pract-linear-regression

June 7, 2023

```
[7]: import pandas as pd
     import numpy as np
     import matplotlib.pyplot as plt
     from sklearn.linear_model import LinearRegression
     from sklearn.metrics import r2_score
     from sklearn.metrics import classification_report,confusion_matrix
```

```
[3]: data=pd.read_csv('50_Startups.csv')
```

```
[4]: print("Descreptive Statistics:\n",data.describe())
```

```
Descreptive Statistics:
            R&D Spend  Administration  Marketing Spend         Profit
count      50.000000       50.000000        50.000000      50.000000
mean    73721.615600   121344.639600    211025.097800  112012.639200
std     45902.256482    28017.802755    122290.310726   40306.180338
min         0.000000    51283.140000         0.000000   14681.400000
25%     39936.370000   103730.875000    129300.132500   90138.902500
50%     73051.080000   122699.795000    212716.240000  107978.190000
75%    101602.800000   144842.180000    299469.085000  139765.977500
max    165349.200000   182645.560000    471784.100000  192261.830000
```
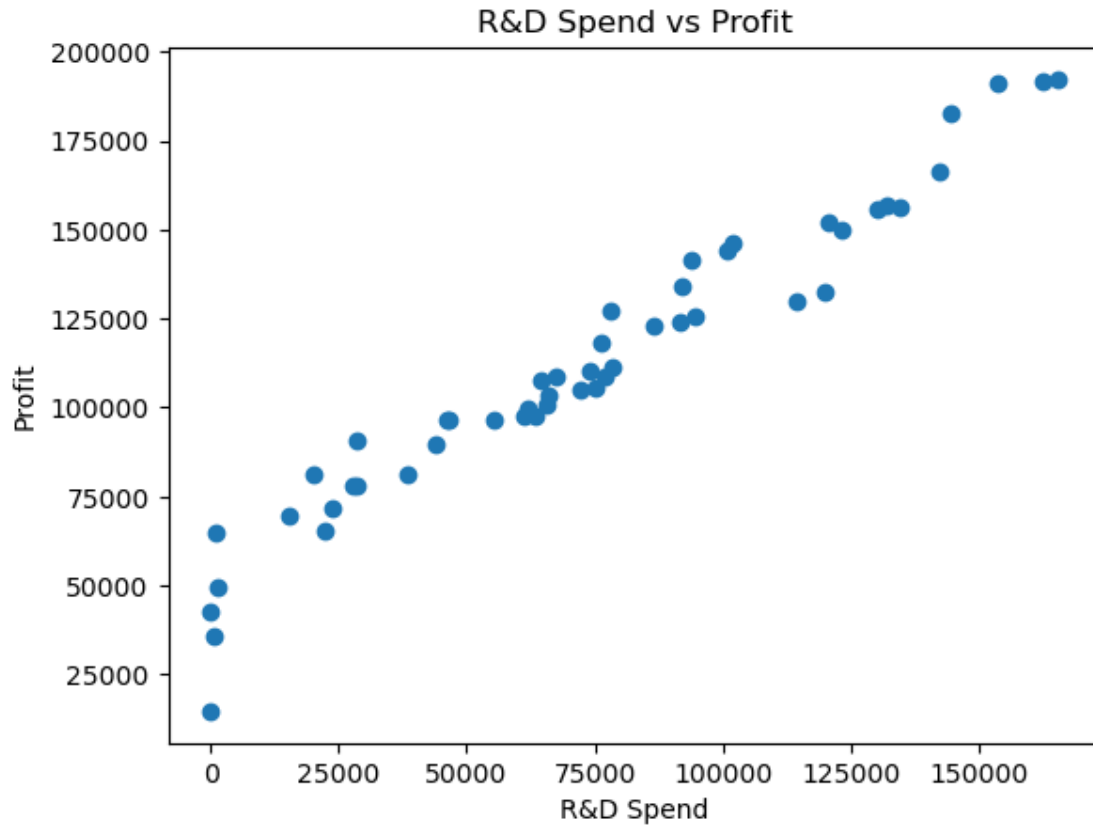
```
[6]: x=data['R&D Spend'].values.reshape(-1,1)
     y=data['Profit'].values
```

```
[8]: plt.scatter(x,y)
     plt.xlabel('R&D Spend')
     plt.ylabel('Profit')
     plt.title('R&D Spend vs Profit')
     plt.show
```

```
[8]: <function matplotlib.pyplot.show(close=None, block=None)>
```

R&D Spend vs Profit

```
[9]:  model=LinearRegression()
      model.fit(x,y)
```

```
[9]:  LinearRegression()
```

```
[10]:  coefficient=model.coef_[0]
       intercept=model.intercept_
       print("Regression Equation: y = {:.2f}x + {:.2f}".format(coefficient,␣
         ↪intercept))
```

```
Regression Equation: y = 0.85x + 49032.90
```

```
[12]:  y_pred=model.predict(x)
       r2=r2_score(y,y_pred)
       print("R2 Statistics:",r2)
```
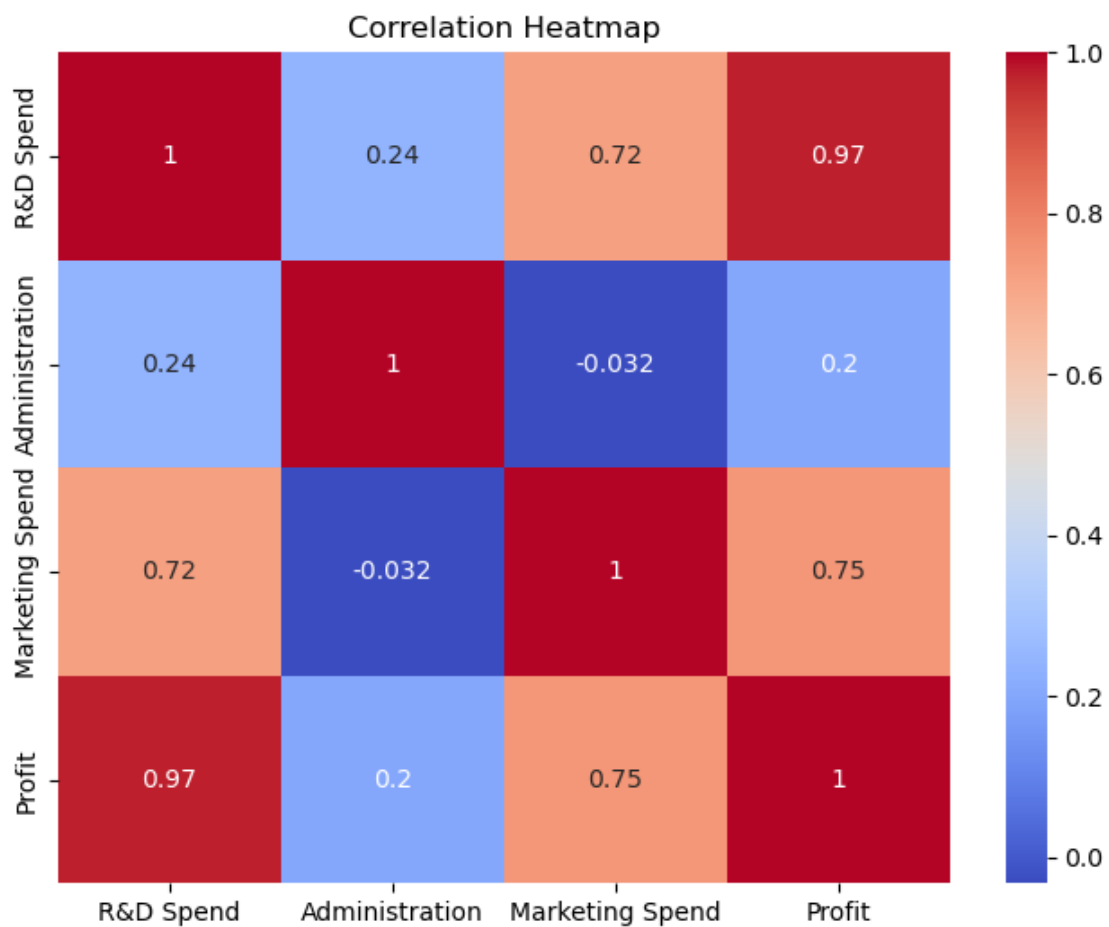
```
R2 Statistics: 0.9465353160804393
```

```
[13]:  corr_matrix=data.corr()
       print(corr_matrix)
```

	R&D Spend	Administration	Marketing Spend	Profit
R&D Spend	1.000000	0.241955	0.724248	0.972900
Administration	0.241955	1.000000	-0.032154	0.200717
Marketing Spend	0.724248	-0.032154	1.000000	0.747766
Profit	0.972900	0.200717	0.747766	1.000000

```python
import seaborn as sns
plt.figure(figsize=(8,6))
sns.heatmap(corr_matrix,annot=True,cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show
```

[15]: <function matplotlib.pyplot.show(close=None, block=None)>