# Forecasting on Sales on Coca-Cola dataset:

* Goal: To Build the most accurate forecasting model for Coca-Cola dataset.
* Data collection: CocaCola\_Sales\_Rawdata.xlsx
* EDA: We take the Coca-Cola dataset and try to bring insights from the line plot.

1. There is an overall upward quadratic trend
2. Seasonality is additive
3. Seasonality is 4

* 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 3 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 4 observations for test and 38 for training. (Total 42 data-points)
* Applying Forecasting: Now I will apply different forecasting methods and look at the RMSE value:

1. Linear Trend: 591.55
2. Exponential Trend: 466.248
3. Quadratic trend: 475.56
4. Additional Seasonality: 1860.024
5. Multiplicative seasonality: 1963.39
6. Additive seasonality with linear trend: 464.98
7. Additive seasonality with quadratic trend: 301.738
8. Additive seasonality with exponential trend: 225.524
9. Multiplicative seasonality with quadratic trend: 581.85

I can see Additive seasonality with exponential trend has the least RMSE value.

So, I will make a Additive seasonality with exponential 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 no information left.

So our Final Forecasted model is:

|  |
| --- |
| 7.394585 |
| 0.024167 |
| -0.07953 |
| 0.136828 |
| 0.104241 |

***Log(sales) = 7.394585+ 0.024167 T - 0.07953 D1+ 0.136828 D2 + 0.104241 D2***