Q3. Time Series Analysis for Sales Forecasting

Can we develop a reliable time series model to forecast future sales volumes and identify seasonal patterns in the Brazilian e-commerce market?

Introduction

For this analysis, we pick olist_orders_dataset and olist_order_items_dataset, using the order timestamps and prices to develop a daily sales time series.

This question interests us because accurate sales forecasting is critical for inventory planning, resource allocation, and customer satisfaction in e-commerce. By identifying seasonal patterns and building predictive models, companies can optimize operations and reduce costs.

Approach / Methodology

The dataset is filtered to include only 'delivered' orders and aggregated daily order counts. This provides a clean and regular time series with which to work. SARIMA modeling is then applied to handle trend and seasonality followed by hyperparameter tuning using AIC (Akaike Information Criterion) to get the best configuration. SARIMA is then compared with Prophet (Facebook Prophet, a modern time series forecasting tool designed to handle seasonality and holidays), while also including the Seasonal Naïve Model as a benchmark, which repeats the sales pattern from the previous week, to provide a baseline for evaluation. The models were evaluated using MAE, RMSE, and MAPE. Ljung-Box test was carried out to check for residuals for autocorrelation.

Data Analysis and Results

Residual Diagnostic:

Apply the Ljung-Box test to the residuals to check for remaining autocorrelation.

```
residuals = final_results.resid
ljung_box_pvalue = acorr_ljungbox(residuals, lags=[10],␣
↪return_df=True)['lb_pvalue'].values[0]
print(f"Ljung-Box p-value: {ljung_box_pvalue:.4f}")
```

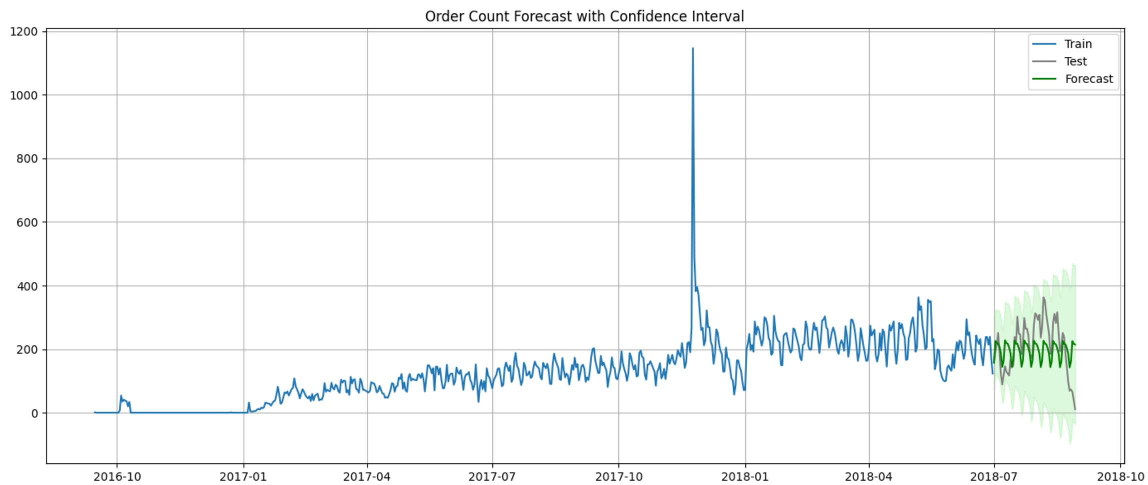
```
Ljung-Box p-value: 0.9130
```

This value indicates, there is no significant auto correlation in residuals.

Sarima Forecast:

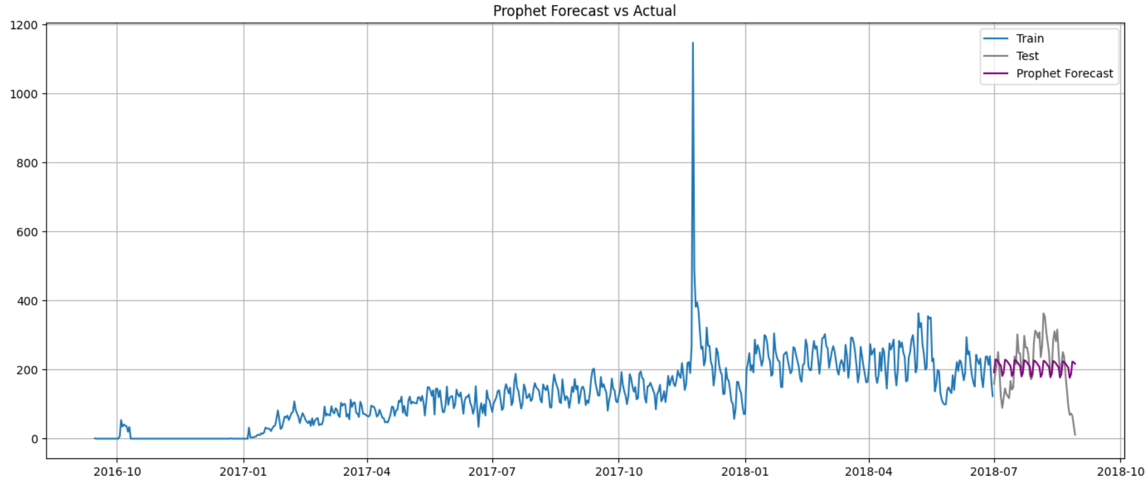
Visualize the SARIMA forecast alongside the actual sales data.

```
plt.figure(figsize=(14, 6))
plt.plot(train.index, train['order_count'], label='Train')
plt.plot(test.index, test['order_count'], label='Test', color='gray')
plt.plot(predicted.index, predicted, label='Forecast', color='green')
plt.fill_between(conf_int.index, conf_int.iloc[:, 0], conf_int.iloc[:, 1],
    color='lightgreen', alpha=0.3)
plt.title('Order Count Forecast with Confidence Interval')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



Prophet Forecast vs Actual:

```
plt.figure(figsize=(14,6))
plt.plot(train.index, train['order_count'], label='Train')
plt.plot(test.index, test['order_count'], label='Test', color='gray')
plt.plot(forecast_test.index, forecast_test['yhat'], label='Prophet Forecast',
    color='purple')
plt.title('Prophet Forecast vs Actual')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



Model	MAE	RMSE	MAPE
SARIMA	63.84%	77.12%	71.21%
Seasonal Naive	58.57%	68.67%	51.57%
Prophet	62.42%	76.73%	73.63%

Discussion

The analysis revealed **weekly seasonality** and an upward **long-term trend** in the Brazilian e-commerce sales data. There was also a sharp sales spike near the end of 2017, likely due to a major event or holiday. Both SARIMA and Prophet models were able to capture this pattern, as can be seen in the plots. The residual component displayed mostly random noise, with some extreme outliers during the sales peak period, which suggests that models may have struggled with irregular sales.

However, on being evaluated using error metrics, the **seasonal naive benchmark outperformed both SARIMA and Prophet** on all three metrics (MAE, RMSE, MAPE). Despite their complexity, the advanced models did not yet meaningfully improve on a simple repeat-last-week strategy. A reason behind this may be that, naïve model fully leverages the strong and stable weekly seasonality seen in the data, while SARIMA and Prophet require further careful parameterization and additional inputs to exceed their performances. Another reason could be the short test window (60 days), which may limit the models' ability to generalize. The residual diagnostics for SARIMA showed no significant autocorrelation, indicating that the model fits the available data well, even if its predictive performance was limited.

The Ljung-Box test (residual diagnostic for SARIMA) showed no significant autocorrelation ($p \approx 0.913$), which means the model adequately captured the patterns in training data.

Future work should incorporate holiday effects, and external regressors (like marketing campaigns), and potentially explore machine-learning approaches to improve predictive performance.