demand_forecast_clean

August 20, 2025

1 Retail Demand Forecasting Analysis

1.1 Executive Summary

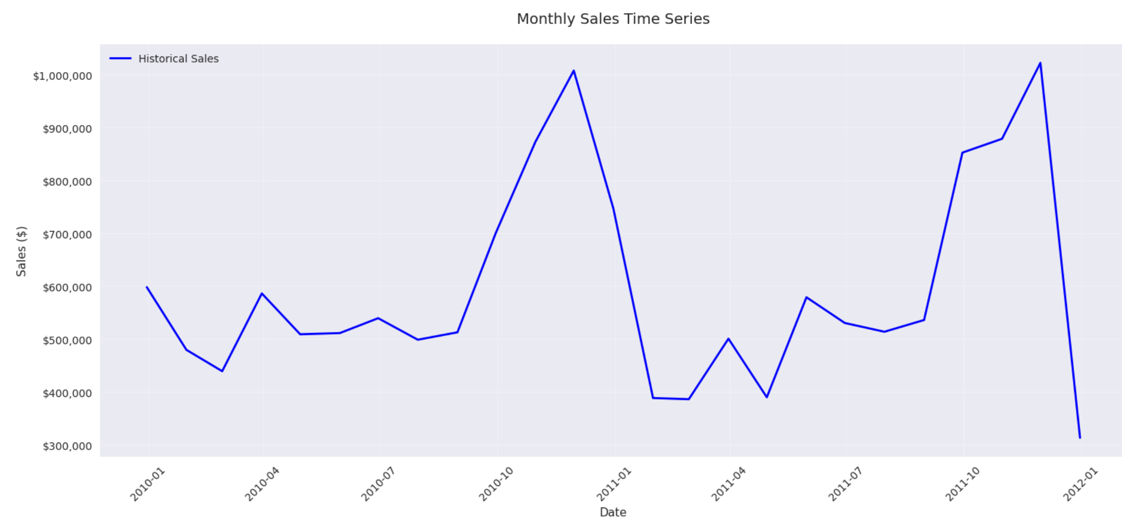
This notebook presents a demand forecasting for the retail business, providing 6-month sales predictions and strategic recommendations. The analysis includes multiple forecasting models, evaluation metrics, and actionable insights to support inventory planning and business decision-making.

1.2 Objectives

1. Predict sales performance for the next 6 months
2. Identify seasonal patterns and trends in demand
3. Determine which products need to be stocked up
4. Provide confidence levels and model limitations
5. Deliver actionable recommendations for business improvement

---

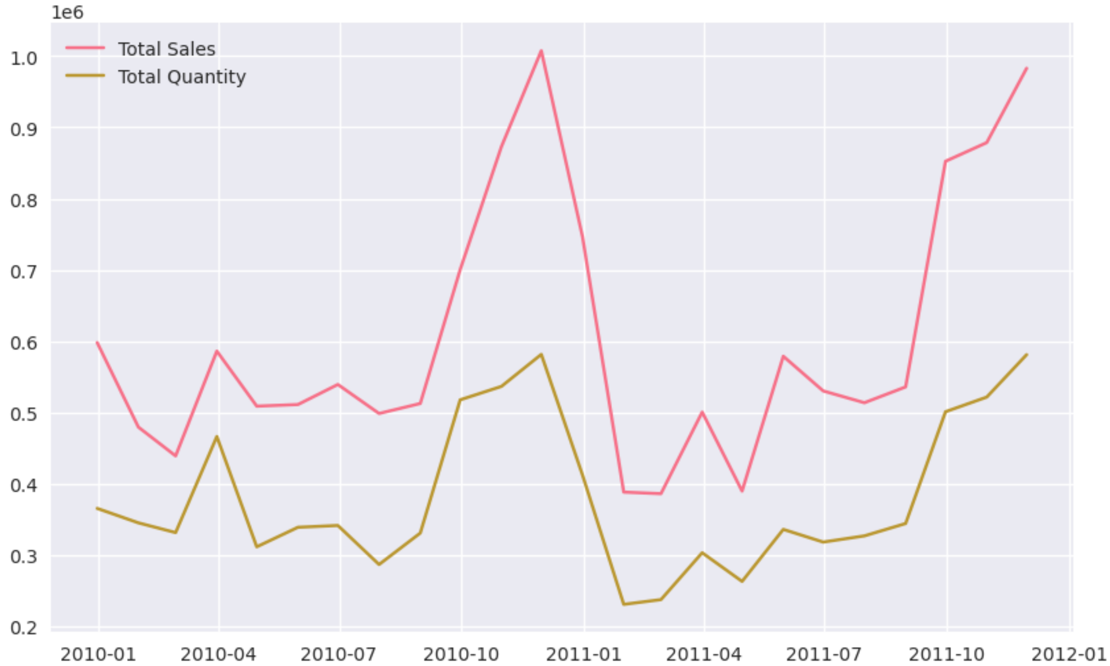
```
 All libraries imported successfully!
 Starting Retail Demand Forecasting Analysis…
============================================================

 Data Overview:
Total months: 25
Date range: 2009-12 to 2011-12
Average monthly sales: $595,649.47
============================================================
```

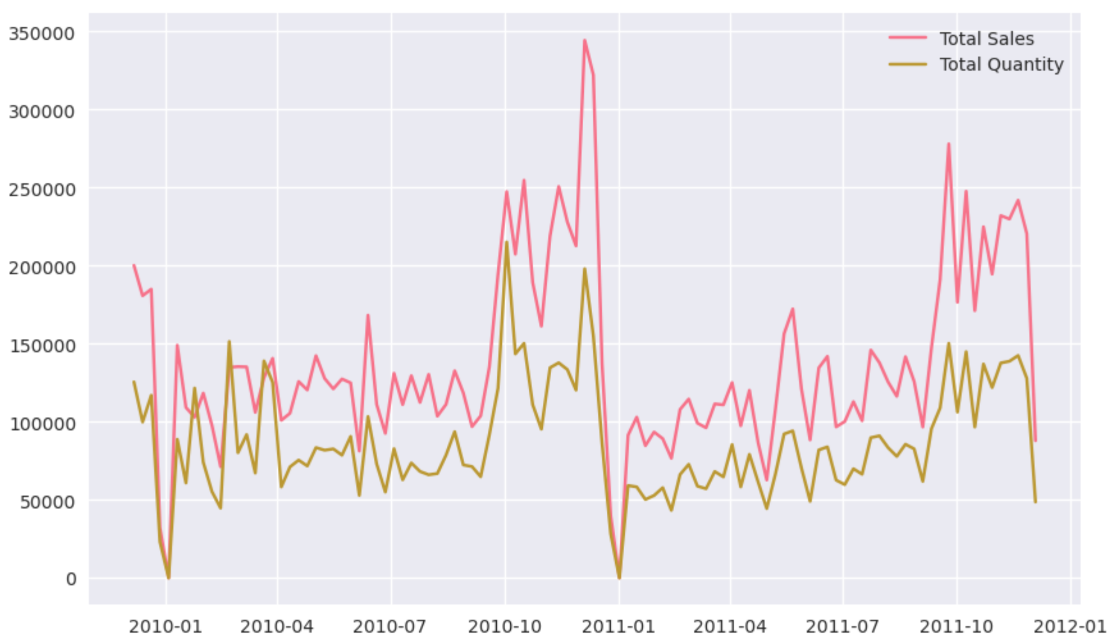


2 2. Data Preprocessing

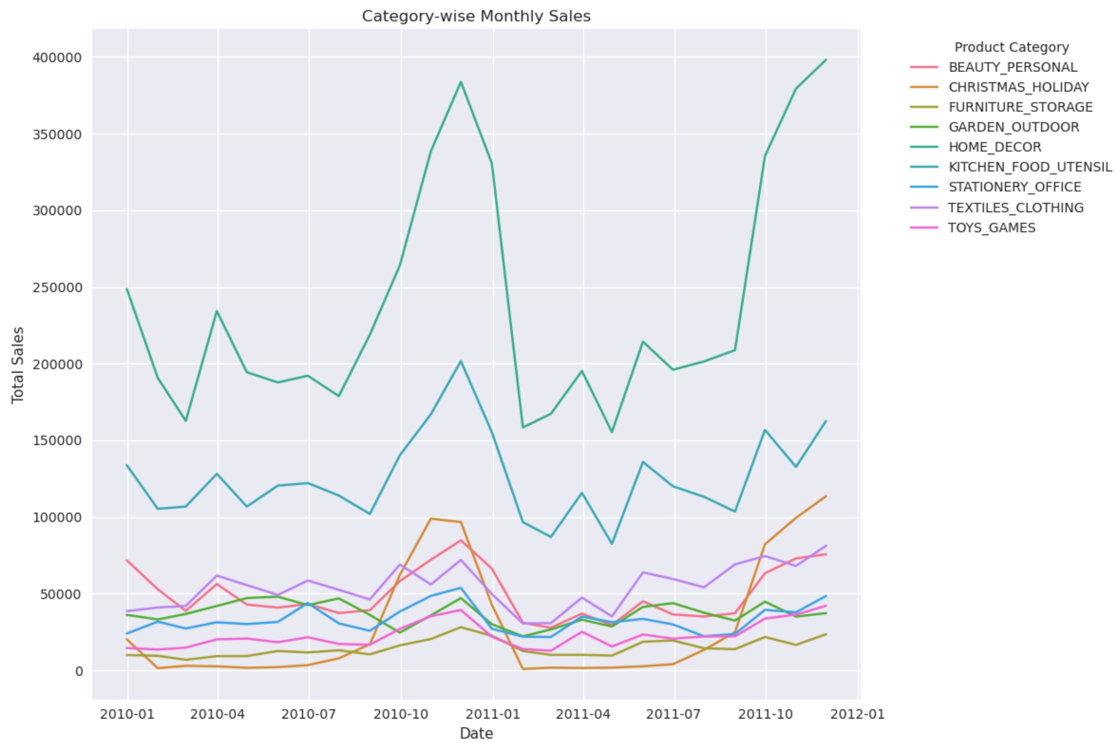
```
 Dataset Overview:
Dataset shape: (738653, 15)
Date range: 2009-12-01 07:45:00 to 2011-11-29 18:14:00
Total transactions: 738,653
Unique customers: 5,789
Unique products: 4,604
Product categories: 9

============================================================
 First few rows of the dataset:

   order_id product_id                 product_description  quantity  \
0    489434      85048  15CM CHRISTMAS GLASS BALL 20 LIGHTS        12
1    489434     79323P                   PINK CHERRY LIGHTS        12
2    489434     79323W                  WHITE CHERRY LIGHTS        12
3    489434      22041         RECORD FRAME 7" SINGLE SIZE         48
4    489434      21232       STRAWBERRY CERAMIC TRINKET BOX        24

           order_date  unit_price  customer_id         country  total_amount  \
0 2009-12-01 07:45:00        6.95      13085.0  United Kingdom          83.4
1 2009-12-01 07:45:00        6.75      13085.0  United Kingdom          81.0
2 2009-12-01 07:45:00        6.75      13085.0  United Kingdom          81.0
3 2009-12-01 07:45:00        2.10      13085.0  United Kingdom         100.8
4 2009-12-01 07:45:00        1.25      13085.0  United Kingdom          30.0

   year  month  quarter day_of_week month_year   product_category
0  2009     12        4     Tuesday    2009-12  CHRISTMAS_HOLIDAY
```

```
1  2009     12        4     Tuesday    2009-12    BEAUTY_PERSONAL
2  2009     12        4     Tuesday    2009-12    BEAUTY_PERSONAL
3  2009     12        4     Tuesday    2009-12         HOME_DECOR
4  2009     12        4     Tuesday    2009-12         HOME_DECOR

 Data Quality Assessment:

Missing values:
Series([], dtype: int64)
 No missing values found!

 Data Types:
order_id                        int64
product_id                     object
product_description            object
quantity                        int64
order_date             datetime64[ns]
unit_price                    float64
customer_id                   float64
country                        object
total_amount                  float64
year                            int64
month                           int64
quarter                         int64
day_of_week                    object
month_year                     object
product_category               object
dtype: object

 Revenue Summary:
Total Revenue: $14,538,763.42
Average Order Value: $19.68
Median Order Value: $10.79

 Date Range Analysis:
Data collection period: 728 days
Years covered: [np.int64(2009), np.int64(2010), np.int64(2011)]
============================================================

 Creating time series aggregations…
 Monthly data created: 24 complete months
 Date range: 2009-12 to 2011-11
 Weekly data created: 105 weeks
```

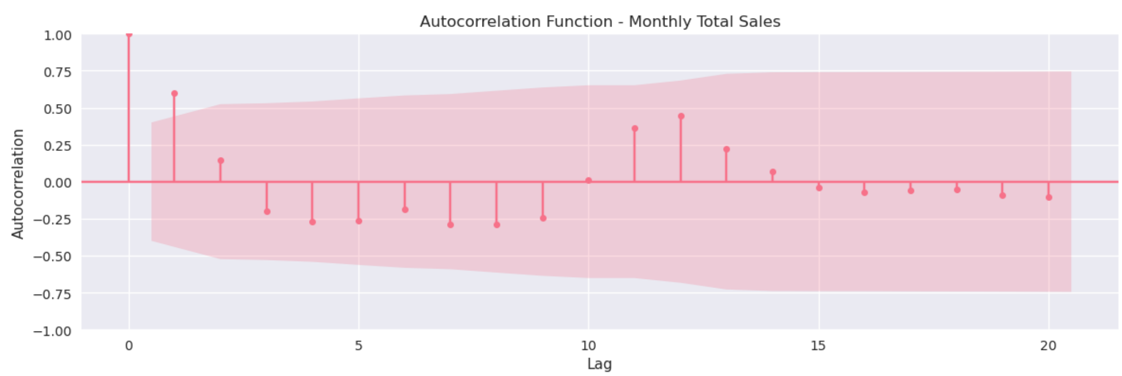


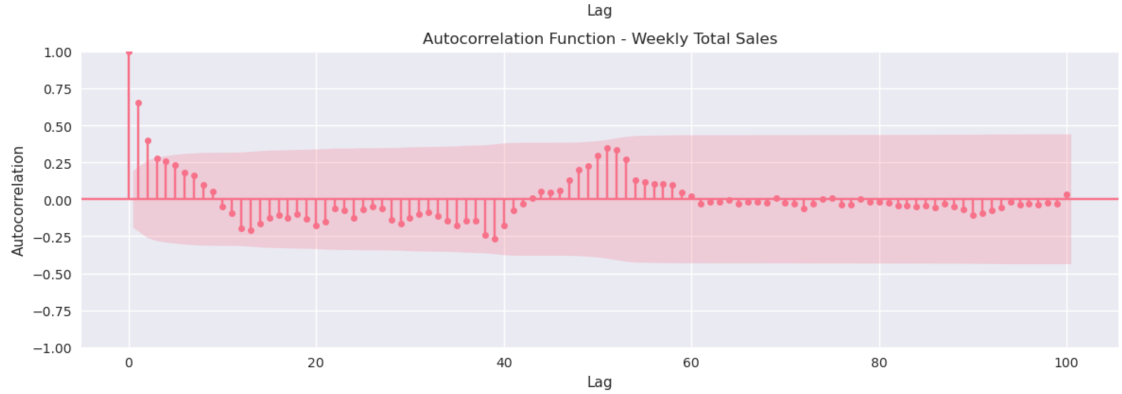



```
Creating category-wise time series…
  Monthly category data created: 216 records
  Weekly category data created: 927 records
```

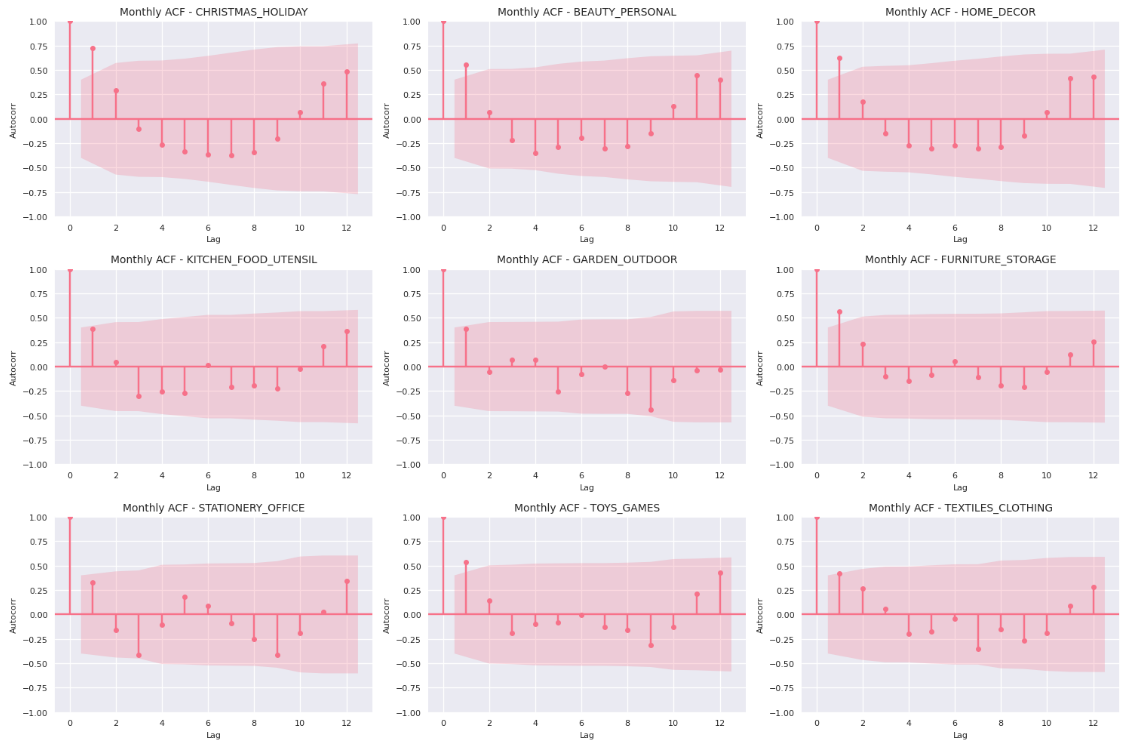


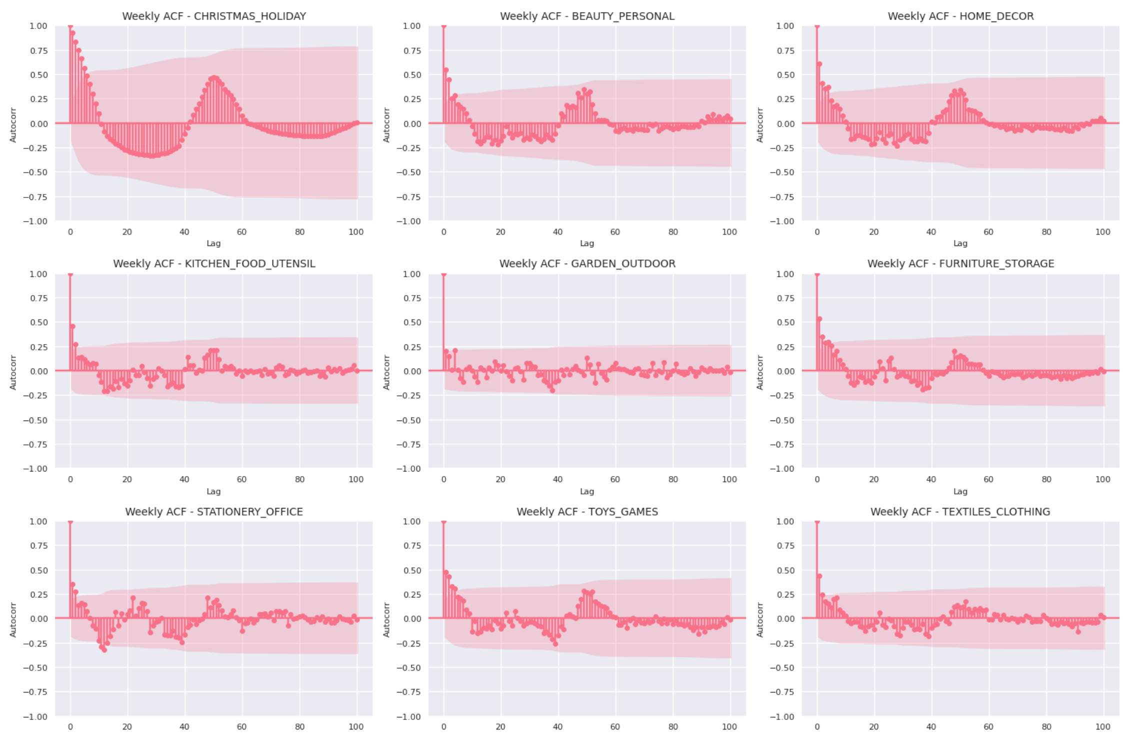
Calculating autocorrelation for monthly revenue...





```
Autocorrelation values for monthly sales (first 12 lags):
Lag 0: 1.000
Lag 1: 0.597
Lag 2: 0.144
Lag 3: -0.199
Lag 4: -0.269
Lag 5: -0.265
Lag 6: -0.188
Lag 7: -0.291
Lag 8: -0.292
Lag 9: -0.246
Lag 10: 0.015
Lag 11: 0.364
Lag 12: 0.448

ADF Test Results for Monthly Total Sales:
ADF Statistic: -3.291
p-value: 0.015
Critical Values:
   1%: -4.012
   5%: -3.104
   10%: -2.691

Calculating autocorrelation for categorical sales…
Found 9 product categories: ['CHRISTMAS_HOLIDAY', 'BEAUTY_PERSONAL',
'HOME_DECOR', 'KITCHEN_FOOD_UTENSIL', 'GARDEN_OUTDOOR', 'FURNITURE_STORAGE',
'STATIONERY_OFFICE', 'TOYS_GAMES', 'TEXTILES_CLOTHING']
```


Monthly ACF - CHRISTMAS_HOLIDAY
Monthly ACF - BEAUTY_PERSONAL
Monthly ACF - HOME_DECOR
Monthly ACF - KITCHEN_FOOD_UTENSIL
Monthly ACF - GARDEN_OUTDOOR
Monthly ACF - FURNITURE_STORAGE
Monthly ACF - STATIONERY_OFFICE
Monthly ACF - TOYS_GAMES
Monthly ACF - TEXTILES_CLOTHING
Autocorr
Lag


Weekly ACF - CHRISTMAS_HOLIDAY
Weekly ACF - BEAUTY_PERSONAL
Weekly ACF - HOME_DECOR
Weekly ACF - KITCHEN_FOOD_UTENSIL
Weekly ACF - GARDEN_OUTDOOR
Weekly ACF - FURNITURE_STORAGE
Weekly ACF - STATIONERY_OFFICE
Weekly ACF - TOYS_GAMES
Weekly ACF - TEXTILES_CLOTHING
Autocorr
Lag

```
Autocorrelation summary for monthly categorical sales:

CHRISTMAS_HOLIDAY (24 data points):
  Lag 0: 1.000
  Lag 1: 0.721
  Lag 2: 0.289
  Lag 3: -0.103
  Lag 4: -0.264
  Lag 5: -0.336
  Lag 6: -0.366

BEAUTY_PERSONAL (24 data points):
  Lag 0: 1.000
  Lag 1: 0.557
  Lag 2: 0.067
  Lag 3: -0.220
  Lag 4: -0.352
  Lag 5: -0.289
  Lag 6: -0.196

HOME_DECOR (24 data points):
  Lag 0: 1.000
  Lag 1: 0.622
  Lag 2: 0.174
  Lag 3: -0.146
  Lag 4: -0.274
  Lag 5: -0.302
  Lag 6: -0.271

KITCHEN_FOOD_UTENSIL (24 data points):
  Lag 0: 1.000
  Lag 1: 0.384
  Lag 2: 0.049
  Lag 3: -0.298
  Lag 4: -0.257
  Lag 5: -0.267
  Lag 6: 0.019

GARDEN_OUTDOOR (24 data points):
  Lag 0: 1.000
  Lag 1: 0.387
  Lag 2: -0.052
  Lag 3: 0.070
  Lag 4: 0.069
  Lag 5: -0.253
  Lag 6: -0.077
```

```
FURNITURE_STORAGE (24 data points):
  Lag 0: 1.000
  Lag 1: 0.565
  Lag 2: 0.235
  Lag 3: -0.097
  Lag 4: -0.145
  Lag 5: -0.089
  Lag 6: 0.055

STATIONERY_OFFICE (24 data points):
  Lag 0: 1.000
  Lag 1: 0.328
  Lag 2: -0.162
  Lag 3: -0.415
  Lag 4: -0.107
  Lag 5: 0.179
  Lag 6: 0.090

TOYS_GAMES (24 data points):
  Lag 0: 1.000
  Lag 1: 0.538
  Lag 2: 0.142
  Lag 3: -0.193
  Lag 4: -0.094
  Lag 5: -0.082
  Lag 6: -0.004

TEXTILES_CLOTHING (24 data points):
  Lag 0: 1.000
  Lag 1: 0.423
  Lag 2: 0.264
  Lag 3: 0.060
  Lag 4: -0.196
  Lag 5: -0.178
  Lag 6: -0.041
```

3 Level 1 — Total Revenue Forecast (CEO view)

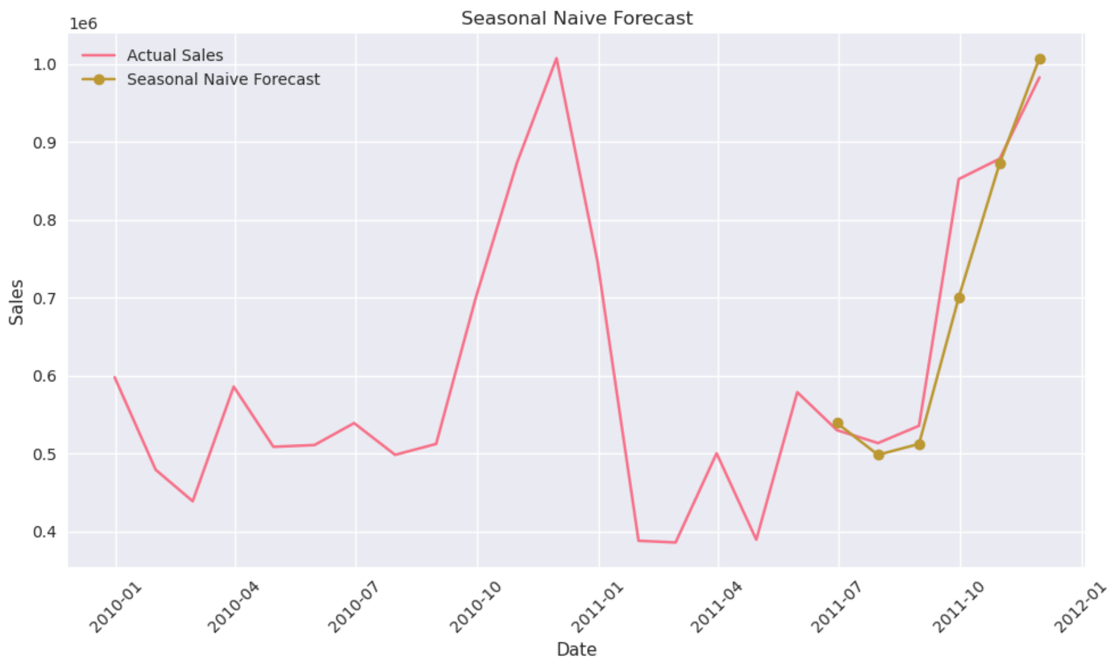
3.1 Backtesting (Model Selection)

1.1 Backtesting (Model Selection) - Rolling origin on last 12–18 months; models: Seasonal-Naïve, ARIMA, SARIMA (m=12). - Collect RMSE/MAE/MAPE; compare

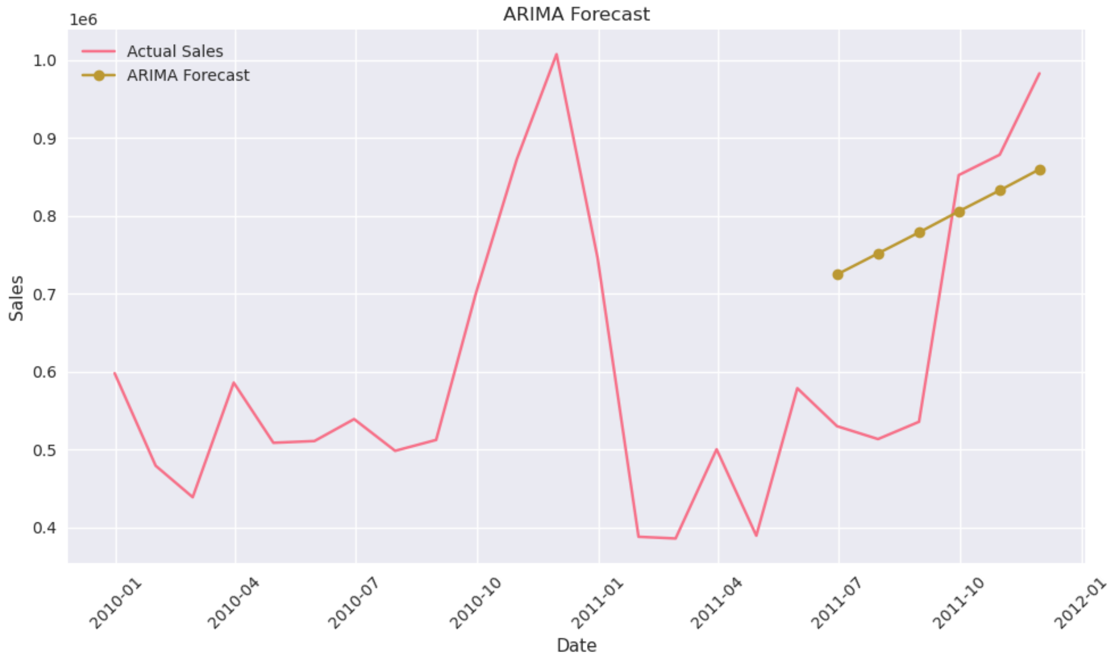
```
Preparing data for forecasting models…
Using clean dataset (December 2011 already removed)
Data Split:
```

- `Training period: 2009-12 to 2011-05 (18 months)`
- `Testing period: 2011-06 to 2011-11 (6 months)`

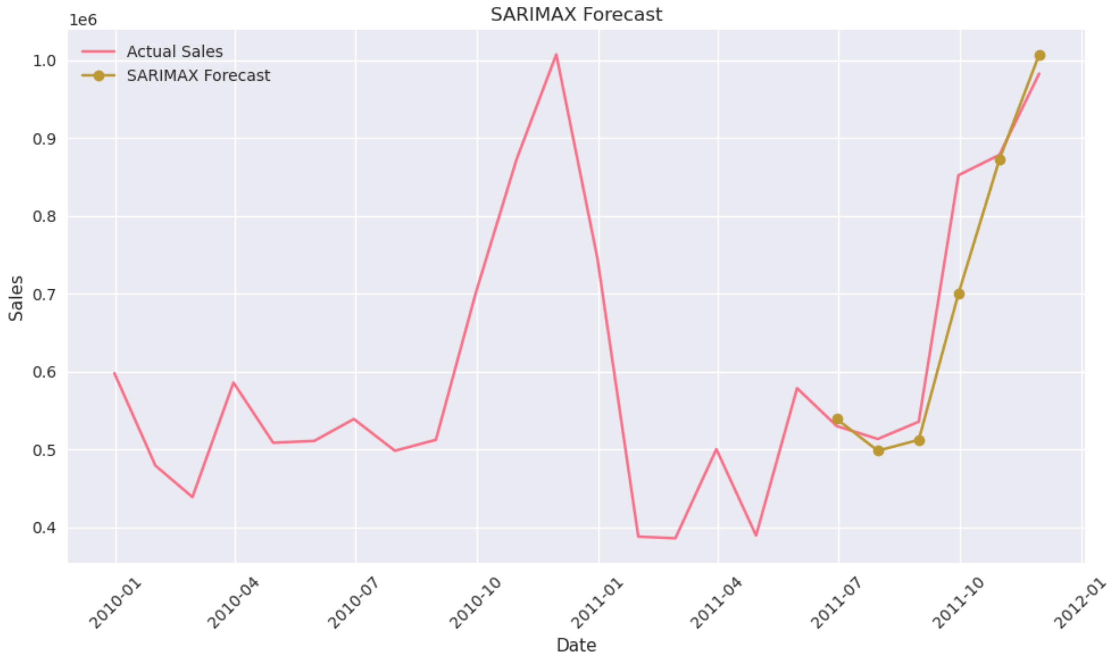
```
 Data prepared for modeling
 Future forecasts will start from December 2011
============================================================

 Model 1: Seasonal Naive Forecasting
This model uses the same month from the previous year as the forecast
 Seasonal Naive Model Results:
  MAE: $38,280.68
  RMSE: $64,004.94
  MAPE: 4.99%
============================================================
```




```
 Model 2: ARIMA (AutoRegressive Integrated Moving Average)
Automatically finding optimal parameters…

 ARIMA Model Results:
  Best parameters: (0, 2, 2)
  AIC: 434.73
  MAE: $148,516.64
  RMSE: $169,819.71
  MAPE: 25.26%
============================================================
```




```
 Model 3: SARIMAX (Seasonal ARIMA with eXogenous variables)
This model accounts for seasonality in the data
 SARIMAX Model Results:
  Best parameters: (2, 0, 1)
  Seasonal parameters: (0, 1, 0, 12)
  AIC: 8.00
  MAE: $38,280.68
  RMSE: $64,004.94
  MAPE: 4.99%
============================================================
```



MODEL PERFORMANCE SUMMARY

==

Model Comparison:

Model	MAE	RMSE	MAPE
Seasonal Naive	38280.68	64004.94	4.99
ARIMA(0, 2, 2)	148516.64	169819.71	25.26
SARIMAX(2, 0, 1)x(0, 1, 0, 12)	38280.68	64004.94	4.99

Best Models by Metric:

- Lowest MAE: SARIMAX(2, 0, 1)x(0, 1, 0, 12) ($38,280.68)
- Lowest RMSE: SARIMAX(2, 0, 1)x(0, 1, 0, 12) ($64,004.94)
- Lowest MAPE: SARIMAX(2, 0, 1)x(0, 1, 0, 12) (4.99%)

Recommended Model: SARIMAX(2, 0, 1)x(0, 1, 0, 12)

This model has the lowest Mean Absolute Error, indicating the best average prediction accuracy.

Key Insights:

- MAE varies by $110,235.96 across models
- Average MAPE across all models: 11.75%
- All models show good forecasting accuracy (MAPE < 15%)

==

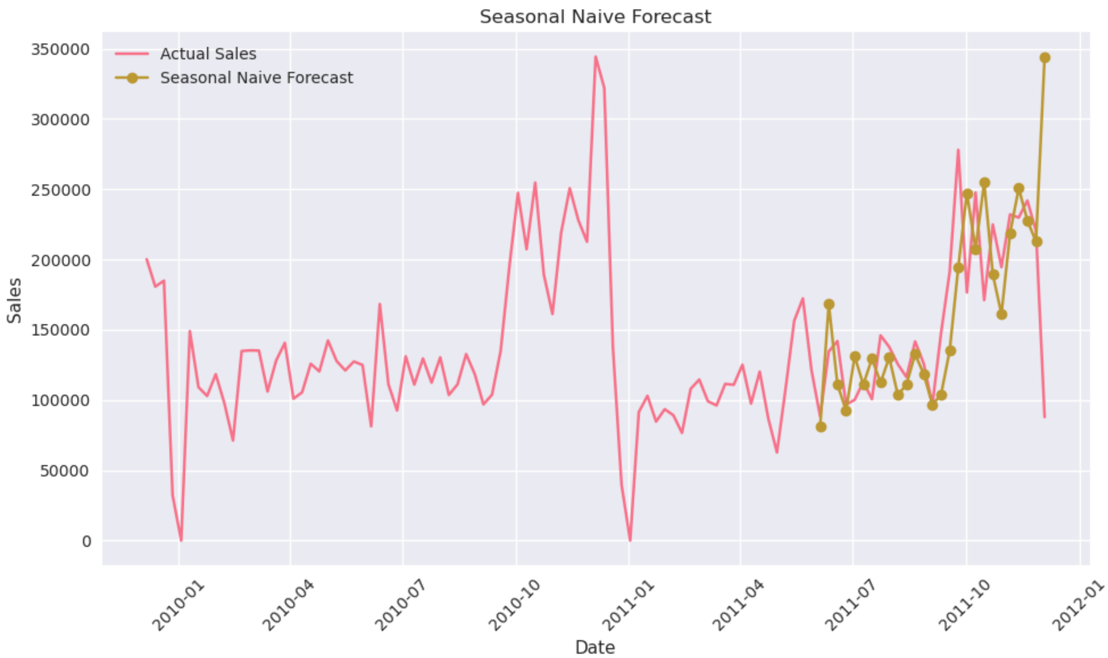
Preparing data for forecasting models…

Using clean dataset (December 2011 already removed)

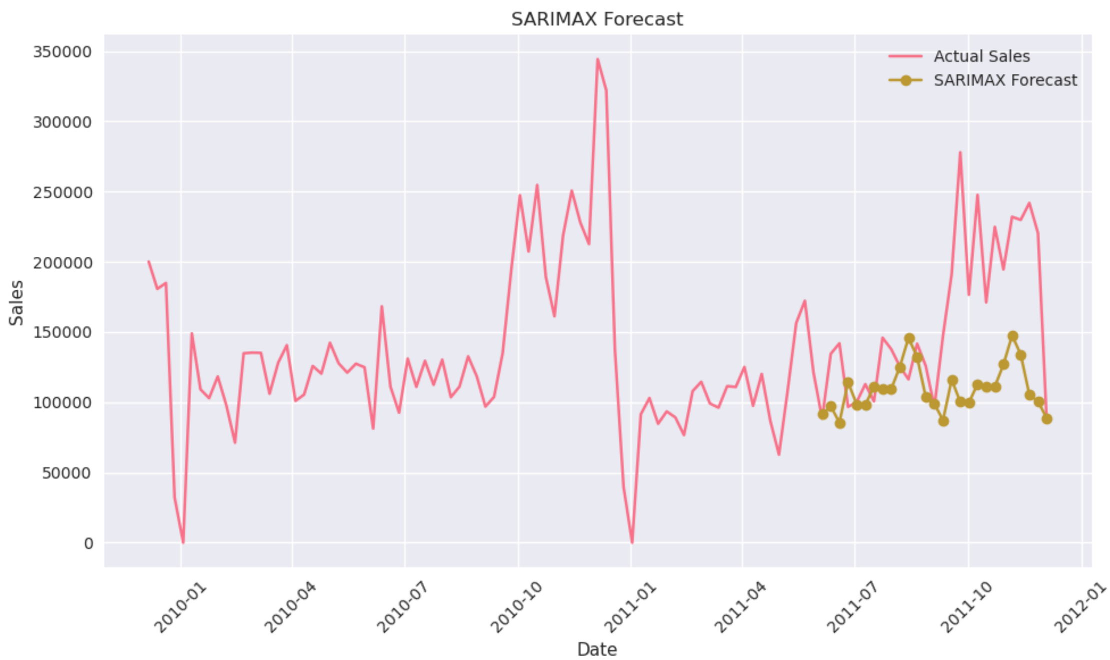
Data Split:
- Training period: 2009-12 to 2011-05 (78 months)
- Testing period: 2011-06 to 2011-12 (27 months)

Data prepared for modeling
Future forecasts will start from December 2011
==

Model 1: Seasonal Naive Forecasting
This model uses the same week from the previous year as the forecast
Seasonal Naive Model Results:
MAE: $36,349.55
RMSE: $60,988.11
MAPE: 26.90%
==



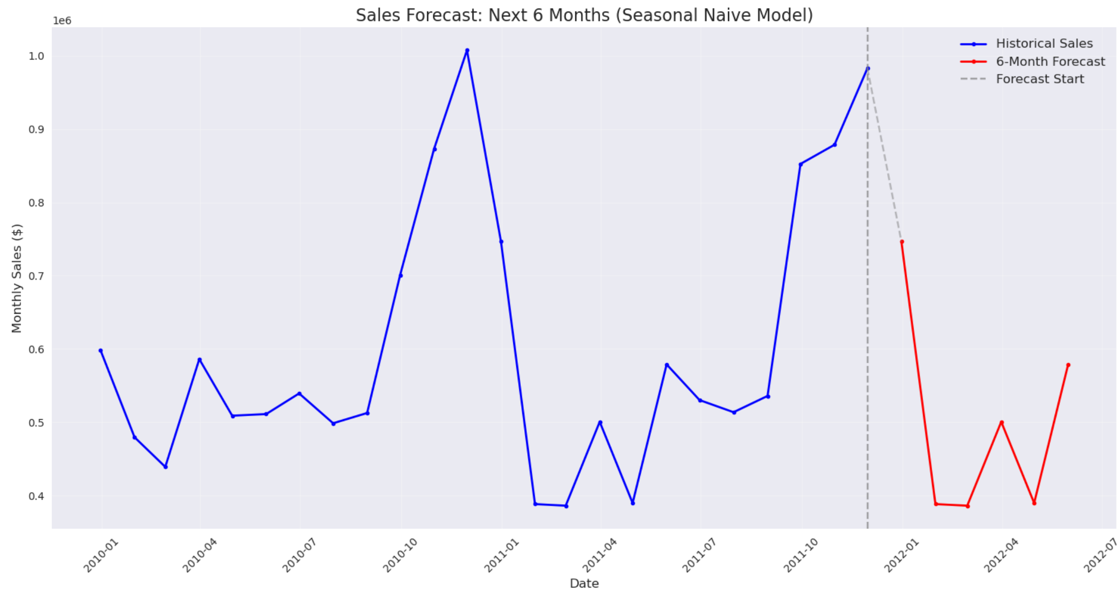
Model 3: SARIMAX (Seasonal ARIMA with eXogenous variables)
This model accounts for seasonality in the data
SARIMAX Model Results:
Best parameters: (0, 1, 2)
Seasonal parameters: (0, 1, 1, 12)
AIC: 1608.72
MAE: $54,583.92
RMSE: $73,077.10
MAPE: 28.32%
==

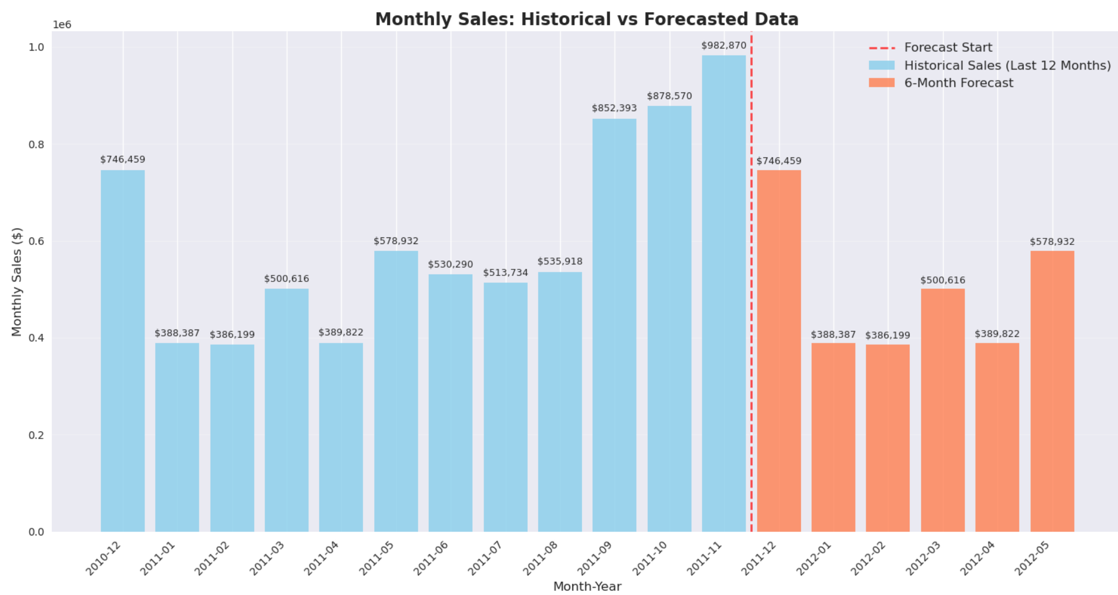


With the data that we have Naive Seasonal is the best model to forecast the next 6 months data (or we can use yearly patternal). It is logical because from the EDA we see that there is no growt between YoY, so we expect the pattern remain the same in our observed data!

3.2 Fit Best Model & Forecast 6 Months

```
 FORECASTING NEXT 6 MONTHS
==================================================
 Forecast Summary:
   Model: Seasonal Naive
   Forecast period: 6 months
   Average monthly forecast: $498,402.47
   Total 6-month forecast: $2,990,414.82

 6-month forecast:
      date  forecasted_sales          model
2011-12-31         746458.82 Seasonal_Naive
2012-01-31         388387.43 Seasonal_Naive
2012-02-29         386198.68 Seasonal_Naive
2012-03-31         500616.20 Seasonal_Naive
2012-04-30         389822.19 Seasonal_Naive
2012-05-31         578931.50 Seasonal_Naive
```





4 Level 2 — Category Forecasts (Middle-Out)

Inventory and merchandising vary by category (Christmas spikes vs core stability).

4.1 Prepare the categorical dataframe

```
Category Monthly Sales Data Shape: (24, 9)
```

```
Available Categories:
['BEAUTY_PERSONAL', 'CHRISTMAS_HOLIDAY', 'FURNITURE_STORAGE', 'GARDEN_OUTDOOR',
'HOME_DECOR', 'KITCHEN_FOOD_UTENSIL', 'STATIONERY_OFFICE', 'TEXTILES_CLOTHING',
'TOYS_GAMES']

First few rows of category data:

product_category  BEAUTY_PERSONAL  CHRISTMAS_HOLIDAY  FURNITURE_STORAGE  \
month_year
2009-12-01               71761.44           20258.20           10017.59
2010-01-01               52920.26            1475.87            9530.33
2010-02-01               38803.38            2993.03            6873.38
2010-03-01               56309.76            2648.19            9295.99
2010-04-01               42949.61            1665.67            9327.05

product_category  GARDEN_OUTDOOR  HOME_DECOR  KITCHEN_FOOD_UTENSIL  \
month_year
2009-12-01              36204.42   248822.76             133782.87
2010-01-01              33203.23   190725.73             105330.18
2010-02-01              36794.59   162667.08             106828.89
2010-03-01              41992.95   234315.55             128173.53
2010-04-01              47169.67   194428.25             106836.02

product_category  STATIONERY_OFFICE  TEXTILES_CLOTHING  TOYS_GAMES
month_year
2009-12-01                 24038.84           38654.20    14600.37
2010-01-01                 31767.38           40988.29    13594.32
2010-02-01                 27334.40           42013.59    14837.22
2010-03-01                 31357.20           61818.26    20223.26
2010-04-01                 30204.03           55608.35    20768.84
```

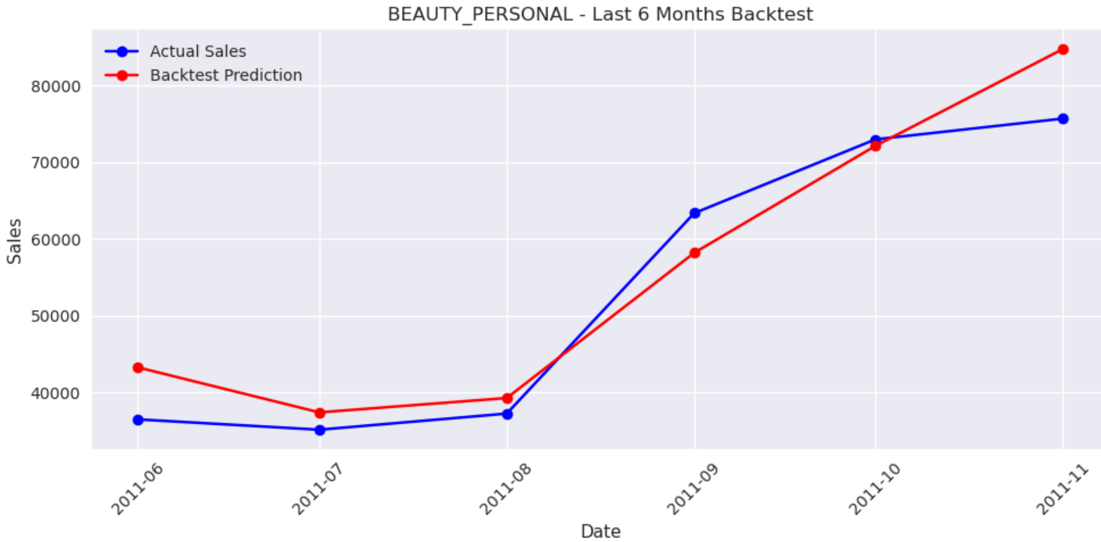
4.2 Backtesting (Model Selection)

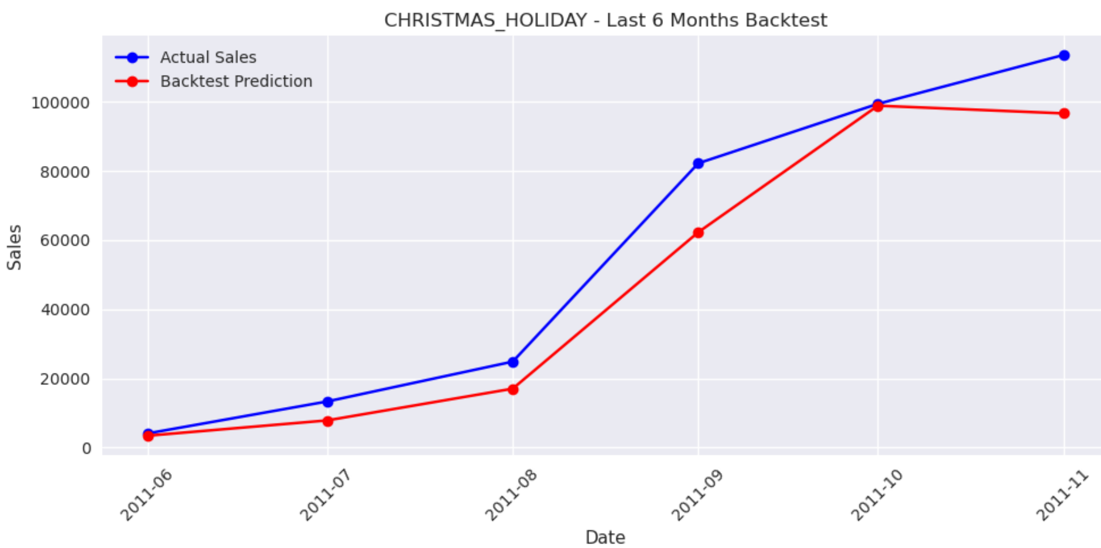
```
Starting backtesting for category-level forecasts…

Backtesting models for BEAUTY_PERSONAL…

Backtesting models for CHRISTMAS_HOLIDAY…

Backtesting models for FURNITURE_STORAGE…

Backtesting models for GARDEN_OUTDOOR…

Backtesting models for HOME_DECOR…

Backtesting models for KITCHEN_FOOD_UTENSIL…

Backtesting models for STATIONERY_OFFICE…
```

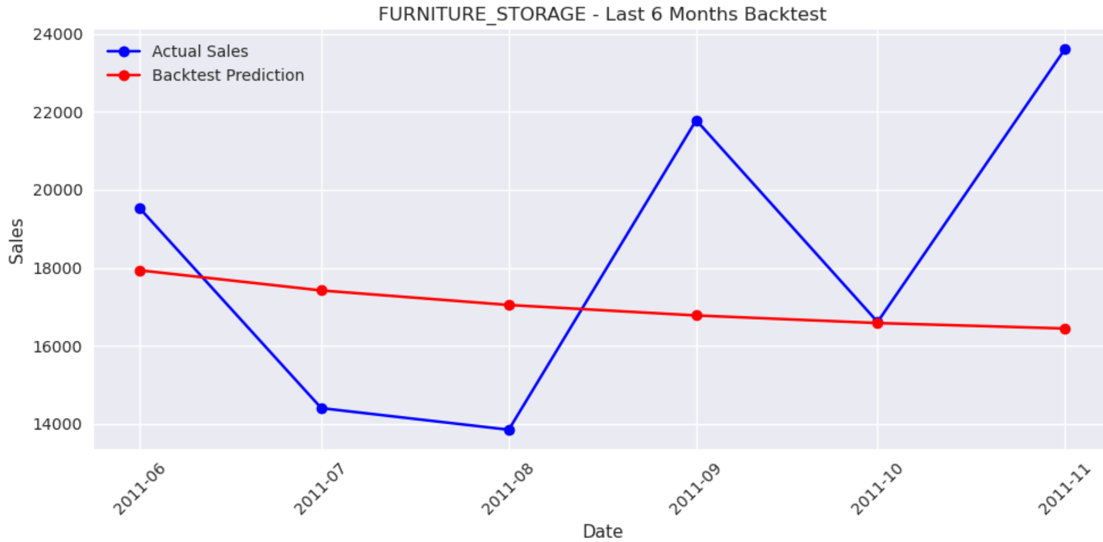
```
Backtesting models for TEXTILES_CLOTHING…

Backtesting models for TOYS_GAMES…

================================================================================
CATEGORY-LEVEL BACKTESTING RESULTS
================================================================================

BEAUTY_PERSONAL:
---------------
  Seasonal_Naive: MAE = 4358.28, RMSE = 5242.52
  ARIMA: MAE = 19078.90, RMSE = 20671.86
  SARIMA: MAE = 9074.28, RMSE = 10899.84
  → Best Model: Seasonal_Naive

CHRISTMAS_HOLIDAY:
-----------------
  Seasonal_Naive: MAE = 8555.65, RMSE = 11366.45
  ARIMA: MAE = 53545.49, RMSE = 69020.92
  SARIMA: MAE = 9189.54, RMSE = 15478.41
  → Best Model: Seasonal_Naive

FURNITURE_STORAGE:
-----------------
  Seasonal_Naive: MAE = 4393.51, RMSE = 4829.16
  ARIMA: MAE = 3342.05, RMSE = 3661.43
  SARIMA: MAE = 8896.80, RMSE = 11086.32
  → Best Model: ARIMA

GARDEN_OUTDOOR:
--------------
  Seasonal_Naive: MAE = 7519.83, RMSE = 10057.29
  ARIMA: MAE = 4827.67, RMSE = 5290.83
  SARIMA: MAE = 5418.63, RMSE = 6760.25
  → Best Model: ARIMA

HOME_DECOR:
----------
  Seasonal_Naive: MAE = 27187.35, RMSE = 35569.19
  ARIMA: MAE = 84450.02, RMSE = 112682.84
  SARIMA: MAE = 45852.92, RMSE = 47563.70
  → Best Model: Seasonal_Naive

KITCHEN_FOOD_UTENSIL:
--------------------
  Seasonal_Naive: MAE = 15732.64, RMSE = 22306.22
  ARIMA: MAE = 20284.97, RMSE = 22256.43
```

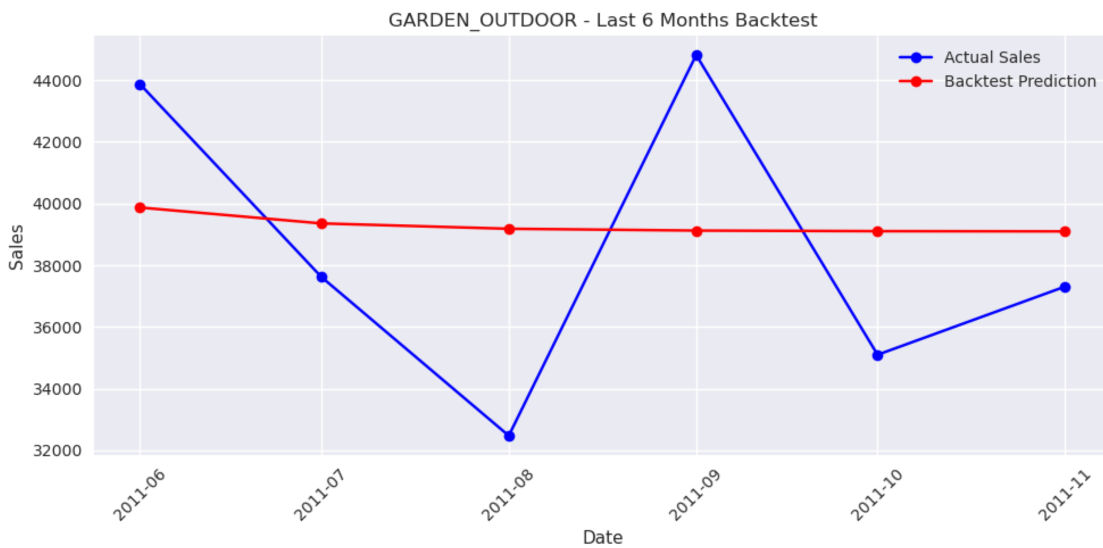
```
  SARIMA: MAE = 36728.45, RMSE = 49948.05
  → Best Model: Seasonal_Naive

STATIONERY_OFFICE:
-----------------
  Seasonal_Naive: MAE = 6923.63, RMSE = 8301.68
  ARIMA: MAE = 8313.95, RMSE = 9258.51
  SARIMA: MAE = 6486.50, RMSE = 7418.53
  → Best Model: SARIMA

TEXTILES_CLOTHING:
-----------------
  Seasonal_Naive: MAE = 8716.69, RMSE = 11455.55
  ARIMA: MAE = 8584.82, RMSE = 9774.54
  SARIMA: MAE = 5982.74, RMSE = 8099.43
  → Best Model: SARIMA

TOYS_GAMES:
----------
  Seasonal_Naive: MAE = 3644.07, RMSE = 4257.31
  ARIMA: MAE = 7876.60, RMSE = 10317.89
  SARIMA: MAE = 6306.59, RMSE = 8232.25
  → Best Model: Seasonal_Naive

============================================================
CATEGORY MODEL SELECTION SUMMARY
============================================================

              Category      Best_Model           MAE          RMSE
0       BEAUTY_PERSONAL  Seasonal_Naive   4358.280000   5242.517454
1     CHRISTMAS_HOLIDAY  Seasonal_Naive   8555.650000  11366.445325
2     FURNITURE_STORAGE           ARIMA   3342.046667   3661.430451
3        GARDEN_OUTDOOR           ARIMA   4827.675000   5290.834028
4            HOME_DECOR  Seasonal_Naive  27187.350000  35569.189466
5  KITCHEN_FOOD_UTENSIL  Seasonal_Naive  15732.635000  22306.222123
6     STATIONERY_OFFICE          SARIMA   6486.498139   7418.528749
7     TEXTILES_CLOTHING          SARIMA   5982.739082   8099.429870
8            TOYS_GAMES  Seasonal_Naive   3644.071667   4257.310219


================================================================================
GENERATING CATEGORY-LEVEL FORECASTS
================================================================================

Generating forecast for BEAUTY_PERSONAL using Seasonal_Naive…

Generating forecast for CHRISTMAS_HOLIDAY using Seasonal_Naive…

Generating forecast for FURNITURE_STORAGE using ARIMA…
```

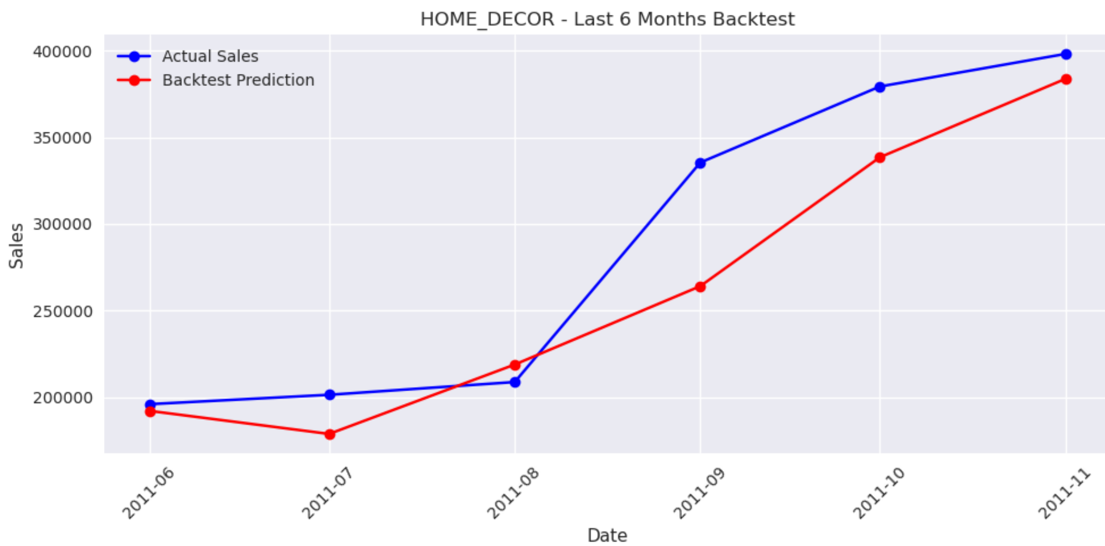
```
Generating forecast for GARDEN_OUTDOOR using ARIMA…

Generating forecast for HOME_DECOR using Seasonal_Naive…

Generating forecast for KITCHEN_FOOD_UTENSIL using Seasonal_Naive…

Generating forecast for STATIONERY_OFFICE using SARIMA…

Generating forecast for TEXTILES_CLOTHING using SARIMA…

Generating forecast for TOYS_GAMES using Seasonal_Naive…

Forecast generation completed!

================================================================================
FORECAST SUMMARY (Next 6 Months)
================================================================================

               Category           Model  Avg_Monthly_Forecast  \
0        BEAUTY_PERSONAL  Seasonal_Naive          39485.875000
1      CHRISTMAS_HOLIDAY  Seasonal_Naive           8584.026667
2      FURNITURE_STORAGE           ARIMA          20840.607063
3         GARDEN_OUTDOOR           ARIMA          37398.103908
4             HOME_DECOR  Seasonal_Naive         203550.908333
5   KITCHEN_FOOD_UTENSIL  Seasonal_Naive         112201.500000
6      STATIONERY_OFFICE          SARIMA          55593.197722
7      TEXTILES_CLOTHING          SARIMA          99509.019875
8             TOYS_GAMES  Seasonal_Naive          18807.631667

   Total_6Month_Forecast  Next_Month_Forecast
0           2.369152e+05         66356.730000
1           5.150416e+04         42752.850000
2           1.250436e+05         22155.427164
3           2.243886e+05         37375.148349
4           1.221305e+06        330526.200000
5           6.732090e+05        155294.250000
6           3.335592e+05         53133.415099
7           5.970541e+05         98289.893237
8           1.128458e+05         21935.000000
```

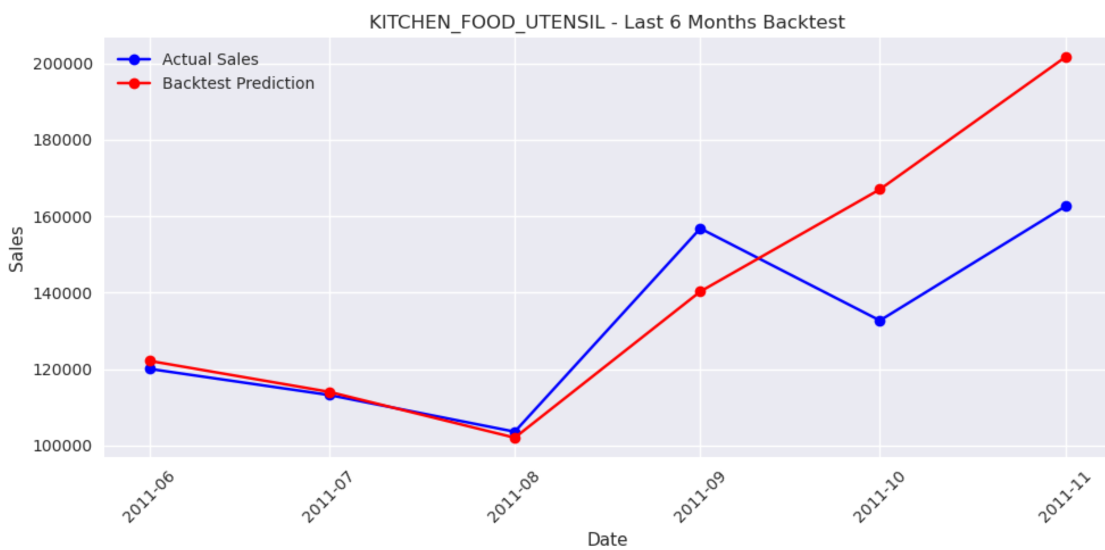

BEAUTY_PERSONAL - Last 6 Months Backtest
Actual Sales
Backtest Prediction
Sales
80000
70000
60000
50000
40000
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
Date

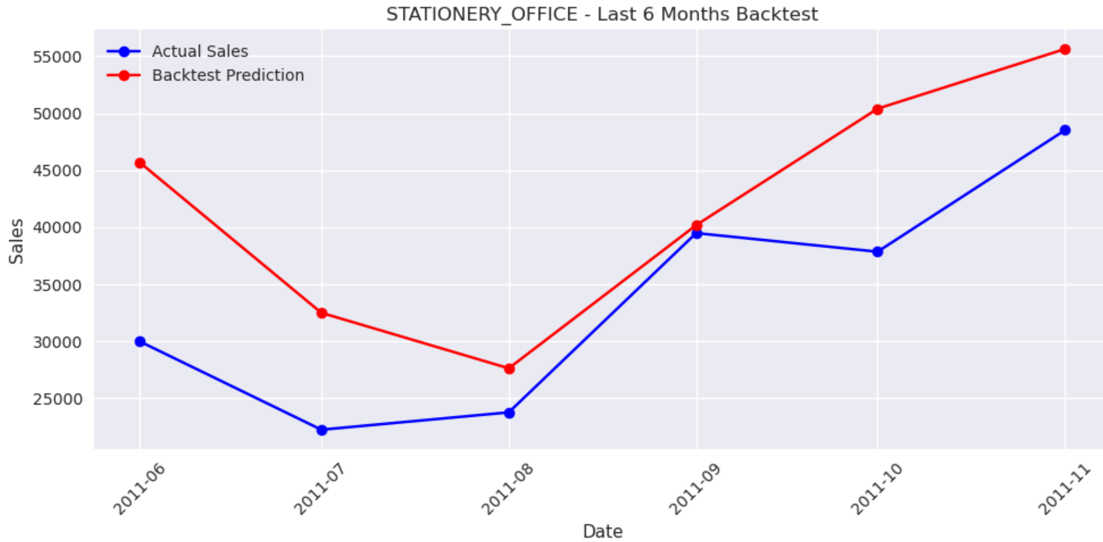

CHRISTMAS_HOLIDAY - Last 6 Months Backtest
Actual Sales
Backtest Prediction
Sales
100000
80000
60000
40000
20000
0
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
Date


FURNITURE_STORAGE - Last 6 Months Backtest
Actual Sales
Backtest Prediction
24000
22000
20000
18000
16000
14000
Sales
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
Date

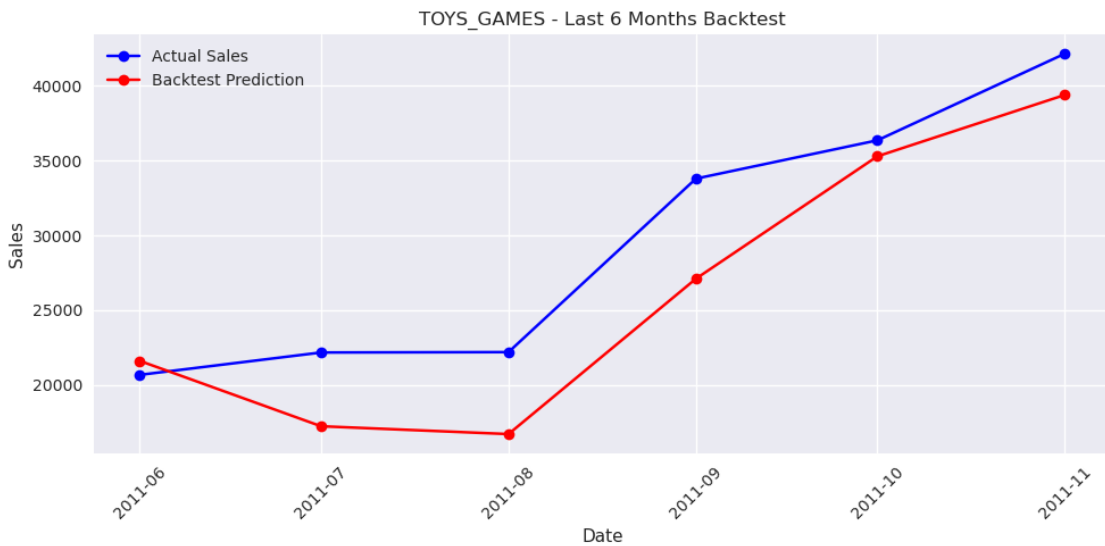

GARDEN_OUTDOOR - Last 6 Months Backtest
Actual Sales
Backtest Prediction
44000
42000
40000
38000
36000
34000
32000
Sales
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
Date


HOME_DECOR - Last 6 Months Backtest
Actual Sales
Backtest Prediction
Sales
400000
350000
300000
250000
200000
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
Date


KITCHEN_FOOD_UTENSIL - Last 6 Months Backtest
Actual Sales
Backtest Prediction
Sales
200000
180000
160000
140000
120000
100000
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
Date


STATIONERY_OFFICE - Last 6 Months Backtest
Actual Sales
Backtest Prediction
Sales
Date
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
25000
30000
35000
40000
45000
50000
55000

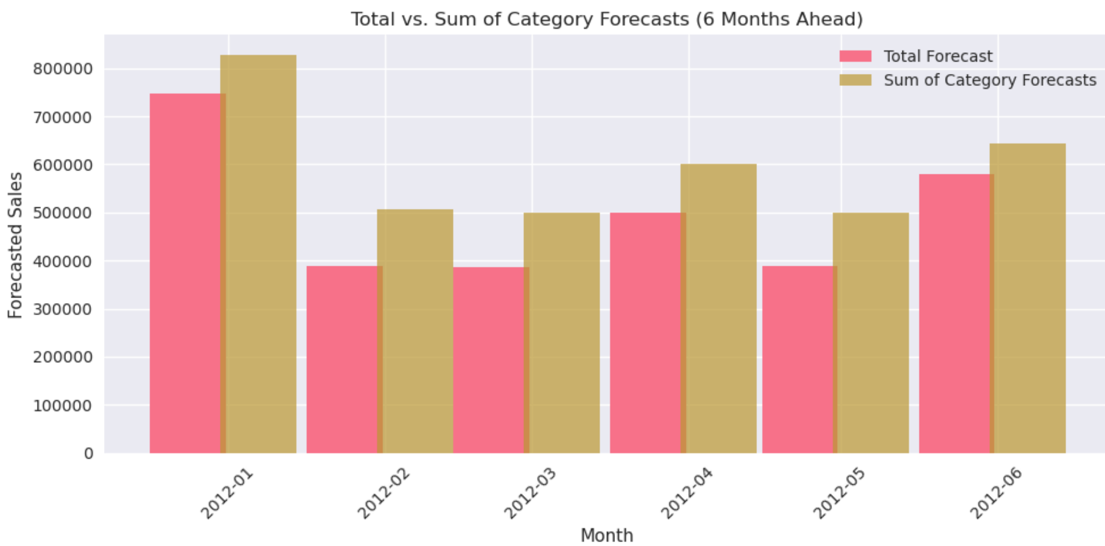

TEXTILES_CLOTHING - Last 6 Months Backtest
Actual Sales
Backtest Prediction
Sales
Date
2011-06
2011-07
2011-08
2011-09
2011-10
2011-11
55000
60000
65000
70000
75000
80000
85000



4.3 Reconciliation Check (recommended)

Compare sum of category forecasts to total; if mismatch material → top-down proportional scaling with guardrails

```
Comparison of Total Sales Forecast vs. Sum of Category Forecasts (Next 6
Months):
            Total_Forecast  Sum_Category_Forecast     Difference
date
2011-12-31       746458.82          827818.913849  -81360.093849
2012-01-31       388387.43          505941.688181 -117554.258181
2012-02-29       386198.68          498902.043749 -112703.363749
2012-03-31       500616.20          600851.855082 -100235.655082
2012-04-30       389822.19          499440.097275 -109617.907275
2012-05-31       578931.50          642870.623270  -63939.123270
```




```
Maximum absolute difference: $117,554.26
Warning: The sum of category forecasts differs materially from the total
forecast. Consider reconciliation.
```

The sum of category forecast is more optimistic but follow the same pattern. So here we have two model that can predict optimistic and pesimistic scenario. For recommendation we will follow pesimistic seenario (Total Forecast)

5 Level 3 - Top SKU Forecast + Long Tail Allocation

Why: Purchase orders happen at SKU; most SKUs are noisy → focus on the few that matter.

5.1 Identify Top SKUs

```
    product_category product_id  total_amount  cum_revenue  cum_revenue_pct
0     BEAUTY_PERSONAL      21915      23903.65     23903.65         0.042538
1     BEAUTY_PERSONAL      22961      18789.20     42692.85         0.075975
2     BEAUTY_PERSONAL      84347      17033.24     59726.09         0.106287
3     BEAUTY_PERSONAL     84029E      15320.11     75046.20         0.133550
4     BEAUTY_PERSONAL    15056BL      12063.66     87109.86         0.155019
..               …          …            …          …               …
883        TOYS_GAMES      21918       2621.55    223254.41         0.766024
884        TOYS_GAMES      23390       2539.40    225793.81         0.774737
885        TOYS_GAMES      22175       2212.40    228006.21         0.782328
886        TOYS_GAMES      22380       1986.48    229992.69         0.789144
887        TOYS_GAMES     75049L       1931.15    231923.84         0.795770

[888 rows x 5 columns]
```

```
       product_category  top_sku_revenue_pct
0        BEAUTY_PERSONAL             0.799018
1      CHRISTMAS_HOLIDAY             0.796723
2      FURNITURE_STORAGE             0.797587
3         GARDEN_OUTDOOR             0.797316
4             HOME_DECOR             0.799639
5   KITCHEN_FOOD_UTENSIL             0.799520
6      STATIONERY_OFFICE             0.799704
7      TEXTILES_CLOTHING             0.796192
8             TOYS_GAMES             0.795770
```

```
product_category
BEAUTY_PERSONAL         124
CHRISTMAS_HOLIDAY        59
FURNITURE_STORAGE        23
GARDEN_OUTDOOR           57
HOME_DECOR              309
KITCHEN_FOOD_UTENSIL    166
STATIONERY_OFFICE        69
TEXTILES_CLOTHING        42
TOYS_GAMES               39
Name: product_id, dtype: int64
```

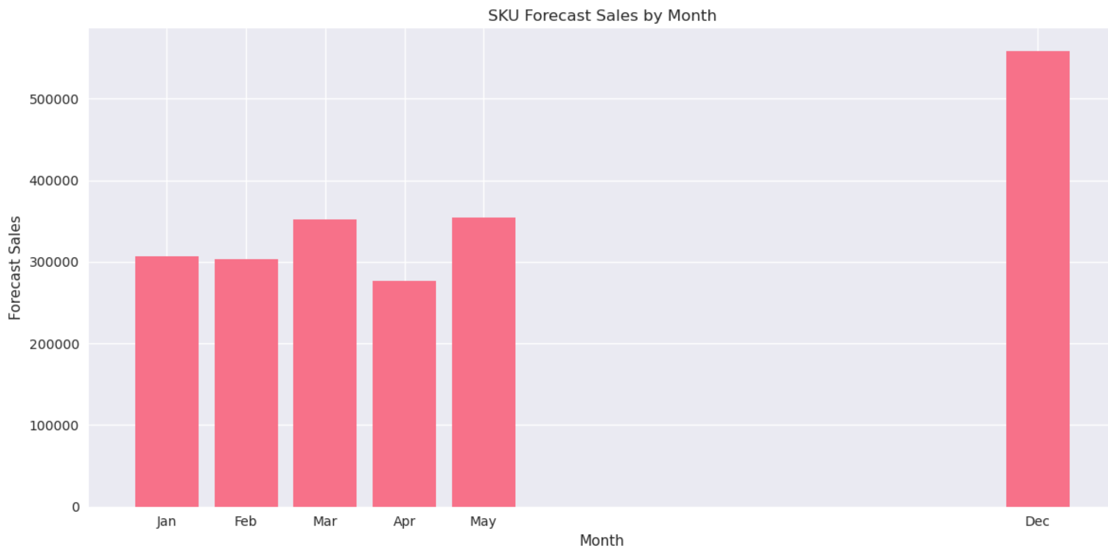
```
product_category
BEAUTY_PERSONAL          796
CHRISTMAS_HOLIDAY        224
FURNITURE_STORAGE         93
GARDEN_OUTDOOR           379
HOME_DECOR              1721
KITCHEN_FOOD_UTENSIL     813
STATIONERY_OFFICE        278
TEXTILES_CLOTHING        300
TOYS_GAMES               151
Name: product_id, dtype: int64
```

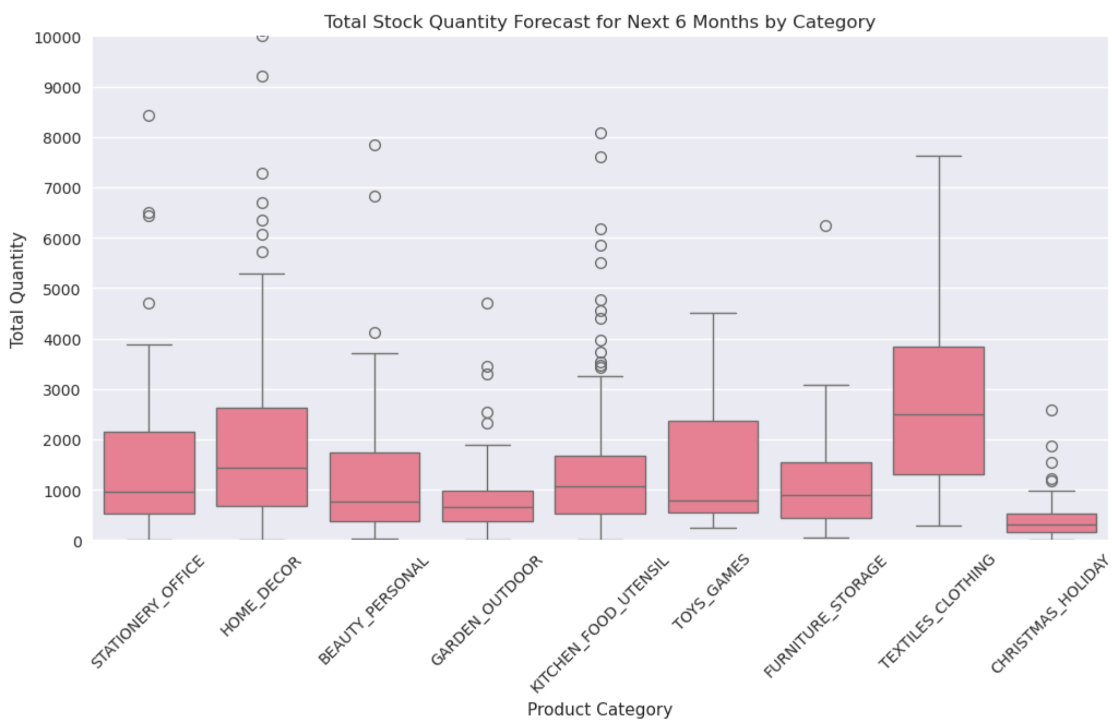
5.2 Forecast each SKU for the next 6 months

```
Forecasting each SKU for the next 6 months: 2011-12 to 2012-05
Could not forecast SKU 22795 in STATIONERY_OFFICE: too many indices for array:
array is 0-dimensional, but 1 were indexed
```

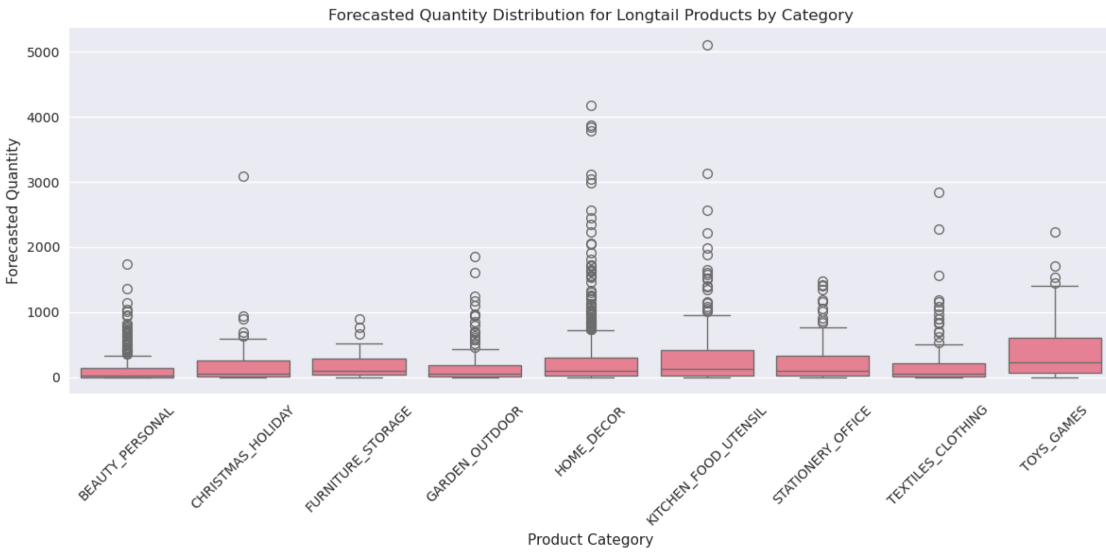
	product_id	product_category	forecast_month	forecast_sales
0	21915	BEAUTY_PERSONAL	2011-12-31	4361.16
1	21915	BEAUTY_PERSONAL	2012-01-31	679.40
2	21915	BEAUTY_PERSONAL	2012-02-29	783.95
3	21915	BEAUTY_PERSONAL	2012-03-31	1490.60
4	21915	BEAUTY_PERSONAL	2012-04-30	1403.12
…	…	…	…	…
3859	75049L	TOYS_GAMES	2012-01-31	33.75

```
3860      75049L      TOYS_GAMES     2012-02-29          38.75
3861      75049L      TOYS_GAMES     2012-03-31         106.60
3862      75049L      TOYS_GAMES     2012-04-30         195.70
3863      75049L      TOYS_GAMES     2012-05-31         563.35

[3864 rows x 4 columns]
```





5.3 Longtail analysis (how much product to keep for longtail products)



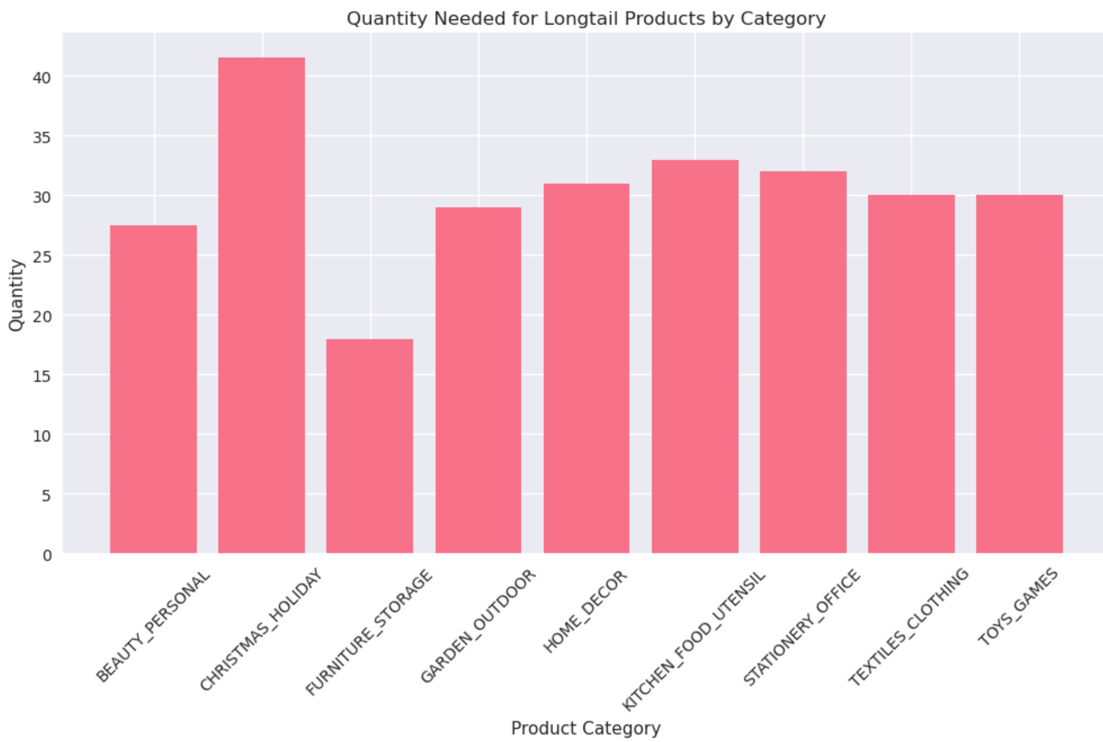

```
Median quantity of longtail products by category and month:

    product_category month_year  quantity
0     BEAUTY_PERSONAL    2009-12       3.0
1     BEAUTY_PERSONAL    2010-01       4.0
2     BEAUTY_PERSONAL    2010-02       3.0
3     BEAUTY_PERSONAL    2010-03       4.0
4     BEAUTY_PERSONAL    2010-04       4.0
..                 …          …        …
211        TOYS_GAMES    2011-07       3.0
212        TOYS_GAMES    2011-08       6.0
213        TOYS_GAMES    2011-09       6.0
214        TOYS_GAMES    2011-10       4.0
215        TOYS_GAMES    2011-11       3.0

[216 rows x 3 columns]
```

We can see that longtail products have fewer quantities sold compared to the top-selling products. We can generalize the quantity needed using a median value for each product category, to optimizing inventory levels for these items.

Since we know its seasonal naive then we can use the data from 2011 as a base line

```
Median quantity of longtail products by category and month:

    product_category month_year  quantity
0     BEAUTY_PERSONAL    2009-12       3.0
1     BEAUTY_PERSONAL    2010-01       4.0
2     BEAUTY_PERSONAL    2010-02       3.0
3     BEAUTY_PERSONAL    2010-03       4.0
```

```
4    BEAUTY_PERSONAL    2010-04       4.0
..               …         …        …
211        TOYS_GAMES    2011-07       3.0
212        TOYS_GAMES    2011-08       6.0
213        TOYS_GAMES    2011-09       6.0
214        TOYS_GAMES    2011-10       4.0
215        TOYS_GAMES    2011-11       3.0

[216 rows x 3 columns]
```



6 Optimization Impact Analysis

This section analyzes the impact of our optimization strategy by comparing: 1. Revenue Analysis: - Calculated revenue from Pareto products (individual quantities) and Longtail products (standardized quantities) - Comparison with the same period from previous year (2010-12 to 2011-05) 2. Inventory Optimization: - Total quantity analysis: Pareto + Longtail products - Comparison with the same period from previous year (2010-12 to 2011-05)

6.1 Measuring Revenue and Inventory Optimization

This section evaluates how much we optimize revenue and inventory by focusing on Pareto (top) and Longtail SKUs. We compare the forecasted results for the next 6 months with the actuals from the same period in the previous year (2010-12 to 2011-05).

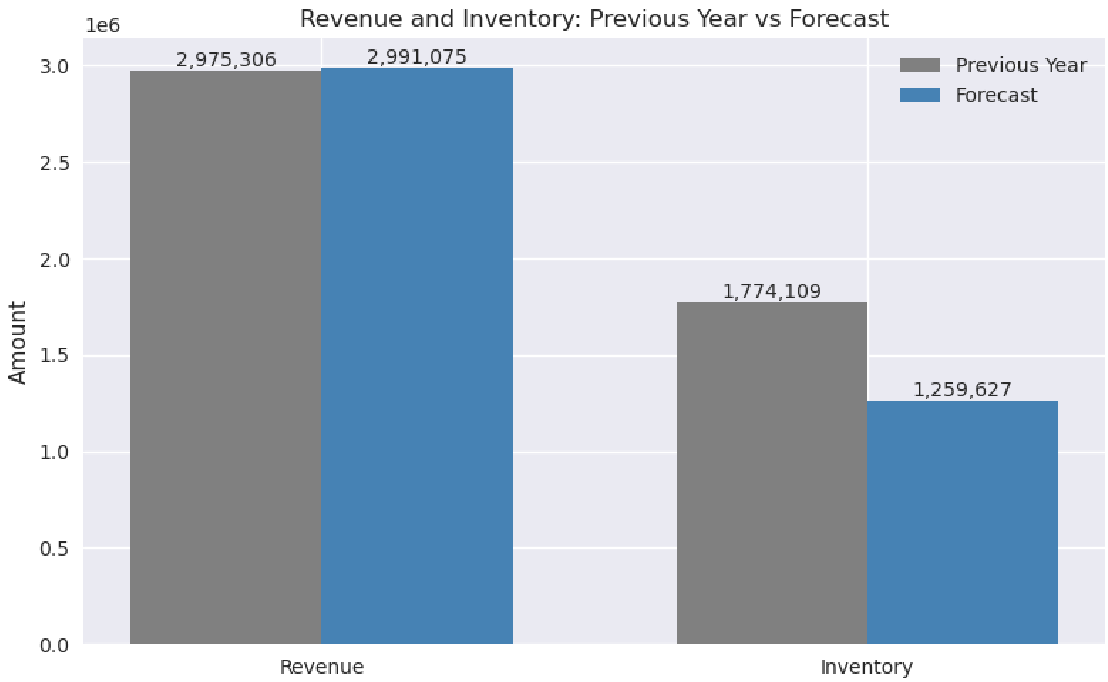
- **Revenue:**
  1. Calculate revenue from Pareto SKUs (each with its own forecasted quantity) and Longtail SKUs (each with the same average forecasted quantity).
  2. Compare to total revenue from the same period last year.
- **Inventory:**
  1. Calculate total forecasted quantity for Pareto and Longtail SKUs.
  2. Compare to total quantity from the same period last year.

```
    product_id  product_category  total_quantity
0        10135  STATIONERY_OFFICE      754.993536
1        15036         HOME_DECOR    10239.626667
2       15044D  STATIONERY_OFFICE      543.947499
3      15056BL    BEAUTY_PERSONAL      783.764706
4       15056N     GARDEN_OUTDOOR     1574.760671
..         …                 …               …
639     85123A         HOME_DECOR    10009.966102
640     85132C         HOME_DECOR      378.051282
641      85150         HOME_DECOR     2094.952941
642      85152         HOME_DECOR     2754.542857
643     85184C         HOME_DECOR      701.474576

[629 rows x 3 columns]
```

```
Forecasted Revenue (Pareto): 2449544.0142629836
Forecasted Revenue (Longtail): 416562.21
Total Forecasted Revenue: 2866106.2242629835
Previous Year Revenue: 2975305.5499999993
Revenue Optimization (%): -3.670188621031404
Forecasted Inventory (Pareto): 1100496.876083287
Forecasted Inventory (Longtail): 122408.0
Total Forecasted Inventory: 1222904.876083287
Previous Year Inventory: 1774109
Inventory Optimization (%): -31.06934939830152
```

So it looks like that the inventory improve massively with the revenue optimization only down by -3.6 %, we can adjust the median value of the longtail product using multiplier since we still have a lot of inventory to spare!

```
Forecasted Revenue (Pareto): 2449544.0142629836
Forecasted Revenue (Longtail): 541530.873
Total Forecasted Revenue: 2991074.8872629837
Previous Year Revenue: 2975305.5499999993
Revenue Optimization (%): 0.5300073218693243
Forecasted Inventory (Pareto): 1100496.876083287
Forecasted Inventory (Longtail): 159130.40000000002
Total Forecasted Inventory: 1259627.2760832869
Previous Year Inventory: 1774109
Inventory Optimization (%): -28.999442757841436
```



We successfully maintain revenue while optimizing inventory levels!