

# Project Title: Monthly Sales Forecasting Using Time Series Analysis

## Introduction:

This project aims to forecast future monthly sales for a business using time series data. The goal is to provide accurate predictions that can help in decision-making, stock management, and overall business planning.

## Step 1: Data Collection and Initial Exploration

In the first step, I searched for a suitable dataset that contains monthly sales figures. Once collected, I performed initial cleaning and formatting of the data using the following tools:

# This dataset was used for the # of a project titled "Sales Forecasting and Demand Prediction Using Time Series Analysis" # at [redacted]																			
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit
0	1	CA-2016-12310	11/01/2016	Standard Class	CS-12320	Chen's Cafe	Consumer	United States	Portland	97201	South	FUR-RC-10001708	Furniture	Bookcases	Bookcase	261.9600	2	0.00	41.9118
1	2	CA-2016-12320	11/01/2016	Standard Class	CS-12320	Chen's Cafe	Consumer	United States	Portland	97201	South	FUR-CN-10000404	Furniture	Chairs	High Back Office Chair	731.9400	3	0.00	210.5200
2	3	CA-2016-12368	6/02/2016	Standard Class	CS-12368	Chen's Cafe	Consumer	United States	Los Angeles	90026	West	CPA-LS-10000100	Office Supplies	Labels	Self-Adhesive Address Labels for Envelopes 3x5	14.6200	2	0.00	4.6714
3	4	US-2015-10066	10/10/2015	Standard Class	SO-10015	Seon O'Donnell	Consumer	United States	Fort Lauderdale	33311	South	FUR-FL-10000277	Furniture	Tables	Breitbart CM450 Series Slim Rectangular Table	957.3775	5	0.45	-380.0110
4	5	US-2015-10066	10/10/2015	Standard Class	SO-10015	Seon O'Donnell	Consumer	United States	Fort Lauderdale	33311	South	OFF-ST-10000780	Office Supplies	Storage	Elkon Fold N Roll Cart System	22.3800	2	0.20	2.5194
5	6	CA-2014-11812	6/06/2014	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Los Angeles	90032	West	FUR-FU-10001467	Furniture	Furniture	Elkon Expressions Wood and Plastic Desk Accents	46.8900	7	0.00	14.1694
6	7	CA-2014-11812	6/06/2014	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Los Angeles	90032	West	OFF-AB-10000203	Office Supplies	Art	Neutral 320	2.2000	4	0.00	1.9404
7	8	CA-2014-11812	6/06/2014	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Los Angeles	90032	West	TEC-PH-10000275	Technology	Phones	Misc 3200 IP Phone VoIP Phone	101.1500	6	0.20	90.7102
8	9	CA-2014-11812	6/06/2014	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Los Angeles	90032	West	OFF-BB-10000393	Office Supplies	Binders	DRL Angle-View Binders with Locking Rings by S.	15.5400	3	0.20	5.7825
9	10	CA-2014-11812	6/06/2014	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Los Angeles	90032	West	OFF-AB-10000402	Office Supplies	Applications	Balco P42200100 6 Colored Large	114.9000	5	0.00	24.4100
10	11	CA-2014-11812	6/06/2014	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Los Angeles	90032	West	FUR-TA-10001539	Furniture	Tables	Chromacraft Rectangular Conference Table	1706.1540	9	0.20	85.3000
11	12	CA-2014-11812	6/06/2014	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Los Angeles	90032	West	TEC-PH-10000333	Technology	Phones	Korbell 250 Conference phone - Chemical black	911.4100	4	0.20	68.9108
12	13	CA-2017-114412	4/15/2017	Standard Class	AA-14880	Andrew Allen	Consumer	United States	Concord	94027	South	OFF-PA-10000265	Office Supplies	Paper	Bonus 1802	15.5300	3	0.20	1.4432
13	14	CA-2016-107108	12/01/2016	Standard Class	BA-11710	Breanna Hoffman	Consumer	United States	Seattle	98105	West	OFF-AB-10000402	Office Supplies	Binders	Fidellium P8300 Photo Comb Binding Machine	407.9740	3	0.20	102.3402
14	15	US-2015-110883	11/02/2015	Standard Class	HP-14815	Harold Redden	Home Office	United States	Fort Worth	76106	Central	OFF-AB-10000211	Office Supplies	Applications	Holmes Replacement Filter for HEPA Air Cleaner	68.0100	5	0.00	-13.0100
15	16	US-2015-110883	11/02/2015	Standard Class	HP-14815	Harold Redden	Home Office	United States	Fort Worth	76106	Central	OFF-BB-10000726	Office Supplies	Binders	Stores DataTech Recycled Plastic Fricited Binders	2.5400	3	0.00	-3.8100
16	17	CA-2014-140880	11/11/2014	Standard Class	PK-18075	Pete Kitz	Consumer	United States	Madison	53711	Central	OFF-ST-10000100	Office Supplies	Storage	Star D-Star Shelving Vertical 5-Shell 17 1/2 x 11	661.0000	6	0.00	13.1716
17	18	CA-2016-107104	5/01/2016	Standard Class	AS-10220	Algebra Jones	Consumer	United States	West Jordan	84064	West	OFF-ST-10000102	Office Supplies	Storage	Fidellium Replacement Filter for HEPA Air Cleaner	55.5000	2	0.00	5.9900
18	19	CA-2014-14338	8/27/2014	Standard Class	ZD-21925	Zachariah Donald	Consumer	United States	San Francisco	94109	West	OFF-AB-10000306	Office Supplies	Art	Neutral 341	6.5000	2	0.20	2.4624
19	20	CA-2014-14338	8/27/2014	Standard Class	ZD-21925	Zachariah Donald	Consumer	United States	San Francisco	94109	West	TEC-PH-10001949	Technology	Phones	Cisco SPA 5010 IP Phone	213.4000	3	0.20	16.0110
20	21	CA-2014-14338	8/27/2014	Standard Class	ZD-21925	Zachariah Donald	Consumer	United States	San Francisco	94109	West	OFF-BB-10000223	Office Supplies	Binders	Cisco SPA 5010 IP Phone	22.2200	4	0.20	2.0408
21	22	CA-2016-17330	12/01/2016	Standard Class	KB-16015	Keri Black	Consumer	United States	Freemont	88025	Central	OFF-AB-10000306	Office Supplies	Art	Wilson Jones Hanging View Binder, White, 11"	18.4600	7	0.00	1.5906
22	23	CA-2016-17330	12/01/2016	Standard Class	KB-16015	Keri Black	Consumer	United States	Freemont	88025	Central	OFF-AB-10001402	Office Supplies	Applications	Access On-Outlet Power Strip, 4' Cord Length	60.3400	7	0.00	15.6888
23	24	US-2017-116689	7/16/2017	Standard Class	SF-20005	Sandra Ferguson	Consumer	United States	Philadelphia	19140	East	FUR-CN-10000774	Furniture	Chairs	Global Deluxe Teaching Chair, Grey	71.3700	2	0.30	1.0106
24	25	CA-2015-10030	9/01/2015	Standard Class	EA-15810	Emily Barnes	Consumer	United States	Orion	48023	West	FUR-TA-10000177	Furniture	Tables	Breitbart CM450 Series Slim Rectangular Table	1044.0100	3	0.00	202.3400
25	26	CA-2016-17375	1/16/2016	Standard Class	EA-15810	Emily Barnes	Consumer	United States	Los Angeles	90049	West	OFF-BB-10001634	Office Supplies	Binders	Wilson Jones Active Use Binders	11.6400	2	0.20	4.2204
26	27	CA-2016-17375	1/16/2016	Standard Class	EA-15810	Emily Barnes	Consumer	United States	Los Angeles	90049	West	TEC-AC-10000327	Technology	Accessories	Innovation R20 Mini Traveler USB 2.0 Flash Drive	90.5700	3	0.00	11.7741
27	28	US-2015-150630	9/17/2015	Standard Class	TR-21520	Tracy Barmann	Consumer	United States	Philadelphia	19140	East	FUR-FU-10000404	Furniture	Bookcases	Rowlands Robin Royal Lanyard Bookcase, Royal	388.4300	7	0.30	160.0502
28	29	US-2015-150630	9/17/2015	Standard Class	TR-21520	Tracy Barmann	Consumer	United States	Philadelphia	19140	East	OFF-BB-10000404	Office Supplies	Binders	Heavy Recycled Paper View Binder for Binding 3x5	44.6100	2	0.70	2.0602
29	30	US-2015-150630	9/17/2015	Standard Class	TR-21520	Tracy Barmann	Consumer	United States	Philadelphia	19140	East	FUR-FU-10000404	Furniture	Furniture	Howard Miller 15 1/4" Diameter Braided Chrome	124.2000	3	0.20	15.5200
30	31	US-2015-150630	9/17/2015	Standard Class	TR-21520	Tracy Barmann	Consumer	United States	Philadelphia	19140	East	OFF-EN-10001308	Office Supplies	Envelopes	Poly Strong Ten Envelopes	3.2400	2	0.20	1.1016
31	32	US-2015-150630	9/17/2015	Standard Class	TR-21520	Tracy Barmann	Consumer	United States	Philadelphia	19140	East	OFF-AB-10000404	Office Supplies	Art	8020200 Model 1800 Electric Pencil Sharpener	86.3400	6	0.20	8.7000
32	33	US-2015-150630	9/17/2015	Standard Class	TR-21520	Tracy Barmann	Consumer	United States	Philadelphia	19140	East	OFF-BB-10001525	Office Supplies	Binders	Access Notebook Covers with Storage Pouch 14 x 10	4.4900	6	0.70	5.7700
33	34	US-2015-150630	9/17/2015	Standard Class	TR-21520	Tracy Barmann	Consumer	United States	Philadelphia	19140	East	OFF-AB-10001603	Office Supplies	Art	Lumber Copiers	15.7600	2	0.20	3.5400
34	35	CA-2017-107272	10/01/2017	Standard Class	MA-17500	Matt Alshouse	Home Office	United States	Houston	77005	Central	OFF-PA-10000349	Office Supplies	Paper	Easy single paper	25.4700	3	0.20	0.9408
35	36	CA-2016-17796	12/01/2016	First Class	GH-14425	Gina Hale	Consumer	United States	Richardson	75080	Central	TEC-PH-10000477	Technology	Phones	GT 30240014	1091.5400	7	0.20	122.4707
36	37	CA-2016-17796	12/01/2016	First Class	GH-14425	Gina Hale	Consumer	United States	Richardson	75080	Central	FUR-FU-10000364	Furniture	Furniture	Electric Archibut's Camp On-Turner Area Lamp	190.9300	5	0.00	147.9600
37	38	CA-2015-114151	12/01/2015	Standard Class	SA-20710	Steve Higgins	Home Office	United States	Houston	77041	Central	OFF-EN-10000286	Office Supplies	Envelopes	#10.4 1/2" x 9 1/2" Premium Diagonal Seam Enve	113.3200	9	0.20	35.4100
38	39	CA-2015-114151	12/01/2015	Standard Class	SA-20710	Steve Higgins	Home Office	United States	Houston	77041	Central	FUR-RC-10000245	Furniture	Bookcases	Atlanta Metals Midsize 3-Shell Bookcases, Conch	532.3900	3	0.30	46.9104
39	40	CA-2015-114151	12/01/2015	Standard Class	SA-20710	Steve Higgins	Home Office	United States	Houston	77041	Central	FUR-RC-10000245	Furniture	Chairs	Global Deluxe Manager's Chair, Dark Grey	214.0500	3	0.30	15.5400
40	41	CA-2015-114151	12/01/2015	Standard Class	SA-20710	Steve Higgins	Home Office	United States	Houston	77041	Central	TEC-PH-10000406	Technology	Phones	Phonetics 10.10 Handset Offer	371.1400	4	0.20	41.7904
41	42	CA-2017-10099	9/10/2017	Standard Class	LC-18830	Linda Casasas	Consumer	United States	Hopewell	80540	Central	TEC-PH-10000406	Technology	Phones	Phonetics 10.10 Handset Offer	371.1400	4	0.20	16.5904

- **Pandas:** for data manipulation and cleaning.
- **Matplotlib & Seaborn:** for visualizing the data trends and patterns.

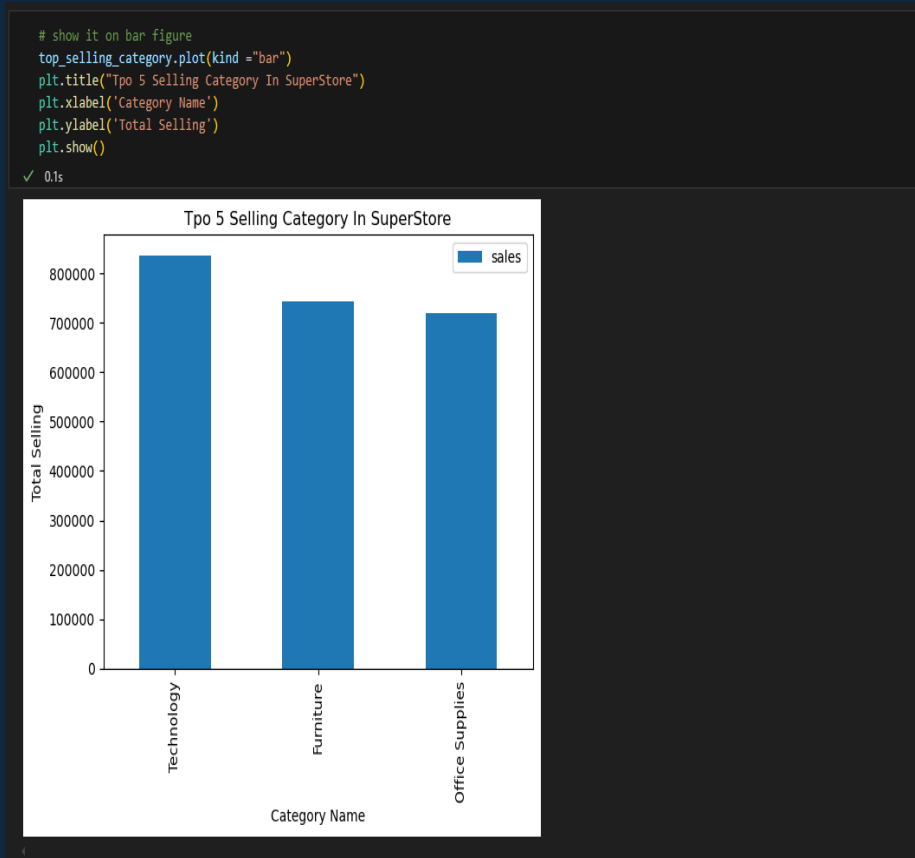
## Tasks:

- Converted the date column to datetime format.
- Aggregated daily data into monthly data using .resample('M').
- Checked for null values and handled them appropriately.

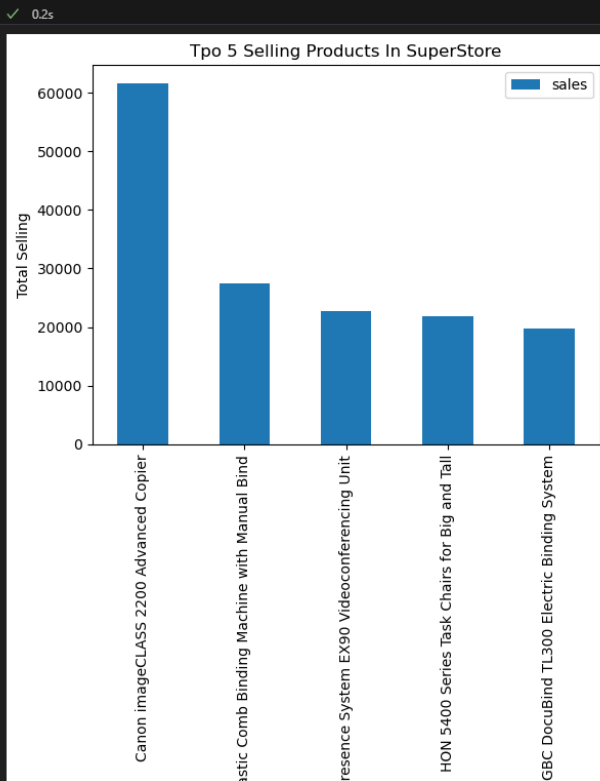
## Step 2: Exploratory Data Analysis (EDA)

The EDA phase helped me extract insights from the dataset and formulate analytical questions such as:

- What are the peak months in terms of sales?
- Are there any seasonal trends?
- Are there outliers or unusual patterns?



```
# show it on bar figure
top_5_selling_products.plot(kind="bar")
plt.title("Top 5 Selling Products In SuperStore")
plt.xlabel('Product Name')
plt.ylabel('Total Selling')
plt.show()
```



### Libraries Used:

- **Matplotlib & Seaborn:** for plotting sales trends and distributions.
- **Statsmodels:** for conducting stationarity tests like ADF (Augmented Dickey-Fuller).

### Insights:

- There is a clear upward trend in sales.
- Some seasonality might exist in the data.

---

### Step 3: Deep Analysis

In this phase, I examined the statistical properties of the time series.

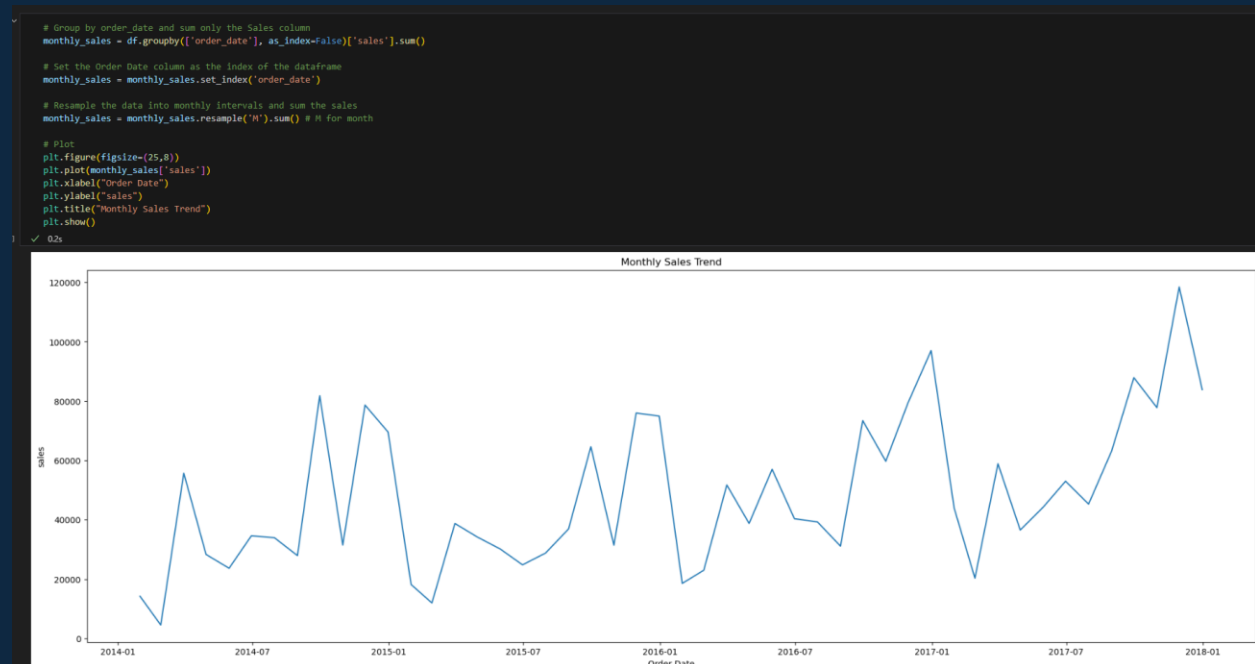
#### Tasks:

- Conducted ADF Test → confirmed the series is non-stationary.
- Differenced the series to make it stationary.

- Used `auto_arima()` to identify the best parameters.

#### Libraries Used:

- **pmdarima:** for `auto_arima()` model selection.
- **Statsmodels:** for SARIMAX modeling and diagnostics.



#### Insights:

- Best model suggested: **ARIMA(0,1,1)**

---

### Step 4: Model Building and Evaluation

Based on the data nature, I concluded that the suitable models should be from the ARIMA family since:

- The data is univariate.
- Exhibits trend and possibly some seasonality.

#### Models Tried:

- ARIMA(1,1,0)
- ARIMA(0,1,1)
- SARIMA(1,1,1,12)

## Evaluation Criteria:

- AIC, BIC, RMSE values
- Residual diagnostics

## Best Model:

- **ARIMA(0,1,1)** had the lowest AIC and stable residuals.

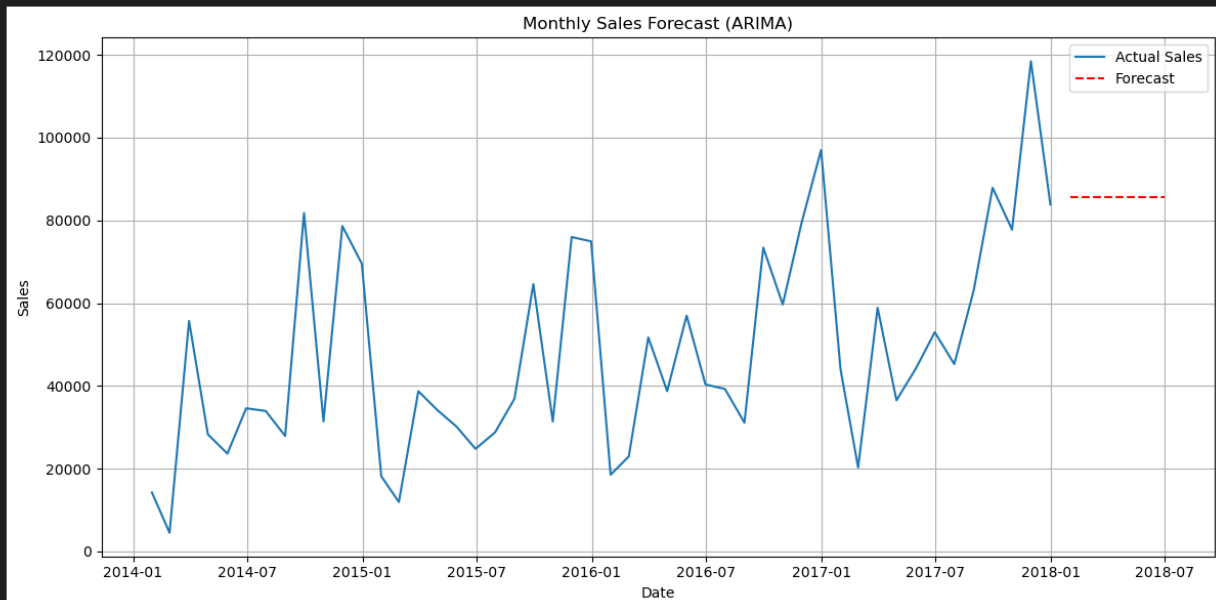
## Libraries Used:

- **statsmodels.tsa.statespace.SARIMAX**

```
SARIMAX Results
=====
Dep. Variable:      Sales      No. Observations:      48
Model:              ARIMA(0, 1, 1)  Log Likelihood:      -540.306
Date:               Thu, 15 May 2025  AIC:                  1084.612
Time:               15:33:18      BIC:                  1088.313
Sample:             01-31-2014    HQIC:                 1086.005
                        - 12-31-2017
Covariance Type:    opg
=====
              coef      std err          z      P>|z|      [0.025      0.975]
-----
ma.L1          -0.6534         0.123      -5.315      0.000      -0.894      -0.412
sigma2      5.981e+08      2.21e-13      2.7e+21      0.000      5.98e+08      5.98e+08
=====
Ljung-Box (L1) (Q):                0.02      Jarque-Bera (JB):                0.15
Prob(Q):                           0.88      Prob(JB):                  0.93
Heteroskedasticity (H):              1.05      Skew:                      -0.00
Prob(H) (two-sided):                0.92      Kurtosis:                  2.73
=====
```

### Warnings:

- [1] Covariance matrix calculated using the outer product of gradients (complex-step).
- [2] Covariance matrix is singular or near-singular, with condition number 5.75e+39. Standard errors may be unstable.



## Step 5: Dashboard Integration

The final model outputs and sales forecasts will be integrated into a dashboard that allows stakeholders to:

- View historical vs. forecasted sales.
- Interact with monthly breakdowns.
- Understand sales patterns easily.

### Tools to Be Used (Suggestion):

- **Power BI** or **Tableau** for the dashboard.
- **Plotly Dash** for a Python-based web app.



## Conclusion

The analysis and modeling phases provided clear insights into sales trends and enabled accurate forecasting. The chosen model (ARIMA 0,1,1) offers a balance of simplicity and accuracy, and the dashboard will further enhance interpretability and business impact.