

```
# Import necessary libraries
import pandas as pd
import numpy as np

# Load the dataset
df = pd.read_csv('Python Project Data -
Supermarket Sales.csv')

# Display the first few rows of the dataset
df.head()

# Check for missing values
df.isnull().sum()

# Handle missing values (e.g., fill with mean,
median, or drop)
df.fillna(df.mean(), inplace=True)

# Check for duplicate rows
df.duplicated().sum()

# Remove duplicate rows
df.drop_duplicates(inplace=True)

# Data type conversion (if necessary)
df['Date'] = pd.to_datetime(df['Date'])

# Feature engineering (e.g., extracting year,
month, day from date)
df['Year'] = df['Date'].dt.year
df['Month'] = df['Date'].dt.month
df['Day'] = df['Date'].dt.day
# Save the cleaned dataset
df.to_csv('cleaned_supermarket_sales.csv',
index=False)
```