## Airline Passengers 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

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

### Load the data

In this step I load the airline passengers dataset using pandas library.

### EDA

In this step I 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 have 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 27.25 and root mean squared error 32.63.

1. Auto Arima

For finding the best params for ARIMA and SARIMAX I did Auto Arima grid search. I got order for ARIMA (2,0,1) 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 304.12, 26.90 and 29.42 for actual, first differencing and second differencing data respectively & root mean squared error 308.085, 32.422, 37.31 for actual , first differencing and second differencing data.

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 52.11, 32.42 and 37.31 for actual, first differencing and second differencing data & mean absolute error 41.099, 26.904 and 29.422 respectively.

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 68.644, 234.76 and 237.5 for actual, first differencing and second differencing data & root mean squared error 71.5, 237.008 and 240.54 respectively.

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 57.32, 27.23 and 27.23 for actual, first differencing and second differencing data & root mean squared error 47.56, 32.64 and 37.54 respectively. Also I have tried to build with multiplicative trend on actual data and got mean absolute error 36.88 and root mean squared error 47.56.

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 304.12, 32.422 and 29.422 for actual, first differencing and second differencing feature respectively & root mean squared error I got 307.66, 19.834 and 16.926 for actual, first differencing and second differencing respectively.

### Model Selection

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

### Conclusion

* Data I found to be 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 I found Triple Exponential Smoothing performed well on actual, first differencing and second differencing features.