##### EDA & Modeling, Sales Forecasting.

The goal of this project is to analyze sales data, uncover patterns, and forecast future sales. The insights gained here will be used to build a predictive model for sales, with 90% of accuracy score using r2 score metrics.

##### ****Dataset Overview****

1. **Source:** souce of datafrom kaggle competition dataset  
   **Number of rows:**2.5 million+
2. **Number of columns:** 29 columns

##### ****Features in the Dataset****

* id: Unique identifier for each entry.
* store\_nbr: Store number indicating the branch.
* family: Product category.
* onpromotion: Indicates whether the product is on promotion.
* year: Year of sale.
* month: Month of sale.
* day: Day of sale.
* + Same in other files

##### ****Steps Conducted in EDA****

**Data Cleaning**

* + Checked for missing values and handled them using appropriate strategies (e.g., mean/mode imputation or dropping rows).
  + Converted categorical variables (family) into numeric representations using label encoding.

**Data Types**

* + Verified data types of all columns to ensure compatibility with the model.
  + Corrected any mismatched data types (e.g., converting onpromotion to numeric).

**Univariate Analysis**

* + Analyzed the distribution of numeric columns using histograms and boxplots.
  + Plotted frequency counts of categorical variables like family using bar charts.

**Bivariate Analysis**

* + Explored relationships between sales (unit\_sales) and other features using scatterplots, correlation heatmaps, and pair plots.

**Seasonality and Trend**

* + Examined sales trends over time using line charts to identify seasonality and growth patterns.

**Outlier Detection**

* + Detected and analyzed outliers in unit\_sales using boxplots.

**Promotions Impact**

* + Compared sales for products on promotion versus not on promotion using grouped bar charts.

##### ****Key Findings****

1. **Sales Trends**
   * Sales exhibit clear seasonality, with peaks during specific months (e.g., holidays).
2. **Product Categories**
   * Certain family categories, such as 'ELECTRONICS' and 'GROCERIES', contribute significantly to sales.
3. **Promotions**
   * Products on promotion tend to have higher sales compared to non-promoted products.
4. **Store Performance**
   * Certain stores consistently outperform others, possibly due to location or customer demographics.

##### ****Conclusion****

The EDA provides insights into key drivers of sales and seasonal trends. These findings inform the preprocessing and feature engineering steps for the forecasting model.