In [1]:
```
import numpy as np
import pandas as pd

file = r'E:\DATA_SCIENCE\Intership\Task-4\advertising.csv'
data = pd.read_csv(file)
data
```

Out[1]:

	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 × 4 columns

In [2]:
```
data.describe()
```

Out[2]:

	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

In [3]:
```
data.isna().sum()
```

Out[3]:
```
TV           0
Radio        0
Newspaper    0
Sales        0
dtype: int64
```

In [4]:
```
data.dtypes
```

Out[4]:
```
TV           float64
Radio        float64
Newspaper    float64
Sales        float64
dtype: object
```

In [5]:
```
data.corr()
```

Out[5]:

	TV	Radio	Newspaper	Sales
TV	1.000000	0.054809	0.056648	0.901208
Radio	0.054809	1.000000	0.354104	0.349631
Newspaper	0.056648	0.354104	1.000000	0.157960
Sales	0.901208	0.349631	0.157960	1.000000

In [6]:
```
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
import statsmodels.api as sm
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score

x =  data[['TV', 'Radio', 'Newspaper']]
y = data["Sales"]


X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, ra
model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

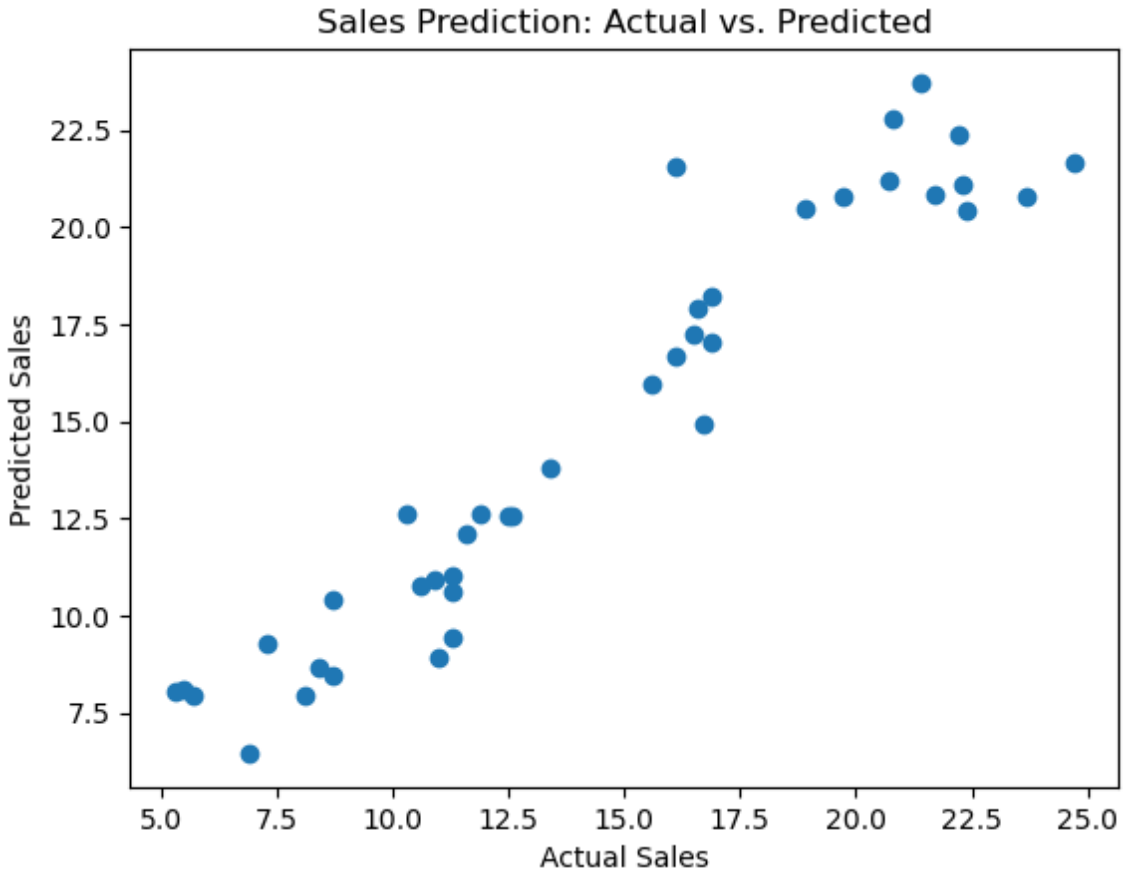
In [7]:
```
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'R-squared Score: {r2}')
```

```
Mean Squared Error: 2.9077569102710923
R-squared Score: 0.9059011844150826
```

In [8]:
```
import matplotlib.pyplot as plt

plt.scatter(y_test, y_pred)
plt.xlabel('Actual Sales')
plt.ylabel('Predicted Sales')
plt.title('Sales Prediction: Actual vs. Predicted')
plt.show()
```



In [ ]: