**EDA Summary**

The EDA project on the Big Mart Sales dataset aimed to gain insights into the factors influencing sales and prepare the data for predictive modelling. The dataset, available on AnalyticsVidhya(<https://datahack.analyticsvidhya.com/contest/practice-problem-big-mart-sales-iii/#About>).

It contains information about various products in Big Mart stores, including attributes like item weight, visibility, type, and outlet details.

1. **Univariate Analysis:**
   * Descriptive statistics were computed for numerical variables such as item weight, visibility, and item outlet sales.
   * Distributions of categorical variables like item type and outlet type were analyzed.
   * Frequency distributions were used to understand the central tendency and variability of the data.
2. **Bivariate Analysis:**
   * Correlation matrices were generated to identify relationships between numerical variables.
   * Scatter plots and Box plot analysis were employed to explore correlations between item visibility, weight, and sales.
3. **Null Value Treatment:**
   * Missing values were identified and handled appropriately. Imputation strategies, such as mean or median replacement, were applied to maintain data integrity.
4. **Outlier Treatment:**
   * Outliers in numerical features were detected using statistical method interquartile range (IQR).
   * Outliers were either removed or transformed based on the nature of the data, ensuring they didn't unduly influence the analysis.
5. **Scaling the Data:**
   * Numerical variables were scaled to a consistent range to avoid biases in machine learning models. Standardization technique was applied.
6. **Encoding of Categorical Variables:**
   * Categorical variables were encoded to numerical format, enabling their use in machine learning algorithms. Techniques like one-hot encoding or label encoding were employed based on the nature of the categorical data.

In summary, the EDA process provided valuable insights into the distribution and relationships within the Big Mart Sales dataset. Addressing missing values, treating outliers, and scaling the data were crucial steps in preparing the dataset for subsequent modeling. The findings from the EDA will guide further feature engineering and model development to predict sales more accurately.