


## Bigmart Sales Prediction ML Project

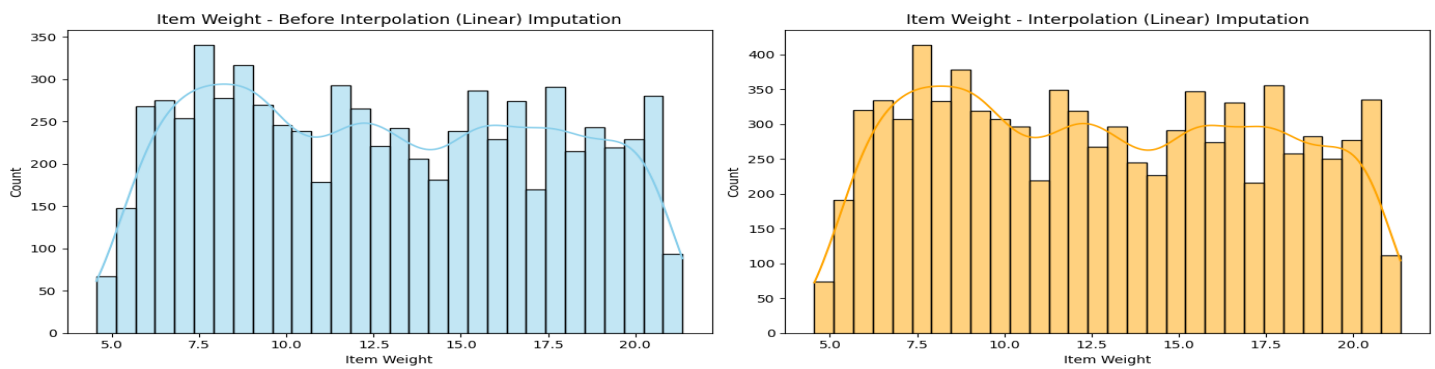
- **Objective:**  
Built a machine learning model to predict Big Mart sales using structured data, enabling actionable insights and improved decision-making.
- **Data Preparation:**  
Cleaned the training dataset by handled missing values by imputations method like interpolation & mean values, applied feature scaling with Standard scaler for standardization of features, encoded categorical variables, and Engineered relevant features to enhance model accuracy.
- **Exploratory Data Analysis (EDA):**  
Performed EDA to explore data distributions, detected outliers, and identify key relationships. Applied log transformation to reduce skewness in sales data. Wherever needed used appropriate chart to understand the data better for consumption.
- **Modeling Strategy:**  
Evaluated multiple regression models including Linear, Lasso, Ridge, ElasticNet, Random Forest, and XGBoost. Used train-test split and selected models based on RMSE and  $R^2$  performance metrics.
- **Hyperparameter Tuning:**  
Optimized XGBoost parameters using GridSearchCV, leading to improved test performance.
- **Results:**  
Final XGBoost model achieved an  $R^2$  of 61%, with RMSE of 1052 on training data and 1148.86 on test data.
- **Conclusion:**  
Successfully developed a robust ML model. Further performance gains could be achieved with additional data on sales skew, outlet-specific pricing anomalies, promotional/holiday effects, and detailed product categorization, Outlet Size based on Revenue or Space data etc.

Here is Screen Shot of Leaderboard

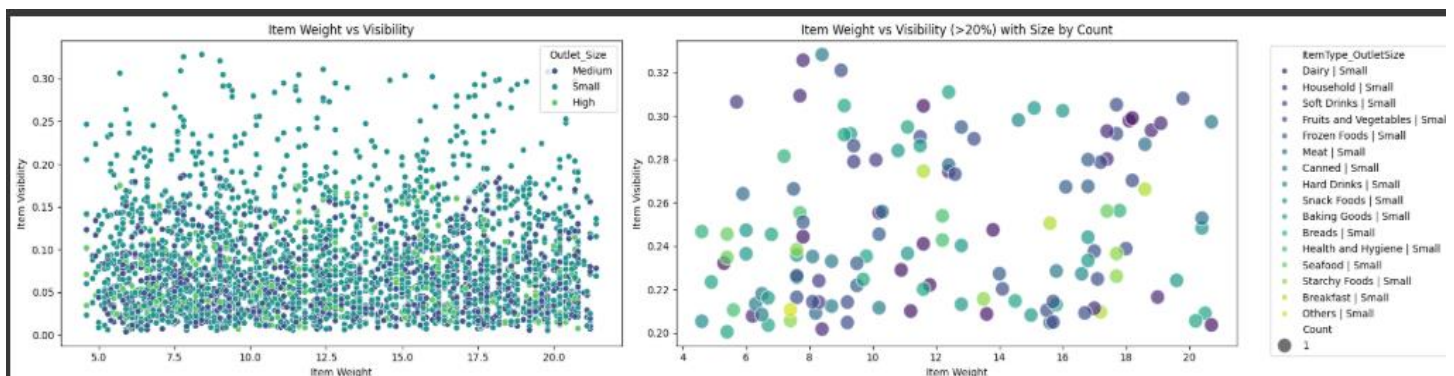
Rank	Username	Score
#585	 You	1148.8620277145

## Graphical Summary on Imputation & EDA

1. Item Weight Imputation based on Interpolation Method used linear. Also taken into consideration Max weight of similar product & item type . Variance approximately similar



2. The spread of Item Visibility across Outlet Sizes indicates that smaller outlets exhibit a significant number of outliers beyond the 20th percentile density, regardless of item weight.



3. Log transformation of Item Visibility was applied to normalize its skewed distribution.

