.

### Split the data

For Building the model I splitted the dataset in to test data and train data. For training the model I used train data and for forecasting I used test data.

### Build the Model

I build various time series models these are given below.

1. Auto Regression

I have first build Auto Regression model with actual data, first differencing data and second differencing data then I got linear curve in forecasting for all data. With mean absolute error 300.61, 31.52 and 359.02 for actual, first differencing and second differencing data respectively & root mean squared error 332.17, 342.23 and 444.5050 for actual, first differencing and second differencing data respectively.I found bell shaped distribution of residuals when trained with second differencing feature.

1. Auto Arima

For finding the best params for ARIMA and SARIMAX I did Auto Arima grid search. I got order for ARIMA (0,1,1) and for SARIMAX for (2,1,1,5).

1. ARIMA

Then I build ARIMA model with suggested parameters by Auto Arima and I found that forecasted result is tend to match the test data. ARIMA model gives better prediction than Auto regression.

With mean absolute error 910.97, 469.49 and 752.16 for actual, first differencing and second differencing data respectively & root mean squared error 1033.93, 609.97 and 836.76 for actual, first differencing and second differencing data respectively. I found residuals are normally distributed when trained model with actual feature.

1. SARIMAX

Then I build SARIMAX model with suggested parameters by Auto Arima. I trained the model with different datasets and I got much similar result as ARIMA model. I got root mean squared error 910.97, 469.4 and 752.16 for actual, first differencing and second differencing feature respectively & mean absolute error 1033.3, 609.97 and 836.76 for actual, first differencing and second differencing feature respectively. I found residuals are normally distributed when trained model with actual feature.

1. Simple Exponential Smoothing

After that I build Simple exponential smoothing model. I build the model with different features separately. I got forecasted result as linear straight line. With mean\_absolute\_error 1043.74, 464.63 and 760.61 for actual, first differencing and second differencing data & root mean squared error 1172.08, 616.406 and 845.70 for actual, first differencing and second differencing feature respectively. I got residuals are normally distributed when trained model with actual and second differencing feature.

1. Double Exponential Smoothing

I build Double Exponential Smoothing with Additive trend for different dataset. I found a linear line with some positive slope it gives better result than Simple Exponential Smoothing. With mean\_absolute\_error 465.60, 488.62 and 734.004 for actual, first differencing and second differencing feature respectively & root mean squared error 574.73, 593.93 and 823.36 actual, first differencing and second differencing feature respectively. I have tried to build with multiplicative trend on actual data and got mean absolute error 405.31 and root mean squared error 477.29. I got residual normally distributed when trained model with second differencing feature.

1. Triple Exponential Smoothing

After that I build Triple Exponential Smoothing model with different features separately. I got better result than Simple and Double Exponential smoothing. With mean absolute error I got 587.24, 221.23 and 416.07 for actual, first differencing and second differencing feature respectively & root mean squared error I got 645.844, 278.88 and427.611for actual, first differencing and second differencing respectively. I found residuals are normally distributed when build with actual feature.

### Model Selection

I trained different time series models and I found Triple Exponential Smoothing gives the minimum error when trained with first differencing and second differencing features best fit curve along with test data. With minimum root mean squared error.

### Conclusion

* When I plotted the actual data then I found data with seasonal and upward trend.
* I trained numbers of time series models to find out which model gives me best result.
* After plotting predicted and test data & checking the model performance with the help of different metrices like mean\_absolute\_error and root\_mean\_squared\_error then I found Triple Exponential Smoothing performed well on actual, first differencing and second differencing features.