Code Review: Seasonal and Residual Component Analysis

Overview

The script analyzes the sales performance for various sub-categories within the Superstore Sales Dataset by calculating and visualizing the seasonal and residual components of monthly sales data. The objective is to evaluate sales against seasonal expectations to identify whether sales meet or exceed expectations for specific months.

Key Components

1. **Library Imports**:
    - The script successfully imports necessary libraries:
        - pandas: For data manipulation and analysis.
        - statsmodels.api: For performing seasonal decomposition of time series data.
        - plotly.graph_objs: For creating interactive visualizations.
2. **Data Loading**:
    - The dataset is loaded using pd.read_excel, and appropriate error handling is applied during date conversion (errors='coerce'), ensuring that any invalid dates are handled gracefully.
3. **Setting Date Index**:
    - The 'Order Date' column is set as the index, which is essential for time series analysis. This allows for efficient resampling of sales data.
4. **Sub-category Filtering**:
    - A list of sub-categories is defined, and the script iterates through each to analyze sales data individually. This modular approach allows for targeted analysis across different product lines.
5. **Monthly Sales Calculation**:
    - The sales data is resampled to monthly totals using resample('ME').sum(). This step is crucial for obtaining a clear view of sales trends over time.
6. **Seasonal Decomposition**:

- The script performs seasonal decomposition using sm.tsa.seasonal_decompose, which separates the time series into seasonal, trend, and residual components. The chosen model is 'additive', which is suitable for data with consistent seasonality.

7. **Output of Seasonal and Residual Values**:
    - The script prints the seasonal and residual values for each sub-category, which provides direct insights into how actual sales deviate from expected seasonal patterns.
8. **Visualization**:
    - The seasonal and residual components are visualized using Plotly, enhancing the user experience with interactive plots. The use of a dark theme improves visibility, especially in presentations.

Insights Provided

- **Seasonal Expectations**: The seasonal component indicates the expected sales trend for each month, allowing for a comparison against actual performance.
- **Residual Analysis**: The residual component highlights any significant deviations from the seasonal expectations, which can indicate potential issues or opportunities in sales.

Suggestions for Improvement

1. **Error Handling**: Consider adding checks to ensure that the data for each sub-category is sufficient for seasonal decomposition. If not, handling the case gracefully would improve robustness.
2. **Documentation**: Adding comments throughout the code to explain each major step would enhance readability and maintainability, especially for others who might work with the code in the future.
3. **Modularization**: Breaking the code into functions could improve organization and facilitate testing. For example, you might create functions for loading data, performing decomposition, and generating visualizations.
4. **Statistical Summary**: Including a statistical summary of the seasonal and residual values (like mean, median, standard deviation) could provide more insights into sales performance trends.

Conclusion

This script effectively analyzes and visualizes the seasonal and residual components of sales data across different sub-categories. By focusing on these elements, it provides valuable insights into sales performance and trends. With a few enhancements, it could be made even more robust and user-friendly for ongoing analysis.