**FUTURE SALES PREDICTION**

PROBLEM DEFINITION AND DESIGN THINKING

PROBLEM STATEMENT:

The problem at hand is to develop a predictive model that leverages historical sales data to forecast future sales for a retail company. The primary objective of this project is to create a tool that empowers the company to optimize inventory management and make data-driven decisions. To achieve this, we need to perform data preprocessing, feature engineering, model selection, training, and evaluation.

DESIGN THINKING:

DATA SOURCE:

The first step is to utilize a dataset containing historical sales data. This dataset should include essential features such as date, product ID, store ID, and sales quantity. This data will serve as the foundation for our predictive model.

DATA PREPROCESSING:

Objective: Clean and preprocess the data to make it suitable for modelling.

STEPS:

1. Handle Missing Values: Identify and handle any missing or null values in the dataset. Depending on the extent of missing data, we may choose to impute values or drop rows/columns.
2. Convert Categorical Features: Convert categorical features like product ID and store ID into numerical representations using techniques such as one-hot encoding or label encoding.
3. Date Parsing: Extract relevant information from the date column, such as day of the week, month, or year. These time-based features can be valuable for forecasting.

FEATURE ENGINEERING:

Objective: Create additional features that can enhance the predictive power of the model.

STEPS:

1. Time-Based Features: Generate features such as day of the week, month, and year from the date column. These features can capture seasonality and trends.
2. Lag Features: Create lag features, where the target variable (sales quantity) from previous time steps is included as a feature. This can help capture autocorrelation in the data.

MODEL SELECTION:

Objective: Choose suitable time series forecasting algorithms for predicting future sales.

STEPS:

1. Initial Model Selection: Start with basic time series models like Moving Averages to establish a baseline. Then, explore more advanced models like ARIMA (Auto Regressive Integrated Moving Average), Exponential Smoothing, or Prophet.
2. Hyperparameter Tuning: Fine-tune the hyperparameters of selected models for optimal performance. This may involve grid search or Bayesian optimization.

MODEL TRAINING:

Objective: Train the selected model using the pre-processed data.

STEPS:

1. Split the dataset into training and validation sets to assess model performance during training.
2. Train the model on the training data, taking into account the time-based structure of the data.

EVALUATION:

Objective: Evaluate the model's performance using appropriate time series forecasting metrics.

STEPS:

1. Calculate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to assess the model's accuracy.
2. Visualize the model's predictions against the actual sales data to gain insights into its performance.

This document outlines the plan for developing a predictive model to forecast future sales for a retail company. The process involves data preprocessing, feature engineering, model selection, training, and evaluation. By following this structured approach, we aim to provide the company with a powerful tool for optimizing inventory management and making informed business decisions based on data-driven sales predictions.