Time Series Analysis:

An introduction to Time Series with python

Sameh BEN FREDJ
XKE, Novembre 2017

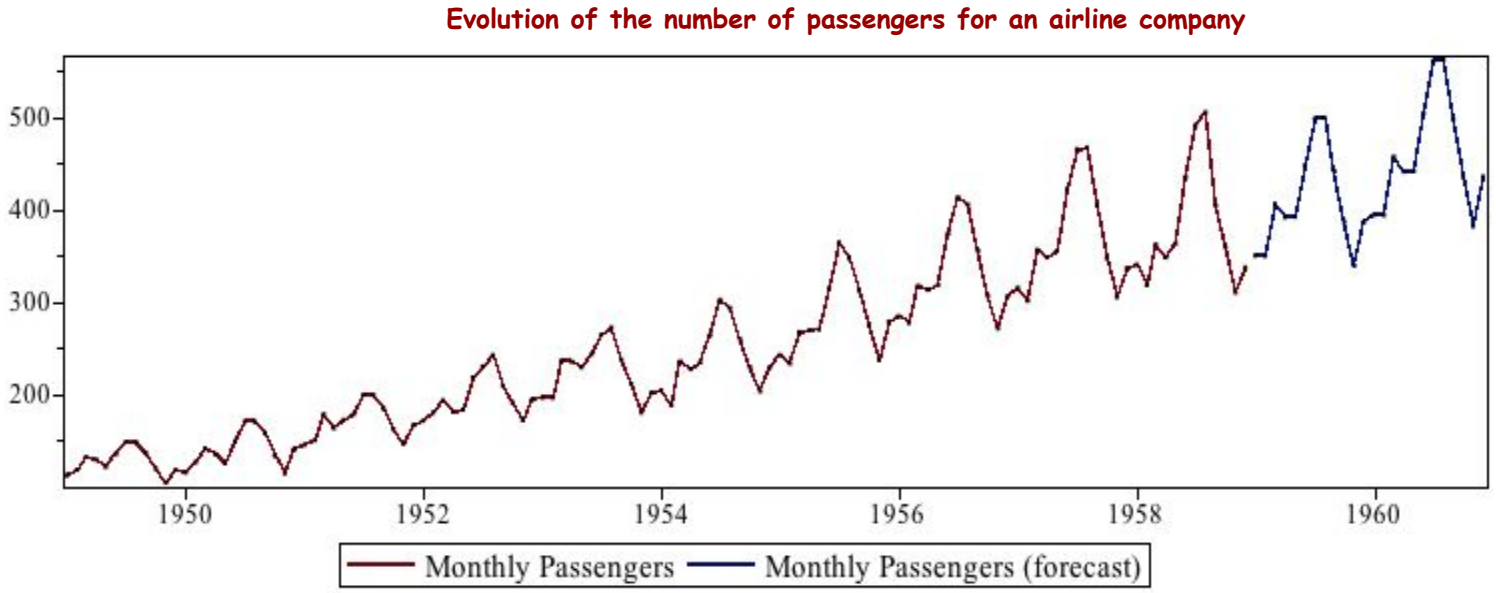
Plan

1) Quick Introduction to Time Series Analysis.

2) Hands On:
    a) Dealing with time in Pandas
    b) Time Series analysis
    c) Prediction/forecasting

WHAT?

A Series of data indexed in Time Order.



What for ?

Time series data are often used to predict / forecast the future instances of a measured variable based on its past observations.

Time Series data are almost everywhere !

- Key Economic Indicator Projections
- Government prediction for unemployment rates
- News: What is going to happen for the stock market after the election of Trump ?
- Companies forecast sales
- Speech recognition
- IoT/ sensor data forecasting/ anomaly detection.
- Air quality/temperature forecasting
- Physiological models for health monitoring (e.g., glucose levels in diabetics)

Some Hypothesis:

Regular spaced time series *Vs* Irregular spaced time series

Univariate *Vs* Multivariate

Tools

2 main librairies

Pandas:

It has many some functionalities dedicated to analyse times series objects which allow to store time data and perform operations on it quickly:

- *Date functionalities*
- *Converting to timestamps*
- *Generating ranges of time series data.*
- *Indexing time data*
- *Shifting/lagging data, resampling, etc.*

Statsmodels:

Python library and its module 'tsa' which contains useful functions for time series analysis, data modelling/estimation and forecasting: e.g ARIMA models.

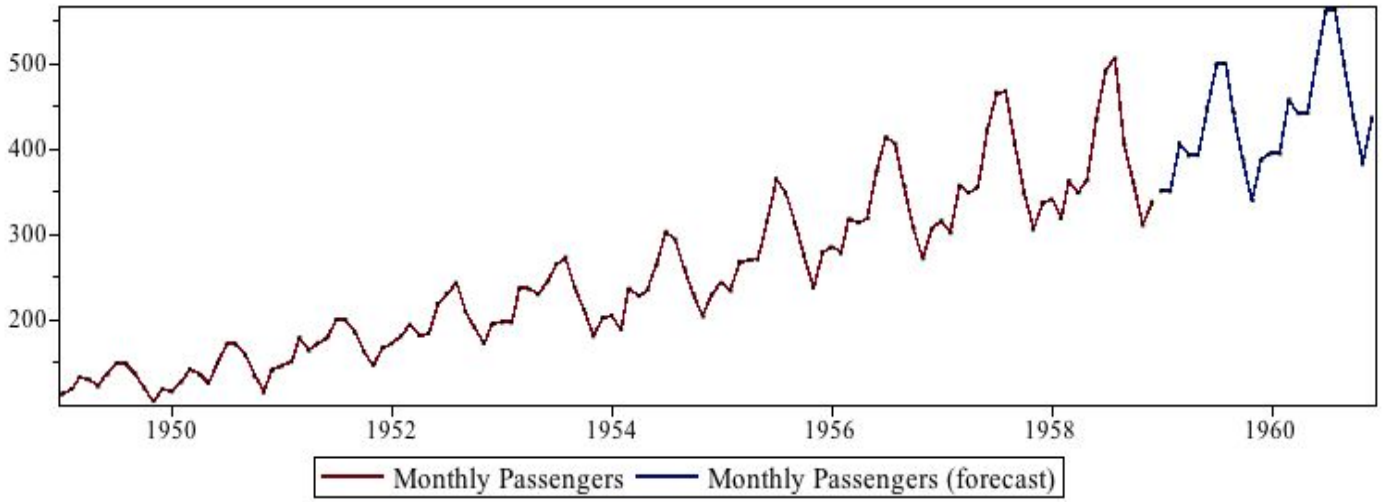
Basic components

Trend

A long-term increase or decrease in the data.

A “changing direction” trend: it goes from an increasing to a decreasing trend

Monthly totals of international airline passengers between 1949 to 1960



Seasonality

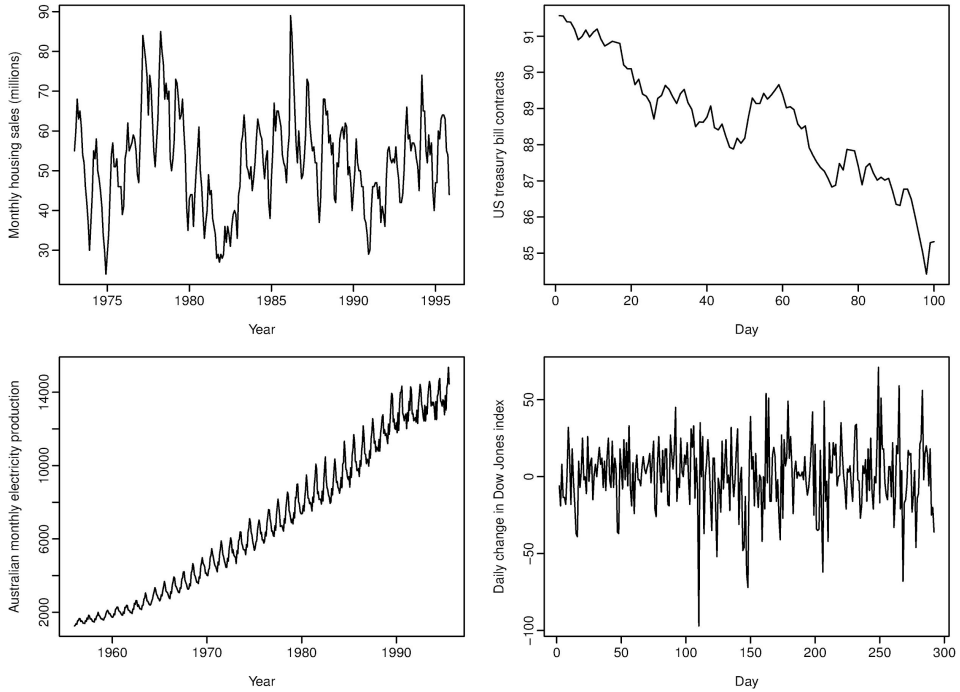
A repeating pattern within any fixed period.

Many type of seasonality:

- Time of Day.
- Daily.
- Weekly.
- Monthly.
- Yearly.

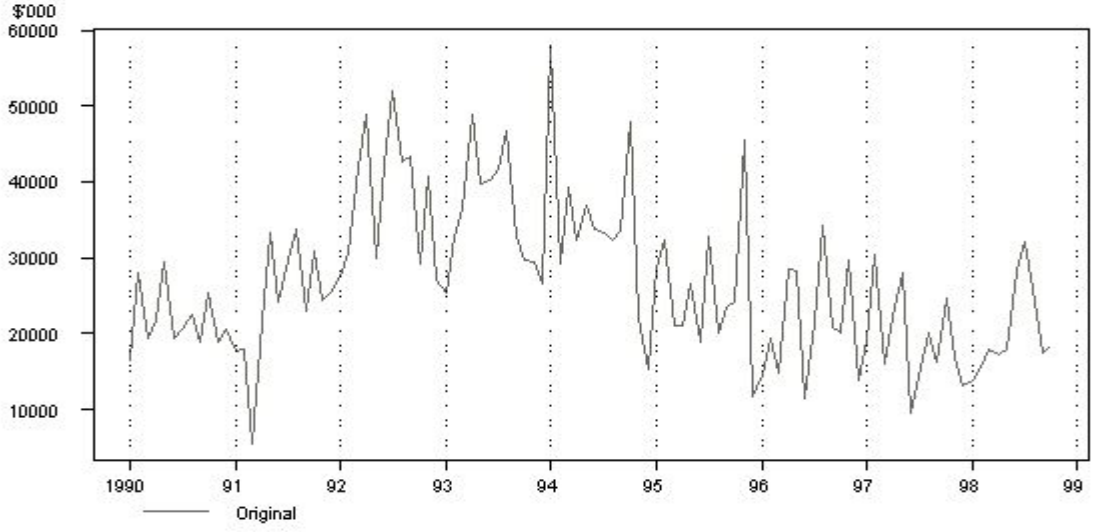
Cyclical

Data exhibit rises and falls that are not of fixed period.



Residuals (The irregular component)

What remains after removing trend and seasonal component of a time series. It results from short term fluctuations in the series which are neither systematic nor predictable.



Time Series stationarity

Time Series Stationarity

TS is said to be stationary if its statistical properties remain constant over the time.

One definition of stationarity is: its statistical components remains constant over time (e.g, mean, variance) .

Most Time Series Models work if the data is stationary.

Trend and seasonality, can sometimes, affect TS stationarity

How to check data stationarity ?

- Visually/plotting rolling means
- Using some statistical tests:

    The Augmented Dickey-Fuller test.

Time Series Modelling

ARIMA

- Stands for **Autoregressive Integrated Moving Averages.**

→ Considered as the most general class of models for forecasting linear time series data.

- ARIMA (p,d,q)

AR: Auto-Regressive model

- The value of a variable in one period is related to its values in previous periods.
- AR(p):

$$AR(p) : X_t = c + \sum_{k=1}^{p} \varphi_k X_{t-k} + \varepsilon_t = c + \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \ldots + \varphi_p X_{t-p} + \varepsilon_t$$

- p: Represents the number of lags of the dependent variable to consider.

MA: Moving Average model

- The value of a variable in one period is related to the residuals from previous periods.
- MA(q):

$$MA(q) : X_t = \mu + \varepsilon_t + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- q: represents the number of lags of residuals to consider.

I: Integrated (differencing)

- Differencing is a way of transforming a nonstationary serie to a stationary one. This is done by subtracting the observation in the current period from the previous one.
- "First differenced”data:

$$y'_t = y_t - y_{t-1}$$

Time Series data exploration and analysis

- ACF: Autocorrelation Function

- PACF: Partial Autocorrelation Function

Thanks!

Q&A