1. Read a file csv into read\_file

Table value:

Country Age YearsExperience Salary

0 France 44.0 1.1 39343.0

1 Spain 27.0 1.3 46205.0

2 Germany 30.0 1.5 37731.0

3 Spain 38.0 2.0 43525.0

4 Germany 40.0 2.2 39891.0

5 France 35.0 2.9 56642.0

6 Spain NaN 3.0 60150.0

7 France 48.0 3.2 54445.0

8 Germany 50.0 NaN 64445.0

9 France 37.0 3.7 57189.0

10 France 44.0 3.9 63218.0

11 Spain 27.0 4.0 NaN

12 Germany 30.0 4.0 56957.0

13 Spain 38.0 4.1 57081.0

14 Germany 40.0 4.5 61111.0

15 France 35.0 4.9 67938.0

16 Spain NaN 5.1 66029.0

17 France 48.0 5.3 83088.0

18 Germany 50.0 5.9 NaN

19 France 37.0 NaN 93940.0

20 France 44.0 6.8 91738.0

21 Spain 27.0 7.1 98273.0

22 Germany 30.0 7.9 101302.0

23 Spain 38.0 8.2 113812.0

24 Germany 40.0 8.7 109431.0

25 France 35.0 9.0 105582.0

26 Spain NaN 9.5 116969.0

27 France 48.0 9.6 112635.0

28 Germany 50.0 10.3 122391.0

29 France 37.0 10.5 121872.0

X array:

[['France' 44.0 1.1]

['Spain' 27.0 1.3]

['Germany' 30.0 1.5]

['Spain' 38.0 2.0]

['Germany' 40.0 2.2]

['France' 35.0 2.9]

['Spain' nan 3.0]

['France' 48.0 3.2]

['Germany' 50.0 nan]

['France' 37.0 3.7]

['France' 44.0 3.9]

['Spain' 27.0 4.0]

['Germany' 30.0 4.0]

['Spain' 38.0 4.1]

['Germany' 40.0 4.5]

['France' 35.0 4.9]

['Spain' nan 5.1]

['France' 48.0 5.3]

['Germany' 50.0 5.9]

['France' 37.0 nan]

['France' 44.0 6.8]

['Spain' 27.0 7.1]

['Germany' 30.0 7.9]

['Spain' 38.0 8.2]

['Germany' 40.0 8.7]

['France' 35.0 9.0]

['Spain' nan 9.5]

['France' 48.0 9.6]

['Germany' 50.0 10.3]

['France' 37.0 10.5]]

y1 array:

[[ 39343.]

[ 46205.]

[ 37731.]

[ 43525.]

[ 39891.]

[ 56642.]

[ 60150.]

[ 54445.]

[ 64445.]

[ 57189.]

[ 63218.]

[ nan]

[ 56957.]

[ 57081.]

[ 61111.]

[ 67938.]

[ 66029.]

[ 83088.]

[ nan]

[ 93940.]

[ 91738.]

[ 98273.]

[101302.]

[113812.]

[109431.]

[105582.]

[116969.]

[112635.]

[122391.]

[121872.]]

2. Eliminate missing data

[['France' 44.0 1.1]

['Spain' 27.0 1.3]

['Germany' 30.0 1.5]

['Spain' 38.0 2.0]

['Germany' 40.0 2.2]

['France' 35.0 2.9]

['Spain' 38.77777777777778 3.0]

['France' 48.0 3.2]

['Germany' 50.0 5.364285714285715]

['France' 37.0 3.7]

['France' 44.0 3.9]

['Spain' 27.0 4.0]

['Germany' 30.0 4.0]

['Spain' 38.0 4.1]

['Germany' 40.0 4.5]

['France' 35.0 4.9]

['Spain' 38.77777777777778 5.1]

['France' 48.0 5.3]

['Germany' 50.0 5.9]

['France' 37.0 5.364285714285715]

['France' 44.0 6.8]

['Spain' 27.0 7.1]

['Germany' 30.0 7.9]

['Spain' 38.0 8.2]

['Germany' 40.0 8.7]

['France' 35.0 9.0]

['Spain' 38.77777777777778 9.5]

['France' 48.0 9.6]

['Germany' 50.0 10.3]

['France' 37.0 10.5]]

[[ 39343. ]

[ 46205. ]

[ 37731. ]

[ 43525. ]

[ 39891. ]

[ 56642. ]

[ 60150. ]

[ 54445. ]

[ 64445. ]

[ 57189. ]

[ 63218. ]

[ 76533.32142857]

[ 56957. ]

[ 57081. ]

[ 61111. ]

[ 67938. ]

[ 66029. ]

[ 83088. ]

[ 76533.32142857]

[ 93940. ]

[ 91738. ]

[ 98273. ]

[101302. ]

[113812. ]

[109431. ]

[105582. ]

[116969. ]

[112635. ]

[122391. ]

[121872. ]]

3. Convert categories data

[[1.0 0.0 0.0 44.0 1.1]

[0.0 0.0 1.0 27.0 1.3]

[0.0 1.0 0.0 30.0 1.5]

[0.0 0.0 1.0 38.0 2.0]

[0.0 1.0 0.0 40.0 2.2]

[1.0 0.0 0.0 35.0 2.9]

[0.0 0.0 1.0 38.77777777777778 3.0]

[1.0 0.0 0.0 48.0 3.2]

[0.0 1.0 0.0 50.0 5.364285714285715]

[1.0 0.0 0.0 37.0 3.7]

[1.0 0.0 0.0 44.0 3.9]

[0.0 0.0 1.0 27.0 4.0]

[0.0 1.0 0.0 30.0 4.0]

[0.0 0.0 1.0 38.0 4.1]

[0.0 1.0 0.0 40.0 4.5]

[1.0 0.0 0.0 35.0 4.9]

[0.0 0.0 1.0 38.77777777777778 5.1]

[1.0 0.0 0.0 48.0 5.3]

[0.0 1.0 0.0 50.0 5.9]

[1.0 0.0 0.0 37.0 5.364285714285715]

[1.0 0.0 0.0 44.0 6.8]

[0.0 0.0 1.0 27.0 7.1]

[0.0 1.0 0.0 30.0 7.9]

[0.0 0.0 1.0 38.0 8.2]

[0.0 1.0 0.0 40.0 8.7]

[1.0 0.0 0.0 35.0 9.0]

[0.0 0.0 1.0 38.77777777777778 9.5]

[1.0 0.0 0.0 48.0 9.6]

[0.0 1.0 0.0 50.0 10.3]

[1.0 0.0 0.0 37.0 10.5]]

4. split training set and test sets

X\_test [[1.0 0.0 0.0 48.0 5.3]

[0.0 0.0 1.0 27.0 7.1]

[1.0 0.0 0.0 44.0 3.9]

[1.0 0.0 0.0 37.0 5.364285714285715]

[0.0 1.0 0.0 40.0 4.5]

[1.0 0.0 0.0 44.0 6.8]]

X\_train [[0.0 0.0 1.0 38.77777777777778 9.5]

[0.0 0.0 1.0 38.0 2.0]

[0.0 1.0 0.0 40.0 8.7]

[0.0 1.0 0.0 30.0 7.9]

[0.0 0.0 1.0 38.0 8.2]

[0.0 1.0 0.0 40.0 2.2]

[0.0 1.0 0.0 30.0 1.5]

[1.0 0.0 0.0 35.0 9.0]

[0.0 0.0 1.0 38.77777777777778 3.0]

[0.0 1.0 0.0 50.0 5.9]

[0.0 0.0 1.0 38.0 4.1]

[1.0 0.0 0.0 48.0 3.2]

[1.0 0.0 0.0 48.0 9.6]

[0.0 0.0 1.0 27.0 1.3]

[0.0 0.0 1.0 38.77777777777778 5.1]

[1.0 0.0 0.0 44.0 1.1]

[1.0 0.0 0.0 35.0 4.9]

[1.0 0.0 0.0 37.0 10.5]

[0.0 1.0 0.0 50.0 10.3]

[1.0 0.0 0.0 37.0 3.7]

[0.0 1.0 0.0 50.0 5.364285714285715]

[0.0 1.0 0.0 30.0 4.0]

[0.0 0.0 1.0 27.0 4.0]

[1.0 0.0 0.0 35.0 2.9]]

y\_test [[83088.]

[98273.]

[63218.]

[93940.]

[61111.]

[91738.]]

y\_train [[116969. ]

[ 43525. ]

[109431. ]

[101302. ]

[113812. ]

[ 39891. ]

[ 37731. ]

[105582. ]

[ 60150. ]

[ 76533.32142857]

[ 57081. ]

[ 54445. ]

[112635. ]

[ 46205. ]

[ 66029. ]

[ 39343. ]

[ 67938. ]

[121872. ]

[122391. ]

[ 57189. ]

[ 64445. ]

[ 56957. ]

[ 76533.32142857]

[ 56642. ]]

A

5. Do linear regression based on years of experience

6. Predict the test set result

[[74897.45133835]

[91602.01494268]

[61905.01297943]

[75494.04289565]

[67473.20084754]

[88817.92100862]]

Predict value for new data points (YearsExperience = 3.1 and YearsExperience = 7.0)

Predit YearEx: 3.1 [[54480.76248861]]

Predit YearEx: 7.0 [[90673.98363133]]

Part (B): Visualization of Data – Training set, Test set, Linear Regression Line, Predicted value for new data points Use matplotlib to visualize: Training set data & Linear Regression fit

Training set data and linear regression fit

Testing set data and linear regression fit

Plot predit value for new data points

![Chart, scatter chart

Description automatically generated]()

![Chart, scatter chart

Description automatically generated]()

![Chart, scatter chart

Description automatically generated]()

Process finished with exit code 0