Name: Vatsal Arya

Roll No.:12

Lab5

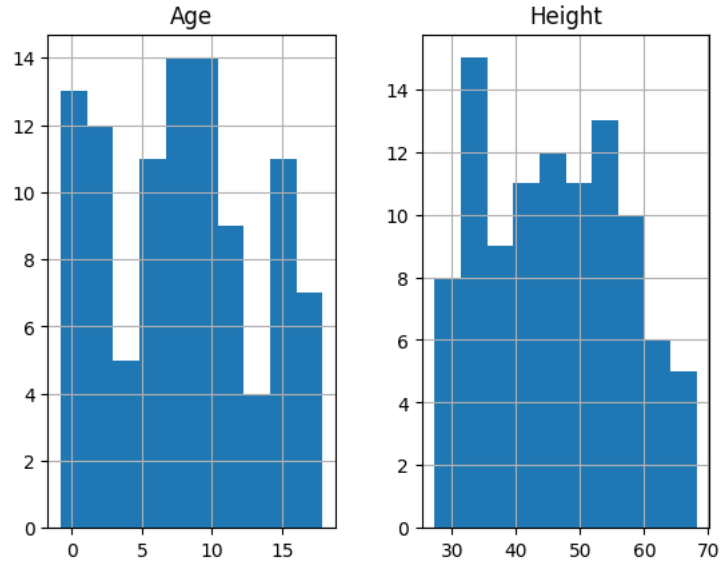
```
import pandas as pd
raw_data = pd.read_pickle('AgesAndHeights.pkl')
print(raw_data)
raw_data.shape
```

```
             Age     Height
0   14.767874  59.627484
1    3.107671  36.146453
2    7.266917  46.912878
3    1.815180  29.125660
4   16.753758  68.170414
..        ...        ...
95   7.323712  46.857505
96   5.591509  39.339990
97   2.625606  32.918925
98   5.519293  40.704154
99  13.117413  55.177407

[100 rows x 2 columns]
(100, 2)
```

```
raw_data.hist()
```

```
array([[<Axes: title={'center': 'Age'}>,
        <Axes: title={'center': 'Height'}>]], dtype=object)
```




```
cleaned_data = raw_data[raw_data['Age']>0]
cleaned_data.shape
```

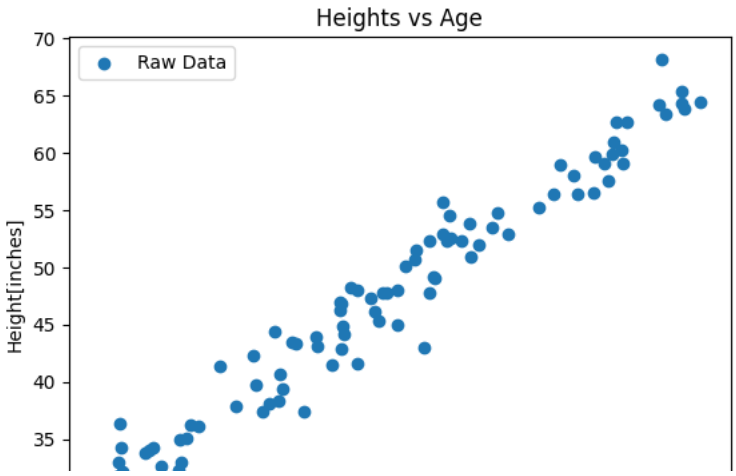
```
(93, 2)
```

```
cleaned_data = raw_data[raw_data['Age']>0]
cleaned_data.shape
```

```
(93, 2)
```

```
import matplotlib.pyplot as plt
ages = cleaned_data['Age']
heights = cleaned_data['Height']
plt.scatter(ages,heights,label='Raw Data')
plt.title("Heights vs Age")
plt.xlabel('Age[Yrs]')
plt.ylabel('Height[inches]')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x78e950cba710>
```




```
from sklearn.model_selection import train_test_split
X_train , y_test = train_test_split(cleaned_data , test_size=0.3, random_state=42)
```

```
def y_hat(ages,params):
  theta1 = params['theta1']
  theta2 = params['theta2']
  return theta1+theta2*ages
  y_hat(5,parameters)


def learn_parameters(data,params):
  x,y = data['Age'] , data['Height']
  x_bar , y_bar = x.mean() , y.mean()
  x,y = x.to_numpy(),y.to_numpy()
  theta2 = sum((x-x_bar)*(y-y_bar))/sum((x-x_bar)**2)
  theta1 = y_bar-theta2*x_bar
  params['theta1'] = theta1
  params['theta2'] = theta2
  new_parameter = {'theta1':0,'theta2':0}



new_parameter = {'theta1':-2,'theta2':1000}
learn_parameters(X_train,new_parameter)
print(new_parameter)
```

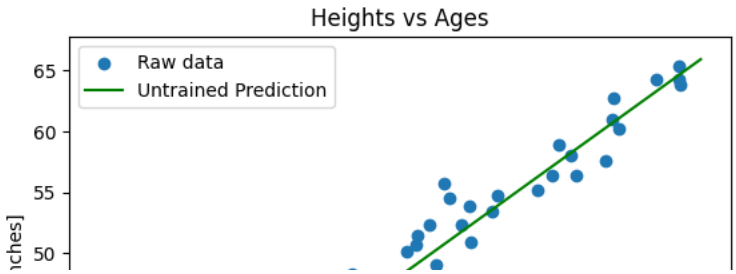
```
{'theta1': 30.317511000776328, 'theta2': 1.9770862462683445}
```

```
spaced_ages = list(range(19))
spaced_untrained_prediction = [y_hat(x,new_parameter) for x in spaced_ages]
print(spaced_untrained_prediction)
```

```
[30.317511000776328, 32.29459724704467, 34.27168349331302, 36.24876973958136, 38.225855985849705, 40.20294223211805, 42.180028478386
```

```
ages = X_train['Age']
heights = X_train['Height']
plt.scatter(ages,heights,label='Raw data')
plt.plot(spaced_ages,spaced_untrained_prediction,label= 'Untrained Prediction', color='green')
plt.title('Heights vs Ages')
plt.xlabel('Age[yrs]')
plt.ylabel('Heights[inches]')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x78e95a1a62c0>
```




```
new_age = int(input('Enter age to predict height:'))
y_hat(new_age,new_parameter)
```

```
Enter age to predict height:20
69.85923592614321
```

```
y_test_Age = y_test['Age']
y_test_Height = y_test['Height']
y_pred_Height = y_hat(y_test_Age,new_parameter)
```

```
import numpy as np
from sklearn.metrics import mean_absolute_error
print('MAE' ,mean_absolute_error(y_test_Height,y_pred_Height))
print('RMSE',np.sqrt(mean_absolute_error(y_test_Height,y_pred_Height)))
```

```
MAE 1.6188718508117308
RMSE 1.2723489500965255
```