Components:

1. Trends: Time varying increasing / decreasing linearly.
2. Seasonality: Exhibiting a regular repeating pattern of some sort, increased volumes during weekends, increased volumes around festivals, etc. Seasonal patterns happen with a periodicity that is known and predictable.
3. Long-Term Cyclicity: Long-term cyclic behaviour. Economic inﬂation-recession cycles are a typical example.
4. Autocorrelation: Dependence of the next value on the past.

Difference between Regression and Time series:

In normal regression problems we assume that the explanatory variables in any two data points are completely independent of each other and they have been sampled from the same distribution.

The independence assumption clearly doesn’t hold for timeseries data. There is an obvious dependence between data points corresponding to observations at successive time-stamps.

Any typical regression dataset has not ’canonical’ ordering of the datapoints in the dataset—any random shufﬂing of the dataset has no effect on the essential nature (the relationship it exhibits between the explanatory variables and the response variable) of the dataset.

Time-series data on the other hand has one canonical ordering—ordered ascending on the timestamp—and that’s the only order that is meaningful for the dataset. Successive datapoints in a time-series will naturally have signiﬁcant correlation that cannot be ignored.

Approach:

1. Plotting: Line/Scatter plot to identify trend, seasonality visually. Presence of outliers
2. Stationarize the data: By removing the Trend, seasonality, cyclicity. This is done by series of transformations(S.
3. Test the residual is stationary, if not then tweak step 2. If stationary check for auto regressive / moving average part.
4. ConstructtheﬁnalmodelbyapplyingtheinversetransformationS−1(.)onthestationary model.
5. Check if the ﬁnal combined model models the data well. To do this we need to compute the residual series ²t =Xt−S−1(M(t)),0≤t≤T that remains after we eliminate the value predicted by the ﬁnal model, and then verify if what remains (the residue) is ’pure’ noise.

Stationary Time Series Analysis and ARMA models:

1. Stationarity: **Time invariant** relationship exists between values seen at different timestamps i.e., the set of values (X1,...,Xn) is identical to the joint distribution of the values (Xt+1,...,Xt+n) for any t, n>0.
2. The other way to interpret the stationarity is: if a series is stationary, then the shape of the time-series plot will look identical irrespective of the time t at which you start collecting / observing the data.
3. In reality, the time series we get is not strongly stationary, so we actually look for weak stationarity. If this is not found, we term the residual as a white noise.
4. Weakly Stationary Time Series:

DataCamp:

Two trending series may show a strong correlation even if they are completely unrelated. This is referred to as "spurious correlation". That's why when you look at the correlation of say, two stocks, you should look at the correlation of their returns and not their levels.

**Autocorrelation**:

* Is measured as correlation between a time series and lagged copy of itself
* With Financial time series, when returns have a negative autocorrelation, we say it is “mean reverting”. Positive autocorrelation : “trend-following”.
* For ex: an autocorrelation of order 3 returns the correlation between a time series at points(t1, t2, t3,…) and its own values lagged by 3 time points, i.e.,(t4, t5, t6,….)
* Used to find repetitive patterns or periodic signal in time series

**Autocorrelation Function**: Shows the entire autocorrelation for different lags. Any significant autocorrelation implies that the series can be forecast from the past.

**Partial autocorrelation:**

* Partial autocorrelation is also correlation of a time series with lagged copy of itself, however it removes the effect of previous time points.
* For ex: a partial autocorrelation function of order 3 returns the correlation between our time series (t1, t2, t3,…) and itself by 3 time points (t4, t5, t6,….) but, only after removing all effects attributable to lags 1 and 2.

**White Noise**:

Constant mean, constant variance with time and Zero autocorrelation at all lags.

**Stationarity**:

Joint distribution does not depend on time.

Strong Stationarity: Entire distribution of data is time-invariant.

Weak Stationarity: Mean, variance and autocorrelation are time-invariant.

**Advanced Topics**:

1. GARCH models
2. Nonlinear models
3. Multivariate Time Series Models
4. Regime switching models
5. State space models and Kalman filtering

To Explore:

Data from Zillow research

Kaggle competitions

Reddit Data

Manipulating Time series data in python

Importing and managing financial data in python

Statistical thinking in Python(Part1)