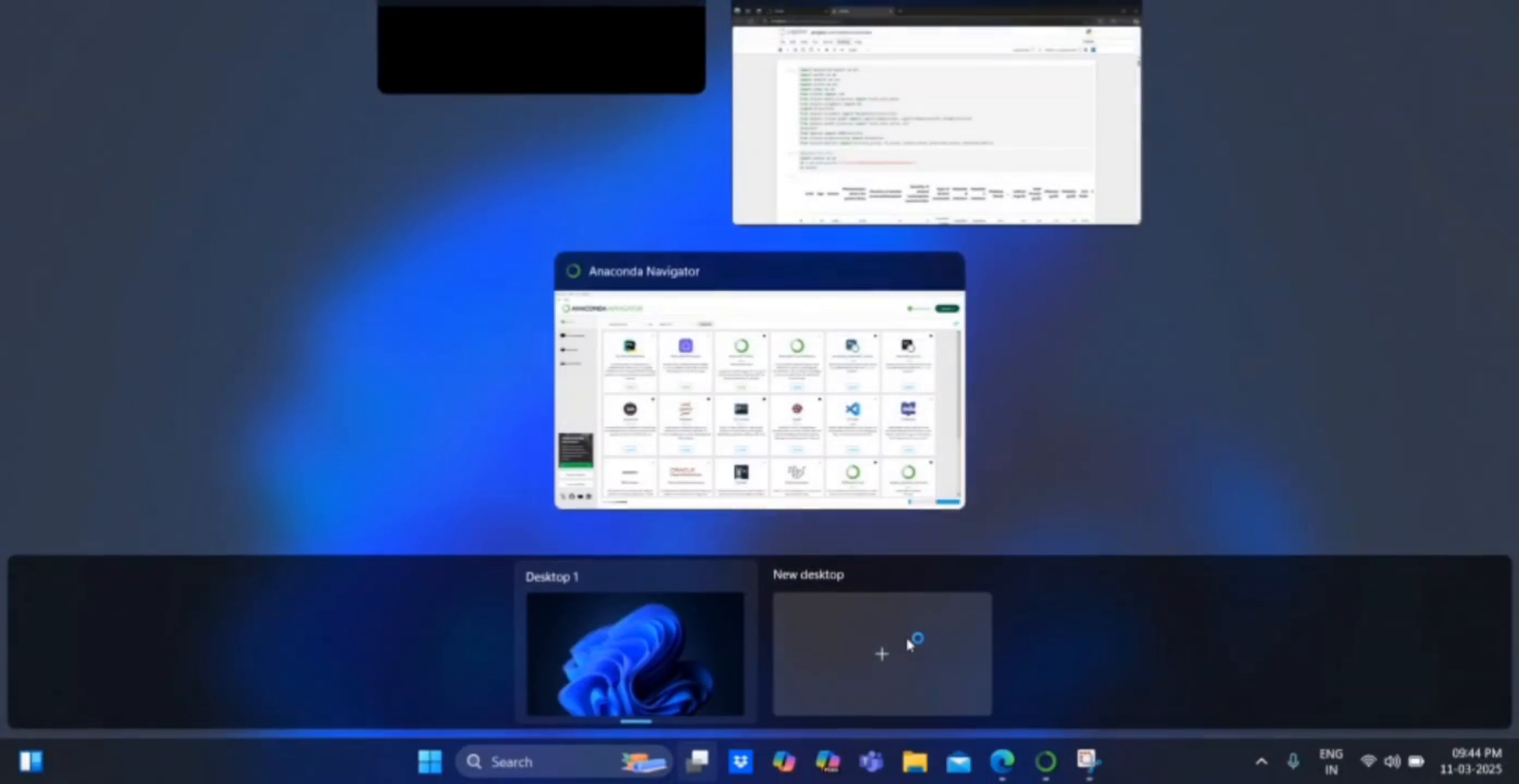

Recycle Bin
Microsoft Edge
VLC media player
c langauge
Control Panel
N Indra (indra k...
PotPlayer 64 bit
TEJO SQL INTERNSHI...
Visual Studio Code
python foundati...
Google Chrome
internship sql ppt KARTHI...
Search
ENG IN
09:44 PM
11-03-2025


Recycle Bin
Microsoft Edge
VLC media player
c langauge
Control Panel
N Indra (indra k...
PotPlayer 64 bit
TEJO SQL INTERNSHI...
Visual Studio Code
python foundati...
Google Chrome
internship sql ppt KARTHI...
Search
ENG IN
09:44 PM
11-03-2025


Anaconda Navigator
Desktop 1
New desktop
Search
ENG
IN
09:44 PM
11-03-2025


Recycle Bin
Microsoft Edge
VLC media player
c langauge
Control Panel
N Indra (indra k...
PotPlayer 64 bit
TEJO SQL INTERNSHI...
Visual Studio Code
python foundati...
Google Chrome
internship sql ppt KARTHI...
Search
ENG IN
09:44 PM
11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

```
from sklearn.metrics import accuracy_score, f1_score, recall_score, precision_score, confusion_matrix
```

[13]:
```
#Reading CSV file
import pandas as pd
df = pd.read_excel(r'C:\Users\uday\Data\HealthCareData.xlsx')
df.head()
```

[13]:

| | S.NO | Age | Gender | Place(location where the patient lives) | Duration of alcohol consumption(years) | Quantity of alcohol consumption (quarters/day) | Type of alcohol consumed | Hepatitis B infection | Hepatitis C infection | Diabetes Result | ... | Indirect (mg/dl) | Total Protein (g/dl) | Albumin (g/dl) | Globulin (g/dl) | A/G Ratio | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 55 | male | rural | 12 | 2 | branded liquor | negative | negative | YES | ... | 3.0 | 6.0 | 3.0 | 4.0 | 0.75 | |
| **1** | 2 | 55 | male | rural | 12 | 2 | branded liquor | negative | negative | YES | ... | 3.0 | 6.0 | 3.0 | 4.0 | 0.75 | |
| **2** | 3 | 55 | male | rural | 12 | 2 | branded liquor | negative | negative | YES | ... | 3.0 | 6.0 | 3.0 | 4.0 | 0.75 | |
| **3** | 4 | 55 | male | rural | 12 | 2 | branded liquor | negative | negative | NO | ... | 3.0 | 6.0 | 3.0 | 4.0 | 0.75 | |
| **4** | 5 | 55 | female | rural | 12 | 2 | branded liquor | negative | negative | YES | ... | 3.0 | 6.0 | 3.0 | 4.0 | 0.75 | |

5 rows × 42 columns

[14]:
```
import pandas as pd
```

Search

ENG IN

09:44 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[14]:
```
import pandas as pd
```

[16]:
```
df = pd.read_excel(r'C:\Users\uday\Data\HealthCareData.xlsx')
```

[17]:
```
print(df)
print(df.shape)
```

```
     S.NO  Age   Gender Place(location where the patient lives)  \
0       1   55     male                                  rural
1       2   55     male                                  rural
2       3   55     male                                  rural
3       4   55     male                                  rural
4       5   55   female                                  rural
..    ...  ...      ...                                    ...
945   946   54   female                                  rural
946   947   72   female                                  urban
947   948   47     male                                  urban
948   949   54   female                                  rural
949   950   52     male                                  rural

     Duration of alcohol consumption(years)  \
0                                        12
1                                        12
2                                        12
3                                        12
4                                        12
..                                      ...
945                                       5
946                                       4
947                                       7
948                                       9
949                                       8

     Quantity of alcohol consumption (quarters/day) Type of alcohol consumed  \
```

Search ENG IN 09:44 PM 11-03-2025

```
3      4   55     male                                rural
4      5   55   female                                rural
..   ...  ...      ...                                  ...
945  946   54   female                                rural
946  947   72   female                                urban
947  948   47     male                                urban
948  949   54   female                                rural
949  950   52     male                                rural

     Duration of alcohol consumption(years)  \
0                                        12
1                                        12
2                                        12
3                                        12
4                                        12
..                                      ...
945                                       5
946                                       4
947                                       7
948                                       9
949                                       8

     Quantity of alcohol consumption (quarters/day) Type of alcohol consumed  \
0                                                 2           branded liquor
1                                                 2           branded liquor
2                                                 2           branded liquor
3                                                 2           branded liquor
4                                                 2           branded liquor
..                                              ...                      ...
945                                               3          country liquor
946                                               3           branded liquor
947                                               3          country liquor
948                                               3          country liquor
949                                               3          country liquor

    Hepatitis B infection Hepatitis C infection Diabetes Result  ...  \
0                negative              negative             YES  ...
```

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[18]:
```
df.shape
```

[18]:
```
(950, 42)
```

[19]:
```
df.isnull().any()
```

[19]:
```
S.NO                                                  False
Age                                                   False
Gender                                                False
Place(location where the patient lives)                True
Duration of alcohol consumption(years)                False
Quantity of alcohol consumption (quarters/day)        False
Type of alcohol consumed                              False
Hepatitis B infection                                 False
Hepatitis C infection                                 False
Diabetes Result                                       False
Blood pressure (mmhg)                                 False
Obesity                                               False
Family history of cirrhosis/ hereditary               False
TCH                                                    True
TG                                                     True
LDL                                                    True
HDL                                                    True
Hemoglobin  (g/dl)                                    False
```

[20]:
```
df.isnull().sum()
```

[20]:
```
S.NO                                                    0
Age                                                     0
Gender                                                  0
Place(location where the patient lives)               134
Duration of alcohol consumption(years)                  0
Quantity of alcohol consumption (quarters/day)          0
Type of alcohol consumed                                0
Hepatitis B infection                                   0
Hepatitis C infection                                   0
Diabetes Result                                         0
```

Search

ENG IN

09:45 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

```
[950 rows x 42 columns]
(950, 42)
```

[18]:
```
df.shape
```

[18]:
```
(950, 42)
```

[19]:
```
df.isnull().any()
```

```
MCV    (femtoliters/cell)                 True
MCH   (picograms/cell)                    True
MCHC   (grams/deciliter)                  True
Total Count                               True
Polymorphs   (%)                         False
Lymphocytes   (%)                        False
Monocytes    (%)                          True
Eosinophils    (%)                        True
Basophils   (%)                           True
Platelet Count   (lakhs/mm)              False
Total Bilirubin     (mg/dl)              False
Direct     (mg/dl)                       False
Indirect      (mg/dl)                     True
Total Protein      (g/dl)                 True
Albumin    (g/dl)                        True
Globulin   (g/dl)                         True
A/G Ratio                                 True
AL.Phosphatase       (U/L)                True
```

[20]:
```
df.isnull().sum()
```

[20]:
```
S.NO                                        0
Age                                         0
Gender                                      0
Place(location where the patient lives)   134
Duration of alcohol consumption(years)      0
Quantity of alcohol consumption (quarters/day)  0
Type of alcohol consumed                    0
Hepatitis B infection                       0
```

Search ENG IN 09:45 PM 11-03-2025

| | |
|---|---|
| A/G Ratio | True |

[20]:
```
df.isnull().sum()
```

[20]:

| | |
|---|---|
| S.NO | 0 |
| Age | 0 |
| Gender | 0 |
| Place(location where the patient lives) | 134 |
| Duration of alcohol consumption(years) | 0 |
| Quantity of alcohol consumption (quarters/day) | 0 |
| Type of alcohol consumed | 0 |
| Hepatitis B infection | 0 |
| Hepatitis C infection | 0 |
| Diabetes Result | 0 |
| Blood pressure (mmhg) | 0 |
| Obesity | 0 |
| Family history of cirrhosis/ hereditary | 0 |
| TCH | 359 |
| TG | 359 |
| LDL | 359 |
| HDL | 368 |
| Hemoglobin (g/dl) | 0 |
| PCV (%) | 30 |
| RBC (million cells/microliter) | 552 |
| MCV (femtoliters/cell) | 9 |
| MCH (picograms/cell) | 658 |
| MCHC (grams/deciliter) | 672 |
| Total Count | 10 |
| Polymorphs (%) | 0 |
| Lymphocytes (%) | 0 |
| Monocytes (%) | 9 |
| Eosinophils (%) | 8 |
| Basophils (%) | 49 |
| Platelet Count (lakhs/mm) | 0 |
| Total Bilirubin (mg/dl) | 0 |
| Direct (mg/dl) | 0 |
| Indirect (mg/dl) | 55 |
| Total Protein (g/dl) | 61 |

```
Albumin  (g/dl)                                                    9
Globulin  (g/dl)                                                  29
A/G Ratio                                                        359
AL.Phosphatase      (U/L)                                         10
SGOT/AST      (U/L)                                                0
SGPT/ALT (U/L)                                                     0
USG Abdomen (diffuse liver or  not)                                0
Predicted Value(Out Come-Patient suffering from liver  cirrosis or not)    54
dtype: int64
```

[21]:
```
df.isnull().sum()
```

[21]:
```
S.NO                                                               0
Age                                                                0
Gender                                                             0
Place(location where the patient lives)                          134
Duration of alcohol consumption(years)                             0
Quantity of alcohol consumption (quarters/day)                     0
Type of alcohol consumed                                           0
Hepatitis B infection                                              0
Hepatitis C infection                                              0
Diabetes Result                                                    0
Blood pressure (mmhg)                                              0
Obesity                                                            0
Family history of cirrhosis/ hereditary                            0
TCH                                                              359
TG                                                               359
LDL                                                              359
HDL                                                              368
Hemoglobin  (g/dl)                                                 0
PCV  (%)                                                          30
RBC  (million cells/microliter)                                  552
MCV   (femtoliters/cell)                                           9
MCH  (picograms/cell)                                            658
MCHC  (grams/deciliter)                                          672
Total Count                                                       10
Polymorphs  (%)                                                    0
Lymphocytes  (%)                                                   0
```

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[21]:
```
df.isnull().sum()
```

[21]:
```
S.NO                                                  0
Age                                                   0
Gender                                                0
Place(location where the patient lives)             134
Duration of alcohol consumption(years)                0
Quantity of alcohol consumption (quarters/day)        0
Type of alcohol consumed                              0
Hepatitis B infection                                 0
Hepatitis C infection                                 0
Diabetes Result                                       0
Blood pressure (mmhg)                                 0
Obesity                                               0
Family history of cirrhosis/ hereditary               0
TCH                                                 359
TG                                                  359
LDL                                                 359
HDL                                                 368
Hemoglobin  (g/dl)                                    0
PCV  (%)                                             30
RBC  (million cells/microliter)                     552
MCV   (femtoliters/cell)                              9
MCH  (picograms/cell)                               658
MCHC  (grams/deciliter)                             672
Total Count                                          10
Polymorphs  (%)                                       0
Lymphocytes  (%)                                      0
Monocytes   (%)                                       9
Eosinophils   (%)                                     8
Basophils  (%)                                       49
Platelet Count  (lakhs/mm)                            0
Total Bilirubin    (mg/dl)                            0
Direct    (mg/dl)                                     0
Indirect     (mg/dl)                                 55
Total Protein     (g/dl)                             61
Albumin   (g/dl)                                      9
```

Search ENG IN 09:46 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[22]:
```
import numpy as np
```

[23]:
```
categorical_features = df.select_dtypes(include=[np.object_])
categorical_features.columns
```

[23]:
```
Index(['Gender', 'Place(location where the patient lives)',
       'Type of alcohol consumed', 'Hepatitis B infection',
       'Hepatitis C infection', 'Diabetes Result', 'Blood pressure (mmhg)',
       'Obesity', 'Family history of cirrhosis/ hereditary', 'TG', 'LDL',
       'Total Bilirubin    (mg/dl)', 'A/G Ratio',
       'USG Abdomen (diffuse liver or  not)',
       'Predicted Value(Out Come-Patient suffering from liver  cirrosis or not)'],
      dtype='object')
```

[24]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

[26]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import math  # For dynamic row calculation

# Load dataset (if applicable)
# df = pd.read_csv("your_file.csv")

# Function to clean numeric values
def clean_numeric(column):
    return pd.to_numeric(column.astype(str)
                         .str.replace(r'[^0-9.]', '', regex=True)  # Keep only numbers & periods
                         .str.replace(r'\.$', '', regex=True),  # Remove trailing periods
                         errors='coerce')

# Apply cleaning to all columns
df = df.apply(clean_numeric)
```

Search ENG IN 09:46 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[24]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

[26]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import math  # For dynamic row calculation

# Load dataset (if applicable)
# df = pd.read_csv("your_file.csv")

# Function to clean numeric values
def clean_numeric(column):
    return pd.to_numeric(column.astype(str)
                         .str.replace(r'[^0-9.]', '', regex=True)  # Keep only numbers & periods
                         .str.replace(r'\.$', '', regex=True),  # Remove trailing periods
                         errors='coerce')

# Apply cleaning to all columns
df = df.apply(clean_numeric)

# Fill missing values
df.fillna(0, inplace=True)

#  * Dynamically adjust subplot grid size
num_cols = len(df.columns)
rows = math.ceil(num_cols / 5)  # Adjust rows based on column count

plt.figure(figsize=(15, rows * 2))  # Adjust figure size dynamically
for i, col in enumerate(df.columns):
    plt.subplot(rows, 5, i + 1)  # Dynamically adjust grid
    sns.boxplot(y=df[col])
    plt.title(col)
```

Search

ENG IN

09:46 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

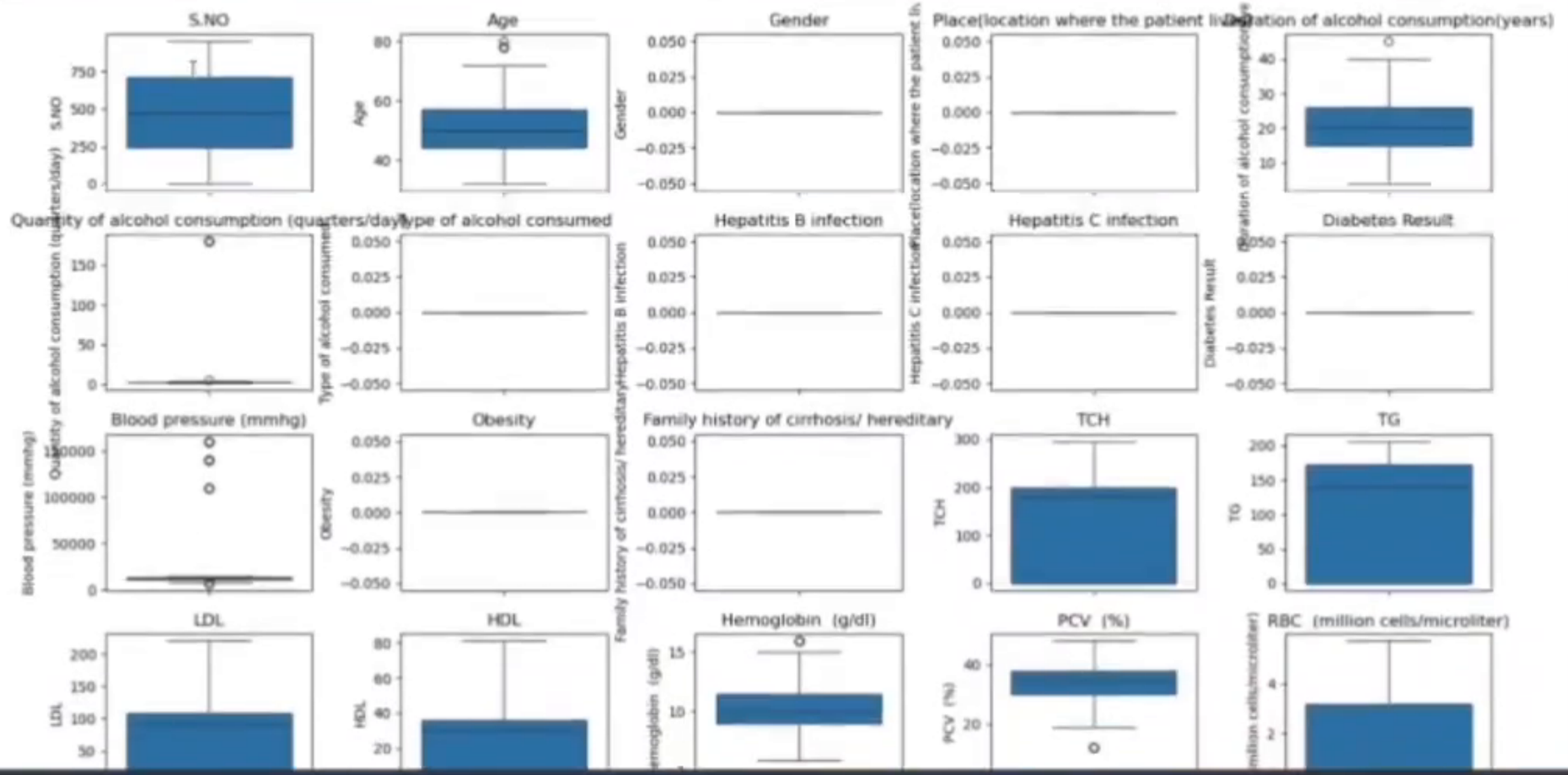
Code JupyterLab Python 3 (ipykernel)

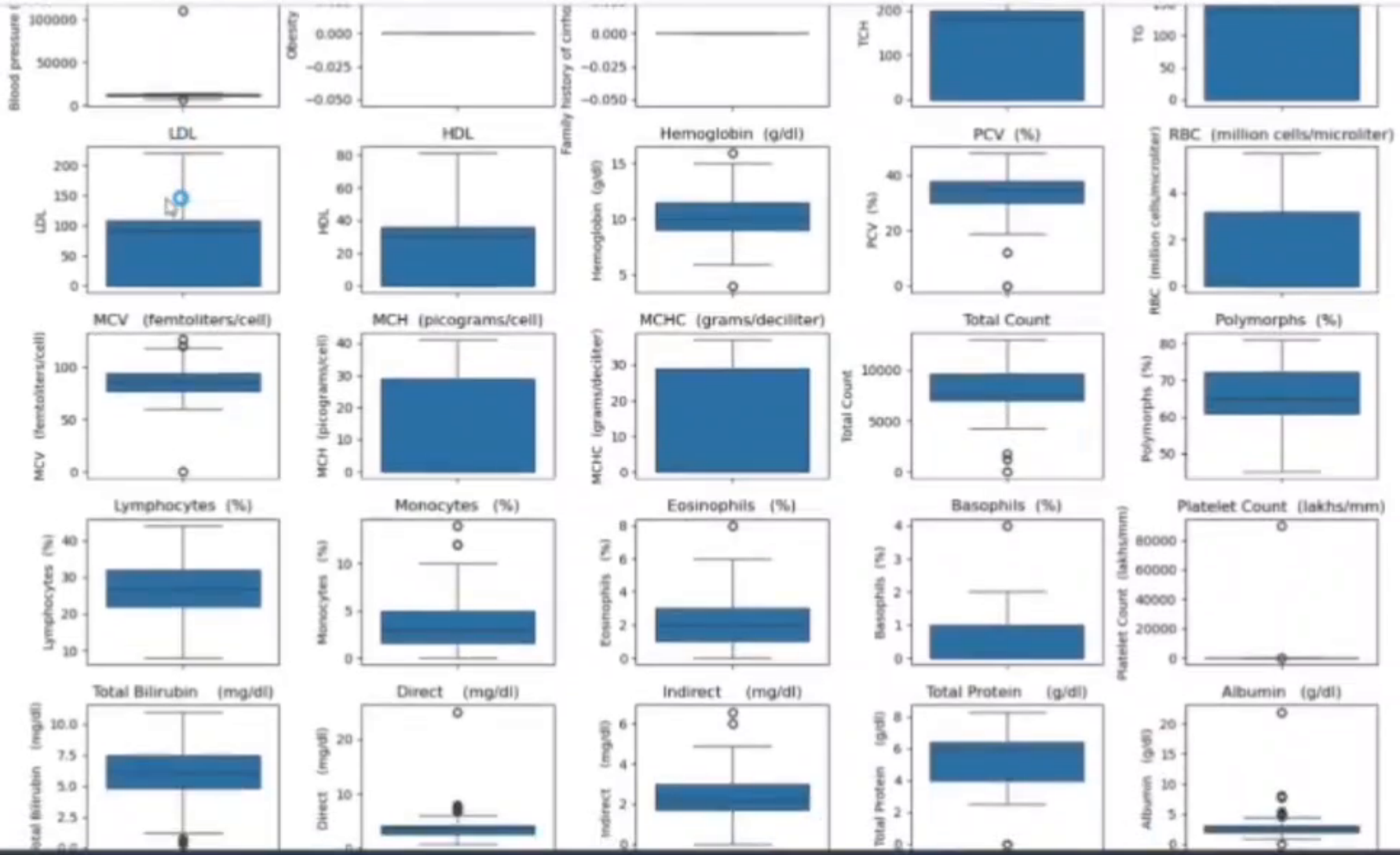
[26]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import math  # For dynamic row calculation

# Load dataset (if applicable)
# df = pd.read_csv("your_file.csv")

# Function to clean numeric values
def clean_numeric(column):
    return pd.to_numeric(column.astype(str)
                         .str.replace(r'[^0-9.]', '', regex=True)  # Keep only numbers & periods
                         .str.replace(r'\.$', '', regex=True),  # Remove trailing periods
                         errors='coerce')

# Apply cleaning to all columns
df = df.apply(clean_numeric)

# Fill missing values
df.fillna(0, inplace=True)

#  * Dynamically adjust subplot grid size
num_cols = len(df.columns)
rows = math.ceil(num_cols / 5)  # Adjust rows based on column count

plt.figure(figsize=(15, rows * 2))  # Adjust figure size dynamically
for i, col in enumerate(df.columns):
    plt.subplot(rows, 5, i + 1)  # Dynamically adjust grid
    sns.boxplot(y=df[col])
    plt.title(col)

plt.tight_layout()
plt.show()
```

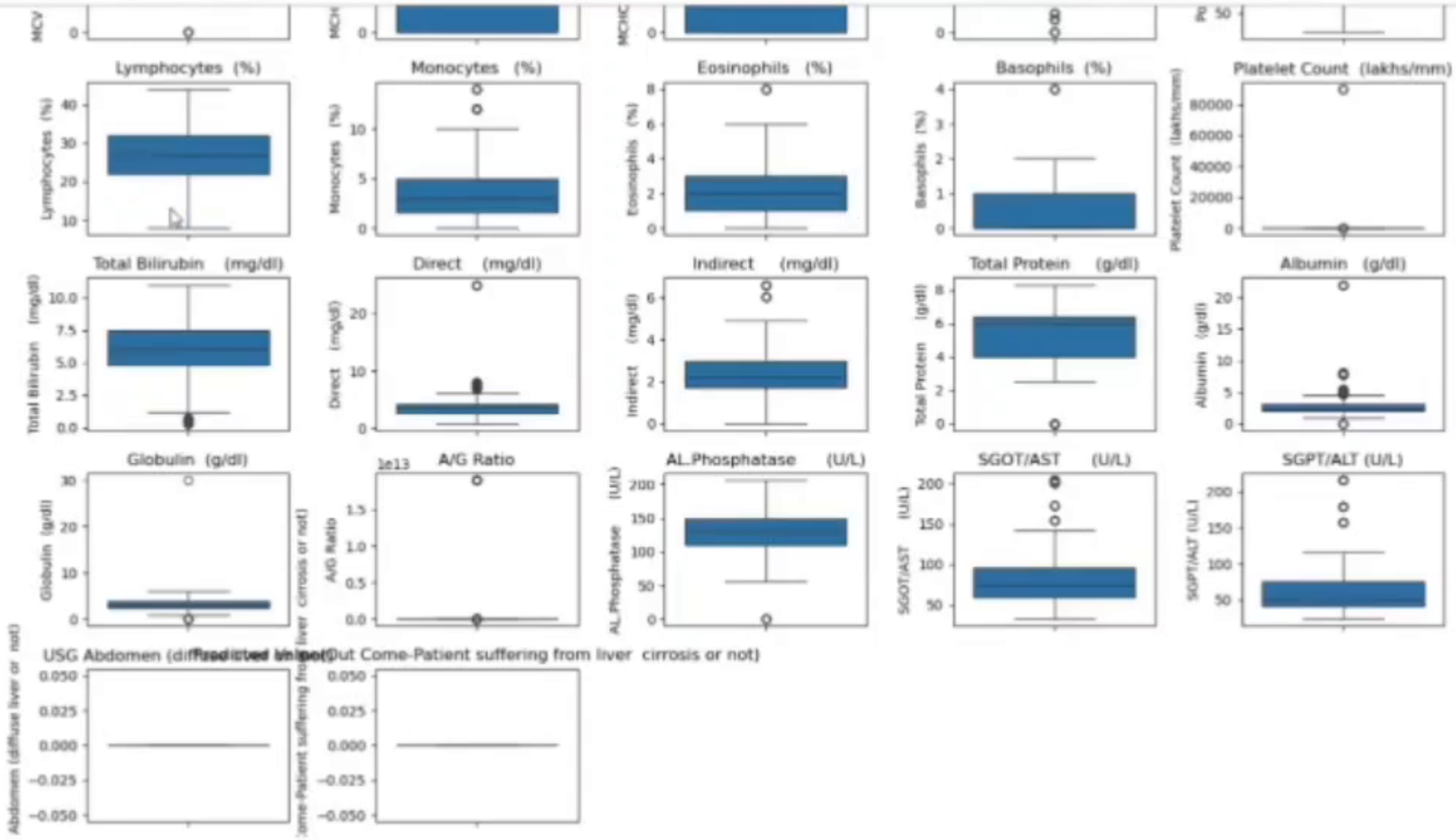
| S.NO | Age | Gender | Place(location where the patient li | ation of alcohol consumption(years) |
|---|---|---|---|---|

Search ENG IN 09:46 PM 11-03-2025

```
    sns.boxplot(y=df[col])
    plt.title(col)

plt.tight_layout()
plt.show()
```




Blood pressure
Obesity
Family history of cirrho
TCH
TG
LDL
HDL
Hemoglobin (g/dl)
PCV (%)
RBC (million cells/microliter)
MCV (femtoliters/cell)
MCH (picograms/cell)
MCHC (grams/deciliter)
Total Count
Polymorphs (%)
Lymphocytes (%)
Monocytes (%)
Eosinophils (%)
Basophils (%)
Platelet Count (lakhs/mm)
Total Bilirubin (mg/dl)
Direct (mg/dl)
Indirect (mg/dl)
Total Protein (g/dl)
Albumin (g/dl)




[27]:
```
df.columns = df.columns.str.strip()
```

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[27]:
```
df.columns = df.columns.str.strip()
```

[28]:
```
print(df.head())
```

```
   S.NO  Age  Gender  Place(location where the patient lives)  \
0     1   55     0.0                                     0.0
1     2   55     0.0                                     0.0
2     3   55     0.0                                     0.0
3     4   55     0.0                                     0.0
4     5   55     0.0                                     0.0

   Duration of alcohol consumption(years)  \
0                                      12
1                                      12
2                                      12
3                                      12
4                                      12

   Quantity of alcohol consumption (quarters/day)  Type of alcohol consumed  \
0                                               2                       0.0
1                                               2                       0.0
2                                               2                       0.0
3                                               2                       0.0
4                                               2                       0.0

   Hepatitis B infection  Hepatitis C infection  Diabetes Result  ...  \
0                    0.0                    0.0              0.0  ...
1                    0.0                    0.0              0.0  ...
2                    0.0                    0.0              0.0  ...
3                    0.0                    0.0              0.0  ...
4                    0.0                    0.0              0.0  ...

   Indirect     (mg/dl)  Total Protein     (g/dl)  Albumin   (g/dl)  \
0                   3.0                       6.0               3.0
1                   3.0                       6.0               3.0
2                   3.0                       6.0               3.0
```

Search ENG IN 09:47 PM 11-03-2025

[28]:
```
print(df.head())
```

```
   S.NO  Age  Gender  Place(location where the patient lives)  \
0     1   55     0.0                                     0.0
1     2   55     0.0                                     0.0
2     3   55     0.0                                     0.0
3     4   55     0.0                                     0.0
4     5   55     0.0                                     0.0

   Duration of alcohol consumption(years)  \
0                                      12
1                                      12
2                                      12
3                                      12
4                                      12

   Quantity of alcohol consumption (quarters/day)  Type of alcohol consumed  \
0                                               2                       0.0
1                                               2                       0.0
2                                               2                       0.0
3                                               2                       0.0
4                                               2                       0.0

   Hepatitis B infection  Hepatitis C infection  Diabetes Result  ...  \
0                    0.0                    0.0              0.0  ...
1                    0.0                    0.0              0.0  ...
2                    0.0                    0.0              0.0  ...
3                    0.0                    0.0              0.0  ...
4                    0.0                    0.0              0.0  ...

   Indirect     (mg/dl)  Total Protein     (g/dl)  Albumin   (g/dl)  \
0                   3.0                      6.0               3.0
1                   3.0                      6.0               3.0
2                   3.0                      6.0               3.0
3                   3.0                      6.0               3.0
4                   3.0                      6.0               3.0
```

| | Predicted Value(Out Come-Patient suffering from liver cirrosis or not) |
|---|---|
| 0 | 0.0 |
| 1 | 0.0 |
| 2 | 0.0 |
| 3 | 0.0 |
| 4 | 0.0 |

[5 rows x 42 columns]

[29]:
```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df.columns = df.columns.str.strip()
def remove_outliers(column):
    q1 = df[column].quantile(0.25)
    q3 = df[column].quantile(0.75)
    iqr = q3 - q1
    upper_limit = q3 + (1.5 * iqr)
    lower_limit = q1 - (1.5 * iqr)
    print(f"{column} - Lower Limit: {lower_limit}, Upper Limit: {upper_limit}")
    df[column] = np.where(df[column] > upper_limit, upper_limit,
                          np.where(df[column] < lower_limit, lower_limit, df[column]))
    sns.boxplot(x=df[column])
    plt.title(f"Boxplot of {column}")
    plt.show()
columns_to_clear = ['Eosinophils (%)', 'Basophils (%)', 'Platelet Count (lakhs/mm)']
```

[30]:
```
from sklearn.model_selection import train_test_split
```

[31]:
```
x = np.array([[1,2], [3,4], [5,6], [7,8], [9,10]])
y = np.array([0, 1, 0, 1, 0])
x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.2, random_state=42)
```

[32]:
```
x_train
```

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[29]:
```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df.columns = df.columns.str.strip()
def remove_outliers(column):
    q1 = df[column].quantile(0.25)
    q3 = df[column].quantile(0.75)
    iqr = q3 - q1
    upper_limit = q3 + (1.5 * iqr)
    lower_limit = q1 - (1.5 * iqr)
    print(f"{column} - Lower Limit: {lower_limit}, Upper Limit: {upper_limit}")
    df[column] = np.where(df[column] > upper_limit, upper_limit,
                          np.where(df[column] < lower_limit, lower_limit, df[column]))
    sns.boxplot(x=df[column])
    plt.title(f"Boxplot of {column}")
    plt.show()
columns_to_clear = ['Eosinophils (%)', 'Basophils (%)', 'Platelet Count (lakhs/mm)']
```

[30]:
```
from sklearn.model_selection import train_test_split
```

[31]:
```
x = np.array([[1,2], [3,4], [5,6], [7,8], [9,10]])
y = np.array([0, 1, 0, 1, 0])
x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.2, random_state=42)
```

[32]:
```
x_train
```

[32]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[33]:
```
x_test
```

[33]:
```
array([[3, 4]])
```

Search ENG IN 09:47 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[30]:
```
from sklearn.model_selection import train_test_split
```

[31]:
```
x = np.array([[1,2], [3,4], [5,6], [7,8], [9,10]])
y = np.array([0, 1, 0, 1, 0])
x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.2, random_state=42)
```

[32]:
```
x_train
```

[32]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[33]:
```
x_test
```

[33]:
```
array([[3, 4]])
```

[34]:
```
y_train
```

[34]:
```
array([0, 0, 0, 1])
```

[35]:
```
y_test
```

[35]:
```
array([1])
```

[36]:
```
df.describe()
```

[36]:

| | S.NO | Age | Gender | Place(location where the patient lives) | Duration of alcohol consumption(years) | Quantity of alcohol consumption (quarters/day) | Type of alcohol consumed | Hepatitis B infection | Hepatitis C infection | Diabetes Result | ... | Indirect (mg/dl) | Total Protein (g/dl) | Albumin (g/dl) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

Search ENG IN 09:48 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[32]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[33]:
```
x_test
```

[33]:
```
array([[3, 4]])
```

[34]:
```
y_train
```

[34]:
```
array([0, 0, 0, 1])
```

[35]:
```
y_test
```

[35]:
```
array([1])
```

[36]:
```
df.describe()
```

[36]:

| | S.NO | Age | Gender | Place(location where the patient lives) | Duration of alcohol consumption(years) | Quantity of alcohol consumption (quarters/day) | Type of alcohol consumed | Hepatitis B infection | Hepatitis C infection | Diabetes Result | ... | Indirect (mg/dl) | Total Protein (g/dl) | Albumin (g/dl) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| : | 950.000000 | 950.000000 | 950.0 | 950.0 | 950.000000 | 950.000000 | 950.0 | 950.0 | 950.0 | 950.0 | ... | 950.000000 | 950.000000 | 950.000000 |
| | 475.500000 | 50.632632 | 0.0 | 0.0 | 20.606316 | 5.158947 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 2.315263 | 5.231368 | 2.770632 |
| | 274.385677 | 8.808272 | 0.0 | 0.0 | 7.980664 | 22.908785 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.206895 | 1.869810 | 2.205275 |
| | 1.000000 | 32.000000 | 0.0 | 0.0 | 4.000000 | 1.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.000000 | 0.000000 | 0.000000 |
| | 238.250000 | 44.000000 | 0.0 | 0.0 | 15.000000 | 2.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.700000 | 4.000000 | 2.000000 |

Search

ENG IN

09:48 PM 11-03-2025

[33]: array([[3, 4]])

[34]:
```
y_train
```

[34]:
```
array([0, 0, 0, 1])
```

[35]:
```
y_test
```

[35]:
```
array([1])
```

[36]:
```
df.describe()
```

[36]:

| | S.NO | Age | Gender | Place(location where the patient lives) | Duration of alcohol consumption(years) | Quantity of alcohol consumption (quarters/day) | Type of alcohol consumed | Hepatitis B infection | Hepatitis C infection | Diabetes Result | ... | Indirect (mg/dl) | Total Protein (g/dl) | Albumin (g/dl) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| : | 950.000000 | 950.000000 | 950.0 | 950.0 | 950.000000 | 950.000000 | 950.0 | 950.0 | 950.0 | 950.0 | ... | 950.000000 | 950.000000 | 950.000000 |
| | 475.500000 | 50.632632 | 0.0 | 0.0 | 20.606316 | 5.158947 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 2.315263 | 5.231368 | 2.770632 |
| | 274.385677 | 8.808272 | 0.0 | 0.0 | 7.980664 | 22.908785 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.206895 | 1.869810 | 2.205275 |
| | 1.000000 | 32.000000 | 0.0 | 0.0 | 4.000000 | 1.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.000000 | 0.000000 | 0.000000 |
| | 238.250000 | 44.000000 | 0.0 | 0.0 | 15.000000 | 2.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.700000 | 4.000000 | 2.000000 |
| | 475.500000 | 50.000000 | 0.0 | 0.0 | 20.000000 | 2.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 2.200000 | 6.000000 | 2.500000 |
| | 712.750000 | 57.000000 | 0.0 | 0.0 | 26.000000 | 3.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 3.000000 | 6.400000 | 3.000000 |
| | 950.000000 | 80.000000 | 0.0 | 0.0 | 45.000000 | 180.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 6.600000 | 8.300000 | 22.000000 |

× 42 columns

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[35]:
```
y_test
```

[35]:
```
array([1])
```

[36]:
```
df.describe()
```

[36]:

| | S.NO | Age | Gender | Place(location where the patient lives) | Duration of alcohol consumption(years) | Quantity of alcohol consumption (quarters/day) | Type of alcohol consumed | Hepatitis B infection | Hepatitis C infection | Diabetes Result | ... | Indirect (mg/dl) | Total Protein (g/dl) | Albumin (g/dl) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 950.000000 | 950.000000 | 950.0 | 950.0 | 950.000000 | 950.000000 | 950.0 | 950.0 | 950.0 | 950.0 | ... | 950.000000 | 950.000000 | 950.000000 |
| | 475.500000 | 50.632632 | 0.0 | 0.0 | 20.606316 | 5.158947 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 2.315263 | 5.231368 | 2.770632 |
| | 274.385677 | 8.808272 | 0.0 | 0.0 | 7.980664 | 22.908785 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.206895 | 1.869810 | 2.205275 |
| | 1.000000 | 32.000000 | 0.0 | 0.0 | 4.000000 | 1.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.000000 | 0.000000 | 0.000000 |
| | 238.250000 | 44.000000 | 0.0 | 0.0 | 15.000000 | 2.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.700000 | 4.000000 | 2.000000 |
| | 475.500000 | 50.000000 | 0.0 | 0.0 | 20.000000 | 2.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 2.200000 | 6.000000 | 2.500000 |
| | 712.750000 | 57.000000 | 0.0 | 0.0 | 26.000000 | 3.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 3.000000 | 6.400000 | 3.000000 |
| | 950.000000 | 80.000000 | 0.0 | 0.0 | 45.000000 | 180.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 6.600000 | 8.300000 | 22.000000 |

× 42 columns

[64]:
```
sns.boxplot(x='Age',y='Diabetes Result',data=df,hue='Gender')
plt.title('Gender vs Diabetes Result', color='red', size=20)
```

Search

ENG IN

09:48 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

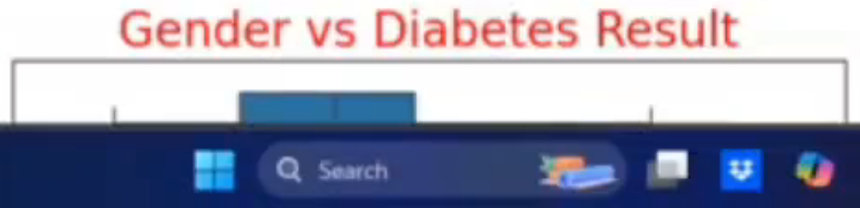
Code JupyterLab Python 3 (ipykernel)

[36]:

| f alcohol n(years) | Quantity of alcohol consumption (quarters/day) | Type of alcohol consumed | Hepatitis B infection | Hepatitis C infection | Diabetes Result | ... | Indirect (mg/dl) | Total Protein (g/dl) | Albumin (g/dl) | Globulin (g/dl) | A/G Ratio | AL.Phosphatase (U/L) | SGOT/AST (U/L) | SGPT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.000000 | 950.000000 | 950.0 | 950.0 | 950.0 | 950.0 | ... | 950.000000 | 950.000000 | 950.000000 | 950.000000 | 9.500000e+02 | 950.000000 | 950.000000 | 950.00 |
| 3.606316 | 5.158947 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 2.315263 | 5.231368 | 2.770632 | 3.146000 | 6.400034e+11 | 131.129474 | 81.794737 | 61.56 |
| 7.980664 | 22.908785 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.206895 | 1.869810 | 2.205275 | 1.433801 | 3.429709e+12 | 30.485448 | 31.106923 | 31.30 |
| 4.000000 | 1.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000e+00 | 0.000000 | 32.000000 | 23.00 |
| 5.000000 | 2.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 1.700000 | 4.000000 | 2.000000 | 2.500000 | 0.000000e+00 | 110.000000 | 59.000000 | 41.00 |
| 0.000000 | 2.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 2.200000 | 6.000000 | 2.500000 | 3.100000 | 7.200000e-01 | 130.000000 | 74.000000 | 49.00 |
| 5.000000 | 3.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 3.000000 | 6.400000 | 3.000000 | 4.000000 | 2.200000e+00 | 150.000000 | 96.000000 | 76.00 |
| 5.000000 | 180.000000 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 6.600000 | 8.300000 | 22.000000 | 30.000000 | 1.900010e+13 | 206.000000 | 204.000000 | 216.00 |

[64]:

```
sns.boxplot(x='Age',y='Diabetes Result',data=df,hue='Gender')
plt.title('Gender vs Diabetes Result',color='red',size=20)
plt.show()
```



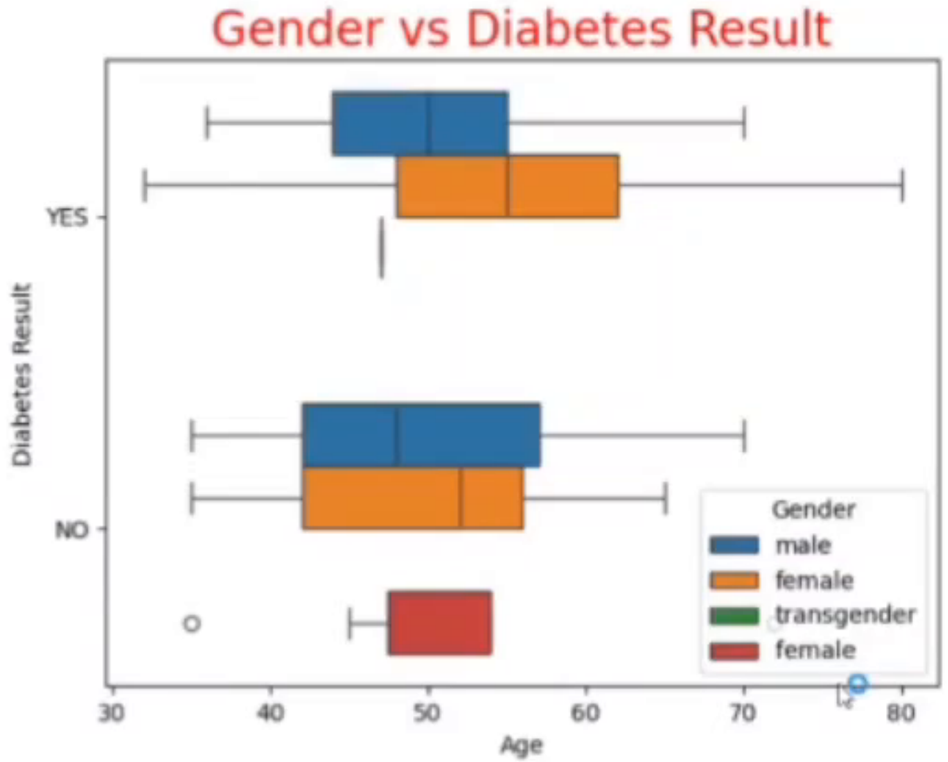

Search ENG IN 09:48 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[64]:
```
sns.boxplot(x='Age',y='Diabetes Result',data=df,hue='Gender')
plt.title('Gender vs Diabetes Result',color='red',size=20)
plt.show()
```

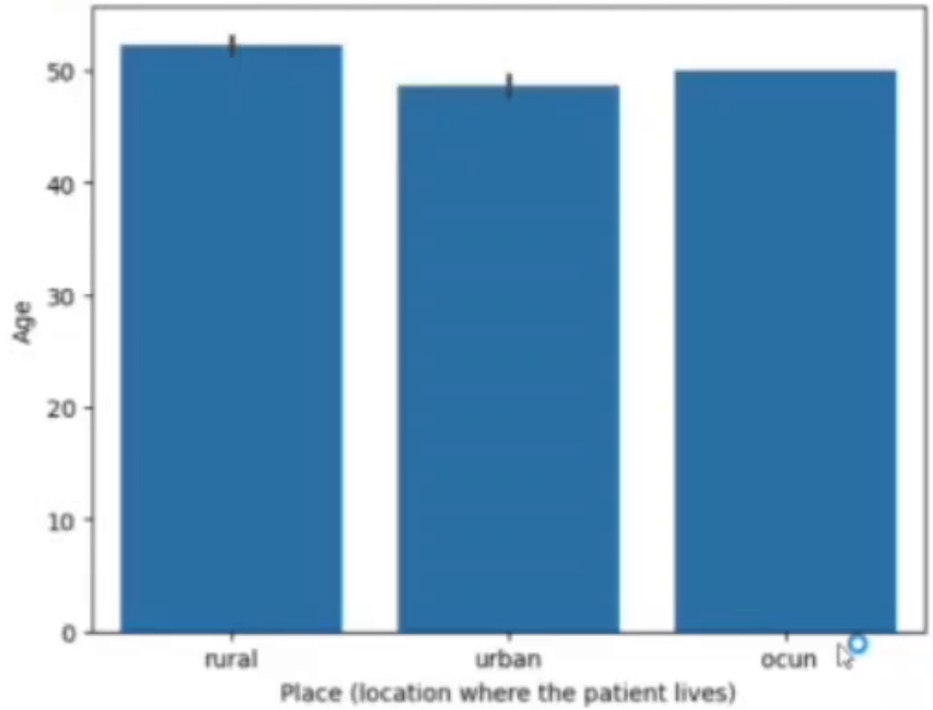



[66]:
```
sns.barplot(x=df['Place(location where the patient lives)'],y=df['Age'])
plt.xlabel('Place (location where the patient lives)')
plt.ylabel('Age')
```

Search ENG IN 09:49 PM 11-03-2025

[66]:
```
sns.barplot(x=df['Place(location where the patient lives)'],y=df['Age'])
plt.xlabel('Place (location where the patient lives)')
plt.ylabel('Age')
```

[66]:
```
Text(0, 0.5, 'Age')
```

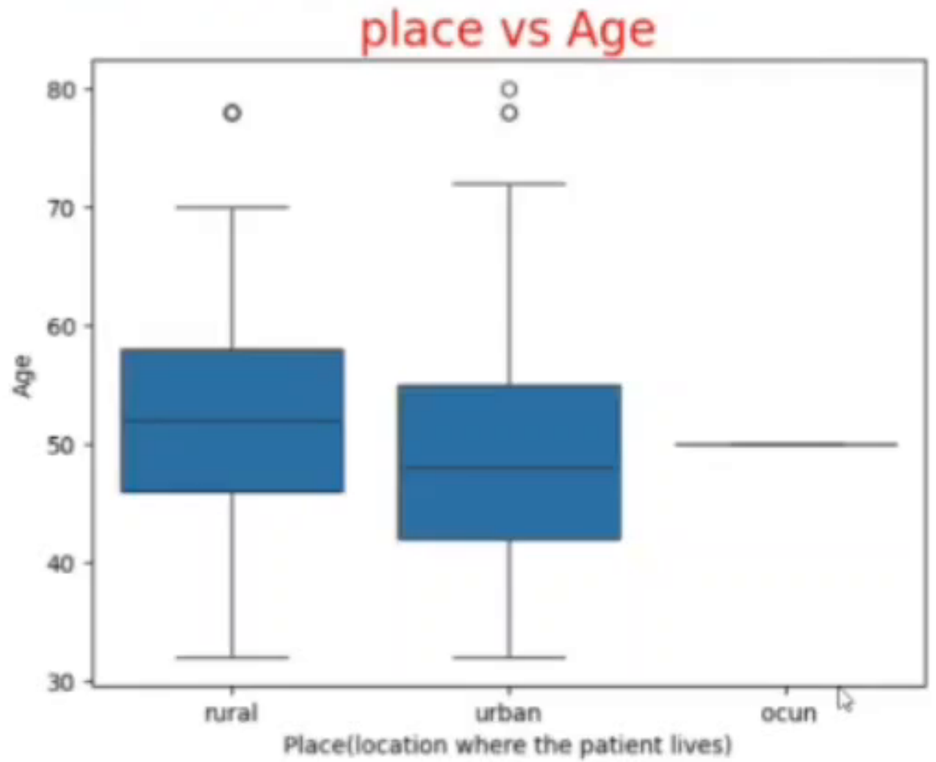



[65]:
```
sns.boxplot(x='Place(location where the patient lives)',y='Age',data=df)
plt.title('place vs Age',color='red',size=20)
```

[65]:
```
Text(0.5, 1.0, 'place vs Age')
```

[65]:
```
sns.boxplot(x='Place(location where the patient lives)',y='Age',data=df)
plt.title('place vs Age',color='red',size=20)
```

[65]:
```
Text(0.5, 1.0, 'place vs Age')
```



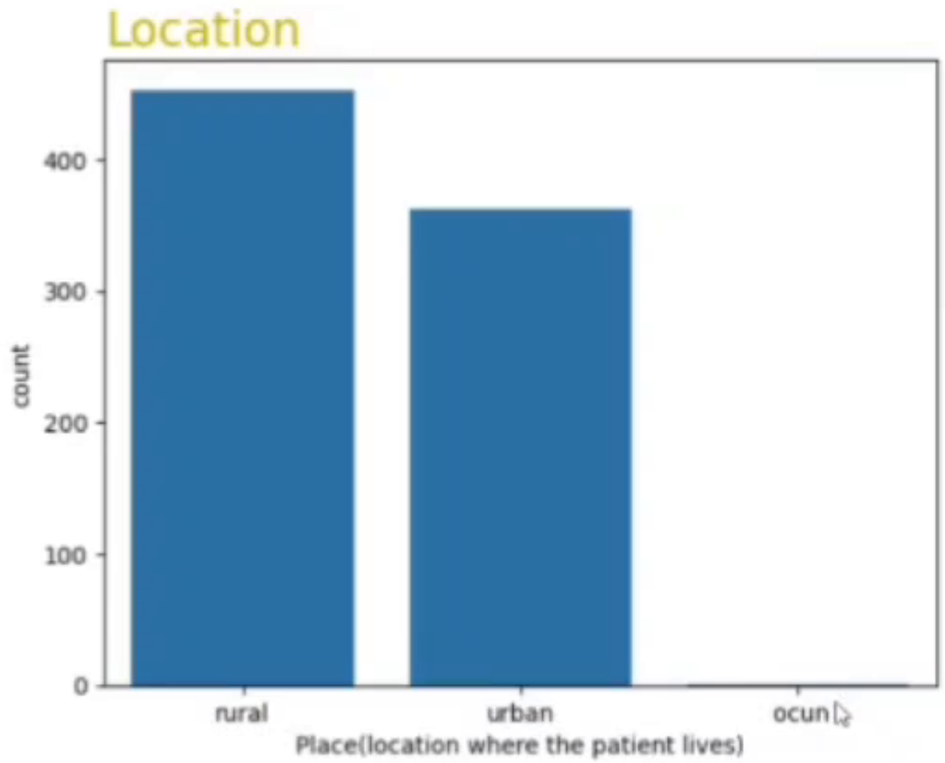

[67]:
```
sns.countplot(data=df,x='Place(location where the patient lives)')
plt.title("Location",color='y',size=20,loc='left')
plt.show()
```

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[67]:
```
sns.countplot(data=df,x='Place(location where the patient lives)')
plt.title("Location",color='y',size=20,loc='left')
plt.show()
```




[42]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import math  # For dynamic row calculation
```

Search ENG IN 09:49 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

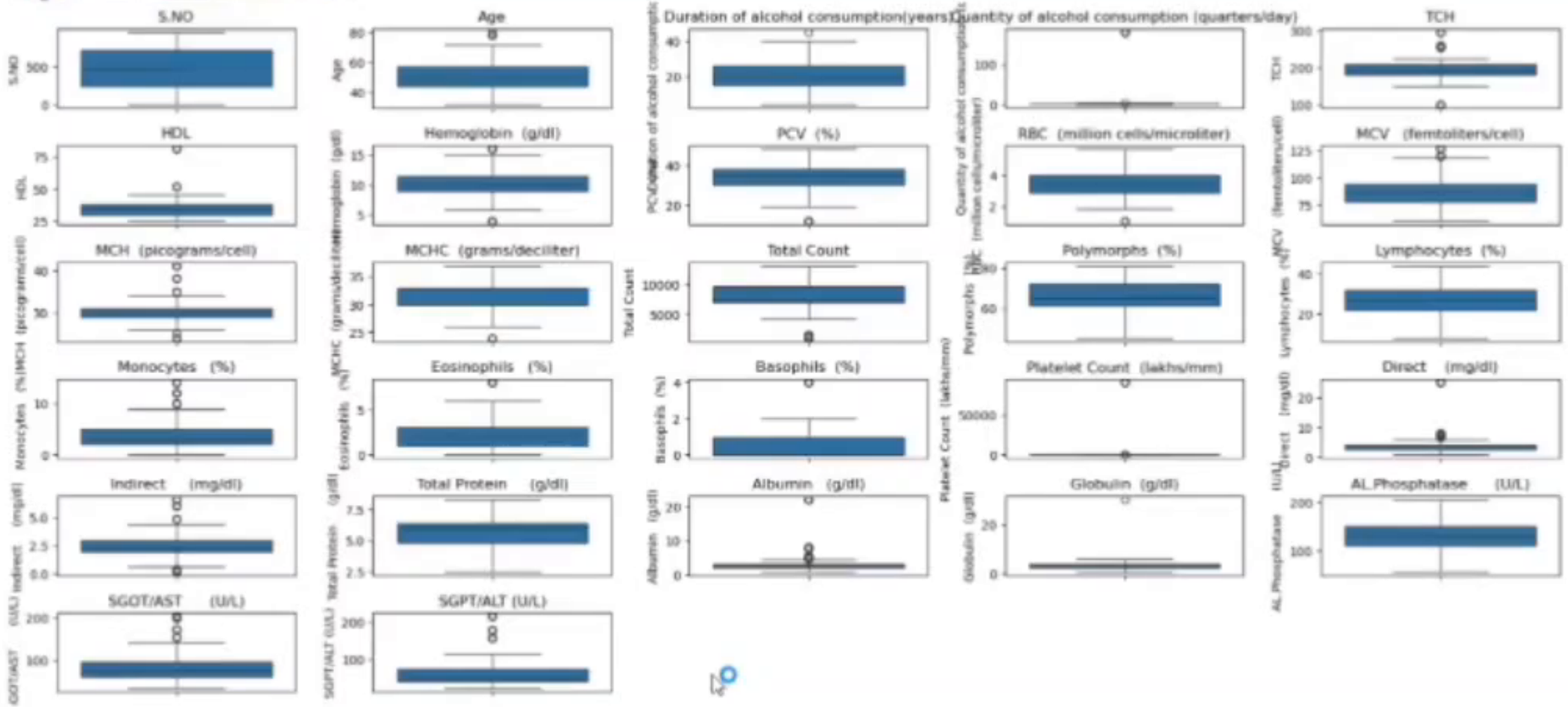
File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[42]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import math  # For dynamic row calculation

# Load dataset (if applicable)
# df = pd.read_csv("your_file.csv")

# Function to clean numeric values
def clean_numeric(column):
    return pd.to_numeric(column.astype(str)
                         .str.replace(r'[^0-9.]', '', regex=True)  # Keep only numbers & periods
                         .str.replace(r'\.$', '', regex=True),  # Remove trailing periods
                         errors='coerce')

# Apply cleaning to all columns
df = df.apply(clean_numeric)

# Fill missing values
df.fillna(0, inplace=True)

#  * Dynamically adjust subplot grid size
num_cols = len(df.columns)
rows = math.ceil(num_cols / 5)  # Adjust rows based on column count


df=pd.read_excel(r'C:\Users\uday\Data\HealthCareData.xlsx')
plt.figure(figsize=(16, 10))
num_cols =5
num_rows =-(-len(df.columns) // num_cols)
fig, axes = plt.subplots(num_rows, num_cols, figsize=(18, 12))
axes = axes.flatten()
for i, col in enumerate(df.select_dtypes(include=['number']).columns):
    sns.boxplot(y=df[col],ax=axes[i])
```

Search ENG IN 09:49 PM 11-03-2025

```
<Figure size 1600x1000 with 0 Axes>
```




[43]:
```
plt.figure(figsize=(15,10))
sns.heatmap(df.select_dtypes(include['number']).corr(), annot=True)
plt.show()
```

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[43]:
```
plt.figure(figsize=(15,10))
sns.heatmap(df.select_dtypes(includes['number']).corr(), annot=True)
plt.show()
```


Search

ENG IN

09:50 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

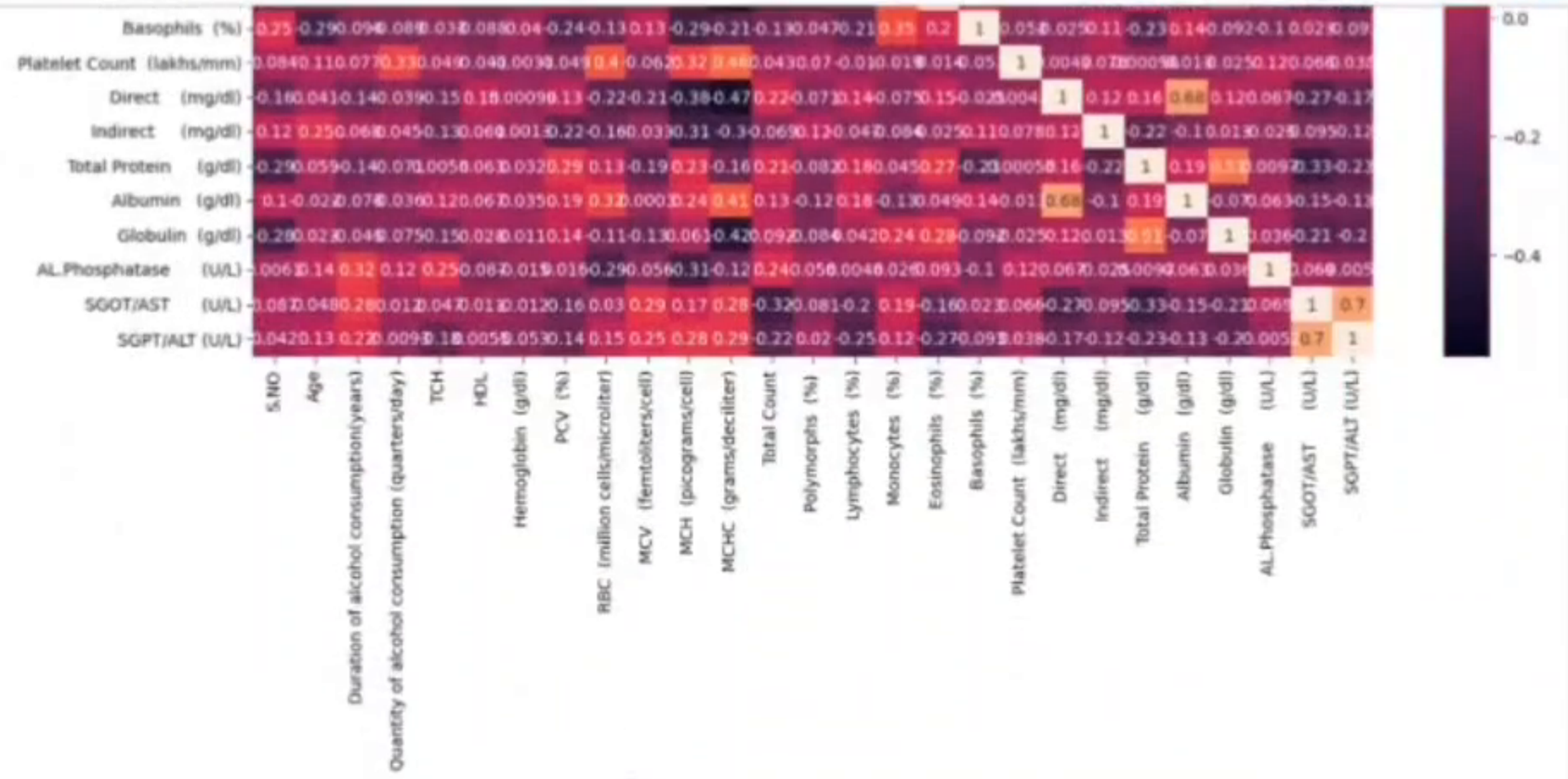
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)




[44]:
```
from sklearn.model_selection import train_test_split
```

[45]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

[46]:
```
x_train
```

[46]:
```
array([[ 9, 10],
```

Search

ENG IN

09:50 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[44]:
```
from sklearn.model_selection import train_test_split
```

[45]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

[46]:
```
x_train
```

[46]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[47]:
```
x_test
```

[47]:
```
array([[3, 4]])
```

[48]:
```
y_train
```

[48]:
```
array([0, 0, 0, 1])
```

[49]:
```
y_test
```

[49]:
```
array([1])
```

[50]:
```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train, y_train)
```

[50]:
```
GaussianNB()
```

[51]:
```
x_train
```

[51]:
```
array([[ 9, 10],
```

Search

ENG IN

09:50 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[47]:
```
x_test
```

[47]:
```
array([[3, 4]])
```

[48]:
```
y_train
```

[48]:
```
array([0, 0, 0, 1])
```

[49]:
```
y_test
```

[49]:
```
array([1])
```

[50]:
```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train, y_train)
```

[50]:
```
GaussianNB()
```

[51]:
```
x_train
```

[51]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[52]:
```
y_train
```

[52]:
```
array([0, 0, 0, 1])
```

[53]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
```

Search

ENG IN

09:50 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[49]:
```
y_test
```

[49]:
```
array([1])
```

[50]:
```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train, y_train)
```

[50]:
```
GaussianNB()
```

[51]:
```
x_train
```

[51]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[52]:
```
y_train
```

[52]:
```
array([0, 0, 0, 1])
```

[53]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
```

[53]:
```
RandomForestClassifier()
```

[54]:
```
x_train
```

[54]:
```
array([[ 9, 10],
```

Search ENG IN 09:51 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[51]:
```
x_train
```

[51]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[52]:
```
y_train
```

[52]:
```
array([0, 0, 0, 1])
```

[53]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
```

[53]:
```
RandomForestClassifier
RandomForestClassifier()
```

[54]:
```
x_train
```

[54]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[55]:
```
y_train
```

[55]:
```
array([0, 0, 0, 1])
```

[56]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(x_train, y_train)
```

Search

ENG IN

09:51 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[53]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
```

[53]:
```
RandomForestClassifier()
```

[54]:
```
x_train
```

[54]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[55]:
```
y_train
```

[55]:
```
array([0, 0, 0, 1])
```

[56]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(x_train, y_train)
```

[56]:
```
KNeighborsClassifier()
```

[57]:
```
print("x_Train", x_train)
print("y_Train", y_train)
```

```
x_Train [[ 9 10]
 [ 5  6]
 [ 1  2]
 [ 7  8]]
y_Train [0 0 0 1]
```

Search ENG IN 09:51 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

[53]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
```

[53]:
```
RandomForestClassifier()
```

[54]:
```
x_train
```

[54]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[55]:
```
y_train
```

[55]:
```
array([0, 0, 0, 1])
```

[56]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(x_train, y_train)
```

[56]:
```
KNeighborsClassifier()
```

[57]:
```
print("x_Train", x_train)
print("y_Train", y_train)
```

```
x_Train [[ 9 10]
 [ 5  6]
 [ 1  2]
 [ 7  8]]
y_Train [0 0 0 1]
```

Search

ENG IN

09:51 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[55]:
```
y_train
```

[55]:
```
array([0, 0, 0, 1])
```

[56]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(x_train, y_train)
```

[56]:
```
KNeighborsClassifier
KNeighborsClassifier()
```

[57]:
```
print("x_Train", x_train)
print("y_Train", y_train)
```

```
x_Train [[ 9 10]
 [ 5  6]
 [ 1  2]
 [ 7  8]]
y_Train [0 0 0 1]
```

[58]:
```
from sklearn.model_selection import train_test_split, GridSearchCV, RandomizedSearchCV
```

[59]:
```
  k = np.random.randint(1,50,60)
```

[60]:
```
params = {'n_neighbors' : k}
```

[61]:
```
from sklearn.model_selection import RandomizedSearchCV
random_search = RandomizedSearchCV(knn,params, n_iter=5, cv=2, n_jobs=-1, verbose=0)
random_search.fit(x_train,y_train)
```

```
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_split.py:776: UserWarning: The least populated class in y has only 1 members, which is less than
n_splits=2.
  warnings.warn(
```

Search ENG IN 09:51 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[57]:
```
print("x_Train", x_train)
print("y_Train", y_train)
```

```
x_Train [[ 9 10]
 [ 5  6]
 [ 1  2]
 [ 7  8]]
y_Train [0 0 0 1]
```

[58]:
```
from sklearn.model_selection import train_test_split, GridSearchCV, RandomizedSearchCV
```

[59]:
```
  k = np.random.randint(1,50,60)
```

[60]:
```
params = {'n_neighbors' : k}
```

[61]:
```
from sklearn.model_selection import RandomizedSearchCV
random_search = RandomizedSearchCV(knn,params, n_iter=5, cv=2, n_jobs=-1, verbose=0)
random_search.fit(x_train,y_train)
```

```
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_split.py:776: UserWarning: The least populated class in y has only 1 members, which is less than n_splits=2.
  warnings.warn(
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_search.py:1102: UserWarning: One or more of the test scores are non-finite: [nan nan nan nan nan]
  warnings.warn(
```

[61]:

▸ RandomizedSearchCV

▸ best_estimator_: KNeighborsClassifier

▸ KNeighborsClassifier

[63]:
```
from sklearn.datasets import load_iris
```

Search ENG IN 09:51 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[59]:
```
k = np.random.randint(1,50,60)
```

[60]:
```
params = {'n_neighbors' : k}
```

[61]:
```
from sklearn.model_selection import RandomizedSearchCV
random_search = RandomizedSearchCV(knn,params, n_iter=5, cv=2, n_jobs=-1, verbose=0)
random_search.fit(x_train,y_train)
```

```
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_split.py:776: UserWarning: The least populated class in y has only 1 members, which is less than n_splits=2.
  warnings.warn(
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_search.py:1102: UserWarning: One or more of the test scores are non-finite: [nan nan nan nan nan]
  warnings.warn(
```

[61]:

▸ RandomizedSearchCV

▸ best_estimator_: KNeighborsClassifier

▸ KNeighborsClassifier

[63]:
```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

# Load the dataset
data = load_iris()
X, y = data.data, data.target  # ✅ Use X (not x)

# Split dataset into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  # ✅ Fix x -> X

# Define hyperparameter search space
K_values = list(range(1, min(20, len(x_train))))
```

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[61]:
```
from sklearn.model_selection import RandomizedSearchCV
random_search = RandomizedSearchCV(knn,params, n_iter=5, cv=2, n_jobs=-1, verbose=0)
random_search.fit(x_train,y_train)
```

```
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_split.py:776: UserWarning: The least populated class in y has only 1 members, which is less than n_splits=2.
  warnings.warn(
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_search.py:1102: UserWarning: One or more of the test scores are non-finite: [nan nan nan nan nan]
  warnings.warn(
```

[61]:

▸ RandomizedSearchCV

▸ best_estimator_: KNeighborsClassifier

▸ KNeighborsClassifier

[63]:
```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

# Load the dataset
data = load_iris()
X, y = data.data, data.target  # ✅ Use X (not x)

# Split dataset into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  # ✅ Fix x -> X

# Define hyperparameter search space
K_values = list(range(1, min(20, len(x_train))))
params = {'n_neighbors': K_values}

# Train KNN model
knn = KNeighborsClassifier(n_neighbors=5)
```

25°C Mostly clear

Search

ENG IN

09:52 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab

Python 3 (ipykernel)

```
C:\Users\uday\Lib\site-packages\sklearn\model_selection\_search.py:1102: UserWarning: One or more of the test scores are non-finite: [nan nan nan nan na
n]
  warnings.warn(
```

[61]:

RandomizedSearchCV

best_estimator_: KNeighborsClassifier

KNeighborsClassifier

[63]:

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

# Load the dataset
data = load_iris()
X, y = data.data, data.target  # ✅ Use X (not x)

# Split dataset into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  # ✅ Fix x -> X

# Define hyperparameter search space
K_values = list(range(1, min(20, len(x_train))))
params = {'n_neighbors': K_values}

# Train KNN model
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(x_train, y_train)

# Evaluate model
print("Train Score:", knn.score(x_train, y_train))
print("Test Score:", knn.score(x_test, y_test))
```

25°C Mostly clear

Search

ENG IN

09:52 PM 11-03-2025

▸ KNeighborsClassifier

[63]:
```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

# Load the dataset
data = load_iris()
X, y = data.data, data.target  # ✅ Use X (not x)

# Split dataset into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  # ✅ Fix x -> X

# Define hyperparameter search space
K_values = list(range(1, min(20, len(x_train))))
params = {'n_neighbors': K_values}

# Train KNN model
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(x_train, y_train)

# Evaluate model
print("Train Score:", knn.score(x_train, y_train))
print("Test Score:", knn.score(x_test, y_test))
```

```
Train Score: 0.9666666666666667
Test Score: 1.0
```

[ ]:

[ ]:

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[63]:
```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

# Load the dataset
data = load_iris()
X, y = data.data, data.target  # ✅ Use X (not x)

# Split dataset into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)  # ✅ Fix x -> X

# Define hyperparameter search space
K_values = list(range(1, min(20, len(x_train))))
params = {'n_neighbors': K_values}

# Train KNN model
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(x_train, y_train)

# Evaluate model
print("Train Score:", knn.score(x_train, y_train))
print("Test Score:", knn.score(x_test, y_test))
```

```
Train Score: 0.9666666666666667
Test Score: 1.0
```

[ ]:

[ ]:

25°C Mostly clear

Search

ENG IN

09:52 PM 11-03-2025

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[51]:
```
x_train
```

[51]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[52]:
```
y_train
```

[52]:
```
array([0, 0, 0, 1])
```

[53]:
```
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier()
rf.fit(x_train, y_train)
```

[53]:
```
RandomForestClassifier
RandomForestClassifier()
```

[54]:
```
x_train
```

[54]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[55]:
```
y_train
```

[55]:
```
array([0, 0, 0, 1])
```

[56]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()
knn.fit(x_train, y_train)
```

[56]:
```
KNeighborsClassifier
```

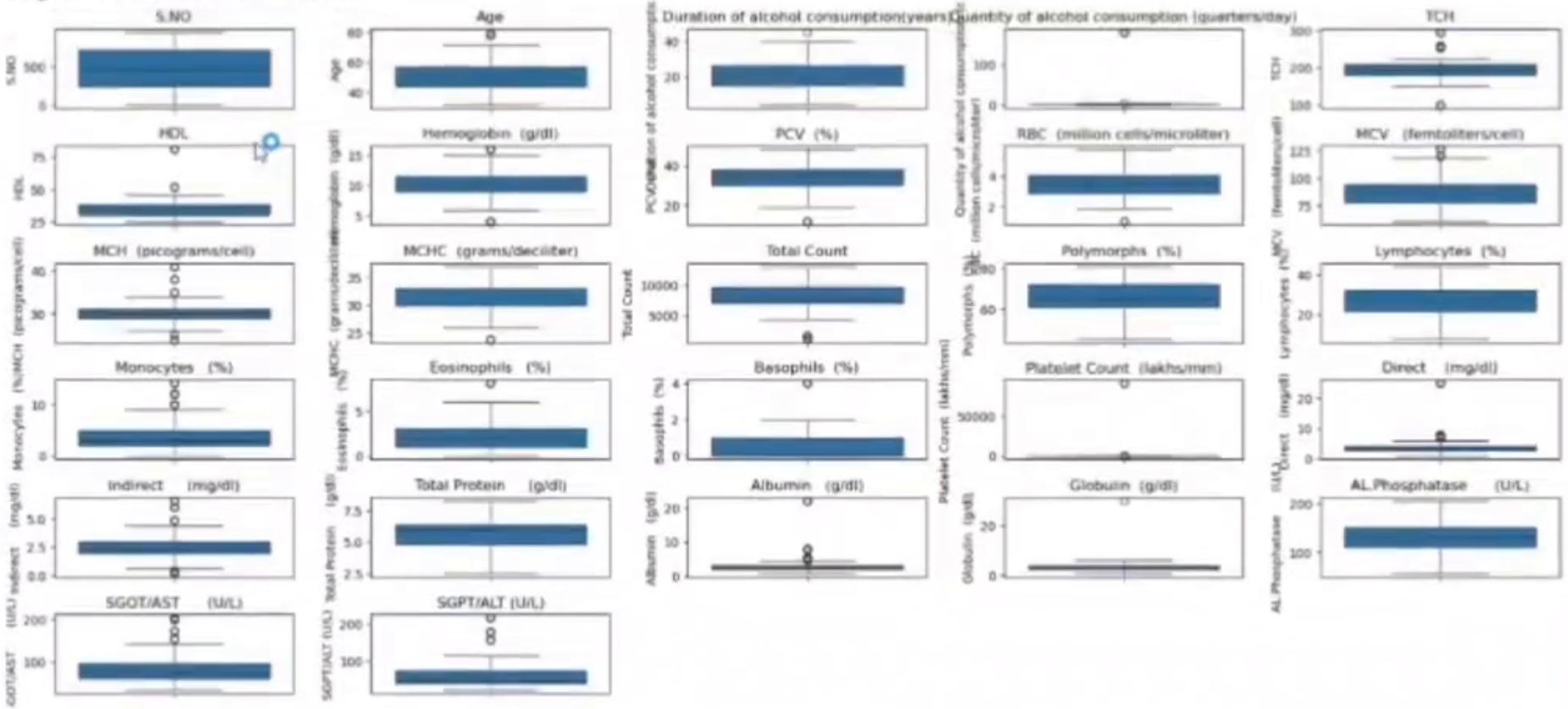
Breaking news Eyesore, public...

Search

ENG IN 09:53 PM 11-03-2025

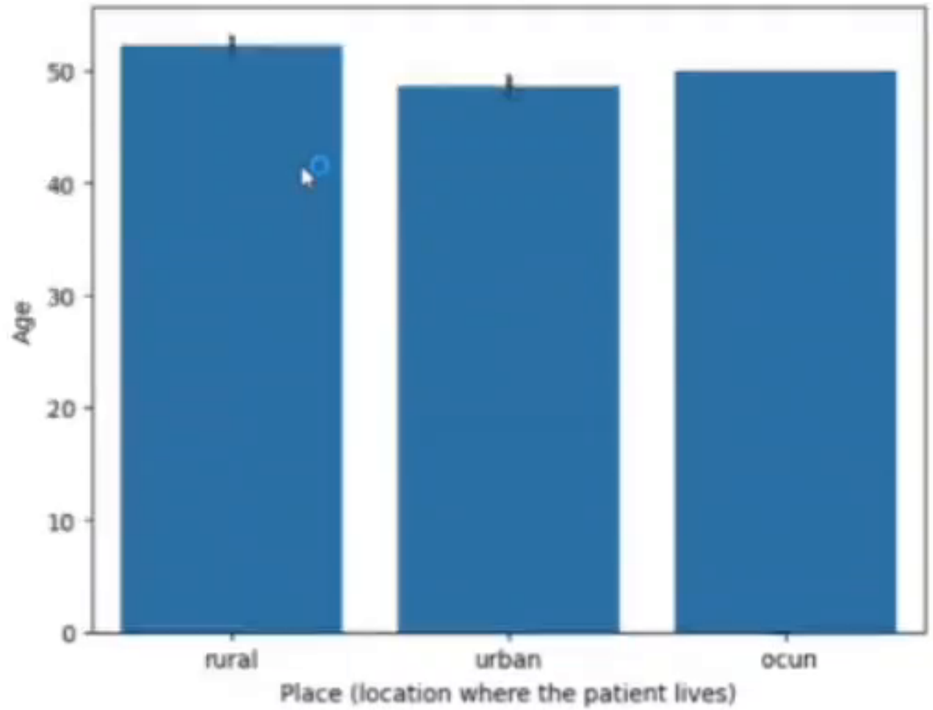
```
plt.show()
```

```
<Figure size 1600x1000 with 0 Axes>
```



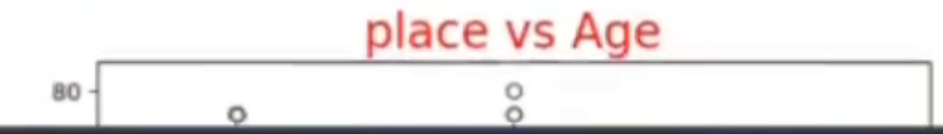

[41]:
```
plt.figure(figsize=(15,10))
sns.heatmap(df.select_dtypes(include=['number']).corr(), annot=True)
plt.show()
```

[56]:
```
Text(0, 0.5, 'Age')
```




[65]:
```
sns.boxplot(x='Place(location where the patient lives)',y='Age',data=df)
plt.title('place vs Age',color='red',size=20)
```

[65]:
```
Text(0.5, 1.0, 'place vs Age')
```



```
3                                                0.0
4                                                0.0

[5 rows x 42 columns]
```

[29]:
```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df.columns = df.columns.str.strip()
def remove_outliers(column):
    q1 = df[column].quantile(0.25)
    q3 = df[column].quantile(0.75)
    iqr = q3 - q1
    upper_limit = q3 + (1.5 * iqr)
    lower_limit = q1 - (1.5 * iqr)
    print(f"{column} - Lower Limit: {lower_limit}, Upper Limit: {upper_limit}")
    df[column] = np.where(df[column] > upper_limit, upper_limit,
                          np.where(df[column] < lower_limit, lower_limit, df[column]))
    sns.boxplot(x=df[column])
    plt.title(f"Boxplot of {column}")
    plt.show()
columns_to_clear = ['Eosinophils (%)', 'Basophils (%)', 'Platelet Count (lakhs/mm)']
```

[30]:
```
from sklearn.model_selection import train_test_split
```

[31]:
```
x = np.array([[1,2], [3,4], [5,6], [7,8], [9,10]])
y = np.array([0, 1, 0, 1, 0])
x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.2, random_state=42)
```

[32]:
```
x_train
```

[32]:
```
array([[ 9, 10],
       [ 5,  6],
       [ 1,  2],
       [ 7,  8]])
```

[20]:

| | |
|---|---|
| S.NO | 0 |
| Age | 0 |
| Gender | 0 |
| Place(location where the patient lives) | 134 |
| Duration of alcohol consumption(years) | 0 |
| Quantity of alcohol consumption (quarters/day) | 0 |
| Type of alcohol consumed | 0 |
| Hepatitis B infection | 0 |
| Hepatitis C infection | 0 |
| Diabetes Result | 0 |
| Blood pressure (mmhg) | 0 |
| Obesity | 0 |
| family history of cirrhosis/ hereditary | 0 |
| TCH | 359 |
| TG | 359 |
| LDL | 359 |
| HDL | 368 |
| Hemoglobin (g/dl) | 0 |
| PCV (%) | 30 |
| RBC (million cells/microliter) | 552 |
| MCV (femtoliters/cell) | 9 |
| MCH (picograms/cell) | 658 |
| MCHC (grams/deciliter) | 672 |
| Total Count | 10 |
| Polymorphs (%) | 0 |
| Lymphocytes (%) | 0 |
| Monocytes (%) | 9 |
| Eosinophils (%) | 8 |
| Basophils (%) | 49 |
| Platelet Count (lakhs/mm) | 0 |
| Total Bilirubin (mg/dl) | 0 |
| Direct (mg/dl) | 0 |
| Indirect (mg/dl) | 55 |
| Total Protein (g/dl) | 61 |
| Albumin (g/dl) | 9 |
| Globulin (g/dl) | 29 |

jupyter project Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[14]:
```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import pickle as pkl
import numpy as np
from sklearn import svm
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KN
eighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression, LogisticRegressionCV, RidgeClassifier
from sklearn.model_selection import train_test_split, Gri
dSearchCV
from xgboost import XGBClassifier
from sklearn.preprocessing import Normalizer
from sklearn.metrics import accuracy_score, f1_score, recall_score, precision_score, confusion_matrix
```

[13]:
```
#Reading CSV file
import pandas as pd
df = pd.read_excel(r'C:\Users\uday\Data\HealthCareData.xlsx')
df.head()
```

[13]:

| | S.NO | Age | Gender | Place(location where the patient lives) | Duration of alcohol consumption(years) | Quantity of alcohol consumption (quarters/day) | Type of alcohol consumed | Hepatitis B infection | Hepatitis C infection | Diabetes Result | ... | Indirect (mg/dl) | Total Protein (g/dl) | Albumin (g/dl) | Globulin (g/dl) | A/G Ratio | A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 55 | male | rural | 12 | 2 | branded liquor | negative | negative | YES | ... | 3.0 | 6.0 | 3.0 | 4.0 | 0.75 | |

Breaking news Eyesore, public...

Search

ENG IN 09:54 PM 11-03-2025