1. Problem Definition & Approach

Objective: Predict item sales across different outlets
Type: Regression (continuous target)
Metrics: RMSE and R^2 score
Challenge: Missing values and categorical variables

2. Data Preprocessing

Data Quality Issues	Data Impacted (%)	Treatment Method
Missing Item Weight	17.00	Linear interpolation
Zero Item Visibility	6.17	Interpolated values
Missing Outlet Size	28.00	Based on Outlet Type patterns

3. Feature Engineering

- **Created Features:** Price/Weight ratio, Store Age, Store Age × Size Interaction, Visibility × MRP Interaction
- **Categorical Handling:** Ordinal Encoding, Fat Content Standardization, Item Type Extraction

4. Feature Selection & Model

- Selected top 6 features via XGBoost importance scores
- Used XGBRFRegressor with 5-fold cross-validation

Rank	Features	Importance (%)
1	Outlet Type	36.7
2	Store Age Size	34.9
3	Item MRP	8.85
4	Outlet Establishment Year	8.24
5	Outlet Identifier	6.40
6	Outlet Size	2.24

5. Implementation & Code Structure

```
project/
   ├── config/           # Configuration files
   ├── data/
   │    ├── raw/         # Original data
   │    └── processed/   # Cleaned datasets
   ├── docs/             # Project documentation
   ├── notebooks/        # Exploratory analysis
   └── src/
        ├── data/        # Data processing
        ├── models/      # Model training
        └── utils/       # Helper functions
```