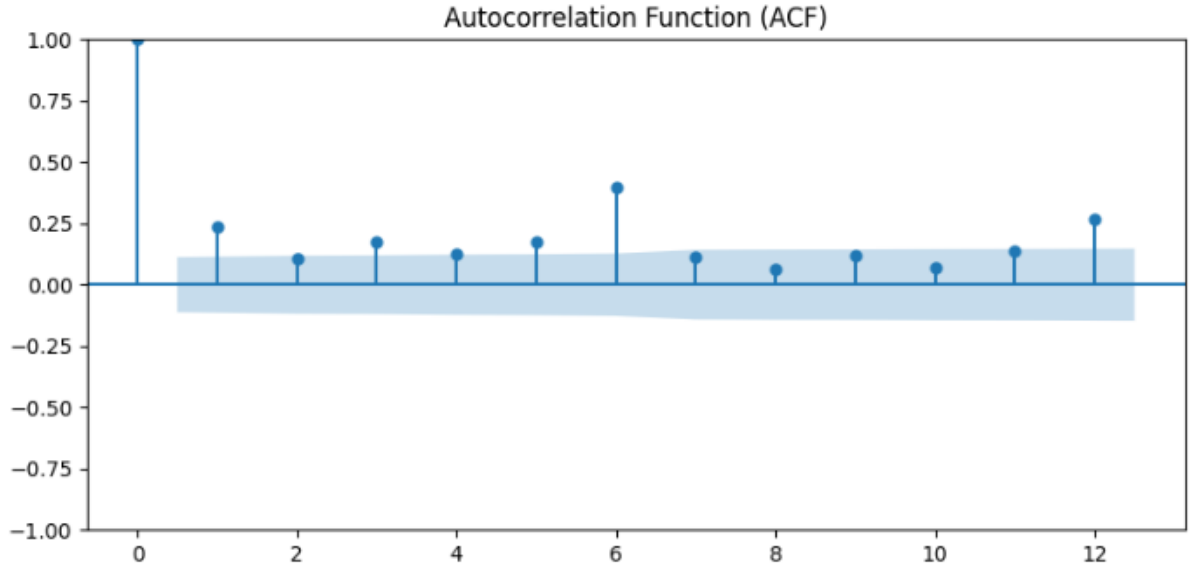
Time Series Analysis

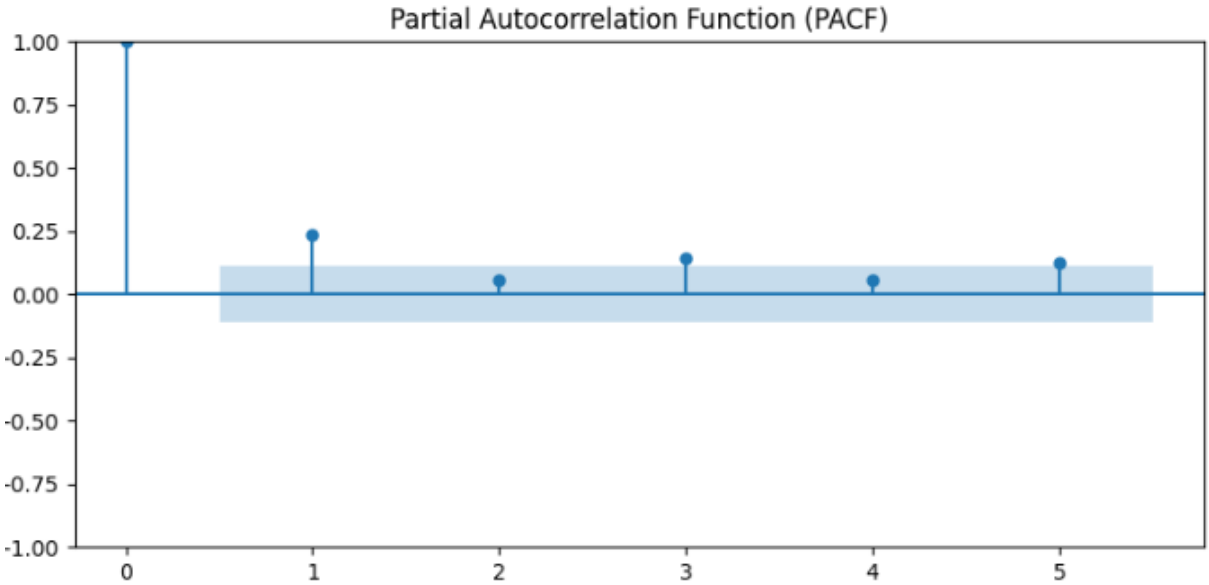
This report is regarding the E-Commerce Customer Data Analysis project.

Time Series Model

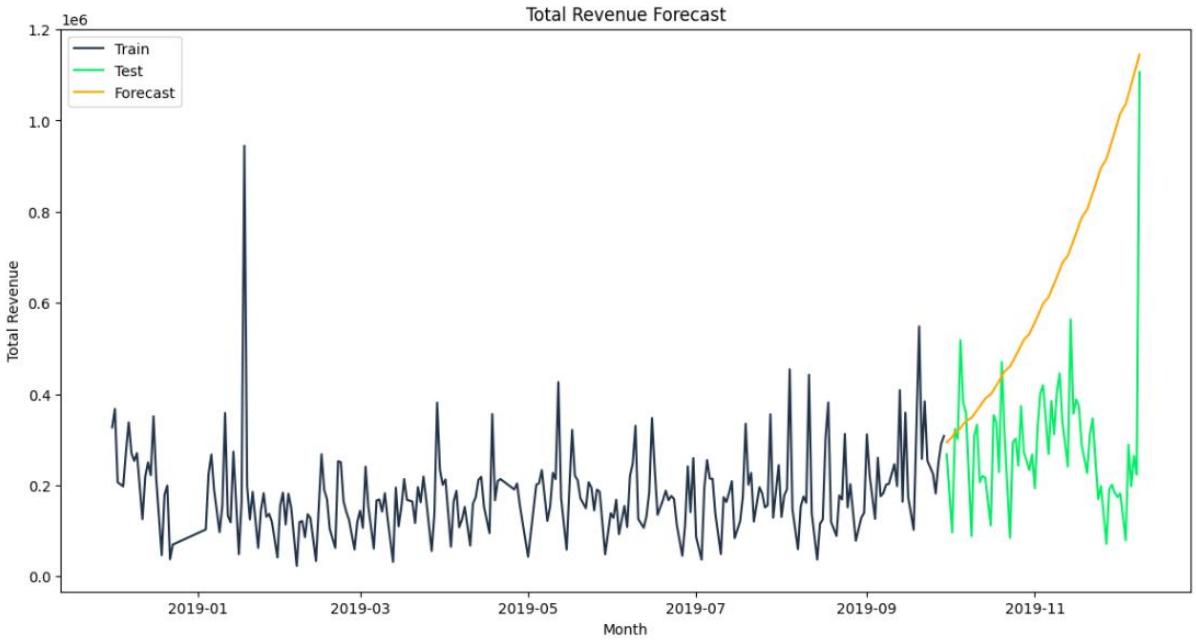
The model chosen for analysis of the given dataset was the Auto Regressive Incremental Moving Average model, or ARIMA Model. This model was chosen as the data does not have clear trends or seasonality because of the limited data we have.

The ARIMA model uses three values (p, d, q) as an order of lags that can be used to modify and fit the model as we require. To determine p, d and q, we graph the Auto Correlation Function and Partial Auto Correlation Functions





Using these graphs we can determine the values of p, d and q, by training the model



The most optimal value of the order is (2, 3, 2). The evaluation metrics used were Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and Root Mean Square Error (RMSE)

AIC: 6540.191279992015
BIC: 6578.432379063262
RMSE: 29936919.3844

This model and forecast tells us hat there is potential seasonality every year in the winter months where sales increases.

Business Recommendations

It is recommended that the website attempt to capitalise on the seasonality by putting end-of-year sales and winter sales on the website to drive sales higher.