

Data Cleaning & Preprocessing Tasks

1 Handling Missing Values

Identify missing values in Sales, Store, and Item.

Fill missing values using:

Median for Sales (to avoid outliers affecting the mean).

Mode for Store and Item (most common value).

Drop rows with too many missing values.

2 Handling Date & Time Data

Convert Date column to datetime format.

Extract new features from Date:

Year, Month, Day, Weekday, Quarter.

Create a Weekend column (1 if Saturday/Sunday, else 0).

3 Removing Duplicates

Check for duplicate rows and remove them.

4 Outlier Detection & Handling

Check for outliers in Sales using:

Boxplots (to detect extreme values).

Z-score or IQR method to remove/fix outliers.

Cap sales at a reasonable max limit.

5 Encoding Categorical Variables

Convert Store and Item into **one-hot encoding** or **label encoding**.

6 Feature Engineering

Create New Features:

Sales per Item = Sales / Item

Sales Growth Rate = (Sales_today - Sales_yesterday) /
Sales_yesterday

Create Lag Features for forecasting (e.g., sales of previous day/week).

Rolling averages for smoothing sales trends.