Team ID	PNT2022TMID52708
Project Name	Early Detection of Chronic Kidney Disease using Machine Learning

Undersatanding Data type and summary of Features

```
In [6]: data['class'].unique()
Out[6]: array(['ckd', 'ckd\t', 'notckd'], dtype=object)

In [7]: data['class']=data['class'].replace('ckd\t','ckd')
        data['class'].unique()
Out[7]: array(['ckd', 'notckd'], dtype=object)

In [8]: catcols=set(data.dtypes[data.dtypes=='O'].index.values)
        print(catcols)

        {'coronary_artery_disease', 'pus_cell_clumps', 'bacteria', 'diabetes_mellitus', 'packed_cell_volume', 'anemia', 'white_blood_ce
        ll_count', 'pedal_edema', 'red_blood_cell_count', 'pus_cell', 'appetite', 'class', 'hypertension', 'red_blood_cells'}

In [9]: for i in catcols:
            print("Column : ",i)
            print(c(data[i]))
            print('\n')

        Column :  coronary_artery_disease
        Counter({'no': 362, 'yes': 34, '\tno': 2, nan: 2})


        Column :  pus_cell_clumps
        Counter({'notpresent': 354, 'present': 42, nan: 4})


        Column :  bacteria
        Counter({'notpresent': 374, 'present': 22, nan: 4})


        Column :  diabetes_mellitus
        Counter({'no': 258, 'yes': 134, '\tno': 3, '\tyes': 2, nan: 2, ' yes': 1})


        Column :  packed_cell_volume
        Counter({nan: 70, '52': 21, '41': 21, '44': 19, '48': 19, '40': 16, '43': 14, '45': 13, '42': 13, '32': 12, '36': 12, '33': 12,
        '28': 12, '50': 12, '37': 11, '34': 11, '35': 9, '29': 9, '30': 9, '46': 9, '31': 8, '39': 7, '24': 7, '26': 6, '38': 5, '47':
        4, '49': 4, '53': 4, '51': 4, '54': 4, '27': 3, '22': 3, '25': 3, '23': 2, '19': 2, '16': 1, '\t?': 1, '14': 1, '18': 1, '17':
        1, '15': 1, '21': 1, '20': 1, '\t43': 1, '9': 1})


        Column :  anemia
        Counter({'no': 339, 'yes': 60, nan: 1})


        Column :  white_blood_cell_count
        Counter({nan: 105, '9800': 11, '6700': 10, '9600': 9, '9200': 9, '7200': 9, '6900': 8, '11000': 8, '5800': 8, '7800': 7, '910
        0': 7, '9400': 7, '7000': 7, '4300': 6, '6300': 6, '10700': 6, '10500': 6, '7500': 5, '8300': 5, '7900': 5, '8600': 5, '5600':
        5, '10200': 5, '5000': 5, '8100': 5, '9500': 5, '6000': 4, '6200': 4, '10300': 4, '7700': 4, '5500': 4, '10400': 4, '6800': 4,
        '6500': 4, '4700': 4, '7300': 3, '4500': 3, '8400': 3, '6400': 3, '4200': 3, '7400': 3, '8000': 3, '5400': 3, '3800': 2, '1140
        0': 2, '5300': 2, '8500': 2, '14600': 2, '7100': 2, '13200': 2, '9000': 2, '8200': 2, '15200': 2, '12400': 2, '12800': 2, '880
        0': 2, '5700': 2, '9300': 2, '6600': 2, '12100': 1, '12200': 1, '18900': 1, '21600': 1, '11300': 1, '\t6200': 1, '11800': 1, '1
        2500': 1, '11900': 1, '12700': 1, '13600': 1, '14900': 1, '16300': 1, '\t8400': 1, '10900': 1, '2200': 1, '11200': 1, '19100':
        1, '\t?': 1, '12300': 1, '16700': 1, '2600': 1, '26400': 1, '4900': 1, '12000': 1, '15700': 1, '4100': 1, '11500': 1, '10800':
        1, '9900': 1, '5200': 1, '5900': 1, '9700': 1, '5100': 1})


        Column :  pedal_edema
        Counter({'no': 323, 'yes': 76, nan: 1})
```

```
Column :  red_blood_cell_count
Counter({nan: 130, '5.2': 18, '4.5': 16, '4.9': 14, '4.7': 11, '3.9': 10, '4.8': 10, '4.6': 9, '3.4': 9, '3.7': 8, '5.0': 8,
'6.1': 8, '5.5': 8, '5.9': 8, '3.8': 7, '5.4': 7, '5.8': 7, '5.3': 7, '4.3': 6, '4.2': 6, '5.6': 6, '4.4': 5, '3.2': 5, '4.1':
5, '6.2': 5, '5.1': 5, '6.4': 5, '5.7': 5, '6.5': 5, '3.6': 4, '6.0': 4, '6.3': 4, '4.0': 3, '4': 3, '3.5': 3, '3.3': 3, '5':
2, '2.6': 2, '2.8': 2, '2.5': 2, '3.1': 2, '2.1': 2, '2.9': 2, '2.7': 2, '3.0': 2, '2.3': 1, '8.0': 1, '3': 1, '2.4': 1, '\t?':
1})


Column :  pus_cell
Counter({'normal': 259, 'abnormal': 76, nan: 65})


Column :  appetite
Counter({'good': 317, 'poor': 82, nan: 1})


Column :  class
Counter({'ckd': 250, 'notckd': 150})


Column :  hypertension
Counter({'no': 251, 'yes': 147, nan: 2})


Column :  red_blood_cells
Counter({'normal': 201, nan: 152, 'abnormal': 47})
```