25 2 8 True label 29 Predicted label print(classification report(y test,rfpred)) precision recall f1-score support 0 0.33 0.20 0.25 10 0.78 0.88 0.83 33 0.72 43 accuracy macro avq 0.56 0.54 0.54 43 weighted avg 0.68 0.72 0.69 43 plot roc curve(rfmodel, X test scaled, y test) plt.show() 1.0 0.8 True Positive Rate 0.2 RandomForestClassifier (AUC = 0.70) 0.0 0.0 0.2 0.4 0.6 0.8 1.0 False Positive Rate The model performs better than decision tree, the precision for 0 slightly better, predicts 2/6. AUC increased to 0.70. Running a Lasso Regression Analysis Your assignment is to run a lasso regression analysis using k-fold cross validation to identify a subset of predictors from a larger pool of predictor variables that best predicts a quantitative response variable. In [34]: df.head() Out[34]: lifeexpectancy employrate urbanrate suicideper100th hivrate 0 48.67 55.70 24.04 6.68 0.40 0 76.92 51.40 46.72 7.70 0.40 2 73.13 50.50 65.22 4.85 0.10 1 3 51.09 14.55 75.70 56.70 2.00 X = df.iloc[:,1:5]y = df.iloc[:,1]X.values, y.values , 6.6843853 , 0.4 Out[36]: (array([[55.70000076, 24.04], 0.4 [51.40000153, 46.72 7.69932985,], [50.5 , 65.22 [58.69999886, 88.92 4.84876966, 0.1], 5.3621788, 0.4 , 14.55467701, [75.69999695**,** 56.7 2.], , 2.1618433 , [58.69999886, 30.46 7.765584 , [58.40000153, 92. 0.5 , 3.74158788, , 8.26289272, , 8.47003012, , 13.09437 , 1.38096464, , 3.37441587, [40.09999847, 63.86 3.74158788**,** 0.1 0.4 [58.69999886, 46.78], 0.1 [61.5 , 88.74 [57.09999847, 67.16 8.47003012,], 0.3], [60.90000153, 51.92 0.1 [66.59999847, 83.7 3.37441587, 3.1], [60.40000153, 88.52 4.41499043, 0.4], , 14.53835678, 0.06 [68.09999847, 27.14 [66.90000153, 39.84 , 3.10860252, 1.4 [53.40000153, 73.46 , 26.8746903 , 0.3 [48.59999847, 97.36 , 15.95385 , 0.2 [56.79999924, 51.7 , 9.38879585, 2.3 [71.59999847, 41.2 , 6.05773974, 1.2 1.4],],],],],],], [81.30000305, 19.56 [83.19999695, 10.4 8.20422173, 1.2], , 14.68093586, 3.3], , 4.96107149, 0.5 [78.90000153, 21.56 [59.09999847, 56.76 [63.5 , 80.4 [55.90000153, 59.62 [58.69999886, 100. 7.21422052, 5.3],],],],],],],],],], , 14.77625 , 0.06 [47.09999847, 57.28 , 10.57191 [56. , 75.66 0.1 [59.09999847, 69.9 , 2.206169 , [56. , 73.5 , 12.36798 , [63.09999847, 86.68 , 8.973104 , [58.69999886, 87.3 , 4.99347591, [58.69999886, 73.92 , 5.83525085, 0.4 0.06], 0.2], 2.5 [58.69999886, 73.92 , [52.70000076, 69.02 , [59.70000076, 65.58 , [42.40000153, 42.72 , [58.20000076, 60.7 , 5.83525085, 0.4],], 6.51953697, 0.9 , 8.50468445, 0.4 , 1.79990363, 0.06 7.304886 , 0.8], , 10.12934971, , 8.9133625, [61.70000076, 39.38 5.], [64.90000153, 20.72 8.9133625 , 0.8], 69.46 , 16.95924 , [56.5 , 1.2 , 10.07194233, [80.69999695, 17. 0.4], [58.69999886, 41.42 8.26289272, 0.4], [56.40000153, 52.36 3.94025898, 0.1 [57.20000076, 63.3 , 16.23437 , 0.1 , 14.09153 [51.20000076, 77.36 0.4 , 8.26289272, 58.69999886, 51.64 0.4 7.74506473, 59. 85.04 , 6.44915676, 2. 71.69999695, 56.42 , 32.74 , 1.57435012, 0.1 [53.5 , 73.64 , 9.211085 , 0.1 [65.09999847, 50.02 , 6.28855467, 1.8 [58.69999886, 57.94 , 8.26289272, 0.4 [49.59999847, 61. , 2.816705 , 0.1 [58.69999886, 83.52 , 8.26289272, 0.4 [58.69999886, 30.84 , 3.57647824, 0.4 [43.09999847, 57.94 , 8.26289272, 0.4 [58.69999886, 93.16 , 8.26289272 , 0.4 [58.69999886, 93.16] [55.59999847, 52.74 1.57435012, 0.1],],],], [59.79999924, 73.2 6.1052819 , 0.9 , , 11.95694065, 0.9 [70.40000153, 12.54 [73.09999847, 60.3 , 4.409532 , 0.3 array([55.70000076, 51.40000153, 50.5 , 58.69999886, 75.69999695, 58.69999886, 58.40000153, 40.09999847, 58.69999886, 61.5 57.09999847, 60.90000153, 66.59999847, 60.40000153, 68.09999847, 66.90000153, 53.40000153, 48.59999847, 56.79999924, 71.59999847, 58.69999886, 58.40000153, 70.40000153, 41.20000076, 46. 64.5 , 63.79999924, 47.29999924, 81.30000305, 83.19999695, 78.90000153, 59.09999847, 63.5 , 55.90000153, 58.69999886, 71.30000305, 68.90000153, 51. , 72.80000305, 63.20000076, , 66.19999695, 64.19999695, 58.69999886, 58.59999847, 59.90000153, 47.09999847, 56. , 59.09999847, 56. 63.09999847, 58.69999886, 58.69999886, 52.70000076, 59.70000076, 42.40000153, 58.20000076, 61.70000076, 64.90000153, 56.5 , 80.69999695, 58.69999886, 56.40000153, 57.20000076, 51.20000076, 58.69999886, 59. , 71.69999695, 55.59999847, 53.5 , 65.09999847, 58.69999886, 49.59999847, 58.69999886, 58.69999886, 43.09999847, 58.69999886, 62.70000076, 81.5 , 65.90000153, 58.90000153, 55.90000153, 56.299999924, 59. , 47.299999924, 73.59999847, 55.40000153, 61.5 , 47.79999924, 37.40000153, 59.90000153, 51.29999924, 46.40000153, 58.20000076, 57.29999924, 38.90000153, 63.70000076, 73.19999695, 58.69999886, 64.59999847, 58.90000153, 65.69999695, 58.90000153, 78.19999695, 56.79999924, 46. , 56.29999924, 66. , 48.59999847, 58.69999886, 53.09999847, 53.5 , 63.59999847, 34.90000153, 83. , 71.80000305, 60.5 , 56.90000153, 45.70000076, 46.79999924, 58.69999886, 42.5 , 46.90000153, 54.5 , 57.90000153, 58.69999886, 44.29999924, 58.69999886, 52.09999847, 58.69999886, 46.20000076, 77. , 74.69999695, 42. , 58.69999886, 61.79999924, 61.29999924, 53.5 , 58.69999886, 65. 58.79999924, 60.40000153, 50.90000153, 58.69999886, 65. 50.70000076, 51. , 58.69999886, 59.79999924, 70.40000153, 73.09999847, 68.30000305, 61. , 48.70000076, 57.59999847, 42.40000153, 76. , 44.20000076, 49.5 , 58.79999924, 79.80000305, 58.69999886, 58.69999886, 58.69999886, 58.69999886, 58.69999886, 58.69999886, 51.20000076, 65.59999847, 58.69999886, 48.70000076, 58.69999886, 63.79999924, 62.40000153, 53.40000153, 55.90000153, 65. , 66. , 41.09999847, 52.5 , 55.09999847, 47.29999924, 44.70000076, 50.90000153, 60.70000076, 64.30000305, 44.79999924, 54.5 , 54.59999847, 78.19999695, 72. , 67.30000305, 63.90000153, 58.69999886, 61.5 , 41.59999847, 42.79999924, 58.5 , 58.69999886, 83.19999695, 54.40000153, 75.19999695, 59.299999924, 62.299999924, 57.5 , 58.69999886, 59.90000153, 71. , 61. , 66.80000305])) 39. X_train, X_test, y_train, y_test = train_test_split(X.values, y.values, test_size=0.2, random_state=0) X train.shape, X test.shape, y train.shape, y test.shape Out[38]: ((170, 4), (43, 4), (170,), (43,)) X train Out[39]: array([[53.40000153, 73.46 , 26.8746903 , 0.3 [61.79999924, 17.24 , 11.65520954, 0.4 [46.90000153, 41. [70.40000153, 65.58 , 2.03417849, 0.2 [58.90000153, 28.38 , 35.75287247, 1.2 [59.90000153, 48.78 , 20.31793022, 3.4 59.69999886, 93.16 , 8.26289272, 0.4 14.55467701, 2. 77.54 , 4.66702461. 93.32 [75.69999695, 56.7 [48.59999847, 77.54 0.4 [59.90000153, 93.32 4.11962 , 0.4 6.597168 , [58.59999847, 63.26 0.3 [58.69999886, 64.78 8.26289272, 0.4 5.83525085, 0.4 [58.69999886, 73.92 7.74506473, 5.2 1.37000155, 0.4 8.26289272, 0.4 8.18837547, 0.6 2.51572084, 0.06 8.26289272, 0.4 4.75108385, 0.1 8.26289272, 0.4 4.41499043, 0.4 [78.90000153, 21.56 4.96107149, 0.5 [80.69999695, 17. , 10.07194233, 0.4 [58.20000076, 60.7 7.304886 , 0.8 [58.69999886, 14.32 8.26289272, 0.4 , [46.40000153, 68.08 4.93004451, 0.3 6.05773974, 1.2 2.9 13.11794853, 20.36959 , 0.7 8.50468445, 0.4 [58.69999886, 54.34 9.50792789, 0.4 [58.69999886, 87.3 4.99347591, 2.5 [53.09999847, 66.96 , 33.34186 , 0.1 , 16.91324806, 0.1 72.80000305, 43.1 In [40]: minmax = MinMaxScaler() In [41]: X train scaled = minmax.fit transform(X train) In [42]: X test scaled = minmax.transform(X test) In [43]: X train scaled Out[43]: array([[0.4179688 , 0.70379464, 0.75027211, 0.00970089], [0.58203127, 0.07633929, 0.32217446, 0.01374293], [0.29101567, 0.34151786, 0.18793913, 0.02586904],[0.75000007, 0.61584821, 0.05155152, 0.00565885],, 0.04607922], [0.52539069, 0.20066964, 1. $\hbox{\tt [0.54492194, 0.42834821, 0.5658418, 0.13500404],}$ $\hbox{\tt [0.52148438, 0.92366071, 0.22675446, 0.01374293],}$ [0.85351562, 0.51674107, 0.40373146, 0.07841552], [0.32421874, 0.74933036, 0.12560892, 0.01374293], [0.54492194, 0.92544643, 0.11021137, 0.01374293],[0.51953125, 0.58995536, 0.17990051, 0.00970089],[0.52148438, 0.60691964, 0.22675446, 0.01374293],[0.52148438, 0.70892857, 0.1584691, 0.01374293],[0.52734378, 0.83303571, 0.21218886, 0.20776071],[0.62109377, 0.71897321, 0.03286937, 0.01374293],[0.52148438, 0.34620536, 0.22675446, 0.01374293],, 0.54754464, 0.22465842, 0.021827], [0.85937505, 0.95133929, 0.06509647, 0. [0.52148438, 0.40602679, 0.22675446, 0.01374293],[0.70312504, 0.19732143, 0.12797336, 0.00161681], [0.52148438, 0.55558036, 0.22675446, 0.01374293], $\hbox{\tt [0.55468756, 0.871875 , 0.11851963, 0.01374293],} \\$ [0.4199219, 0.80401786, 0.34329059, 0.00970089],[0.3417969 , 0.48928571, 0.27728485, 0.00161681], [0.27343752, 0.54888393, 0.30976316, 1.], [0.91601571, 0.12455357, 0.13387995, 0.01778496],[0.95117187, 0.07366071, 0.2776399, 0.01374293],[0.5117188 , 0.56138393, 0.19980739, 0.02991108], [0.52148438, 0.04375 , 0.22675446, 0.01374293], [0.28125005, 0.64375 , 0.13300721, 0.00970089], [0.77343752, 0.34375 , 0.16472733, 0.04607922], [0.93359387, 0.08861607, 0.36331878, 0.11479386],[0.48437501, 0.64419643, 0.5672949, 0.02586904],[0.54101567, 0.61584821, 0.23355565, 0.01374293], [0.52148438, 0.49040179, 0.26177515, 0.01374293],[0.52148438, 0.85825893, 0.13479143, 0.09862571], [0.41210937, 0.63125 , 0.93218239, 0.00161681], , 0.35633921, 0.00161681], [0.37109377, 0.2875 , 0.4890625 , 0.00905952, 0.01374293], [0.25 [0.66406254, 0.55513393, 0.19643106, 0.05820534],[0.63085947, 0.70401786, 0.36674653, 0.01374293],[0.52148438, 0.46473214, 0.38015049, 0.00161681],[0.6582031 , 0.98169643, 0.03647136, 0.01374293], [0.50585944, 0.74553571, 0.11496376, 0.00970089],[0.70507814, 0.18683036, 0.4032724 , 0. [0.29882813, 0.67745536, 0.25357902, 0.00161681],[0.55664066, 0.66919643, 0.2252967 , 0.01778496], , 0.68638393, 0.22675446, 0.01374293], [0.52148438, 1. , 0.22675446, 0.01374293], [0.29492186, 0.52321429, 0.40996393, 0.], [0.36523441, 0.68325893, 0.09887904, 0.00161681],[0.7617188, 0.19464286, 0.32212137, 0.01374293],[0.45117187, 0.05245536, 0.78485175, 0. [0.43945315, 0.35803571, 0.22165332, 0.03799515],[0.47656256, 0.46830357, 0.10516626, 0.00161681],[0.36132815, 0.61183036, 0.13072109, 0.00161681],[0.47851565, 0.65915179, 0.47136766, 0.04607922],[0.27734378, 0.50915179, 0.05366775, 0.00161681],[0.52148438, 0.53058036, 0.21752611, 0.01374293],[0.52148438, 0.53058036, 0.11734878, 0.01374293], [0.49023437, 0.63348214, 0.36265555, 0.00970089],[0.26757816, 0.24308036, 0.22675446, 0.03799515],[0.52148438, 0.81607143, 0.22675446, 0.01374293],[0.4980469, 0.9140625, 0.40324183, 0.01778496],[0.52148438, 0.16026786, 0.13450862, 0.01374293],[0.90234374, 0.16875, 0.34343864, 0.22392886],[0.20312504, 0.98125 , 0.22675446, 0.01374293], [0.52148438, 0.37321429, 0.22675446, 0.01374293],[0.52148438, 0.87633929, 0.1451624, 0.01374293],[0.5117188, 0.47879464, 0.11859042, 0.06628941],[0.52148438, 0.53058036, 0.10908358, 0.01374293],[0.68945323, 0.18861607, 0.26529644, 0.01374293], [0.5917969 , 0.79575893, 0.27356385, 0.021827], [0.60742188, 0.85133929, 0.24673148, 0.00565885], [0.05664066, 0.63058036, 0.21589654, 0.01374293], [0.46289067, 0.15223214, 0.18235378, 0.01374293],[0.49414064, 0.62589286, 0.52727793, 0. [0.52148438, 0.13549107, 0.12828801, 0.01374293],, 0.01374293], [0.13476566, 0.75915179, 0. [0.37500004, 0.74732143, 0.39070394, 0.01374293],[0.75000007, 0.02388393, 0.33066164, 0.03395311],[0.66796873, 0.26294643, 0.40818539, 0.01374293],[0.4199219, 0.70580357, 0.25342547, 0.00161681],[0.46875003, 0.70424107, 0.34222346, 0. [0.52929688, 0.6640625, 0.05638931, 0.01374293],[0.81250002, 0.91361607, 0.31573228, 0.00970089],[0.13671876, 0.22589286, 0.17057939, 0.01374293],[0.17968753, 0.41339286, 0.32727992, 0.01374293],[0.96679693, 0.26830357, 0.20371554, 0.05012126],[0.52539069, 0.79308036, 0.62453508, 0. [0.8789063, 0.29508929, 0.29109738, 0.46240905],[0.6796876 , 0.30066964, 0.38546468, 0.5755861], [0.56640628, 0.27924107, 0.33240828, 0.5432498], [0.51562506, 0.91071429, 0.21276603, 0.01778496], $\hbox{\tt [0.4609375~,~0.47254464,~0.03861734,~0.00161681],}$ [0.29882813, 0.63727679, 0.56145603, 0. [0.20507814, 0.53058036, 0.22675446, 0.01374293],[0.52148438, 0.19464286, 0.22529335, 0.01374293],[0.29882813, 0.36875 , 0.26744659, 0.04203719], [0.64453129, 0.084375, 0.12048192, 0.01374293],[0.65625001, 0.35691964, 0.18592757, 0.03395311],[0.37695313, 0.90691964, 0.16118613, 0.00565885],[0.2480469 , 0.72008929 , 0.5779229 , 0.03799515], [0.56445319, 0.46339286, 0.03317774, 0.00161681], [0.2890625, 0.9359375, 0.12234876, 0.00161681],[0.52539069, 0.28883929, 0.27210815, 0.00970089],[0.76757823, 0.31450893, 0.37542719, 0.18755053],[0.58007817, 0.3234375, 0.27925467, 0.19967664],[0.52148438, 0.46026786, 0.22675446, 0.01374293],[0.3789063, 0.40535714, 0.21090242, 0.01374293],[0.52148438, 0.67723214, 0.19692418, 0.01374293],[0.32617191, 0.53058036, 0.40133975, 0.01374293],[0.62890623, 0.56852679, 0.2720654, 0.13500404],[0.27343752, 0.85446429, 0.15855529, 0.00161681], [0.83398437, 0.24754464, 0.27885095, 0.021827], [0.56054692, 0.82745536, 0.30700265, 0.00161681],[0.32421874, 0.97053571, 0.44308778, 0.00565885],[0.79687511, 0.36495536, 0.47007398, 0.00161681],[0.19531251, 0.29508929, 0.21997772, 0.52708165],[0.40429691, 0.65424107, 0.17771688, 0.03395311], [0.52148438, 0.140625, 0.15022956, 0.01374293],[0.39257812, 0.52209821, 0.36252183, 0. , 0.22675446, 0.01374293], [0.52148438, 1. [0.96289074, 0.10223214, 0.22510415, 0.04607922],[0.99609381, 0.21339286, 0.15656075, 0.00565885],[0.52148438, 0.40892857, 0.26416884, 0.01374293],[0.70898448, 0.68080357, 0.0409958, 0.01374293],[0.52929688, 0.51741071, 0.19725713, 0.21180275],[0.57617191, 0.03147321, 0.40352021, 0.05820534],[0.52148438, 1. , 0.30799396, 0.01374293],[0.43945315, 0.53058036, 0.22675446, 0.01374293],[0.61523441, 0.78125 , 0.27845696, 0.00565885], [0.37500004, 0.80379464, 0.17439777, 0.01374293],[0.36914068, 0.42366071, 0.20898181, 0.14308812],[0.17773436, 0.56183036, 0.43635727, 0.7170574], [0.52148438, 0.78191964, 0.15056578, 0.01374293], , 0. , 0.4072829 , 0.130962], [0.66406254, 0.29151786, 0.83436084, 0.02586904], [0.63671876, 0.58348214, 0.12753989, 0.01374293],[0.47460939, 0.41830357, 0.19292436, 0.02991108],[0.46679693, 0.42633929, 0.54065799, 0. [0.66210944, 0.21696429, 0.20708716, 0.09862571], [0.77734386, 0.09375 , 0.28044689, 0.44219887], [0.54296877, 0.70089286, 0.16606461, 0.03395311],[0.4980469, 0.29486607, 0.14098651, 0.00161681],[0.34374999, 0.56473214, 0.07356263, 0.00161681],[0.59960943, 0.42611607, 0.05719735, 0.02991108], [0.63476566, 0.8390625 , 0.12870253, 0.01576395], [0.52148438, 0.93549107, 0.16556924, 0.01374293],[0.46679693, 0.40669643, 0.03706741, 0.07437348],[0.43750006, 0.64263393, 0.52749283, 0.04203719], [0.4179688 , 0.51517857, 0.29377983, 0. [0.53320314, 0.88772321, 0.16351554, 0.00565885],[0.60937505, 0.71540179, 0.13237876, 0.01778496],[0.52148438, 0.16049107, 0.12169422, 0.01374293], [0.64257819, 0.11517857, 0.24505105, 0.02991108], [0.52343752, 0.5171875, 0.33531203, 0.00565885],[0.30859375, 0.64799107, 0.16371871, 0.00565885],[0.64648438, 0.4421875, 0.17121975, 0.07033145],[0.57617191, 0.45825893, 0.31488648, 0.00565885], [0.7207032, 0.18169643, 0.22732203, 0.13500404],[0.51562506, 0.26875 , 0.4315201 , 0.00565885], [0.57617191, 0.87433036, 0.23258088, 0.00161681], [0.90234374, 0.22857143, 0.73183425, 0.00565885],[0.18749998, 0.62611607, 0.08284803, 0. [0.77539061, 0.51361607, 0.17573721, 0.07841552],[0.62304694, 0.35267857, 0.17350751, 0.12691997],[0.48632818, 0.30647321, 0.62309827, 0.], [0.46875003, 0.72834821, 0.29170312, 0.00161681], [0.62109377, 0.30535714, 0.33132425, 0.06224737]])In [44]: X_test_scaled Out[44]: array([[0.2109375 , 0.65044643, 0.09456375, 0. [0.37109377, 0.87098214, 0.28045069, 0.01374293],[0.10546879, 0.62723214, 0.47343244, 0.01374293],[0.52148438, 0.55982143, 0.28948773, 0.01374293],[0.80273439, 0.55691964, 0.1183661, 0.00970089],[0.21679686, 0.53058036, 0.22675446, 0.01374293],[0.61914068, 0.53058036, 0.70892101, 0.00161681],[0.4199219, 0.91830357, 0.22675446, 0.01374293],[0.52148438, 0.22388393, 0.05514251, 0.01374293],, 0.22675446, 0.01374293], [0.52734378, 1. [0.20312504, 0.36071429, 0.04496177, 0. [0.64453129, 0.74866071, 0.29876585, 0.00161681],[0.52343752, 0.696875, 0.77838546, 0.03799515],[0.61718751, 1. , 0.22675446, 0.01374293], [0.52148438, 0.228125 , 0.09493374, 0.01374293], $\hbox{\tt [0.84374999, 0.753125 , 0.03351489, 0.01374293],}$ [0.24023437, 0.35 , 0.43140442, 0.01374293], [0.67578126, 0.81808036, 0.08925007, 0.12287793],[0.56640628, 0.60848214, 0.04840977, 0. [0.23828128, 0.53058036, 0.22675446, 0.01374293],[0.36914068, 0.16227679, 0.35640691, 1.04446241],, 0.02879464, 0.34000531, 0.26030719], [1. [0.51757816, 0.4265625, 0.33797016, 0.01374293],[0.44140625, 0.17924107, 0.06884484, 0.00565885], [0.52148438, 0.53058036, 0.11295265, 0.01374293], $\hbox{\tt [0.48437501, 0.4609375, 0.25842417, 0.09054163],}$, 0.25107468, 0.00161681], [0.59375007, 1. [0.6816407 , 0.32857143, 0.0817732 , 0.0541633], $\hbox{\tt [0.45703131, 0.21361607, 0.51706445, 0.00970089],}$ [0.52148438, 0.53058036, 0.22675446, 0.01374293], [0.1582031, 0.59665179, 0.09957799, 0.00161681],[0.49218754, 0.59040179, 0.45097832, 0.00161681], [0.55468756, 0.06852679, 0.25474442, 0.02991108],[0.80468749, 0.125 , 0.30199268, 0.25222312], [0.57226564, 0.79709821, 0.22397292, 0.00565885],[0.52148438, 0.24464286, 0.22675446, 0.01374293], $\hbox{\tt [0.46679693, 0.54933036, 0.15278952, 0.01374293],}$ [0.32617191, 0.56830357, 0.37792048, 0.00161681],[0.54492194, 0.56852679, 0.28588507, 0.00565885],[0.78125005, 0.25580357, 0.27132559, 0.05012126],[0.47460939, 0.16808036, 0.21538295, 0.95149555],[0.40039065, 0.74464286, 0.15996632, 0.01374293],[0.64453129, 0.85 , 0.33692916, 0.00161681]]) In [45]: lasso = LassoCV(verbose=True, cv=5, random state=0) In [46]: lasso.fit(X train scaled, y train) [Parallel(n jobs=1)]: Using backend SequentialBackend with 1 concurrent workers. [Parallel(n jobs=1)]: Done 5 out of 5 | elapsed: 0.0s finished Out[46]: LassoCV(cv=5, random state=0, verbose=True) In [47]: lassopred = lasso.predict(X test scaled) lassopred Out[47]: array([42.81604395, 51.00784471, 37.42144623, 58.70014357, 73.08574318, 43.11574318, 63.69514546, 53.50534471, 58.70014357, 58.99984471, 42.41644623, 64.99384471, 58.80004395, 63.59524318, 58.70014357, 75.18364166, 44.31454395, 66.59224318, 60.99784471, 44.21464546, 50.90794623, 83.17564166, 58.50034471, 54.60424318, 58.70014357, 56.80204395, 62.39644623, 66.89194623, 55.40344623, 58.70014357, 40.11874318, 57.20164546, 60.39844623, 73.18564166, 61.29754395, 58.70014357, 55.90294623, 48.71014546, 59.89894623, 71.98684471, 56.30254395, 52.50634471, 64.99384471]) mse = mean_squared_error(y_test,lassopred) mse Out[48]: 9.484787969598183e-05 In [49]: rmse = np.sqrt(mse) Out[49]: 0.00973898761145027 r2score = r2_score(y_test, lassopred) r2score Out[50]: 0.9999989897302989 fig, ax = plt.subplots(figsize=(10,8)) sns.regplot(x=y_test, y=lassopred, ax=ax) plt.title("Plot to compare actual vs predicted", fontsize=20) plt.ylabel("Predicted") plt.xlabel("Actual") plt.show() Plot to compare actual vs predicted 80 70 50 50 60 80 70 40 Actual The results are very good with minimum RMSE score and very high R2 score. This means the exploratory variables used are strongly related to life expectency. Running a k-means Cluster Analysis Your assignment is to run a k-means cluster analysis to identify subgroups of observations in your data set that have similar patterns of response on a set of clustering variables. df.head() lifeexpectancy employrate urbanrate suicideper100th hivrate 24.04 48.67 55.70 6.68 0.40 76.92 51.40 46.72 7.70 0.40 2 73.13 50.50 65.22 4.85 0.10 73.13 58.70 5.36 51.09 75.70 56.70 14.55 2.00 df.drop(["politycat"], axis=1, inplace=True) In [54]: Out[54]: lifeexpectancy employrate urbanrate suicideper100th hivrate 0 48.67 55.70 24.04 6.68 0.40 76.92 51.40 46.72 7.70 0.40 2 4.85 73.13 50.50 65.22 0.10 73.13 58.70 88.92 5.36 0.40 4 51.09 75.70 56.70 14.55 2.00 75.18 71.00 27.84 208 11.65 0.40 209 72.83 32.00 71.90 8.26 0.40 210 65.49 39.00 30.64 6.27 0.40 211 49.02 61.00 12.02 35.42 13.50 212 51.38 66.80 37.34 13.91 14.30 213 rows × 5 columns X = df.iloc[:, 0:4]X.values

Out[56]: array([[48.673 , 55.70000076, 24.04 , 6.6843853], , 51.40000153, 46.72 , 7.69932985], [76.918 [73.131 , 50.5 , 65.22 , 4.84876966], [73.131 , 58.69999886, 88.92 , 5.3621788], [51.093 , 75.69999695, 56.7 , 14.55467701], [73.131 , 58.69999886, 30.46 , 2.1618433], [75.901 , 58.40000153, 92. , 7.765584], [74.241 , 40.09999847, 63.86 , 3.74158788], , 51.20000076, 77.36 , 14.09153], 58.69999886, 51.64 , 8.26289272], [75.133 , 65.90000153. 29.84 [48.132 7.563692091 [69.927 , 58.90000153, 28.38 , 35.75287247], , 55.90000153, 46.84 , 1.51924801], [62.095 , 7.06018448], [73.126 , 56.29999924, 47.88 [55.442 , 79.80000305, 18.34 , 13.11794853], , 58.69999886, 32.32 , 8.26289272], [73.131 , 8.21094799], [74.641 , 58.69999886, 27.84 scaler = StandardScaler() X_scaled = scaler.fit_transform(X) In [59]: X scaled Out[59]: array([[-2.3224868 , -0.3072641 , -1.41175552, -0.4717681], [0.73866884, -0.75567586, -0.43511132, -0.30162019],[0.32823881, -0.84952966, 0.36153408, -0.77949544], [0.32823881, 0.00558118, 1.38210143, -0.69342621], [-2.06021041, 1.778372 , -0.00535342, 0.84762787], [0.32823881, 0.00558118, -1.13529804, -1.22993869], [0.62844773, -0.02570309, 1.51473213, -0.29051317], [0.44853913, -1.9340608, 0.30296988, -0.96510624],[0.5574597 , 0.00558118, -0.43252761, -0.20714306], [1.27937005, 0.29757041, 1.37435029, -0.17241801], [1.1652473 , -0.16126978, 0.44507419, 0.60281824], [0.06899702, 0.23500147, -0.21118937, -1.3608472], [0.59799333, 0.82940755, 1.15731824, -1.0266599], [0.53697613, 0.18286056, 1.36487667, -0.85221531], [-0.1255427, 0.98583029, -1.27826359, 0.8448919],[0.72967341, 0.86069242, -0.73137729, -1.07122154],[0.02672934, -0.54711221, 0.71636425, 2.91298663],[1.07366733, -1.04766529, 1.74554398, 1.08218883], $[\ 0.64698049,\ -0.19255425,\ -0.22066299,\ -0.01839375],$ [-1.51961758, 1.35081667, -0.67281309, -0.57682058], [0.32823881, 0.00558118, 1.85922743, -0.20714306], [-0.31618079, -0.02570309, -0.96218915, 1.01324625], [-0.3776315 , 1.2256788 , 0.37703637 , -1.25134075], [0.60341226, -1.81935055, -0.40410674, 0.39198183],[-1.83369898, -1.31879787, 0.11866489, 0.28758317], [0.36692999, 0.61041588, 1.23827464, -0.79152592], [0.85647646, 0.53741852, 0.77492845, -1.36268508], [0.35424969, -1.18323158, 0.61473813, -0.0472705], [-1.58919669, 2.36235085, -1.6046729, -0.21697881], [-2.13412466, 2.56048568, -1.99912002, 0.86879423], [-0.7561982, 2.11207431, -1.51854907, -0.76066887],[-2.00417863, 0.04729387, -0.00276971, -0.38294516],[1.18237114, 0.50613406, 1.01521393, 0.10100031], [0.43932695, -0.28640765, 0.12038736, -0.6479691], [0.32823881, 0.00558118, 1.85922743, -0.20714306], [-2.35229093, 1.3195326, -0.7856353, 0.67893634], [-2.22711357, 1.06925607, -1.29807207, -0.20376039], [0.97731868, -0.79738875, 1.36143171, 0.11288281], [0.36346188, 1.47595534, -0.59099545, 1.24302479],[0.39023141, 0.47484959, 0.76114864, -0.76961587],[-0.97989179, 0.97540227, -1.23778539, -0.79587174],[-2.35239931, 0.78769466, -0.98458134, 0.87417295],[-1.37894206, 0.57913101, 0.19445385, 0.06290708], [0.32823881, 0.00558118, 0.04804335, -0.85919351], [0.99801901, -0.00484704, 0.27713273, -0.48638944], [-1.59591617, 0.13071965, -0.34640378, 1.81379448], [0.70853957, -1.20408803, 0.01962249, 0.88477293], [0.97981139, -0.27597963, 0.81110046, 0.17994655],[1.02836504, 0.04729387, 0.56306383, -1.22250782],[0.82179529, -0.27597963, 0.71808672, 0.48104435],[0.35695915, -0.62010957, 0.52516935, -0.4994037], [0.59929387, 0.1098632, 0.37703637, -0.16660847],[0.33951019, -1.69421228, -0.60735898, -1.29061519],[0.22690475, -0.04655953, 0.16689423, -0.36774577],[-2.0607523, 0.31842685, -0.75118577, 0.1057546],[-0.92180082, 0.65212877, -1.55472108, -0.09809663], [0.51183228, -0.22383872, 0.54411659, 1.250735], [-1.17356448, 2.29978112, -1.71491139, 0.09613068], [0.32823881, 0.00558118, -0.66333946, -0.20714306], [-0.09292072, -0.23426674, -0.19224213, -0.93180051],[1.07019921, -0.15084136, 0.27885521, 1.12921593], [1.2394867, -0.77653231, 0.88430571, 0.76998471],[0.54521291, 0.00558118, -0.22324671, -0.20714306],[-0.801934 , 0.03686585, 1.21502121, -0.29395307], [-1.25842498, 1.3612447 , -0.01741076, -0.51120242], [0.39391628, -0.31769252, -0.1758786 , -1.32842756], [1.11756069, -0.53668419, 0.72411539, -0.04818566], [-0.63665653, 0.67298482, -0.29300701, -0.53812617],[0.32823881, 0.00558118, 0.04804335, -0.20714306],[1.06347973, -0.94338346, 0.1798128, -1.12015599],[0.32823881, 0.00558118, 1.1495671, -0.20714306],[0.63440856, 0.00558118, -1.11893451, -0.99278565], [1.05524295, -1.62121532, 0.04804335, -0.20714306], [0.65456699, 0.00558118, 1.56468395, -0.20714306], [0.11592498, 0.42270868, -0.35501616, -1.21769195], [-1.73464087, 2.3832069, -0.96391162, -0.34445348], [-2.38111966, 0.75641059, -1.16199642, -0.32435885],[-0.01900646, 0.02643782, -1.22486682, 4.40134795],[-0.86782824, -0.28640765, -0.42994389, -1.33766502],[0.32769691, -0.24469516, -0.3851595, -0.40876816], [1.37170868, 0.03686585, 1.85922743, -0.20714306], [0.46728864, -1.18323158, 0.45971524, 1.78765561], [1.26820704, 1.55938032, 1.52592822, 0.32315871], [-0.50551833, -0.33854856, -1.174915, 1.52308479],[-0.0798069, 0.29757041, -0.23099785, 0.31811781],[0.31122335, -1.13109067, 0.50105468, -0.58283185], [-0.11492159, -2.2156214, 0.42095952, 1.263041], [1.13305884, 0.13071965, 0.19445385, 0.14527136], [1.24804861, -0.76610428, 1.50009107, -0.5979259], [1.27373436, -1.27708498, 0.48469115, -0.76587031], [0.32780529, -0.04655953, -0.15176393, -0.85179339],[1.44052914, -0.14041334, 0.41579209, 1.583956471, [0.35771781, -2.05919867, 0.92995134, -1.55858429],[-0.33438841, 0.5269905, 0.04804335, 2.6665369],[-1.40549484, 1.51766743, -1.51682659, 0.24127161], $\hbox{\tt [0.32823881, 0.00558118, -0.55912963, -0.20714306],}$ [-0.13616381, 0.6208439, 0.25215682, -0.79845518], [1.14227103, 0.02643782, 1.06085956, 2.16360247], [0.48484599, 0.73555375, 1.7886059, -1.34121747], [-0.25884847, 0.02643782, -0.8846777, 0.06316187],[-0.28377556, 2.03907656, -1.11721203, 2.80309825],[0.35078157, -0.19255425, 0.48641363, 1.82245486], [0.27502488, -1.31879787, 1.29770008, -0.61360551], [-2.37418343, -0.24469516, -1.35060761, -0.27491651], [-1.44321062, 0.76683862, 0.14277956, -0.38786848], [0.50782227, -1.04766529, 0.89205685, -0.80996365], [0.32823881, 0.00558118, -1.83031732, -0.20714306], [0.230698 , -0.57839708, 0.43646181, 3.99715961], [1.06868191, -0.53668419, 1.10306023, 0.48740439], [1.1739176, 0.51656208, 1.85922743, -0.20714306],[0.51421661, -2.47632596, 0.43387809, -0.27185551], [-0.36679363, 2.53962963, -1.17577624, -0.6254928], [-1.72239408, 1.37167352, -1.63739995, 0.11286015], [0.44637156, 0.19328858, 0.58287232, -0.21583122], [0.73108234, -0.18212583, -0.81663988, 2.15503918], [-2.02216949, -1.35008234, -1.06123155, -0.20714306],[1.03302532, -1.23537249, 1.61205205, -0.82939399],[0.32823881, 0.00558118, 0.61387689, -0.38492955],[1.12677288, -1.68378426, 0.04804335, -0.20714306],[-1.24856252, -1.22494407, -0.68142547, -0.43847984], [0.35446645, -0.43240236, -0.61769384, -0.23754549], [0.74257048, -0.077844 , 0.8774158 , -0.87340808], [-0.12185783, 0.00558118, -1.47634839, -0.79399647], [-0.08511746, -1.49607705, -0.64869841, 1.01255681],[0.32823881, 0.00558118, 1.85922743, 0.27703884],[-0.17387959, -0.6826789, 0.01531629, 0.60202124],[0.48452085, 0.00558118, 0.14450203, -0.20714306], $[\ 0.22191933,\ -1.29794143,\ -0.03463553,\ -1.23872814],$ [-2.1527658 , 1.91393869, -0.86056303, 0.17633636], [-0.53207111, 1.67409017, -1.04400678, 0.10334846], [-0.82772813, -1.73592517, -0.86056303, -0.247532], [0.32823881, 0.00558118, 0.04804335, -0.90845352],[-0.14169112, 0.32885487, -1.70457654, 0.36155367],[1.15224186, 0.27671396, 1.07636185, -0.22372085],[0.70984012, -0.53668419, 1.54401423, -0.20714306],[0.69564251, 0.00558118, 0.34258684, -0.20714306], [1.14357157, 0.66255679, 1.28047532, 0.44949066], [0.42718854, 0.0160094, -0.00363095, 0.43985267], [-1.671998 , 0.18286056, -1.73471988, -0.0403248], [-1.97502477, -0.80781677, -0.36448978, -0.31306686],[0.32823881, 0.00558118, 0.04804335, -0.88539414],[1.19158332, 0.66255679, 0.88947314, 0.22203992], [0.31122335, -0.82867322, 0.63713033, -0.9692719], [-0.50562671, -0.79738875, -0.88984513, 0.56517324],[0.32823881, 0.00558118, 1.01779764, -0.66122246], [0.65304969, 0.12029123, 0.70516815, -0.56885049], [-0.79239668, 1.2256788, -1.90696753, 0.41213665], [0.25735915, 1.50723941, 0.14966946, -0.85313037], [0.42133609, 1.00668713, 0.62765671, -1.3142521], [-0.14667654, 0.24542949, 0.34861551, -1.27006532],[0.65283293, -1.03723687, 0.19359262, 0.69379619], [0.42014392, -0.95381149, -0.11128573, 0.09401464],[-0.13865652, 0.0160094, 0.68966586, 3.0805403], [-1.58887156, 2.20592811, -1.65720843, 0.606771],[0.32823881, 0.00558118, -1.05520288, -0.20714306],[0.4918906 , 0.00558118, -1.24812025, -0.2158512], [0.23633369, 0.00558118, -0.42133151, 0.0158441], [0.25378266, 0.00558118, -1.45653991, -0.66322627], [0.32823881, 0.00558118, 1.61032957, -0.57180286], [-0.58918667, 0.00558118, 0.16086556, 0.16674298], [0.41277417, -0.77653231, 1.10219899, -0.51918537],[-1.16879582, 0.72512573, -0.62200003, -0.45046857],[0.47899354, 0.00558118, -0.20602194, 0.70708689], [0.32823881, -1.03723687, 0.04804335, 0.83337345], [0.32823881, 0.00558118, -0.10697954, 0.00157785], [-2.41775165, 0.53741852, -0.82094607, 0.41608576], [1.1947263, 0.39142421, 1.85922743, -0.06219622], [0.57913544, -0.54711221, -0.01138209, 0.1923236],[1.00127037, -0.28640765, -0.35415492, 1.66370069],[-0.24389221, 0.66255679, -1.67357196, -0.84052019],[-2.0465547, 0.76683862, -0.87434284, 3.41414975],[-1.87553315, -1.82977897, 0.16861671, 1.04207548], [1.22485558, -0.64096601, 0.87397085, -0.60519589], [0.52440421, -0.36983343, -1.79672903, 3.11907896], [-0.93751573, -1.18323158, -0.5763544, 0.03537924],[0.04992237, -1.45436417, 0.77923464, 1.88579706], [-2.31760976, -0.80781677, -1.3729998, 0.56557677],[1.22864883, 0.21414503, 1.19349025, 0.2711307], [1.32608126, 0.58955983, 0.71722549, 0.62720018],[-1.29007155, 2.03907656, -1.34802389, 0.48828675], [0.43607559, 1.39252956, -1.01214096, 0.05849785],[-0.82664434, 0.90240531, -1.27137368, 0.02256451],[-1.4132981, 0.54784694, -0.63836355, -0.52449127],[0.24001857, 0.00558118, -1.37988971, -0.83329497],[0.00234413, 0.29757041, -1.87768543, 0.84636884], [0.47823489, -1.77763806, 0.41665333, -1.06481565], [0.42014392, -1.65249979, 0.5105283, -0.99499074], [-0.55450549, -0.01527507, -0.35329369, 0.45569498], [0.32823881, 0.00558118, 0.04804335, -0.26214342], [-1.73258168, 2.56048568, -1.88802029, 0.46782433],[-0.1743131, -0.44283039, 0.48038496, 1.58523726],[0.69835198, 1.72623108, 0.90669791, -1.35883784], [1.09111629, 0.06815031, 1.42602458, -0.58404275], [0.91348365, 0.38099579, 1.07119442, 0.07183771], [0.74809779, -0.11955689, 1.5276507, 0.84470971], [-0.19674749, -0.11955689, -0.86142427, -0.71831422], [0.09912629, 0.00558118, -1.38075095, -0.75692201],[0.4659881 , 0.13071965, 1.57157385, -0.90173197], [0.55041509, 1.28824774, -1.24812025, 0.36123728],[0.29583358, -2.77874342, 0.64918766, -0.20714306],[-0.4995575 , -2.04877064, -1.12754689, -0.54194266], [-2.28433751, 0.24542949, -0.92171095, 0.42254653], [-2.02867221, 0.85026439, -0.83903207, 0.73875906]]) wcss = []for i in range (1,11): kmeans = KMeans(n clusters=i,init='k-means++',random state=0) kmeans.fit(X scaled) wcss.append(kmeans.inertia) plt.figure(figsize=(10,8)) plt.plot(range(1,11), wcss, linewidth=2) plt.xlabel("Number of Clusters") plt.ylabel("WCSS") plt.show() 800 700 600 400 300 200 Number of Clusters kmeans = KMeans(n clusters=3,init='k-means++',random_state=0) kmeans.fit(X scaled) KMeans(n clusters=3, random state=0) In [64]: df segm kmeans = df.copy() df segm kmeans lifeexpectancy employrate urbanrate suicideper100th hivrate 0 48.67 55.70 24.04 6.68 0.40 76.92 51.40 46.72 7.70 0.40 2 73.13 50.50 65.22 4.85 0.10 3 73.13 58.70 88.92 5.36 0.40 4 51.09 75.70 56.70 14.55 2.00 208 75.18 71.00 27.84 11.65 0.40 209 72.83 32.00 71.90 8.26 0.40 65.49 39.00 210 30.64 6.27 0.40 211 49.02 61.00 35.42 12.02 13.50 212 51.38 66.80 37.34 13.91 14.30 213 rows × 5 columns df segm kmeans['Segment K-Means'] = kmeans.labels df segm kmeans lifeexpectancy employrate urbanrate suicideper100th hivrate Segment K-Means 0 48.67 55.70 24.04 0.40 1 6.68 76.92 51.40 46.72 7.70 0.40 2 2 73.13 50.50 4.85 2 65.22 0.10 3 2 73.13 58.70 88.92 5.36 0.40 4 51.09 75.70 56.70 14.55 2.00 1 208 75.18 71.00 27.84 11.65 0.40 1 209 71.90 8.26 2 72.83 32.00 0.40 210 65.49 39.00 30.64 6.27 0.40 1 49.02 61.00 35.42 12.02 13.50 211 37.34 212 51.38 66.80 13.91 14.30 213 rows × 6 columns df segm analysis = df segm kmeans.groupby(['Segment K-Means']).mean() df_segm_analysis lifeexpectancy employrate urbanrate suicideper100th hivrate **Segment K-Means** 70.66 55.48 20.89 1.08 0 54.39 60.17 64.15 32.37 8.98 3.43 2 75.52 56.35 71.10 6.99 0.45 x axis = df segm kmeans['urbanrate'] y axis = df segm kmeans['lifeexpectancy'] plt.figure(figsize=(10,8)) sns.scatterplot(x=x_axis,y=y_axis,hue=df_segm_kmeans["Segment K-Means"],palette='viridis') plt.title('Segmentation K-Means') plt.show() Segmentation K-Means 85 Segment K-Means 80 75 70 lifeexpectancy 65 60 55 50 20 60 100 40 80 urbanrate From the graph, at least 2 segments (yellow vs green) is clearly separated. The overlapping purple points are problematic. Python code done by Dennis Lam