

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv('pima-indians-diabetes - pima-indians-diabetes.csv')
df
```

	# 1. Number of times pregnant	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8
0	# 2. Plasma glucose concentration a 2 hours in...	NaN							
1	# 3. Diastolic blood pressure (mm Hg)	NaN							
2	# 4. Triceps skin fold thickness (mm)	NaN							
3	# 5. 2-Hour serum insulin (mu U/ml)	NaN							
4	# 6. Body mass index (weight in kg/(height in ...	NaN							
...
771	10	101.0	76.0	48.0	180.0	32.9	0.171	63.0	0.0
772	2	122.0	70.0	27.0	0.0	36.8	0.340	27.0	0.0
773	5	121.0	72.0	23.0	112.0	26.2	0.245	30.0	0.0
774	1	126.0	60.0	0.0	0.0	30.1	0.349	47.0	1.0
775	1	93.0	70.0	31.0	0.0	30.4	0.315	23.0	0.0

```
df.head(12)
```

	# 1. Number of times pregnant	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8
0	# 2. Plasma glucose concentration a 2 hours in...	NaN							
1	# 3. Diastolic blood pressure (mm Hg)	NaN							
2	# 4. Triceps skin fold thickness (mm)	NaN							
3	# 5. 2-Hour serum insulin (mu U/ml)	NaN							
4	# 6. Body mass index (weight in kg/(height in ...	NaN							
5	# 7. Diabetes pedigree function	NaN							
6	# 8. Age (years)	NaN							
7	# 9. Class variable (0 or 1)	NaN							
8	6	148.0	72.0	35.0	0.0	33.6	0.627	50.0	1.0
9	1	85.0	66.0	29.0	0.0	26.6	0.351	31.0	0.0
10	8	183.0	64.0	0.0	0.0	23.3	0.672	32.0	1.0

✓ a) Data Cleaning

```
data = df[8: ]
data
```

# 1. Number of times pregnant	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8
8	6	148.0	72.0	35.0	0.0	33.6	0.627	50.0
9	1	85.0	66.0	29.0	0.0	26.6	0.351	31.0
10	8	183.0	64.0	0.0	0.0	23.3	0.672	32.0
11	1	89.0	66.0	23.0	94.0	28.1	0.167	21.0
12	0	137.0	40.0	35.0	168.0	43.1	2.288	33.0
...
771	10	101.0	76.0	48.0	180.0	32.9	0.171	63.0
772	2	122.0	70.0	27.0	0.0	36.8	0.340	27.0
773	5	121.0	72.0	23.0	112.0	26.2	0.245	30.0
774	1	126.0	60.0	0.0	0.0	30.1	0.349	47.0
775	1	93.0	70.0	31.0	0.0	30.4	0.315	23.0

768 rows × 9 columns

```
# Renaming columns
# df.rename(columns=['# 1. Number of times pregnant', '# 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test',
data.columns = ['# 1. Number of times pregnant', '# 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test',
data
```

# 1. Number of times pregnant	# 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test	# 3. Diastolic blood pressure (mm Hg)	# 4. Triceps skin fold thickness (mm)	# 5. 2-Hour serum insulin (mu U/ml)	# 6. Body mass index (weight in kg/(height in m)^2)	# 7. Diabetes pedigree function	# 8. Age (years)	# 9. Class variable (0 or 1)
8	6	148.0	72.0	35.0	0.0	33.6	0.627	50.0
9	1	85.0	66.0	29.0	0.0	26.6	0.351	31.0
10	8	183.0	64.0	0.0	0.0	23.3	0.672	32.0
11	1	89.0	66.0	23.0	94.0	28.1	0.167	21.0
12	0	137.0	40.0	35.0	168.0	43.1	2.288	33.0
...
771	10	101.0	76.0	48.0	180.0	32.9	0.171	63.0
772	2	122.0	70.0	27.0	0.0	36.8	0.340	27.0
773	5	121.0	72.0	23.0	112.0	26.2	0.245	30.0
774	1	126.0	60.0	0.0	0.0	30.1	0.349	47.0
775	1	93.0	70.0	31.0	0.0	30.4	0.315	23.0

768 rows × 9 columns

data.isnull().sum()

```
# 1. Number of times pregnant          0
# 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test    3
# 3. Diastolic blood pressure (mm Hg)      1
# 4. Triceps skin fold thickness (mm)      1
# 5. 2-Hour serum insulin (mu U/ml)       1
# 6. Body mass index (weight in kg/(height in m)^2)   2
# 7. Diabetes pedigree function          1
# 8. Age (years)                         0
# 9. Class variable (0 or 1)              0
dtype: int64
```

data.shape

(768, 9)

data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 8 to 775
Data columns (total 9 columns):
 # Column                                         Non-Null Count Dtype
 --- -----
 0 # 1. Number of times pregnant                 768 non-null   object
 1 # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test 765 non-null   float64
```

```

2 # 3. Diastolic blood pressure (mm Hg)           767 non-null   float64
3 # 4. Triceps skin fold thickness (mm)          767 non-null   float64
4 # 5. 2-Hour serum insulin (mu U/ml)           767 non-null   float64
5 # 6. Body mass index (weight in kg/(height in m)^2) 766 non-null   float64
6 # 7. Diabetes pedigree function                767 non-null   float64
7 # 8. Age (years)                             768 non-null   float64
8 # 9. Class variable (0 or 1)                  768 non-null   float64
dtypes: float64(8), object(1)
memory usage: 54.1+ KB

```

```
data.describe()
```

	# 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test	# 3. Diastolic blood pressure (mm Hg)	# 4. Triceps skin fold thickness (mm)	# 5. 2-Hour serum insulin (mu U/ml)	# 6. Body mass index (weight in kg/(height in m)^2)	# 7. Diabetes pedigree function	# 8. Age (years)	# 9. Class variable (0 or 1)
count	765.000000	767.000000	767.000000	767.000000	766.000000	767.000000	768.000000	768.000000
mean	120.773856	69.114733	20.518905	79.701434	31.985117	0.472309	33.240885	0.348958
std	31.893226	19.366734	15.955203	115.287146	7.891759	0.331328	11.760232	0.476951
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000	21.000000	0.000000
25%	99.000000	63.000000	0.000000	0.000000	27.300000	0.244000	24.000000	0.000000
50%	117.000000	72.000000	23.000000	29.000000	32.000000	0.374000	29.000000	0.000000
75%	140.000000	80.000000	32.000000	126.500000	36.575000	0.626500	41.000000	1.000000

```
data.fillna(data.select_dtypes(include=['number']).mean(), inplace=True)
```

```
data.info()
```

```

Է <class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 8 to 775
Data columns (total 9 columns):
 # Column                                         Non-Null Count Dtype
 --- -----
 0 # 1. Number of times pregnant                768 non-null   object
 1 # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test 768 non-null   float64
 2 # 3. Diastolic blood pressure (mm Hg)          768 non-null   float64
 3 # 4. Triceps skin fold thickness (mm)          768 non-null   float64
 4 # 5. 2-Hour serum insulin (mu U/ml)           768 non-null   float64
 5 # 6. Body mass index (weight in kg/(height in m)^2) 768 non-null   float64
 6 # 7. Diabetes pedigree function                768 non-null   float64
 7 # 8. Age (years)                             768 non-null   float64
 8 # 9. Class variable (0 or 1)                  768 non-null   float64
dtypes: float64(8), object(1)
memory usage: 54.1+ KB
C:\Users\shubham\AppData\Local\Temp\ipykernel_12880\209208995.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

```

See the caveats in the documentation: [```
data.shape
```](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-data.fillna(data.select_dtypes(include=['number']).mean(), inplace=True)</a></p>
</div>
<div data-bbox=)

```
Է (768, 9)
```

```

Check duplicates
data.duplicated().sum() # next data.drop_duplicates if dup

```

```
Է np.int64(0)
```

```
data
```

| # 1. Number of times pregnant | # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test | # 3. Diastolic blood pressure (mm Hg) | # 4. Triceps skin fold thickness (mm) | # 5. 2-Hour serum insulin (mu U/ml) | # 6. Body mass index (weight in kg/(height in m)^2) | # 7. Diabetes pedigree function | # 8. Age (years) | # 9. Class variable (0 or 1) |
|-------------------------------|-------------------------------------------------------------------------------|---------------------------------------|---------------------------------------|-------------------------------------|-----------------------------------------------------|---------------------------------|------------------|------------------------------|
| 8                             | 6                                                                             | 148.0                                 | 72.0                                  | 35.0                                | 0.0                                                 | 33.6                            | 0.627            | 50.0                         |
| 9                             | 1                                                                             | 85.0                                  | 66.0                                  | 29.0                                | 0.0                                                 | 26.6                            | 0.351            | 31.0                         |
| 10                            | 8                                                                             | 183.0                                 | 64.0                                  | 0.0                                 | 0.0                                                 | 23.3                            | 0.672            | 32.0                         |
| 11                            | 1                                                                             | 89.0                                  | 66.0                                  | 23.0                                | 94.0                                                | 28.1                            | 0.167            | 21.0                         |
| 12                            | 0                                                                             | 137.0                                 | 40.0                                  | 35.0                                | 168.0                                               | 43.1                            | 2.288            | 33.0                         |
| ...                           | ...                                                                           | ...                                   | ...                                   | ...                                 | ...                                                 | ...                             | ...              | ...                          |
| 771                           | 10                                                                            | 101.0                                 | 76.0                                  | 48.0                                | 180.0                                               | 32.9                            | 0.171            | 63.0                         |
| 772                           | 2                                                                             | 122.0                                 | 70.0                                  | 27.0                                | 0.0                                                 | 36.8                            | 0.340            | 27.0                         |
| 773                           | 5                                                                             | 121.0                                 | 72.0                                  | 23.0                                | 112.0                                               | 26.2                            | 0.245            | 30.0                         |
| 774                           | 1                                                                             | 126.0                                 | 60.0                                  | 0.0                                 | 0.0                                                 | 30.1                            | 0.349            | 47.0                         |
| 775                           | 1                                                                             | 93.0                                  | 70.0                                  | 31.0                                | 0.0                                                 | 30.4                            | 0.315            | 23.0                         |

768 rows × 9 columns

✓ b) Data Integration (merge)

```
subset1 = data.iloc[:200, :]
subset1
```

| # 1. Number of times pregnant | # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test | # 3. Diastolic blood pressure (mm Hg) | # 4. Triceps skin fold thickness (mm) | # 5. 2-Hour serum insulin (mu U/ml) | # 6. Body mass index (weight in kg/(height in m)^2) | # 7. Diabetes pedigree function | # 8. Age (years) | # 9. Class variable (0 or 1) |
|-------------------------------|-------------------------------------------------------------------------------|---------------------------------------|---------------------------------------|-------------------------------------|-----------------------------------------------------|---------------------------------|------------------|------------------------------|
| 8                             | 6                                                                             | 148.0                                 | 72.0                                  | 35.0                                | 0.0                                                 | 33.6                            | 0.627            | 50.0                         |
| 9                             | 1                                                                             | 85.0                                  | 66.0                                  | 29.0                                | 0.0                                                 | 26.6                            | 0.351            | 31.0                         |
| 10                            | 8                                                                             | 183.0                                 | 64.0                                  | 0.0                                 | 0.0                                                 | 23.3                            | 0.672            | 32.0                         |
| 11                            | 1                                                                             | 89.0                                  | 66.0                                  | 23.0                                | 94.0                                                | 28.1                            | 0.167            | 21.0                         |
| 12                            | 0                                                                             | 137.0                                 | 40.0                                  | 35.0                                | 168.0                                               | 43.1                            | 2.288            | 33.0                         |
| ...                           | ...                                                                           | ...                                   | ...                                   | ...                                 | ...                                                 | ...                             | ...              | ...                          |
| 203                           | 5                                                                             | 158.0                                 | 84.0                                  | 41.0                                | 210.0                                               | 39.4                            | 0.395            | 29.0                         |
| 204                           | 1                                                                             | 105.0                                 | 58.0                                  | 0.0                                 | 0.0                                                 | 24.3                            | 0.187            | 21.0                         |
| 205                           | 3                                                                             | 107.0                                 | 62.0                                  | 13.0                                | 48.0                                                | 22.9                            | 0.678            | 23.0                         |
| 206                           | 4                                                                             | 109.0                                 | 64.0                                  | 44.0                                | 99.0                                                | 34.8                            | 0.905            | 26.0                         |
| 207                           | 4                                                                             | 148.0                                 | 60.0                                  | 27.0                                | 318.0                                               | 30.9                            | 0.150            | 29.0                         |

200 rows × 9 columns

```
subset2 = data.iloc[150: 400, :]
subset2
```

|     | # 1.<br>Number of<br>times<br>pregnant | # 2. Plasma glucose<br>concentration a 2<br>hours in an oral<br>glucose tolerance<br>test | # 3.<br>Diastolic<br>blood<br>pressure (mm<br>Hg) | # 4. Triceps<br>skin fold<br>thickness<br>(mm) | # 5. 2-Hour<br>serum<br>insulin (mu<br>U/ml) | # 6. Body mass<br>index (weight<br>in kg/(height<br>in m)^2) | # 7.<br>Diabetes<br>pedigree<br>function | # 8.<br>Age<br>(years) | # 9. Class<br>variable<br>(0 or 1) |
|-----|----------------------------------------|-------------------------------------------------------------------------------------------|---------------------------------------------------|------------------------------------------------|----------------------------------------------|--------------------------------------------------------------|------------------------------------------|------------------------|------------------------------------|
| 158 | 1                                      | 136.0                                                                                     | 74.0                                              | 50.0                                           | 204.000000                                   | 37.4                                                         | 0.399                                    | 24.0                   | 0.0                                |
| 159 | 4                                      | 114.0                                                                                     | 65.0                                              | 0.0                                            | 0.000000                                     | 21.9                                                         | 0.432                                    | 37.0                   | 0.0                                |
| 160 | 9                                      | 156.0                                                                                     | 86.0                                              | 28.0                                           | 79.701434                                    | 34.3                                                         | 1.189                                    | 42.0                   | 1.0                                |
| 161 | 1                                      | 153.0                                                                                     | 82.0                                              | 42.0                                           | 485.000000                                   | 40.6                                                         | 0.687                                    | 23.0                   | 0.0                                |
| 162 | 8                                      | 188.0                                                                                     | 78.0                                              | 0.0                                            | 0.000000                                     | 47.9                                                         | 0.137                                    | 43.0                   | 1.0                                |
| ... | ...                                    | ...                                                                                       | ...                                               | ...                                            | ...                                          | ...                                                          | ...                                      | ...                    | ...                                |
| 403 | 2                                      | 127.0                                                                                     | 58.0                                              | 24.0                                           | 275.000000                                   | 27.7                                                         | 1.600                                    | 25.0                   | 0.0                                |
| 404 | 3                                      | 96.0                                                                                      | 56.0                                              | 34.0                                           | 115.000000                                   | 24.7                                                         | 0.944                                    | 39.0                   | 0.0                                |
| 405 | 0                                      | 131.0                                                                                     | 66.0                                              | 40.0                                           | 0.000000                                     | 34.3                                                         | 0.196                                    | 22.0                   | 1.0                                |
| 406 | 3                                      | 82.0                                                                                      | 70.0                                              | 0.0                                            | 0.000000                                     | 21.1                                                         | 0.389                                    | 25.0                   | 0.0                                |
| 407 | 3                                      | 193.0                                                                                     | 70.0                                              | 31.0                                           | 0.000000                                     | 34.9                                                         | 0.241                                    | 25.0                   | 1.0                                |

250 rows x 9 columns

```
inner_merge= subset1.merge(subset2, how="inner")
inner_merge
```

| # 1.<br>Number<br>of<br>times<br>pregnant | # 2. Plasma glucose<br>concentration a<br>2 hours in an oral<br>glucose tolerance<br>test | # 3.<br>Diastolic<br>blood<br>pressure<br>(mm Hg) | # 4.<br>Triceps<br>skin fold<br>thickness<br>(mm) | # 5. 2-Hour<br>serum<br>insulin (mu<br>U/ml) | # 6. Body mass<br>index (weight<br>in kg/(height<br>in m) <sup>2</sup> ) | # 7.<br>Diabetes<br>pedigree<br>function | # 8.<br>Age<br>(years) | # 9. Class<br>variable<br>(0 or 1) |     |
|-------------------------------------------|-------------------------------------------------------------------------------------------|---------------------------------------------------|---------------------------------------------------|----------------------------------------------|--------------------------------------------------------------------------|------------------------------------------|------------------------|------------------------------------|-----|
| 0                                         | 1                                                                                         | 136.0                                             | 74.0                                              | 50.0                                         | 204.000000                                                               | 37.400000                                | 0.399                  | 24.0                               | 0.0 |
| 1                                         | 4                                                                                         | 114.0                                             | 65.0                                              | 0.0                                          | 0.000000                                                                 | 21.900000                                | 0.432                  | 37.0                               | 0.0 |
| 2                                         | 9                                                                                         | 156.0                                             | 86.0                                              | 28.0                                         | 79.701434                                                                | 34.300000                                | 1.189                  | 42.0                               | 1.0 |
| 3                                         | 1                                                                                         | 153.0                                             | 82.0                                              | 42.0                                         | 485.000000                                                               | 40.600000                                | 0.687                  | 23.0                               | 0.0 |
| 4                                         | 8                                                                                         | 188.0                                             | 78.0                                              | 0.0                                          | 0.000000                                                                 | 47.900000                                | 0.137                  | 43.0                               | 1.0 |
| 5                                         | 7                                                                                         | 152.0                                             | 88.0                                              | 44.0                                         | 0.000000                                                                 | 50.000000                                | 0.337                  | 36.0                               | 1.0 |
| 6                                         | 2                                                                                         | 99.0                                              | 52.0                                              | 15.0                                         | 94.000000                                                                | 24.600000                                | 0.637                  | 21.0                               | 0.0 |
| 7                                         | 1                                                                                         | 109.0                                             | 56.0                                              | 21.0                                         | 135.000000                                                               | 25.200000                                | 0.833                  | 23.0                               | 0.0 |
| 8                                         | 2                                                                                         | 88.0                                              | 74.0                                              | 19.0                                         | 53.000000                                                                | 29.000000                                | 0.229                  | 22.0                               | 0.0 |
| 9                                         | 17                                                                                        | 163.0                                             | 72.0                                              | 41.0                                         | 114.000000                                                               | 40.900000                                | 0.817                  | 47.0                               | 1.0 |
| 10                                        | 4                                                                                         | 151.0                                             | 90.0                                              | 38.0                                         | 0.000000                                                                 | 29.700000                                | 0.294                  | 36.0                               | 0.0 |
| 11                                        | 7                                                                                         | 102.0                                             | 74.0                                              | 40.0                                         | 105.000000                                                               | 37.200000                                | 0.204                  | 45.0                               | 0.0 |
| 12                                        | 0                                                                                         | 114.0                                             | 80.0                                              | 34.0                                         | 285.000000                                                               | 44.200000                                | 0.167                  | 27.0                               | 0.0 |
| 13                                        | 2                                                                                         | 100.0                                             | 64.0                                              | 23.0                                         | 0.000000                                                                 | 29.700000                                | 0.368                  | 21.0                               | 0.0 |
| 14                                        | 0                                                                                         | 131.0                                             | 88.0                                              | 0.0                                          | 0.000000                                                                 | 31.600000                                | 0.743                  | 32.0                               | 1.0 |
| 15                                        | 6                                                                                         | 104.0                                             | 74.0                                              | 18.0                                         | 156.000000                                                               | 29.900000                                | 0.722                  | 41.0                               | 1.0 |
| 16                                        | 3                                                                                         | 148.0                                             | 66.0                                              | 25.0                                         | 0.000000                                                                 | 32.500000                                | 0.256                  | 22.0                               | 0.0 |
| 17                                        | 4                                                                                         | 120.0                                             | 68.0                                              | 0.0                                          | 0.000000                                                                 | 29.600000                                | 0.709                  | 34.0                               | 0.0 |
| 18                                        | 4                                                                                         | 110.0                                             | 66.0                                              | 0.0                                          | 0.000000                                                                 | 31.900000                                | 0.471                  | 29.0                               | 0.0 |
| 19                                        | 3                                                                                         | 111.0                                             | 90.0                                              | 12.0                                         | 78.000000                                                                | 28.400000                                | 0.495                  | 29.0                               | 0.0 |
| 20                                        | 6                                                                                         | 102.0                                             | 82.0                                              | 0.0                                          | 0.000000                                                                 | 30.800000                                | 0.180                  | 36.0                               | 1.0 |
| 21                                        | 6                                                                                         | 134.0                                             | 70.0                                              | 23.0                                         | 130.000000                                                               | 35.400000                                | 0.542                  | 29.0                               | 1.0 |
| 22                                        | 2                                                                                         | 87.0                                              | 0.0                                               | 23.0                                         | 0.000000                                                                 | 28.900000                                | 0.773                  | 25.0                               | 0.0 |
| 23                                        | 1                                                                                         | 79.0                                              | 60.0                                              | 42.0                                         | 48.000000                                                                | 43.500000                                | 0.678                  | 23.0                               | 0.0 |
| 24                                        | 2                                                                                         | 75.0                                              | 64.0                                              | 24.0                                         | 55.000000                                                                | 29.700000                                | 0.370                  | 33.0                               | 0.0 |
| 25                                        | 8                                                                                         | 179.0                                             | 72.0                                              | 42.0                                         | 130.000000                                                               | 32.700000                                | 0.719                  | 36.0                               | 1.0 |
| 26                                        | 6                                                                                         | 85.0                                              | 78.0                                              | 0.0                                          | 0.000000                                                                 | 31.200000                                | 0.382                  | 42.0                               | 0.0 |
| 27                                        | 0                                                                                         | 129.0                                             | 110.0                                             | 46.0                                         | 130.000000                                                               | 67.100000                                | 0.319                  | 26.0                               | 1.0 |
| 28                                        | 5                                                                                         | 143.0                                             | 78.0                                              | 0.0                                          | 0.000000                                                                 | 45.000000                                | 0.190                  | 47.0                               | 0.0 |
| 29                                        | 5                                                                                         | 130.0                                             | 82.0                                              | 0.0                                          | 0.000000                                                                 | 39.100000                                | 0.956                  | 37.0                               | 1.0 |
| 30                                        | 6                                                                                         | 87.0                                              | 80.0                                              | 0.0                                          | 0.000000                                                                 | 23.200000                                | 0.084                  | 32.0                               | 0.0 |
| 31                                        | 0                                                                                         | 119.0                                             | 64.0                                              | 18.0                                         | 92.000000                                                                | 34.900000                                | 0.725                  | 23.0                               | 0.0 |
| 32                                        | 1                                                                                         | 0.0                                               | 74.0                                              | 20.0                                         | 23.000000                                                                | 27.700000                                | 0.299                  | 21.0                               | 0.0 |
| 33                                        | 5                                                                                         | 73.0                                              | 60.0                                              | 0.0                                          | 0.000000                                                                 | 26.800000                                | 0.268                  | 27.0                               | 0.0 |
| 34                                        | 4                                                                                         | 141.0                                             | 74.0                                              | 0.0                                          | 0.000000                                                                 | 27.600000                                | 0.244                  | 40.0                               | 0.0 |
| 35                                        | 7                                                                                         | 194.0                                             | 68.0                                              | 28.0                                         | 0.000000                                                                 | 35.900000                                | 0.745                  | 41.0                               | 1.0 |
| 36                                        | 8                                                                                         | 181.0                                             | 68.0                                              | 36.0                                         | 495.000000                                                               | 30.100000                                | 0.615                  | 60.0                               | 1.0 |
| 37                                        | 1                                                                                         | 128.0                                             | 98.0                                              | 41.0                                         | 58.000000                                                                | 31.985117                                | 1.321                  | 33.0                               | 1.0 |
| 38                                        | 8                                                                                         | 109.0                                             | 76.0                                              | 39.0                                         | 114.000000                                                               | 27.900000                                | 0.640                  | 31.0                               | 1.0 |
| 39                                        | 5                                                                                         | 139.0                                             | 80.0                                              | 35.0                                         | 160.000000                                                               | 31.600000                                | 0.361                  | 25.0                               | 1.0 |
| 40                                        | 3                                                                                         | 111.0                                             | 62.0                                              | 0.0                                          | 0.000000                                                                 | 22.600000                                | 0.142                  | 21.0                               | 0.0 |
| 41                                        | 9                                                                                         | 123.0                                             | 70.0                                              | 44.0                                         | 94.000000                                                                | 33.100000                                | 0.374                  | 40.0                               | 0.0 |
| 42                                        | 7                                                                                         | 159.0                                             | 66.0                                              | 0.0                                          | 0.000000                                                                 | 30.400000                                | 0.383                  | 36.0                               | 1.0 |
| 43                                        | 11                                                                                        | 135.0                                             | 0.0                                               | 0.0                                          | 0.000000                                                                 | 52.300000                                | 0.578                  | 40.0                               | 1.0 |
| 44                                        | 8                                                                                         | 85.0                                              | 55.0                                              | 20.0                                         | 0.000000                                                                 | 24.400000                                | 0.136                  | 42.0                               | 0.0 |
| 45                                        | 5                                                                                         | 158.0                                             | 84.0                                              | 41.0                                         | 210.000000                                                               | 39.400000                                | 0.395                  | 29.0                               | 1.0 |
| 46                                        | 1                                                                                         | 105.0                                             | 58.0                                              | 0.0                                          | 0.000000                                                                 | 24.300000                                | 0.187                  | 21.0                               | 0.0 |

|    |   |       |      |      |            |           |       |      |     |
|----|---|-------|------|------|------------|-----------|-------|------|-----|
| 47 | 3 | 107.0 | 62.0 | 13.0 | 48.000000  | 22.900000 | 0.678 | 23.0 | 1.0 |
| 48 | 4 | 109.0 | 64.0 | 44.0 | 99.000000  | 34.800000 | 0.905 | 26.0 | 1.0 |
| 49 | 4 | 148.0 | 60.0 | 27.0 | 318.000000 | 30.900000 | 0.150 | 29.0 | 1.0 |

inner\_merge.shape

(50, 9)

```
outer_merge= subset1.merge(subset2, how="outer")
outer_merge
```

| # 1. Number of times pregnant | # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test | # 3. Diastolic blood pressure (mm Hg) | # 4. Triceps skin fold thickness (mm) | # 5. 2-Hour serum insulin (mu U/ml) | # 6. Body mass index (weight in kg/(height in m)^2) | # 7. Diabetes pedigree function | # 8. Age (years) | # 9. Class variable (0 or 1) |     |
|-------------------------------|-------------------------------------------------------------------------------|---------------------------------------|---------------------------------------|-------------------------------------|-----------------------------------------------------|---------------------------------|------------------|------------------------------|-----|
| 0                             | 0                                                                             | 78.0                                  | 88.0                                  | 29.0                                | 40.000000                                           | 36.9                            | 0.434            | 21.0                         | 0.0 |
| 1                             | 0                                                                             | 84.0                                  | 64.0                                  | 22.0                                | 66.000000                                           | 35.8                            | 0.545            | 21.0                         | 0.0 |
| 2                             | 0                                                                             | 86.0                                  | 68.0                                  | 32.0                                | 0.000000                                            | 35.8                            | 0.238            | 25.0                         | 0.0 |
| 3                             | 0                                                                             | 93.0                                  | 60.0                                  | 25.0                                | 92.000000                                           | 28.7                            | 0.532            | 22.0                         | 0.0 |
| 4                             | 0                                                                             | 93.0                                  | 100.0                                 | 39.0                                | 72.000000                                           | 43.4                            | 1.021            | 35.0                         | 0.0 |
| ...                           | ...                                                                           | ...                                   | ...                                   | ...                                 | ...                                                 | ...                             | ...              | ...                          |     |
| 395                           | 9                                                                             | 156.0                                 | 86.0                                  | 28.0                                | 79.701434                                           | 34.3                            | 1.189            | 42.0                         | 1.0 |
| 396                           | 9                                                                             | 164.0                                 | 84.0                                  | 21.0                                | 0.000000                                            | 30.8                            | 0.831            | 32.0                         | 1.0 |
| 397                           | 9                                                                             | 165.0                                 | 88.0                                  | 0.0                                 | 0.000000                                            | 30.4                            | 0.302            | 49.0                         | 1.0 |
| 398                           | 9                                                                             | 171.0                                 | 110.0                                 | 24.0                                | 240.000000                                          | 45.4                            | 0.721            | 54.0                         | 1.0 |
| 399                           | 9                                                                             | 184.0                                 | 85.0                                  | 15.0                                | 0.000000                                            | 30.0                            | 1.213            | 49.0                         | 1.0 |

400 rows × 9 columns

data

| # 1. Number of times pregnant | # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test | # 3. Diastolic blood pressure (mm Hg) | # 4. Triceps skin fold thickness (mm) | # 5. 2-Hour serum insulin (mu U/ml) | # 6. Body mass index (weight in kg/(height in m)^2) | # 7. Diabetes pedigree function | # 8. Age (years) | # 9. Class variable (0 or 1) |     |
|-------------------------------|-------------------------------------------------------------------------------|---------------------------------------|---------------------------------------|-------------------------------------|-----------------------------------------------------|---------------------------------|------------------|------------------------------|-----|
| 8                             | 6                                                                             | 148.0                                 | 72.0                                  | 35.0                                | 0.0                                                 | 33.6                            | 0.627            | 50.0                         | 1.0 |
| 9                             | 1                                                                             | 85.0                                  | 66.0                                  | 29.0                                | 0.0                                                 | 26.6                            | 0.351            | 31.0                         | 0.0 |
| 10                            | 8                                                                             | 183.0                                 | 64.0                                  | 0.0                                 | 0.0                                                 | 23.3                            | 0.672            | 32.0                         | 1.0 |
| 11                            | 1                                                                             | 89.0                                  | 66.0                                  | 23.0                                | 94.0                                                | 28.1                            | 0.167            | 21.0                         | 0.0 |
| 12                            | 0                                                                             | 137.0                                 | 40.0                                  | 35.0                                | 168.0                                               | 43.1                            | 2.288            | 33.0                         | 1.0 |
| ...                           | ...                                                                           | ...                                   | ...                                   | ...                                 | ...                                                 | ...                             | ...              | ...                          |     |
| 771                           | 10                                                                            | 101.0                                 | 76.0                                  | 48.0                                | 180.0                                               | 32.9                            | 0.171            | 63.0                         | 0.0 |
| 772                           | 2                                                                             | 122.0                                 | 70.0                                  | 27.0                                | 0.0                                                 | 36.8                            | 0.340            | 27.0                         | 0.0 |
| 773                           | 5                                                                             | 121.0                                 | 72.0                                  | 23.0                                | 112.0                                               | 26.2                            | 0.245            | 30.0                         | 0.0 |
| 774                           | 1                                                                             | 126.0                                 | 60.0                                  | 0.0                                 | 0.0                                                 | 30.1                            | 0.349            | 47.0                         | 1.0 |
| 775                           | 1                                                                             | 93.0                                  | 70.0                                  | 31.0                                | 0.0                                                 | 30.4                            | 0.315            | 23.0                         | 0.0 |

768 rows × 9 columns

Not Applicable -

c) Data Transformation , d) Error Correction

e) Model Building

```
x = data.drop(columns="# 9. Class variable (0 or 1)")
x.head(4)
```

| # 1. Number of times pregnant | # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test | # 3. Diastolic blood pressure (mm Hg) | # 4. Triceps skin fold thickness (mm) | # 5. 2-Hour serum insulin (mu U/ml) | # 6. Body mass index (weight in kg/(height in m) <sup>2</sup> ) | # 7. Diabetes pedigree function | # 8. Age (years) |
|-------------------------------|-------------------------------------------------------------------------------|---------------------------------------|---------------------------------------|-------------------------------------|-----------------------------------------------------------------|---------------------------------|------------------|
| 8                             | 6                                                                             | 148.0                                 | 72.0                                  | 35.0                                | 0.0                                                             | 33.6                            | 0.627            |
| 9                             | 1                                                                             | 85.0                                  | 66.0                                  | 29.0                                | 0.0                                                             | 26.6                            | 0.351            |
| 10                            | 8                                                                             | 183.0                                 | 64.0                                  | 0.0                                 | 0.0                                                             | 23.3                            | 0.672            |

```
y = data['# 9. Class variable (0 or 1)']
y.head(4)
```

```
→ 8 1.0
 9 0.0
 10 1.0
 11 0.0
Name: # 9. Class variable (0 or 1), dtype: float64
```

```
from sklearn.model_selection import train_test_split
```

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,random_state=42)
x_train.head(4)
```

```
print("x train ", x_train.shape, "and x test ", x_test.shape)
```

→ x train (614, 8) and x test (154, 8)

```
y_train.head(3)
#classification
```

```
→ 68 0.0
 626 1.0
 354 0.0
Name: # 9. Class variable (0 or 1), dtype: float64
```

```
from sklearn.tree import DecisionTreeClassifier
```

```
model = DecisionTreeClassifier(criterion = "gini")
model
```

DecisionTreeClassifier()  

```
model.fit(x_train, y_train)
```

```
→ ▾ DecisionTreeClassifier ⓘ ⓘ
DecisionTreeClassifier()
```

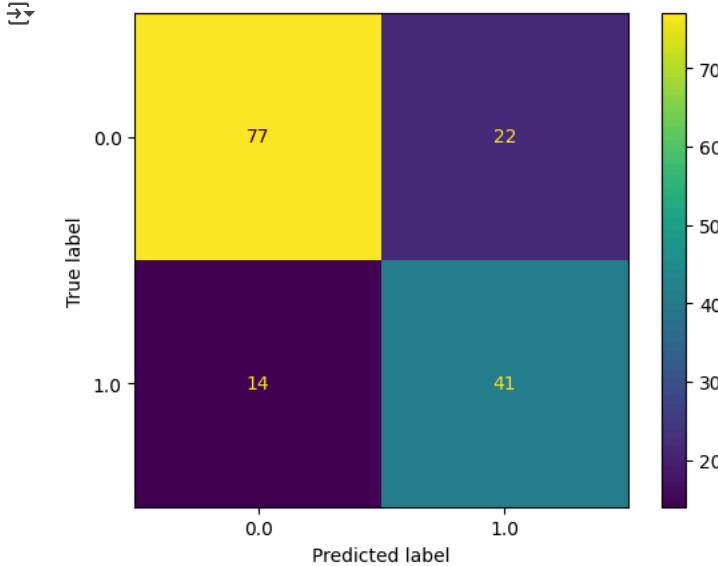
```
y_pred = model.predict(x_test)
y_pred
```

```
print(y_pred.shape, y_test.shape, x_test.shape)
```

→ (154,) (154,) (154, 8)

```
from sklearn.metrics import ConfusionMatrixDisplay
```

```
ConfusionMatrixDisplay.from_predictions(y_test, y_pred)
plt.show()
```



```
from sklearn.metrics import accuracy_score, precision_score, recall_score
acc = accuracy_score(y_test, y_pred)
```

```
pre = precision_score(y_test, y_pred)
```

```
rec = recall_score(y_test, y_pred)
```

```
print(f" Accuracy : {acc : .4f} ; Precision : {pre: .4f} ; Recall : {rec : .4f} ")
```

→ Accuracy : 0.7662 ; Precision : 0.6508 ; Recall : 0.7455

```
data.iloc[[0,1,2]]
```

|   | # 1. Number of times pregnant | # 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test | # 3. Diastolic blood pressure (mm Hg) | # 4. Triceps skin fold thickness (mm) | # 5. 2-Hour serum insulin (mu U/ml) | # 6. Body mass index (weight in kg/(height in m)^2) | # 7. Diabetes pedigree function | # 8. Age (years) | # 9. Class variable (0 or 1) |
|---|-------------------------------|-------------------------------------------------------------------------------|---------------------------------------|---------------------------------------|-------------------------------------|-----------------------------------------------------|---------------------------------|------------------|------------------------------|
| 8 | 6                             | 148.0                                                                         | 72.0                                  | 35.0                                  | 0.0                                 | 33.6                                                | 0.627                           | 50.0             | 1.0                          |
| 9 | 1                             | 85.0                                                                          | 66.0                                  | 29.0                                  | 0.0                                 | 26.6                                                | 0.351                           | 31.0             | 0.0                          |

```
gini_index0=x.iloc[[0]]
```

```
gini_index1=x.iloc[[1]]
```

```
gini_index2=x.iloc[[2]]
```

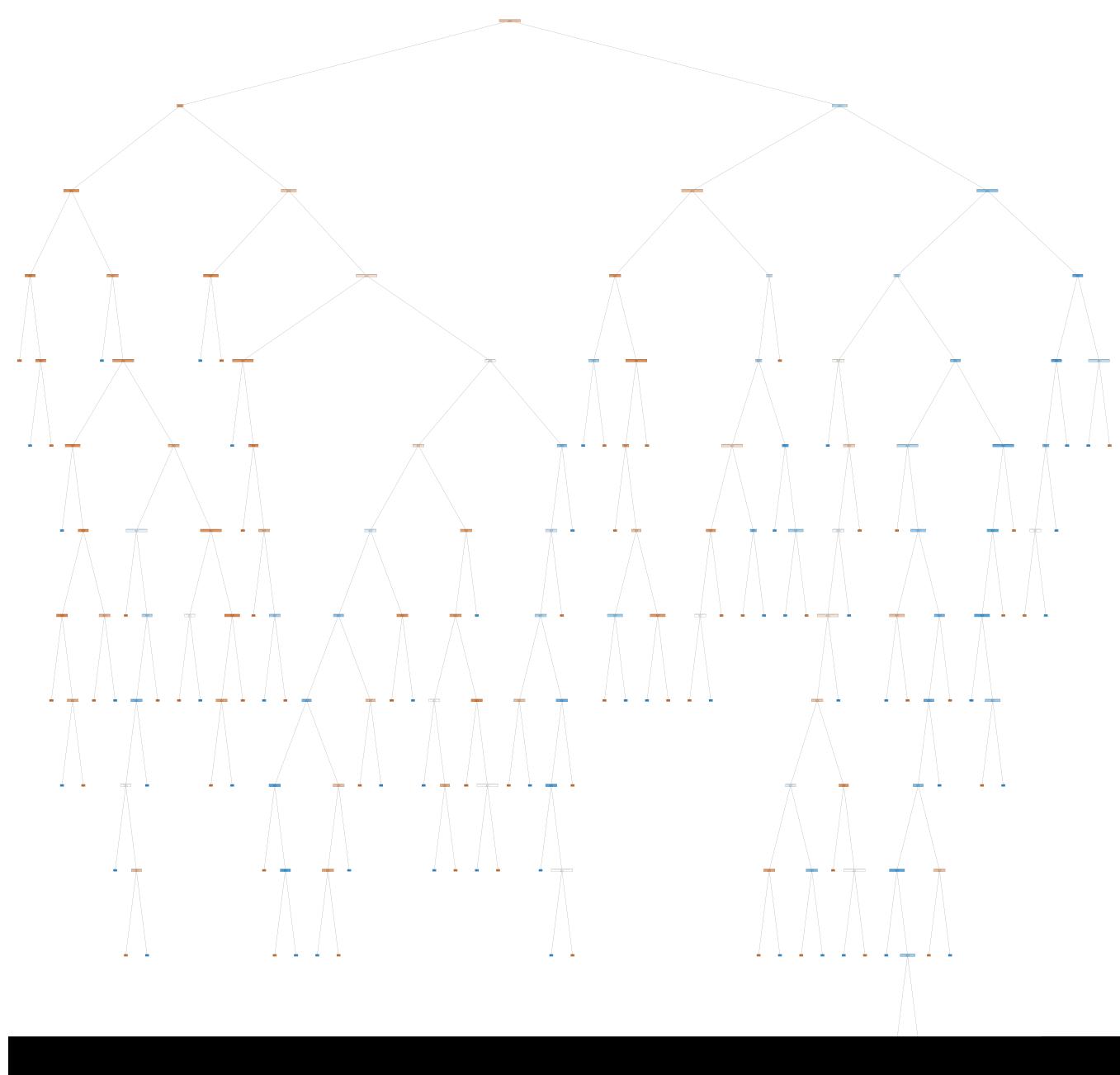
```
print(model.predict(gini_index0), model.predict(gini_index1), model.predict(gini_index2))
```

→ [1.] [0.] [1.]

1- Yes Diabetes

0- No Diabetes

```
from sklearn.tree import plot_tree
plt.figure(figsize=(250,250))
plot_tree(model , feature_names=x.columns, class_names=['No Diabetes', 'Yes'], filled=True)
plt.show()
```



```
import numpy as np

Input guidelines (range & units)
INPUT_GUIDELINES = {
 "Pregnancies": (0, 15, "times"),
 "Glucose Level": (50, 200, "mg/dL"),
 "Blood Pressure": (40, 130, "mmHg"),
 "Skin Thickness": (0, 100, "mm"),
 "Insulin Level": (0, 300, "μU/ml"),
 "BMI": (10, 50, "kg/m²"),
 "Diabetes Pedigree": (0.0, 2.5, "factor"),
 "Age": (10, 100, "years")
}
```

```
Function to take user input and predict diabetes
def predict_diabetes(model):
 print("\n◆ Enter the following health metrics:\n")

 user_input = []

 for feature, (min_val, max_val, unit) in INPUT_GUIDELINES.items():
 while True:
 try:
 value = float(input(f"{feature} ({min_val}-{max_val} {unit}): "))

 if min_val <= value <= max_val:
 user_input.append(value)
 break
 else:
 print(f"⚠ Out of range! Enter a value between {min_val} and {max_val} {unit}.")
 except ValueError:
 print("✖ Invalid input! Please enter a numeric value.")

 # Convert to NumPy array & reshape for model
 user_input = np.array(user_input).reshape(1, -1)

 # Predict diabetes (0 = No, 1 = Yes)
 prediction = model.predict(user_input)[0]
 result = "Yes (Diabetes Detected) ❓" if prediction == 1 else "No (Healthy) ✓"

 print(f"\n🔍 Prediction: {result}")

Call function with trained model
predict_diabetes(model)

Pregnancies (0-15 times): 5
Glucose Level (50-200 mg/dL): 180
Blood Pressure (40-130 mmHg): 90
Skin Thickness (0-100 mm): 35
Insulin Level (0-300 mu U/ml): 150
BMI (10-50 kg/m²): 32.0
Diabetes Pedigree (0.0-2.5 factor): 1.2
Age (10-100 years): 50
🔍 Prediction: Yes (Diabetes Detected) ❓

Pregnancies (0-15 times): 1
Glucose Level (50-200 mg/dL): 95
Blood Pressure (40-130 mmHg): 70
Skin Thickness (0-100 mm): 20
Insulin Level (0-300 mu U/ml): 80
BMI (10-50 kg/m²): 22.5
Diabetes Pedigree (0.0-2.5 factor): 0.3
Age (10-100 years): 28
🔍 Prediction: No (Healthy) ✓

Pregnancies (0-15 times): 3
Glucose Level (50-200 mg/dL): 130
Blood Pressure (40-130 mmHg): 85
Skin Thickness (0-100 mm): 27
Insulin Level (0-300 mu U/ml): 110
BMI (10-50 kg/m²): 28.3
Diabetes Pedigree (0.0-2.5 factor): 0.8
Age (10-100 years): 40
🔍 Prediction: [Depends on Model Output]
```



◆ Enter the following health metrics:

🔍 Prediction: No (Healthy) ✓  
C:\Users\shubham\AppData\Local\ Packages\PythonSoftwareFoundation.Python.3.11\_qbz5n2kfra8p0\LocalCache\local-packages\Python311\site  
warnings.warn(

-----EXTRA PART-----

```
plt.figure(figsize=(20, 20))
data.boxplot()
```

```
plt.xticks(rotation=45)
plt.title("Box Plot for Outlier Detection")
plt.show()
```

