TIME SERIES

Course Objectives

What is time series

What is time series analysis and forecasting

How to validate a time series forecasting model

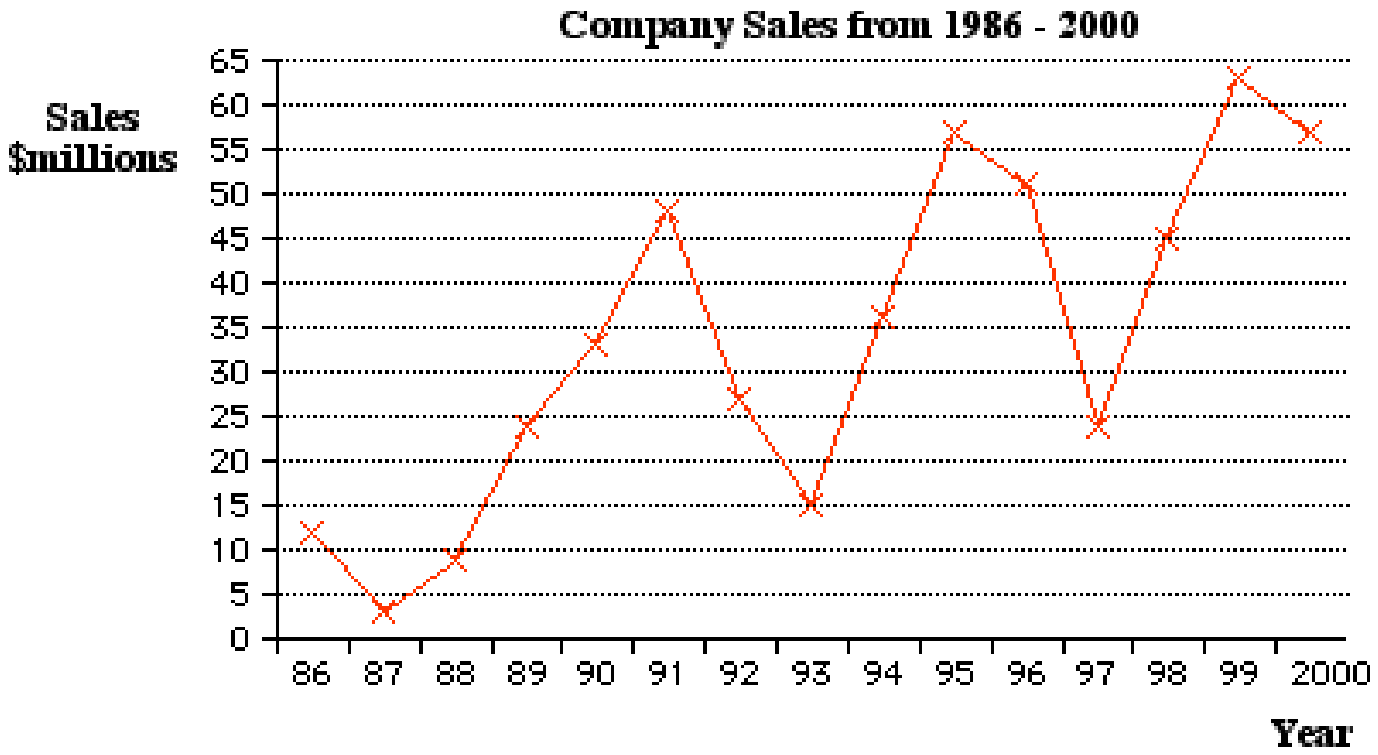
Learning Outcomes

At the end of the course, you will be able to

- Understand what is time series
- How to perform time series analysis
- How to perform time series forecasting
- Understand the technique to validate a time series forecast

Time Series

- A sequence of observations over a certain period



Usually the observations are taken at regular intervals
(e.g.: hours, days, weeks, months, years)

Time Series

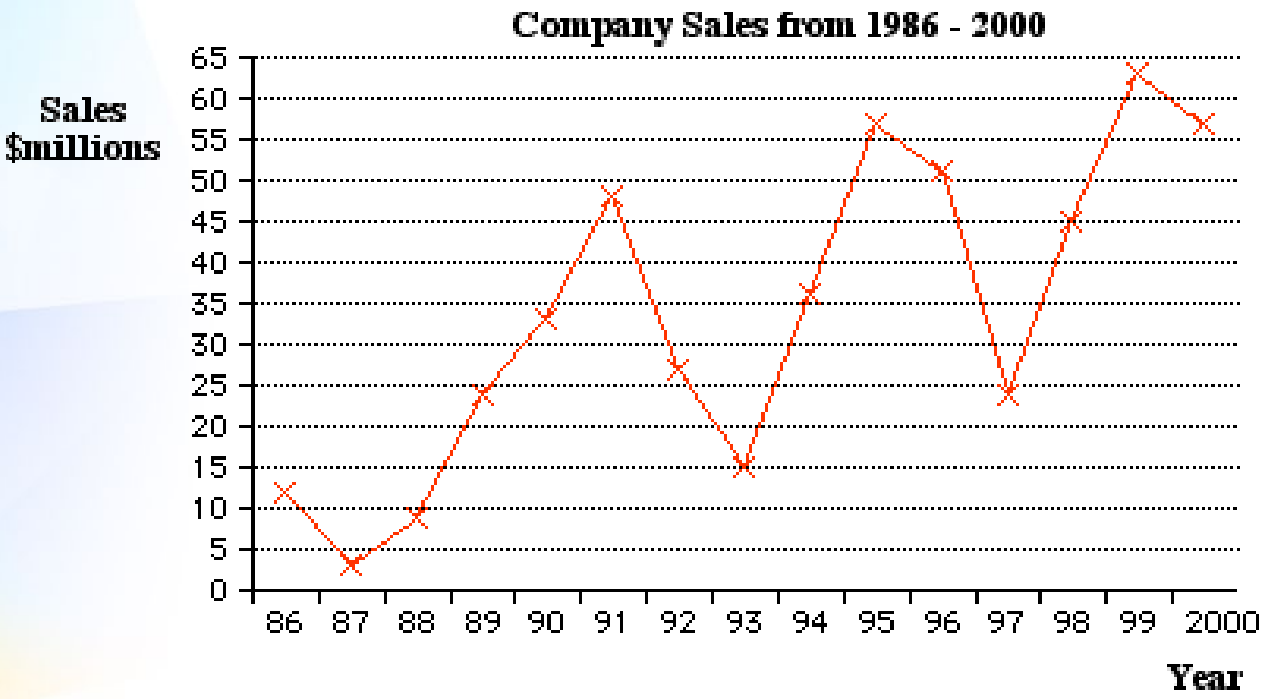
- Work with time based data
  - Years, months, days, hours
- A **time series** is a sequence of data indexed by time.

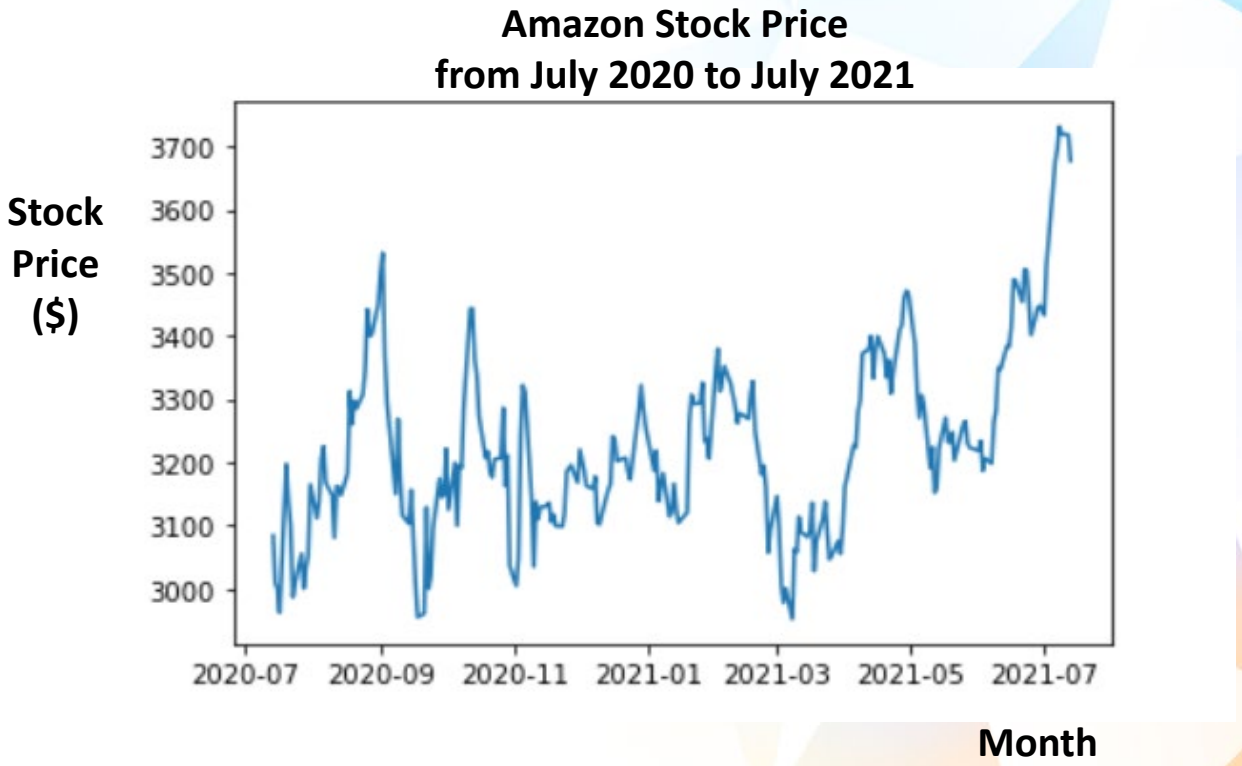
Date	Open	High	Low	Close	Adj Close	Volume
7/14/2020	3089	3127.38	2950	3084	3084	7231900
7/15/2020	3080.23	3098.35	2973.18	3008.87	3008.87	5788900
7/16/2020	2971.06	3032	2918.23	2999.9	2999.9	6394200
7/17/2020	3009	3024	2948.45	2961.97	2961.97	4761300
7/20/2020	3000.2	3201.36	2994.02	3196.84	3196.84	7598200
7/21/2020	3232.49	3240.58	3105.72	3138.29	3138.29	6121300
7/22/2020	3125	3150	3065.26	3099.91	3099.91	4104200
7/23/2020	3098.27	3098.27	2970	2986.55	2986.55	5656900
7/24/2020	2930	3031.58	2888	3008.91	3008.91	5632400
7/27/2020	3062	3098	3015.77	3055.21	3055.21	4170500
7/28/2020	3054.27	3077.09	2995.76	3000.33	3000.33	3126700
7/29/2020	3030.99	3039.16	2996.77	3033.53	3033.53	2974100
7/30/2020	3014	3092	3005	3051.88	3051.88	6128300
7/31/2020	3244	3246.82	3151	3164.68	3164.68	8085500

index

Time Series Visualization

- Usually plotted as line graphs against time





Purpose- Time Series Modeling

Time Series Analysis

Predict future value of an item

Date	Open	High	Low	Close	Adj Close	Volume
7/14/2020	3089	3127.38	2950	3084	3084	7231900
7/15/2020	3080.23	3098.35	2973.18	3008.87	3008.87	5788900
7/16/2020	2971.06	3032	2918.23	2999.9	2999.9	6394200
7/17/2020	3009	3024	2948.45	2961.97	2961.97	4761300
7/20/2020	3000.2	3201.36	2994.02	3196.84	3196.84	7598200
7/21/2020	3232.49	3240.58	3105.72	3138.29	3138.29	6121300
7/22/2020	3125	3150	3065.26	3099.91	3099.91	4104200
7/23/2020	3098.27	3098.27	2970	2986.55	2986.55	5656900
7/24/2020	2930	3031.58	2888	3008.91	3008.91	5632400
7/27/2020	3062	3098	3015.77	3055.21	3055.21	4170500
7/28/2020	3054.27	3077.09	2995.76	3000.33	3000.33	3126700
7/29/2020	3030.99	3039.16	2996.77	3033.53	3033.53	2974100
7/30/2020	3014	3092	3005	3051.88	3051.88	6128300
7/31/2020	3244	3246.82	3151	3164.68	3164.68	8085500

index

No other independent variables

Time Series Analysis

- Analyse historical data

- Example, average value

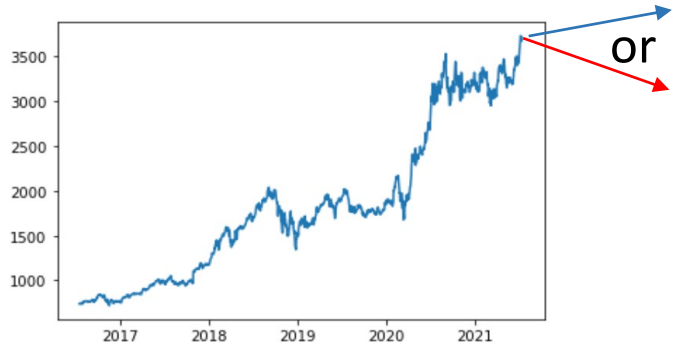
$$\hat{Y} = \frac{Y_t + Y_{t-1} + .. + Y_{t-m+1}}{m}$$

m ~ most recently observation

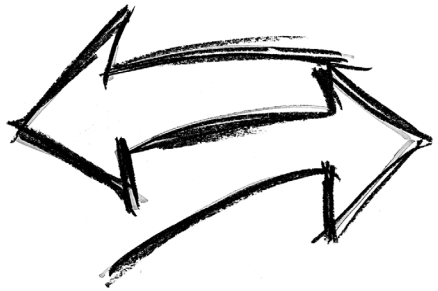
Month	Expenditure
1	2000
2	2500
3	2200
4	2600
5	2400
6	2500

The average expenditure for the last 6 months is **$2366.67**

Time Series Forecasting



to predict future data movement



which technique is the best?

many, many different techniques

try out and select the one that seems to work best

Time Series Forecasting – Moving Average

- Moving Average
  - The next observation is the mean of a selected set of past observations which changes across time

$$\hat{Y}_t = \frac{Y_{t-1} + Y_{t-2} + .. + Y_{t-w}}{w}$$

$\hat{Y}_t$ = moving average for period t
Y_{t-1} = actual value for period t−1
w = window or interval

Moving Average Example

Period	Demand	3 Period Moving Average
1	7	-
2	14	-
3	11	-
4	19	10.6667
5	9	14.6667
6	8	13
7	12	12
8	11	9.6667
9	7	10.3333
10	10	10
11	10	9.3333
12		9

3 period moving average

$$\hat{Y}_t = \frac{Y_{t-1} + Y_{t-2} + Y_{t-3}}{3}$$

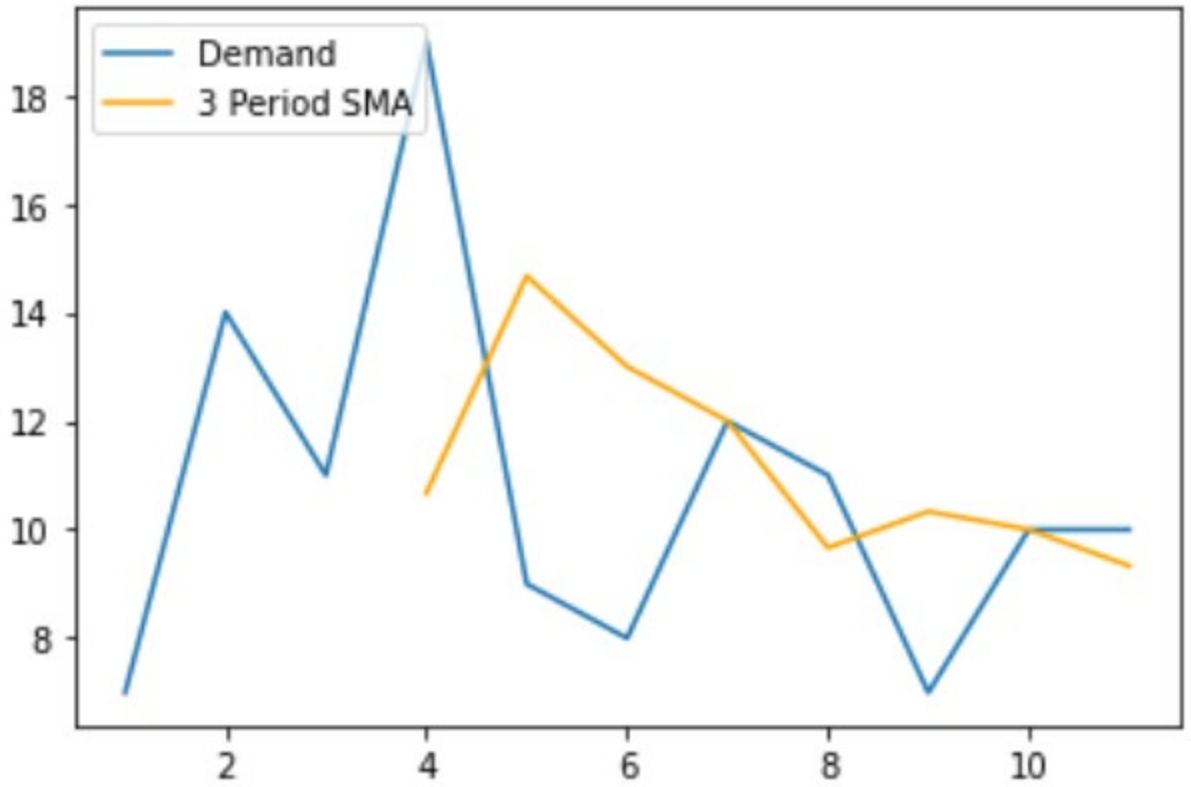
$$\frac{7 + 14 + 11}{3}$$

$$\hat{Y}_4 = \frac{Y_3 + Y_2 + Y_1}{3}$$

$$\frac{14 + 11 + 19}{3}$$

$$\frac{7 + 10 + 10}{3}$$

Data Visualization – Moving Average



Demand of Product over time

Moving Average Consideration

- Simple to use
- Good for identifying trends

Disadvantages

- Didn’t factor in past trends or seasonality
- Assume all data points used are of the same weightage

Weighted Moving Average

- Assign a **heavier weighting** to more current data points since they are more relevant than data points in the distant past.

$$\hat{Y}_t = \frac{3}{(1+2+3)} * Y_{t-1} + \frac{2}{(1+2+3)} * Y_{t-2} + \frac{1}{(1+2+3)} * Y_{t-3}$$

E.g. Time Period = 3

Assign the weight 3, 2, 1 to the recent period, with the most recent period receives a bigger weightage

Weighted Moving Average

Period	Demand	3 Period Weighted Moving Average
1	7	-
2	14	-
3	11	-
4	19	11.3333
5	9	15.5000
6	8	12.6667
7	12	10.1667
8	11	10.1667
9	7	10.8333
10	10	9.1667
11	10	9.1667
12		9.5000

Time Period = 3

$$\hat{Y}_t = \frac{3}{(1+2+3)} * Y_{t-1} + \frac{2}{(1+2+3)} * Y_{t-2} + \frac{1}{(1+2+3)} * Y_{t-3}$$

$$= (\frac{3}{6} * 8) + (\frac{2}{6} * 9) + (\frac{1}{6} * 19)$$

$$\hat{Y}_7 = \frac{3}{(1+2+3)} * Y_6 + \frac{2}{(1+2+3)} * Y_5 + \frac{1}{(1+2+3)} * Y_4$$

Limitations of Forecasting

- May do a good job in predicting potential “trends”

- But there are uncertainties especially in predicting immediate data movements

Validating The Model

How accurate is the forecast using this model?

Should I use 200 days or 50 days period for moving average

Measuring Forecast Accuracy- MSE or RMSE

Mean Squared Error

$$MSE = \sum_{t=1}^{n} e_t^2 / n$$

Let $e_t = Y_t - \hat{Y}_t$ be the errors of forecast.

(the difference between the actual and predicted value)

Root Mean Squared Error

$$RMSE = \sqrt{MSE}$$

Measuring Forecast Accuracy

Period	Actual Value	Predicted Value (3 Period Moving Average)	e	e^2
1	7	-		
2	14	-		
3	11	-		
4	19	10.6667	-8.3333	69.4439
5	9	14.6667	5.6667	32.1115
6	8	13	5	25.0000
7	12	12	0	0.0000
8	11	9.6667	-1.3333	1.7777
9	7	10.3333	3.3333	11.1109
10	10	10	0	0.0000
11	10	9.3333	-0.6667	0.4445
12		9		
			Total	**139.8884**

$$e_t = Y_t - \hat{Y}_t$$

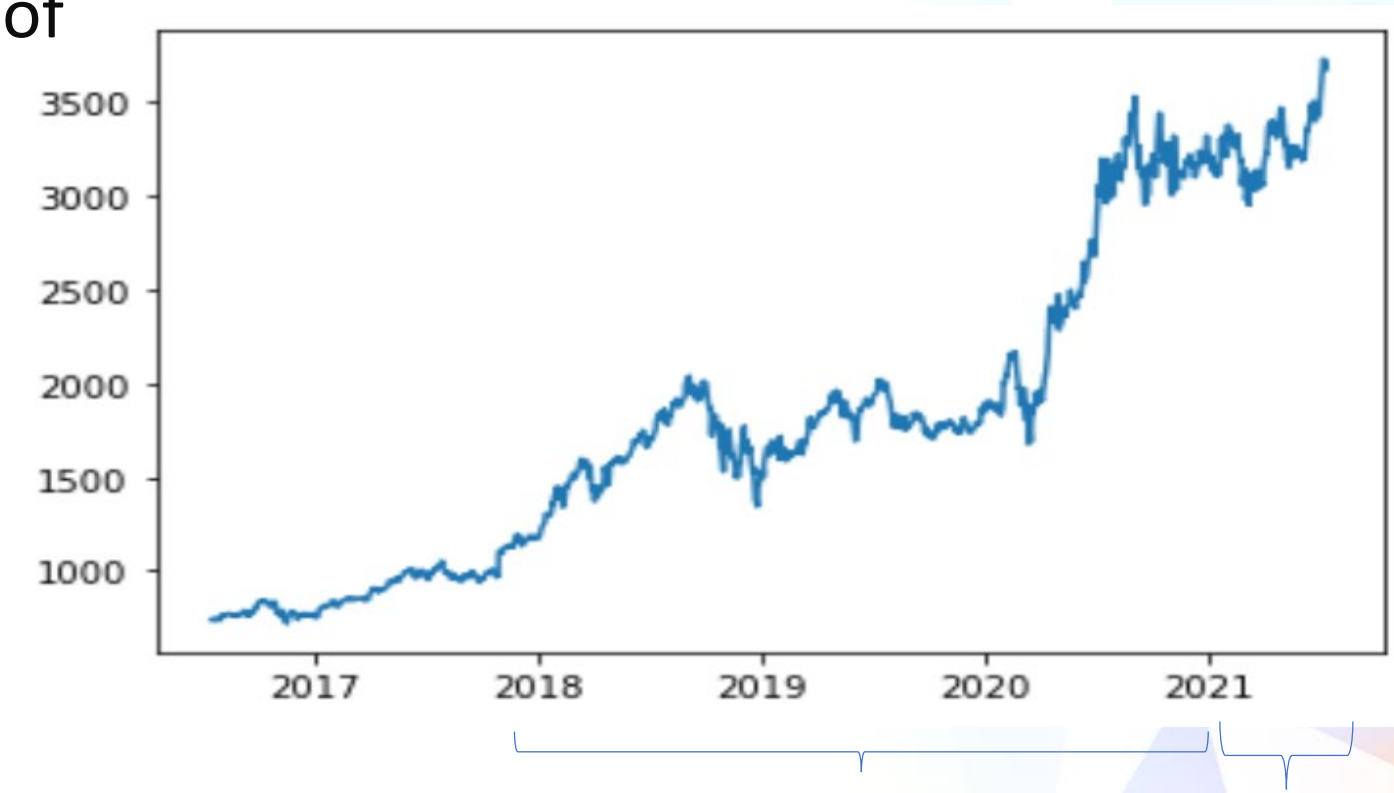
(difference between the actual and predicted values)

The **mean squared error** for the period between 4 and 11 is **139.8884 /8 = 17.4861**

The **root mean squared error** is $\sqrt{17.4861}$ = **4.1816**

Selection of Data

- Select most recent period of available data for training



Building the Model

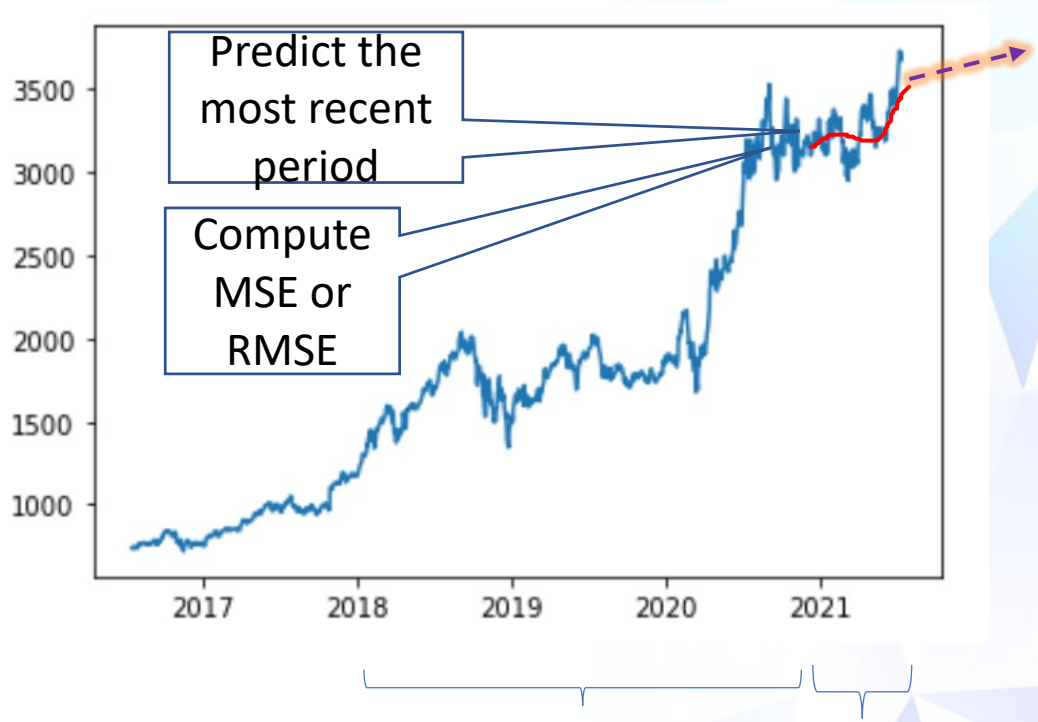
Extract historical data; determine the period of data to be used

Train Time Series Model

Predict Most Recent Period (which data is available)

Compare the prediction with the actual outcome

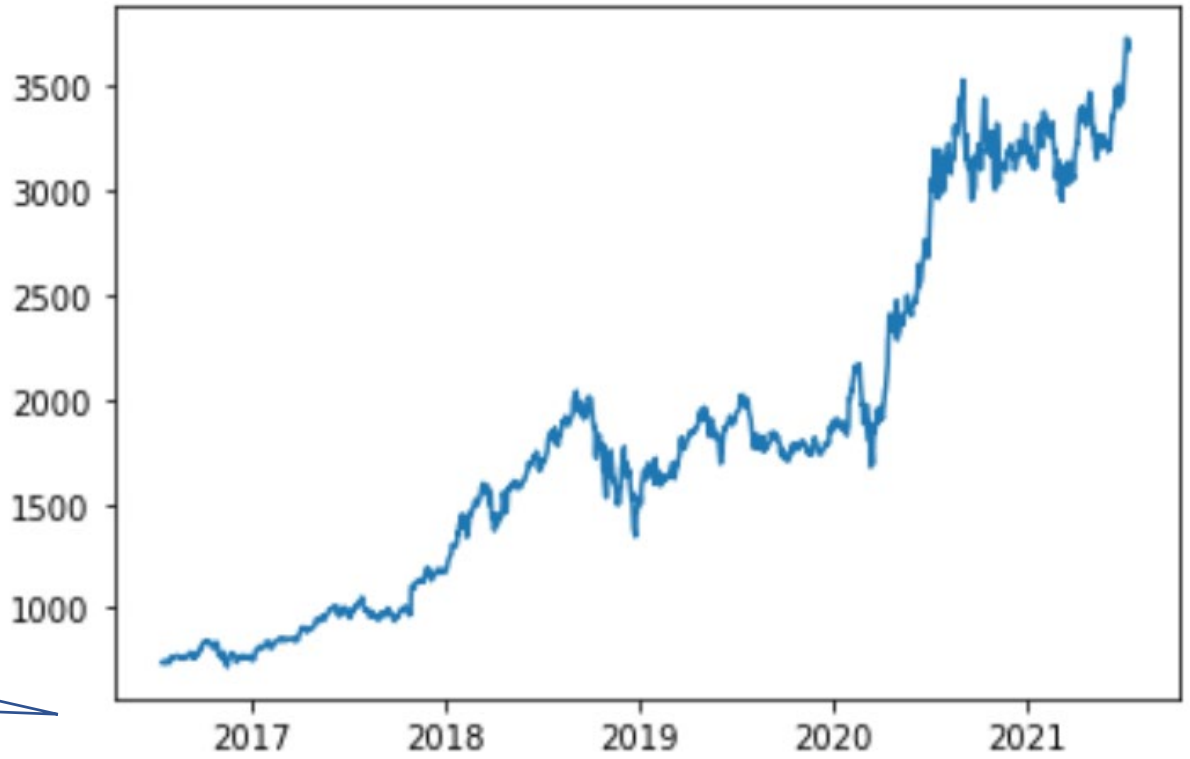
Predict the future!



When to Retrain?

- Predictions over time becomes less and less accurate
- Retrain the model when the actual data is available

When 2022 data is available, retrain the model with the latest data!



What Have We Learnt?

- Understand what is time series
- How to perform time series analysis
- How to perform time series forecasting using moving average technique
- Understand the technique to validate a time series forecast