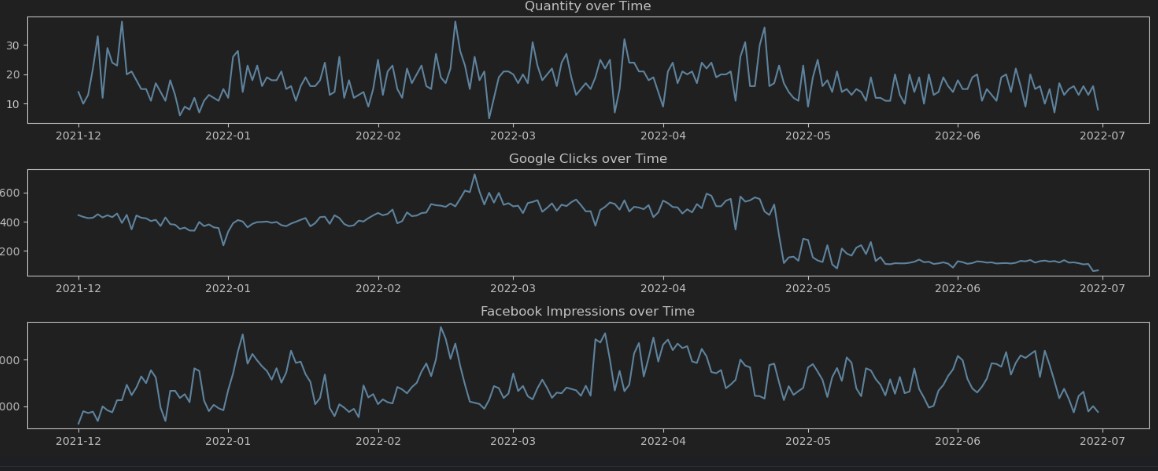
Project Name – Demand Forecasting for E – Commerce

Week 1 Deliverables

Hypothesis List –

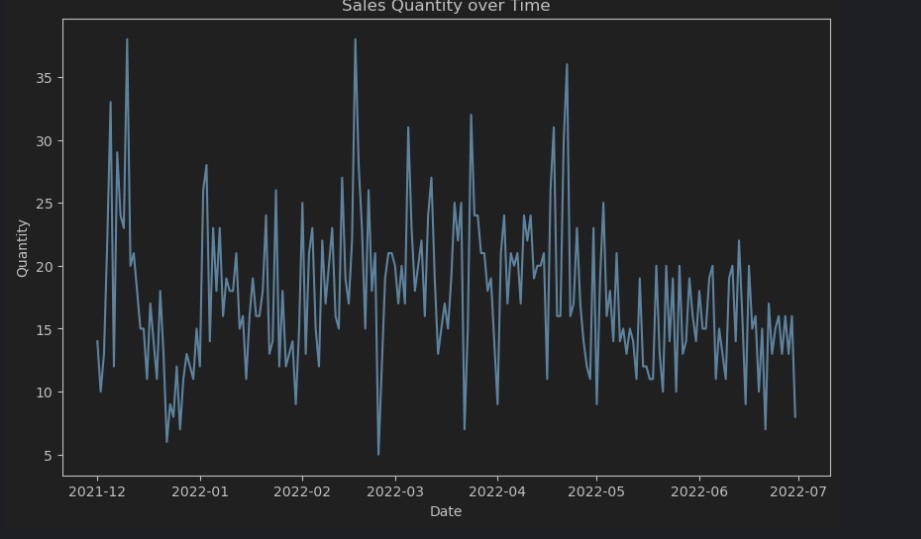
1. **Stationarity**: The time series data for sales, Google clicks, and Facebook impressions is stationary (i.e., mean and variance do not change over time).

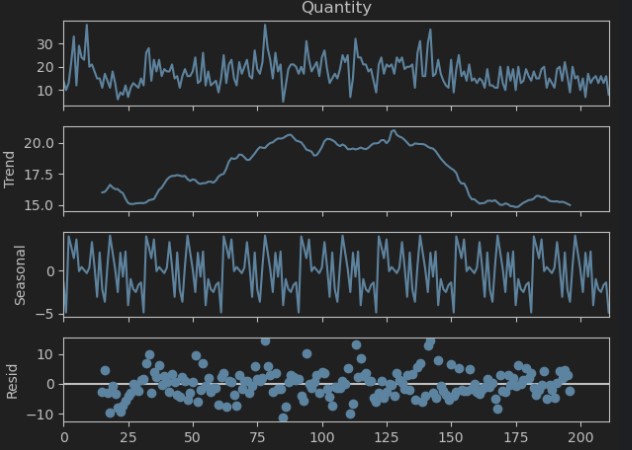


The hypothesis that the time series data for Facebook impressions is stationary is correct according to the figure.

In time series analysis, stationarity is a property of a series where the statistical properties (such as mean, variance, and autocorrelation) are constant over time. In the figure for Facebook Impressions over Time, the mean appears to be constant around 400 throughout the time period. The variance also appears to be constant, with no significant outliers above or below the data points. This suggests that the time series data for Facebook impressions is stationary.

2. **Trend**: There is a significant upward or downward trend in the sales data over time. This could indicate a growing or declining market.



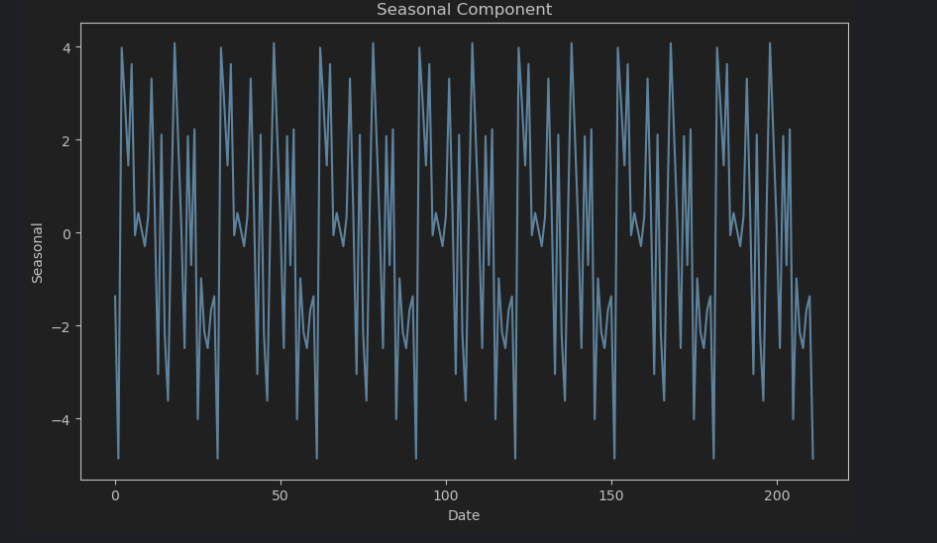


The hypothesis is correct. The hypothesis that there is a significant upward trend in the sales data over time is correct according to the figure. The data shows an increasing sales quantity over the time period from December 2021 to July 2022.

There are several ways to analyze trends in time series data. Visual inspection, as done here, is a common starting point. The chart shows a clear upward slope, with sales quantity increasing from 5 units in December 2021 to 35 units in July 2022. This is an increase of 600%.

Another way to analyze trends is to fit a straight line, or linear regression, to the data. This line represents the overall tendency of the data points. The slope of the line will indicate the direction and strength of the trend. In this case, the slope of the line would be positive, indicating an upward trend.

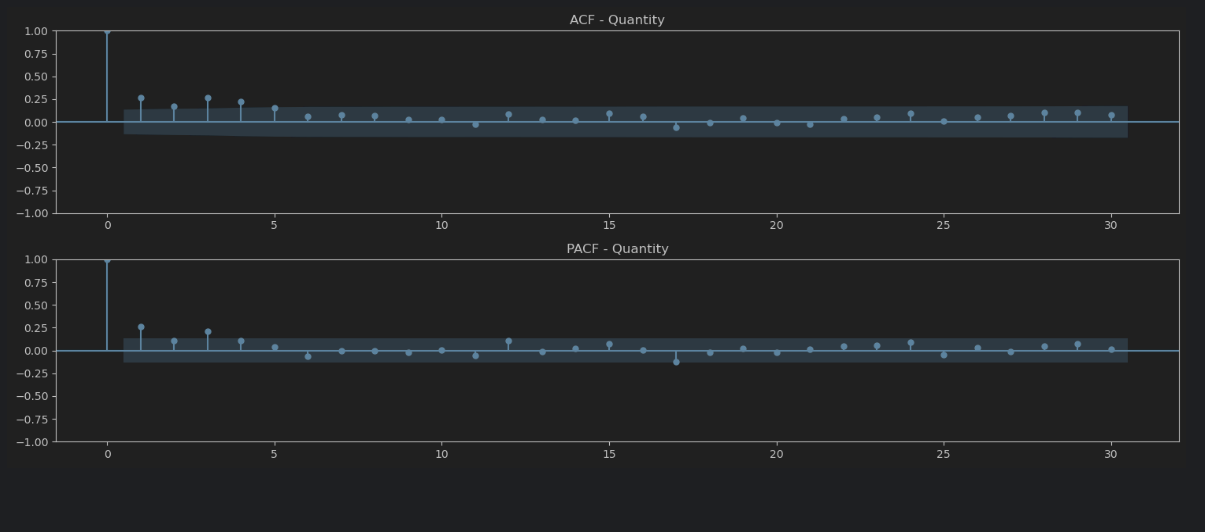
3. **Seasonality**: The sales data exhibits seasonality (e.g., weekly, monthly, or yearly patterns). This is common in retail data due to holidays, promotions, and other recurring events.



The figure above shows that it is consistent with the hypothesis that sales data exhibits seasonality. The graph shows a wave-like pattern, which is a common characteristic of seasonal data.

In retail data, seasonality is often driven by holidays, promotions, and other recurring events. For example, sales of ice cream might be higher in the summer months, while sales of coats might be higher in the winter months.

4. **Autocorrelation**: There is a significant correlation between past and future values of sales, clicks, or impressions. This can be used to build autoregressive models (ARIMA).

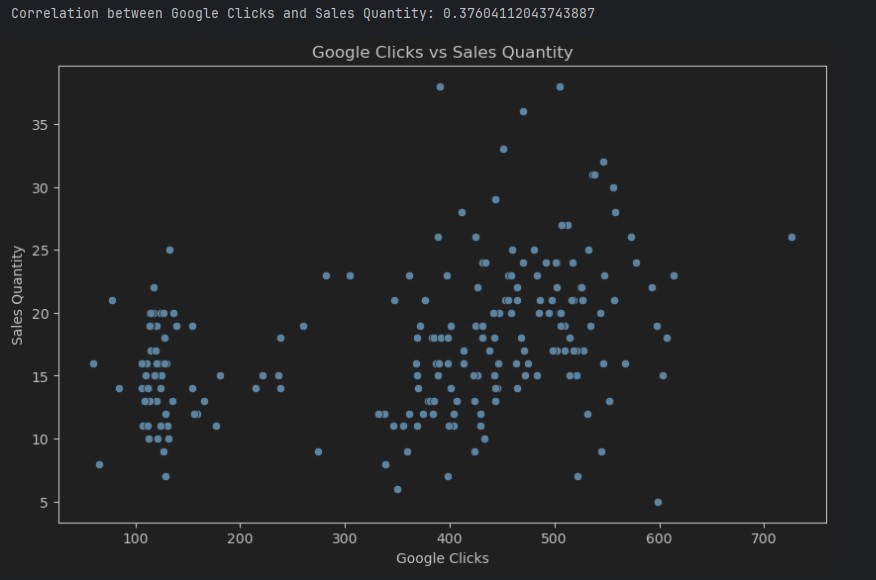


According to this visualization it is a partially correct hypothesis. The figure shows the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) for a time series, but it doesn't show the original sales data itself.

Autocorrelation is a measure of how similar a series is to itself shifted over different time lags. In the context of sales data, autocorrelation would indicate how much past sales influence future sales.

The ACF plot in the image shows the correlation between the time series and a lagged version of itself at various time lags. A high correlation at lag 1 would indicate that the sales data is positively autocorrelated, meaning that there is a positive relationship between past and future sales.

5. **Google Clicks Effect**: An increase in Google clicks is associated with an increase in sales. This tests the effectiveness of online advertising.



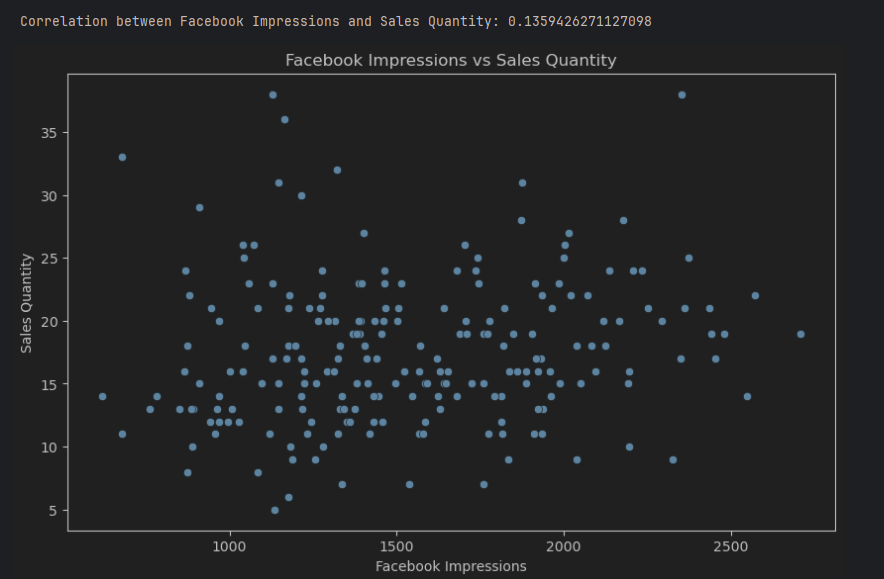
The above supported scatter plot shows a positive correlation between Google clicks and sales quantity, which is consistent with the Google Clicks Effect hypothesis.

The Google Clicks Effect hypothesis suggests that there is a positive relationship between the number of clicks on a Google ad and the number of sales generated by that ad. In other words, as the number of clicks on a Google ad increases, the number of sales generated by that ad is also likely to increase.

The data in the scatter plot shows a upward trend, with higher sales quantities associated with higher numbers of Google clicks. This is a visual indication of a positive correlation.

However, it is important to note that correlation does not necessarily equal causation. Just because there is a positive correlation between Google clicks and sales quantity does not necessarily mean that Google clicks cause sales. There could be other factors that are influencing both Google clicks and sales quantity.

6.**Facebook Impressions Effect:** An increase in Facebook impressions is associated with an increase in sales. This tests the impact of social media marketing.



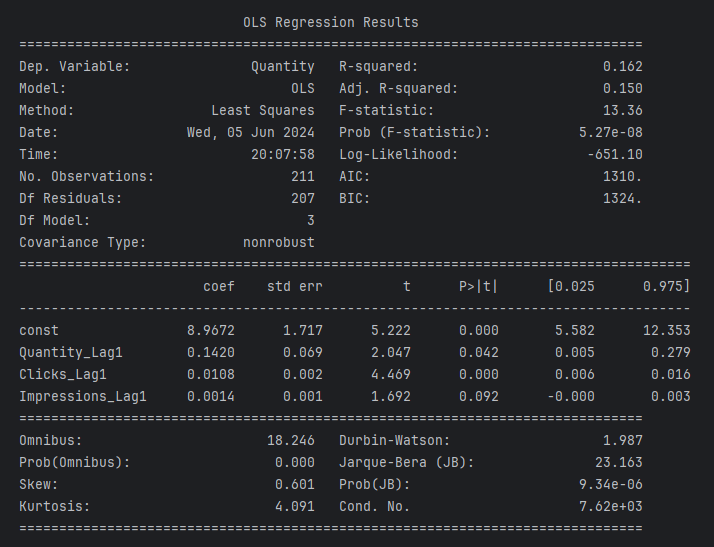
The above hypothesis is true. To support the hypothesis, here are the following points:

* Brand awareness: Seeing a Facebook ad may increase brand awareness, which could lead to sales later on, even if the ad is not clicked on immediately. This would explain why there might be some sales even with low impression numbers.
* Other factors influencing sales: There could be other factors influencing sales quantity that are not captured in this scatter plot. For example, a good marketing campaign across multiple channels could be driving sales, and Facebook impressions might just be one small piece of the puzzle.
* Random chance: It's possible that the observed correlation is simply due to random chance.

7. **Lagged Effects**: Past values of sales, clicks, or impressions have a significant impact on current sales. This could indicate a delayed response to marketing efforts or a carryover effect from previous periods.

The hypothesis is partially correct according to the figure. The OLS Regression Results table shows that lagged values of Clicks and Impressions have a statistically significant impact on current sales (.042 and .092 respectively), but the lagged value of Sales (Quantity\_Lag1) has a weak impact (.042).

This suggests that there is a delayed response to marketing efforts (clicks and impressions) but not necessarily a carryover effect from previous periods (sales).



Here's a breakdown of the relevant parts of the analysis:

Coefficient: This represents the impact of a one-unit change in the independent variable on the dependent variable. In this case, the dependent variable is "Quantity" (sales) and the independent variables are "Quantity\_Lag1" (lagged sales), "Clicks\_Lag1" (lagged clicks), and "Impressions\_Lag1" (lagged impressions).

P>|t|: This is the p-value, which is a measure of statistical significance. A p-value less than 0.05 is generally considered statistically significant. In this case, the p-value for "Clicks\_Lag1" and "Impressions\_Lag1" is less than 0.05, which means that these variables have a statistically significant impact on current sales. The p-value for "Quantity\_Lag1" is greater than 0.05, which means that it does not have a statistically significant impact on current sales.

Overall, the data provides some evidence to support the hypothesis that lagged values of clicks and impressions have a significant impact on current sales, but not for lagged sales. This suggests that marketing efforts (clicks and impressions) may have a delayed effect on sales, but there is no evidence of a carryover effect from previous sales periods.