Aditi Rao 118A2088

Please find dataset 2 in the shared folder of Day 5. Apply simple linear regression on this dataset and estimate profit if expenditure on administration is 90,000 Rs. Upload py or ipynb file here. Write answer in Q.7

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

```
from google.colab import files
uploaded = files.upload()
```

```
Choose Files  No file chosen    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving data2.csv to data2.csv
```

```
dataset = pd.read_csv('data2.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values
print(X)
print(y)
```

```
[[136897.8 ]
 [151377.59]
 [101145.55]
 [118671.85]
 [ 91391.77]
 [ 99814.71]
 [147198.87]
 [145530.06]
 [148718.95]
 [108679.17]
 [110594.11]
 [ 91790.61]
 [127320.38]
 [135495.07]
 [156547.42]
 [122616.84]
 [121597.55]
 [145077.58]
 [114175.79]
 [153514.11]
 [113867.3 ]
 [153773.43]
 [122782.75]
 [105751.03]
 [ 99281.34]
 [139553.16]
 [144135.98]
 [127864.55]
 [182645.56]
 [153032.06]
 [115641.28]
 [152701.92]
 [129219.61]
 [103057.49]
 [157693.92]
 [ 85047.44]
 [127056.21]
 [ 51283.14]
 [ 65947.93]
 [ 82982.09]
 [118546.05]
 [ 84710.77]
 [ 96189.63]
 [127382.3 ]
 [154806.14]
 [124153.04]
 [115816.21]
 [135426.92]
```

```
 [ 51743.15]
 [116983.8 ]]
[192261.83 191792.06 191050.39 182901.99 166187.94 156991.12 156122.51
 155752.6  152211.77 149759.96 146121.95 144259.4  141585.52 134307.35
 132602.65 129917.04 126992.93 125370.37 124266.9  122776.86 118474.03
 111313.02 110352.25 108733.99 108552.04 107404.34 105733.54 105008.31
 103282.38 101004.64  99937.59  97483.56  97427.84  96778.92  96712.8
  96479.51  90708.19  89949.14  81229.06  81005.76  78239.91  77798.83
  71498.49  69758.98  65200.33  64926.08  49490.75  42559.73  35673.41
  14681.4 ]
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 1/3, random_state = 0)
```

```
print(X_test)
print(y_test)
```

```
[[182645.56]
 [ 91790.61]
 [110594.11]
 [ 84710.77]
 [101145.55]
 [127864.55]
 [ 65947.93]
 [152701.92]
 [122782.75]
 [ 91391.77]
 [103057.49]
 [ 85047.44]
 [144135.98]
 [157693.92]
 [114175.79]
 [145530.06]
 [156547.42]]
[103282.38 144259.4  146121.95  77798.83 191050.39 105008.31  81229.06
  97483.56 110352.25 166187.94  96778.92  96479.51 105733.54  96712.8
 124266.9  155752.6  132602.65]
```

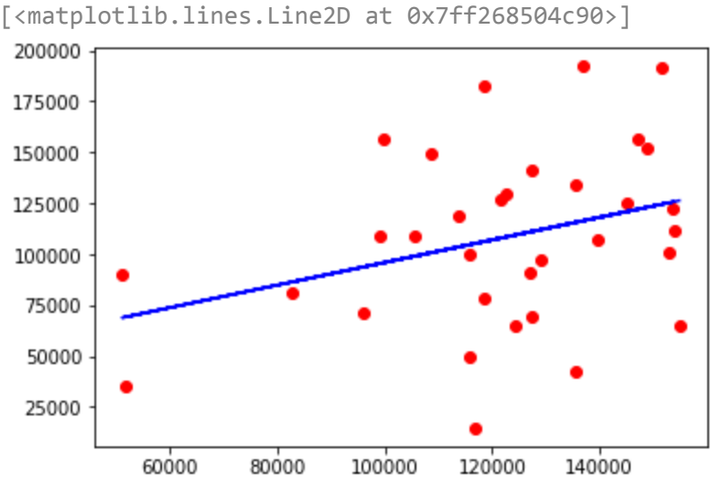
```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)

y_pred = regressor.predict(X_test)
```

```
print(y_pred)
```

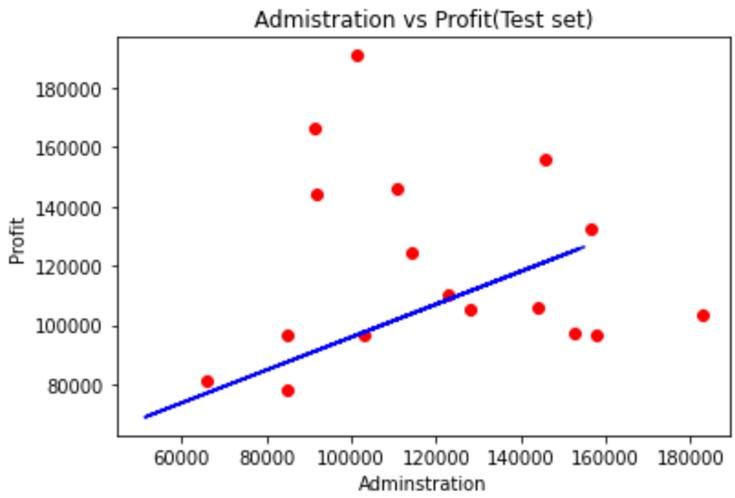
```
[141777.39163335  91338.13890839 101777.1352274   87407.6778932
  96531.64991562 111365.03426681  76991.25447765 125153.80837822
 108543.8099472   91116.717934    97593.08712796  87594.58442074
 120398.32051007 127925.1791137  103765.54931711 121172.26131834
 127288.68541413]
```

```
#visualizing training set results
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
```

```
[<matplotlib.lines.Line2D at 0x7ff268504c90>]
```



```
# Visualising the Test set results
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Admistration vs Profit(Test set)')
plt.xlabel('Adminstration')
plt.ylabel('Profit')
plt.show()
```




```
import math
from sklearn.metrics import mean_squared_error
rmse = math.sqrt(mean_squared_error(y_test, y_pred))
print(rmse)
```

```
38079.86621904332
```

```
from statistics import mean
y_test_mean = mean(y_test)
rmse = rmse / y_test_mean
print(rmse)
```

```
2.667658428667361e-06
```

```
a = [[90000]]
sal = regressor.predict(a)
print(sal)
```

```
[90344.05955089]
```