Start-Tech Academy

Time Series and Forecasting

Time Series

A time series is a series of data points indexed (or listed or graphed) in time order.

Year	Profit
2001	50000
2002	60000
2003	70000
2004	60000
2005	75000

	A	B
1	Time	Temperature
2	10:00 AM	72
3	10:05 AM	76
4	10:10 AM	83
5	10:15 AM	87
6	10:20 AM	91
7	10:25 AM	99
8	10:30 AM	99
9	10:35 AM	99
10	10:40 AM	107
11	10:45 AM	113
12	10:50 AM	119
13	10:55 AM	123
14	11:00 AM	127
15	11:05 AM	134
16	11:10 AM	139

City	Latitude	Temperature
Miami, FL	26	83
Houston, TX	30	82
Mobile, AL	31	82
Phoenix, AZ	33	92
Dallas, TX	33	85
Los Angeles, CA	34	75
Memphis, TN	35	81
Norfolk, VA	37	77
San Francisco, CA	38	64
Baltimore, MD	39	76
Kansas City, MO	39	76
Washington, DC	39	74
Pittsburgh, PA	40	71
Cleveland, OH	41	70
New York, NY	41	76
Boston, MA	42	72
Syracuse, NY	43	68

Time Series and Forecasting

Time Series Forecasting

Time series forecasting is the use of a model to predict future values based on previously observed values

	Year	Sales
3	2012	15000
4	2013	15500
5	2014	16000
6	2015	16500
7	2016	17000
8	2017	17500
9	2018	18000
10	2019	18500
11	2020	19000
12	2021	19500
13	2023	20500
14	2024	21000

Forecasting Application

Retail Industry

Forecasting Application

Energy Industry

Forecasting Application

Government

Forecasting Application

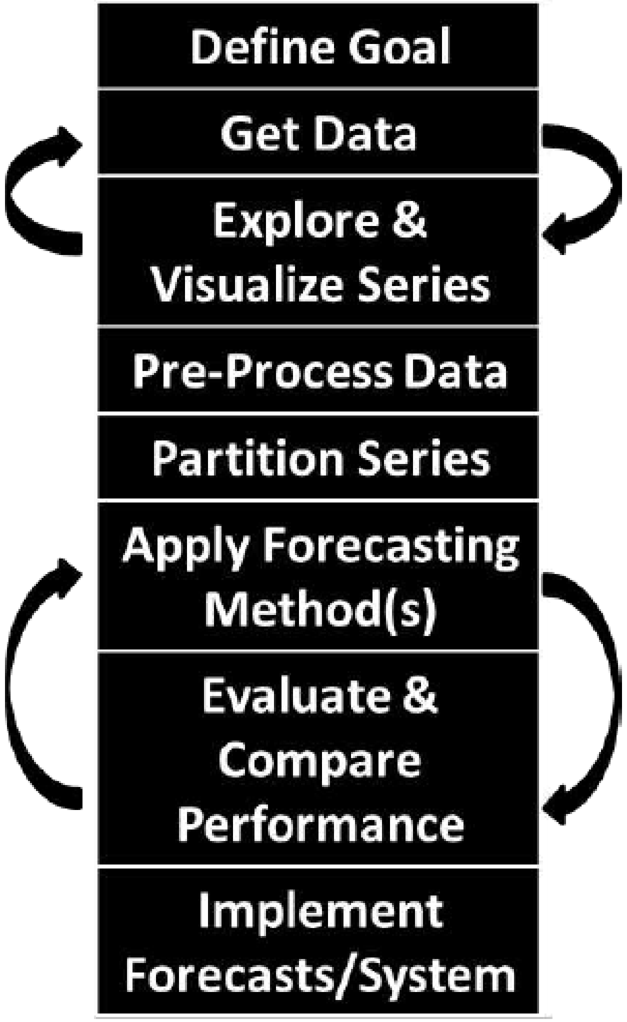
Financial Organization

Forecasting Application

Agriculture

Forecasting Process

Steps



Forecasting Process

Goal Definition

One must first determine

1. The purpose of generating forecasts,
2. The type of forecasts that are needed,
3. How the forecasts will be used by the organization,
4. What are the costs associated with forecast errors,
5. What data will be available in the future, and more

Forecasting Process

Goal Definition

Forecasting Process

Goal Definition Example

AMTRAK –US Railway company
Ridership Data
From Jan 1991 to March 2004

Forecasting Process

Goal Definition Example

AMTRAK

Descriptive

Evaluating the effect of some event, such as airport closures due to inclement weather, or the opening of a new large national highway. This goal is retrospective in nature, and is therefore descriptive or even explanatory.

Predictive

Amtrak might have is to forecast future monthly ridership on its trains for purposes of pricing. Using demand data to determine pricing is called "revenue management" and is a popular practice by airlines and hotel chains

Forecasting Horizon

Basic Notation

$t = 1, 2, 3, \ldots$	An index denoting the time period of interest. $t = 1$ is the first period in a series.
$y_1, y_2, y_3, \ldots, y_n$	A series of n values measured over n time periods, where y_t denotes the value of the series at time period t. For example, for a series of daily average temperatures, $t = 1, 2, 3, \ldots$ denotes day 1, day 2, and day 3; y_1, y_2, and y_3 denote the temperatures on days 1,2, and 3.
F_t	The forecasted value for time period t.
F_{t+k}	The k-step-ahead forecast when the forecasting time is t. If we are currently at time period t, the forecast for the next time period $(t+1)$ is denoted F_{t+1}.
e_t	The forecast error for time period t, which is the difference between the actual value and the forecast at time t, and equal to $y_t - F_t$

Forecasting Horizon

Forecasting Horizon

The forecast horizon k is the number of periods ahead that we must forecast, and F_{t+k} is a k-step-ahead forecast.

In the Amtrak ridership example, one-month-ahead forecasts (F_{t+1}) might be sufficient for revenue management (for creating flexible pricing), whereas longer term forecasts, such as three-month ahead (F_{t+3}), are more likely to be needed for scheduling and procurement purposes.

Forecasting

Forecasting Use and Automation

It is important to understand how the forecasts will be used

- Over-prediction vs under-prediction
- Presented to management or to the technical department
- Forecasts be used directly or will they be "adjusted" in some way before use
- Should forecasts be numerical or binary ("event"/"non-event“)

Automation requirement

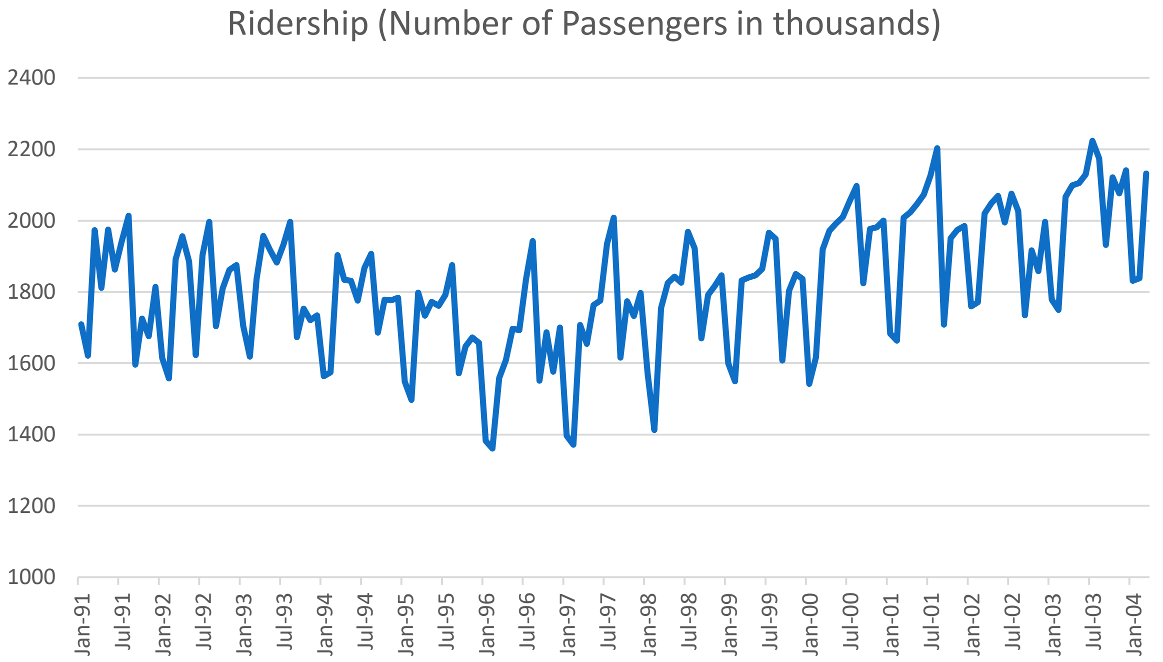
1. How many series need to be forecasted?
2. Is the forecasting an ongoing process or a one time event?
3. Which data and software will be available during the forecasting period?
4. What forecasting expertise will be available at the organization during the forecasting period?

Visualizing Time Series

The most basic and informative plot for visualizing a time series is the **time plot**.
it is a line chart of the series values (y1, y2, . . .) over time (t = 1, 2, . . .), with temporal labels (e.g., calendar date) on the horizontal axis

Time Plot

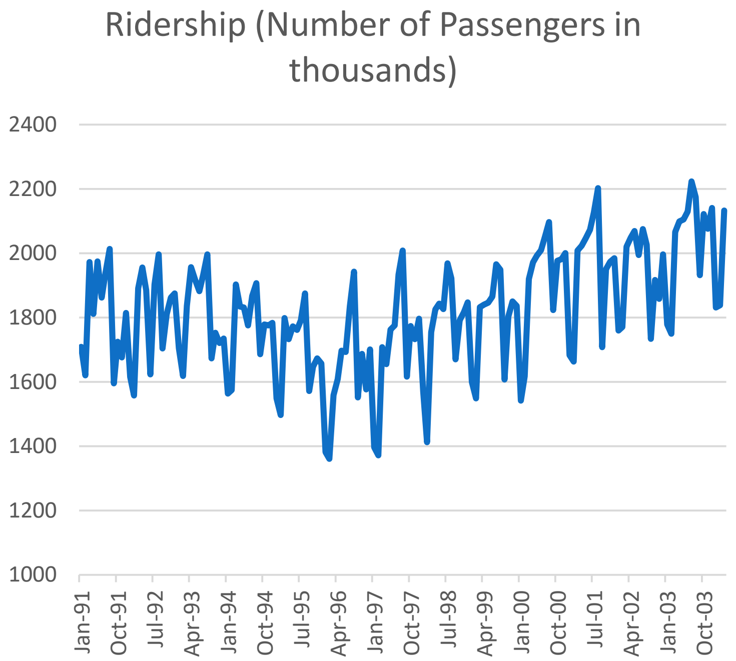
Month	Ridership
Jan-91	1709
Feb-91	1621
Mar-91	1973
Apr-91	1812
May-91	1975
Jun-91	1862
Jul-91	1940
Aug-91	2013
Sep-91	1596
Oct-91	1725
Nov-91	1676
Dec-91	1814
Jan-92	1615
Mar-04	2132

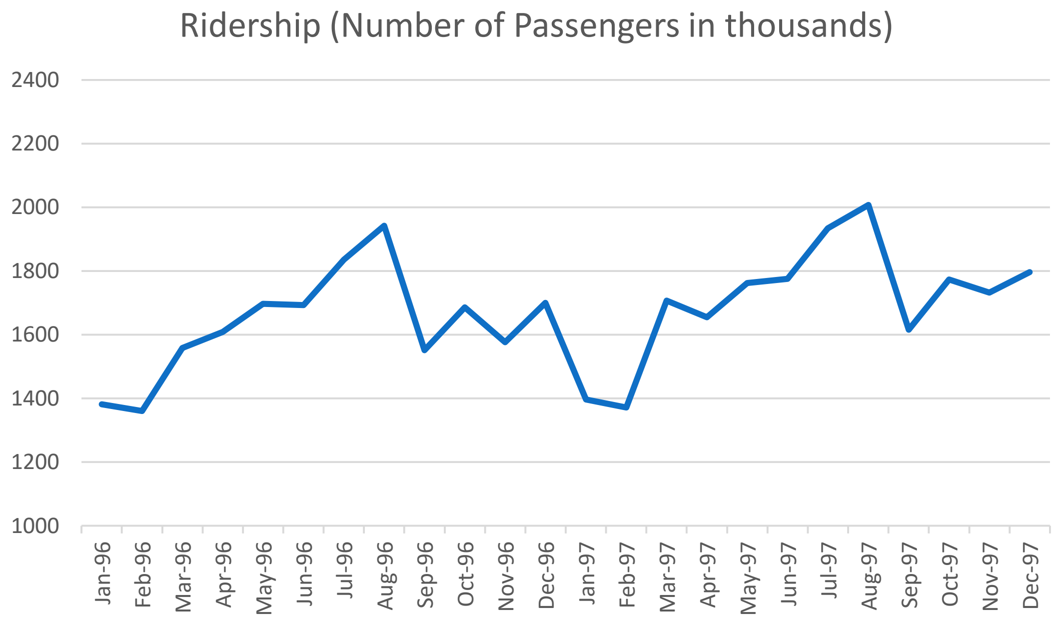


Visualizing Time Series

Zooming in to a shorter period within the series can reveal patterns that are hidden when viewing the entire series. This is especially important when the time series is long.

Zooming in

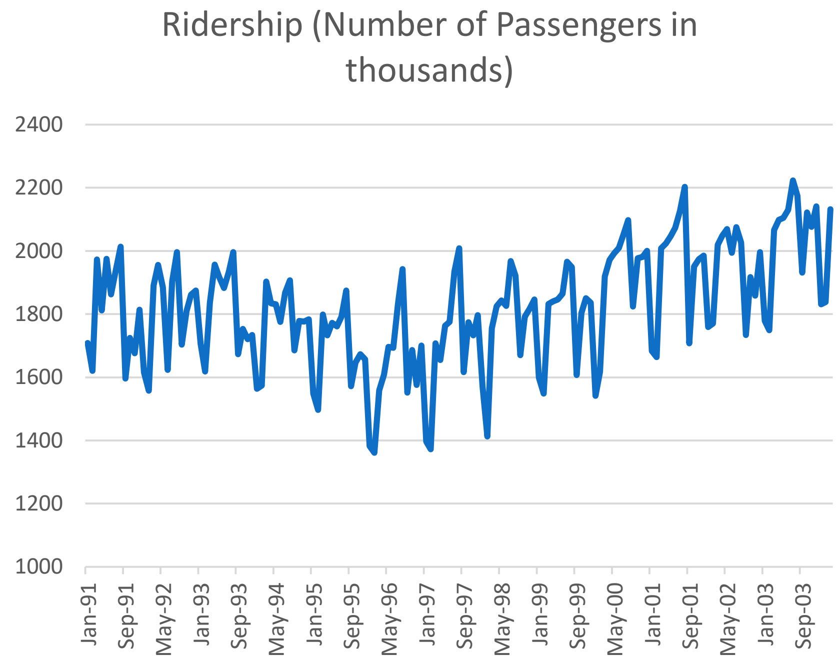


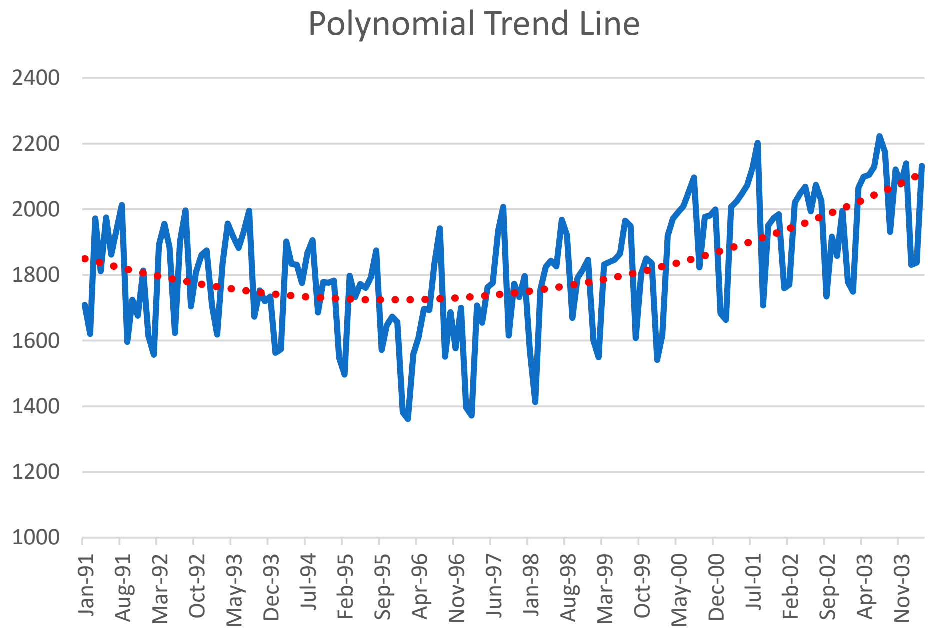


Visualizing Time Series

Another possibility for better capturing the shape of the trend is to add a trend line
By trying different trend lines one can see what type of trend (e.g., linear, exponential, cubic) best approximates the data.

Adding Trend Lines

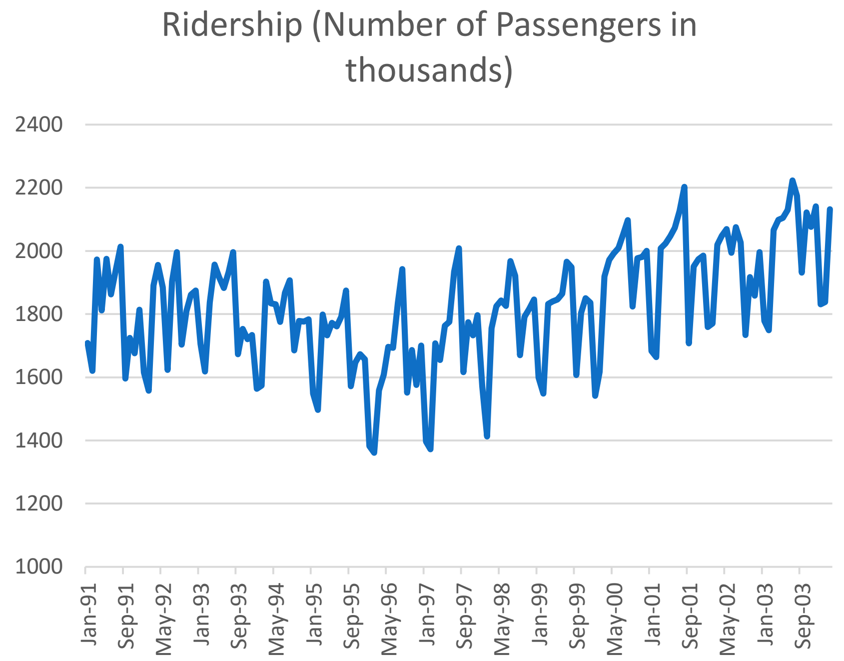


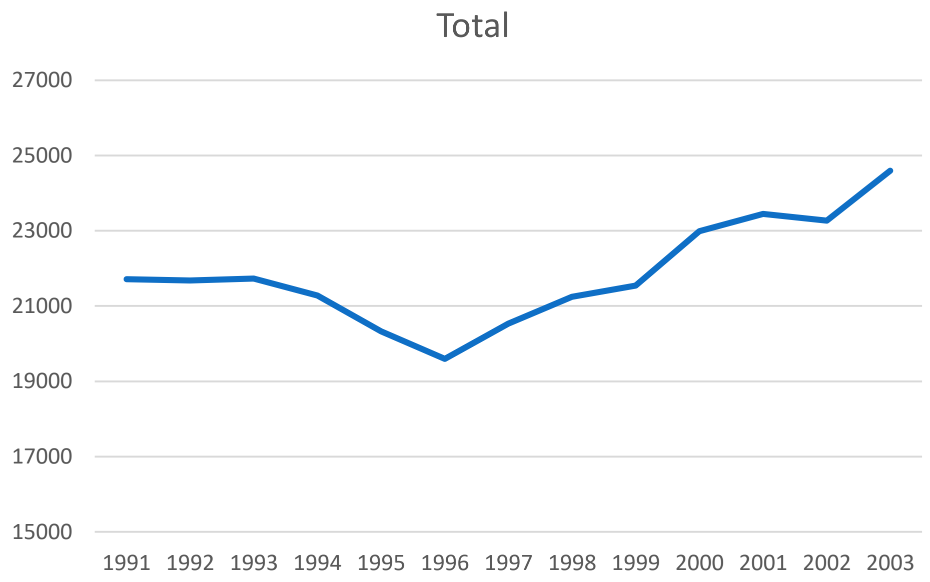


Visualizing Time Series

It is often easier to see trends in the data when seasonality is suppressed.
Suppressing seasonal patterns can be done by plotting the series at a cruder time scale

Suppressing Seasonality

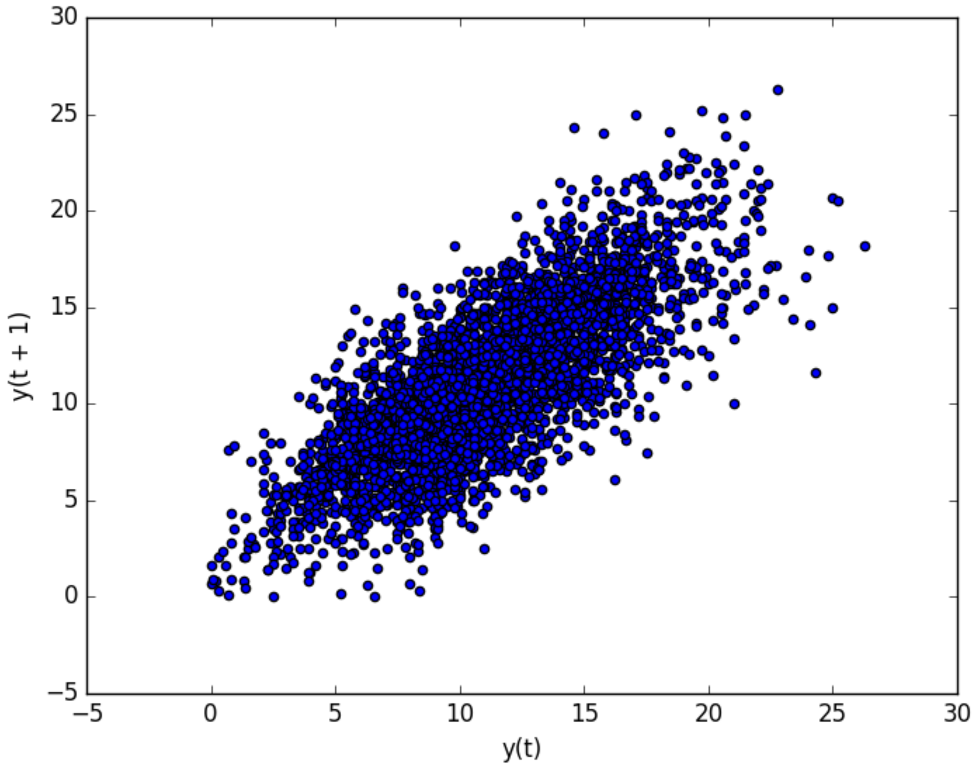


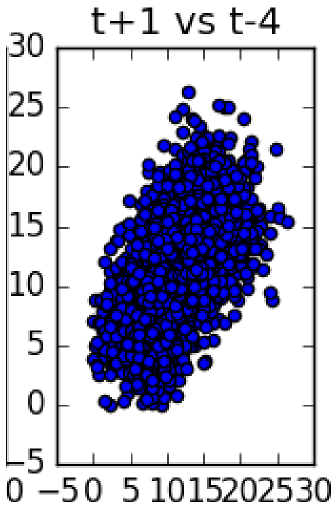


Visualizing Time Series

Time series modeling assumes a relationship between an observation and the previous observation. Previous observations in a time series are called lags. A useful type of plot to explore the relationship between each observation and a lag of that observation is called the scatter plot.

Lag Scatter Plot





Time Series – Feature Engineering

What we will learn

- What is Feature Engineering
- The rationale and goals of feature engineering time series data.
- How to develop basic date-time based input features.
- How to develop more sophisticated lag and sliding window summary statistics features.

Time Series – Feature Engineering

Basics

A time series dataset must be transformed to be modeled as a supervised learning problem.

time 1	value 1
time 2	value 2
time 3	value 3

input 1	output 1
input 2	output 2
input 3	output 3

Input variables are also called features in the field of machine learning, and the task before us is to create or invent new input features from our time series dataset.

Time Series – Feature Engineering

Types of Feature

- **Date Time Features:** these are components of the time step itself for each observation.
- **Lag Features:** these are values at prior time steps.
- **Window Features:** these are a summary of values over a fixed window of prior time steps.

Date	Footfall
10 Jan 20	853
11 Jan 20	1376
12 Jan 20	1289
13 Jan 20	657

Weekend	Footfall 7 d ago	Last 7 Day avg
0	785	985
1	1456	972
1	1145	995
0	764	970

Time Series – Feature Engineering

Window Features

- **Rolling window:** add a summary of the values at previous time Steps
- **Expanding Window:** Another type of window that may be useful includes all previous data in the series.

Date	Footfall
10 Jan 20	853
11 Jan 20	1376
12 Jan 20	1289
13 Jan 20	657

Last 7 Day avg	Maximum till date
985	1195
972	1195
995	1376
970	1376

Time Series – Resampling

Resampling

- **Resampling –** Changing frequency of available data to match the frequency of required forecast
- **Types**
  1. Upsampling
  2. Downsampling

Resampling

Upsampling

- **Upsampling –** Where you increase the frequency of the samples, such as from minutes to seconds.

Quarter	Footfall
Quarter 1	853
Quarter 2	1376
Quarter 3	1289
Quarter 4	657

Month	Footfall
1	240
2	260
3	353
4	433
5	467
6	476
7	500
8	450
9	339
10	140
11	217
12	300

Resampling

Downsampling

- **Downsampling –** Where you decrease the frequency of the samples, such as from days to months.

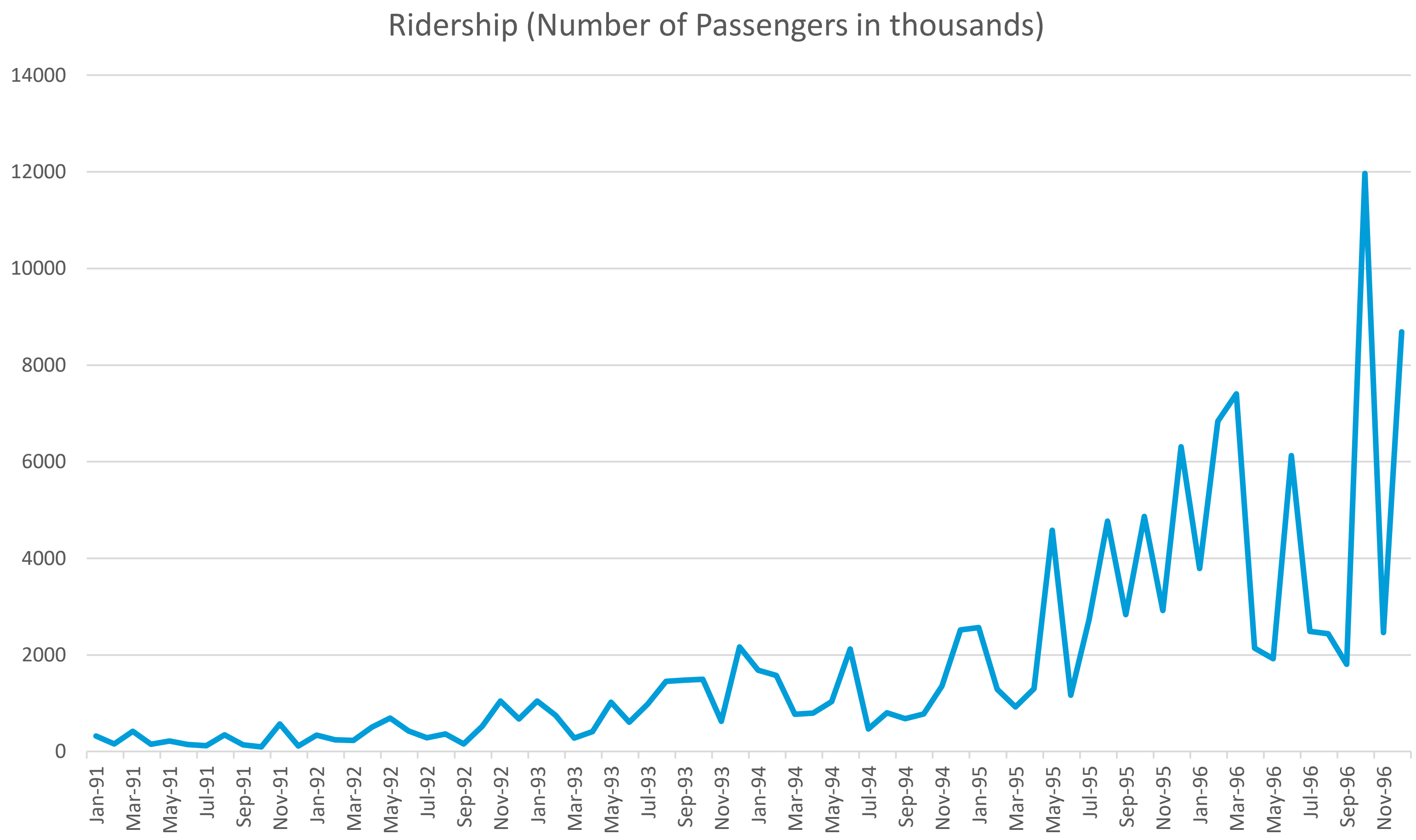
Quarter	Footfall
Y 1 – Q1	240
Y 1 – Q2	260
Y 1 – Q3	353
Y 1 – Q4	433
Y 2 – Q1	467
Y 2 – Q2	476
Y 2 – Q3	500
Y 2 – Q4	450
Y 3 – Q1	339
Y 3 – Q2	140
Y 3 – Q3	217
Y 3 – Q4	300

Quarter	Footfall
Year 1	1286
Year 2	1793
Year 3	1289

Power Transformation

To better identify the shape of a trend, it is useful to change the scale of the series

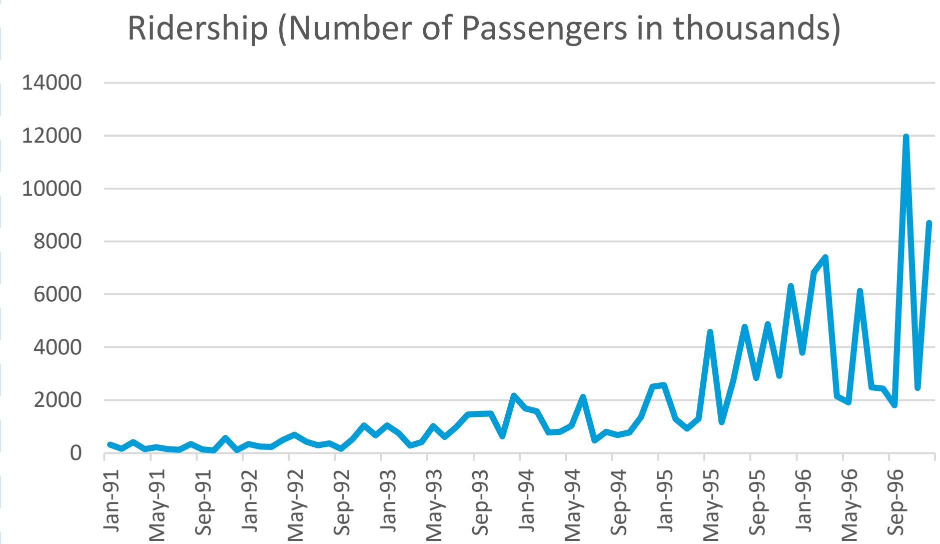
Changing the Scale

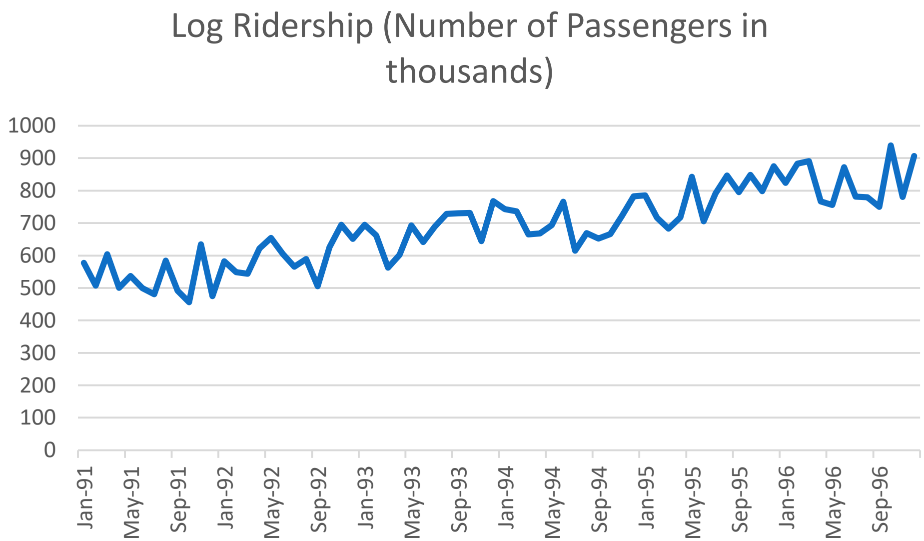


Power Transformation

To better identify the shape of a trend, it is useful to change the scale of the series
If the trend on the new scale appears more linear, then the trend in the original series is closer to an new function trend.

Changing the Scale





Time Series – Moving Average

Moving Average Smoothening

What we will cover

- What is moving average smoothening? Why?
- Centered Window vs Trailing Window MA
- Feature engineering and forecasting using MA

Time Series – Moving Average

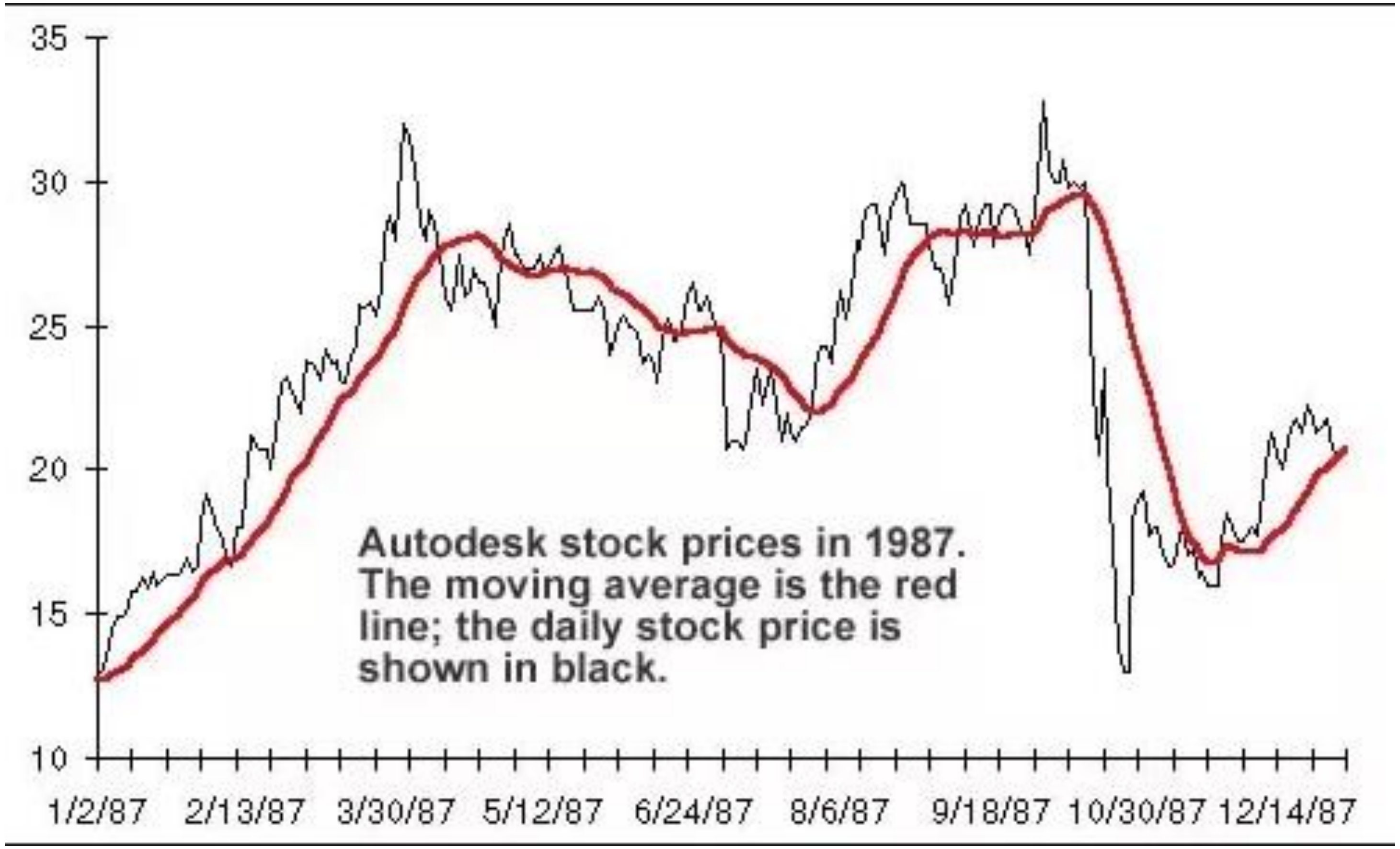
Moving Average Smoothening

MA smoothening is creating a new series where values are averages of the raw observations

Month	Temp. (°F)	Moving average
Jan	39	
Feb	42	
Mar	50	44
Apr	60	51
May	71	60
Jun	79	70
Jul	85	78
Aug	81	82
Sep	76	81

Time Series – Moving Average

Moving Average Smoothening

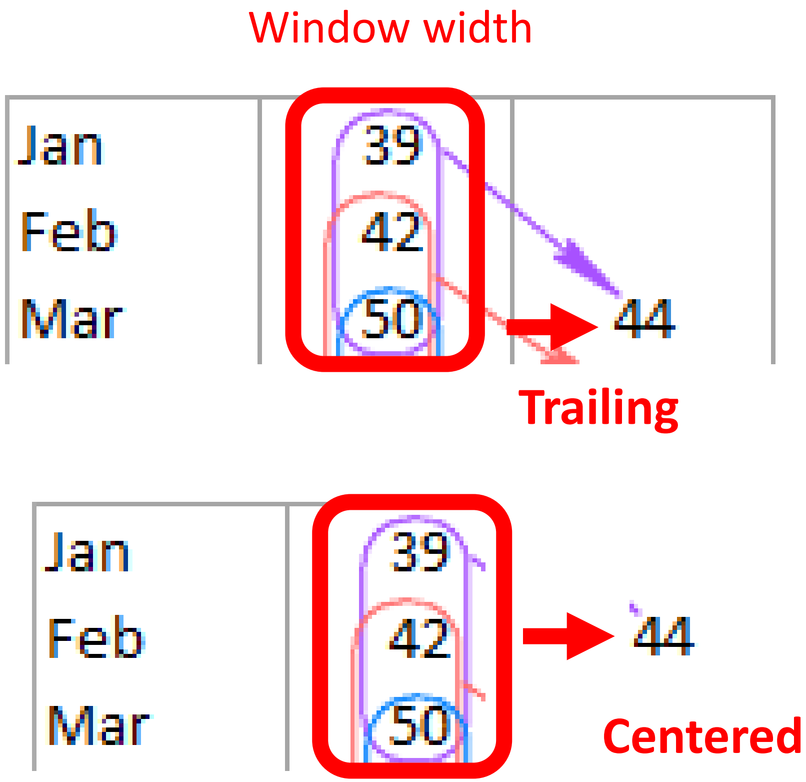


Time Series – Moving Average

Moving Average Smoothening

MA smoothening is creating a new series where values are averages of the raw observations

Month	Temp. (°F)	Moving average
Jan	39	
Feb	42	
Mar	50	44
Apr	60	51
May	71	60
Jun	79	70
Jul	85	78
Aug	81	82
Sep	76	81



Time Series – Moving Average

Moving Average Smoothening

Trailing MA can be used in 2 ways

1. Feature Engineering

Month	Temp. (°F)	Moving average
Jan	39	
Feb	42	
Mar	50	44
Apr	60	51
May	71	60

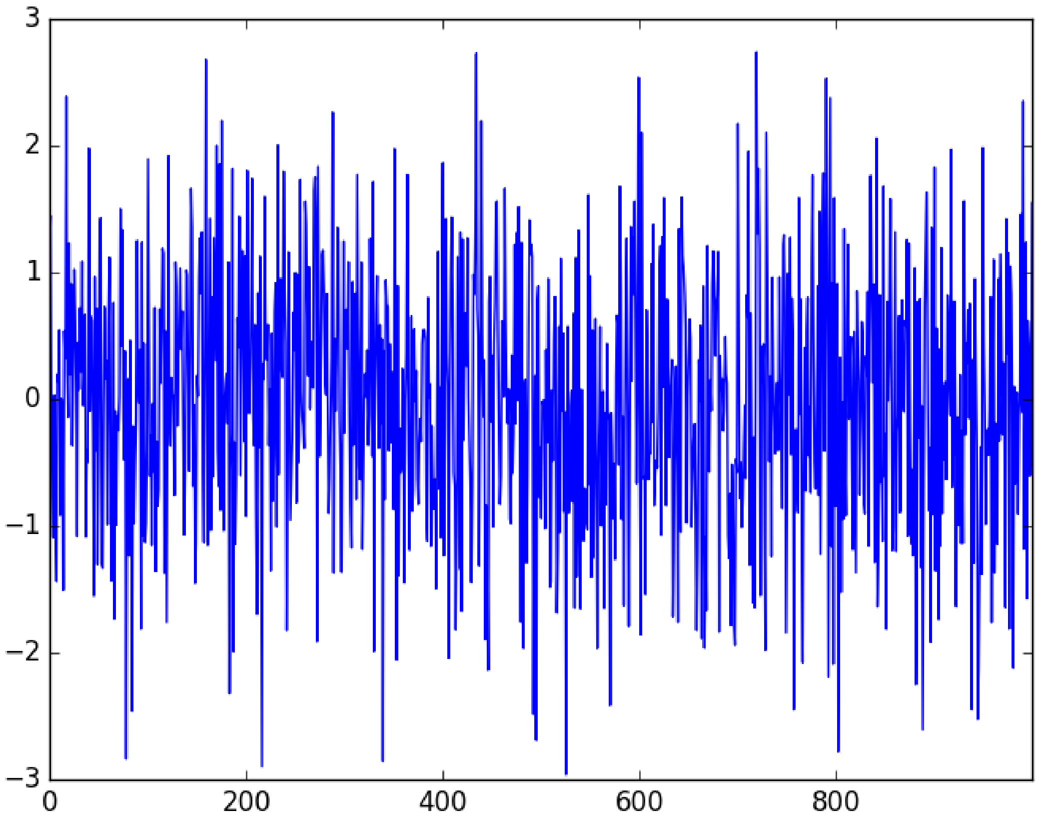
2. Forecasting

   Forecating for month of June can be 60, which is trailing MA for previous 3 months

Time Series – White Noise

White noise is a sequence of random numbers

White Noise



Time Series – White Noise

White Noise Why?

1. To identify whether the series to be forecasted is white noise or not
2. The error values of our predictions are white noise or not

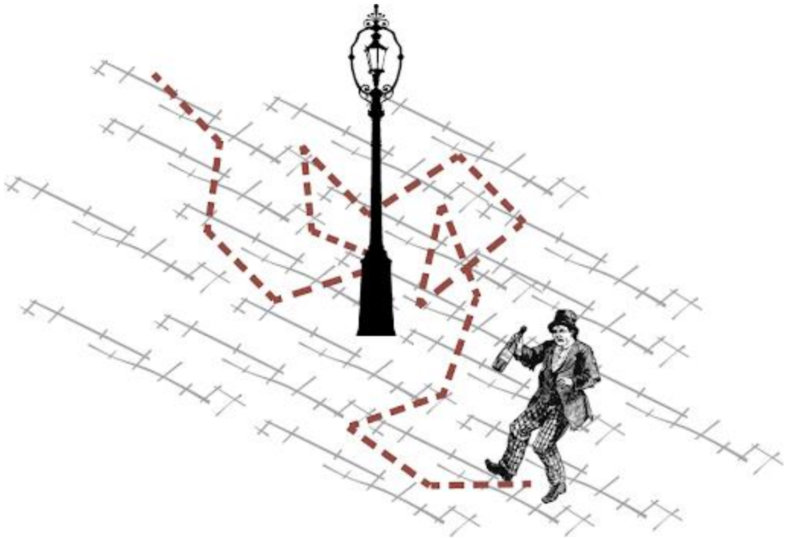
Summary

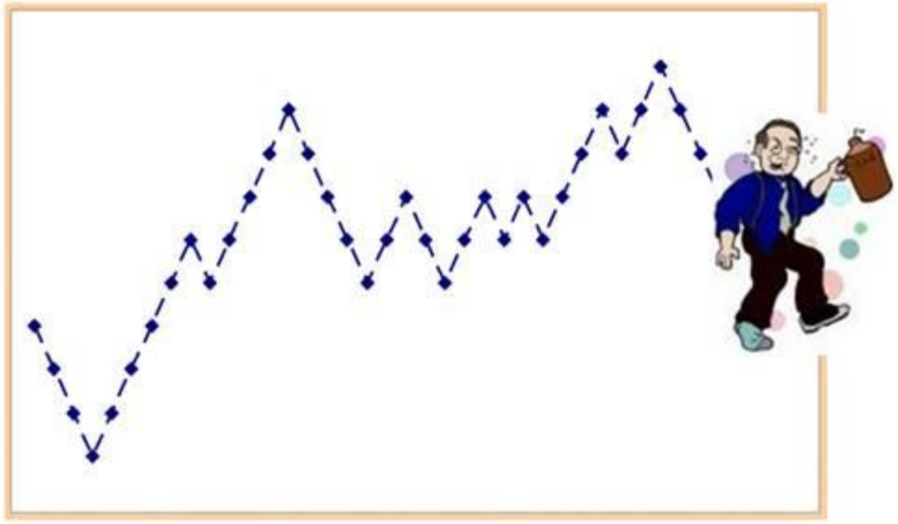
- Series should not be white noise
- Error values should be white noise

Time Series – Random Walk

Random Walk

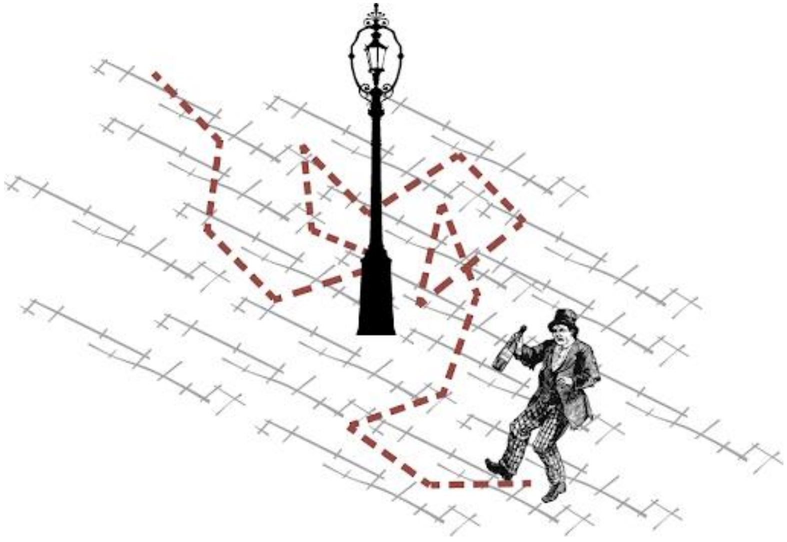
Next value of the series is the modification of the previous value in the sequence

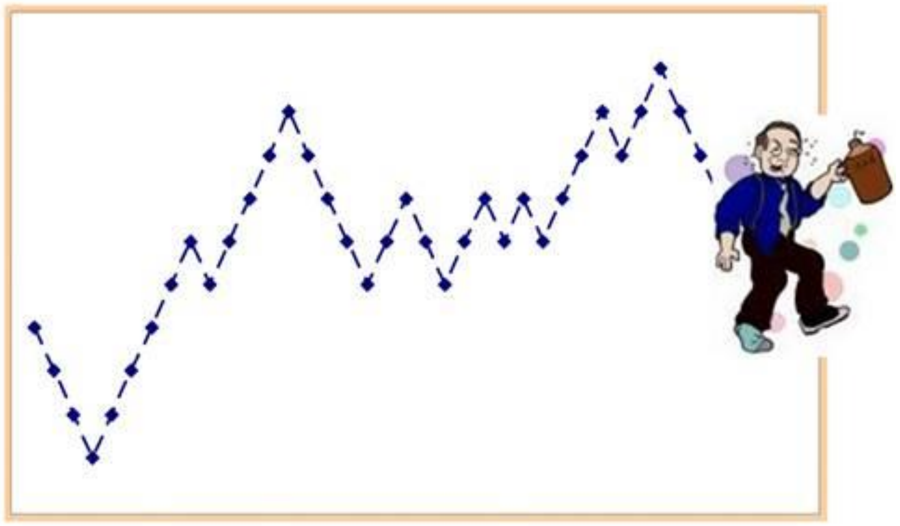




Time Series – Random Walk

Random Walk





In the case of Random Walk, naïve forecasting gives the best result

Time Series – Smoothing

Simple exponential smoothing

1. Moving Average smoothing

   Simple averaging

2. Exponential smoothing

   weighted averaging

$$F_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots$$

Smoothing constant

Time Series – Differencing

Removing trend and seasonality

Differencing is a method for removing trend and seasonality

1. Lag 1 Differencing

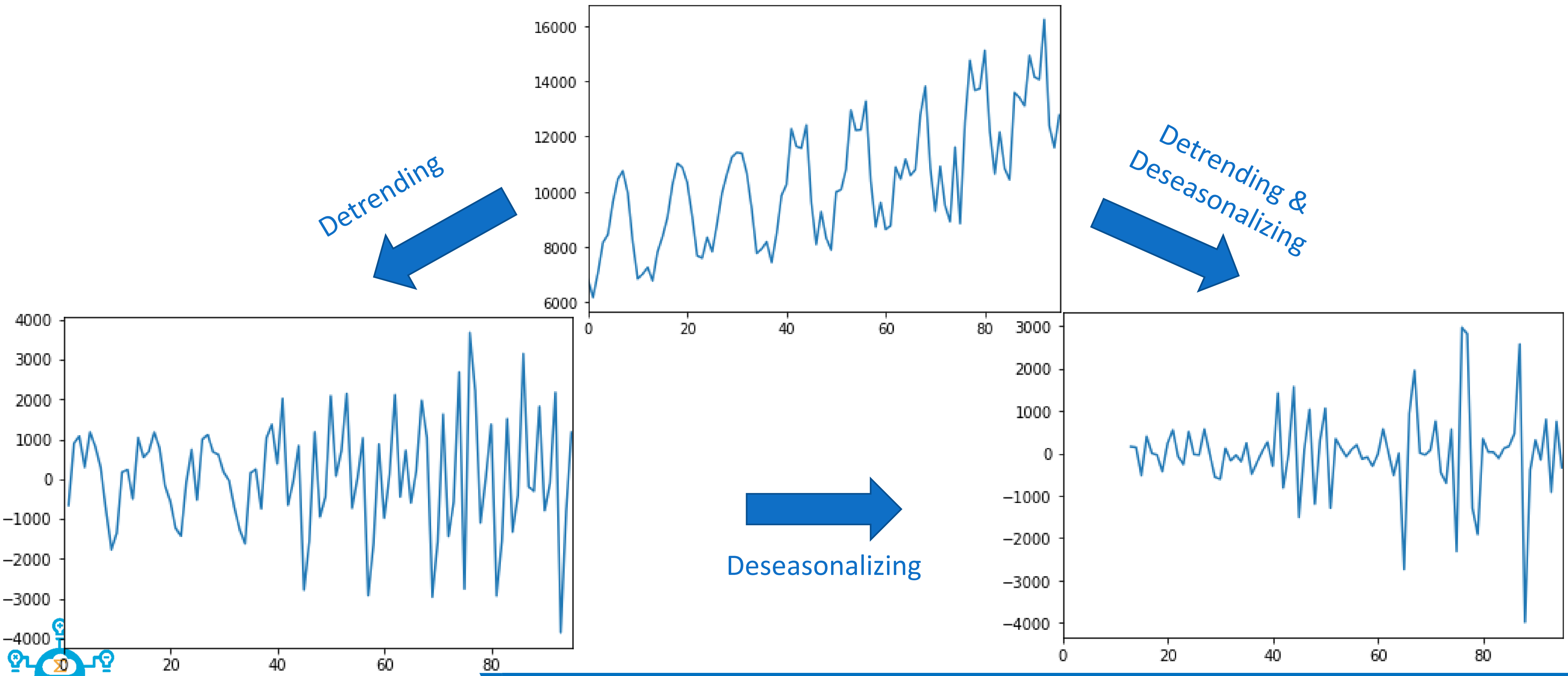
$$Y_t - Y_{t-1}$$

2. Lag k Differencing

$$Y_t - Y_{t-k}$$

	Month	MilesMM	lag1	Difference
0	1963-01-01	6827	NaN	NaN
1	1963-02-01	6178	6827.0	-649.0
2	1963-03-01	7084	6178.0	906.0
3	1963-04-01	8162	7084.0	1078.0
4	1963-05-01	8462	8162.0	300.0

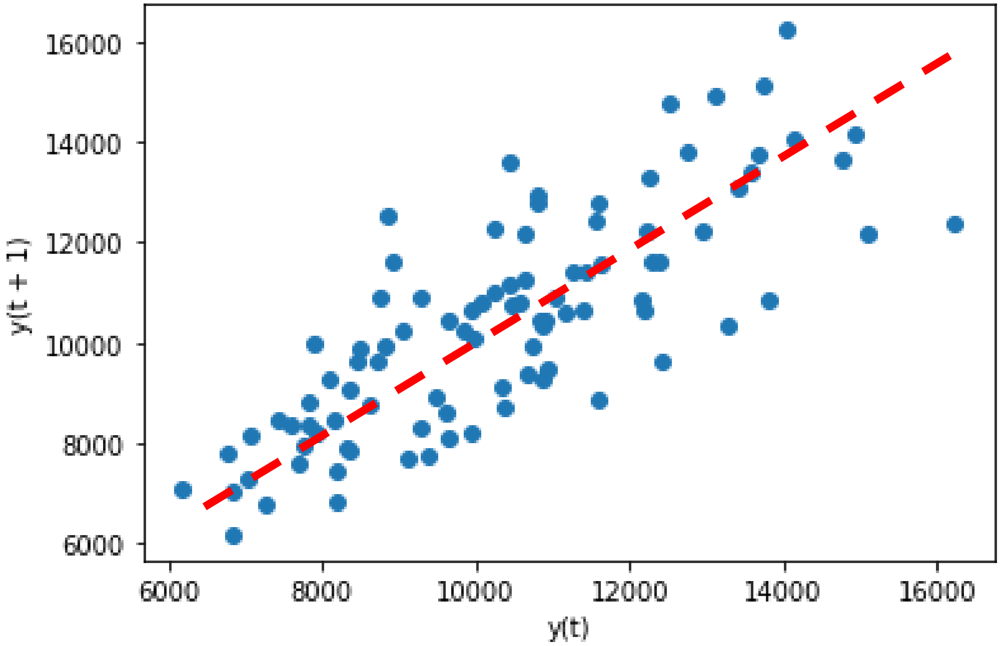
Time Series – Differencing



Time Series – Auto regression model

Auto regression

Use only on series without trend and seasonality



Simple regression model

$$\hat{Y} = B_0 + B_1 \times X_1$$

Time Series – Auto regression model

$$\hat{Y} = B_0 + B_1 \times X_1$$

Steps

1. **Step 1**

   Feed a set of X and Y values to the model

2. **Step 2**

   Model tries to fit a straight line on the given values. Finds value of B0 and B1 by minimizing total error

3. **Step 3**

   Use B0 and B1 values to predict future values of Y

Fore auto regression model – lag values of same variable are used as input values

Time Series – Moving Average model

It is different from moving average smoothing method

Moving Average

Input ~~Model~~ $\hat{Y}$ + Residual

Residual ~~Model~~ Forecasted Residual

Time Series – Moving Average model

Steps

1. **Step 1**

   Build a forecasting model (AR or Persistence model)

2. **Step 2**

   Find residual/ Forecast errors

3. **Step 3**

   Build a forecast model on residuals

4. **Step 4**

   Use forecasted residuals to update the initial forecast

If initial forecasting model is AR this technique is also called ARMA

Time Series – ACF and PACF

ACF and PACF

1. **Correlation-**

   Relation between 2 variables. Measured by Pearson's Correlation Coef.

Coef. Value	Relationship	Inference
+1	Positive	If x increases, y also increases
-1	Negative	If x increases, y decreases
0	No Relationship	If x increases, no effect on y

For time series analysis, instead of 2 different variables,

X is lagged value of y

So, therefore it is called AutoCorrelation => Correlation with itself

Time Series – ACF and PACF

ACF and PACF

Lag 1

	date	births	lag1
	1959-01-01	35	NaN
	1959-01-02	32	35.0
	1959-01-03	30	32.0
	1959-01-04	31	30.0
	1959-01-05	44	31.0
	1959-01-06	29	44.0
	1959-01-07	45	29.0

Correlation Coef = 0.8

Lag 2

	date	births	lag2
	1959-01-01	35	NaN
	1959-01-02	32	NaN
	1959-01-03	30	35.0
	1959-01-04	31	32.0
	1959-01-05	44	30.0
	1959-01-06	29	31.0
	1959-01-07	45	44.0

Correlation Coef = 0.6

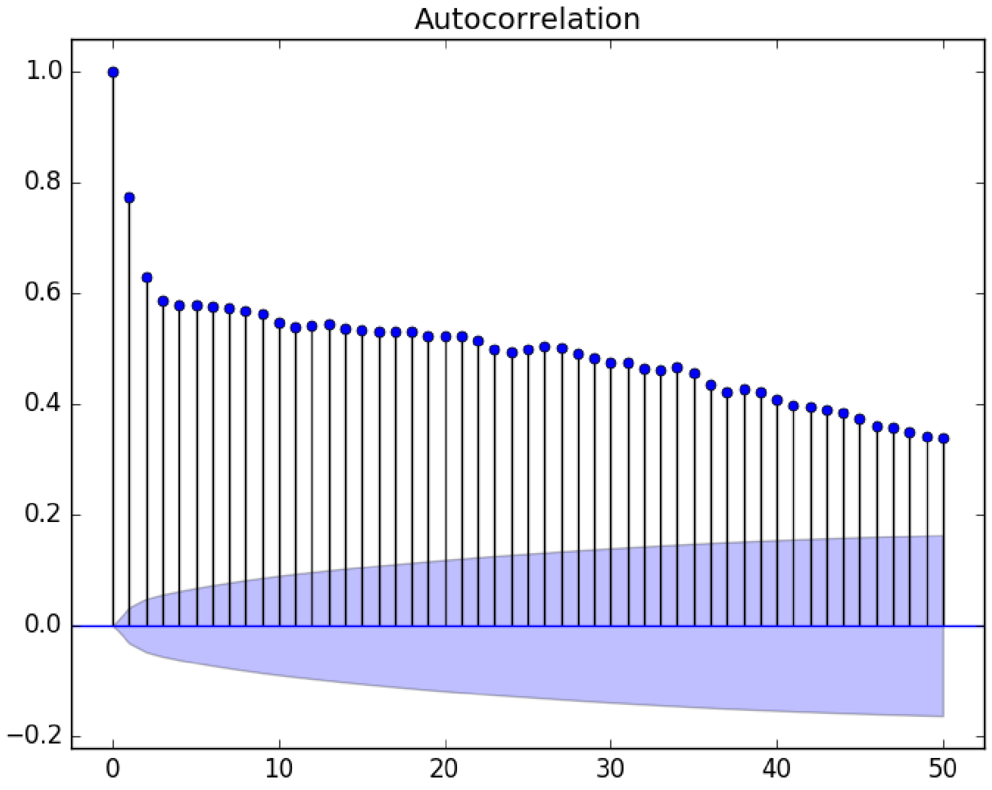
Lag 3

	date	births	lag3
	1959-01-01	35	NaN
	1959-01-02	32	NaN
	1959-01-03	30	NaN
	1959-01-04	31	35.0
	1959-01-05	44	32.0
	1959-01-06	29	30.0
	1959-01-07	45	31.0

Correlation Coef = 0.4

Time Series – ACF and PACF

Autocorrelation



Cone of 95% confidence

Lagged values with correlation coef. outside this cone can be used for MA

Time Series – ACF and PACF

Partial Autocorrelation

Lag 1

date	births	lag1
1959-01-01	35	NaN
1959-01-02	32	35.0
1959-01-03	30	32.0
1959-01-04	31	30.0
1959-01-05	44	31.0
1959-01-06	29	44.0
1959-01-07	45	29.0

Correlation
Coef = 0.8

Lag 2

date	births	lag2
1959-01-01	35	NaN
1959-01-02	32	NaN
1959-01-03	30	35.0
1959-01-04	31	32.0
1959-01-05	44	30.0
1959-01-06	29	31.0
1959-01-07	45	44.0

Correlation
Coef = 0.6

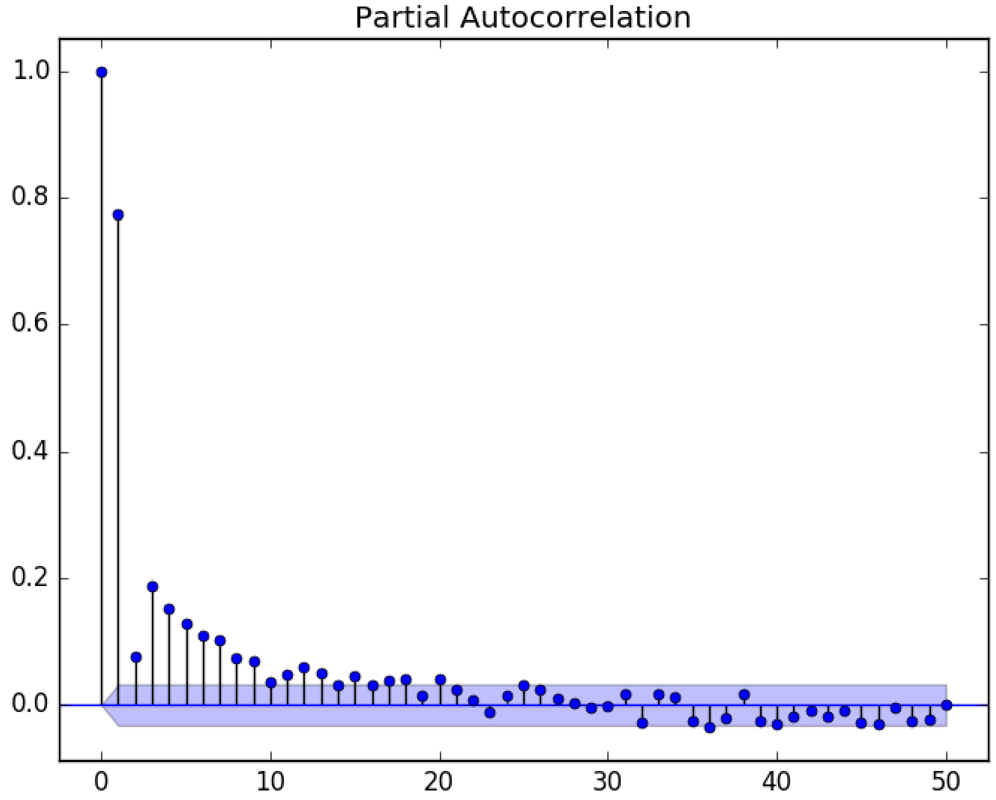
Lag 1 –Lag 2

date	births	lag1	lag2
1959-01-01	35	NaN	NaN
1959-01-02	32	35.0	NaN
1959-01-03	30	32.0	35.0
1959-01-04	31	30.0	32.0
1959-01-05	44	31.0	30.0
1959-01-06	29	44.0	31.0
1959-01-07	45	29.0	44.0

Correlation
Coef = 0.8

Time Series – ACF and PACF

Partial Autocorrelation



Lagged values with correlation coef. outside this cone can be used for Autoregression

Time Series – ARIMA

ARIMA

ARIMA - Auto Regression Integrated Moving Average

Integrated – To handle trend and seasonality

It integrates differencing method into ARMA to handle any type of series

1. **Step 1**

   Use differencing to remove trend and seasonality

2. **Step 2**

   Use AR to do forecasting

3. **Step 3**

   Use MA on residuals to update the forecast

4. **Step 4**

   Add Trend and Seasonality to get the forecasted value

Time Series – ARIMA

(AR I MA)

| | |

p d q

Parameter

1. **P – Order of autoregression**

   how many lag variables to choose

2. **D – Order of Integration**

   Number of differencing needed

3. **Q – Order of Moving Average**

   What is the window size for Moving Average