In [32]:
```
os.chdir(r'C:\Users\admin\Desktop\2213557-MA 336')
```

In [33]:
```
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
```

In [34]:
```
dataset = pd.read_csv('advertising.csv')
```

Exploratory Data Analysis (EDA)

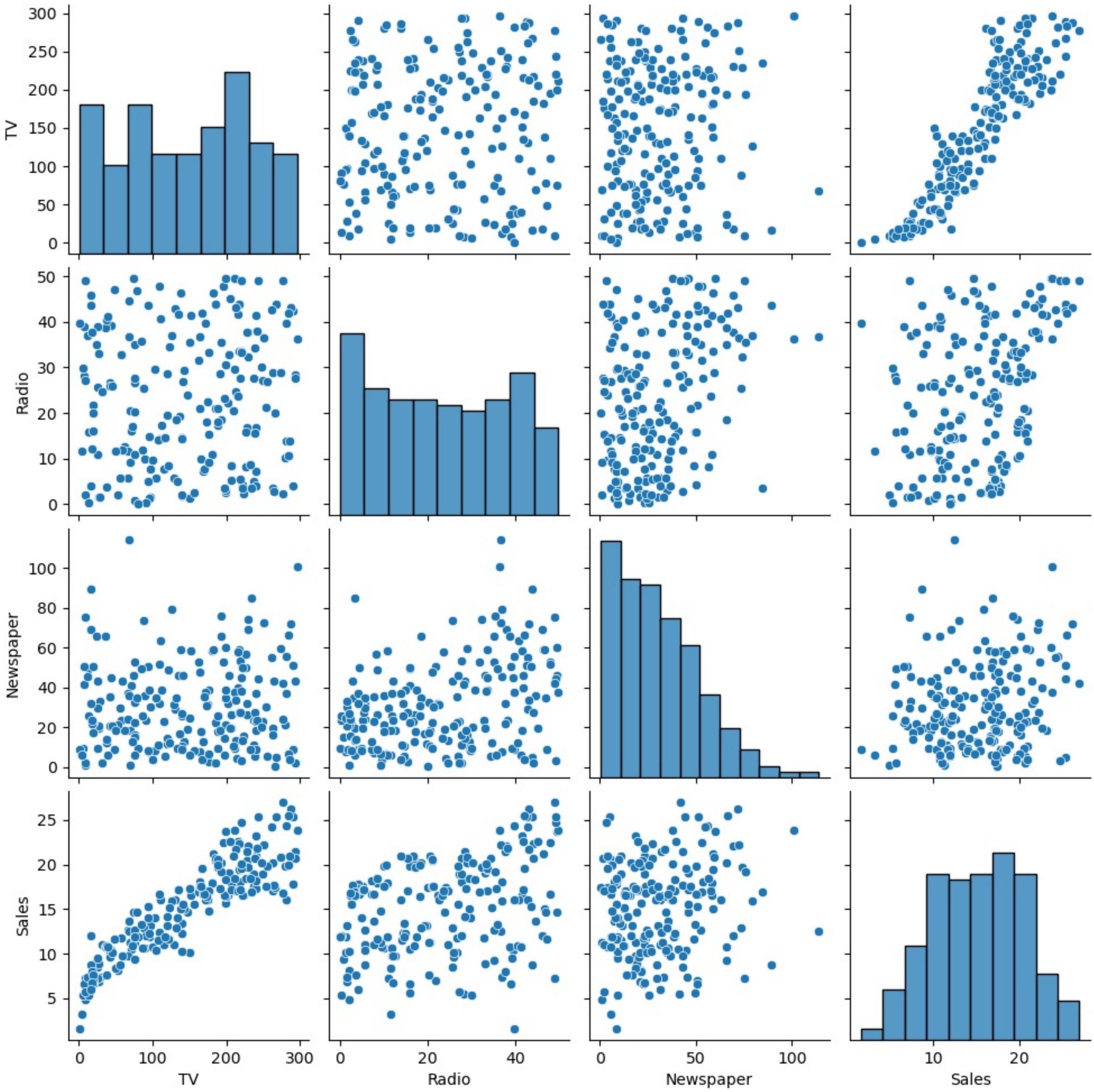
In [35]:
```
# the first few rows of the dataset
print(dataset.head())
```

```
      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
```

In [36]:
```
# Checking the missing values
print(dataset.isnull().sum())
```

```
TV           0
Radio        0
Newspaper    0
Sales        0
dtype: int64
```

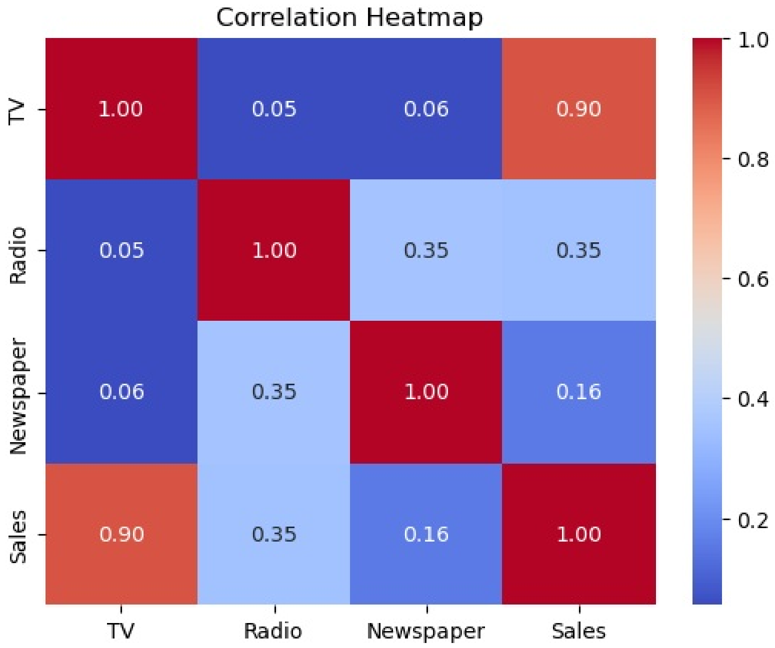
In [37]:
```
# Explore data distribution
sns.pairplot(dataset)
plt.show()
```



In [38]:

```
# Create a correlation matrix
correlation_matrix = dataset.corr()

# Plot a heatmap of the correlation matrix
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```



In both correlation analysis and regression, TV stands out as the advertising channel with the strongest and most positive impact on Sales, followed by Radio and then Newspaper. This reinforces the conclusion that increasing TV advertising spending is likely to have a more positive impact on Sales in this dataset.

TV as the Independent Variable

In [39]:

```
# TV as the independent variable
X_tv = dataset[['TV']]
y_tv = dataset['Sales']

# Split the data into training and testing sets
X_tv_train, X_tv_test, y_tv_train, y_tv_test = train_test_split(X_tv, y_tv, test_size=0.2, random_state=42)

# Initialize the model
model_tv = LinearRegression()

# Train the model
model_tv.fit(X_tv_train, y_tv_train)

# Make predictions on the test set
y_tv_pred = model_tv.predict(X_tv_test)

# Evaluate the model
mse_tv = mean_squared_error(y_tv_test, y_tv_pred)
r2_tv = r2_score(y_tv_test, y_tv_pred)

print(f'TV - Mean Squared Error: {mse_tv}')
print(f'TV - R-squared: {r2_tv}')
```

```
TV - Mean Squared Error: 6.101072906773963
TV - R-squared: 0.802561303423698
```

here a lower MSE 6.10 indicates better predictive performance and R-squared of 0.80 is quite high, indicating that 80% of the variance in Sales can be explained by the variation in TV spending.

Radio as the Independent Variable

In [29]:

```
# Radio as the independent variable
X_radio = dataset[['Radio']]
y_radio = dataset['Sales']

# Split the data into training and testing sets
X_radio_train, X_radio_test, y_radio_train, y_radio_test = train_test_split(X_radio, y_radio, test_size=0.2, ra

# Initialize the model
model_radio = LinearRegression()

# Train the model
model_radio.fit(X_radio_train, y_radio_train)

# Make predictions on the test set
y_radio_pred = model_radio.predict(X_radio_test)

# Evaluate the model
```

```
mse_radio = mean_squared_error(y_radio_test, y_radio_pred)
r2_radio = r2_score(y_radio_test, y_radio_pred)

print(f'Radio - Mean Squared Error: {mse_radio}')
print(f'Radio - R-squared: {r2_radio}')
```

```
Radio - Mean Squared Error: 27.595581948583174
Radio - R-squared: 0.10697088619553419
```

The higher MSE suggests that the model's predictions are less accurate compared to the TV model and A low R-squared (0.11) indicates that only 11% of the variance in Sales can be explained by the variation in Radio spending.

Newspaper as the Independent Variable

In [30]:

```
# Newspaper as the independent variable
X_newspaper = dataset[['Newspaper']]
y_newspaper = dataset['Sales']

# Split the data into training and testing sets
X_newspaper_train, X_newspaper_test, y_newspaper_train, y_newspaper_test = train_test_split(X_newspaper, y_news

# Initialize the model
model_newspaper = LinearRegression()

# Train the model
model_newspaper.fit(X_newspaper_train, y_newspaper_train)

# Make predictions on the test set
y_newspaper_pred = model_newspaper.predict(X_newspaper_test)

# Evaluate the model
mse_newspaper = mean_squared_error(y_newspaper_test, y_newspaper_pred)
r2_newspaper = r2_score(y_newspaper_test, y_newspaper_pred)

print(f'Newspaper - Mean Squared Error: {mse_newspaper}')
print(f'Newspaper - R-squared: {r2_newspaper}')
```

```
Newspaper - Mean Squared Error: 30.759376922769615
Newspaper - R-squared: 0.004586344085821592
```

similar to Radio, the higher MSE suggests less accurate predictions. The R-squared value of nearly zero indicates that Newspaper spending does not effectively predict Sales.

#Interpretation- The TV advertising spending has a strong positive correlation with Sales, as evidenced by the low MSE and high R-squared. Radio advertising, on the other hand, seems to have a weaker correlation with Sales, as indicated by the higher MSE and lower R-squared. Newspaper advertising, in this dataset, does not appear to have a significant linear relationship with Sales, as both MSE and R-squared are not favorable. In summary, based on this analysis, increasing TV advertising spending is likely to have a more positive impact on Sales compared to Radio or Newspaper advertising in this particular dataset.