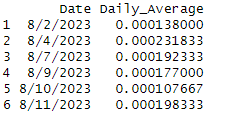
**Time Series Forecasting**

Time series forecasting plays an important role in various fields such as finance, marketing, weather prediction, etc. It allows us to take a look into what the future may look like for our predicted value. We will try to forecast the average daily engagement rates based on the previously retrieved data.

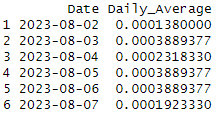
Time series forecasting in marketing analysis is valuable for predicting future consumer behavior, enabling businesses to strategically plan marketing campaigns and allocate resources effectively. By accurately anticipating trends and demand patterns over time, companies can optimize their promotional efforts, inventory management, and overall marketing strategy to align with the dynamic preferences of their target audience.

To perform the task, we will work on R, as it easily provides the statistical tools needed.

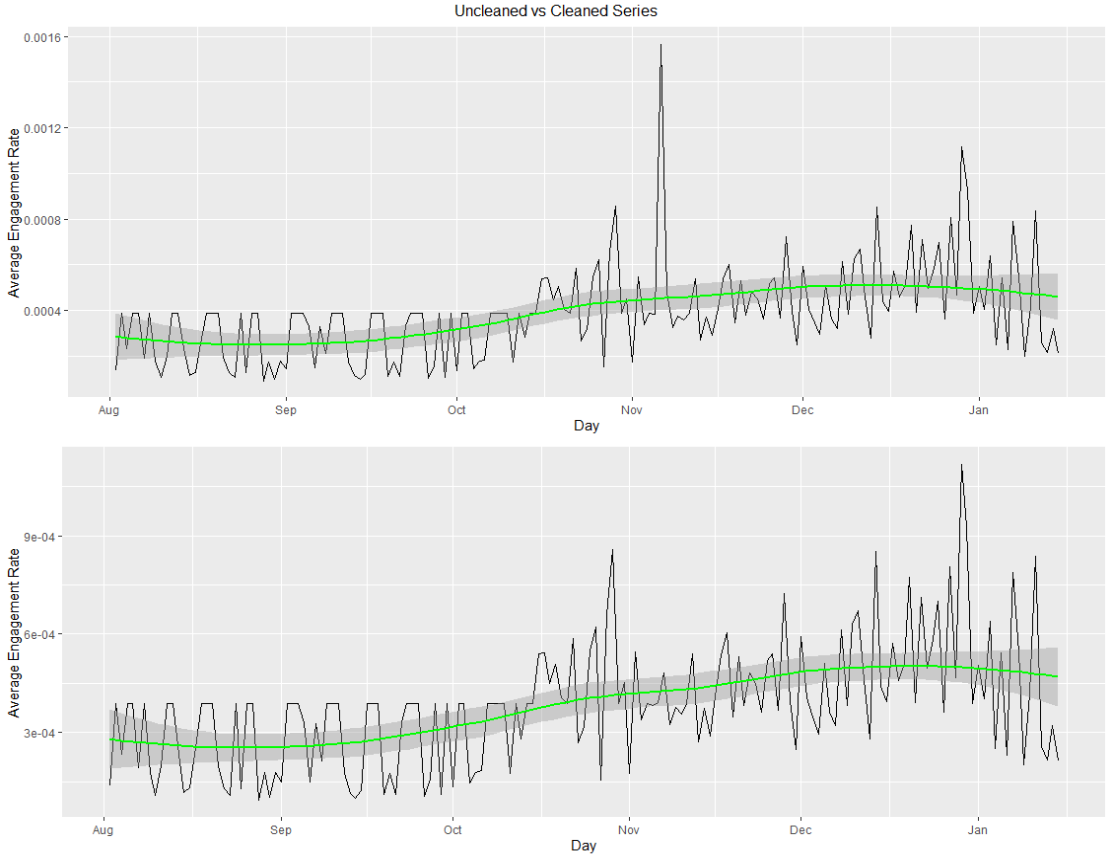
First, we print the data and we notice that we have some missing days.

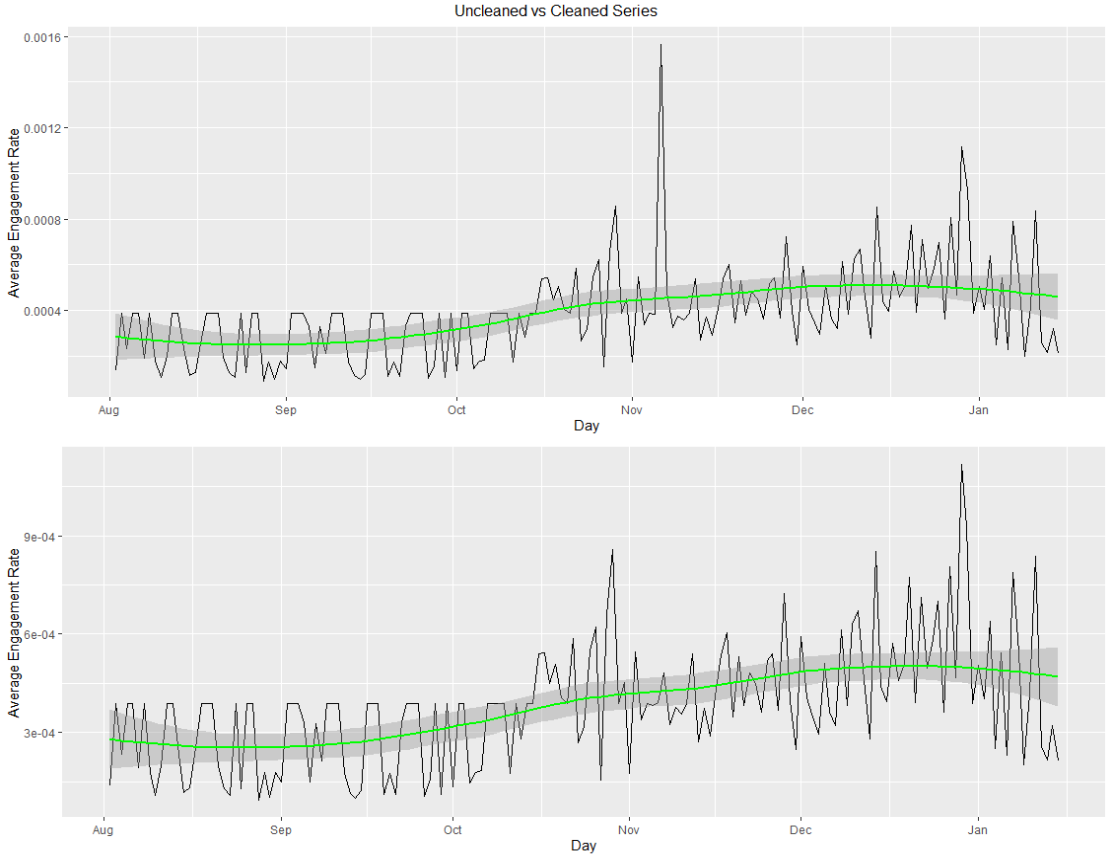


To solve this issue, we will fill the dataframe with the missing dates and we will assign them the average of the “Daily\_Average” column.

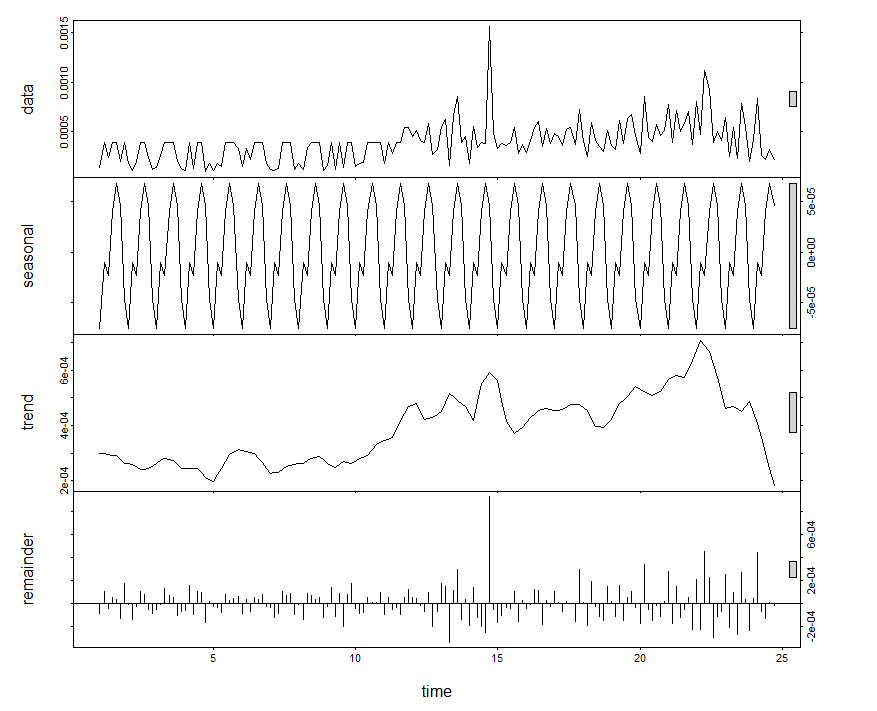


Next, we create a Time Series object to plot the evolution of the daily average engagement rate and utilize the provided “clean” function to remove any outliers that may be present. The pre-cleaning and post-cleaning plots are provided below respectively:

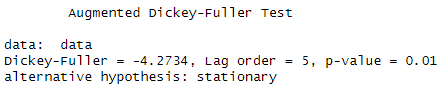




We decompose the time series object to check for any trend and seasonality in the data. The trend is ambiguous and although we can see a pattern for seasonality it may be deceiving so we will look into it further when working on the forecasting model.



One of the main assumptions of time series forecasting is the need for the data to be stationary. There are several ways to check for this: to graphically identify stationarity or to perform the Augmented Dickey-Fuller test whose H0 is that the data is not stationary. The test results in a p-value < 0.05, successfully rejecting the null hypothesis.

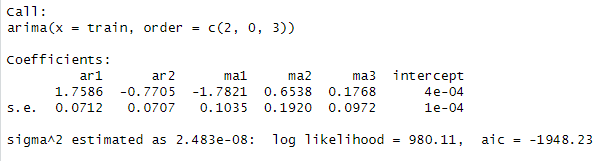


We split the data into training and testing sets as follows: 134 and 33 observations respectively.

**Auto Arima Model:**

An ARIMA (Auto Regressive Integrated Moving Average) model is a time series forecasting method that combines autoregressive (AR) and moving average (MA) components, along with. It is widely used for modeling and predicting future values based on past observations, providing a flexible and powerful tool for capturing temporal patterns in data.

We have several ways of finding the best fitted ARIMA (Auto Regressive Integrate Moving Average) model. We can either go with the classical way of reading the ACF and PACF plots, but it could be time consuming and inaccurate, or we can use the Auto-Arima function to try several combinations and return the best one based on the AIC metric. It is also not very accurate, so we can use it as a guiding point from which we can try different combinations, the best retrieved model is the following:

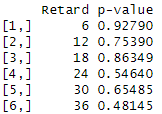
****

The best fitted-model found is the ARIMA (2,0,3) with no moving differencing components since the data is already stationary. The t-test shows that the components are all significant:

**t-statistics:**

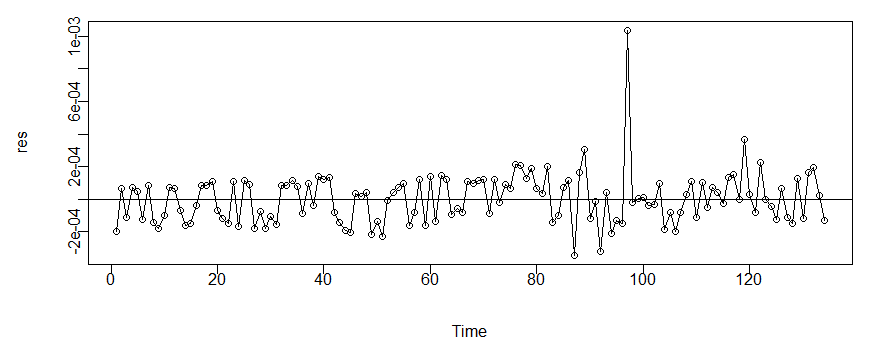
****

**Ljung-Box test:**

****

The null hypothesis is not rejected, there is no autocorrelation in the residuals.

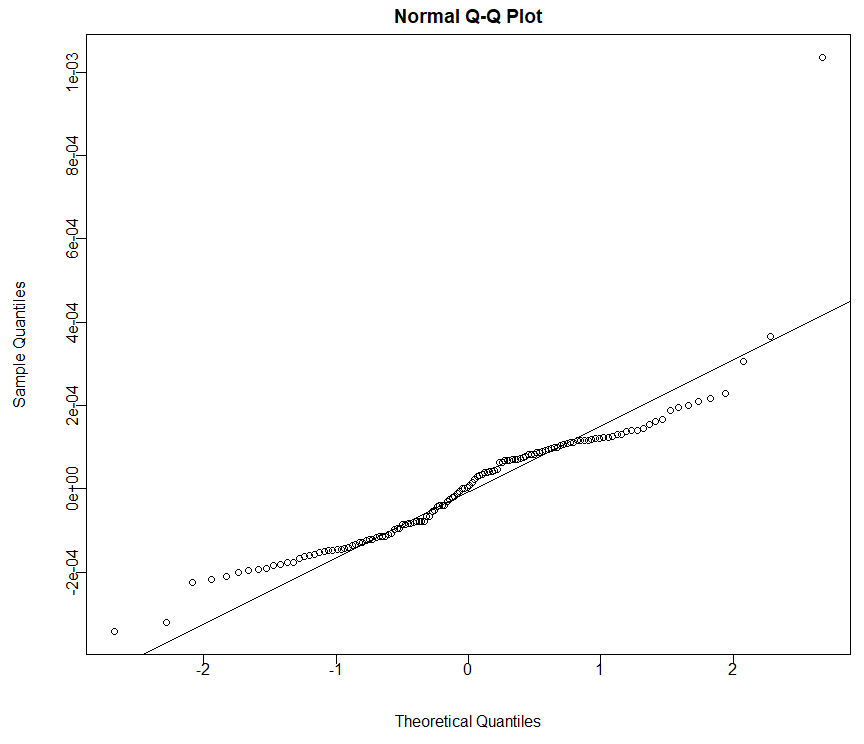
**Heteroskedasticity test:**

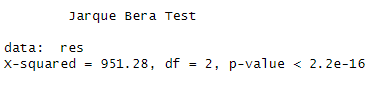
****

The plot shows that there is no heteroskedasticity in the data.

**Normality:**

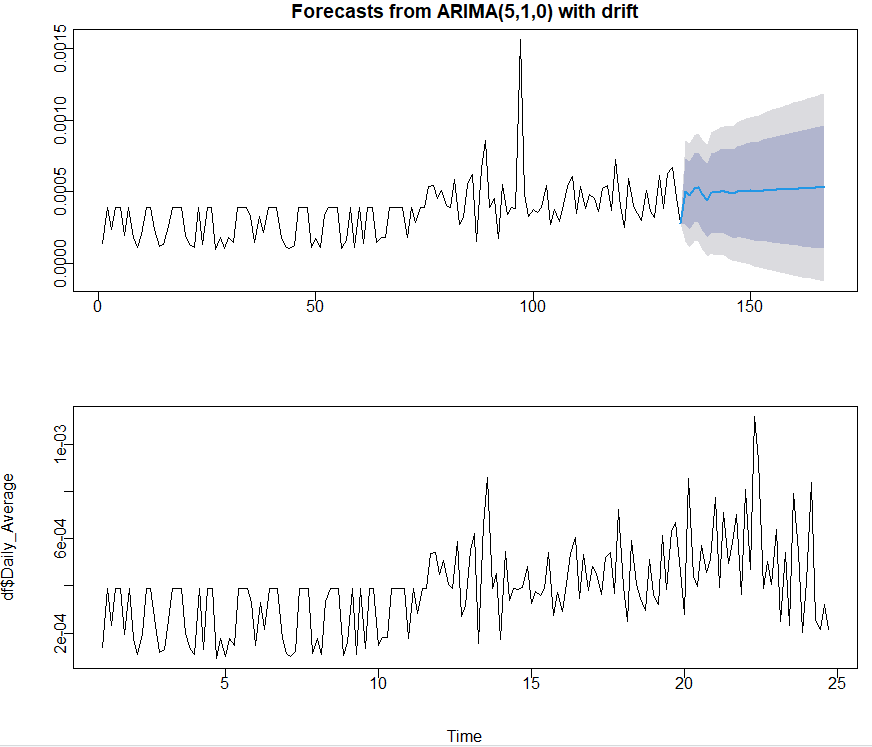
To check for normality, we perform two tests. The first is the QQ-plot to check graphically and the second is the Jarque-Bera Test to check for normality statistically. Both tests show that the distribution of the residuals is normal.

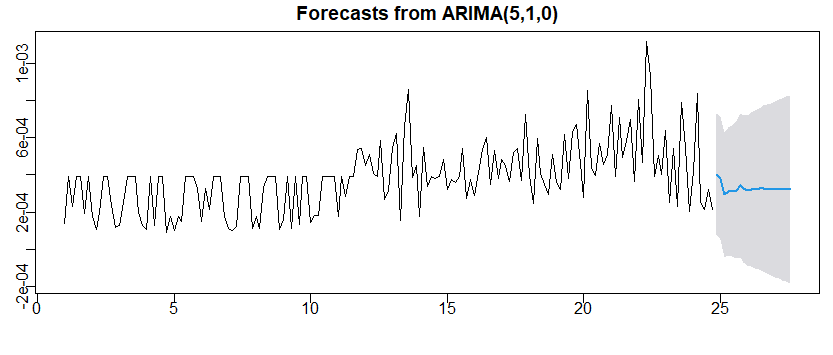




**Prediction:**

We will use the resulting model to predict the data of the testing set, then forecast it into the future. The blue line is the forecasted data and the dark grey area is the 95% confidence interval.



**Forecasting  
**

The forecasting 95% confidence interval is quite wide, this is due to the data itself and the learning of the model on said data. It gives an overall idea about how the daily average engagement rate will move in the upcoming weeks.