**PHASE-1**

**PROBLEM DEFINITION:**

The problem is to develop a predictive model that uses historical sales data to forecast future sales for a retail company. The objective is to create a tool that enables the company to optimize inventory management and make informed business decisions based on datadriven sales prediction.

**Step 1: Data Collection**

-From the given dataset https://www.kaggle.com/datasets/chakradharmattal/future-sales-prediction

- Ensure that your dataset is clean, handle missing values, outliers, and format it properly for analysis.

**Step 2: Exploratory Data Analysis (EDA) or preprocessing**

- Conducting EDA to understand the data better. Visualizing the data, computing summary statistics, and analyzing correlations.

- Identifying trends, seasonality, and patterns in the historical sales data.

**Step 3: Feature Engineering**

- Create new features or transform existing ones that can help capture important information for sales prediction. Consider the following features:

- Lag features: Past sales data for the same or related products.

- Time-related features: Day of the week, month, year, and holidays.

- Product-related features: Pricing, product category, promotions, and inventory levels.

**Step 4: Data Splitting**

- Split the dataset into training and validation sets. Typically, we can reserve the most recent data for validation and testing.

- Data should be time-ordered, so that we can train on earlier data and validate on more recent data.

**Step 5: Model Selection**

- Choose an appropriate predictive modeling technique. In this case, time series forecasting models like SARIMA (Seasonal AutoRegressive Integrated Moving Average) or Prophet are often suitable choices.

**Step 6: Model Training**

- Train the selected model using the training dataset. Pay attention to the hyperparameters of the chosen model.

- If using a time series model, consider tuning hyperparameters like seasonality, trend order, and differencing.

**Step 7: Model Evaluation**

- Assess the model's performance on the validation dataset using relevant evaluation metrics for sales forecasting, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or Mean Absolute Percentage Error (MAPE).

- Compare the model's performance to a baseline model for context.