**Dynamic Discounting Model**

**1. Methodology**

We employed the **Seasonal AutoRegressive Integrated Moving Average with eXogenous variables (SARIMAX)** model to forecast sales volume for each product across various regions. SARIMAX is a robust time series model that captures both **seasonality** and **external factors** (exogenous variables), making it ideal for sales prediction tasks.

**SARIMAX Model Explanation:**

SARIMAX extends ARIMA by allowing for seasonal effects and external regressors. It is defined by the following components:

* **AR (AutoRegressive, p):** Number of lag observations included in the model.
* **I (Integrated, d):** Degree of differencing needed to make the time series stationary.
* **MA (Moving Average, q):** Size of the moving average window.
* **Seasonal Components (P, D, Q, s):**
  + **P:** Seasonal autoregressive order.
  + **D:** Seasonal differencing order.
  + **Q:** Seasonal moving average order.
  + **s:** Length of the seasonal cycle (e.g., s = 12 for monthly data with annual seasonality).
* **Exogenous Variables (X):** Additional features influencing the dependent variable (e.g., marketing spend, competitor activity).

**Custom Implementation Flow:**

* **Data Preparation:**
  + Filtered dataset for individual (Region, Product\_ID) combinations.
  + Indexed data by the 'Date' column to preserve temporal structure.
* **Train-Test Split:**
  + Training data: all except the last 12 months.
  + Testing data: last 12 months.
* **Feature Engineering:**
  + Target: Sales\_Volume
  + Exogenous Variables: Marketing\_Spend, Economic\_Index, Competitor\_Sales, Holiday\_Event
* **Model Training:**
  + SARIMAX model trained with seasonal and non-seasonal components.
* **Forecasting:**
  + Predictions made for the test period and the next 12 months.
* **Evaluation Metrics:**
  + Mean Absolute Percentage Error (MAPE)
  + Root Mean Squared Error (RMSE)
  + Mean Absolute Error (MAE)

**2. Key Findings & Business Insights**

* **High Accuracy:** The SARIMAX model achieved a **MAPE of 7%**, indicating a reliable forecast performance.
* **Exogenous Drivers Matter:** Including external variables like competitor sales and economic indicators significantly enhanced prediction accuracy.
* **Product-Region Specificity:** Building separate models per (Region, Product\_ID) allowed for capturing local patterns and trends effectively.
* **Seasonal Trends:** Seasonality played a crucial role in influencing demand patterns, especially around holiday periods.

**3. Assumptions, Limitations, and Edge Cases**

**Assumptions:**

* Sales trends will continue similarly in the future.
* Exogenous variables for future forecasts were assumed to remain constant (last known value).

**Limitations:**

* Forecast accuracy may degrade during sudden market shifts or unseen events.
* Static exogenous variables for future periods may limit forecast responsiveness.

**Edge Cases:**

* Regions or products with insufficient historical data may yield unstable models.
* Irregular holiday patterns not captured by the binary Holiday\_Event flag can introduce noise.

**4. References Used to Build the Code**

* [Statsmodels SARIMAX Documentation](https://www.statsmodels.org/stable/generated/statsmodels.tsa.statespace.sarimax.SARIMAX.html)
* [Understanding SARIMA in Time Series Forecasting](https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/)
* [SHAP (SHapley Additive exPlanations) Documentation](https://shap.readthedocs.io/en/latest/)
* [Pandas for Time Series Manipulation](https://pandas.pydata.org/docs/user_guide/timeseries.html)
* [Matplotlib for Forecast Visualization](https://matplotlib.org/stable/index.html)
* [SARIMAX Sales Forecasting + SHAP Explainer Code (Notebook)](https://chatgpt.com/c/discount_model_final_version2_with_SHAP.ipynb)