Understanding Data type and Summary of Features

Team ID	PNT2022TMID17480
Project Title	Early Detection of Chronic Kidney Disease using Machine Learning

```
{
 "nbformat": 4,
 "nbformat_minor": 0,
 "metadata": {
  "colab": {
   "provenance": [],
   "collapsed_sections": []
  },
  "kernelspec": {
   "name": "python3",
   "display_name": "Python 3"
  },
  "language_info": {
   "name": "python"
  }
 },
 "cells": [
  {
   "cell_type": "code",
   "source": [
    "data = pd.read_csv(\"/content/chronickidneydisease.csv\");"
   ],
   "metadata": {
    "id": "A2aBgR4sRSsU"
   },
   "execution_count": null,
   "outputs": []
  },
```

```json
{
  "cell_type": "markdown",
  "source": [
    "#Understaning the datatype and summary"
  ],
  "metadata": {
    "id": "ed86CRu7SG5r"
  }
},
{
  "cell_type": "code",
  "source": [
    "data.dtypes"
  ],
  "metadata": {
    "colab": {
      "base_uri": "https://localhost:8080/"
    },
    "id": "D2vLOokESGKT",
    "outputId": "56c71e26-39b3-4b1e-ea38-11a5a7592f80"
  },
  "execution_count": null,
  "outputs": [
    {
      "output_type": "execute_result",
      "data": {
        "text/plain": [
          "id              int64\n",
          "age            float64\n",
          "bp             float64\n",
          "sg             float64\n",
          "al             float64\n",
          "su             float64\n",
          "rbc             object\n",
          "pc              object\n",
          "pcc             object\n",
          "ba              object\n",
          "bgr            float64\n",
          "bu             float64\n",
          "sc             float64\n",
```

```
      "sod          float64\n",
      "pot          float64\n",
      "hemo           float64\n",
      "pcv            object\n",
      "wc             object\n",
      "rc            object\n",
      "htn            object\n",
      "dm             object\n",
      "cad            object\n",
      "appet            object\n",
      "pe            object\n",
      "ane             object\n",
      "classification     object\n",
      "dtype: object"
     ]
    },
    "metadata": {},
    "execution_count": 4
   }
  ]
 },
 {
  "cell_type": "code",
  "source": [
   "data.describe()"
  ],
  "metadata": {
   "colab": {
    "base_uri": "https://localhost:8080/",
    "height": 364
   },
   "id": "-j6uRsumS74T",
   "outputId": "a4f160b5-e061-45cc-df9c-2db7eaf7944e"
  },
  "execution_count": null,
  "outputs": [
   {
    "output_type": "execute_result",
    "data": {
     "text/plain": [
```

"   id   age   bp   sg   al   su \\\n",
"count 400.000000 391.000000 388.000000 353.000000 354.000000 351.000000 \n",
"mean   199.500000   51.483376   76.469072   1.017408   1.016949   0.450142 \n",
"std   115.614301   17.169714   13.683637   0.005717   1.352679   1.099191 \n",
"min   0.000000   2.000000   50.000000   1.005000   0.000000   0.000000 \n",
"25%   99.750000   42.000000   70.000000   1.010000   0.000000   0.000000 \n",
"50%   199.500000   55.000000   80.000000   1.020000   0.000000   0.000000 \n",
"75%   299.250000   64.500000   80.000000   1.020000   2.000000   0.000000 \n",
"max   399.000000   90.000000 180.000000   1.025000   5.000000   5.000000 \n",
"\n",
"   bgr   bu   sc   sod   pot   hemo \n",
"count 356.000000 381.000000 383.000000 313.000000 312.000000 348.000000 \n",
"mean   148.036517   57.425722   3.072454 137.528754   4.627244   12.526437 \n",
"std   79.281714   50.503006   5.741126   10.408752   3.193904   2.912587 \n",
"min   22.000000   1.500000   0.400000   4.500000   2.500000   3.100000 \n",
"25%   99.000000   27.000000   0.900000 135.000000   3.800000   10.300000 \n",
"50%   121.000000   42.000000   1.300000 138.000000   4.400000   12.650000 \n",
"75%   163.000000   66.000000   2.800000 142.000000   4.900000   15.000000 \n",
"max   490.000000 391.000000   76.000000 163.000000   47.000000   17.800000 "
],
"text/html": [
"\n",
" <div id=\"df-a1553464-1772-4f99-b421-bb832b30ccf0\">\n",
"   <div class=\"colab-df-container\">\n",

```
"      <div>\n",
"<style scoped>\n",
"    .dataframe tbody tr th:only-of-type {\n",
"        vertical-align: middle;\n",
"    }\n",
"\n",
"    .dataframe tbody tr th {\n",
"        vertical-align: top;\n",
"    }\n",
"\n",
"    .dataframe thead th {\n",
"        text-align: right;\n",
"    }\n",
"</style>\n",
"<table border=\"1\" class=\"dataframe\">\n",
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"      <th></th>\n",
"      <th>id</th>\n",
"      <th>age</th>\n",
"      <th>bp</th>\n",
"      <th>sg</th>\n",
"      <th>al</th>\n",
"      <th>su</th>\n",
"      <th>bgr</th>\n",
"      <th>bu</th>\n",
"      <th>sc</th>\n",
"      <th>sod</th>\n",
"      <th>pot</th>\n",
"      <th>hemo</th>\n",
"    </tr>\n",
"  </thead>\n",
"  <tbody>\n",
"    <tr>\n",
"      <th>count</th>\n",
"      <td>400.000000</td>\n",
"      <td>391.000000</td>\n",
"      <td>388.000000</td>\n",
"      <td>353.000000</td>\n",
"      <td>354.000000</td>\n",
```

"   <td>351.000000</td>\n",
"   <td>356.000000</td>\n",
"   <td>381.000000</td>\n",
"   <td>383.000000</td>\n",
"   <td>313.000000</td>\n",
"   <td>312.000000</td>\n",
"   <td>348.000000</td>\n",
"   </tr>\n",
"   <tr>\n",
"   <th>mean</th>\n",
"   <td>199.500000</td>\n",
"   <td>51.483376</td>\n",
"   <td>76.469072</td>\n",
"   <td>1.017408</td>\n",
"   <td>1.016949</td>\n",
"   <td>0.450142</td>\n",
"   <td>148.036517</td>\n",
"   <td>57.425722</td>\n",
"   <td>3.072454</td>\n",
"   <td>137.528754</td>\n",
"   <td>4.627244</td>\n",
"   <td>12.526437</td>\n",
"   </tr>\n",
"   <tr>\n",
"   <th>std</th>\n",
"   <td>115.614301</td>\n",
"   <td>17.169714</td>\n",
"   <td>13.683637</td>\n",
"   <td>0.005717</td>\n",
"   <td>1.352679</td>\n",
"   <td>1.099191</td>\n",
"   <td>79.281714</td>\n",
"   <td>50.503006</td>\n",
"   <td>5.741126</td>\n",
"   <td>10.408752</td>\n",
"   <td>3.193904</td>\n",
"   <td>2.912587</td>\n",
"   </tr>\n",
"   <tr>\n",
"   <th>min</th>\n",

"   <td>0.000000</td>\n",
"   <td>2.000000</td>\n",
"   <td>50.000000</td>\n",
"   <td>1.005000</td>\n",
"   <td>0.000000</td>\n",
"   <td>0.000000</td>\n",
"   <td>22.000000</td>\n",
"   <td>1.500000</td>\n",
"   <td>0.400000</td>\n",
"   <td>4.500000</td>\n",
"   <td>2.500000</td>\n",
"   <td>3.100000</td>\n",
"   </tr>\n",
"   <tr>\n",
"   <th>25%</th>\n",
"   <td>99.750000</td>\n",
"   <td>42.000000</td>\n",
"   <td>70.000000</td>\n",
"   <td>1.010000</td>\n",
"   <td>0.000000</td>\n",
"   <td>0.000000</td>\n",
"   <td>99.000000</td>\n",
"   <td>27.000000</td>\n",
"   <td>0.900000</td>\n",
"   <td>135.000000</td>\n",
"   <td>3.800000</td>\n",
"   <td>10.300000</td>\n",
"   </tr>\n",
"   <tr>\n",
"   <th>50%</th>\n",
"   <td>199.500000</td>\n",
"   <td>55.000000</td>\n",
"   <td>80.000000</td>\n",
"   <td>1.020000</td>\n",
"   <td>0.000000</td>\n",
"   <td>0.000000</td>\n",
"   <td>121.000000</td>\n",
"   <td>42.000000</td>\n",
"   <td>1.300000</td>\n",
"   <td>138.000000</td>\n",

```
"        <td>4.400000</td>\n",
"        <td>12.650000</td>\n",
"      </tr>\n",
"      <tr>\n",
"        <th>75%</th>\n",
"        <td>299.250000</td>\n",
"        <td>64.500000</td>\n",
"        <td>80.000000</td>\n",
"        <td>1.020000</td>\n",
"        <td>2.000000</td>\n",
"        <td>0.000000</td>\n",
"        <td>163.000000</td>\n",
"        <td>66.000000</td>\n",
"        <td>2.800000</td>\n",
"        <td>142.000000</td>\n",
"        <td>4.900000</td>\n",
"        <td>15.000000</td>\n",
"      </tr>\n",
"      <tr>\n",
"        <th>max</th>\n",
"        <td>399.000000</td>\n",
"        <td>90.000000</td>\n",
"        <td>180.000000</td>\n",
"        <td>1.025000</td>\n",
"        <td>5.000000</td>\n",
"        <td>5.000000</td>\n",
"        <td>490.000000</td>\n",
"        <td>391.000000</td>\n",
"        <td>76.000000</td>\n",
"        <td>163.000000</td>\n",
"        <td>47.000000</td>\n",
"        <td>17.800000</td>\n",
"      </tr>\n",
"  </tbody>\n",
"</table>\n",
"</div>\n",
"      <button class=\"colab-df-convert\"
onclick=\"convertToInteractive('df-a1553464-1772-4f99-b421-
bb832b30ccf0')\"\n",
"              title=\"Convert this dataframe to an interactive table.\"\n",
```

"   style=\"display:none;\">\n",
"   \n",
" <svg xmlns=\"http://www.w3.org/2000/svg\" height=\"24px\"viewBox=\"0 0 24 24\"\n",
"   width=\"24px\">\n",
"   <path d=\"M0 0h24v24H0V0z\" fill=\"none\"/>\n",
"   <path d=\"M18.56 5.44l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94zm-11 1L8.5 8.5l.94-2.06 2.06-.94-2.06-.94L8.5 2.5l-.94 2.06-2.06.94zm10 10l.94 2.06.94-2.06 2.06-.94-2.06-.94-.94-2.06-.94 2.06-2.06.94z\"/><path d=\"M17.41 7.96l-1.37-1.37c-.4-.4-.92-.59-1.43-.59-.52 0-1.04.2-1.43.59L10.3 9.45l-7.72 7.72c-.78.78-.78 2.05 0 2.83L4 21.41c.39.39.9.59 1.41.59.51 0 1.02-.2 1.41-.59l7.78-7.78 2.81-2.81c.8-.78.8-2.07 0-2.86zM5.41 20L4 18.59l7.72-7.72 1.47 1.35L5.41 20z\"/>\n",
" </svg>\n",
"   </button>\n",
"   \n",
" <style>\n",
"   .colab-df-container {\n",
"   display:flex;\n",
"   flex-wrap:wrap;\n",
"   gap: 12px;\n",
"   }\n",
"\n",
"   .colab-df-convert {\n",
"   background-color: #E8F0FE;\n",
"   border: none;\n",
"   border-radius: 50%;\n",
"   cursor: pointer;\n",
"   display: none;\n",
"   fill: #1967D2;\n",
"   height: 32px;\n",
"   padding: 0 0 0 0;\n",
"   width: 32px;\n",
"   }\n",
"\n",
"   .colab-df-convert:hover {\n",
"   background-color: #E2EBFA;\n",
"   box-shadow: 0px 1px 2px rgba(60, 64, 67, 0.3), 0px 1px 3px 1px rgba(60, 64, 67, 0.15);\n",
"   fill: #174EA6;\n",

```
    "    }\n",
"\n",
"    [theme=dark] .colab-df-convert {\n",
"      background-color: #3B4455;\n",
"      fill: #D2E3FC;\n",
"    }\n",
"\n",
"    [theme=dark] .colab-df-convert:hover {\n",
"      background-color: #434B5C;\n",
"      box-shadow: 0px 1px 3px 1px rgba(0, 0, 0, 0.15);\n",
"      filter: drop-shadow(0px 1px 2px rgba(0, 0, 0, 0.3));\n",
"      fill: #FFFFFF;\n",
"    }\n",
"  </style>\n",
"\n",
"      <script>\n",
"        const buttonEl =\n",
"          document.querySelector('#df-a1553464-1772-4f99-b421-
bb832b30ccf0 button.colab-df-convert');\n",
"        buttonEl.style.display =\n",
"          google.colab.kernel.accessAllowed ? 'block' : 'none';\n",
"\n",
"        async function convertToInteractive(key) {\n",
"          const element = document.querySelector('#df-a1553464-1772-
4f99-b421-bb832b30ccf0');\n",
"          const dataTable =\n",
"            await
google.colab.kernel.invokeFunction('convertToInteractive',\n",
"                                     [key], {});\n",
"          if (!dataTable) return;\n",
"\n",
"          const docLinkHtml = 'Like what you see? Visit the ' +\n",
"            '<a target=\"_blank\"
href=https://colab.research.google.com/notebooks/data_table.ipynb>data table
notebook</a>'\n",
"            + ' to learn more about interactive tables.';\n",
"          element.innerHTML = '';\n",
"          dataTable['output_type'] = 'display_data';\n",
"          await google.colab.output.renderOutput(dataTable, element);\n",
"          const docLink = document.createElement('div');\n",
```

```
                  "          docLink.innerHTML = docLinkHtml;\n",
                  "          element.appendChild(docLink);\n",
                  "        }\n",
                  "      </script>\n",
                  "    </div>\n",
                  "  </div>\n",
                  "  "
                ]
              },
              "metadata": {},
              "execution_count": 5
            }
          ]
        },
        {
          "cell_type": "code",
          "source": [
            "data.info()"
          ],
          "metadata": {
            "colab": {
              "base_uri": "https://localhost:8080/"
            },
            "id": "wOVyTOWmTCm6",
            "outputId": "80bf4f18-7613-4915-b1ff-a421cf94dc6a"
          },
          "execution_count": null,
          "outputs": [
            {
              "output_type": "stream",
              "name": "stdout",
              "text": [
                "<class 'pandas.core.frame.DataFrame'>\n",
                "RangeIndex: 400 entries, 0 to 399\n",
                "Data columns (total 26 columns):\n",
                " #   Column         Non-Null Count  Dtype  \n",
                "---  ------         --------------  -----  \n",
                " 0   id             400 non-null    int64  \n",
                " 1   age            391 non-null    float64\n",
                " 2   bp             388 non-null    float64\n",
```

" 3   sg   353 non-null   float64\n",
" 4   al   354 non-null   float64\n",
" 5   su   351 non-null   float64\n",
" 6   rbc   248 non-null   object \n",
" 7   pc   335 non-null   object \n",
" 8   pcc   396 non-null   object \n",
" 9   ba   396 non-null   object \n",
" 10 bgr   356 non-null   float64\n",
" 11 bu   381 non-null   float64\n",
" 12 sc   383 non-null   float64\n",
" 13 sod   313 non-null   float64\n",
" 14 pot   312 non-null   float64\n",
" 15 hemo   348 non-null   float64\n",
" 16 pcv   330 non-null   object \n",
" 17 wc   295 non-null   object \n",
" 18 rc   270 non-null   object \n",
" 19 htn   398 non-null   object \n",
" 20 dm   398 non-null   object \n",
" 21 cad   398 non-null   object \n",
" 22 appet   399 non-null   object \n",
" 23 pe   399 non-null   object \n",
" 24 ane   399 non-null   object \n",
" 25 classification 400 non-null   object \n",
"dtypes: float64(11), int64(1), object(14)\n",
"memory usage: 81.4+ KB\n"
    ]
  }
 ]
 }
]
}