

```
In [150... import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler

df = pd.read_csv("/home/csl4/Downloads/ObesityDataSet_raw_and_data_sinthetic.csv")
print(df)
```

	Gender	Age	Height	Weight	family_history_with_overweight	\			
0	Female	21.000000	1.620000	64.000000		yes			
1	Female	21.000000	1.520000	56.000000		yes			
2	Male	23.000000	1.800000	77.000000		yes			
3	Male	27.000000	1.800000	87.000000		no			
4	Male	22.000000	1.780000	89.800000		no			
...	...	...	...	...		...			
2106	Female	20.976842	1.710730	131.408528		yes			
2107	Female	21.982942	1.748584	133.742943		yes			
2108	Female	22.524036	1.752206	133.689352		yes			
2109	Female	24.361936	1.739450	133.346641		yes			
2110	Female	23.664709	1.738836	133.472641		yes			
...	...	...	...	...		...			
2106	yes	3.0	3.0	Sometimes	no	2.000000	no	0.000000	1.000000
2107	yes	3.0	3.0	Sometimes	yes	3.000000	yes	3.000000	0.000000
2	no	2.0	3.0	Sometimes	no	2.000000	no	2.000000	1.000000
3	no	3.0	3.0	Sometimes	no	2.000000	no	2.000000	0.000000
4	no	2.0	1.0	Sometimes	no	2.000000	no	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...
2106	yes	3.0	3.0	Sometimes	no	1.728139	no	1.676269	0.906247
2107	yes	3.0	3.0	Sometimes	no	2.005130	no	1.341390	0.599270
2108	yes	3.0	3.0	Sometimes	no	2.054193	no	1.414209	0.646288
2109	yes	3.0	3.0	Sometimes	no	2.852339	no	1.139107	0.586035
2110	yes	3.0	3.0	Sometimes	no	2.863513	no	1.026452	0.714137
...	...	...	...	...	...	...	...	...	...
2106	Sometimes	Public_Transportation		MTRANS			NObeyesdad		
1	Sometimes	Public_Transportation					Normal_Weight		
2	Frequently	Public_Transportation					Normal_Weight		
3	Frequently		Walking				Overweight_Level_I		
4	Sometimes	Public_Transportation					Overweight_Level_II		
...	...	...	...	...	...	...	...	...	...
2106	Sometimes	Public_Transportation					Obesity_Type_III		
2107	Sometimes	Public_Transportation					Obesity_Type_III		
2108	Sometimes	Public_Transportation					Obesity_Type_III		
2109	Sometimes	Public_Transportation					Obesity_Type_III		
2110	Sometimes	Public_Transportation					Obesity_Type_III		

[2111 rows x 17 columns]

```
In [151... print(df.columns)

Index(['Gender', 'Age', 'Height', 'Weight', 'family_history_with_overweight',
       'FAVC', 'FCVC', 'NCP', 'CAEC', 'SMOKE', 'CH20', 'SCC', 'FAF', 'TUE',
       'CALC', 'MTRANS', 'NObeyesdad'],
      dtype='object')
```

```
In [152... print(df.index)

RangeIndex(start=0, stop=2111, step=1)
```

```
In [153... print(df.dtypes)
```

```
Gender          object
Age            float64
Height         float64
Weight          float64
family_history_with_overweight    object
FAVC           object
FCVC           float64
NCP            float64
CAEC           object
SMOKE          object
CH20           float64
SCC            object
FAF            float64
TUE            float64
CALC           object
MTRANS          object
NObeyesdad     object
dtype: object
```

In [154]: `print(df.info)`

```
<bound method DataFrame.info of
family_history_with_overweight \>
   Gender      Age      Height      Weight      family_h
0   Female  21.000000  1.620000  64.000000      yes
1   Female  21.000000  1.520000  56.000000      yes
2     Male  23.000000  1.800000  77.000000      yes
3     Male  27.000000  1.800000  87.000000      no
4     Male  22.000000  1.780000  89.800000      no
...
2106  Female  20.976842  1.710730  131.408528      yes
2107  Female  21.982942  1.748584  133.742943      yes
2108  Female  22.524036  1.752206  133.689352      yes
2109  Female  24.361936  1.739450  133.346641      yes
2110  Female  23.664709  1.738836  133.472641      yes

   FAVC      FCVC      NCP      CAEC      SMOKE      CH20      SCC      FAF      TUE \>
0   no     2.0     3.0  Sometimes     no  2.000000     no  0.000000  1.000000
1   no     3.0     3.0  Sometimes    yes  3.000000    yes  3.000000  0.000000
2   no     2.0     3.0  Sometimes    no  2.000000     no  2.000000  1.000000
3   no     3.0     3.0  Sometimes    no  2.000000     no  2.000000  0.000000
4   no     2.0     1.0  Sometimes    no  2.000000     no  0.000000  0.000000
...
2106  yes     3.0     3.0  Sometimes    no  1.728139     no  1.676269  0.906247
2107  yes     3.0     3.0  Sometimes    no  2.005130     no  1.341390  0.599270
2108  yes     3.0     3.0  Sometimes    no  2.054193     no  1.414209  0.646288
2109  yes     3.0     3.0  Sometimes    no  2.852339     no  1.139107  0.586035
2110  yes     3.0     3.0  Sometimes    no  2.863513     no  1.026452  0.714137

   CALC      MTRANS      NObeyesdad
0       no  Public_Transportation  Normal_Weight
1  Sometimes  Public_Transportation  Normal_Weight
2 Frequently  Public_Transportation  Normal_Weight
3 Frequently        Walking  Overweight_Level_I
4 Sometimes  Public_Transportation  Overweight_Level_II
...
2106  Sometimes  Public_Transportation  Obesity_Type_III
2107  Sometimes  Public_Transportation  Obesity_Type_III
2108  Sometimes  Public_Transportation  Obesity_Type_III
2109  Sometimes  Public_Transportation  Obesity_Type_III
2110  Sometimes  Public_Transportation  Obesity_Type_III
```

[2111 rows x 17 columns]>

In [155]: `print(df.size)`  
`print(df.shape)`

35887  
(2111, 17)

In [156... df.loc[2,'Weight'] #loc is label based # loc[row,col]

Out[156... 77.0

In [157... df.loc[4]

Out[157... Gender Male  
Age 22.0  
Height 1.78  
Weight 89.8  
family\_history\_with\_overweight no  
FAVC no  
FCVC 2.0  
NCP 1.0  
CAEC Sometimes  
SMOKE no  
CH20 2.0  
SCC no  
FAF 0.0  
TUE 0.0  
CALC Sometimes  
MTRANS Public\_Transportation  
NObeyesdad Overweight\_Level\_II  
Name: 4, dtype: object

In [158... print(df["Gender"].head())

0 Female  
1 Female  
2 Male  
3 Male  
4 Male  
Name: Gender, dtype: object

In [159... df.loc[1:3]

Out[159... Gender Age Height Weight family\_history\_with\_overweight FAVC FCVC NCP  
1 Female 21.0 1.52 56.0 yes no 3.0 3.0 So  
2 Male 23.0 1.80 77.0 yes no 2.0 3.0 So  
3 Male 27.0 1.80 87.0 no no 3.0 3.0 So

In [160... df.iloc[1:3] #iloc is index based and last value is exclusive # loc is label based

Out[160... Gender Age Height Weight family\_history\_with\_overweight FAVC FCVC NCP  
1 Female 21.0 1.52 56.0 yes no 3.0 3.0 So  
2 Male 23.0 1.80 77.0 yes no 2.0 3.0 So

In [161... # Data filtering

In [162... print(df[df.Age == 32])

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	\			
149	Female	32.0	1.67	90.0		yes	yes	3.0			
229	Male	32.0	1.75	120.0		yes	no	3.0			
278	Female	32.0	1.57	57.0		yes	yes	3.0			
	NCP	CAEC	SMOKE	CH20	SCC	FAF	TUE	CALC	MTRANS	\	
149	1.0	Sometimes	no	2.0	no	2.0	0.0	Sometimes	Automobile		
229	3.0	Sometimes	no	3.0	no	0.0	2.0	no	Automobile		
278	3.0	Sometimes	no	2.0	no	0.0	0.0	Sometimes	Automobile		
	NObeyesdad										
149	Obesity_Type_I										
229	Obesity_Type_II										
278	Normal_Weight										

```
In [163]: print(df[(df.Age == 21) & (df.Gender == "Female")])
# using & condition for filtering
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	\
0	Female	21.0	1.620000	64.000000		yes	no
1	Female	21.0	1.520000	56.000000		yes	no
11	Female	21.0	1.720000	80.000000		yes	yes
35	Female	21.0	1.500000	65.000000		yes	no
37	Female	21.0	1.600000	48.000000		no	yes
39	Female	21.0	1.750000	88.000000		yes	yes
40	Female	21.0	1.670000	75.000000		yes	yes
42	Female	21.0	1.660000	64.000000		yes	yes
45	Female	21.0	1.530000	65.000000		yes	no
49	Female	21.0	1.550000	50.000000		no	yes
50	Female	21.0	1.610000	54.500000		yes	yes
60	Female	21.0	1.550000	49.000000		yes	yes
64	Female	21.0	1.660000	57.000000		yes	yes
65	Female	21.0	1.620000	69.000000		yes	yes
97	Female	21.0	1.520000	42.000000		no	no
98	Female	21.0	1.520000	42.000000		no	no
100	Female	21.0	1.690000	63.000000		no	yes
102	Female	21.0	1.550000	57.000000		no	yes
109	Female	21.0	1.650000	88.000000		yes	yes
162	Female	21.0	1.630000	60.000000		yes	yes
211	Female	21.0	1.630000	51.000000		no	yes
218	Female	21.0	1.500000	42.300000		yes	no
219	Female	21.0	1.600000	68.000000		yes	yes
220	Female	21.0	1.750000	78.000000		yes	no
222	Female	21.0	1.720000	66.500000		yes	yes
231	Female	21.0	1.630000	66.000000		yes	yes
236	Female	21.0	1.660000	59.000000		no	yes
244	Female	21.0	1.540000	49.000000		yes	no
286	Female	21.0	1.600000	61.000000		no	yes
296	Female	21.0	1.530000	53.000000		no	yes
325	Female	21.0	1.550000	58.000000		no	yes
346	Female	21.0	1.540000	47.000000		yes	no
373	Female	21.0	1.540000	56.000000		no	yes
523	Female	21.0	1.520000	42.000000		no	yes
527	Female	21.0	1.520000	42.000000		no	yes
659	Female	21.0	1.520000	42.000000		no	yes
663	Female	21.0	1.520000	42.000000		no	yes
774	Female	21.0	1.754813	77.929204		yes	yes
846	Female	21.0	1.754497	77.956921		yes	yes
847	Female	21.0	1.752944	77.965532		yes	yes
900	Female	21.0	1.618148	68.981403		yes	yes
937	Female	21.0	1.753578	77.979170		yes	yes

	FCVC	NCP	CAEC	SMOKE	CH20	SCC	FAF	TUE	\
0	2.000000	3.000000	Sometimes	no	2.000000	no	0.000000	1.000000	
1	3.000000	3.000000	Sometimes	yes	3.000000	yes	3.000000	0.000000	
11	2.000000	3.000000	Frequently	no	2.000000	yes	2.000000	1.000000	
35	2.000000	3.000000	Sometimes	no	2.000000	no	2.000000	2.000000	
37	2.000000	3.000000	Sometimes	no	1.000000	no	1.000000	0.000000	
39	2.000000	3.000000	Sometimes	no	3.000000	no	3.000000	0.000000	
40	2.000000	3.000000	Sometimes	no	2.000000	no	1.000000	0.000000	
42	1.000000	3.000000	Sometimes	no	1.000000	no	0.000000	0.000000	
45	2.000000	3.000000	Sometimes	no	1.000000	no	0.000000	1.000000	
49	2.000000	3.000000	Sometimes	no	2.000000	no	0.000000	0.000000	
50	3.000000	3.000000	Sometimes	no	3.000000	no	0.000000	1.000000	
60	2.000000	3.000000	Sometimes	no	3.000000	no	3.000000	1.000000	
64	2.000000	3.000000	Frequently	no	1.000000	no	1.000000	1.000000	
65	1.000000	3.000000	Frequently	no	2.000000	no	0.000000	1.000000	
97	3.000000	1.000000	Frequently	no	1.000000	no	0.000000	0.000000	
98	3.000000	1.000000	Frequently	no	1.000000	no	0.000000	0.000000	
100	3.000000	1.000000	Sometimes	no	1.000000	no	0.000000	0.000000	
102	2.000000	4.000000	Frequently	no	2.000000	yes	2.000000	0.000000	
109	3.000000	1.000000	Sometimes	no	3.000000	no	2.000000	1.000000	
162	3.000000	3.000000	Always	yes	2.000000	no	2.000000	0.000000	
211	2.000000	1.000000	Sometimes	no	1.000000	no	1.000000	1.000000	

218	1.000000	1.000000	Sometimes	no	2.000000	no	3.000000	0.000000
219	2.000000	3.000000	Sometimes	no	3.000000	no	1.000000	0.000000
220	2.000000	3.000000	Frequently	no	2.000000	yes	0.000000	2.000000
222	3.000000	4.000000	Always	no	3.000000	no	0.000000	2.000000
231	3.000000	1.000000	Sometimes	yes	3.000000	no	0.000000	0.000000
236	1.000000	3.000000	Always	no	2.000000	yes	3.000000	0.000000
244	2.000000	1.000000	Sometimes	no	2.000000	yes	2.000000	0.000000
286	2.000000	3.000000	Sometimes	no	1.000000	no	1.000000	1.000000
296	3.000000	3.000000	Sometimes	no	2.000000	no	2.000000	1.000000
325	2.000000	1.000000	Sometimes	no	1.000000	no	1.000000	0.000000
346	3.000000	3.000000	Always	no	1.000000	no	2.000000	0.000000
373	2.000000	1.000000	Sometimes	no	2.000000	no	0.000000	2.000000
523	3.000000	1.000000	Frequently	no	1.000000	no	0.000000	0.000000
527	3.000000	1.000000	Frequently	no	1.000000	no	0.000000	0.000000
659	3.000000	1.000000	Frequently	no	1.000000	no	0.000000	0.000000
663	3.000000	1.000000	Frequently	no	1.000000	no	0.000000	0.000000
774	2.915279	1.104642	Sometimes	no	1.530046	no	1.360635	0.000000
846	2.446872	2.372311	Sometimes	no	1.810310	no	0.094893	1.650778
847	2.839048	2.106010	Sometimes	no	1.639202	no	1.117311	0.967919
900	1.142468	3.000000	no	no	2.197732	no	0.827506	0.572877
937	2.273548	2.390070	Sometimes	no	1.648404	no	0.874643	1.102696

	CALC	MTRANS	NObeyesdad
0	no	Public_Transportation	Normal_Weight
1	Sometimes	Public_Transportation	Normal_Weight
11	Sometimes	Public_Transportation	Overweight_Level_II
35	Sometimes	Public_Transportation	Overweight_Level_II
37	Sometimes	Public_Transportation	Normal_Weight
39	Sometimes	Public_Transportation	Overweight_Level_II
40	Sometimes	Public_Transportation	Overweight_Level_I
42	no	Public_Transportation	Normal_Weight
45	no	Public_Transportation	Overweight_Level_II
49	Sometimes	Public_Transportation	Normal_Weight
50	Sometimes	Walking	Normal_Weight
60	Sometimes	Public_Transportation	Normal_Weight
64	no	Public_Transportation	Normal_Weight
65	no	Public_Transportation	Overweight_Level_I
97	Sometimes	Public_Transportation	Insufficient_Weight
98	Sometimes	Public_Transportation	Insufficient_Weight
100	Sometimes	Public_Transportation	Normal_Weight
102	Sometimes	Automobile	Normal_Weight
109	no	Public_Transportation	Obesity_Type_I
162	Sometimes	Public_Transportation	Normal_Weight
211	no	Public_Transportation	Normal_Weight
218	no	Public_Transportation	Normal_Weight
219	Sometimes	Public_Transportation	Overweight_Level_I
220	Frequently	Public_Transportation	Overweight_Level_I
222	no	Public_Transportation	Normal_Weight
231	Sometimes	Public_Transportation	Normal_Weight
236	no	Automobile	Normal_Weight
244	Sometimes	Public_Transportation	Normal_Weight
286	Sometimes	Public_Transportation	Normal_Weight
296	Sometimes	Public_Transportation	Normal_Weight
325	Sometimes	Public_Transportation	Normal_Weight
346	no	Public_Transportation	Normal_Weight
373	Sometimes	Public_Transportation	Normal_Weight
523	Sometimes	Public_Transportation	Insufficient_Weight
527	Sometimes	Public_Transportation	Insufficient_Weight
659	Sometimes	Public_Transportation	Insufficient_Weight
663	Sometimes	Public_Transportation	Insufficient_Weight
774	Sometimes	Public_Transportation	Overweight_Level_I
846	Sometimes	Public_Transportation	Overweight_Level_I
847	Sometimes	Public_Transportation	Overweight_Level_I
900	Sometimes	Public_Transportation	Overweight_Level_I
937	Sometimes	Public_Transportation	Overweight_Level_I

```
In [164]: print(df.describe()) # describe the data with count,mean,std,etc.
```

	Age	Height	Weight	FCVC	NCP	\
count	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000	
mean	24.312600	1.701677	86.586058	2.419043	2.685628	
std	6.345968	0.093305	26.191172	0.533927	0.778039	
min	14.000000	1.450000	39.000000	1.000000	1.000000	
25%	19.947192	1.630000	65.473343	2.000000	2.658738	
50%	22.777890	1.700499	83.000000	2.385502	3.000000	
75%	26.000000	1.768464	107.430682	3.000000	3.000000	
max	61.000000	1.980000	173.000000	3.000000	4.000000	
	CH20	FAF	TUE			
count	2111.000000	2111.000000	2111.000000			
mean	2.008011	1.010298	0.657866			
std	0.612953	0.850592	0.608927			
min	1.000000	0.000000	0.000000			
25%	1.584812	0.124505	0.000000			
50%	2.000000	1.000000	0.625350			
75%	2.477420	1.666678	1.000000			
max	3.000000	3.000000	2.000000			

```
In [165]: df.isnull()
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCF
0	False	False	False	False		False	False	False
1	False	False	False	False		False	False	False
2	False	False	False	False		False	False	False
3	False	False	False	False		False	False	False
4	False	False	False	False		False	False	False
...	...	...	...	...	...	...	...	...
2106	False	False	False	False		False	False	False
2107	False	False	False	False		False	False	False
2108	False	False	False	False		False	False	False
2109	False	False	False	False		False	False	False
2110	False	False	False	False		False	False	False

2111 rows × 17 columns

```
In [166]: print(df.isnull().values.any()) # any NaN value is present or not ?
```

False

```
In [167]: # Data Transformation
```

```
In [168]: categorical_cols = df.select_dtypes(include="object").columns
numerical_cols = df.select_dtypes(include=["int64", "float64"]).columns
```

```
In [169]: le = LabelEncoder()
```

```
for col in categorical_cols:
    df[col] = le.fit_transform(df[col])

print(df)
```

	Gender	Age	Height	Weight	family_history_with_overweight	\
0	0	21.000000	1.620000	64.000000		1
1	0	21.000000	1.520000	56.000000		1
2	1	23.000000	1.800000	77.000000		1
3	1	27.000000	1.800000	87.000000		0
4	1	22.000000	1.780000	89.800000		0
...	...	...	...	...		...
2106	0	20.976842	1.710730	131.408528		1
2107	0	21.982942	1.748584	133.742943		1
2108	0	22.524036	1.752206	133.689352		1
2109	0	24.361936	1.739450	133.346641		1
2110	0	23.664709	1.738836	133.472641		1

	FAVC	FCVC	NCP	CAEC	SMOKE	CH20	SCC	FAF	TUE	CALC	\
0	0	2.0	3.0	2	0	2.000000	0	0.000000	1.000000	3	
1	0	3.0	3.0	2	1	3.000000	1	3.000000	0.000000	2	
2	0	2.0	3.0	2	0	2.000000	0	2.000000	1.000000	1	
3	0	3.0	3.0	2	0	2.000000	0	2.000000	0.000000	1	
4	0	2.0	1.0	2	0	2.000000	0	0.000000	0.000000	2	
...	...	...	...	...	...	...	...	...	...	...	...
2106	1	3.0	3.0	2	0	1.728139	0	1.676269	0.906247	2	
2107	1	3.0	3.0	2	0	2.005130	0	1.341390	0.599270	2	
2108	1	3.0	3.0	2	0	2.054193	0	1.414209	0.646288	2	
2109	1	3.0	3.0	2	0	2.852339	0	1.139107	0.586035	2	
2110	1	3.0	3.0	2	0	2.863513	0	1.026452	0.714137	2	

	MTRANS	NObeyesdad
0	3	1
1	3	1
2	3	1
3	4	5
4	3	6
...	...	...
2106	3	4
2107	3	4
2108	3	4
2109	3	4
2110	3	4

[2111 rows x 17 columns]

```
In [170... X = df.drop("NObeyesdad", axis=1)
y = df["NObeyesdad"]
```

```
In [171... scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

```
In [172... X_scaled = pd.DataFrame(X_scaled, columns=X.columns)
```

```
In [173... X_scaled.head()
```

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC
0	-1.011914	-0.522124	-0.875589	-0.862558		0.472291 -2.759769
1	-1.011914	-0.522124	-1.947599	-1.168077		0.472291 -2.759769
2	0.988227	-0.206889	1.054029	-0.366090		0.472291 -2.759769
3	0.988227	0.423582	1.054029	0.015808		-2.117337 -2.759769
4	0.988227	-0.364507	0.839627	0.122740		-2.117337 -2.759769

In [ ]:

