**Sales**

**Non-Seasonal Time Series Analysis**

**Dataset – ‘**sales.csv’

**Dataset Dates –** From 1992-01 to 2019-01

**Source of Data** – Kaggle

**Code** – Completed in RStudio

**Reading the Data**

The first thing that I did, was to import and read the data. I then summarised it and received the following information.



The data consists of two columns, the date, and the number of sales for the product. I plot the data as a time series.

Graphical user interface

Description automatically generated with medium confidence

**Stationarity**

A stationary process is when the mean and variance don’t change overtime, and when there is no trend to be seen. You can tell by looking at the plot that the data is non-stationary. We will need to make the data stationary if we want to use an arima model or if we want to forecast for future values.

To conform for stationarity, I performed an adf test.

We reject the initial adf test as the null hypothesis is not rejected. This is because the p-value is 0.4402.

Text

Description automatically generated with medium confidence

To make the data stationary, we will use the first difference.



Text

Description automatically generated

Again, we see that the p-value is too high, so the null hypothesis is rejected.

We try again with the second difference.



Text

Description automatically generated

We can accept the null hypothesis with a p-value of 0.01. This means that the data is now stationary, as can be seen in the time series plotted below.

Timeline

Description automatically generated

Now that we have a stationary model, we can conduct our pacf and acf tests to determine the inputs for the arima model.

**Arima model and Forecasting**

First, the pacf test.

The pacf series below is the model for the initial, non-stationary data.

A picture containing graphical user interface

Description automatically generated

Then after two differences we can see the change in the data.



Chart, box and whisker chart

Description automatically generated

From this model, we take 7 as our ar term.

Then to get the ma value, we use the acf test.

Chart, histogram

Description automatically generated

The series above is the original data. It is clearly not stationary from the series.

Chart

Description automatically generated

This is the acf after second differencing. From this, we take 6 as our ma term.

From the data that we have received, the arima model will use the values (7,2,6) with 2 being the second order differencing.

We can now see this being used for forecasting below.

Text

Description automatically generated with medium confidence

Graphical user interface, chart

Description automatically generated

As you can see, there has become a slight upward trend as the model has become forecasted.