



{Code Clause};

CODE CLAUSE

Chris Aryan

Data Science Intern



TASK

AIM :

To develop a time series forecasting model to predict the demand for products in a retail store.

DESCRIPTION :

Using historical sales data to forecast future demand, considering seasonality and trends.



TASK

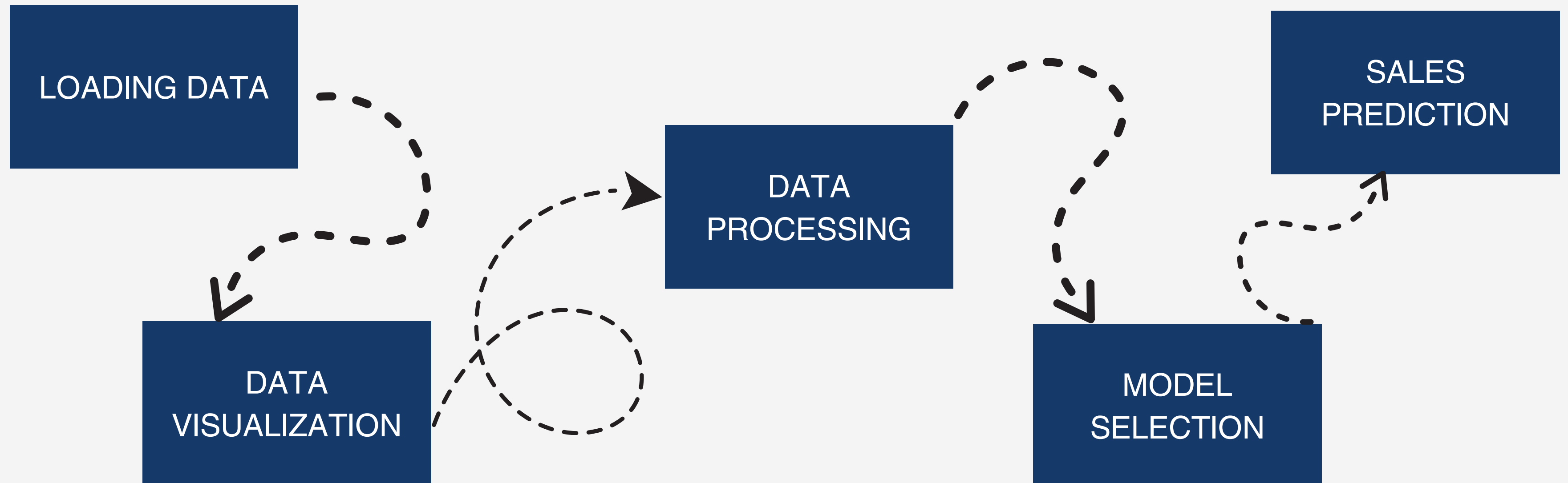
TECHNOLOGIES USED :

Python, Pandas, Numpy, Matplotlib, Statsmodels, Augmented Dickey-Fuller test, autocorrelation, SARIMA Model.

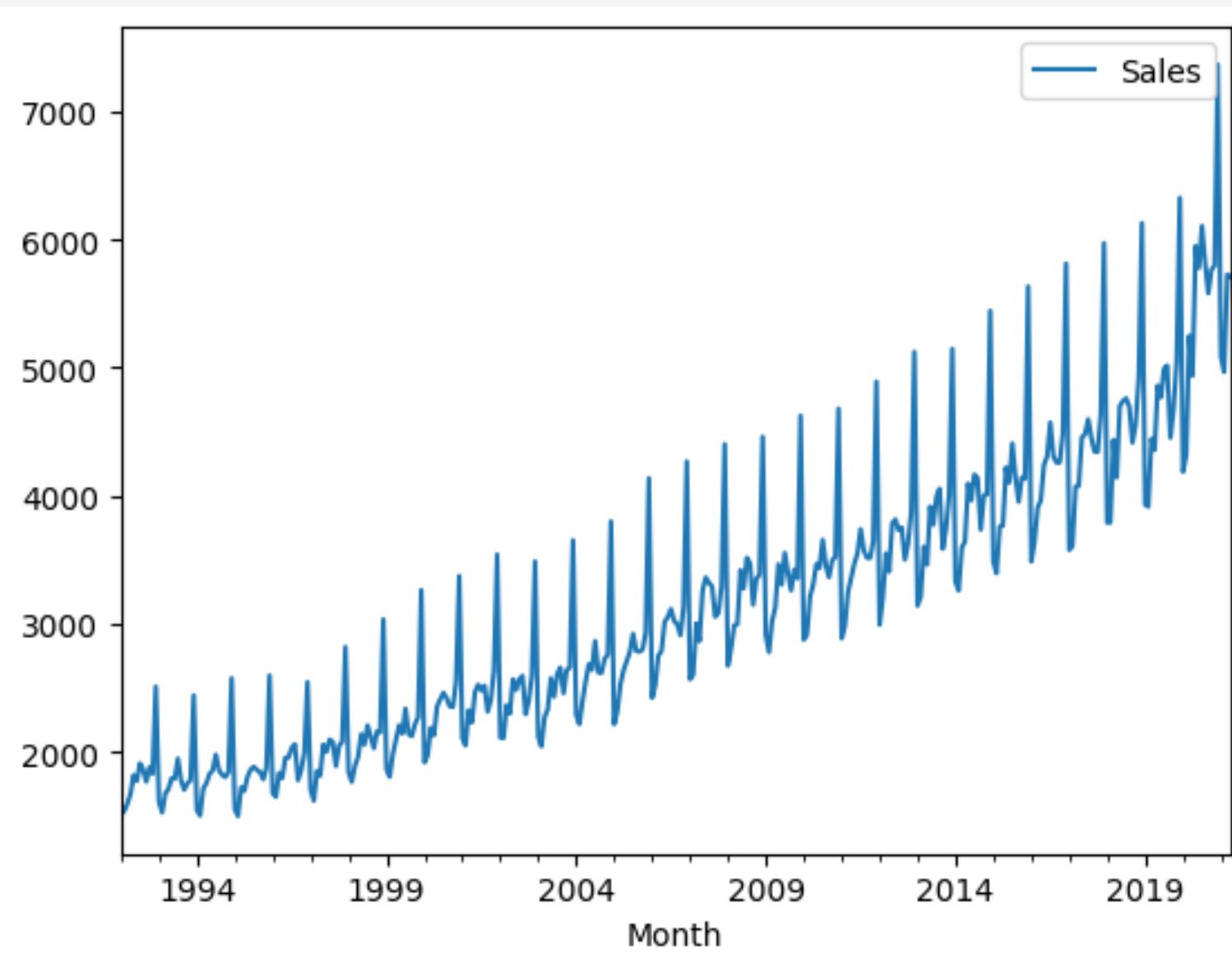
DATASET :

The dataset is taken from Kaggle. Please find the link below [dataset](#).

FLOWCHART



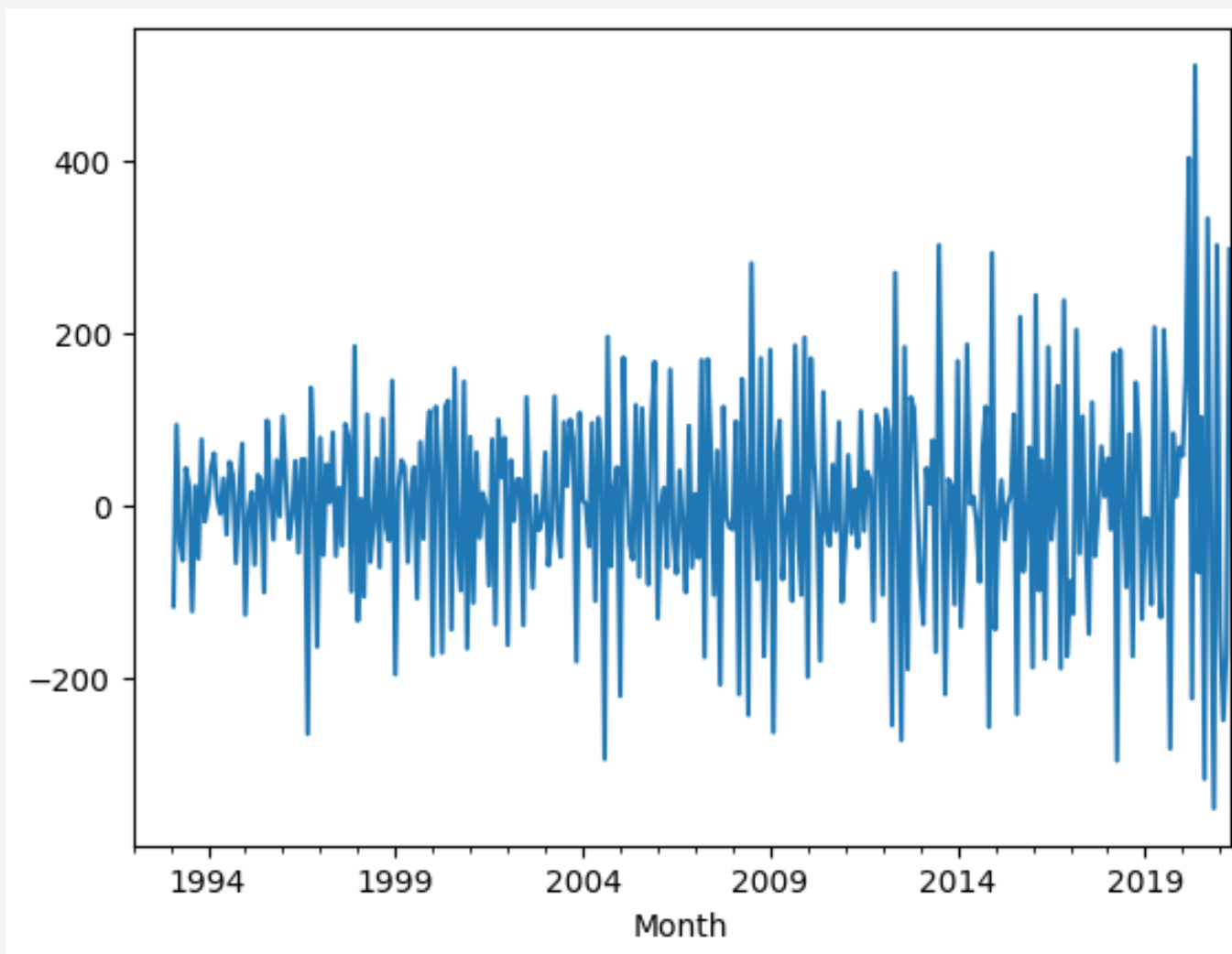
DATA VISUALIZATION



- From the graph, we observe that the data is non stationary.
- Applying Dickey-Fuller test gives us a p-value of 0.99 which is much greater than the threshold value of 0.5, thus proving that the data is non stationary

DATA PROCESSING

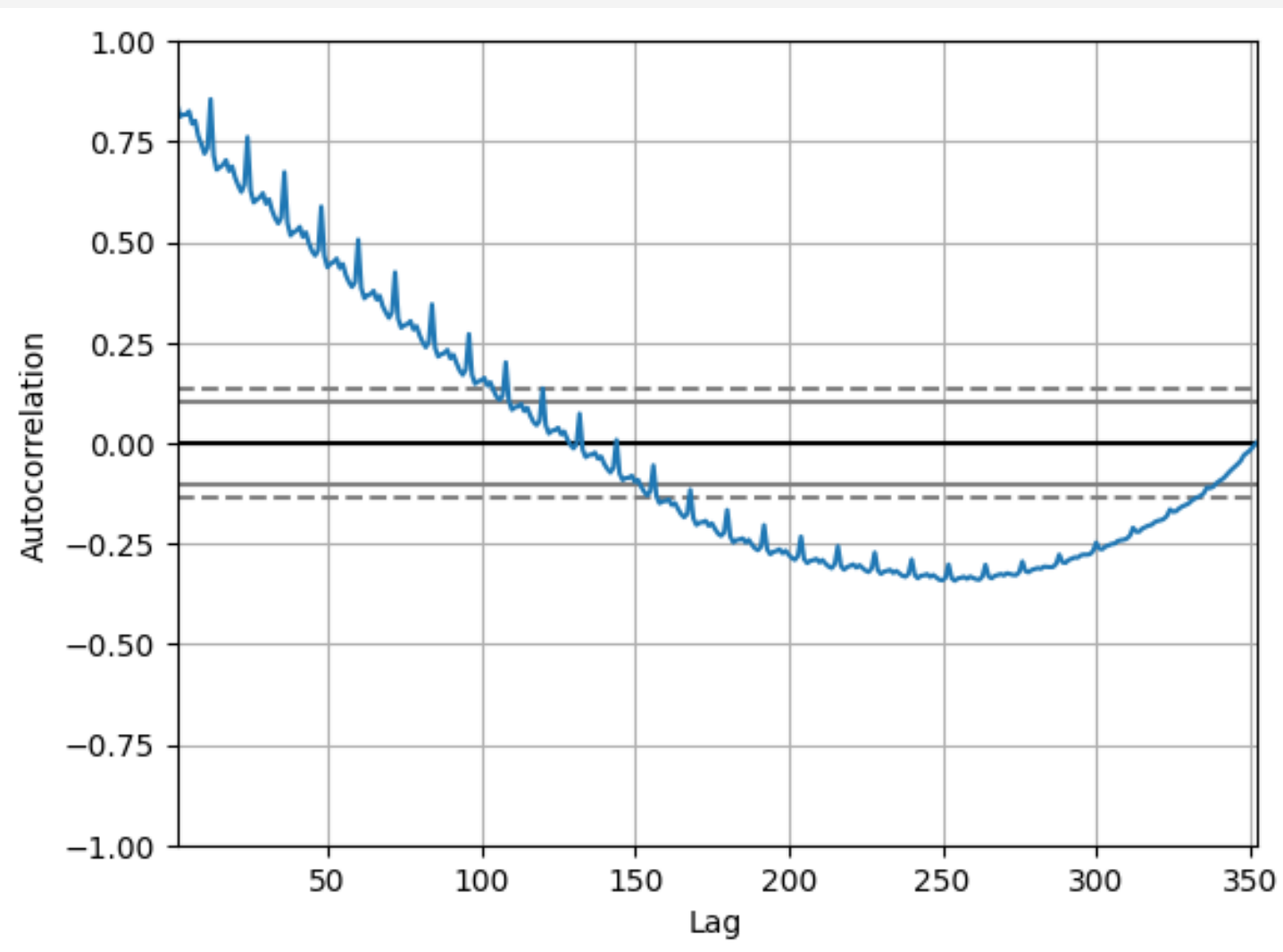
- Applying differencing, in order to make the data stationary.



- After applying differencing, we can observe the stationary behavior of the data.
- Further, applying Dickey-Fuller test, gives us a p-value of $1.14e^{-05}$.

DATA ANALYSIS

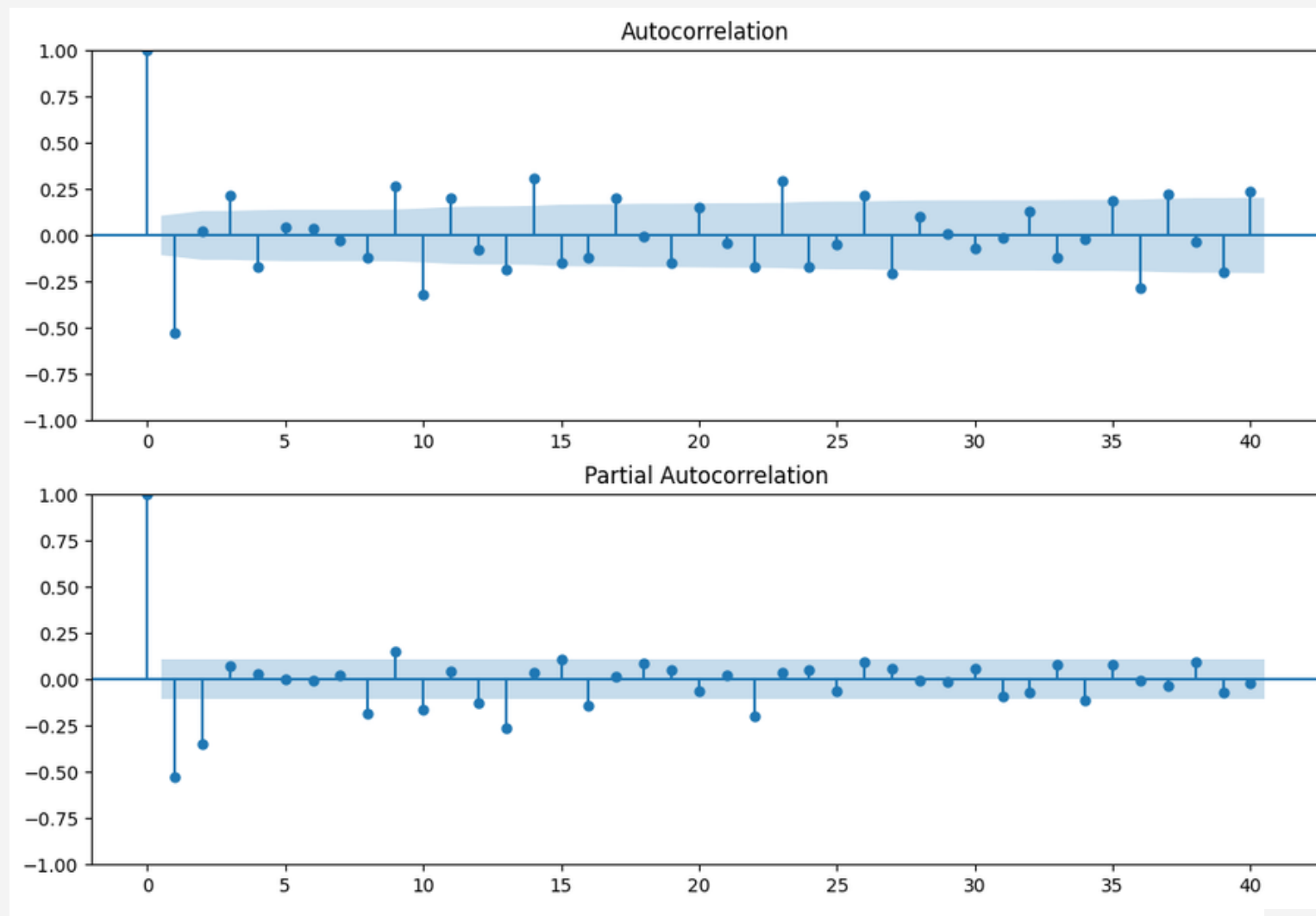
- Plotting autocorrelation v/s lags.



- The high positive correlation suggests strong correlation between each observation and its nearby lagged values.
- The gradual decrease shows that the strength of the correlation diminishes as the lag increases. This behavior is due to diminishing seasonality.
- The convergence to 0 suggests that the trend or seasonality effects have diminished, and the data is becoming more stationary.

DATA ANALYSIS

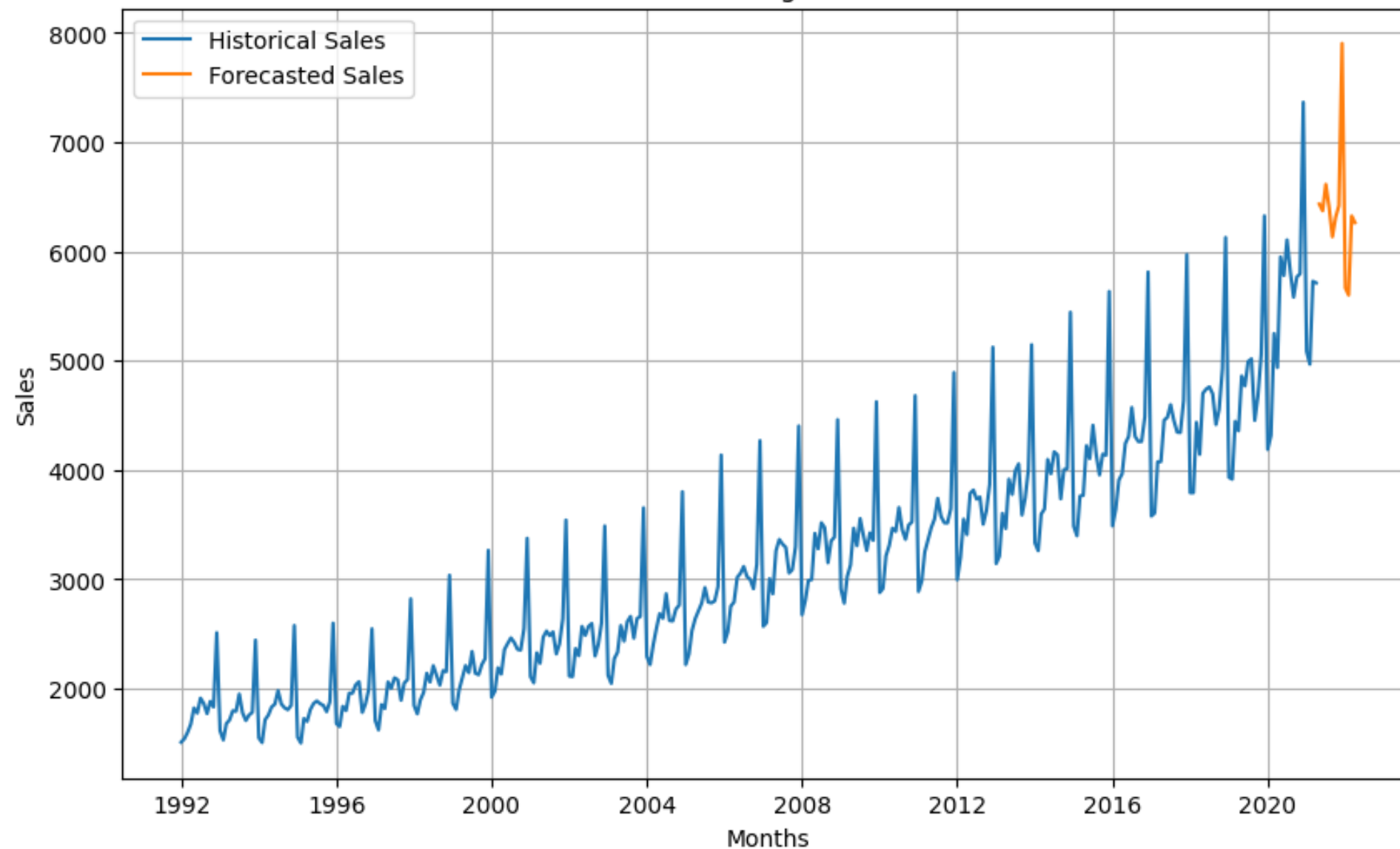
- Plotting autocorrelation and partial autocorrelation v/s lags.



- The constant but non-zero autocorrelation suggests that there may be weak or insignificant correlation in the processed data.
- We observe that the partial autocorrelation plot is almost constant near to zero. This signifies that there is no significant autocorrelation between the observations at different lags and the data is stationary..

SALES FORECASTING

Sales Forecasting with SARIMA



- We train the historical sales data in our model.
- Forecasting sales for the next 12 months.
- We can conclude that the sales in the period of 01-04-2021 to 01-04-2022 is in the range of 5500 to 8000 approximately.
- We can conclude that the sales in the period of 01-04-2021 to 01-04-2022 is in the range of 5500 to 8000 approximately.



{Code Clause};

THANKYOU

Chris Aryan

Data Science Intern