

2024

SALES FORECASTING AND TREND ANALYSIS: A TIME SERIES APPROACH



Gary Waiyaki

TABLE OF CONTENTS

- 01** Tile Page
- 02** Table of Contents
- 03** Introduction and Project Overview
- 04** Data Collection and Preparation
- 05** Exploratory Data Analysis (EDA)
- 06** Decomposition and Stationarity Testing
- 07** SARIMA Model Selection and Forecasting
- 08** Comparative Analysis Using Prophet Model
- 09** Trends and Patterns Prophet Model Components
- 10** Conclusion and Recommendation

INTRODUCTION

In the competitive landscape of retail, the ability to predict future sales with accuracy is not just an advantage; it is a necessity for survival and growth.

This report presents a detailed exploration into the realm of sales forecasting by leveraging the strengths of two sophisticated time series analysis models: Seasonal Autoregressive Integrated Moving Average (SARIMA) and Prophet. SARIMA excels in capturing the nuances of seasonality and trends in time series data, while Prophet offers flexibility in handling holiday effects and is robust to missing data and shifts in trends.

The combined use of these models allows us to address the dual objectives of the project: firstly, to untangle and understand the historical sales trends and patterns with a fine-tooth comb; secondly, to extend this understanding into the future, making predictions that can inform key business strategies. The analysis encompasses several years of sales data from a major retailer, focusing on two primary product categories: furniture and office supplies.

This project goes beyond mere prediction. It delves into the why and how of sales fluctuations, identifying critical periods of high and low demand, and examining the underlying factors that drive these variations. Through rigorous data preprocessing, exploratory analysis, and model tuning, we present a narrative that combines statistical rigor with business acumen.

The outcomes of this project are designed to serve as a cornerstone for inventory management, budget allocation, and strategic planning, ensuring that business decisions are data-driven and forward-looking. As we navigate through the complex patterns that underpin sales data, we pave the way for a more predictive and proactive business model.

DATA COLLECTION AND PREPARATION

Data Collection:

- The data was **sourced** from [Kaggle](#) and includes historical records of furniture and office supplies sales.
- **Time frame:** Data spanning from January 2015 to December 2022, providing a comprehensive view of sales trends over the years.
- **Attributes:** The dataset included variables such as 'Order Date', 'Sales', 'Quantity', and 'Category'.

Data Cleaning:

- **Handling Missing Values:** We checked for and imputed missing values in the dataset to maintain the continuity of the time series.
- **Duplicate Removal:** Any duplicate records found in the dataset were removed to prevent skewing of the analysis results.

Data Transformation:

- **Date Parsing:** Converted 'Order Date' from string format to Python datetime object for time series analysis compatibility.
- **Indexing:** Set the 'Order Date' as the DataFrame index, which is a common practice for time series manipulation in pandas.
- **Aggregation:** Sales data was aggregated monthly to analyze and forecast at a consistent time interval.

EXPLORATORY DATA ANALYSIS (EDA)

Trend Analysis:

Visual Inspection: Created a line plot to visually inspect the overall trends in furniture sales over time. This helped in identifying any long-term upward or downward trends in the sales data.

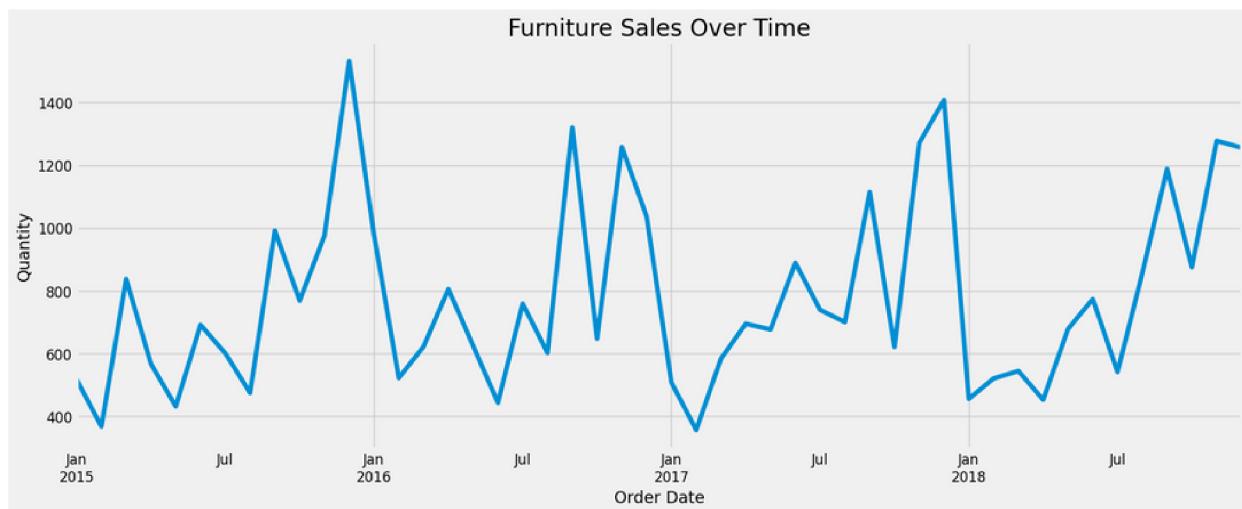


Fig 1: Furniture Sales Over Time

Plot Interpretation:

There are distinguishable patterns from January 2015 to July 2018 when we plot the data.

- Seasonality: noticeable peaks and troughs suggest seasonal patterns in furniture sales. Peaks are observed around July each year, which indicates a higher demand during mid-year, due to events like summer sales or financial planning for businesses.
- Trend: there is a general upward trend in the quantity of furniture sales over the observed period with some fluctuations. The lowest points consistently rise yearly, indicating that the worst-performing months are still improving over time.

DECOMPOSITION

Decomposition: Performed seasonal decomposition to separate the time series into trend, seasonal, and residual components, facilitating a clearer understanding of the underlying patterns.

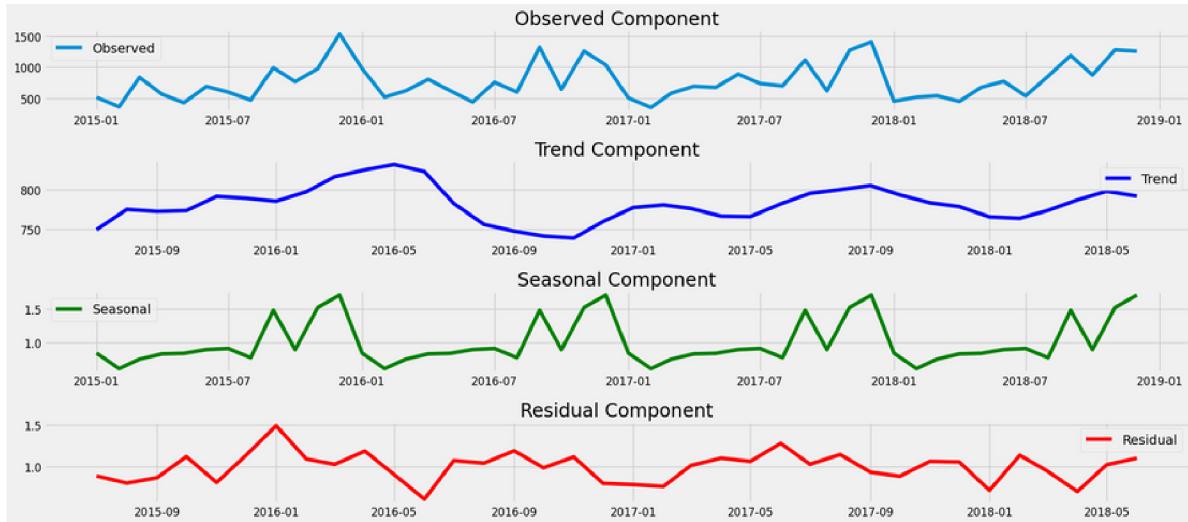


Fig 2: Seasonal Decomposition of Furniture Sales

Plot Interpretation:

- From the plot above we can see that our time series furniture sales data has obvious seasonality.

Testing for Stationarity

To apply time series forecasting models effectively, we tested the data for stationarity. The Augmented Dickey-Fuller test results revealed our time series data was stationary and ready for the SARIMA model.

The Augmented Dickey-Fuller (ADF) test results indicate the following:

- ADF Statistic: a value of -5.321971 is highly negative
- p-value: a value of 0.000005 (essentially 0) is far below the common significance level (0.05)

Given these results, we reject the null hypothesis of the ADF Test suggesting that the time series furniture data is not stationary.

SARIMA MODEL SELECTION AND FORECASTING

- We carefully selected and tuned the SARIMA model parameters using the `auto_arima` function from the `pmdarima` package in Python. This is a powerful tool for automatically determining both the non-seasonal and seasonal parameters for a SARIMA model.
- The `auto_arima` function from the `pmdarima` package in Python is a powerful tool for automatically determining both the non-seasonal and seasonal parameters for a SARIMA model.
- We employed the SARIMA model to forecast future sales, carefully interpreting the confidence intervals and forecasted values.

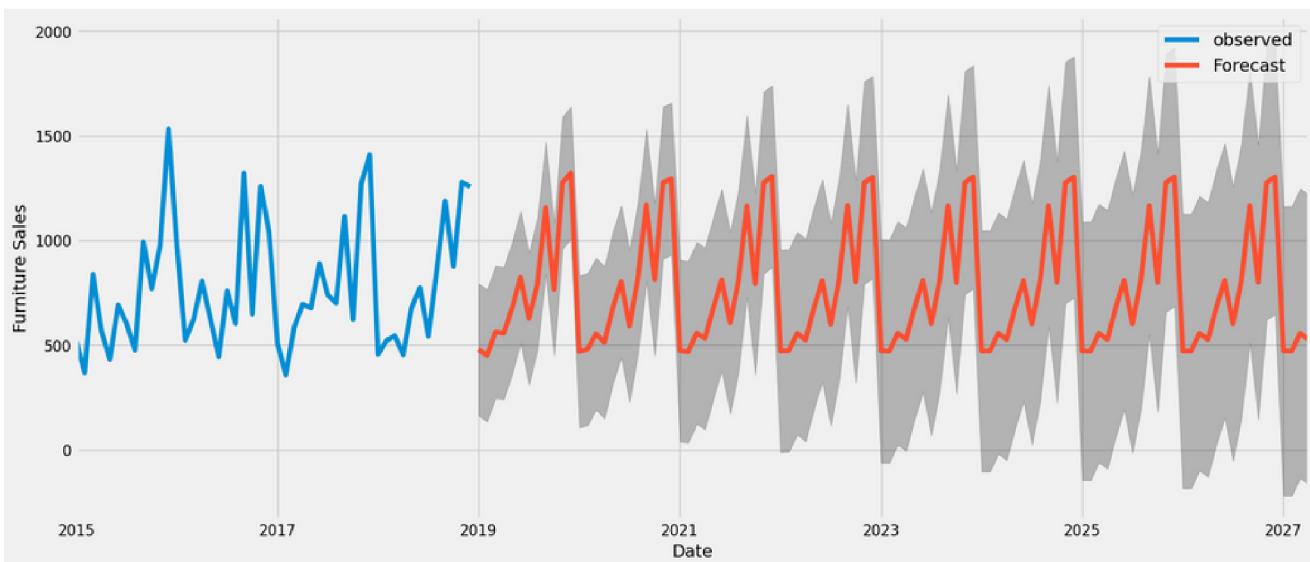


Fig 3: Future Furniture Sales Forecasts using SARIMA Model

Plot Interpretation:

- The plot shows our model captured furniture sales seasonality. As we forecast future sales, the confidence intervals generated by our model widen showing a natural pattern of uncertainty the further out we are into the future.
- The above time series analysis of furniture sales makes me curious about how other categories compare over time. Therefore, let's compare the time series of furniture sales and office supplies sales.

COMPARATIVE ANALYSIS USING PROPHET MODEL

A comparative analysis was conducted between furniture and office supplies sales. The trend comparisons and future sales predictions provided valuable business insights.

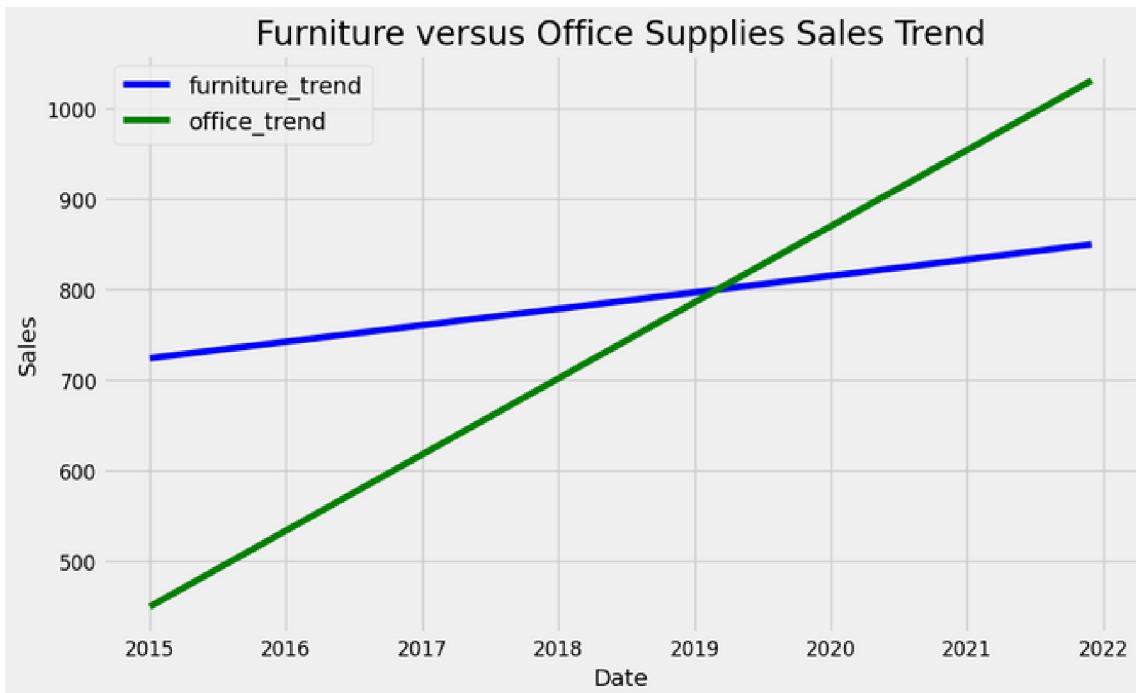


Fig 4: Furniture versus Office Sales Trend

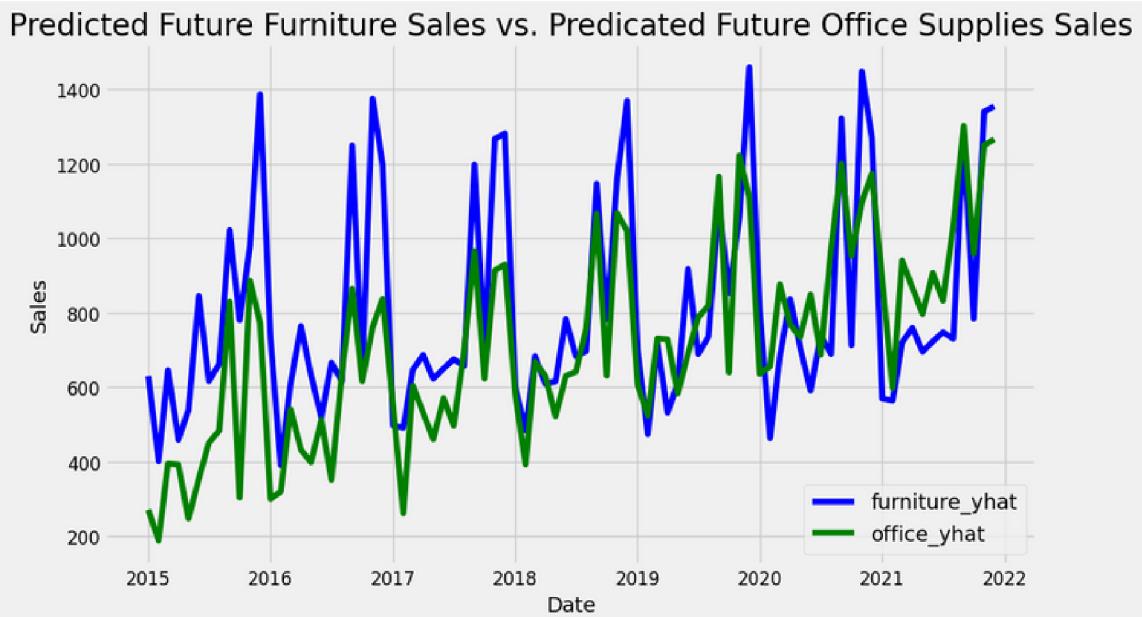


Fig 5: Predicted Furniture Sales versus Office Sales

TRENDS AND PATTERNS PROPHET MODEL COMPONENTS

We will now use the Prophet Models to inspect different trends of these two categories in the data.

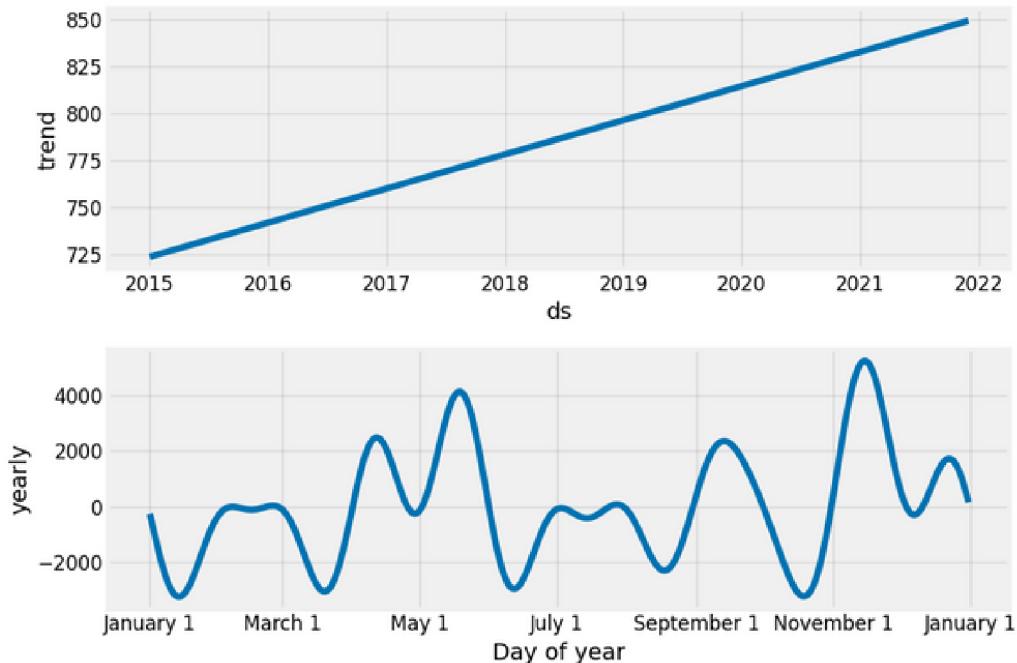


Fig 6: Prophet Furniture Forecast Model Components

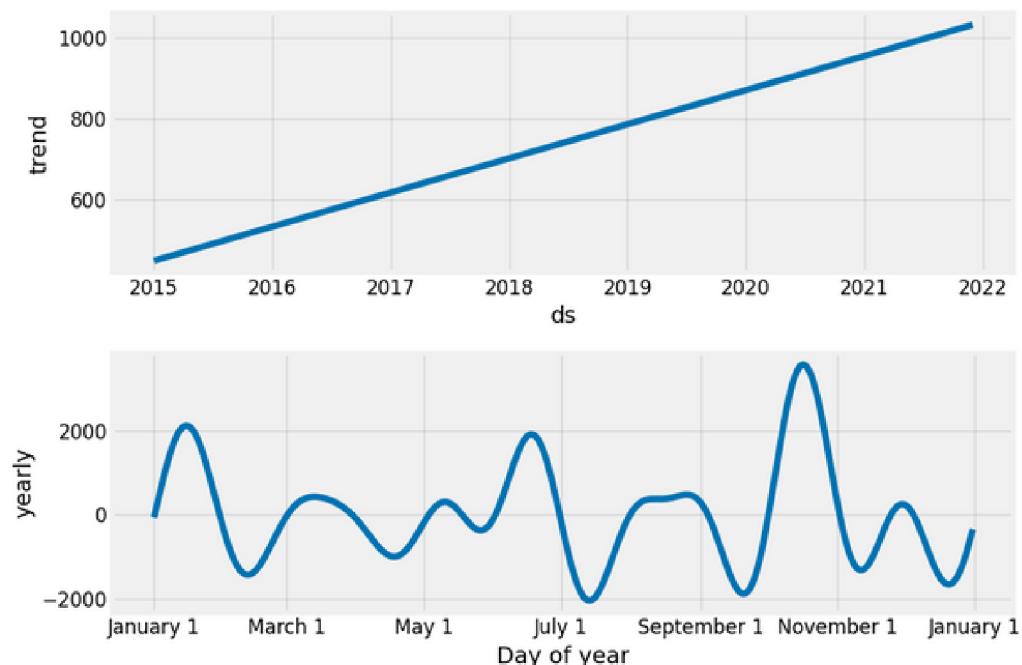


Fig 7: Prophet Office Supplies Forecast Model Components

CONCLUSION

- The sales for both furniture and office supplies have been increasing linearly over time. However, the office supplies' growth trend outpaces that of furniture sales.
- The worst month for furniture sales is April whereas the worst month for office supplies sales is August. The best month for furniture sales is December and the best month for office supplies sales is November.

RECOMMENDATIONS

- The time series analysis provided us with a clear understanding of the sales patterns and forecasted trends. Recommendations include:
 - Strategic stocking of inventory before peak seasons and potentially expanding the office supplies line, which shows a promising growth trend.
 - Further analysis could explore external factors that may impact sales, such as economic indicators or marketing efforts.