```
In [5]:
          import pandas as pd
          import numpy as np
          import matplotlib.pyplot as plt
          from sklearn.model_selection import GridSearchCV, train_test_split, cross_val_score
          from sklearn.preprocessing import StandardScaler
          from sklearn.metrics import r2_score
          import scipy.stats as stats
          import seaborn as sns
          %matplotlib inline
In [6]:
          df = pd.read_csv('CKD dataset.csv')
          data = df
          data.head()
Out[6]:
                                                                                                                        Unname
                                                                             wbcc
                                                                                    rbcc
                                                 normal notpresent notpresent ... 44.0 7800.0 5.2
           NaN 48.0 80.0 1.020
                                           NaN
                7.0 50.0 1.020 4.0 0.0
                                           NaN
                                                 normal notpresent notpresent ... 38.0 6000.0 NaN
                                                                                                          good
                                                 normal notpresent notpresent ...
                                                                             31. 0 7500. 0
                          1.010 2.0 3.0
                                         normal
         3 NaN 48.0 70.0 1.005 4.0 0.0 normal
                                               abnormal
                                                         present notpresent ... 32.0 6700.0 3.9
                 51.0 80.0
                          1. 010 2. 0 0. 0 normal
                                                 normal notpresent notpresent ... 35.0 7300.0
        5 rows x 26 columns
In [7]:
          data.shape
         (400, 26)
Out[7]:
```

```
In [7]:
         data.shape
        (400, 26)
Out[7]:
In [8]:
         df.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 400 entries, 0 to 399
        Data columns (total 26 columns):
                        0 non-null float64
        age
        bp
                        391 non-null float64
        sg
                        388 non-null float64
        al
                        353 non-null float64
                        354 non-null float64
        su
                        351 non-null float64
        rbc
                        248 non-null object
        pc
                        335 non-null object
        pcc
                        396 non-null object
        ba
                        396 non-null object
        bgr
                        356 non-null float64
        bu
                        381 non-null float64
        SC
                        383 non-null float64
        sod
                        313 non-null float64
        pot
                        312 non-null float64
        hemo
                        348 non-null float64
        pcv
        wbcc
                        329 non-null float64
        rbcc
                        294 non-null float64
                        269 non-null float64
        htn
                        398 non-null object
        dm
                        398 non-null object
        cad
        appet
                        398 non-null object
                        399 non-null object
        pe
                        399 non-null object
        ane
                        399 non-null object
        class
                        400 non-null object
        Unnamed: 25
        dtypes: float64(15), object(11)
        memory usage: 81.3+ KB
```

In [9]: df.describe()

	age	bp	sg	al	su	rbc	bu	sc	sod	pot	hemo	
count	0.0	391. 000000	388. 000000	353. 000000	354. 000000	351. 000000	356.000000	381. 000000	383. 000000	313. 000000	312. 000000	348
mean	NaN	51. 483376	76. 469072	1. 017408	1. 016949	0. 450142	148. 036517	57. 425722	3. 072454	137. 528754	4. 627244	12
std	NaN	17. 169714	13. 683637	0.005717	1. 352679	1. 099191	79. 281714	50. 503006	5.741126	10.408752	3. 193904	8
min	NaN	2. 000000	50.000000	1. 005000	0.000000	0.000000	22.000000	1. 500000	0.400000	4. 500000	2.500000	
25%	NaN	42.000000	70.000000	1. 010000	0.000000	0.000000	99. 000000	27. 000000	0. 900000	135. 000000	3. 800000	10
50%	NaN	55.000000	80.000000	1. 020000	0.000000	0.000000	121. 000000	42. 000000	1. 300000	138. 000000	4. 400000	12
75%	NaN	64. 500000	80.000000	1. 020000	2.000000	0.000000	163. 000000	66.000000	2.800000	142. 000000	4. 900000	15
max	NaN	90.000000	180. 000000	1. 025000	5.000000	5. 000000	490.000000	391. 000000	76.000000	163. 000000	47. 000000	17
	mean std min 25% 50%	count 0.0 mean NaN std NaN min NaN 25% NaN 50% NaN 75% NaN	count 0.0 391.0000000 mean NaN 51.483376 std NaN 17.169714 min NaN 2.0000000 25% NaN 42.0000000 50% NaN 55.0000000 75% NaN 64.5000000	count 0.0 391.000000 388.000000 mean NaN 51.483376 76.469072 std NaN 17.169714 13.683637 min NaN 2.000000 50.000000 25% NaN 42.000000 70.000000 50% NaN 55.000000 80.000000 75% NaN 64.500000 80.000000	count 0.0 391.000000 388.000000 353.000000 mean NaN 51.483376 76.469072 1.017408 std NaN 17.169714 13.683637 0.005717 min NaN 2.000000 50.000000 1.005000 25% NaN 42.000000 70.000000 1.010000 50% NaN 55.000000 80.000000 1.020000 75% NaN 64.500000 80.000000 1.020000	count 0. 0 391.000000 388.000000 353.000000 354.000000 mean NaN 51.483376 76.469072 1.017408 1.016949 std NaN 17.169714 13.683637 0.005717 1.352679 min NaN 2.000000 50.000000 1.005000 0.000000 25% NaN 42.000000 70.000000 1.010000 0.000000 50% NaN 55.000000 80.000000 1.020000 2.000000 75% NaN 64.500000 80.000000 1.020000 2.000000	count 0. 0 391.000000 388.000000 353.000000 354.000000 351.000000 mean NaN 51.483376 76.469072 1.017408 1.016949 0.450142 std NaN 17.169714 13.683637 0.005717 1.352679 1.099191 min NaN 2.000000 50.000000 1.005000 0.000000 0.000000 25% NaN 42.000000 70.000000 1.010000 0.000000 0.000000 50% NaN 55.000000 80.000000 1.020000 0.000000 0.000000 75% NaN 64.500000 80.000000 1.020000 2.000000 0.000000	count 0. 0 391.000000 388.000000 353.000000 354.000000 351.000000 356.000000 mean NaN 51.483376 76.469072 1.017408 1.016949 0.450142 148.036517 std NaN 17.169714 13.683637 0.005717 1.352679 1.099191 79.281714 min NaN 2.000000 50.000000 1.005000 0.000000 0.000000 22.000000 25% NaN 42.000000 70.000000 1.010000 0.000000 0.000000 121.000000 50% NaN 55.000000 80.000000 1.020000 2.000000 0.000000 163.000000	count 0.0 391.000000 388.000000 353.000000 354.000000 351.000000 356.000000 381.000000 mean NaN 51.483376 76.469072 1.017408 1.016949 0.450142 148.036517 57.425722 std NaN 17.169714 13.683637 0.005717 1.352679 1.099191 79.281714 50.503006 min NaN 2.000000 50.000000 1.005000 0.000000 0.000000 22.000000 1.500000 25% NaN 42.000000 80.000000 1.020000 0.000000 0.000000 121.000000 42.000000 75% NaN 64.500000 80.000000 1.020000 2.000000 0.000000 163.000000 66.000000	count 0. 0 391.000000 388.000000 353.000000 354.000000 351.000000 356.000000 381.000000 383.000000 mean NaN 51.483376 76.469072 1.017408 1.016949 0.450142 148.036517 57.425722 3.072454 std NaN 17.169714 13.683637 0.005717 1.352679 1.099191 79.281714 50.503006 5.741126 min NaN 2.000000 50.000000 1.005000 0.000000 0.000000 22.000000 1.500000 0.400000 25% NaN 42.000000 70.000000 1.020000 0.000000 0.000000 99.000000 27.000000 0.900000 50% NaN 55.000000 80.000000 1.020000 2.000000 0.000000 163.00000 66.000000 2.800000	count 0. 0 391.000000 388.000000 353.000000 354.000000 351.000000 356.000000 381.000000 383.000000 313.000000 mean NaN 51.483376 76.469072 1.017408 1.016949 0.450142 148.036517 57.425722 3.072454 137.528754 std NaN 17.169714 13.683637 0.005717 1.352679 1.099191 79.281714 50.503006 5.741126 10.408752 min NaN 2.000000 50.00000 1.005000 0.000000 0.000000 22.000000 1.500000 0.400000 4.500000 25% NaN 42.000000 70.000000 1.010000 0.000000 0.000000 99.000000 27.000000 0.900000 135.000000 50% NaN 55.000000 80.000000 1.020000 2.000000 0.000000 163.000000 66.000000 2.800000 142.000000	count 0.0 391.000000 388.000000 353.000000 354.000000 351.000000 356.000000 381.000000 383.000000 313.000000 312.000000 mean NaN 51.483376 76.469072 1.017408 1.016949 0.450142 148.036517 57.425722 3.072454 137.528754 4.627244 std NaN 17.169714 13.683637 0.005717 1.352679 1.099191 79.281714 50.503006 5.741126 10.408752 3.193904 min NaN 2.000000 50.00000 1.005000 0.000000 0.000000 22.000000 1.500000 0.400000 4.500000 2.500000 25% NaN 42.000000 70.00000 1.010000 0.000000 0.000000 99.000000 27.000000 0.900000 135.00000 4.400000 50% NaN 55.00000 80.00000 1.020000 2.000000 0.000000 121.000000 42.000000 1.300000 142.000000 4.900000

```
data.shape
           (400, 26)
 Out[7]:
 In [8]:
            df.info()
           <class 'pandas.core.frame.DataFrame'>
           RangeIndex: 400 entries, 0 to 399
           Data columns (total 26 columns):
                             0 non-null float64
           age
           bp
                             391 non-null float64
                             388 non-null float64
           sg
                             353 non-null float64
           al
                             354 non-null float64
           su
           rbc
                             351 non-null float64
                             248 non-null object
           pc
                             335 non-null object
           pcc
                             396 non-null object
           ba
                             396 non-null object
           bgr
           bu
                             356 non-null float64
                             381 non-null float64
           SC
                             383 non-null float64
           sod
                             313 non-null float64
           pot
                             312 non-null float64
           hemo
                             348 non-null float64
           pcv
                             329 non-null float64
           wbcc
                             294 non-null float64
           rbcc
                             269 non-null float64
           htn
                             398 non-null object
           dm
                             398 non-null object
           cad
                             398 non-null object
           appet
                             399 non-null object
           pe
                             399 non-null object
           ane
           class
                             399 non-null object
           Unnamed: 25
                             400 non-null object
           dtypes: float64(15), object(11)
           memory usage: 81.3+ KB
 In [9]:
            df.describe()
                                                                       rbe
                                                                                   bu
                            bp
                                                   al
                                                                                                                   pot
 Out[9]:
                                                                                                        sod
                                                                                                                            hemo
                 age
                                       sg
                                                              su
                     391.000000
                 0.0
                                388.000000
                                           353.000000 354.000000
                                                                 351.000000
                                                                           356.000000
                                                                                       381. 000000
                                                                                                 383.000000
                                                                                                            313.000000
                                                                                                                       312.000000
           count
                      51. 483376
                                 76.469072
                                              1.017408
                                                         1.016949
                                                                   0.450142
                                                                            148. 036517
                                                                                       57.425722
                                                                                                   3.072454
                                                                                                            137. 528754
                                                                                                                        4.627244
                                                                                                                                  12
                NaN
            mean
                       17.169714
                                 13.683637
                                             0.005717
                                                        1. 352679
                                                                    1. 099191
                                                                                       50.503006
                                                                                                             10.408752
                                                                                                                         3.193904
                NaN
                                                                             79. 281714
                                                                                                    5.741126
             std
                                             1.005000
                       2.000000
                                                        0.000000
                                                                  0.000000
                                                                                        1.500000
                                                                                                             4.500000
                NaN
                                 50.000000
                                                                            22.000000
                                                                                                   0.400000
                                                                                                                        2.500000
                      42.000000
                                 70.000000
                                              1.010000
                                                        0.000000
                                                                  0.000000
                                                                            99.000000
                                                                                                            135.000000
                                                                                                                                   10
                                                                                       27.000000
                                                                                                   0.900000
                                                                                                                        3.800000
            25%
                NaN
                                 80.000000
                      55.000000
                                                        0.000000
                                                                  0.000000
                                                                            121.000000
                                                                                                   1. 300000
                                                                                                            138.000000
                                                                                                                                  12
                NaN
                                             1.020000
                                                                                       42.000000
                                                                                                                        4.400000
                      64.500000
                                 80.000000
                                             1.020000
                                                                  0.000000
                                                                                                   2.800000 142.000000
                                                                                                                                  15
            75%
                NaN
                                                        2.000000
                                                                            163.000000
                                                                                       66.000000
                                                                                                                        4.900000
                     90.000000
                                                                  5.000000
                                                                                                  76.000000 163.000000
                                                                                                                                   17
                                180.000000
                                             1.025000
                                                        5.000000
                                                                           490.000000
                                                                                      391.000000
                                                                                                                       47.000000
            max NaN
In [10]:
            df.isna().sum()
                             400
           age
Out[10]:
                               9
           bp
                              12
           sg
           al
                              47
           su
                              46
                              49
           rbc
                             152
           pc
                              65
           pcc
                               4
           ba
           bgr
                               4
           bu
                              44
                              19
           SC
                              17
           sod
                              87
           pot
           hemo
                              88
                              52
           pcv
           wbcc
                              71
                             106
           rbcc
           htn
                             131
                               2
           dm
           cad
           appet
           pe
           ane
           class
           Unnamed: 25
           dtype: int64
 In [ ]:
```

In [7]: