Exploratory and explanatory data analysis for the Advertising dataset

Descriptive Data Analysis: Descriptive data analysis focuses on understanding the basic characteristics of the dataset. In this context, the following questions fall under descriptive data analysis:

1. What is the distribution of TV advertising spending?
2. What is the distribution of radio advertising spending?
3. What is the distribution of newspaper advertising spending?
4. What is the distribution of sales figures?

Exploratory Data Analysis: Exploratory data analysis involves examining relationships, patterns, and trends within the data. The following questions from your list fall under exploratory data analysis:

1. Is there a correlation between TV advertising spending and sales?
2. Is there a correlation between radio advertising spending and sales?
3. Is there a correlation between newspaper advertising spending and sales?

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
data= pd.read_csv("/home/c4leb/Desktop/C4LEB/Desktop/python_clas
```

Univariate, bivariate, and multivariate explorations are terms commonly used in data analysis and statistics to describe different ways of examining and analyzing data based on the number of variables involved. Here's an explanation of each:

UNIVARIATE EXPLORATION

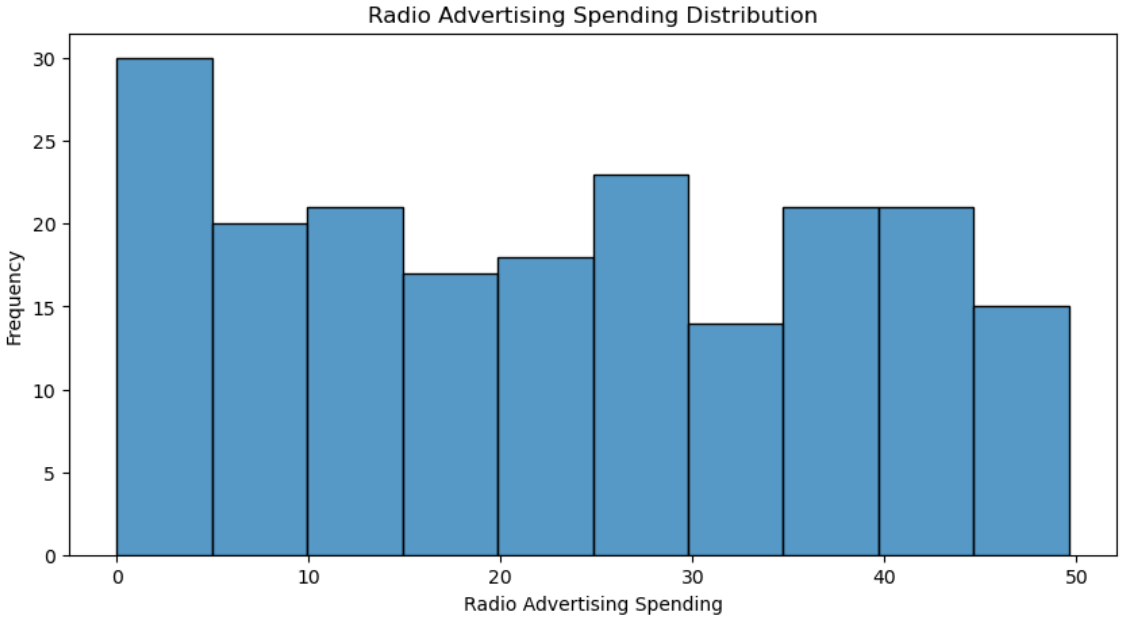
1. Univariate Exploration:

   - Univariate exploration focuses on a single variable at a time. It involves analyzing and visualizing the distribution and characteristics of a single data attribute (variable) in isolation.
   - Common techniques for univariate exploration include:
     - Histograms: To visualize the distribution of numerical data.
     - Bar charts: For visualizing the distribution of categorical data.
     - Summary statistics (mean, median, mode, variance, etc.) to describe the central tendency and variability of the data.
     - Box plots to visualize the spread and central tendency of data.
     - Probability density functions (PDFs) for continuous data.

   Univariate exploration helps you understand the characteristics of individual variables, such as their distribution, central tendency, and spread.

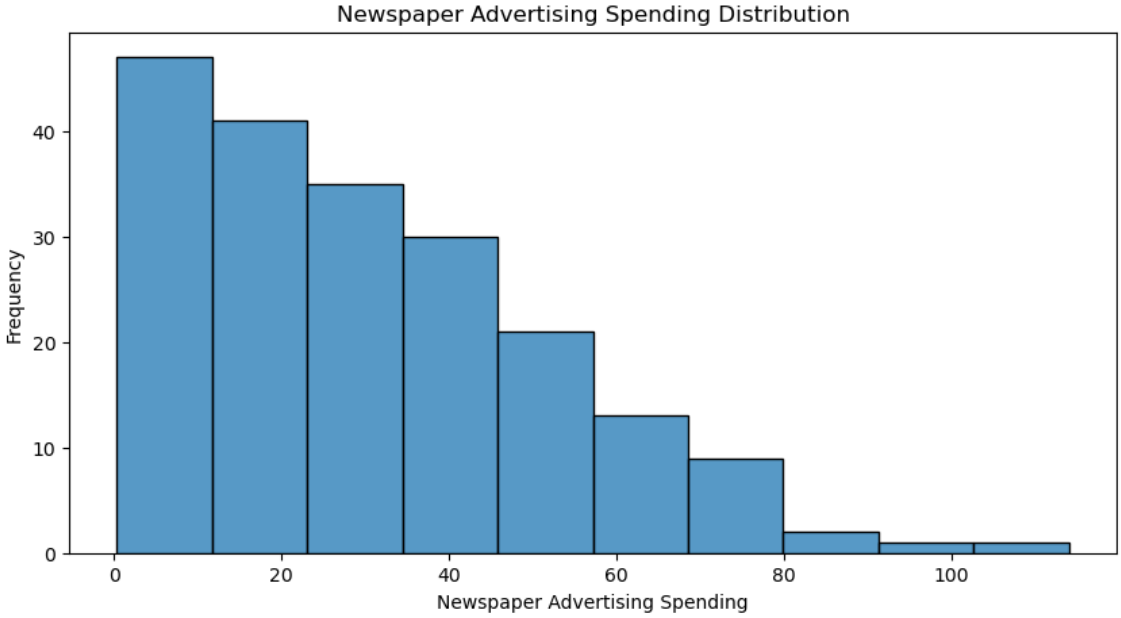
In [3]:

```
# Distribution of Radio advertising spending
plt.figure(figsize=(10, 5))
sns.histplot(data['Radio'], bins=10)
plt.title('Radio Advertising Spending Distribution')
plt.xlabel('Radio Advertising Spending')
plt.ylabel('Frequency')
```



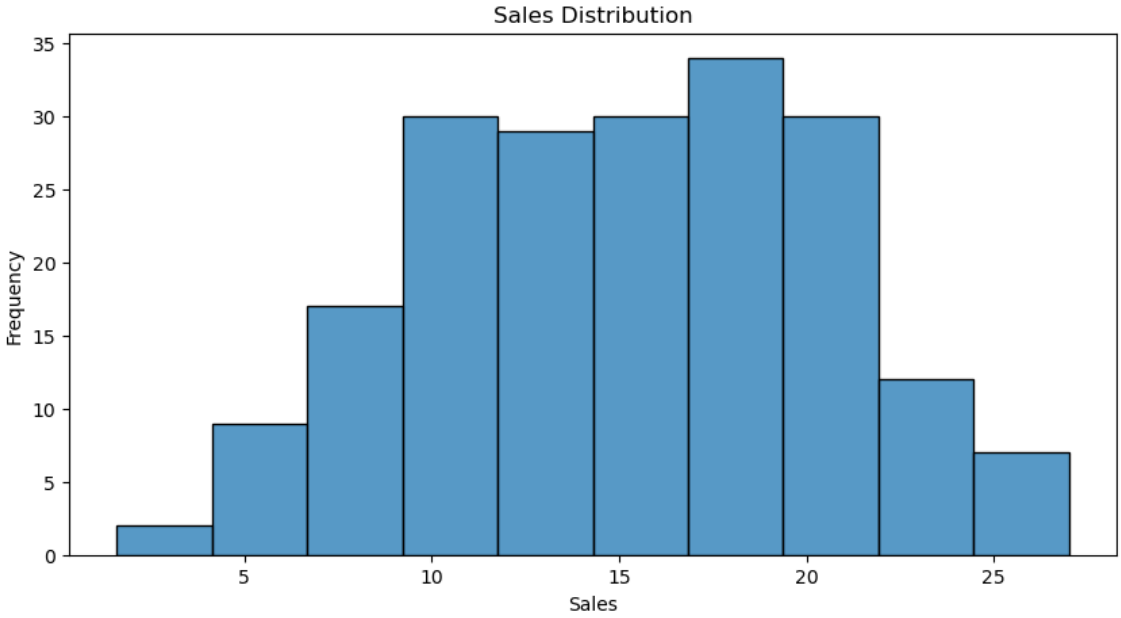
In [4]:

```
# Distribution of Newspaper advertising spending
plt.figure(figsize=(10, 5))
sns.histplot(data['Newspaper'], bins=10)
plt.title('Newspaper Advertising Spending Distribution')
plt.xlabel('Newspaper Advertising Spending')
plt.ylabel('Frequency')
```



In [5]:

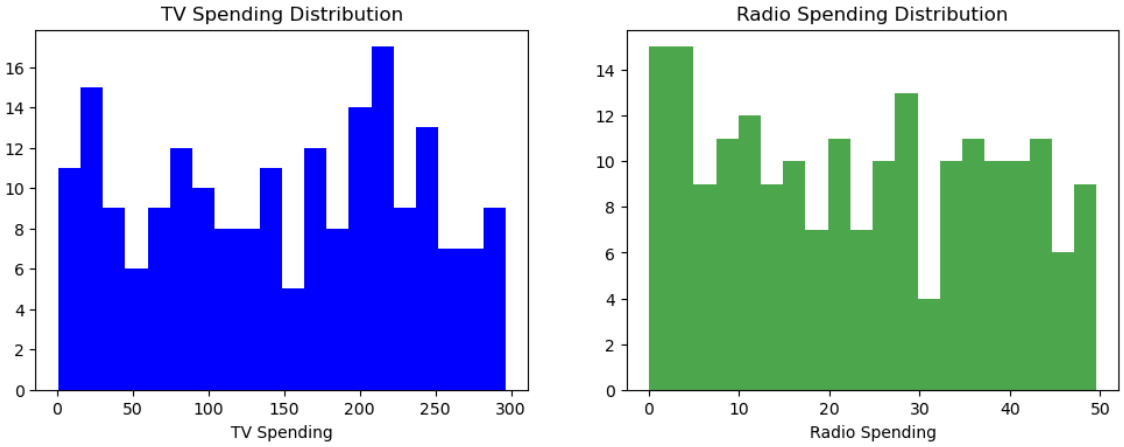
```
# Distribution of Sales
plt.figure(figsize=(10, 5))
sns.histplot(data['Sales'], bins=10)
plt.title('Sales Distribution')
plt.xlabel('Sales')
plt.ylabel('Frequency')
```



In [58]:

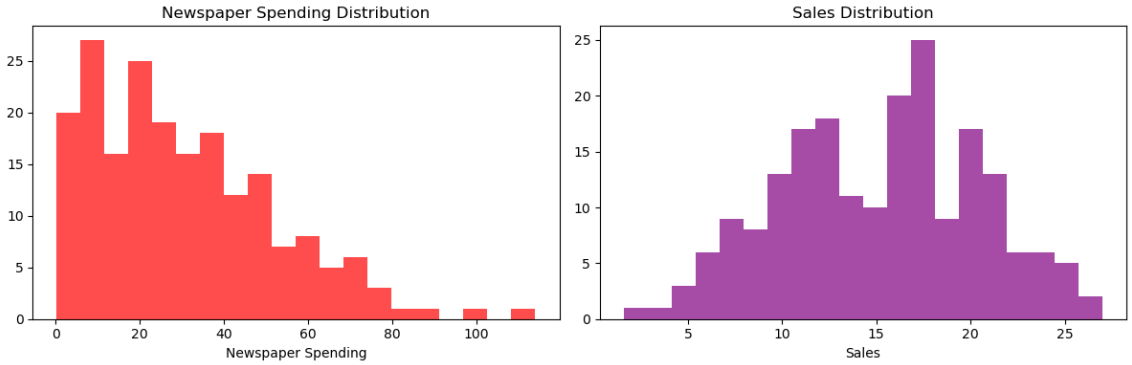
```
# Histograms to visualize distributions
plt.figure(figsize=(12, 4))
plt.subplot(121)
plt.hist(data['TV'], bins=20, color='blue', alpha=1)
plt.title('TV Spending Distribution')
plt.xlabel('TV Spending')
plt.subplot(122)
plt.hist(data['Radio'], bins=20, color='green', alpha=0.7)
plt.title('Radio Spending Distribution')
plt.xlabel('Radio Spending')
```

Out[58]: Text(0.5, 0, 'Radio Spending')



In [62]:

```
plt.figure(figsize=(12, 4))
plt.subplot(121)
plt.hist(data['Newspaper'], bins=20, color='red', alpha=0.7)
plt.title('Newspaper Spending Distribution')
plt.xlabel('Newspaper Spending')
plt.subplot(122)
plt.hist(data['Sales'], bins=20, color='purple', alpha=0.7)
plt.title('Sales Distribution')
plt.xlabel('Sales')
plt.tight_layout()
```

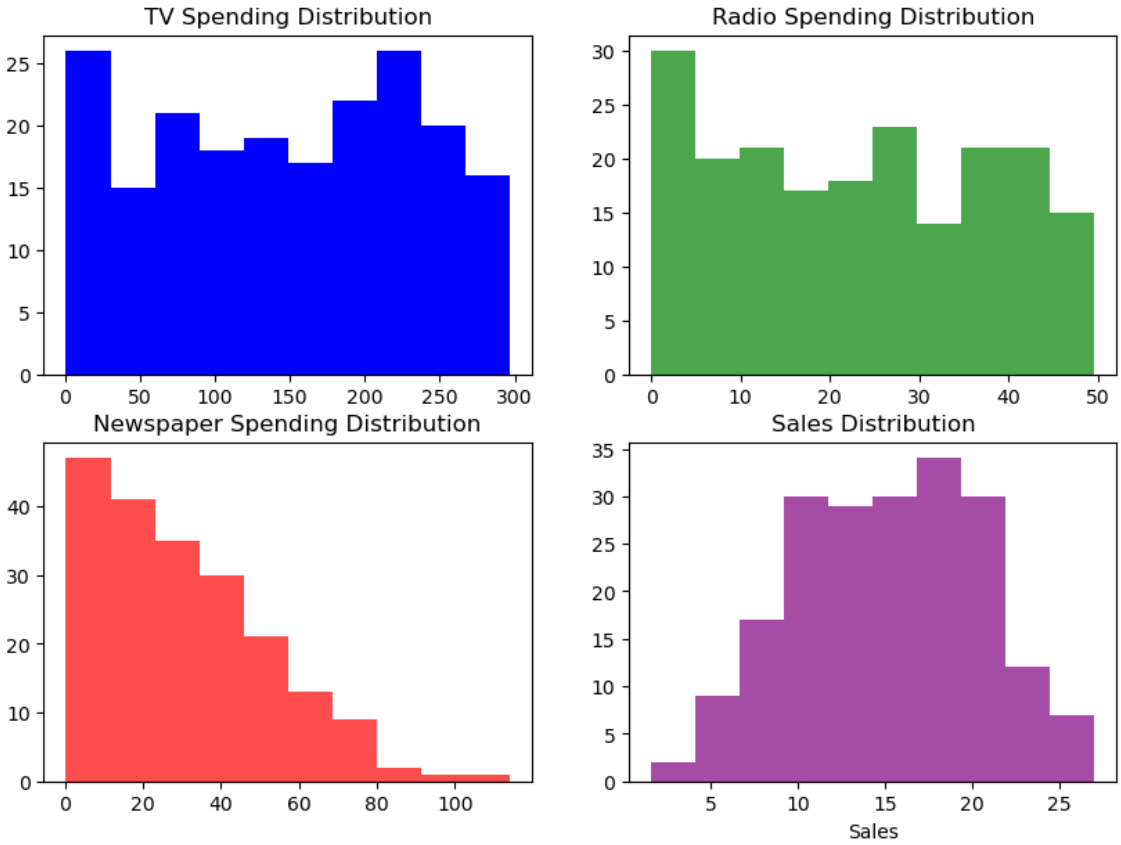


USING AX(axis) to set the subplots

In [81]:

```
fig, ax = plt.subplots(nrows=2, ncols=2, figsize=(10,7))
# Histograms to visualize distributions

ax[0, 0].hist(data['TV'], bins=10, color='blue', alpha=1)
ax[0, 0].set_title('TV Spending Distribution')
plt.xlabel('TV Spending')
# plt.subplot(122)
ax[0, 1].hist(data['Radio'], bins=10, color='green', alpha=0.7)
ax[0, 1].set_title('Radio Spending Distribution')
plt.xlabel('Radio Spending')

# plt.figure(figsize=(12, 4))
# plt.subplot(121)
ax[1, 0].hist(data['Newspaper'], bins=10, color='red', alpha=0.7
ax[1, 0].set_title('Newspaper Spending Distribution')
plt.xlabel('Newspaper Spending')
# plt.subplot(122)
ax[1, 1].hist(data['Sales'], bins=10, color='purple', alpha=0.7)
ax[1, 1].set_title('Sales Distribution')
plt.xlabel('Sales')
# plt.tight_layout()
```

Out[81]: Text(0.5, 0, 'Sales')



BIVARIATE EXPLORATION

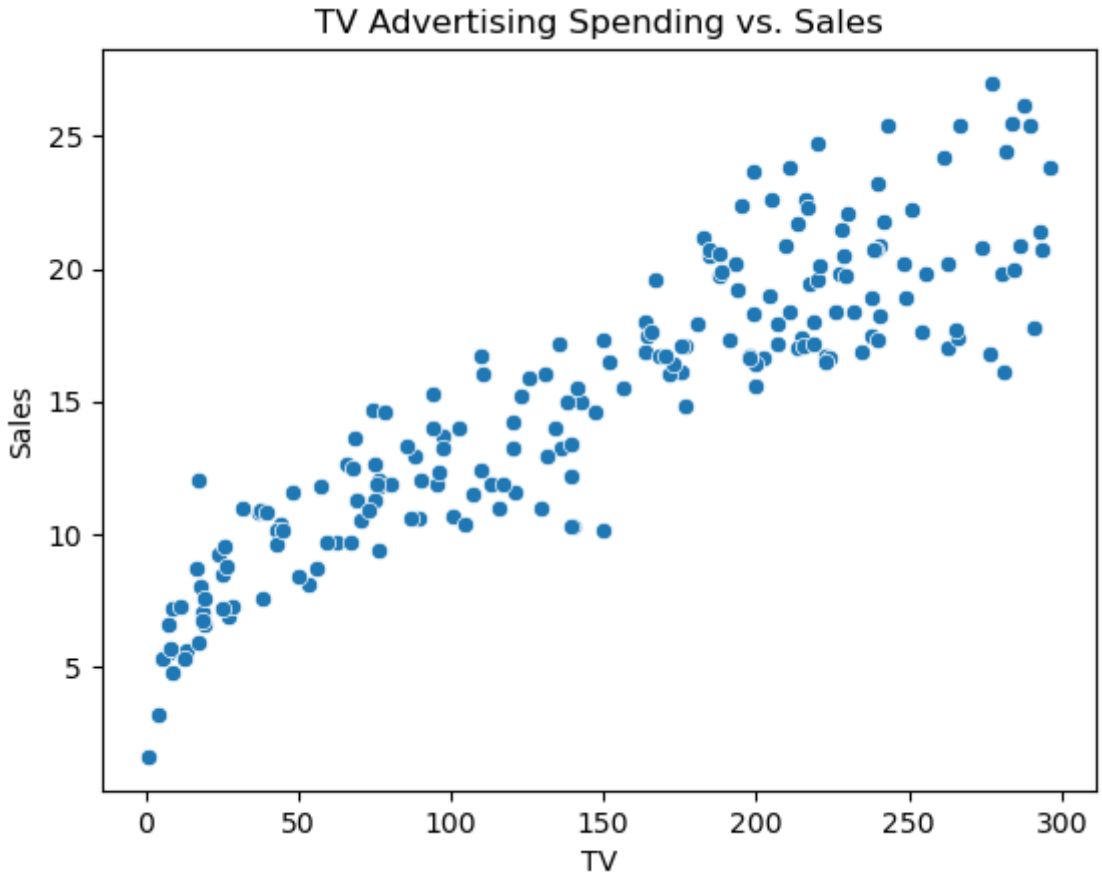
2. Bivariate Exploration:

- Bivariate exploration involves examining the relationship between two variables. It seeks to understand how one variable is related to another and how they interact.
- Common techniques for bivariate exploration include:
  - Scatter plots: To visualize the relationship between two numerical variables.

- Bar charts or stacked bar charts: To visualize relationships between a categorical variable and a numerical variable.
- Correlation analysis: To quantify the strength and direction of the relationship between two numerical variables.
- Cross-tabulations or contingency tables: To examine relationships between two categorical variables.

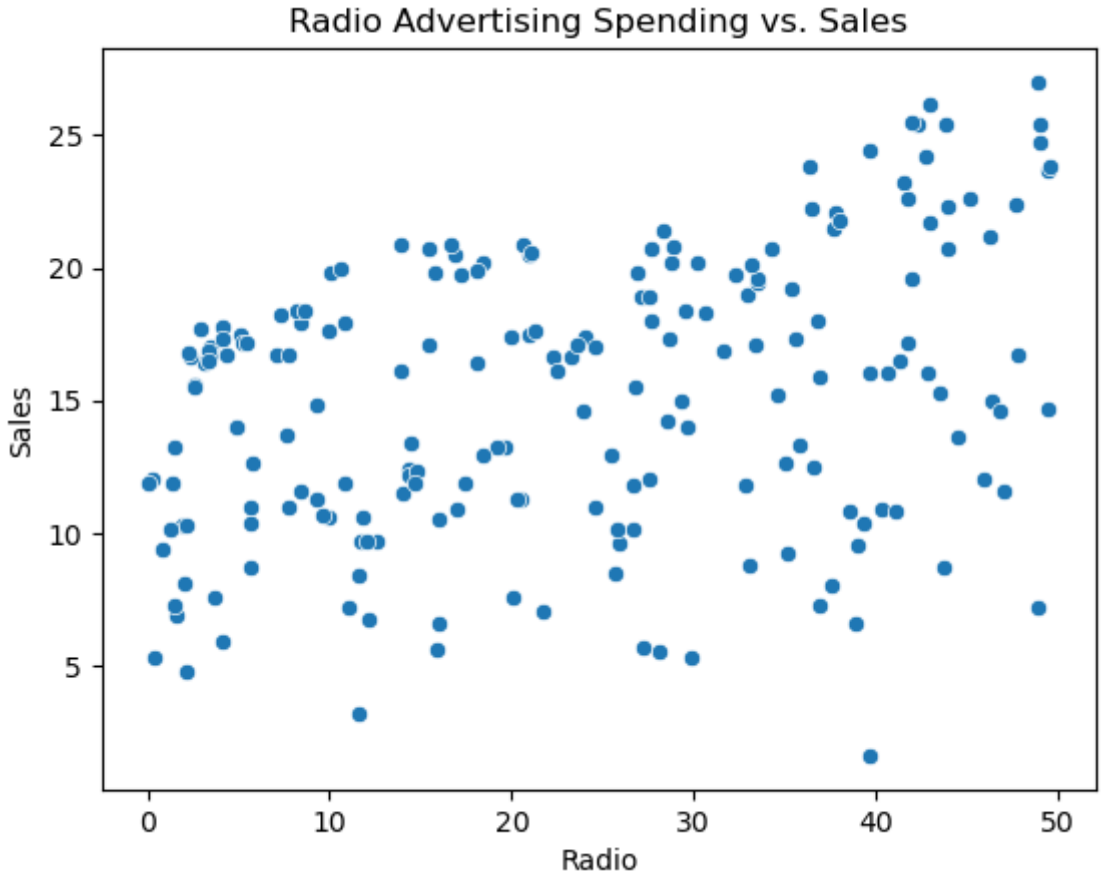
Bivariate exploration helps you uncover patterns, dependencies, and associations between pairs of variables, which can be valuable for understanding cause-and-effect relationships or making predictions.

In [7]:
```
# Correlation between TV advertising spending and sales
sns.scatterplot(x='TV', y='Sales', data=data)
plt.title('TV Advertising Spending vs. Sales')
plt.show()
```



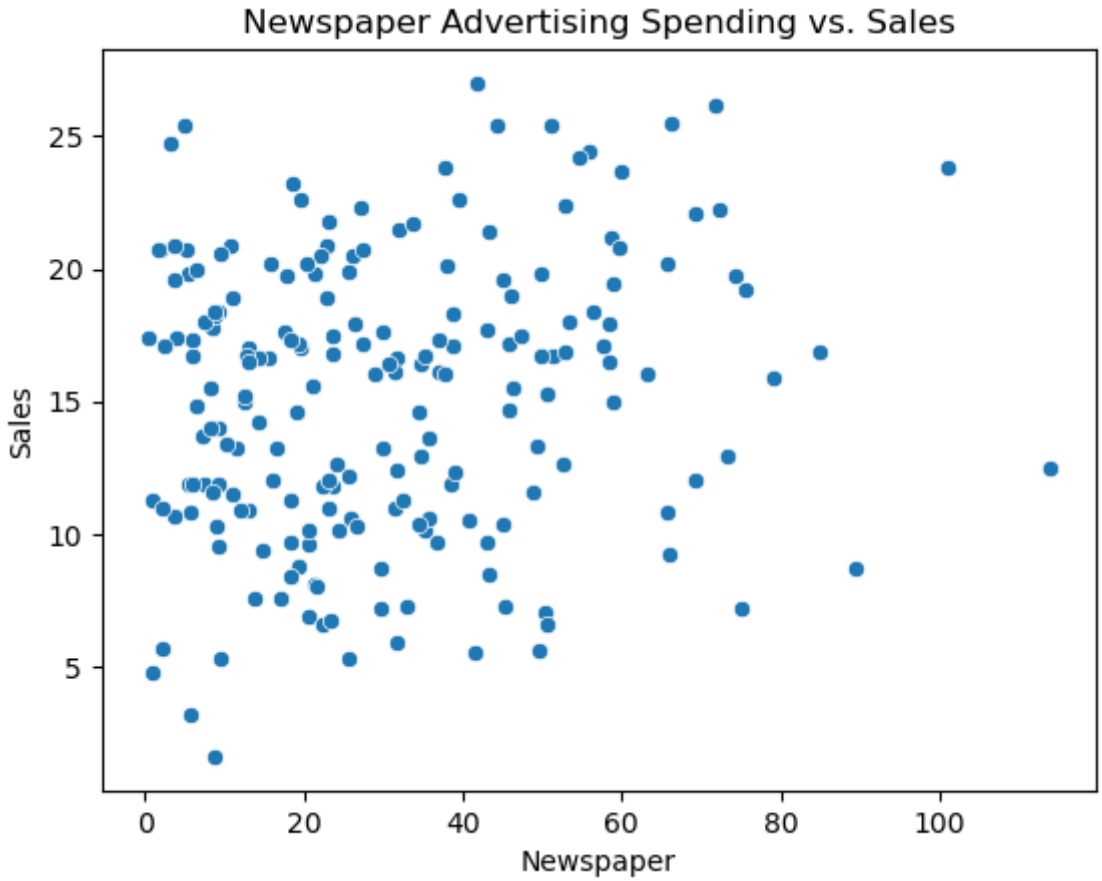
In [8]:

```
1
2 # Correlation between Radio advertising spending and sales
3 sns.scatterplot(x='Radio', y='Sales', data=data)
4 plt.title('Radio Advertising Spending vs. Sales')
```



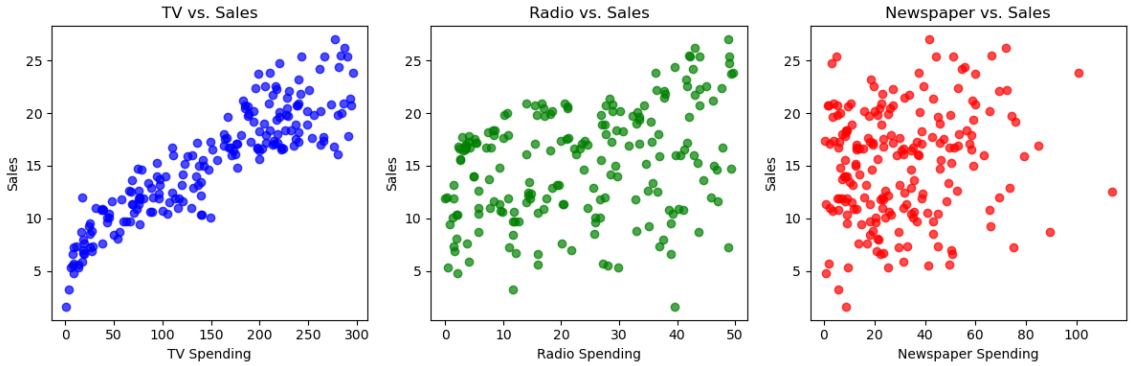
In [9]:

```
# Correlation between Newspaper advertising spending and sales
sns.scatterplot(x='Newspaper', y='Sales', data=data)
plt.title('Newspaper Advertising Spending vs. Sales')
plt.show()
```



In [30]:

```
# Scatter plot to visualize relationships
plt.figure(figsize=(12, 4))
plt.subplot(131)
plt.scatter(data['TV'], data['Sales'], color='blue', alpha=0.7)
plt.title('TV vs. Sales')
plt.xlabel('TV Spending')
plt.ylabel('Sales')
plt.subplot(132)
plt.scatter(data['Radio'], data['Sales'], color='green', alpha=0
plt.title('Radio vs. Sales')
plt.xlabel('Radio Spending')
plt.ylabel('Sales')
plt.subplot(133)
plt.scatter(data['Newspaper'], data['Sales'], color='red', alpha
plt.title('Newspaper vs. Sales')
plt.xlabel('Newspaper Spending')
plt.ylabel('Sales')
plt.tight_layout()
plt.show()
```

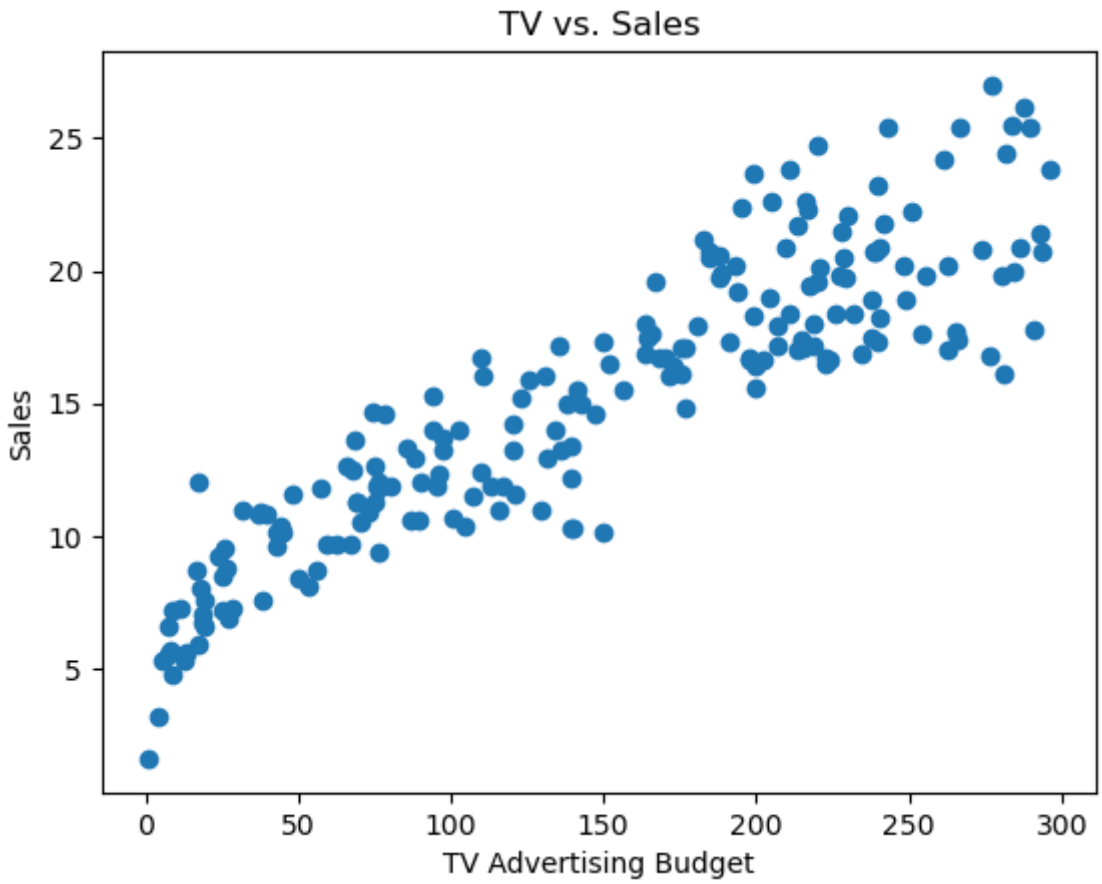


In [ ]:

```
1
```

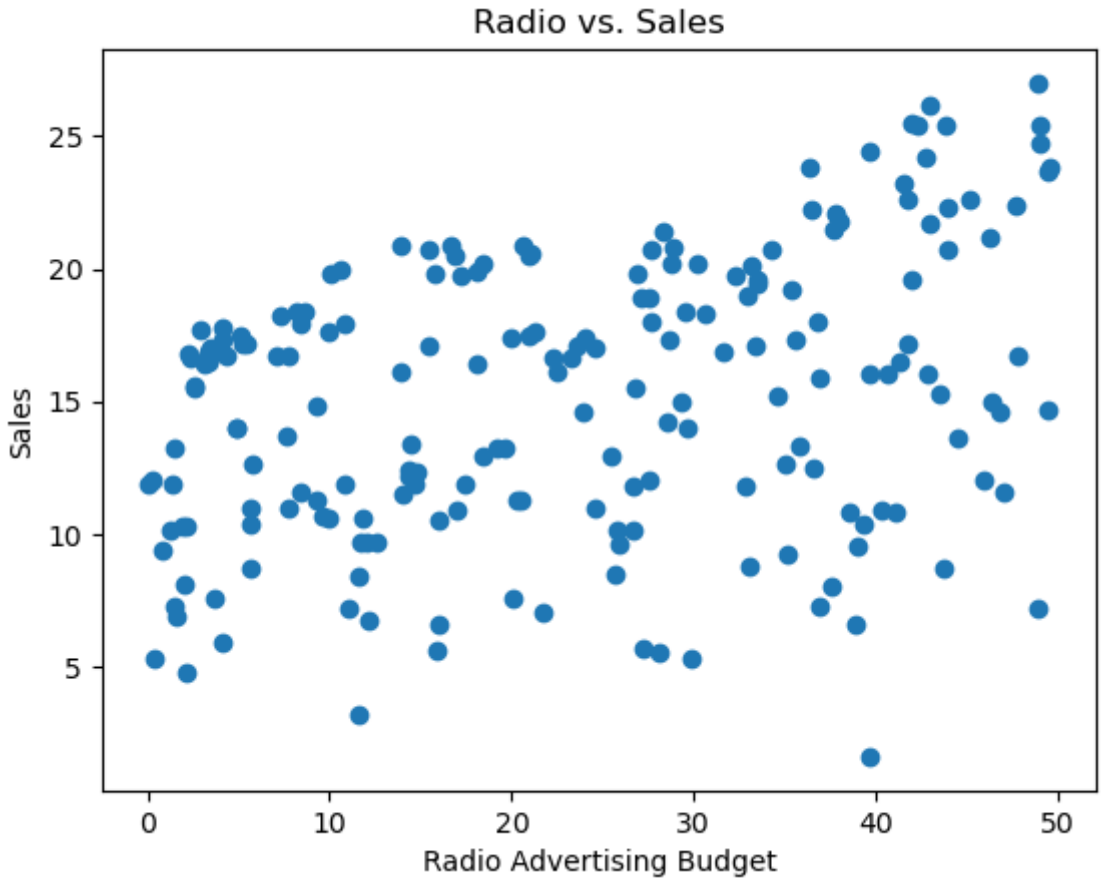
In [15]:

```
# Bivariate Plots
# Scatter plot of TV vs. Sales
plt.scatter(data['TV'], data['Sales'])
plt.xlabel('TV Advertising Budget')
plt.ylabel('Sales')
plt.title('TV vs. Sales')
```



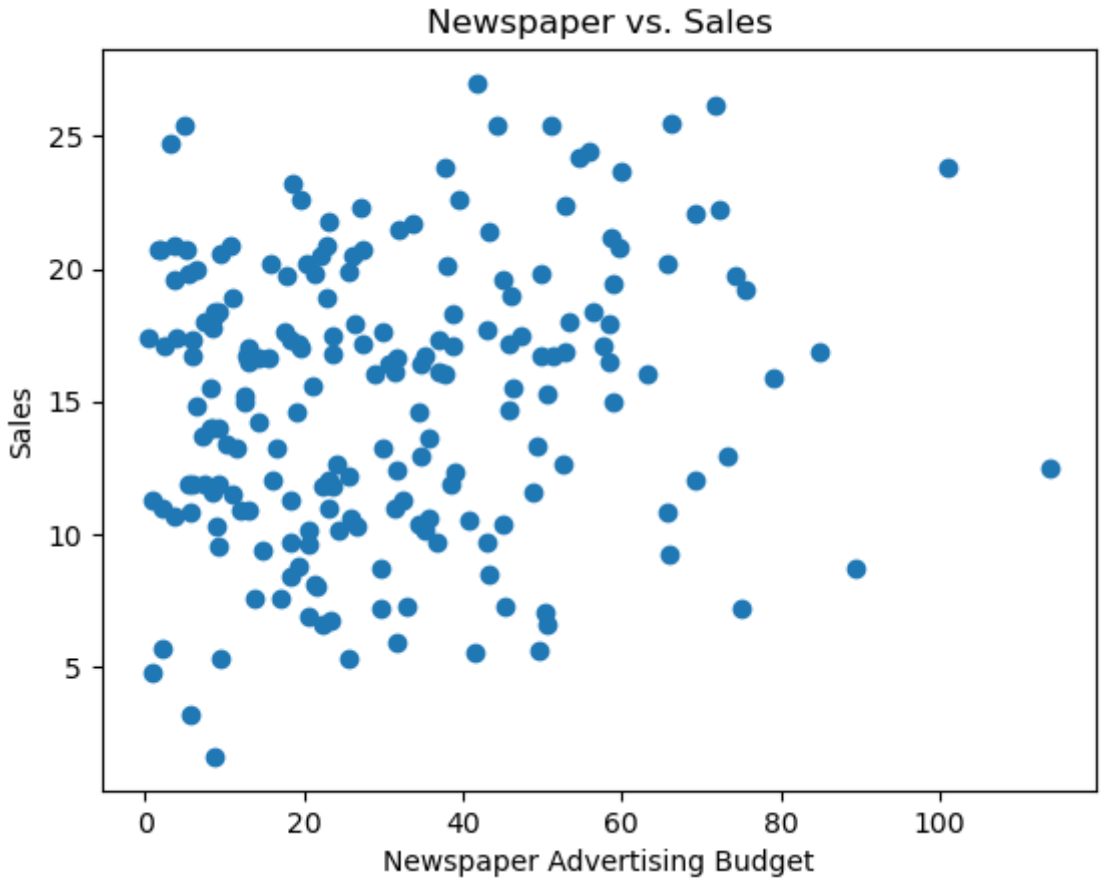
In [16]:

```
# Scatter plot of Radio vs. Sales
plt.scatter(data['Radio'], data['Sales'])
plt.xlabel('Radio Advertising Budget')
plt.ylabel('Sales')
plt.title('Radio vs. Sales')
```



In [17]:

```
# Scatter plot of Newspaper vs. Sales
plt.scatter(data['Newspaper'], data['Sales'])
plt.xlabel('Newspaper Advertising Budget')
plt.ylabel('Sales')
plt.title('Newspaper vs. Sales')
```



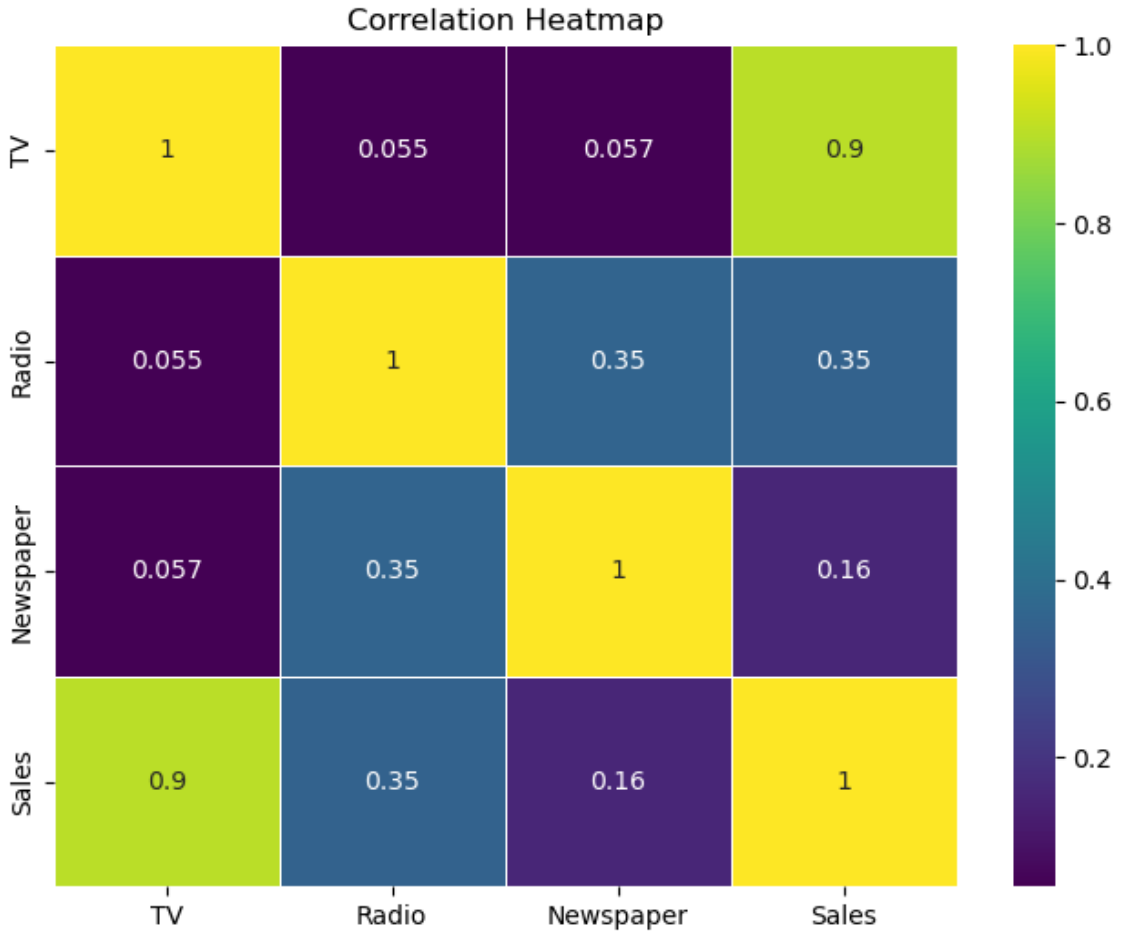
Multivariate Exploration

3. Multivariate Exploration:
    - Multivariate exploration involves the analysis of three or more variables simultaneously. It aims to understand how multiple variables interact with each other and how they collectively affect the data.
    - Common techniques for multivariate exploration include:
        - Multivariate scatter plots or 3D scatter plots: To visualize the relationships between three or more numerical variables.
        - Multivariate regression analysis: To model the relationships between multiple independent variables and a dependent variable.
        - Principal Component Analysis (PCA) or Factor Analysis: To reduce the dimensionality of data while preserving as much information as possible.
        - Clustering and classification techniques: To group data points into clusters or classes based on multiple variables.

Multivariate exploration is essential for understanding complex datasets with multiple interacting factors. It allows you to uncover patterns and dependencies that may not be apparent in univariate or bivariate analyses.

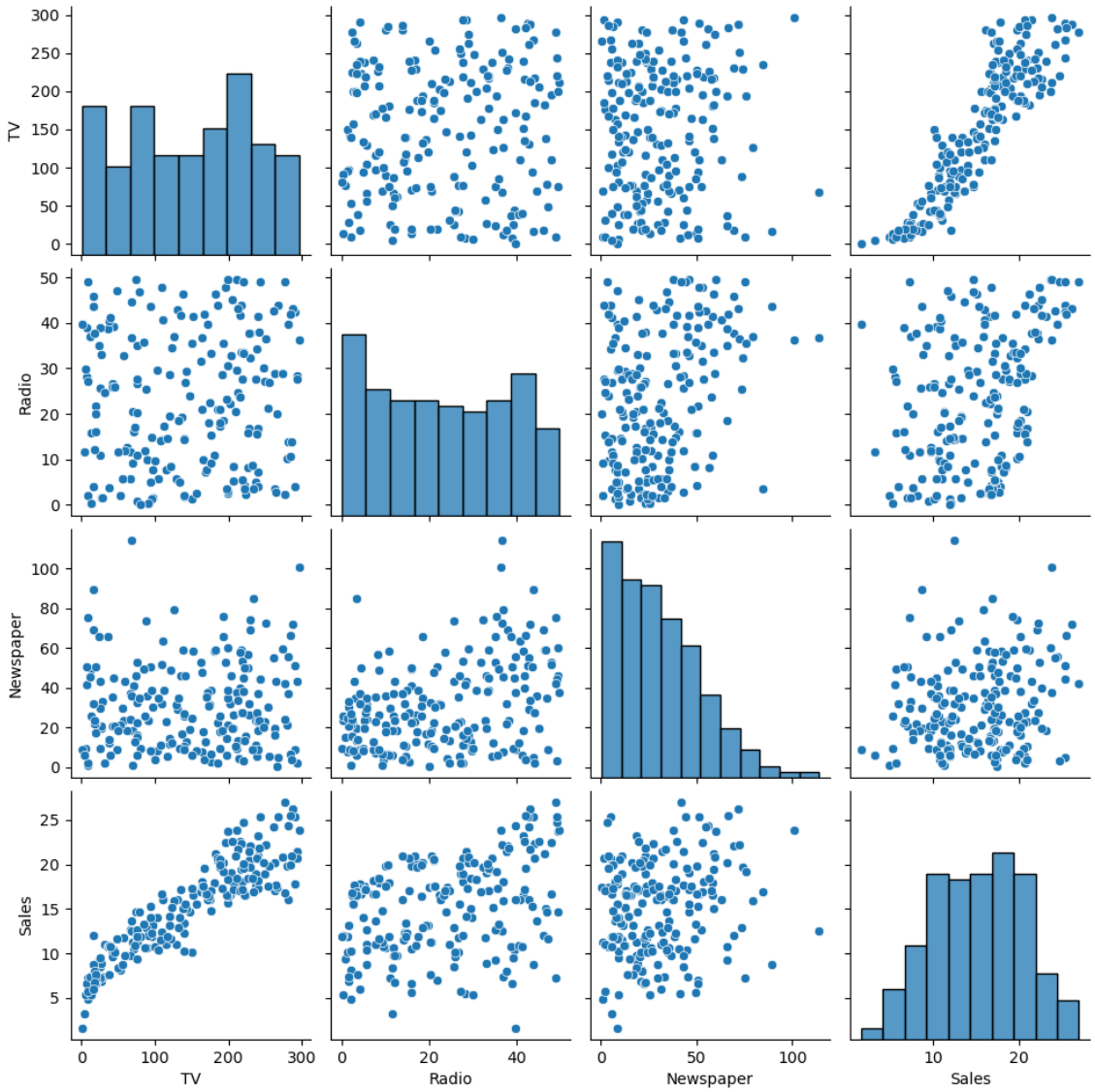
In [21]:

```
import seaborn as sns

# Correlation matrix
corr_matrix = data.corr()

# Heatmap to visualize correlations
plt.figure(figsize=(8, 6))
sns.heatmap(corr_matrix, annot=True, cmap='viridis', linewidths=(
plt.title('Correlation Heatmap')
plt.show()
```



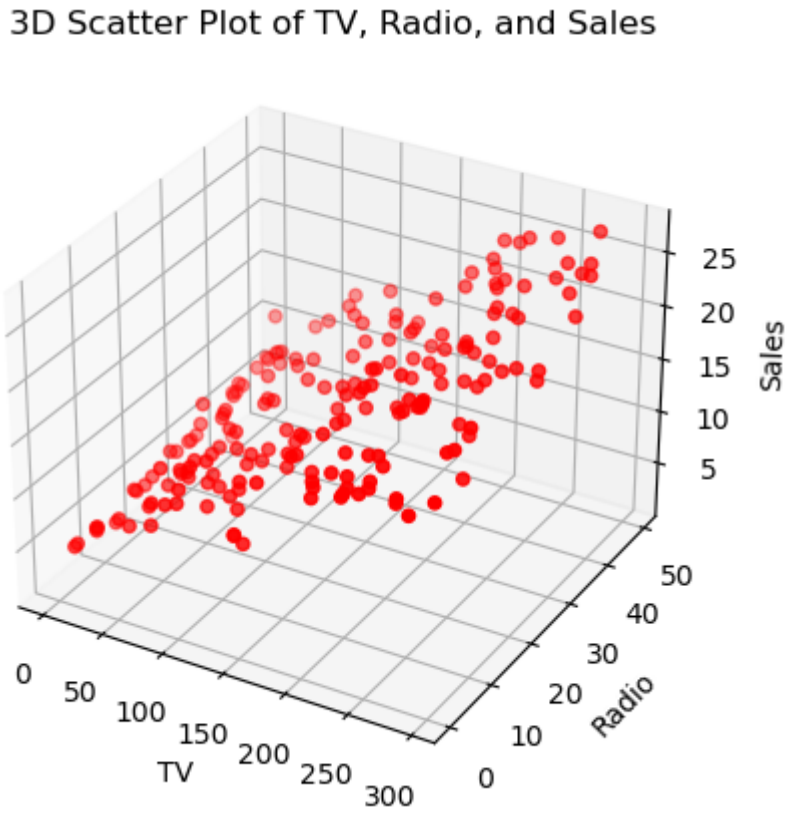
In [22]:

```
1
2 # Multivariate Plot
3 # Pairplot to visualize relationships between all variables
4 sns.pairplot(data)
5 plt.show()
```



In [23]:

```
# Multivariate Plot
# 3D Scatter plot of TV, Radio, and Sales
from mpl_toolkits.mplot3d import Axes3D
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(data['TV'], data['Radio'], data['Sales'], c='r', mark
ax.set_xlabel('TV')
ax.set_ylabel('Radio')
ax.set_zlabel('Sales')
plt.title('3D Scatter Plot of TV, Radio, and Sales')
```



In summary, univariate exploration focuses on individual variables, bivariate exploration examines relationships between two variables, and multivariate exploration delves into the interactions and patterns among three or more variables. These approaches are fundamental to data analysis and are used to gain insights, make predictions, and inform decision-making in various fields, including statistics, data science, and machine learning.

In [ ]:

```
1
```