Experiment 4: Linear Regression

Step 1: Import Libraries and Load Data

```
import warnings
warnings.filterwarnings('ignore')

import numpy as np
import pandas as pd

advertising = pd.read_csv("advertising.csv")
advertising.head()
```

Output:

```
   TV   | Radio | Newspaper | Sales
-------|-------|-----------|-------
230.1  | 37.8  | 69.2      | 22.1
44.5   | 39.3  | 45.1      | 10.4
17.2   | 45.9  | 69.3      | 9.3
151.5  | 41.3  | 58.5      | 18.5
180.8  | 10.8  | 58.4      | 12.9
```

Step 2: Visualize Data

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.pairplot(advertising, x_vars=['TV', 'Newspaper', 'Radio'], y_vars='Sales', size=4, aspect=1, kind='scatter')
plt.show()

sns.heatmap(advertising.corr(), cmap="YlGnBu", annot=True)
plt.show()
```

Output:

```
Pairplot and Heatmap (visual graphs, not shown in HTML)
```

Step 3: Prepare Data for Model

```
X = advertising['TV']
y = advertising['Sales']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size = 0.7, test_size = 0.3, random_state = 100)

X_train.head()
y_train.head()
```

Output:

```
X_train:
0    230.1
1     44.5
2     17.2
3    151.5
4    180.8

y_train:
0    22.1
1    10.4
2     9.3
3    18.5
4    12.9
```

Step 4: Build Model using statsmodels

```
import statsmodels.api as sm

X_train_sm = sm.add_constant(X_train)

lr = sm.OLS(y_train, X_train_sm).fit()
print(lr.params)
print(lr.summary())
```

Output:

```
const    6.989666
TV       0.046592
dtype: float64

OLS Regression Results:
==============================================================================
Dep. Variable:                      y   R-squared:                       0.815
Model:                            OLS   Adj. R-squared:                  0.813
Method:                 Least Squares   F-statistic:                     601.5
Date:                ...                Prob (F-statistic):           1.58e-52
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          6.9897      0.457     15.288      0.000       6.087       7.892
TV             0.0466      0.002     24.525      0.000       0.043       0.050
==============================================================================
```