Model Development Phase

Date	4 June 2024
Team ID	SWTID1719938571
Project Title	Walmart Sales Analysis for Retail Industry with Machine Learning
Maximum Marks	5 Marks

Feature Selection Report

In the forthcoming update, each feature will be accompanied by a brief description. Users will indicate whether it's selected or not, providing reasoning for their decision. This process will streamline decision-making and enhance transparency in feature selection.

Feature	Description	Selected (Yes/No)	Reasoning
Store	Each stores contain number of department-ide sales for each store	Yes	This feature identifies the store number, allowing us to segment data based on different store locations
Department	The department numbers	Yes	The department number specifies which department within the store the data corresponds to. Different departments within a store might have different sales patterns, product categories, and customer preferences.

Isholiday	Whether holiday is there is not	Yes	This binary feature indicates whether the week contains a special holiday. Holidays often lead to increased consumer spending, so this feature helps capture the impact of holidays on sales.
Temperature	Average temperature in the region	Yes	Average temperature in the region where the store is located. Weather conditions can influence consumer behavior and affect sales of certain products (e.g., clothing, seasonal items).
Fuel Price	Cost of fuel in the region	No	While fuel price can impact consumer purchasing power and transportation costs, its direct effect on sales may be less pronounced or variable across different regions or stores.
CPI	The consumer price index	Yes	CPI measures changes in the price level of a market basket of consumer goods and services purchased by households. It reflects inflationary pressures and can affect consumer spending behavior.
Unemployment	The unemployment rate	No	Although unemployment rates can influence consumer spending and economic conditions, the relationship may not be direct or consistent enough across different regions to impact sales significantly.
Year	Features denotes calendar year in which sales is recorded	Yes	Adding the year can help capture any long-term trends or seasonality that might not be captured by the Date feature alone
Size	The size feature refers to the physical dimensions or	Yes	This feature likely refers to the size of the store. Larger stores may have more inventory, a wider selection of products, and attract more customers, leading to

	square footage of the Walmart store		potentially higher sales. Conversely, smaller stores might have a more focused product offering and cater to a specific demographic.
Dayofweek	Indicates the specific day of the week on which the sales data is recorded	Yes	Incorporating the day of the week can help capture any weekly patterns or trends in sales. For instance, weekends might see higher foot traffic and sales compared to weekdays due to more people being off work and available for shopping.
Month	Represents the calendar month in which the sales data is recorded	Yes	Similar to the day of the week, including the month allows us to capture seasonal patterns in sales.
Date	Represents the calendar date in which the sales data is recorded	Yes	This feature indicates the week for which the data is recorded. Seasonality, holidays, and other time-related factors can significantly influence sales patterns, making this feature crucial for forecasting.
Weekly_sales	Sales for the given department in the given store	Yes	This is the target variable we aim to predict in the train.csv file. This is the feature that we will be using to train the model