# Forecasting on Airlines dataset:

* Goal: To Build the most accurate forecasting model for Airlines dataset.
* Data collection: Airlines.xlsx
* EDA: We take the airlines dataset and try to bring insights from the line plot.

1. There is an overall upward linear trend
2. There is multiplicative seasonality
3. Seasonality is 12

* Pre- Processing:

1. I will add T(trend component) to capture linear trend
2. I will add T^2(trend component) to capture quadratic trend (if any)
3. I will Log(y) to capture multiplicative seasonality and exponential trend
4. I will 11 Dummy variables, to capture seasonality components

* Partition Series: I will partition the pre-processed dataset into training and test. The test data will contain the recent observation(Seasonality). Here it will be 12 observation for test and 84 for training. (Total 96 datapoints)
* Applying Forecasting: Now I will apply different forecasting methods and look at the RMSE value:

1. Linear Trend: 53.199
2. Exponential Trend: 46.05
3. Quadratic trend: 48.05
4. Additional Seasonality: 132.82
5. Multiplicative seasonality: 140.06
6. Additive seasonality with linear trend: 35.35
7. Additive seasonality with quadratic trend: 26.36
8. Multiplicative seasonality with linear trend: 10.519
9. Multiplicative seasonality with quadratic trend: 18.37

I can see that Multiplicative Seasonality with linear trend has the least RMSE value. I also visualized the same during EDA.

So, I will make a Multiplicative seasonality with linear trend model using all the data-points and forecast it on a new data sheet.

Now, After forecasting, I will look into the errors and run an auto-correlation model to generate ACF plot. The ACF plot I obtained is:

I can find there are still some information left, So I will make an ARIMA model using the error value.

The coefficient obtained is:

Screen Clipping

E(t+1) = -3.223\*10^-16 + 0.671 E(t)

I will take the residuals obtained from the ARIMA model and again look at the ACF plot:

The obtained ACF plot is:

I cannot find any more significant information is left in the errors.

So our Final Forecasted model is:

***log(passengers) = (4.666292 + 0.011227 T + 0.015443 D1+ 0.014741 D2 +0.146693 D3 +0.102613 D4+ 0.084863 D5+0.197479 D6+ 0.292506 D7+0.279162 D8+0.157696 D9+0.016358 D10+-0.12532 D11) + ( -3.223\*10^-16 + 0.671 E(t))***