Task 6 – Data Wrangling & Preprocessing

Topics Covered

1. Handling Missing Data
   - Drop or fill missing values.
2. Outlier Detection
   - Using IQR method or visualization.
3. Feature Scaling
   - Normalization, Standardization.
4. Encoding Categorical Data
   - One-hot encoding, label encoding.

Program 1

```
import pandas as pd

data = {'Name': ['A', 'B', 'C', 'D'],
        'Age': [25, None, 30, 28]}
df = pd.DataFrame(data)

# Fill missing values with mean
df['Age'].fillna(df['Age'].mean(), inplace=True)
print(df)
```

Program 2

```
from sklearn.preprocessing import StandardScaler, LabelEncoder
import pandas as pd

data = {'Name': ['A', 'B', 'C', 'D'],
        'Salary': [25000, 40000, 35000, 30000]}
df = pd.DataFrame(data)

# Feature Scaling
scaler = StandardScaler()
df['Salary_scaled'] = scaler.fit_transform(df[['Salary']])

# Label Encoding
encoder = LabelEncoder()
df['Name_encoded'] = encoder.fit_transform(df['Name'])

print(df)
```

🌟 Key Takeaways

- Learned techniques for handling missing data.
- Practiced outlier detection methods.
- Understood feature scaling for better ML model performance.
- Applied encoding techniques for categorical variables.
- Wrangling ensures datasets are clean and ready for ML pipelines.