```
df.dropna(inplace=True)
```

[22]:
```
value_sales = df['TV'].value_counts()
print("\nsales:")
print(value_sales)
```

```
sales:
199.8    2
109.8    2
17.2     2
177.0    2
```



```
222.4    2
         ..
139.3    1
216.8    1
199.1    1
26.8     1
232.1    1
Name: TV, Length: 190, dtype: int64
```

[25]:
```
label_encoder = LabelEncoder()
df['Sales'] = label_encoder.fit_transform(df['Sales'])
df
```

[25]:

| | TV | Radio | Newspaper | Sales |
|---|---|---|---|---|
| 0 | 230.1 | 37.8 | 69.2 | 106 |
| 1 | 44.5 | 39.3 | 45.1 | 28 |
| 2 | 17.2 | 45.9 | 69.3 | 40 |
| 3 | 151.5 | 41.3 | 58.5 | 66 |
| 4 | 180.8 | 10.8 | 58.4 | 80 |
| .. | ... | ... | ... | ... |
| 195 | 38.2 | 3.7 | 13.8 | 14 |
| 196 | 94.2 | 4.9 | 8.1 | 52 |
| 197 | 177.0 | 9.3 | 6.4 | 56 |
| 198 | 283.6 | 42.0 | 66.2 | 118 |
| 199 | 232.1 | 8.6 | 8.7 | 84 |

[200 rows x 4 columns]

[27]:
```
from sklearn.linear_model import LinearRegression
model=LinearRegression()
```

[32]:
```
X = df.drop('Newspaper', axis=1)
y = df['Newspaper']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)
```

[33]:
```
print("\n X_test info")
print(X_test.info())
```

```
 X_test info
<class 'pandas.core.frame.DataFrame'>
Int64Index: 40 entries, 95 to 76
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   TV      40 non-null     float64
```

# 1 FUTURE SALES PREDICTION - ADS_PHASE 3

CHARAN SAI REDDY (Team member)

## 3.1 Problem Statement : Loading and Preprocessing

In this part you will begin building your project by loading and preprocessing the dataset.

Begin building the future sales prediction by loading and preprocessing the dataset.

[15]:
```
#importing necessary libraries
import pandas as pd
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.impute import SimpleImputer
from sklearn.model_selection import train_test_split
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

# Import the Netflix dataset
file_path = r"C:\Users\gayat\OneDrive\Desktop\sales_data.csv" # Remove the 'r'␣
↪from the parentheses
encoding = "ISO-8859-1"
df = pd.read_csv(file_path, encoding=encoding)
df.head()
```

[15]:

| | TV | Radio | Newspaper | Sales |
|---|---|---|---|---|
| 0 | 230.1 | 37.8 | 69.2 | 22.1 |
| 1 | 44.5 | 39.3 | 45.1 | 10.4 |
| 2 | 17.2 | 45.9 | 69.3 | 12.0 |
| 3 | 151.5 | 41.3 | 58.5 | 16.5 |
| 4 | 180.8 | 10.8 | 58.4 | 17.9 |

[16]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   TV         200 non-null    float64
 1   Radio      200 non-null    float64
 2   Newspaper  200 non-null    float64
 3   Sales      200 non-null    float64
dtypes: float64(4)
memory usage: 6.4 KB
```



[17]:
```
df.head()
```

[17]:

| | TV | Radio | Newspaper | Sales |
|---|---|---|---|---|
| 0 | 230.1 | 37.8 | 69.2 | 22.1 |
| 1 | 44.5 | 39.3 | 45.1 | 10.4 |
| 2 | 17.2 | 45.9 | 69.3 | 12.0 |
| 3 | 151.5 | 41.3 | 58.5 | 16.5 |
| 4 | 180.8 | 10.8 | 58.4 | 17.9 |

[18]:
```
df.isnull()
```

[18]:

| | TV | Radio | Newspaper | Sales |
|---|---|---|---|---|
| 0 | False | False | False | False |
| 1 | False | False | False | False |
| 2 | False | False | False | False |
| 3 | False | False | False | False |
| 4 | False | False | False | False |