First 5 rows of raw data:

Age Workclass Final\_Weight Education Highest\_Grade Marital\_Status Occupation Relationship Race Sex Capital\_Gain Capital\_Loss Hours\_per\_Week Native\_Country Income\_Class

0 50 Self-emp-not-inc 83311 Bachelors 13 Married-civ-spouse Exec-managerial Husband White Male 0 0 13 United-States <=50K

1 38 Private 215646 HS-grad 9 Divorced Handlers-cleaners Not-in-family White Male 0 0 40 United-States <=50K

2 53 Private 234721 11th 7 Married-civ-spouse Handlers-cleaners Husband Black Male 0 0 40 United-States <=50K

3 28 Private 338409 Bachelors 13 Married-civ-spouse Prof-specialty Wife Black Female 0 0 40 Cuba <=50K

4 37 Private 284582 Masters 14 Married-civ-spouse Exec-managerial Wife White Female 0 0 40 United-States <=50K

Last 5 rows of raw data:

Age Workclass Final\_Weight Education Highest\_Grade Marital\_Status Occupation Relationship Race Sex Capital\_Gain Capital\_Loss Hours\_per\_Week Native\_Country Income\_Class

32555 27 Private 257302 Assoc-acdm 12 Married-civ-spouse Tech-support Wife White Female 0 0 38 United-States <=50K

32556 40 Private 154374 HS-grad 9 Married-civ-spouse Machine-op-inspct Husband White Male 0 0 40 United-States >50K

32557 58 Private 151910 HS-grad 9 Widowed Adm-clerical Unmarried White Female 0 0 40 United-States <=50K

32558 22 Private 201490 HS-grad 9 Never-married Adm-clerical Own-child White Male 0 0 20 United-States <=50K

32559 52 Self-emp-inc 287927 HS-grad 9 Married-civ-spouse Exec-managerial Wife White Female 15024 0 40 United-States >50K

Shape of raw data:

(32560, 15)

Any missing data:

False

Data Statistical Description:

Age Final\_Weight Highest\_Grade Capital\_Gain Capital\_Loss Hours\_per\_Week

count 32560.000000 3.256000e+04 32560.000000 32560.000000 32560.000000 32560.000000

mean 38.581634 1.897818e+05 10.080590 1077.615172 87.306511 40.437469

std 13.640642 1.055498e+05 2.572709 7385.402999 402.966116 12.347618

min 17.000000 1.228500e+04 1.000000 0.000000 0.000000 1.000000

25% 28.000000 1.178315e+05 9.000000 0.000000 0.000000 40.000000

50% 37.000000 1.783630e+05 10.000000 0.000000 0.000000 40.000000

75% 48.000000 2.370545e+05 12.000000 0.000000 0.000000 45.000000

max 90.000000 1.484705e+06 16.000000 99999.000000 4356.000000 99.000000

Data Types Information:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 32560 entries, 0 to 32559

Data columns (total 15 columns):

Age 32560 non-null int64

Workclass 32560 non-null object

Final\_Weight 32560 non-null int64

Education 32560 non-null object

Highest\_Grade 32560 non-null int64

Marital\_Status 32560 non-null object

Occupation 32560 non-null object

Relationship 32560 non-null object

Race 32560 non-null object

Sex 32560 non-null object

Capital\_Gain 32560 non-null int64

Capital\_Loss 32560 non-null int64

Hours\_per\_Week 32560 non-null int64

Native\_Country 32560 non-null object

Income\_Class 32560 non-null object

dtypes: int64(6), object(9)

memory usage: 3.7+ MB

None

Check Columns:

Index(['Age', 'Workclass', 'Final\_Weight', 'Education', 'Highest\_Grade',

'Marital\_Status', 'Occupation', 'Relationship', 'Race', 'Sex',

'Capital\_Gain', 'Capital\_Loss', 'Hours\_per\_Week', 'Native\_Country',

'Income\_Class'],

dtype='object')

Age : [50 38 53 28 37 49 52 31 42 30 23 32 40 34 25 43 54 35 59 56 19 39 20 45

22 48 21 24 57 44 41 29 18 47 46 36 79 27 67 33 76 17 55 61 70 64 71 68

66 51 58 26 60 90 75 65 77 62 63 80 72 74 69 73 81 78 88 82 83 84 85 86

87]

Workclass : [' Self-emp-not-inc' ' Private' ' State-gov' ' Federal-gov' ' Local-gov'

' ?' ' Self-emp-inc' ' Without-pay' ' Never-worked']

Final\_Weight : [ 83311 215646 234721 ... 34066 84661 257302]

Education : [' Bachelors' ' HS-grad' ' 11th' ' Masters' ' 9th' ' Some-college'

' Assoc-acdm' ' Assoc-voc' ' 7th-8th' ' Doctorate' ' Prof-school'

' 5th-6th' ' 10th' ' 1st-4th' ' Preschool' ' 12th']

Highest\_Grade : [13 9 7 14 5 10 12 11 4 16 15 3 6 2 1 8]

Marital\_Status : [' Married-civ-spouse' ' Divorced' ' Married-spouse-absent'

' Never-married' ' Separated' ' Married-AF-spouse' ' Widowed']

Occupation : [' Exec-managerial' ' Handlers-cleaners' ' Prof-specialty'

' Other-service' ' Adm-clerical' ' Sales' ' Craft-repair'

' Transport-moving' ' Farming-fishing' ' Machine-op-inspct'

' Tech-support' ' ?' ' Protective-serv' ' Armed-Forces'

' Priv-house-serv']

Relationship : [' Husband' ' Not-in-family' ' Wife' ' Own-child' ' Unmarried'

' Other-relative']

Race : [' White' ' Black' ' Asian-Pac-Islander' ' Amer-Indian-Eskimo' ' Other']

Sex : [' Male' ' Female']

Capital\_Gain : [ 0 14084 5178 5013 2407 14344 15024 7688 34095 4064 4386 7298

1409 3674 1055 3464 2050 2176 2174 594 20051 6849 4101 1111

8614 3411 2597 25236 4650 9386 2463 3103 10605 2964 3325 2580

3471 4865 99999 6514 1471 2329 2105 2885 25124 10520 2202 2961

27828 6767 2228 1506 13550 2635 5556 4787 3781 3137 3818 3942

914 401 2829 2977 4934 2062 2354 5455 15020 1424 3273 22040

4416 3908 10566 991 4931 1086 7430 6497 114 7896 2346 3418

3432 2907 1151 2414 2290 15831 41310 4508 2538 3456 6418 1848

3887 5721 9562 1455 2036 1831 11678 2936 2993 7443 6360 1797

1173 4687 6723 2009 6097 2653 1639 18481 7978 2387 5060]

Capital\_Loss : [ 0 2042 1408 1902 1573 1887 1719 1762 1564 2179 1816 1980 1977 1876

1340 2206 1741 1485 2339 2415 1380 1721 2051 2377 1669 2352 1672 653

2392 1504 2001 1590 1651 1628 1848 1740 2002 1579 2258 1602 419 2547

2174 2205 1726 2444 1138 2238 625 213 1539 880 1668 1092 1594 3004

2231 1844 810 2824 2559 2057 1974 974 2149 1825 1735 1258 2129 2603

2282 323 4356 2246 1617 1648 2489 3770 1755 3683 2267 2080 2457 155

3900 2201 1944 2467 2163 2754 2472 1411]

Hours\_per\_Week : [13 40 16 45 50 80 30 35 60 20 52 44 15 25 38 43 55 48 58 32 70 2 22 56

41 28 36 24 46 42 12 65 1 10 34 75 98 33 54 8 6 64 19 18 72 5 9 47

37 21 26 14 4 59 7 99 53 39 62 57 78 90 66 11 49 84 3 17 68 27 85 31

51 77 63 23 87 88 73 89 97 94 29 96 67 82 86 91 81 76 92 61 74 95]

Native\_Country : [' United-States' ' Cuba' ' Jamaica' ' India' ' ?' ' Mexico' ' South'

' Puerto-Rico' ' Honduras' ' England' ' Canada' ' Germany' ' Iran'

' Philippines' ' Italy' ' Poland' ' Columbia' ' Cambodia' ' Thailand'

' Ecuador' ' Laos' ' Taiwan' ' Haiti' ' Portugal' ' Dominican-Republic'

' El-Salvador' ' France' ' Guatemala' ' China' ' Japan' ' Yugoslavia'

' Peru' ' Outlying-US(Guam-USVI-etc)' ' Scotland' ' Trinadad&Tobago'

' Greece' ' Nicaragua' ' Vietnam' ' Hong' ' Ireland' ' Hungary'

' Holand-Netherlands']

Income\_Class : [' <=50K' ' >50K']

First 5 rows of raw data:

Age Workclass Final\_Weight Education Highest\_Grade Marital\_Status Occupation Relationship Race Sex Capital\_Gain Capital\_Loss Hours\_per\_Week Native\_Country Income\_Class

0 38 Private 89814 HS-grad 9 Married-civ-spouse Farming-fishing Husband White Male 0 0 50 United-States <=50K.

1 28 Local-gov 336951 Assoc-acdm 12 Married-civ-spouse Protective-serv Husband White Male 0 0 40 United-States >50K.

2 44 Private 160323 Some-college 10 Married-civ-spouse Machine-op-inspct Husband Black Male 7688 0 40 United-States >50K.

3 18 ? 103497 Some-college 10 Never-married ? Own-child White Female 0 0 30 United-States <=50K.

4 34 Private 198693 10th 6 Never-married Other-service Not-in-family White Male 0 0 30 United-States <=50K.

Last 5 rows of raw data:

Age Workclass Final\_Weight Education Highest\_Grade Marital\_Status Occupation ... Race Sex Capital\_Gain Capital\_Loss Hours\_per\_Week Native\_Country Income\_Class

16275 39 Private 215419 Bachelors 13 Divorced Prof-specialty ... White Female 0 0 36 United-States <=50K.

16276 64 ? 321403 HS-grad 9 Widowed ? ... Black Male 0 0 40 United-States <=50K.

16277 38 Private 374983 Bachelors 13 Married-civ-spouse Prof-specialty ... White Male 0 0 50 United-States <=50K.

16278 44 Private 83891 Bachelors 13 Divorced Adm-clerical ... Asian-Pac-Islander Male 5455 0 40 United-States <=50K.

16279 35 Self-emp-inc 182148 Bachelors 13 Married-civ-spouse Exec-managerial ... White Male 0 0 60 United-States >50K.

[5 rows x 15 columns]

Shape of raw data:

(16280, 15)

Any missing data:

False

Data Statistical Description:

Age Final\_Weight Highest\_Grade Capital\_Gain Capital\_Loss Hours\_per\_Week

count 16280.000000 1.628000e+04 16280.000000 16280.00000 16280.000000 16280.000000

mean 38.768305 1.894334e+05 10.073096 1081.97156 87.904668 40.392260

std 13.849192 1.057177e+05 2.567511 7584.16416 403.117078 12.479715

min 17.000000 1.349200e+04 1.000000 0.00000 0.000000 1.000000

25% 28.000000 1.167162e+05 9.000000 0.00000 0.000000 40.000000

50% 37.000000 1.778295e+05 10.000000 0.00000 0.000000 40.000000

75% 48.000000 2.383840e+05 12.000000 0.00000 0.000000 45.000000

max 90.000000 1.490400e+06 16.000000 99999.00000 3770.000000 99.000000

Data Types Information:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 16280 entries, 0 to 16279

Data columns (total 15 columns):

Age 16280 non-null int64

Workclass 16280 non-null object

Final\_Weight 16280 non-null int64

Education 16280 non-null object

Highest\_Grade 16280 non-null int64

Marital\_Status 16280 non-null object

Occupation 16280 non-null object

Relationship 16280 non-null object

Race 16280 non-null object

Sex 16280 non-null object

Capital\_Gain 16280 non-null int64

Capital\_Loss 16280 non-null int64

Hours\_per\_Week 16280 non-null int64

Native\_Country 16280 non-null object

Income\_Class 16280 non-null object

dtypes: int64(6), object(9)

memory usage: 1.9+ MB

None

Check Columns:

Index(['Age', 'Workclass', 'Final\_Weight', 'Education', 'Highest\_Grade',

'Marital\_Status', 'Occupation', 'Relationship', 'Race', 'Sex',

'Capital\_Gain', 'Capital\_Loss', 'Hours\_per\_Week', 'Native\_Country',

'Income\_Class'],

dtype='object')

Age : [38 28 44 18 34 29 63 24 55 65 36 26 58 48 43 20 37 40 72 25 45 22 23 54

32 46 56 17 39 52 21 42 33 30 47 41 19 69 50 31 59 49 51 27 57 61 64 79

73 53 77 80 62 35 68 66 75 60 67 71 70 90 81 74 78 82 83 85 76 84 89 88

87]

Workclass : [' Private' ' Local-gov' ' ?' ' Self-emp-not-inc' ' Federal-gov'

' State-gov' ' Self-emp-inc' ' Without-pay' ' Never-worked']

Final\_Weight : [ 89814 336951 160323 ... 349230 321403 83891]

Education : [' HS-grad' ' Assoc-acdm' ' Some-college' ' 10th' ' Prof-school'

' 7th-8th' ' Bachelors' ' Masters' ' Doctorate' ' 11th' ' 5th-6th'

' Assoc-voc' ' 9th' ' 12th' ' 1st-4th' ' Preschool']

Highest\_Grade : [ 9 12 10 6 15 4 13 14 16 7 3 11 5 8 2 1]

Marital\_Status : [' Married-civ-spouse' ' Never-married' ' Widowed' ' Divorced'

' Separated' ' Married-spouse-absent' ' Married-AF-spouse']

Occupation : [' Farming-fishing' ' Protective-serv' ' Machine-op-inspct' ' ?'

' Other-service' ' Prof-specialty' ' Craft-repair' ' Adm-clerical'

' Exec-managerial' ' Tech-support' ' Sales' ' Priv-house-serv'

' Transport-moving' ' Handlers-cleaners' ' Armed-Forces']

Relationship : [' Husband' ' Own-child' ' Not-in-family' ' Unmarried' ' Wife'

' Other-relative']

Race : [' White' ' Black' ' Asian-Pac-Islander' ' Other' ' Amer-Indian-Eskimo']

Sex : [' Male' ' Female']

Capital\_Gain : [ 0 7688 3103 6418 7298 3908 14084 5178 15024 99999 2597 2907

4650 6497 1055 5013 27828 4934 4064 3674 2174 10605 3418 114

2580 3411 4508 4386 8614 13550 6849 2463 3137 2885 2964 1471

10566 2354 1424 1455 3325 4416 25236 594 2105 4787 2829 401

4865 1264 1506 10520 3464 2653 20051 4101 1797 2407 3471 1086

1848 14344 1151 2993 2290 15020 9386 2202 3818 2176 5455 11678

7978 7262 6514 41310 3456 7430 2414 2062 34095 1831 6723 5060

15831 2977 2346 3273 2329 9562 2635 4931 1731 6097 914 7896

5556 1409 3781 3942 2538 3887 25124 7443 5721 1173 4687 6612

6767 2961 991 2036 2936]

Capital\_Loss : [ 0 1721 1876 2415 1887 625 1977 2057 1429 1590 1485 2051 2377 1672

1628 1902 1602 1741 2444 1408 2001 2042 1740 1825 1848 1719 3004 2179

1573 2205 1258 2339 1726 2258 1340 1504 2559 1668 1974 1980 1564 2547

2002 1669 1617 323 3175 2472 2174 1579 2129 1510 1735 2282 1870 1411

1911 1651 1092 1762 2457 2231 2238 653 1138 2246 2603 2392 1944 1380

2465 1421 3770 1594 213 2149 2824 1844 2467 2163 1816 1648]

Hours\_per\_Week : [50 40 30 32 10 39 35 48 25 20 45 47 6 43 90 54 60 38 36 18 24 44 56 28

16 41 22 55 14 33 37 8 12 70 15 75 52 84 42 80 68 99 65 5 17 72 53 29

96 21 46 3 1 23 49 67 76 7 2 58 26 34 4 51 78 63 31 92 77 27 85 13

19 98 62 66 57 11 86 59 9 64 73 61 88 79 89 74 69]

Native\_Country : [' United-States' ' ?' ' Peru' ' Guatemala' ' Mexico'

' Dominican-Republic' ' Ireland' ' Germany' ' Philippines' ' Thailand'

' Haiti' ' El-Salvador' ' Puerto-Rico' ' Vietnam' ' South' ' Columbia'

' Japan' ' India' ' Cambodia' ' Poland' ' Laos' ' England' ' Cuba'

' Taiwan' ' Italy' ' Canada' ' Portugal' ' China' ' Nicaragua'

' Honduras' ' Iran' ' Scotland' ' Jamaica' ' Ecuador' ' Yugoslavia'

' Hungary' ' Hong' ' Greece' ' Trinadad&Tobago'

' Outlying-US(Guam-USVI-etc)' ' France']

Income\_Class : [' <=50K.' ' >50K.']

KNeighborsClassifier

Accuracy: 0.8236

Precision: 0.6440

Recall: 0.5668

F1\_Score: 0.6030

AUC: 0.7350

GaussianNaiveBayes

Accuracy: 0.5189

Precision: 0.3238

Recall: 0.9527

F1\_Score: 0.4834

AUC: 0.6687

SVMClassifier

Accuracy: 0.8522

Precision: 0.7365

Recall: 0.5829

F1\_Score: 0.6508

AUC: 0.7592

DecisionTreeClassifier

Accuracy: 0.8196

Precision: 0.6218

Recall: 0.6032

F1\_Score: 0.6124

AUC: 0.7449

RandomForestClassifier

Accuracy: 0.8399

Precision: 0.6923

Recall: 0.5803

F1\_Score: 0.6314

AUC: 0.7503

AdaBoostClassifier

Accuracy: 0.8595

Precision: 0.7479

Recall: 0.6110

F1\_Score: 0.6726

AUC: 0.7737

GradientBoostingClassifier

Accuracy: 0.8711

Precision: 0.7964

Recall: 0.6102

F1\_Score: 0.6910

AUC: 0.7810

LinearDiscriminantAnalysis

S:\Python37\lib\site-packages\sklearn\discriminant\_analysis.py:388: UserWarning: Variables are collinear.

warnings.warn("Variables are collinear.")

Accuracy: 0.8452

Precision: 0.7114

Recall: 0.5801

F1\_Score: 0.6391

AUC: 0.7536

MultilayerPerceptronClassifier

Accuracy: 0.8491

Precision: 0.7027

Recall: 0.6264

F1\_Score: 0.6624

AUC: 0.7722

LogisticRegression

Accuracy: 0.8517

Precision: 0.7290

Recall: 0.5923

F1\_Score: 0.6536

AUC: 0.7621