**Objective**

The objective of this project was to build a robust sales forecasting model for a retail business. The company was looking to improve its ability to predict future sales, enabling better decision-making related to inventory management

The goal was to reduce the costs associated with overstocking and stockouts by predicting future sales more accurately, enabling efficient decision-making around inventory replenishment.

**Problem Statement**

The goal was to develop a predictive model that could forecast sales at product level, help the business optimize its inventory management and enhance profitability.

1. Adapt to external shocks like the COVID-19 pandemic.
2. Optimize inventory levels across Product/Style.

**Steps Involved:**

**1. Data Collection & Exploration:**

* + 1. Historical Sales Data: Detailed daily sales information for products across stores.
    2. Markdown Data: Records of discounts offered on various products during different sales events.
    3. Event Data: Information on major retail events like End of Season Sales (EOSS) and festivals.
    4. External Factors: Data on external events, particularly the COVID-19 pandemic, which had a direct impact on foot traffic and sales.

We performed Exploratory Data Analysis (EDA) to understand sales patterns, identify seasonal trends, assess the effect of markdowns, and find correlations between different features and sales.

**2. Feature Engineering:**

Once the data was clean, we moved on to feature engineering, which involved creating new, more meaningful variables to feed into the forecasting model. This included:

* + 1. Lag Features: Previous weeks’ and months' sales to capture trends.
    2. Rolling Averages: Average sales over moving windows (e.g., 7-day, 30-day) to smooth fluctuations. – if we don’t have data for particular months
    3. Categorical Encoding: Encoding store types, locations, and product categories to make them machine-readable.
    4. Event Flags: Markdown was used as it is, Binary variables indicating whether a promotional event (e.g., EOSS) was active.
    5. COVID-19 Impact Variables: Flags and indices to account for the pandemic's impact on sales.

**3. Model Selection:**

Given the presence of seasonality and external factors, I chose SARIMAX (Seasonal AutoRegressive Integrated Moving Average with exogenous regressors) as the primary model. SARIMAX was well-suited for the problem due to its ability to capture both seasonal and trend components while allowing the incorporation of external factors (like promotions) that influence sales.

Rationale for SARIMAX:

SARIMAX can handle seasonal effects in the data, which was critical given that the sales patterns were influenced by holidays and promotions.

The model also allowed me to add exogenous variables, like marketing campaigns and external economic conditions, which improved the accuracy of the forecasts.

Compared to machine learning models like XGBoost, SARIMAX’s interpretability was crucial for understanding which factors influenced sales the most.

**Execution Flow:** order of execution when integrating Optuna with SARIMAX is as follows:

1. Define Optuna Study:
   * Define the search space for SARIMAX hyperparameters.
   * Write the objective function, where SARIMAX is created, trained, and evaluated for each trial.
2. Run Optuna Optimization:
   * Optuna iterates through the trials, each time:
     + Suggesting a new set of hyperparameters.
     + Creating a SARIMAX model using those parameters.
     + Fitting the model on the training data and evaluating its performance (e.g., using MAPE).
3. Extract Best Hyperparameters:
   * Optuna provides the best parameters after the trials are complete.
4. Build Final SARIMAX Model:
   * Use the best hyperparameters to build and train the final SARIMAX model on your entire dataset.
   * Evaluate its performance and deploy for production use.

**Key Benefits of Using Optuna in SARIMAX:**

* Automated Hyperparameter Tuning: Optuna intelligently searches the hyperparameter space and avoids the exhaustive grid search approach.
* Optimization of Forecast Accuracy: By minimizing MAPE through Optuna, you can ensure that the SARIMAX model's accuracy is maximized.
* Efficient Search: Optuna uses Bayesian optimization and other techniques to speed up the search process and converge on the best parameters efficiently.

**4. Model Evaluation and Validation:**

Using the optimized SARIMAX model, we trained on the full training dataset and validated it on a test set. Key steps included:

* **MAPE Calculation:** The final model’s predictions were compared against the actual sales, with MAPE as the evaluation metric. The resulting MAPE of ~7% indicated that the model was effective in predicting sales with minimal error.
* **Residual Analysis:** The residuals of the model were analyzed to ensure no significant autocorrelation or pattern, confirming that the model was capturing the essential trends and seasonality.

**5. Forecasting and Business Impact:**

The optimized SARIMAX model was deployed to forecast future sales. The forecasts were provided monthly for next 6 months, enabling the business to:

* **Inventory Optimization:** Better align inventory with predicted demand, reducing overstock and stockout issues.
* **Resource Allocation:** Allocate resources (e.g., manpower, storage space) based on expected sales volumes.
* **Breakdown of Sales Qty by Sales Mix on Style/ Colour level.**

**Project Impact on Business:**

The SARIMAX model, optimized using Optuna, had a profound impact on the retail business:

1. **Reduced Inventory Costs:** The accurate forecasts allowed the company to reduce overstock by 15%, leading to substantial cost savings.
2. **Improved Sales Campaigns:** The ability to forecast sales during promotions enabled better planning, resulting in a 10% increase in campaign efficiency/Revenue.
3. **Enhanced Resource Planning:** Stores could better prepare for high-demand periods, improving customer satisfaction and reducing lost sales.
4. **Informed Decision Making:** The business gained a deeper understanding of how external factors (e.g., economic conditions, holidays) influence sales, allowing for more data-driven decisions.