Final Project Presentation

Data Analysis with Pandas
BMI of 500 People

Ankara University Artificial Intelligence Technology

Melisa Gözet
21823408
Data Analysis with Python
12^{th} January 2022

AGENDA

- Data Intake Report
- Data Analysis with Pandas
- Conclusion

Data Intake Report

Data Storage Location : https://www.kaggle.com/yersever/500-person-gender-height-weight-bodymassindex

Tabular Data Details : 500_Person_Gender_Height_Weight_Index.csv

Total number of observations	**500 Rows**
Total number of files	**1**
Total number of features	**4 Columns**
Base format of the file	**.csv**
Size of the data	**15.8+ KB**

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Gender  500 non-null    object
 1   Height  500 non-null    int64
 2   Weight  500 non-null    int64
 3   Index   500 non-null    int64
dtypes: int64(3), object(1)
memory usage: 15.8+ KB
```

```
[9] df.isnull().sum()
```

```
Gender    0
Height    0
Weight    0
Index     0
dtype: int64
```

```
df
```

	Gender	Height	Weight	Index
0	Male	174	96	4
1	Male	189	87	2
2	Female	185	110	4
3	Female	195	104	3
4	Male	149	61	3
...	...	...	...	...
495	Female	150	153	5
496	Female	184	121	4
497	Female	141	136	5
498	Male	150	95	5
499	Male	173	131	5

500 rows × 4 columns

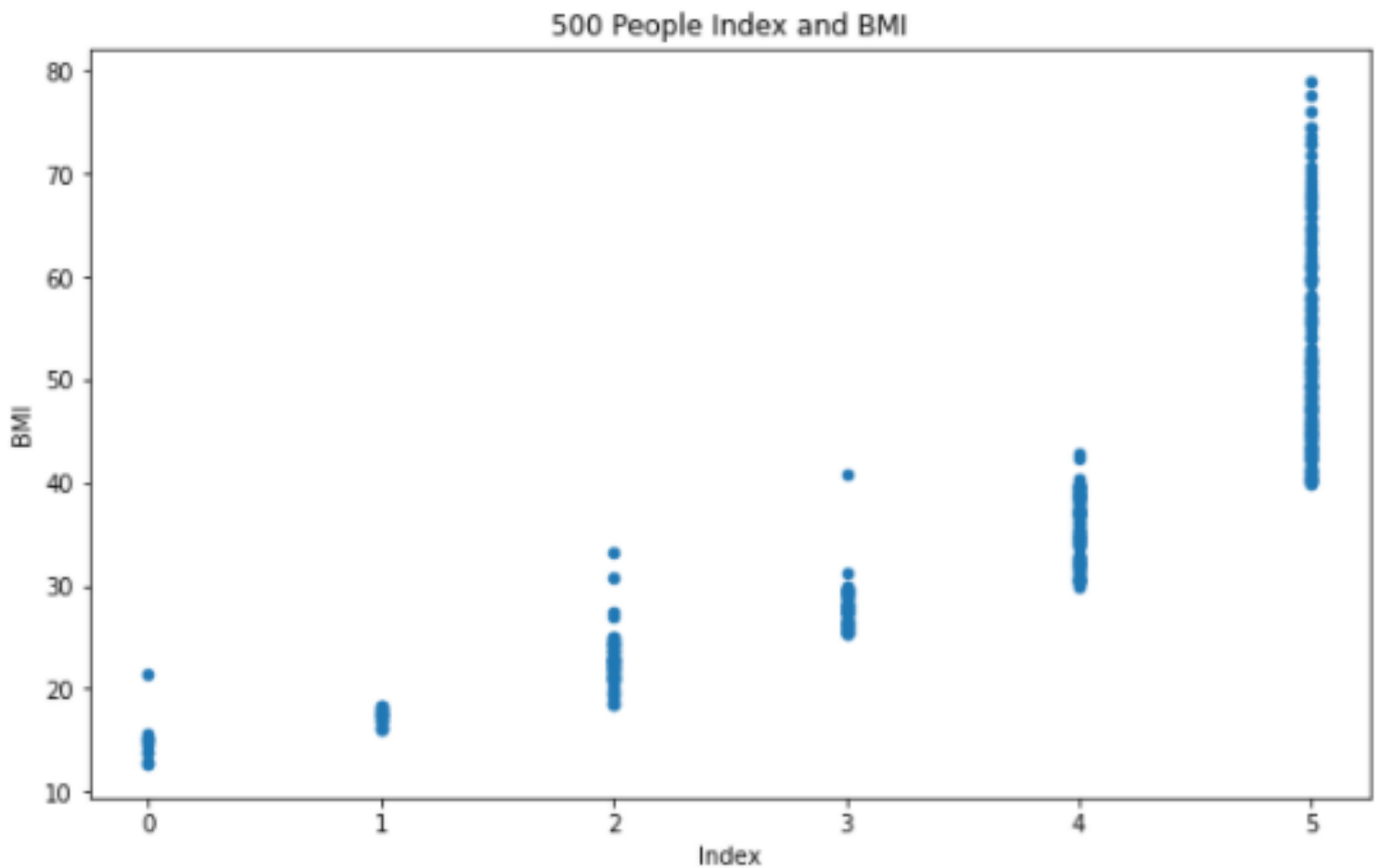
DATA DESCRIPTION

- **Gender : Male / Female**
- **Height : Number (cm)**
- **Weight : Number (kg)**
- **Index : [0 1 2 3 4 5]**

```python
#Making a BMI Column using the formula for BMI
df["BMI"] = round(df["Weight"]/((df["Height"])/100)**2, 2)    #Weight(Kg)/Height(m)^2
df.head()
```

	Gender	Height	Weight	Index	BMI
0	Male	174	96	4	31.71
1	Male	189	87	2	24.36
2	Female	185	110	4	32.14
3	Female	195	104	3	27.35
4	Male	149	61	3	27.48

```
df.plot(x="Index", y="BMI", kind="scatter",
        figsize=(10,6),
        title="500 People Index and BMI");
```



BMI	Nutritional status
Below 18.5	Underweight
18.5–24.9	Normal weight
25.0–29.9	Pre-obesity
30.0–34.9	Obesity class I
35.0–39.9	Obesity class II
Above 40	Obesity class III

```
def index(x):
    if x["BMI"] <18.5:
        return 0
    elif 18.5<= x["BMI"] < 25:
        return 1
    elif 25<= x["BMI"] < 30:
        return 2
    elif 30<= x["BMI"] < 35:
        return 3
    elif 35<= x["BMI"] < 40:
        return 4
    elif x["BMI"] >= 40:
        return 5
df["New_Index"] = df.apply(index, axis=1)
df.head()
```

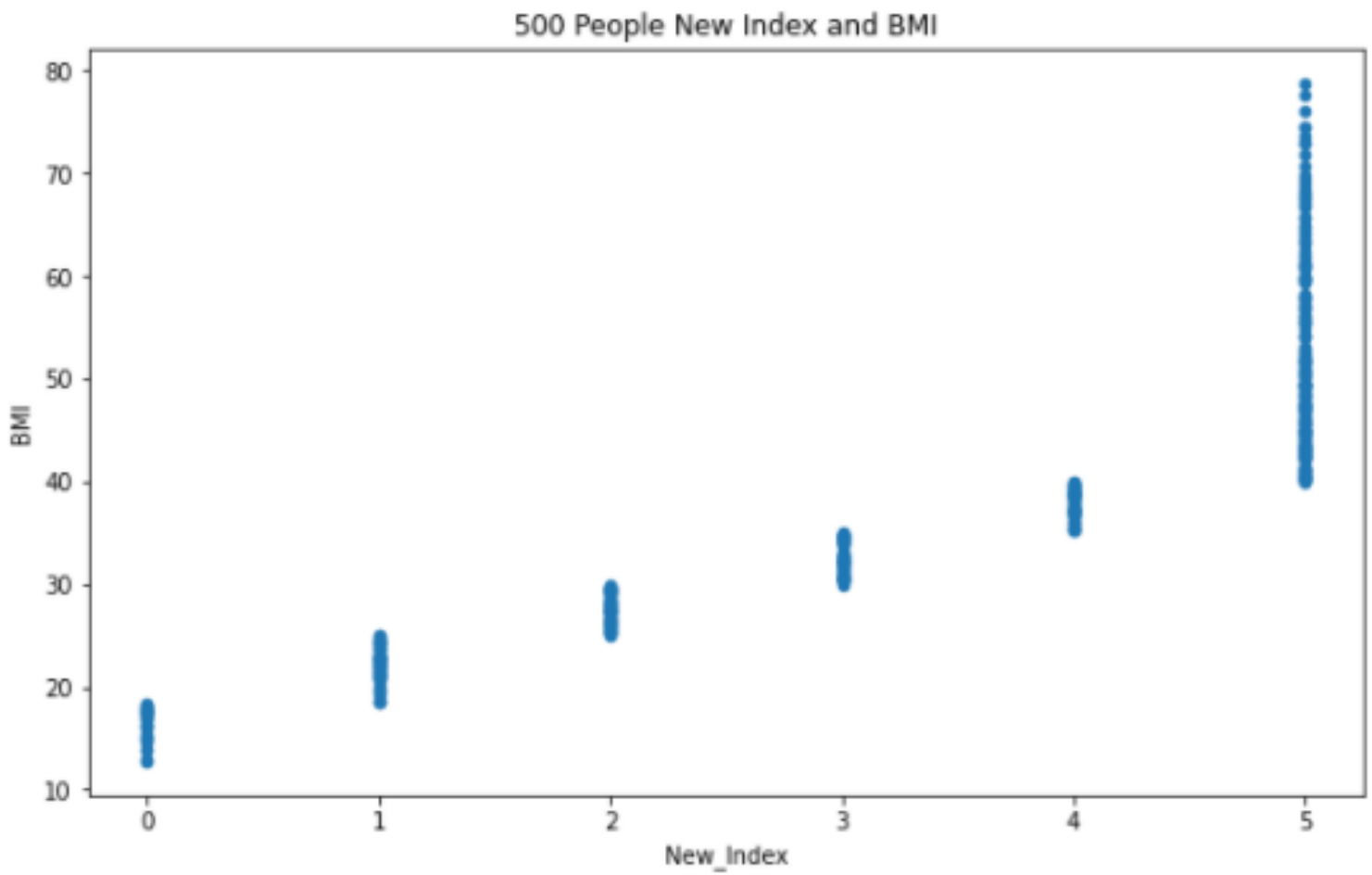
	Gender	Height	Weight	Index	BMI	New_Index
0	Male	174	96	4	31.71	3
1	Male	189	87	2	24.36	1
2	Female	185	110	4	32.14	3
3	Female	195	104	3	27.35	2
4	Male	149	61	3	27.48	2

https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi

```python
def status(x):
    if x["New_Index"] == 0:
        return "Underweight"
    elif x["New_Index"] == 1:
        return "Normal weight"
    elif x["New_Index"] == 2:
        return "Pre-Obesity"
    elif x["New_Index"] == 3:
        return "Obese Class I"
    elif x["New_Index"]== 4:
        return "Obese Class II"
    elif x["New_Index"] == 5:
        return "Obese Class III"
df["Status"] = df.apply(status, axis=1)
df.head()
```

	Gender	Height	Weight	BMI	New_Index	Status
0	Male	174	96	31.71	3	Obese Class I
1	Male	189	87	24.36	1	Normal weight
2	Female	185	110	32.14	3	Obese Class I
3	Female	195	104	27.35	2	Pre-Obesity
4	Male	149	61	27.48	2	Pre-Obesity

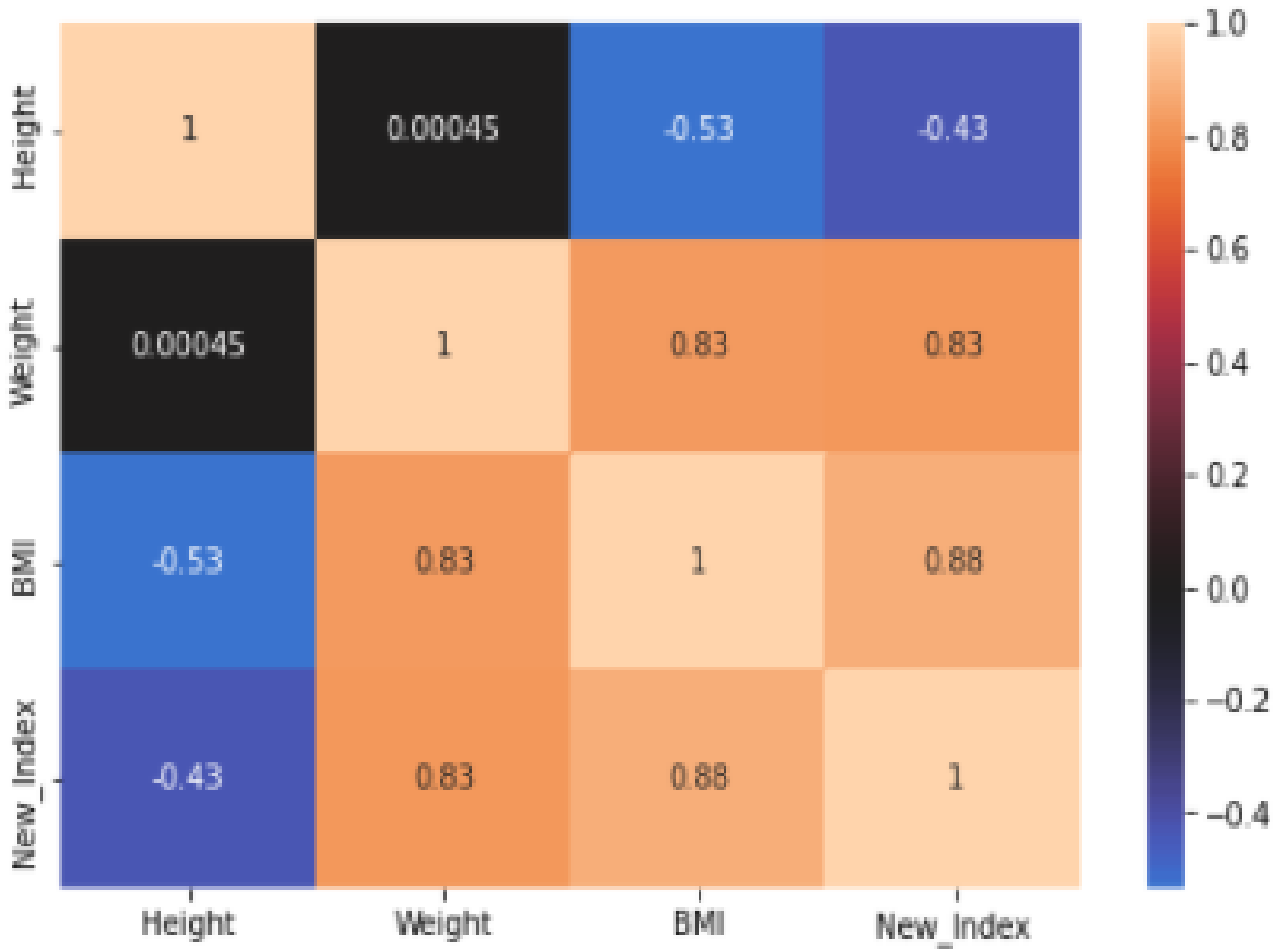
```
df.plot(x="New_Index", y="BMI", kind="scatter",
        figsize=(10,6),
        title="500 People New Index and BMI");
```



```
df.describe().T
```

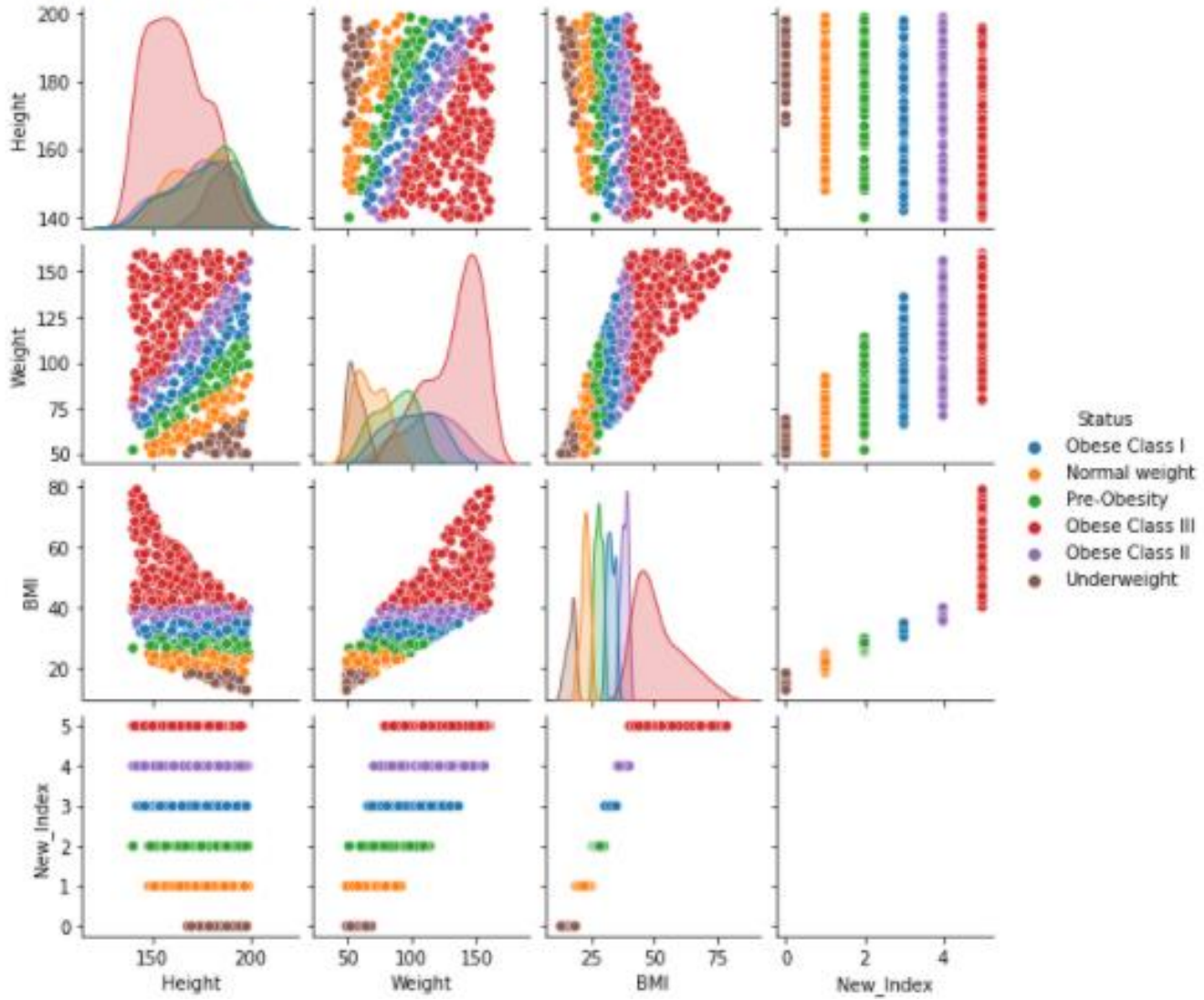
	count	mean	std	min	25%	50%	75%	max
Height	500.0	169.94400	16.375261	140.00	156.0000	170.500	184.0000	199.00
Weight	500.0	106.00000	32.382607	50.00	80.0000	106.000	136.0000	160.00
BMI	500.0	37.76572	13.965550	12.75	27.1675	36.955	46.3875	78.85
New_Index	500.0	3.33600	1.705556	0.00	2.0000	4.000	5.0000	5.00

```
plt.subplots(figsize=(8,5))
sns.heatmap(df.corr(), annot = True, center = 0);
```



```
sns.pairplot(df, hue='Status', size=2);
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:2076: UserWarning: The `size` p
  warnings.warn(msg, UserWarning)
```



New_Index

```
[33] df[df["New_Index"].isna()]
```

Gender	Height	Weight	BMI	New_Index	Status

```
print(df["New_Index"].value_counts())
print()
print(df["New_Index"].value_counts(normalize = True)*100)
```

```
5    202
2     69
4     65
3     65
1     65
0     34
Name: New_Index, dtype: int64

5    40.4
2    13.8
4    13.0
3    13.0
1    13.0
0     6.8
Name: New_Index, dtype: float64
```

Status

```
[52] df[df["Status"].isna()]
```

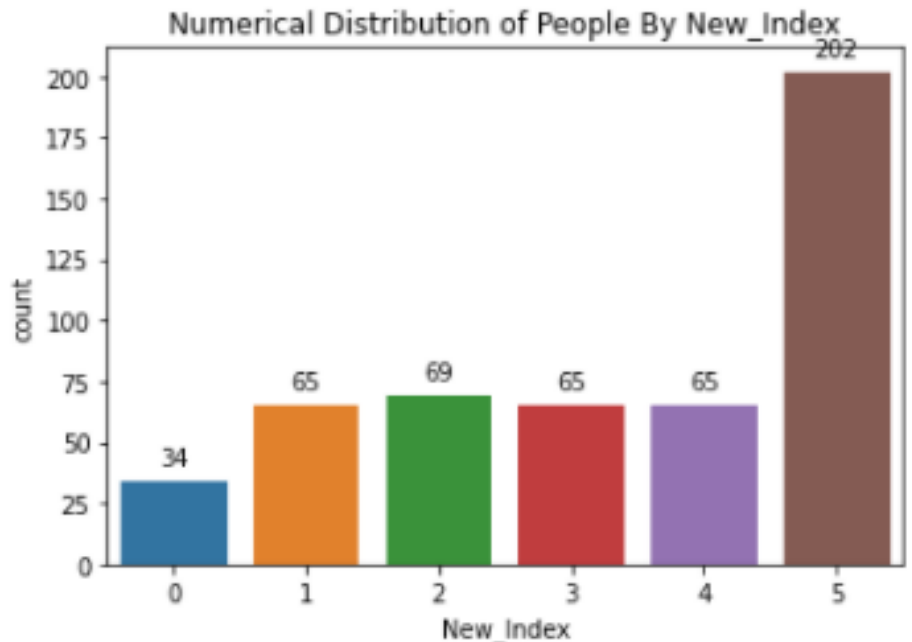
Gender	Height	Weight	BMI	New_Index	Status

```
print(df["Status"].value_counts())
print()
print(df["Status"].value_counts(normalize = True)*100)
```

```
Obese Class III    202
Pre-Obesity         69
Normal weight       65
Obese Class II      65
Obese Class I       65
Underweight         34
Name: Status, dtype: int64

Obese Class III    40.4
Pre-Obesity        13.8
Normal weight      13.0
Obese Class II     13.0
Obese Class I      13.0
Underweight         6.8
Name: Status, dtype: float64
```

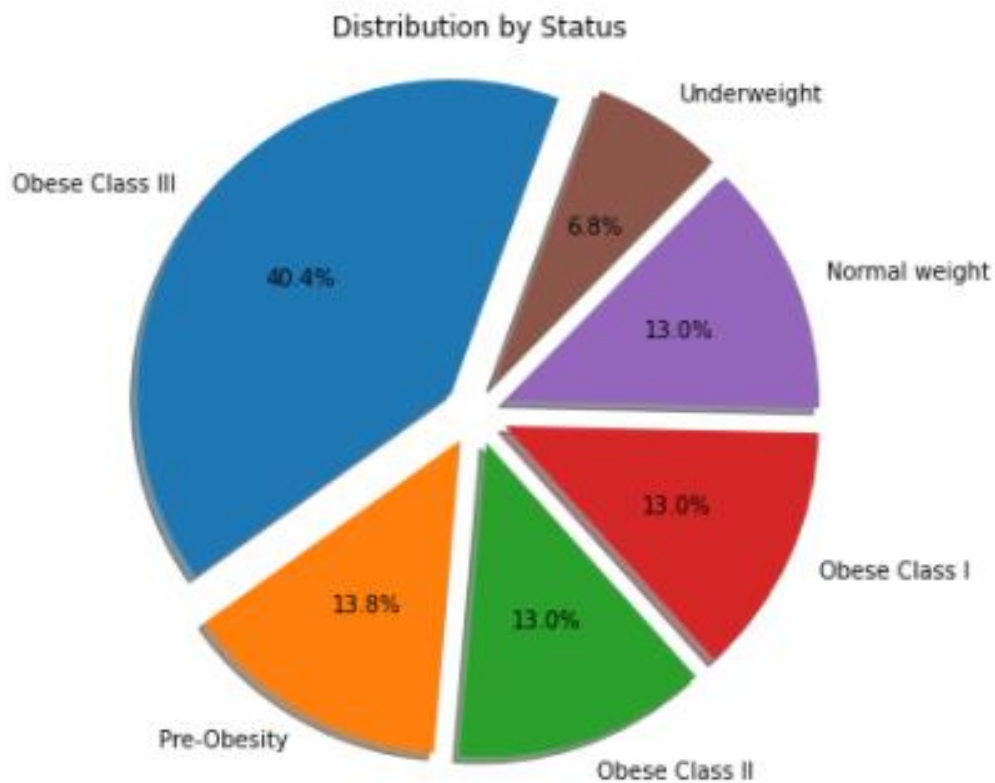
```
ax = sns.countplot(df['New_Index']);

for p in ax.patches:
    ax.annotate(format(p.get_height()), (p.get_x() + p.get_width() / 2., p.get_height()),
       ha = 'center', va = 'center', xytext = (0, 10), textcoords = 'offset points')

ax.set_title("Numerical Distribution of People By New_Index");
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the f
  FutureWarning
```



```
print(df["Status"].value_counts())
print()

label =  "Obese Class III", "Pre-Obesity", "Obese Class II", "Obese Class I","Normal weight", "Underweight"
explode = (0.1, 0.1, 0.1, 0.1, 0.1, 0.1)
sizes = df["Status"].value_counts(normalize = True)*100

fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels = label, autopct="%1.1f%%", shadow=True, startangle=70)
ax1.axis("equal")
plt.title("Distribution by Status")
plt.rcParams['figure.figsize'] = [6, 6]
plt.show();
```

```
Obese Class III    202
Pre-Obesity         69
Normal weight       65
Obese Class II      65
Obese Class I       65
Underweight         34
Name: Status, dtype: int64
```



Gender

```
[37] df[df["Gender"].isna()]
```

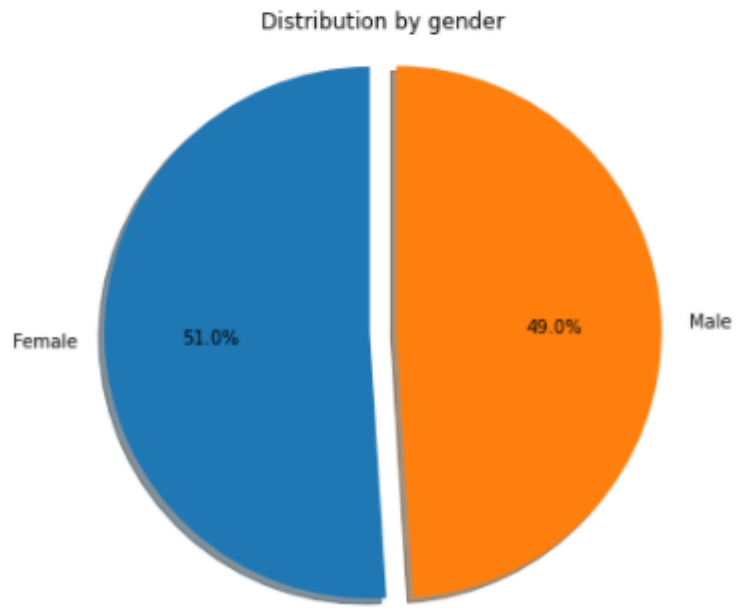
	Gender	Height	Weight	BMI	New_Index	Status

```
[38] print(df["Gender"].value_counts())
     print()
     print(df["Gender"].value_counts(normalize = True)*100)
```

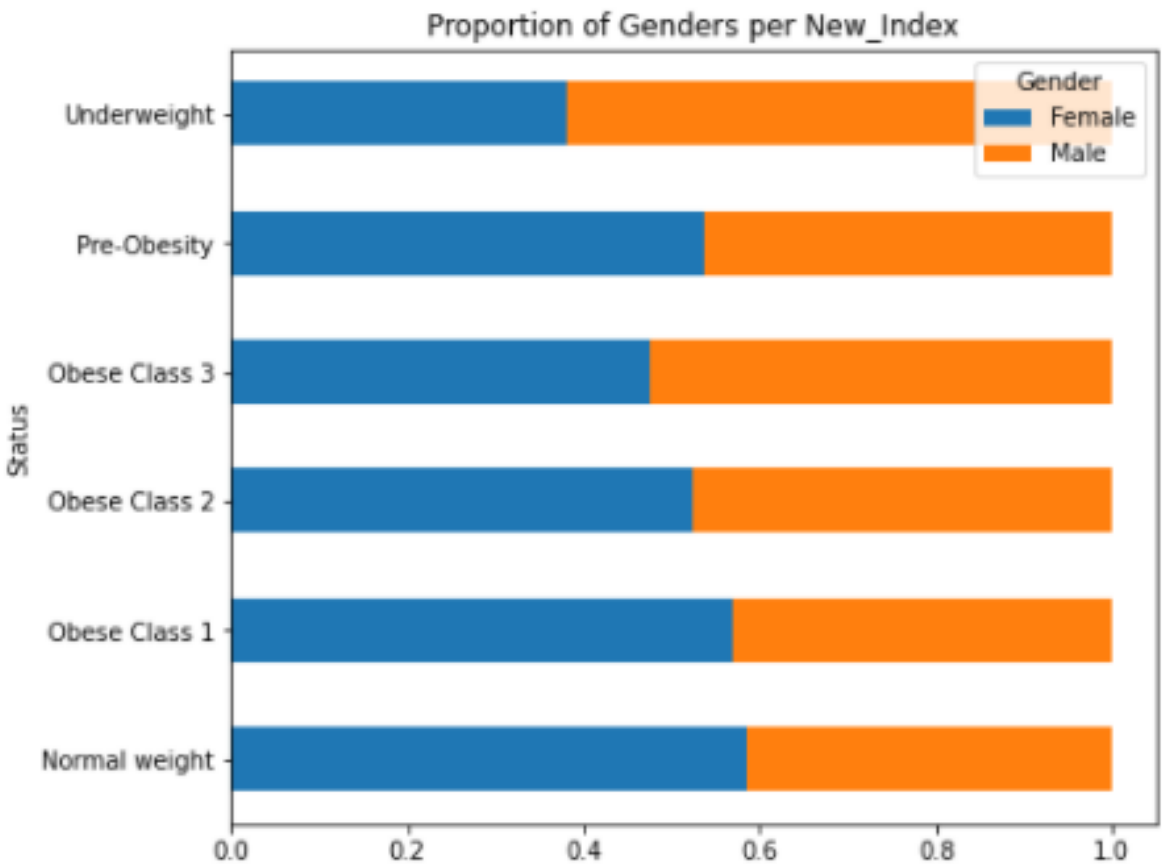
```
Female    255
Male      245
Name: Gender, dtype: int64

Female    51.0
Male      49.0
Name: Gender, dtype: float64
```

```
print(df["Gender"].value_counts())
print()

label =  "Female", "Male"
explode = (0, 0.1)
sizes = df["Gender"].value_counts(normalize = True)*100

fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels = label, autopct='%1.1f%%', shadow=True, startangle=90)
ax1.axis("equal")
plt.title("Distribution by gender")
plt.rcParams['figure.figsize'] = [6, 6]
plt.show();
```

```
Female    255
Male      245
Name: Gender, dtype: int64
```

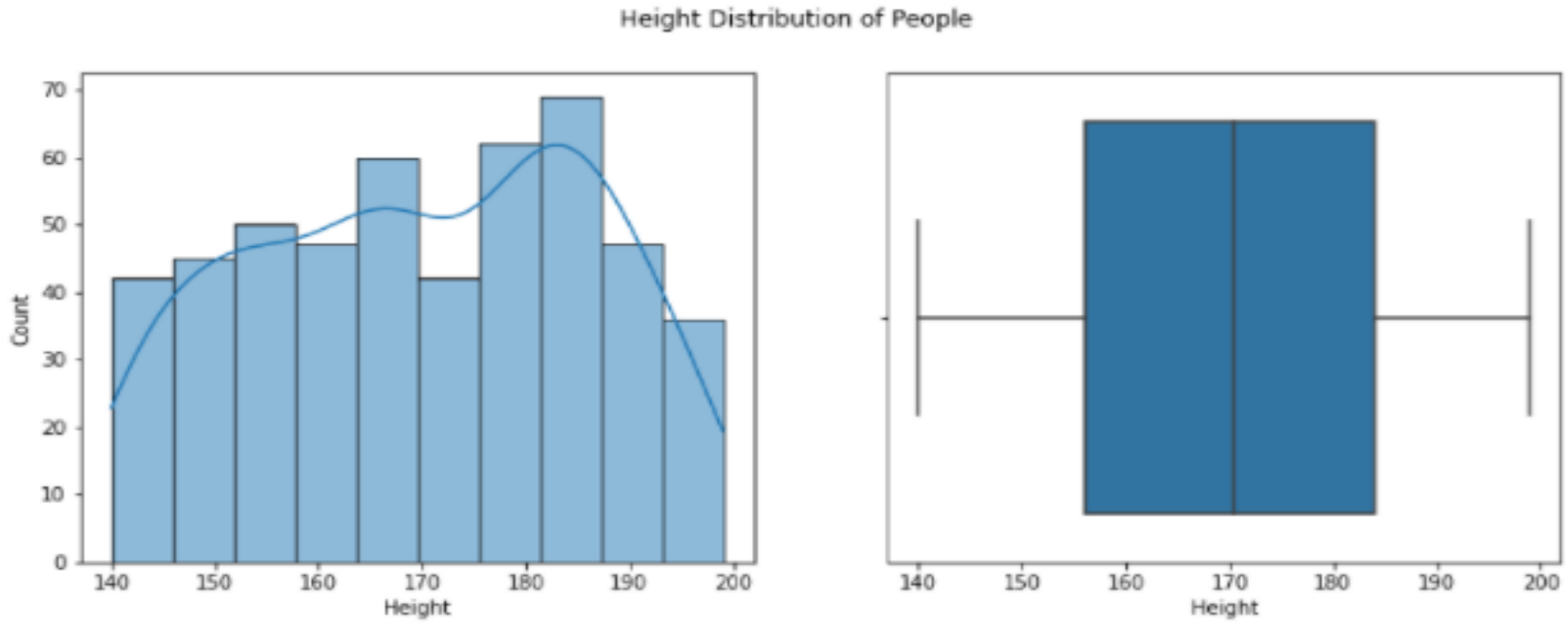


```
pd.crosstab(index = df["Status"], columns = df["Gender"], normalize = "index").\
plot(kind = "barh", stacked = True, figsize = (7,6), title = "Proportion of Genders per New_Index");
```



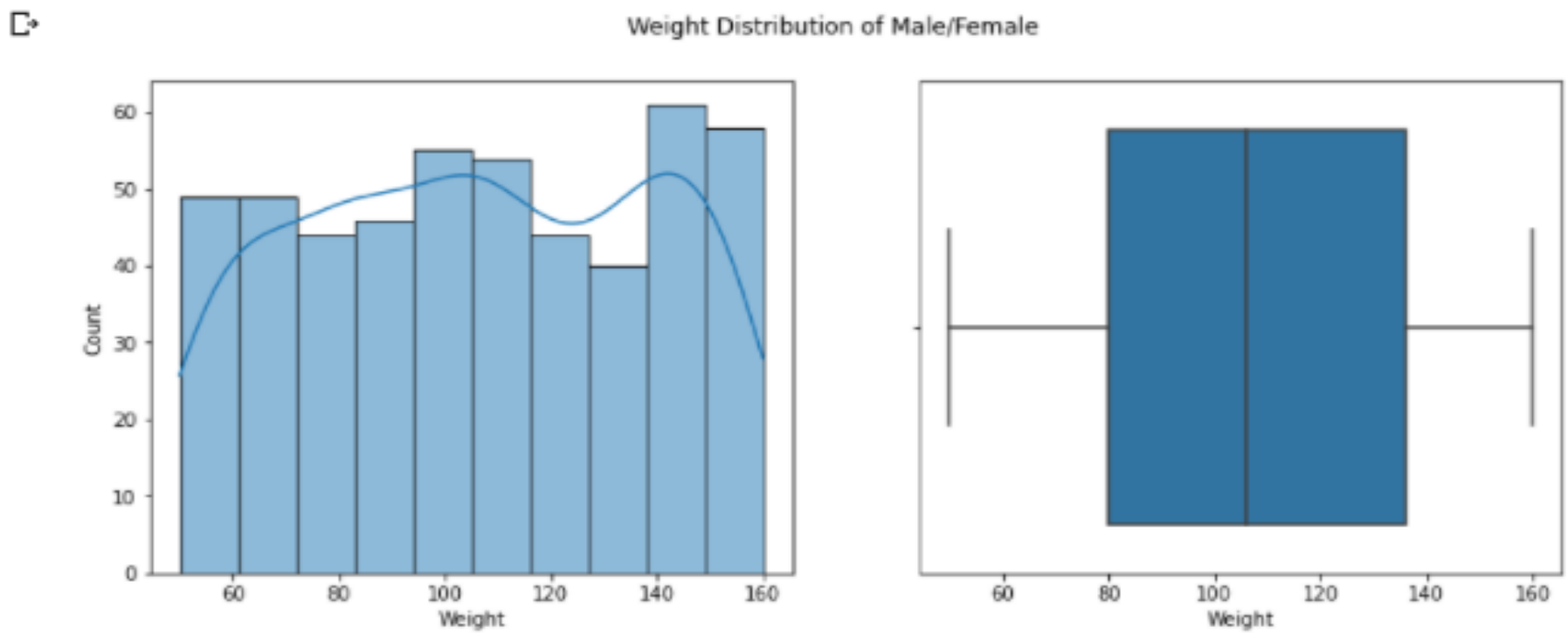
Height

```
fig, ax = plt.subplots(1,2, figsize = (13,5));

sns.histplot(df["Height"], kde = True, ax = ax[0]);
sns.boxplot(x = "Height", data = df, ax = ax[1]);

plt.suptitle("Height Distribution of People");
```



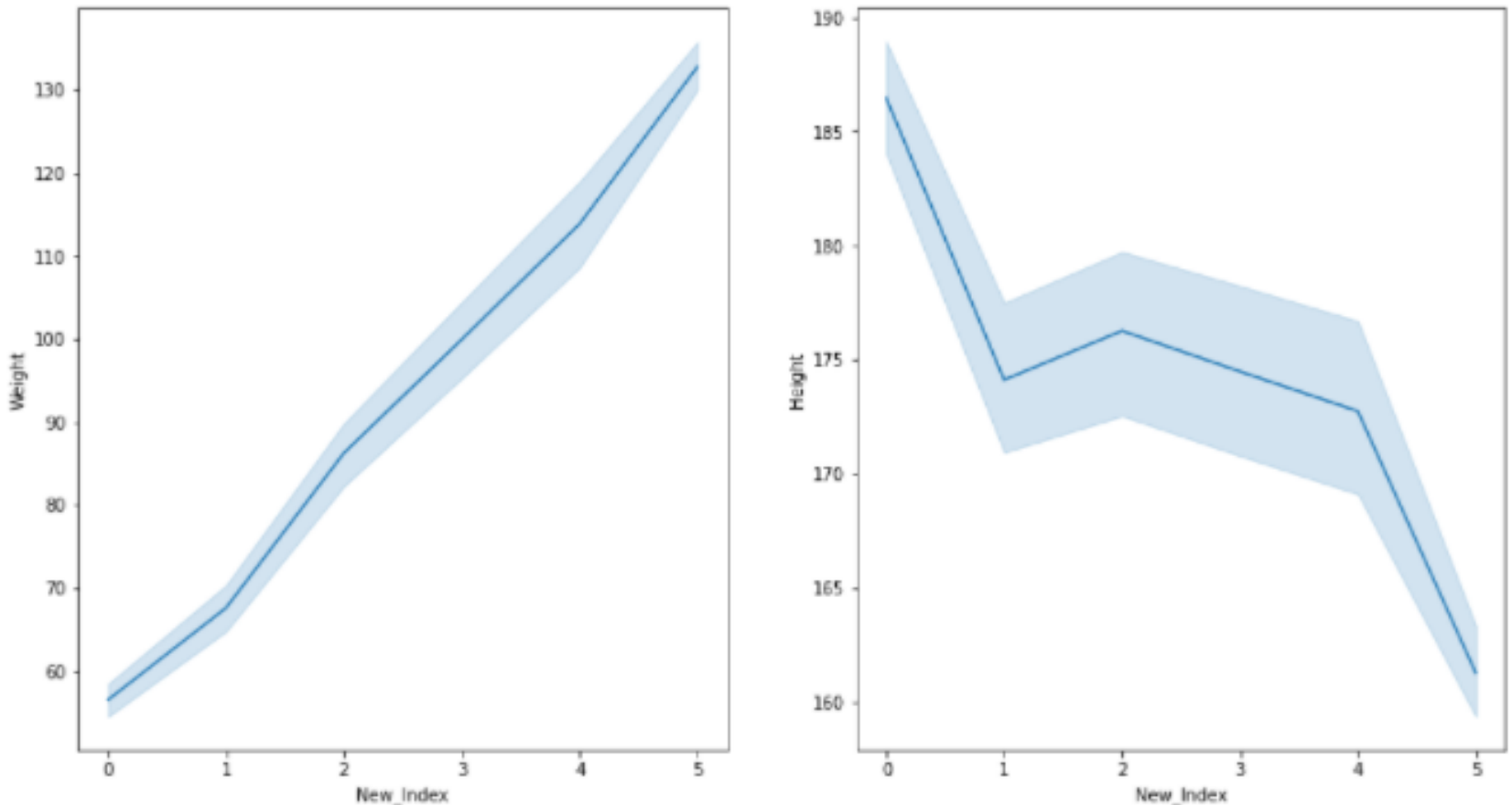
Weight

```
fig, ax = plt.subplots(1,2, figsize = (13,5));

sns.histplot(df["Weight"], kde = True, ax = ax[0]);
sns.boxplot(x = "Weight", data = df, ax = ax[1]);

plt.suptitle("Weight Distribution of People");
```

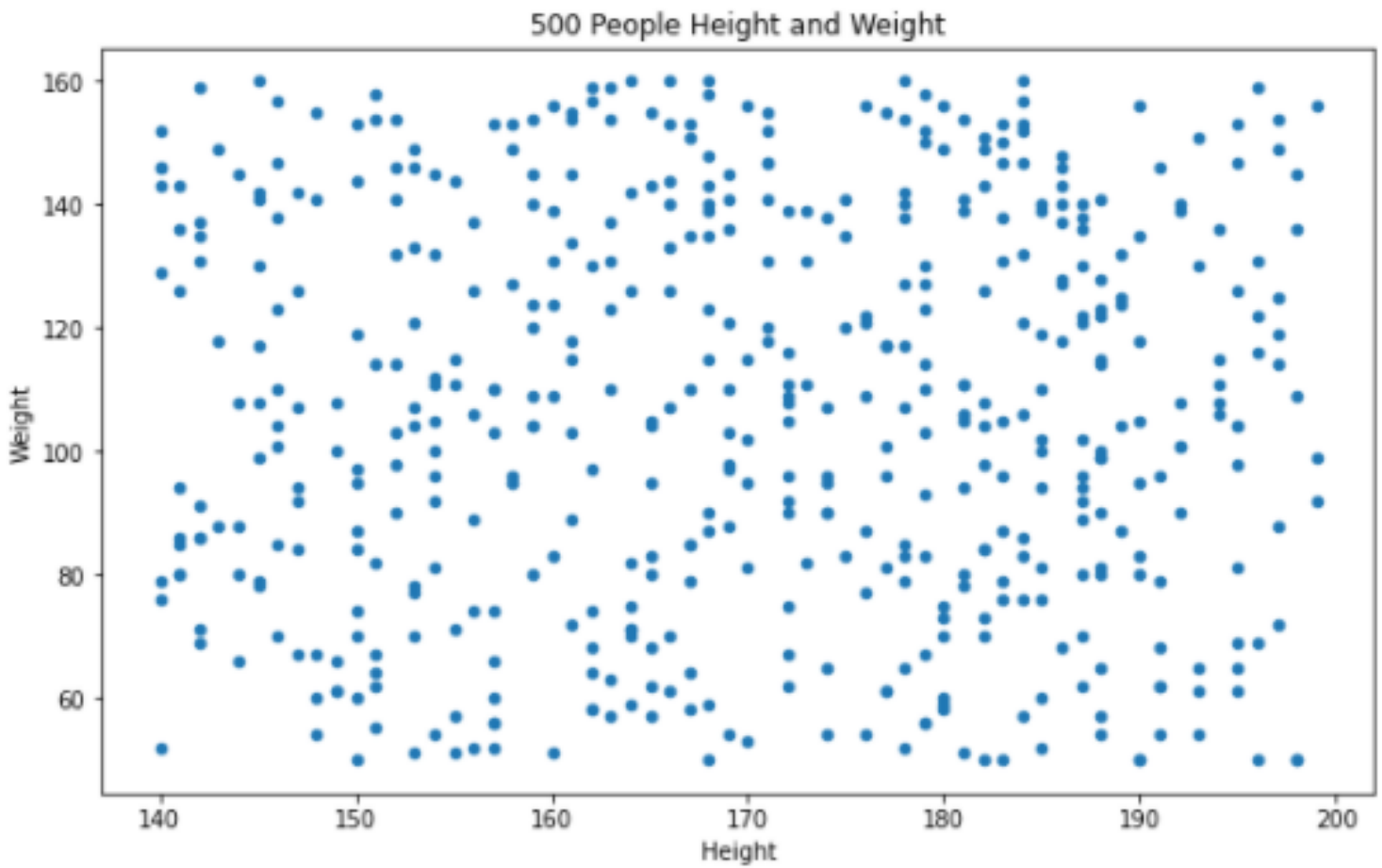


```
fig, ax= plt.subplots(1,2, figsize=(15,8))
sns.lineplot(df['New_Index'],df['Weight'], ax=ax[0])
sns.lineplot(df['New_Index'],df['Height'], ax=ax[1])
plt.show();
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keywor
  FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keywor
  FutureWarning
```

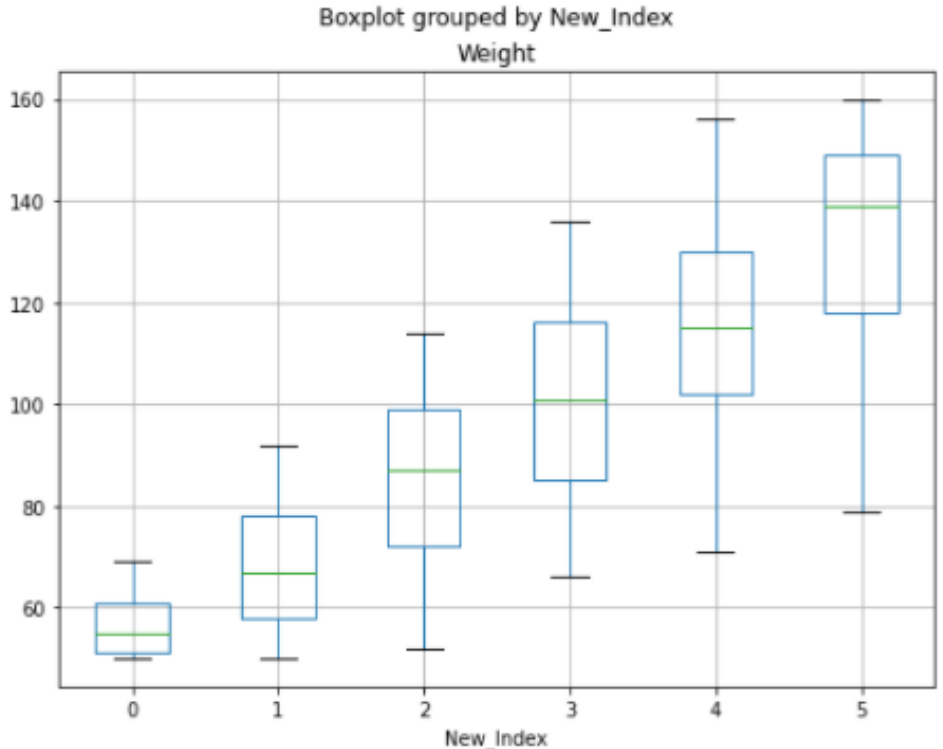


```
df.plot(x='Height', y='Weight', kind='scatter',
        figsize=(10,6),
        title='500 People Height and Weight');
```



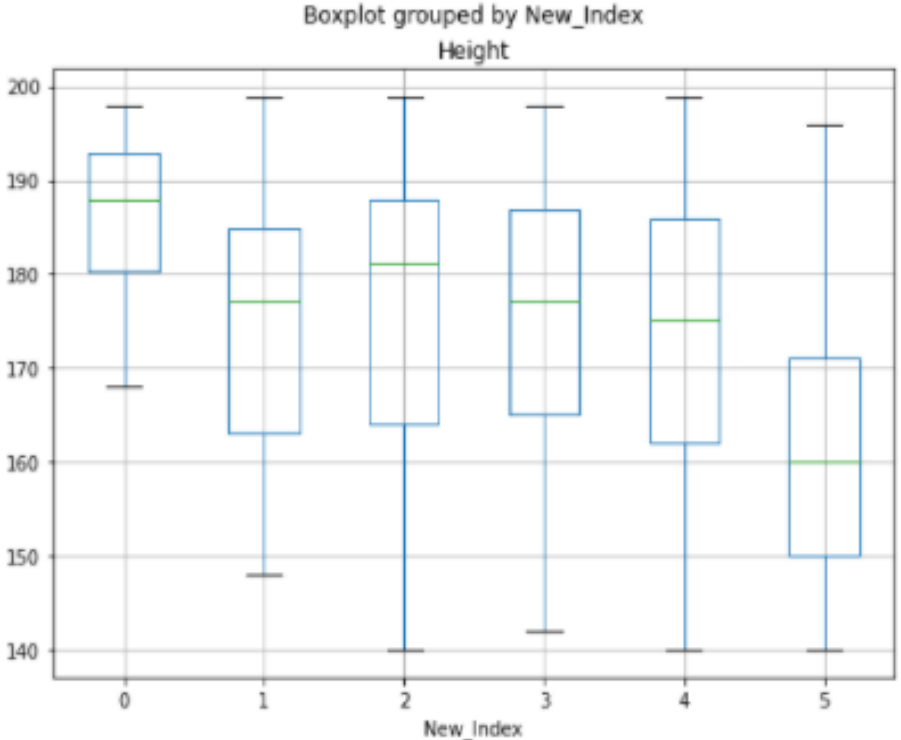
```
df.boxplot(column="Weight", by="New_Index", figsize=(8,6));
```

```
/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py:83: Visit
  return array(a, dtype, copy=False, order=order)
```



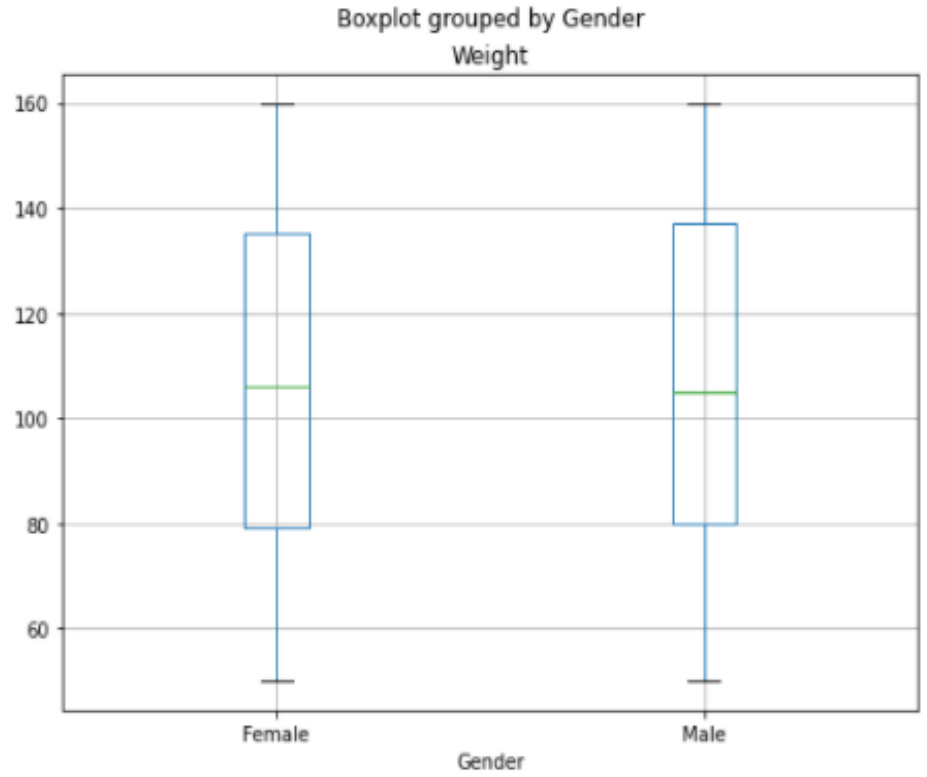

```
df.boxplot(column="Height", by="New_Index", figsize=(8,6));
```

```
/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py:83: VisibleD
  return array(a, dtype, copy=False, order=order)
```



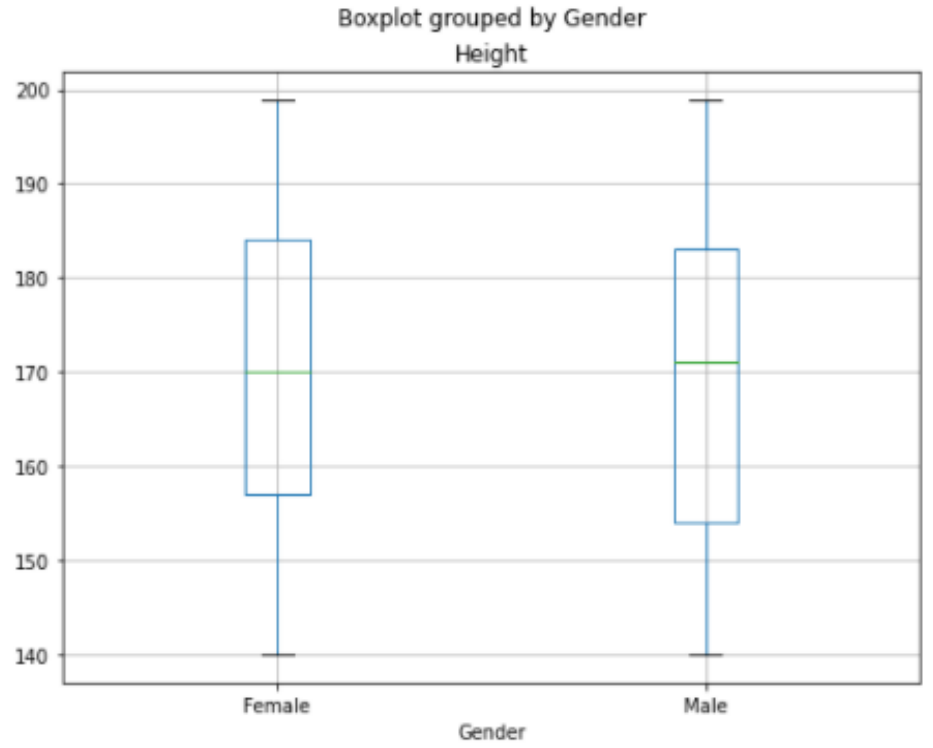
```
df.boxplot(column="Weight", by="Gender", figsize=(8,6));
```

```
/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py:83:
  return array(a, dtype, copy=False, order=order)
```

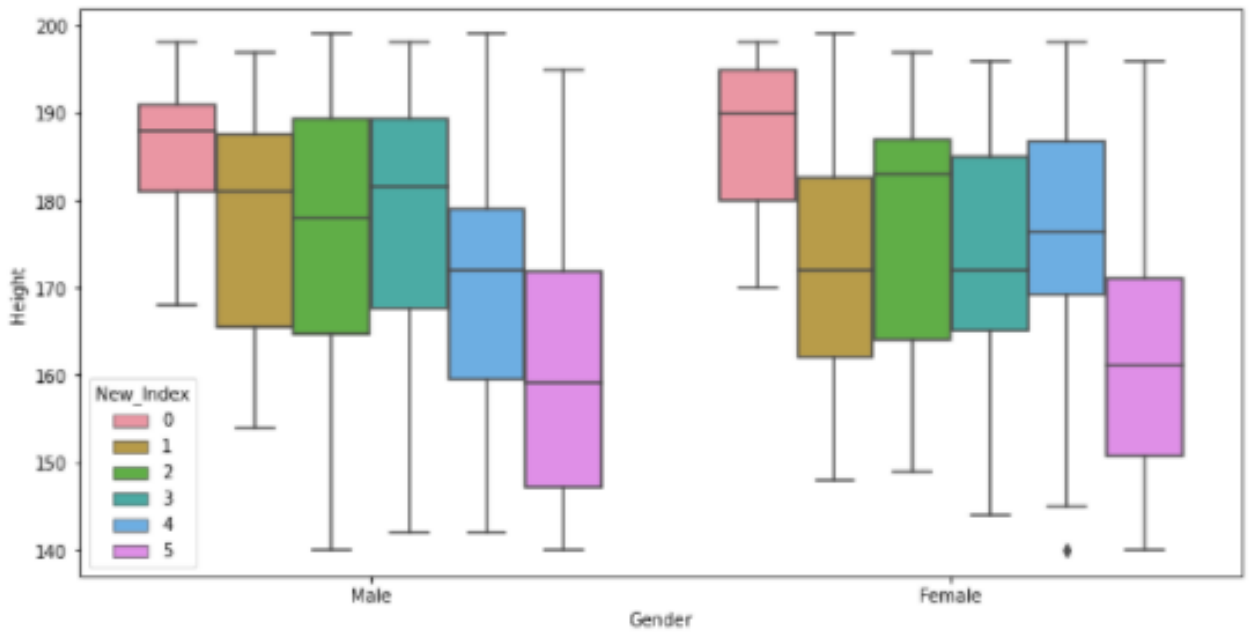



```
df.boxplot(column="Height", by="Gender", figsize=(8,6));
```

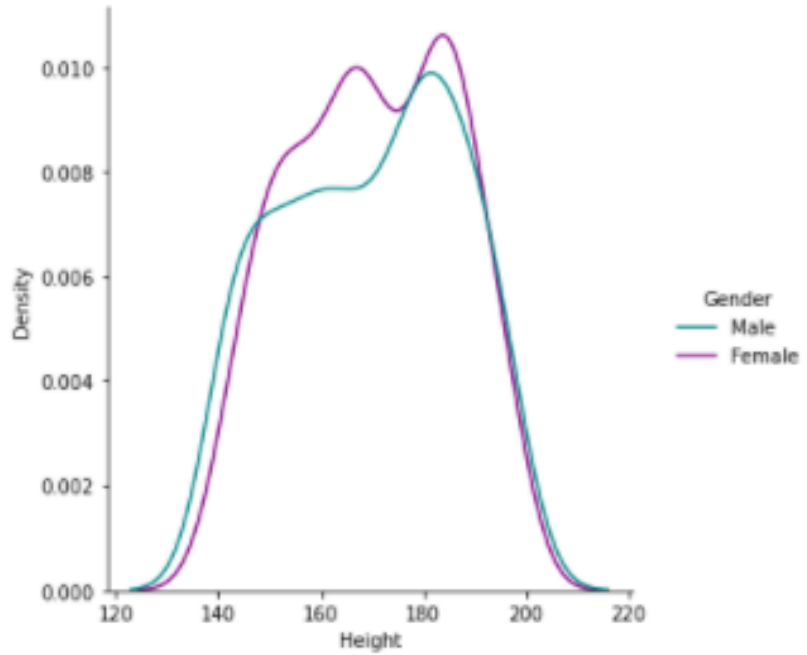
```
/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py:83:
  return array(a, dtype, copy=False, order=order)
```



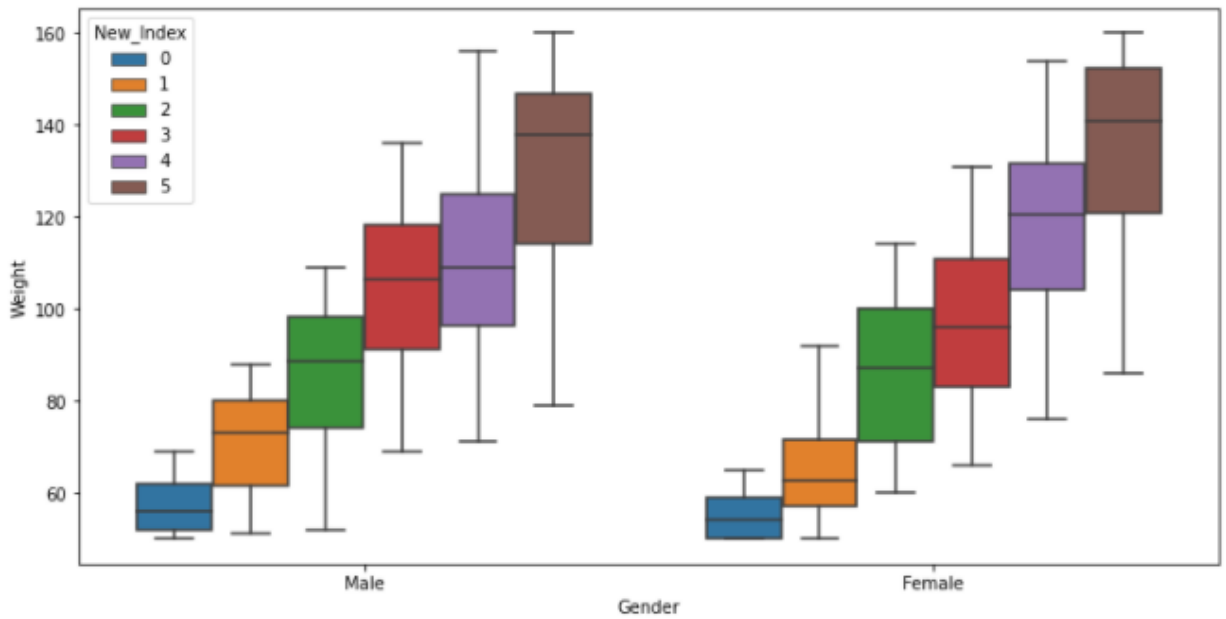
```
fig = plt.figure(figsize=(12, 6))
rows = 1
columns = 1


fig.add_subplot(rows, columns, 1)
sns.boxplot(data=df, x=df["Gender"], y="Height",  hue = df["New_Index"]);
```



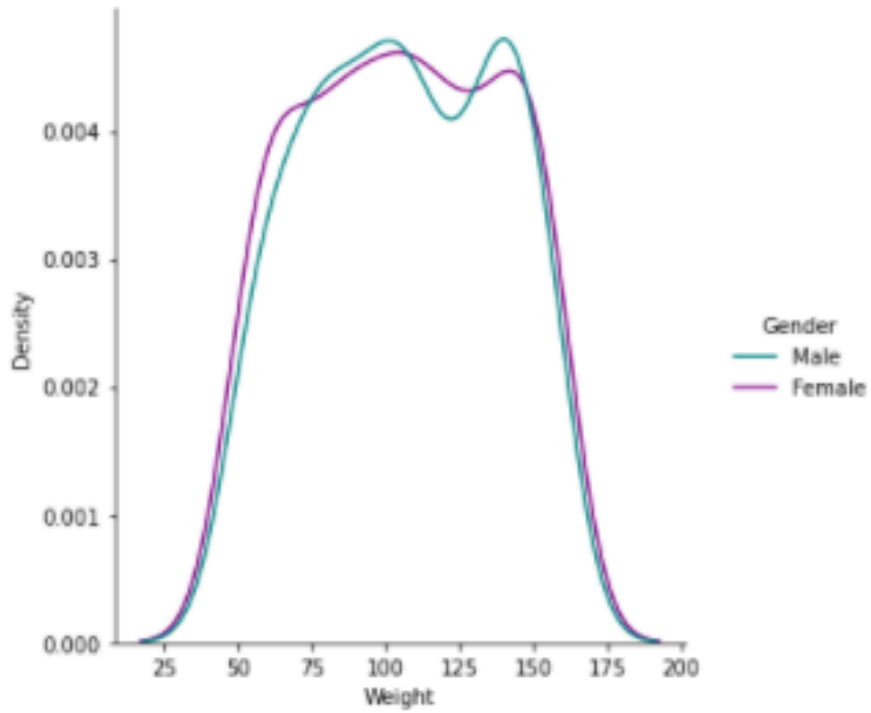

```
color = ["teal", "purple"]
sns.set_palette(sns.color_palette(color))
sns.displot(data=df,x="Height",kind="kde",hue="Gender");
```



```
fig = plt.figure(figsize=(12, 6))
rows = 1
columns = 1


fig.add_subplot(rows, columns, 1)
sns.boxplot(data=df, x=df["Gender"], y="Weight",  hue = df["New_Index"]);
```




```
color = ["teal", "purple"]
sns.set_palette(sns.color_palette(color))
sns.displot(data=df,x="Weight",kind="kde",hue="Gender");
```



```
df1 = df.groupby(["Gender", "New_Index","Status"]).size().reset_index().rename(columns = {0 : 'count'})
df1
```

	Gender	New_Index	Status	count
0	Female	0	Underweight	13
1	Female	1	Normal weight	38
2	Female	2	Pre-Obesity	37
3	Female	3	Obese Class I	37
4	Female	4	Obese Class II	34
5	Female	5	Obese Class III	96
6	Male	0	Underweight	21
7	Male	1	Normal weight	27
8	Male	2	Pre-Obesity	32
9	Male	3	Obese Class I	28
10	Male	4	Obese Class II	31
11	Male	5	Obese Class III	106

```
gender = df.groupby(["Gender", "New_Index", "Status"]).mean()
gender
```

Gender	New_Index	Status	Height	Weight	BMI
Female	0	Underweight	187.076923	55.153846	15.840769
	1	Normal weight	171.894737	65.000000	21.913158
	2	Pre-Obesity	177.027027	86.702703	27.462973
	3	Obese Class I	172.405405	97.324324	32.418108
	4	Obese Class II	175.411765	116.823529	37.696471
	5	Obese Class III	161.989583	135.260417	52.079792
Male	0	Underweight	186.095238	57.380952	16.578571
	1	Normal weight	177.259259	71.259259	22.582222
	2	Pre-Obesity	175.406250	85.750000	27.671250
	3	Obese Class I	177.250000	103.428571	32.641071
	4	Obese Class II	169.838710	110.677419	38.000645
	5	Obese Class III	160.650943	130.632075	51.056321

```
df.groupby("New_Index").agg([np.mean,np.median])
```

	Height		Weight		BMI	
	mean	median	mean	median	mean	median
New_Index						
0	186.470588	188	56.529412	55	16.296471	16.92
1	174.123077	177	67.600000	67	22.191077	22.22
2	176.275362	181	86.260870	87	27.559565	27.47
3	174.492308	177	99.953846	101	32.514154	32.36
4	172.753846	175	113.892308	115	37.841538	38.04
5	161.287129	160	132.831683	139	51.542723	48.92

```
df.groupby("Gender").agg([np.mean,np.median])
```

	Height		Weight		BMI		New_Index	
	mean	median	mean	median	mean	median	mean	median
Gender								
Female	170.227451	170	105.698039	106	37.394392	35.74	3.290196	4
Male	169.648980	171	106.314286	105	38.152204	38.39	3.383673	4

Conclusion

- **People who are very tall are generally Underweight (Index 0) and people with shorter height are generally Obese Class III (Index 5).**
- **The mean weight of males is heavier than that of females.**
- **The mean height of males is slighly less than that of females.**
- **The distributions for height and weight don’t have any outliers.**
- **The height of people varies between a minimum of 140 cm and a maximum of 199 cm. The mean height is 169 cm. Mean height of female is 170 cm and the mean height of males is 169 cm.**
- **The weight of people varies between a minimum of 50 kg and a maximum of 160 kg. The mean height is 106 kg. Mean height of female is 105 kg and the mean height of males is 106 kg.**

THANK YOU ☺