

Retail Sales Trend & Seasonality Analysis

Abstract

In this project, we performed a comprehensive time series analysis on daily retail sales data collected from multiple outlets of a nationwide retail company. The objective was to understand the underlying structure of the sales data by identifying long-term trends, recurring seasonal patterns, and random fluctuations. The analysis involved visual inspection of the time series, decomposition into trend, seasonality, and residual components, autocorrelation and partial autocorrelation analysis (ACF & PACF), and stationarity testing using the Augmented Dickey-Fuller (ADF) test. The results revealed a clear upward trend, strong weekly seasonality, and non-stationarity in the original series. This statistical validation prepares the dataset for accurate future forecasting using time series models such as ARIMA and SARIMA.

Problem Statement

Retail sales data is influenced by multiple time-dependent factors such as weekends, seasonal shopping behavior, festivals, and promotional campaigns. Without understanding whether the data exhibits a long-term trend, repeating seasonal patterns, or randomness, building forecasting models can lead to misleading predictions.

The goal of this project is to statistically analyze retail sales data to:

- Identify trends and seasonal effects
 - Measure temporal dependence using autocorrelation
 - Test whether the dataset satisfies stationarity assumptions
 - Prepare the data for reliable forecasting models
-

Table of Contents

1. Dataset Loading and Preprocessing
 2. Time Series Visualization
 3. Seasonal Decomposition (Trend, Seasonality, Residuals)
 4. Autocorrelation Function (ACF) Analysis
 5. Partial Autocorrelation Function (PACF) Analysis
 6. Stationarity Testing using Augmented Dickey-Fuller (ADF) Test
 7. Interpretation of Results
 8. Summary
 9. References
-

Dataset Loading and Preprocessing

The retail sales dataset was loaded using Pandas. The Date column was converted into a datetime format and set as the index to ensure compatibility with time series analysis methods.

Key preprocessing steps:

- Converted date column to datetime
- Set date as time index
- Verified data consistency and continuity

This ensured the dataset was correctly structured for time-dependent analysis.

Time Series Visualization

A time series plot was generated to visually inspect sales behavior over time.

Observations:

- Sales values show an overall increasing pattern
- Regular fluctuations suggest recurring seasonal behavior
- Short-term variations indicate noise and promotional effects

This visualization provided an initial understanding of the data dynamics.

Seasonal Decomposition

The sales series was decomposed using an **additive model** with a weekly seasonal period.

The decomposition produced four components:

- **Observed:** Original sales data
- **Trend:** Long-term upward movement in sales
- **Seasonal:** Repeating weekly pattern indicating higher sales on specific days
- **Residual:** Random noise not explained by trend or seasonality

Key Insight:

The presence of a strong trend and consistent seasonality confirms that sales behavior is systematic rather than random.

Autocorrelation Function (ACF) Analysis

The ACF plot was used to analyze the correlation between current sales values and their past values.

Findings:

- High autocorrelation at lower lags
- Gradual decay of correlation

- Repeating spikes indicating seasonality

This suggests that past sales strongly influence future sales and confirms non-stationary behavior.

Partial Autocorrelation Function (PACF) Analysis

PACF analysis was performed to measure direct relationships between observations at specific lags.

Observations:

- Significant spikes at early lags
- Limited influence beyond a certain lag

This information is useful for selecting appropriate AR terms when building forecasting models like ARIMA.

Stationarity Testing using ADF Test

The Augmented Dickey-Fuller (ADF) test was conducted to statistically verify stationarity.

Results:

- ADF Statistic ≈ -0.41
- p-value ≈ 0.90

Since the p-value is greater than 0.05, the null hypothesis cannot be rejected.

Conclusion:

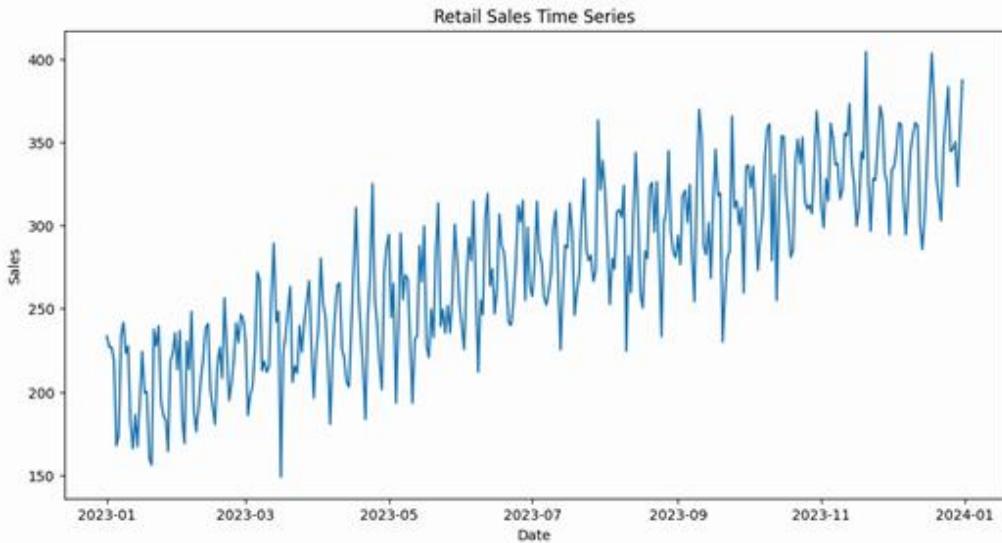
The time series is **non-stationary**, primarily due to the presence of trend and seasonality.

```
In [14]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

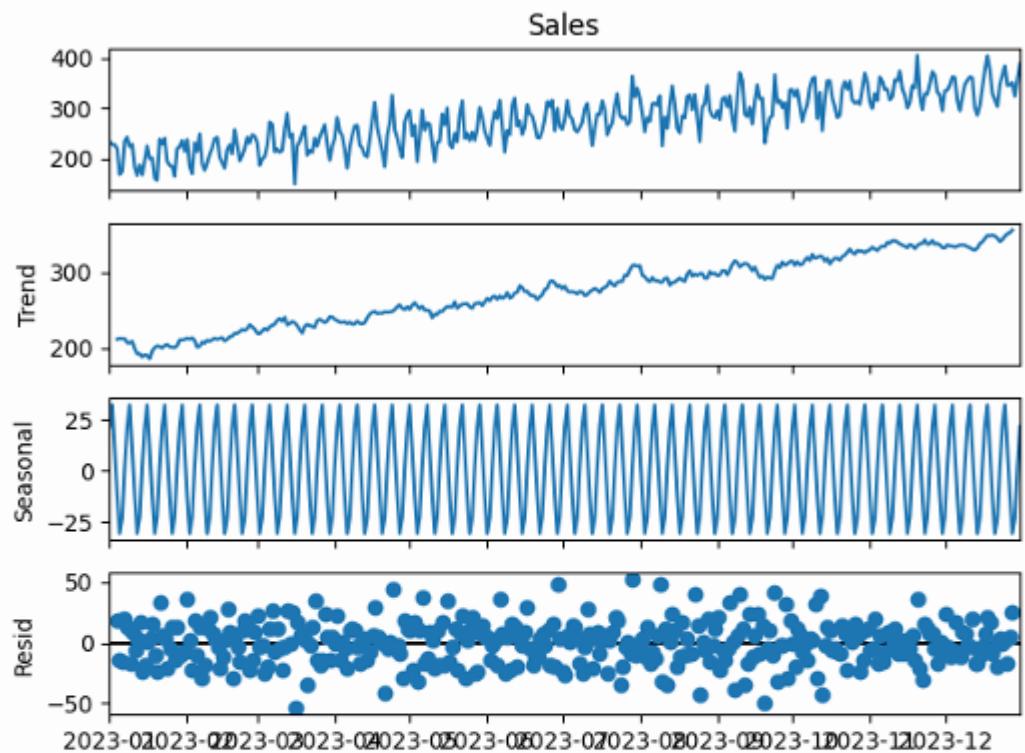
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller

# Load dataset
df = pd.read_csv(r"C:\Users\sanja\Downloads\retail_sales.csv")
df['Date'] = pd.to_datetime(df['Date'])
df.set_index('Date', inplace=True)
```

```
In [10]: # 1. Time Series Plot
plt.figure(figsize=(12,6))
plt.plot(df['Sales'])
plt.title("Retail Sales Time Series")
plt.xlabel("Date")
plt.ylabel("Sales")
plt.show()
```



```
In [11]: # 2. Decomposition
decomposition = seasonal_decompose(df['Sales'], model='additive', period=7)
decomposition.plot()
plt.show()
```



```
In [12]: # 3. ACF & PACF
plot_acf(df['Sales'], lags=30)
plt.show()

plot_pacf(df['Sales'], lags=30)
plt.show()

# 4. Stationarity Test
adf_test = adfuller(df['Sales'])
print("ADF Statistic:", adf_test[0])
print("p-value:", adf_test[1])
```

