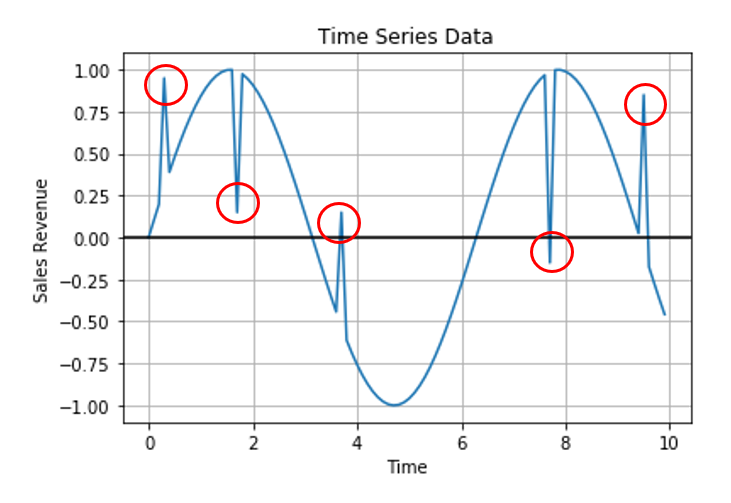
What is Anomaly?

In time series data, an anomaly or outlier can be termed as a data point which is not following the common collective trend or seasonal or cyclic pattern of the entire data and is significantly distinct from rest of the data.



From the above time series plot, we can see that, 5 data points which are significantly different from the overall series is highlighted in red circle. So these 5 anomaly data points does not follow the overall sinusoidal nature of the time series and hence can be termed as time series anomaly.

**Why is Time Series Anomaly Detection so important?**

in order to estimate the “new normal” and regroup and restructure business strategies and decision making process, it is very important to keep track of anomalies in every sector.

From Sales and Marketing, to Supply Chain and Manufacturing, every stage of the business requires sufficient information, especially about these anomalies to shape their process and maximize productivity and outcome.

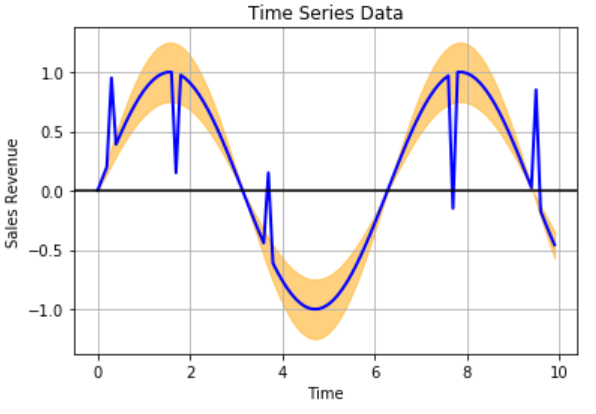
wherever we have a common pattern in data, especially for time series data, it is very important to segregate the outliers and spare time and attention to study these.

**How to do Time Series Anomaly Detection?**

From a very high level and in a very generic way, time series anomaly detection can be done by three main ways:

1. By Predictive Confidence Level Approach
2. Statistical Profiling Approach
3. Clustering Based Unsupervised Approach

**Time Series Anomaly Detection By Predictive Confidence Level Approach:**

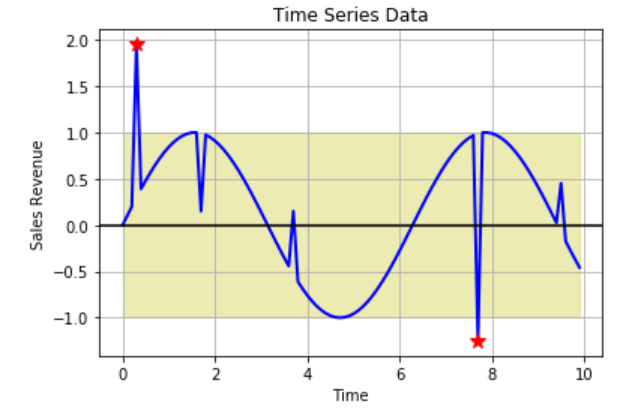


One way of doing anomaly detection with time series data is by building a predictive model using the historical data to estimate and get a sense of the overall common trend, seasonal or cyclic pattern of the time series data.

Using the predictive model to forecast future values and based on the error rates (which can be calculated using MAPE — Mean Absolute Percentage Error), we can come up with a confidence interval, or a confidence band for the predicted values and any actual data point which is falling beyond this confidence band is an anomaly.

The main advantage of this approach is finding local outlier but the main disadvantage is, this approach highly depends on the efficiency of the predictive model. Any loop hole in the predictive model can give false positives and false negatives.

**Time Series Anomaly Detection by Statistical Profiling Approach**



has been in use effectively in the field of economics and finance sectors.

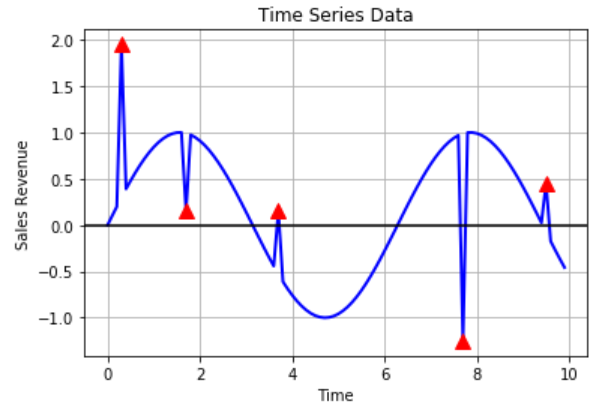
Generating a statistical model or profile of the given data can be the fastest and the most useful approach, as this method can provide a more controlled and explainable outcome

This can be done by calculating statistical values like mean or median moving average of the historical data and using a standard deviation to come up with a band of statistical values which can define the uppermost bound and the lower most bound and anything falling beyond these ranges can be an anomaly.

this approach is very handy and can always be the baseline approach, instead of going with any sophisticated and complex methods which require a lot of fine tuning and may not be explainable.

This is very effective for highly volatile time series as well, as most of the time series predictive model algorithms fail when the data is highly volatile.

main drawback of this approach is detecting the local outliers. As we see in the previous figure, out of five obvious anomaly points only 2 most significant anomaly points got detected.

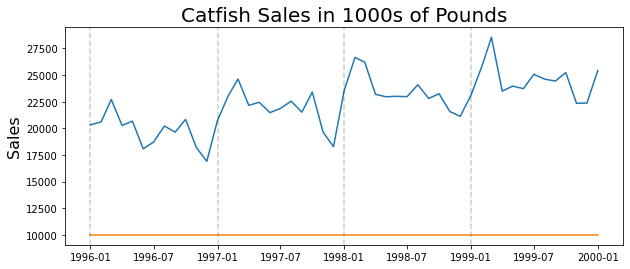
**Time Series Anomaly Detection By Clustering Based Unsupervised Approach:**

**Unsupervised approaches are extremely useful for anomaly detection as it does not require any labelled data,**

**when Density Based Spatial Clustering of Applications with Noise (DBSCAN) becomes the natural choice. DBSCAN does not require any predefined number of clusters and has only two parameters (minimum number of points in a cluster and epsilon, distance between clusters), so it is very easy to tune and very fast in performance.**

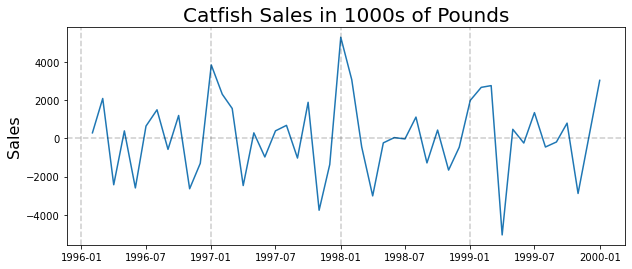
**Data Set:**

**It is Catfish Sales monthly sales Data from Jan, 1996 to Jan, 2000**

****

We can See that there is Trend As well AS Seasonality In the Data Set.

**Removing The Trend:**

****

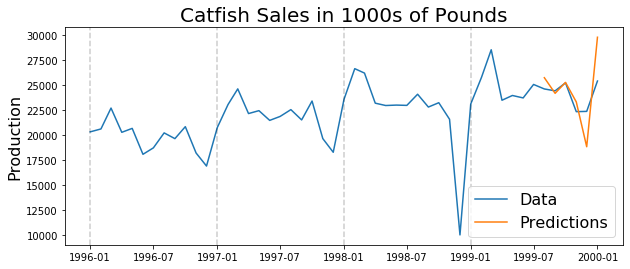
After Removing the Trend We can See Data is Center Around 0 And There is no Upward or Downward Trend. But we still have the Seasonality.

**ACF and PACF:**

**## Based on PACF, we should start with a seasonal AR process**

**## Based on ACF, we should start with a seasonal MA process, with Period 1 year**

**Prediction Without Anomaly Detection.**

****

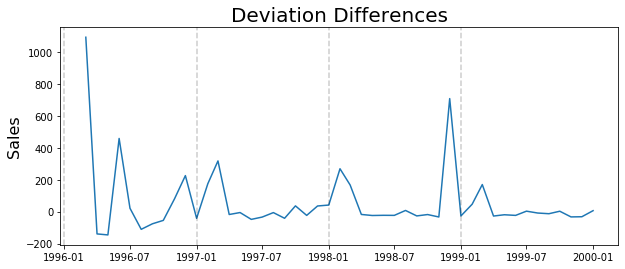
Prediction Is Highly Skewed.

First Couple of months a OK but As soon as we get to the December Time Period It Is higly Under Predicted. Because IT is Taking previour Years December Data for Prediction.

And to this loc prediction we over correct and predicted next Month Vary High. Hence ThisWill create the Problem in future Prediction.

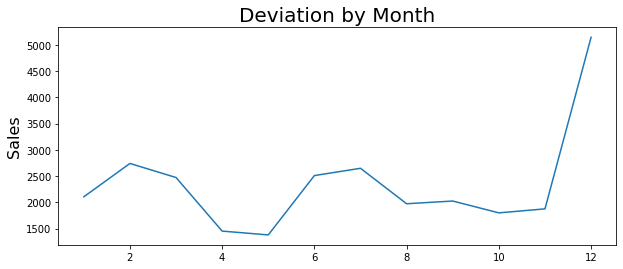
**Detecting The Anamoly:**

1. **Deviation Method:** We will calculate the Std Dev For the Each Window Size, in our Case Window Is on year and goes so on, As soon as we reach the Anomaly our Std. Dev goes Up.

****

**We need to Ingnore the High SD in Beigning as it is calculated with very less data. It is one of the Drawback as it is very difficult to use when anomaly occur very early.**

1. **Seasonal Method:**

****

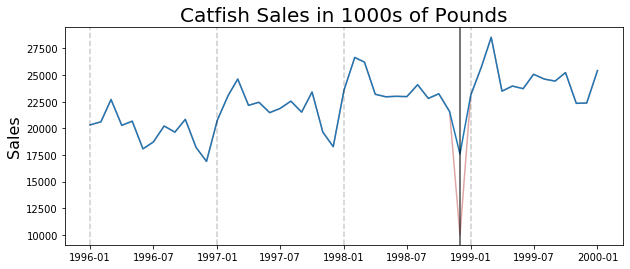
**This is used when we know when there is seasonality in our Data.**

**This chart show what is the STD dev when we only consider the data point in the given month.**

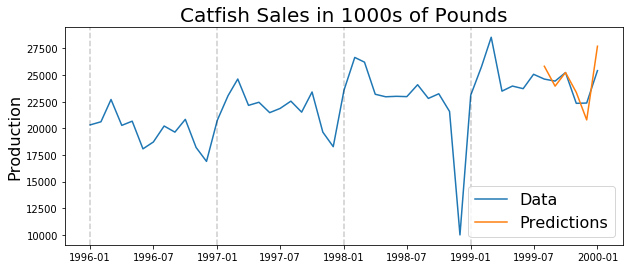
**So the Anomaly occur in the Month Of December.**

**Now in the Next Step we will find the which December has the anomaly, By collecting all December data and eliminating one December data at a time and Calculate the SD of the Other three December, And the December month which change the SD the Highest contain the Anomaly.**

**Then We will replace the Decemeber with Anomaly with the Local mean of All other remaing Decemebr.**

****

**Prediction after Detecting and Treating the Anomaly:**

****

**We find the must close Prediction as compare to earlier one.**