Inventory Forecasting & Sales Analysis Project
Self-Initiated | Mar 2025

Project Overview

This project simulates a real-world inventory environment using synthetic retail sales data across 1,000 SKUs, 5 product categories, and 10 vendors over approximately one year. The objective is to analyze sales performance, identify inventory challenges, and build foundational forecasting insights to support inventory planning and replenishment decisions.

Dataset Summary

- **Data Source:** Simulated synthetic data
- **Structure:** ~12 months of weekly SKU-level sales
- **Fields:** SKU_ID, Category, Vendor, Week, Store_Location, Units_Sold, Unit_Cost, Lead_Time_Days
- **Total Records:** 52,000

Methodology & Analysis

1. Data Cleaning

- Removed missing `Units_Sold` and `Store_Location` values
- Standardized data types (e.g., categorical conversion, numeric fields)
- Exported cleaned dataset for analysis and modeling

```
# Loading datasets
df = pd.read_csv(r'C:\Users\brivi\Downloads\My Portfolio\inventory-forecasting-project\data\unstructured_sku_sales_data.csv')

# Previewing top 5 rows
df.head()
```

[7]:

	SKU_ID	Category	Vendor	Week	Store_Location	Units_Sold	Unit_Cost	Lead_Time_Days
0	SKU_0359	Electronics	Vendor_1	42	Store_10	18.0	75.95	29
1	SKU_0062	Toys	Vendor_10	51	Store_13	16.0	14.50	7
2	SKU_0458	Home & Kitchen	Vendor_1	38	Store_15	5.0	109.42	22
3	SKU_0346	Clothing	Vendor_7	35	Store_12	25.0	98.35	7
4	SKU_0188	Electronics	Vendor_5	23	Store_9	25.0	188.68	16

```
[9]: # Checking if there are missing values
df.isnull().sum()
```

```
[9]: SKU_ID               0
Category             0
Vendor               0
Week                 0
Store_Location    1040
Units_Sold        1040
Unit_Cost            0
Lead_Time_Days       0
dtype: int64
```

```
[11]: # Both the Store_Location and Units_Sold have same number of missing values: 1040
# Since 1040 missing values are significant portion of the total number of data
# We'll investigate further by chyecking if all 1040 rows belong to the same week/SKU
# Or if they're concentrated in a few vendors or categories
df[df['Units_Sold'].isnull()]['Vendor'].value_counts()
```

```
[11]: Vendor
Vendor_2     130
Vendor_3     114
Vendor_10    109
Vendor_1     108
Vendor_7     105
Vendor_9     105
Vendor_8     103
Vendor_5      99
Vendor_6      91
Vendor_4      76
Name: count, dtype: int64
```

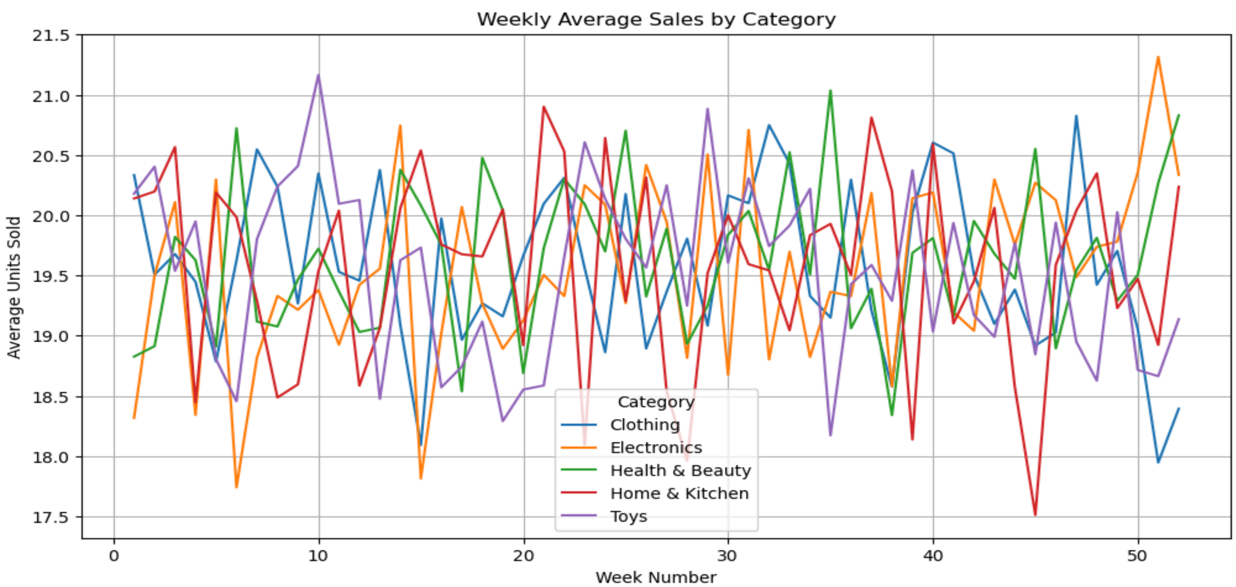
```
[46]: df = df.dropna(subset=['Store_Location', 'Units_Sold'])
df.isnull().sum()
```

```
[46]: SKU_ID            0
Category          0
Vendor            0
Week              0
Store_Location    0
Units_Sold        0
Unit_Cost         0
Lead_Time_Days    0
dtype: int64
```

2. Exploratory Data Analysis (EDA)

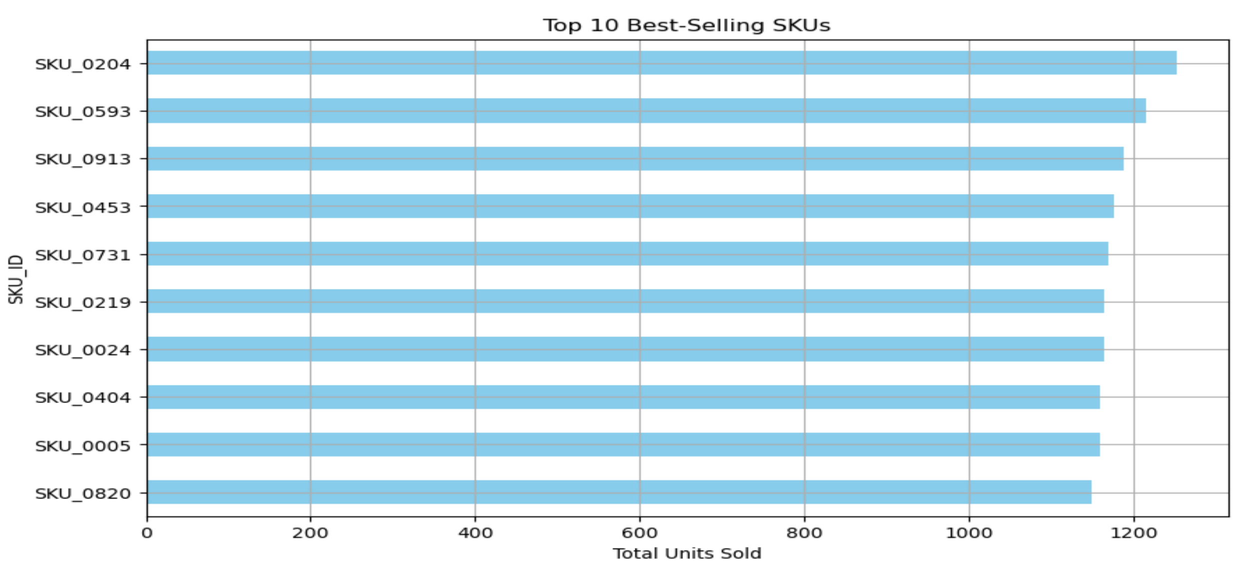
a. Weekly Average Sales by Category

Line plot analysis reveals consistent demand across all five categories with minor fluctuations, reflecting a fairly stable demand environment for weekly inventory planning.

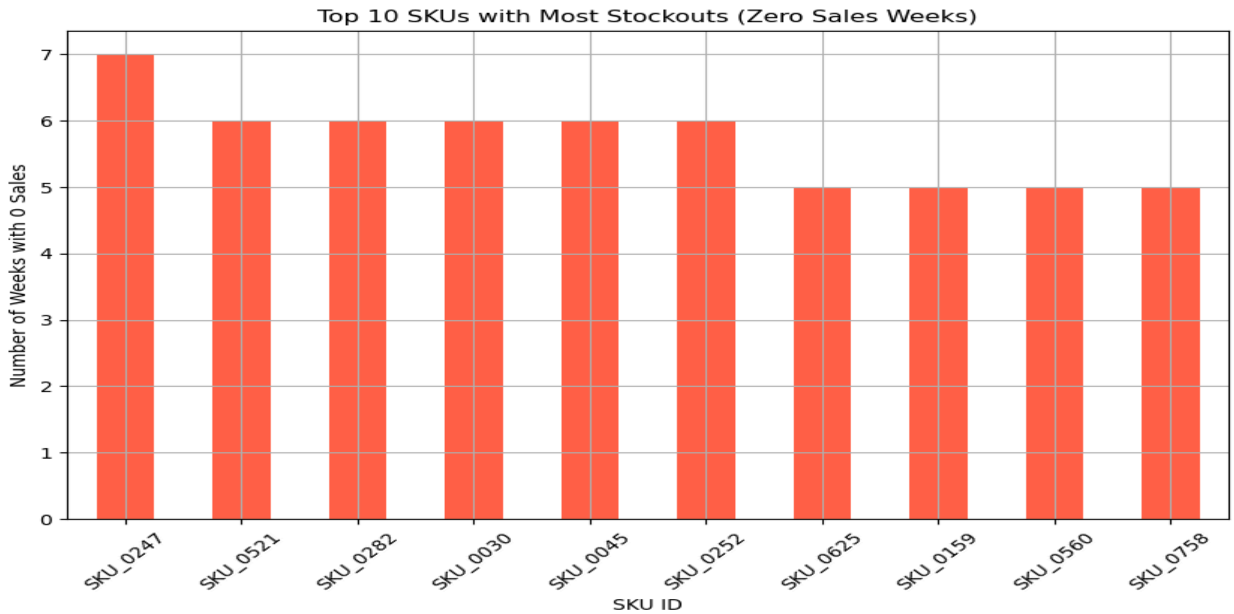


b. Top 10 Best-Selling SKUs

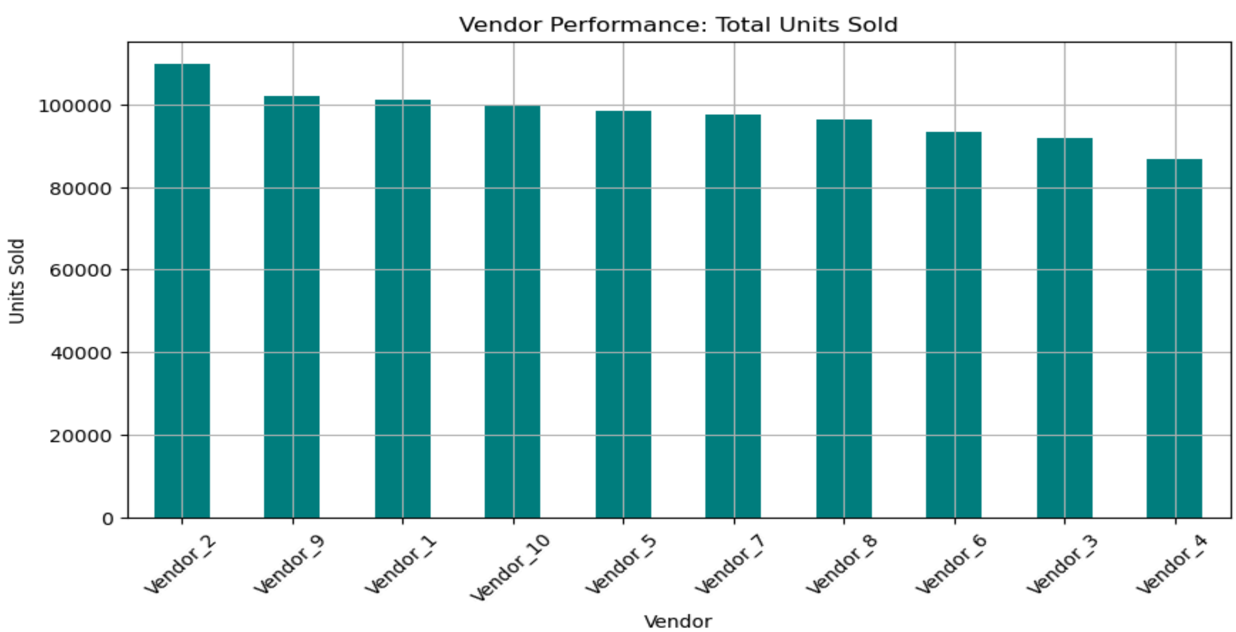
Horizontal bar chart shows that SKU_0204 led total annual sales, followed closely by other high-volume SKUs. These products should be prioritized in reorder strategies and monitored closely.



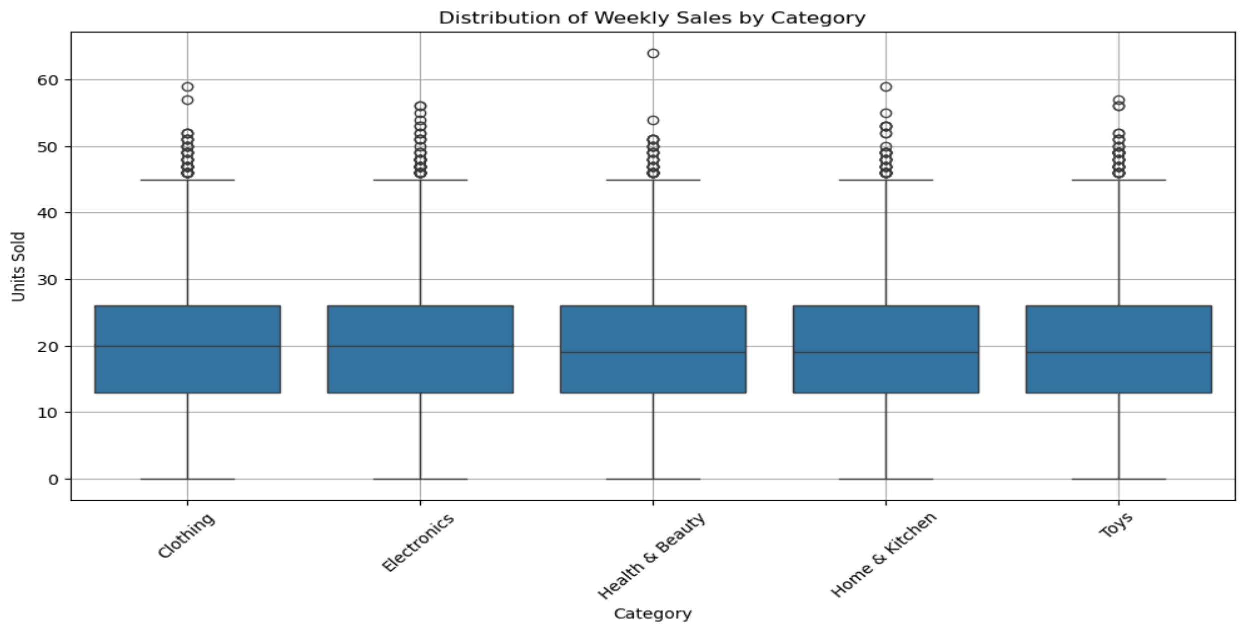
c. Stockout Frequency by SKU
Several SKUs, such as SKU_0247 and SKU_0521, had 5+ weeks of zero sales, suggesting stockout risks or poor demand. These require either better forecasting or strategic phasing out.



d. Vendor Performance Summary
Vendor_2 had the highest total units sold, followed by Vendor_9 and Vendor_1. Vendor performance evaluation helps prioritize partners for reliable replenishment and lead time planning.



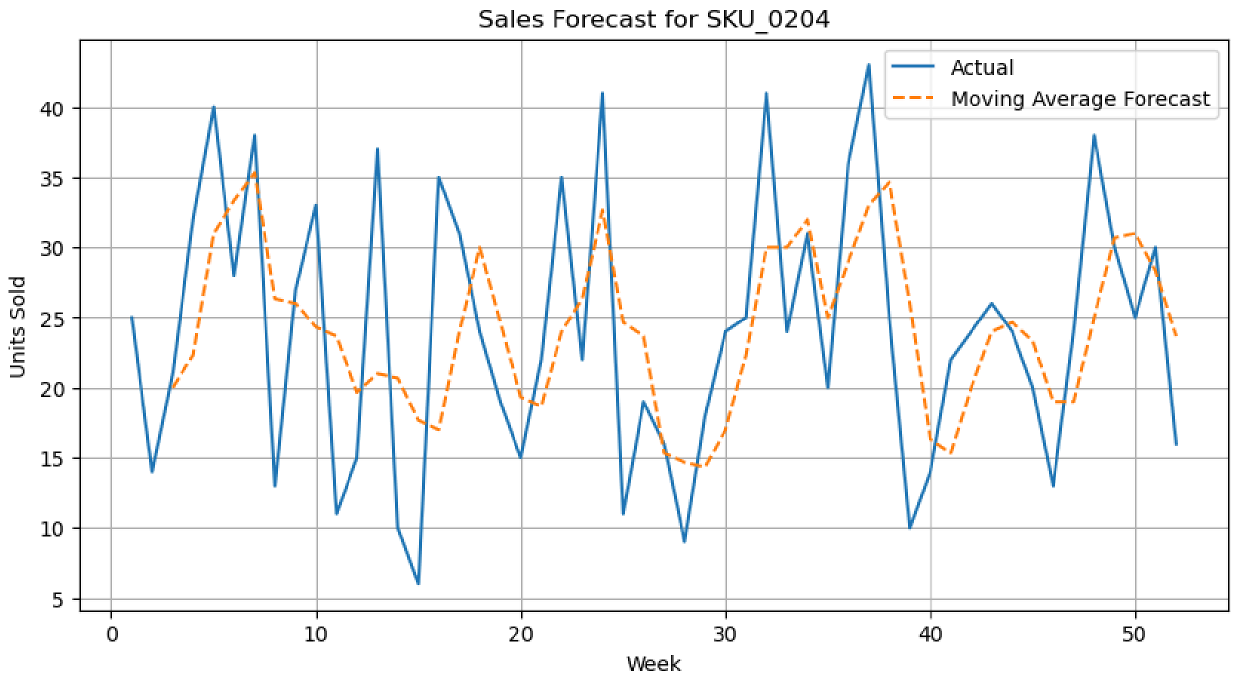
e. Weekly Sales Variability by Category
Box plot highlights that most categories have similar sales spread with noticeable outliers. Some categories like Health & Beauty showed higher variability, useful for safety stock considerations.



📊 Forecasting

✅ Model: 3-Week Moving Average

- Applied a 3-week rolling window to forecast weekly demand for SKU_0204
- This method smooths short-term fluctuations while retaining recent trend information
- **Evaluation Metrics:** MAE = 6.74, RMSE = 8.04
- Suitable for SKUs with relatively steady weekly demand and minimal seasonality



🎓 Skills Demonstrated

- Data Cleaning (pandas)
- Exploratory Data Analysis (matplotlib, seaborn)
- Forecasting with Moving Averages
- Vendor and SKU-level performance evaluation
- Visual storytelling and data interpretation

Contact

This project is part of my portfolio as an aspiring inventory planning and supply chain data analyst. Charts, notebooks, and dashboard files available upon request or GitHub repository.