**Sales Data Analysis and Cleaning**

**Overview**

This project focuses on cleaning, merging, and analyzing sales data from multiple sources to uncover trends such as monthly sales performance. The processed data is used to generate a cleaned dataset and visualizations for insights.

**Datasets**

* **sales.csv**: Transaction records including date, time, product code, supermarket number, sales amount, and units sold.
* **promotion.csv**: Promotion details by product code, supermarket, and week.
* **item.csv**: Product information (code, description, category).
* **supermarkets.csv**: Supermarket metadata (name, location, size).

**Dependencies**

* Python 3.x
* Libraries: pandas, matplotlib, jupyter

**Installation**

pip install pandas matplotlib jupyter

**Usage**

1. Clone the repository.
2. Place datasets in Data Sciene Internship Assignment Datasets/.
3. Run the Jupyter notebook:

jupyter notebook

**Data Cleaning Steps**

1. **Handling Missing Values**:
   * Dropped rows with missing critical fields (e.g., amount, units).
   * Removed rows with nulls in items\_df and promotion\_df.
2. **Column Standardization**:
   * Renamed columns for consistency (e.g., supermarketNo → supermarket).
   * Stripped whitespace and converted column names to lowercase.
3. **Data Type Conversion**:
   * Formatted time in sales\_df to HH:MM .
   * Converted amount and units to numeric types.
4. **Merging Data**:
   * Combined sales\_df with items\_df on code.
   * Joined the result with promotion\_df on code and supermarket.
   * Merged with supermarkets\_df to include location/size details.
5. **Feature Engineering**:
   * Created revenue column (amount \* units).

**Output**

* **cleaned\_sales\_data.csv**: Final merged and cleaned dataset ready for analysis.