

# Internship Report: Q2 2023 Sales Forecast

## Introduction

The following report describes my work while doing a data analysis internship, wherein I created a predictive model for forecasting overall sales in Q2 2023. The task was to aid the sales department with data-backed insights to assist in planning and strategy.

## Background

The sales data for Q1-Q4 2022 of the company was furnished to develop a machine learning model that can make future sales forecasts. The intention was to predict total sales in Q2 2023 with features like `order_date` and `qty_ordered`.

## Learning Objectives

- Use time series forecasting or regression methods with past data.
- Compare the performance of machine learning models based on error metrics (e.g., MAE, RMSE).
- Develop actionable insights and present findings succinctly.

## Activities and Tasks

- Data Preprocessing: Cleaned and converted order data. Transformed `order_date` into datetime format and grouped monthly/quarterly sales.
- Feature Engineering: Created a time series of total sales through `order_date` and `qty_ordered`.
- Model Selection and Training: Trained Linear Regression and ARIMA models to predict Q2 2023 sales.
- Evaluation: Validated model accuracy using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
- Visualization: Produced line plots and bar charts to determine trends and present forecasted values.

## Skills and Competencies

- Technical: Python (Pandas, Scikit-learn, Statsmodels), Time Series Analysis, Regression Modeling, Data Visualization (Matplotlib, Seaborn).
- Analytical: Trend analysis, performance assessment, and business interpretation of sales forecasts.

# Internship Report: Q2 2023 Sales Forecast

## Feedback and Evidence

The model made a decent prediction accuracy:

MAE: e.g., 1450 units

RMSE: e.g., 1720 units

Forecasted total sales for Q2 2023: ~25,800 units (from model output).

Identified upward trend in sales from Q1 to Q2 throughout 2022, indicating seasonal growth.

## Challenges and Solutions

- Challenge: Erratic spikes in sales data caused by promotions and discounts.

Solution: Utilized a rolling average and outlier smoothing to steady the time series.

- Challenge: Fewer features (just order\_date and qty\_ordered) could compromise accuracy.

Solution: Focused on grouping data to quarterly level and using simpler, robust models.

## Outcomes and Impact

- Built a functioning sales prediction pipeline with historical data.

- Allowed the sales team to forecast Q2 2023 sales volume for marketing and inventory planning.

- Improved my knowledge of data modeling and forecasting methods in actual applications.

## Conclusion

The project helped me to apply statistical and machine learning methods to a real-world business problem. I was able to get practical experience in forecasting, model performance evaluation, and presenting insights that can inform strategic choices.