This report analyzes and forecasts retail sales data using a combination of ARIMA and LSTM models. The analysis involves data preprocessing, model fitting, evaluation, and extended forecasting.

1. Data Loading and Preprocessing:

. The dataset, 'retail_sales.csv', was loaded into a pandas DataFrame.
. The 'Date' column was converted to datetime objects for proper time series analysis.
. The 'Date' column was set as the index.
. Missing values were filled using the forward fill method ('ffill').
. The data was resampled to a monthly frequency ('M') and the sales values were summed for each month.

2. Data Exploration:

. The first few rows of the dataset were displayed using `df.head()`.
. Descriptive statistics of the data were calculated using `df.describe()`.
. The presence of missing values was checked using `df.isnull().sum()`.
. The time series data was plotted using `df.plot()` to visualize trends and seasonality.

3. Stationarity Test (ADF Test):

. The Augmented Dickey.Fuller (ADF) test was performed on the 'Sales' column to determine if the time series data was stationary.
. The ADF test aims to detect the presence of a unit root, which indicates non.stationarity.
. The p.value from the test was used to determine stationarity. If the p.value is less than 0.05, the data is considered stationary; otherwise, it is non.stationary.

4. Autocorrelation and Partial Autocorrelation (ACF/PACF) Plots:

. Autocorrelation and partial autocorrelation plots were generated using `autocorrelation_plot` and `plot_acf`, `plot_pacf` from `statsmodels.graphics.tsaplots`.
. These plots help identify the order of the ARIMA model based on the patterns observed in the correlations.

5. ARIMA Model Fitting:

. A SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous regressors) model was fitted to the 'Sales' data using the `SARIMAX` function from `statsmodels`.
. The model was configured with specific orders for the AR, I, MA, seasonal AR, I, MA, and seasonal period components. These orders were determined based on the ACF/PACF plots.
. The model was fitted using the `fit()` method, and its summary was printed.

6. LSTM Model:

. Data Scaling: The data was scaled to the range of 0 to 1 using `MinMaxScaler` to improve the performance of the LSTM model.
. Data Preparation for LSTM: A dataset suitable for the LSTM model was created using a windowing approach, where past values were used as input and future values as target.
. LSTM Model Building: An LSTM model was built using the `Sequential` model from Keras. The model consisted of multiple LSTM layers and dense layers.
. Model Training: The model was trained on the training data using the `fit()` method.
. Model Evaluation: The model's performance was evaluated on both the training and testing sets using metrics like RMSE, MAE, and R^2.

7. Model Evaluation:

. The performance of both the ARIMA and LSTM models was evaluated using the following metrics:
    . RMSE (Root Mean Squared Error): Measures the average difference between the predicted and actual values.
    . MAE (Mean Absolute Error): Measures the average absolute difference between the predicted and actual values.
    . R^2 (Coefficient of Determination): Represents the proportion of variance explained by the model.

8. Forecasting:

. The fitted SARIMAX model was used to generate forecasts for future periods.
. Predictions were made for a specific range of dates, and the `dynamic=True` argument was used to ensure that predictions were dynamically updated using previous predictions.

. The forecasts were plotted against the actual sales data to visualize the model's performance.

9. Extended Forecasting:

. The forecast was extended beyond the original dataset by creating a future DataFrame.
. Predictions were made for the entire extended dataset, including future dates.
. The extended forecast was plotted along with the actual sales data.

Conclusion:

This report demonstrated the use of ARIMA and LSTM models for time series forecasting of retail sales data. Both models were able to predict future sales with reasonable accuracy. The choice of the best model depends on the specific dataset and the desired performance metrics. The SARIMAX model is generally suitable for time series data with seasonality, while LSTM models excel at capturing complex non.linear patterns. Further exploration of different model configurations and hyperparameter tuning can improve the forecasting accuracy.