CODE

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Load the dataset
data= pd.read_csv('sales_pred.csv', encoding='ISO-8859-1')  # Load the dataset
print(data)

# Data Preprocessing
print(data.head())
print(data.isnull().sum())
print(data.describe())

# Check correlation
sns.heatmap(data.corr(), annot=True, cmap='coolwarm')
plt.show()

# Split into features and target variable
X = data[['TV', 'Radio', 'Newspaper']]  # Features
y = data['Sales']  # Target variable

# Split into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Initialize and train the model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Compare predicted vs actual
comparison_df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
print(comparison_df.head())

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f'Mean Squared Error: {mse}')
print(f'R-squared: {r2}')

# Plot actual vs predicted sales
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred)
plt.plot([y.min(), y.max()], [y.min(), y.max()], color='red', linestyle='--')  # Line of perfect prediction
plt.xlabel('Actual Sales')
plt.ylabel('Predicted Sales')
plt.title('Actual vs Predicted Sales')
plt.show()

# Print model coefficients
print('Coefficients:', model.coef_)
```

```
print('Intercept:', model.intercept_)
```

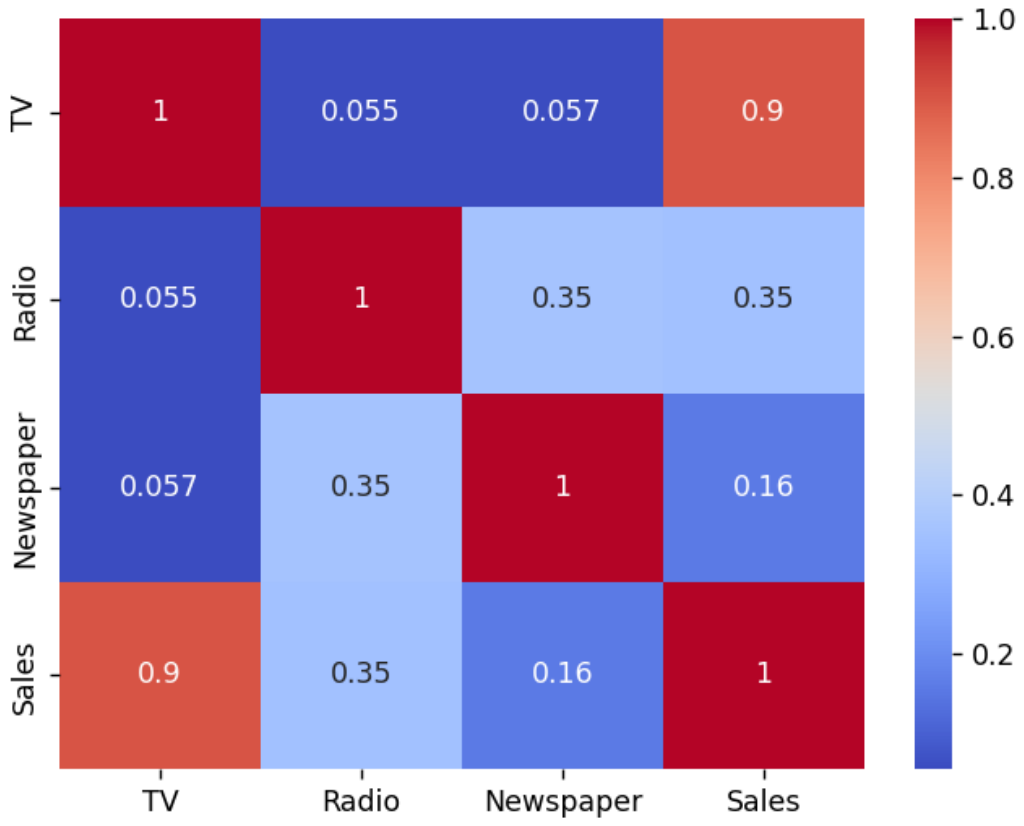
OUTPUT

```
  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
..     ...    ...       ...    ...
195   38.2    3.7      13.8    7.6
196   94.2    4.9       8.1   14.0
197  177.0    9.3       6.4   14.8
198  283.6   42.0      66.2   25.5
199  232.1    8.6       8.7   18.4

[200 rows x 4 columns]
     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
TV          0
Radio       0
Newspaper   0
Sales       0
```

dtype: int64

	TV	Radio	Newspaper	Sales
count	200.000000	200.000000	200.000000	200.000000
mean	147.042500	23.264000	30.554000	15.130500
std	85.854236	14.846809	21.778621	5.283892
min	0.700000	0.000000	0.300000	1.600000
25%	74.375000	9.975000	12.750000	11.000000
50%	149.750000	22.900000	25.750000	16.000000
75%	218.825000	36.525000	45.100000	19.050000
max	296.400000	49.600000	114.000000	27.000000



	Actual	Predicted
95	16.9	17.034772
15	22.4	20.409740

30 21.4 23.723989

158 7.3 9.272785

128 24.7 21.682719

Mean Squared Error: 2.9077569102710896

R-squared: 0.9059011844150826

Coefficients: [0.05450927 0.10094536 0.00433665]

Intercept: 4.714126402214127