**#MULTIPLE LINEAR REGRESSION:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn import linear\_model

df=pd.read\_csv('D:\Gandhi\Machine Learning\MplsDemo.csv')

df.head()

|  | **neighborhood** | **population** | **white** | **black** | **foreignBorn** | **hhIncome** | **poverty** | **collegeGrad** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | Cedar Riverside | 8247 | 0.353 | 0.464 | 0.408 | 18892 | 0.060 | 0.258 |
| **1** | Phillips West | 5184 | 0.199 | 0.538 | 0.318 | 18404 | 0.042 | 0.211 |
| **2** | Downtown West | 7141 | 0.561 | 0.211 | 0.203 | 67086 | 0.057 | 0.551 |
| **3** | Downtown East | 1674 | 0.543 | 0.221 | 0.221 | 70669 | 0.071 | 0.577 |
| **4** | Shingle Creek | 3249 | 0.407 | 0.259 | 0.140 | 59414 | 0.110 | 0.247 |

df.describe()

|  | **population** | **white** | **black** | **foreignBorn** | **hhIncome** | **poverty** | **collegeGrad** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | 84.000000 | 84.000000 | 84.000000 | 84.000000 | 84.000000 | 84.000000 | 84.000000 |
| **mean** | 4781.023810 | 0.634286 | 0.161143 | 0.140940 | 60798.595238 | 0.064571 | 0.476571 |
| **std** | 2762.714607 | 0.224411 | 0.148176 | 0.086556 | 26432.312248 | 0.021531 | 0.189707 |
| **min** | 240.000000 | 0.087000 | 0.004000 | 0.029000 | 17469.000000 | 0.031000 | 0.122000 |
| **25%** | 2909.250000 | 0.515000 | 0.052750 | 0.080000 | 42027.500000 | 0.049000 | 0.341750 |
| **50%** | 4518.000000 | 0.694000 | 0.122000 | 0.125500 | 57302.000000 | 0.059000 | 0.479500 |
| **75%** | 5985.000000 | 0.818250 | 0.225250 | 0.172500 | 75482.250000 | 0.076750 | 0.622250 |
| **max** | 16022.000000 | 0.902000 | 0.656000 | 0.408000 | 118750.000000 | 0.135000 | 0.837000 |

x=df[['population','white','black','foreignBorn','hhIncome','poverty']]

y=df['collegeGrad']

x.head()

|  | **population** | **white** | **black** | **foreignBorn** | **hhIncome** | **poverty** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | 8247 | 0.353 | 0.464 | 0.408 | 18892 | 0.060 |
| **1** | 5184 | 0.199 | 0.538 | 0.318 | 18404 | 0.042 |
| **2** | 7141 | 0.561 | 0.211 | 0.203 | 67086 | 0.057 |
| **3** | 1674 | 0.543 | 0.221 | 0.221 | 70669 | 0.071 |
| **4** | 3249 | 0.407 | 0.259 | 0.140 | 59414 | 0.110 |

y.head()

Out[23]:

0 0.258

1 0.211

2 0.551

3 0.577

4 0.247

Name: collegeGrad, dtype: float64

from sklearn.model\_selection import train\_test\_split

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y,test\_size=0.2,random\_state=0)

reg=linear\_model.LinearRegression()

reg.fit(x\_train, y\_train)

LinearRegression(copy\_X=True, fit\_intercept=True, n\_jobs=1, normalize=False)

coef\_df=pd.DataFrame(reg.coef\_,x.columns,columns=['coeffs'])

coef\_df

|  | **coeffs** |
| --- | --- |
| **population** | 0.000002 |
| **white** | 0.907465 |
| **black** | 0.317999 |
| **foreignBorn** | 0.430267 |
| **hhIncome** | 0.000002 |
| **poverty** | 0.041487 |

y\_pred=reg.predict(x\_test)

y\_pred

array([0.45144664, 0.70535902, 0.13087826, 0.66928949, 0.64569031,

0.58420521, 0.47726211, 0.54288046, 0.09034178, 0.45397585,

0.42357102, 0.60546549, 0.40075218, 0.4866974 , 0.6312033 ,

0.18674455, 0.64400333])

comparision=pd.DataFrame({'Actual': y\_test,'Predicted': y\_pred})

comparision

|  | **Actual** | **Predicted** |
| --- | --- | --- |
| **30** | 0.443 | 0.451447 |
| **40** | 0.672 | 0.705359 |
| **43** | 0.216 | 0.130878 |
| **50** | 0.654 | 0.669289 |
| **22** | 0.540 | 0.645690 |
| **54** | 0.499 | 0.584205 |
| **2** | 0.551 | 0.477262 |
| **56** | 0.370 | 0.542880 |
| **26** | 0.151 | 0.090342 |
| **8** | 0.345 | 0.453976 |
| **69** | 0.394 | 0.423571 |
| **13** | 0.537 | 0.605465 |
| **66** | 0.356 | 0.400752 |
| **77** | 0.587 | 0.486697 |
| **16** | 0.655 | 0.631203 |
| **27** | 0.181 | 0.186745 |
| **75** | 0.463 | 0.644003 |

from sklearn import metrics

print('MAE',metrics.mean\_absolute\_error(y\_test,y\_pred))

print('MSE',metrics.mean\_squared\_error(y\_test,y\_pred))

print('RMSE',np.sqrt(metrics.mean\_squared\_error(y\_test,y\_pred)))

MAE 0.07076474656611298

MSE 0.0075859371344015845

RMSE 0.08709728545943085